NMF-Based Spectral Deconvolution with a Web Platform GC Mixture Touch

Yasuyuki Zushi*

ABSTRACT: Complete separation of chemicals in a complex mixture is far from being achieved even with the current high-performance separation technology, such as gas chromatography—mass spectrometry (GC—MS). Several deconvolution techniques based on multivariate curve resolution (MCR), or model peak methods, which are represented by AMDIS, have been developed to address the above-mentioned issue. The model peak methods have been developed to provide easy-to-use tools, including AMDIS, but are limited for MCR with approximation methods. The objective of this study was to provide an easy-to-use deconvolution tool based on the MCR approach for GC—MS data. The spectral deconvolution tool based on non-negative matrix factorization (NMF), which calculates outputs using an approximation method, was implemented as a free web platform, namely, GC Mixture Touch, clarifying the effects of the parameters required for the deconvolution. The GC Mixture Touch was applied to the actual mixture sample of road dust spiked with chemical standards. The recommended parameter settings for smoothing of the chromatogram, the number of ranks, and the NMF algorithm required for the deconvolution were clarified through the study. The performance with the suggested parameters was evaluated with respect to compound identification for the actual sample. All of the test compounds in the sample were correctly identified with the GC Mixture Touch, outperforming AMDIS with respect to the identification. The GC Mixture Touch is easy to use on the web even for users without programming skills. This is expected to enhance the application of the NMF-based deconvolution, and it should prove helpful in finding the compounds hidden in complex mixtures that are difficult to find using conventional approaches.

INTRODUCTION

Thousands or tens of thousands of chemicals are found in the environment or artificial products, known as environmental mixtures or technical mixtures, respectively. These mixtures are analyzed using sophisticated technologies for separation, such as gas chromatography—mass spectrometry (GC—MS), liquid chromatography—mass spectrometry (LC-MS), or LC-diode array detection (DAD), to identify the wide range of the chemicals that they contain. Nevertheless, complete separation of all of the chemicals contained in these mixtures is not achievable in practical cases, such as analysis of oil products, cigarette smoke, or environmental samples. To overcome this limitation of the instruments in performing the chemical separation, techniques of “spectral deconvolution” have been developed. The term “peak deconvolution” is also used in the same sense because peaks are separated simultaneously with spectra separation in the data recorded by a multivariate detector, such as MS or DAD. The deconvolution techniques are roughly classified into the following two categories: model peak methods and multivariate curve resolution (MCR), which is also known as self-modeling curve resolution. The MCR is further divided into the type of unique and rational resolutions. The unique resolution attempts to identify the factors for single species that are uniquely defined according to the mathematical principle. Meanwhile, the rational resolution attempts to yield factors for single species that best approximate the true profiles. MCR-alternative least square (ALS) and iterative target transformation factor analysis (ITTFA), which are categorized under rational resolution, use the least square of the measurement matrix, along with multiplied matrices of concentration and spectral profiles to obtain approximated results of deconvoluted spectrum through iterative calculation. In the historical flow of the development, both MCR-ALS and ITTFA added non-negative constraints that suppress the generation of negative values in the profiles during the iteration process for approximation. The non-negative constraint allows us to obtain realistic outputs in which the spectrum and concentrations basically consist of positive values. Approximation approaches have also been developed in...
environmental analysis as positive matrix factorization (PMF), in acoustic analysis as independent component analysis (ICA), and in image analysis as non-negative matrix factorization (NMF). In particular, NMF is widely used because the process is well generalized and formulated by Lee and Seung. Recently, NMF was applied for deconvolution of spectra and for classification of chromatographic patterns in measurements recorded by state-of-the-art technology of GC × GC-HR-TOFMS. Apart from MCR, model peak methods are implemented in the software AMDIS and ADAP-GC which were developed for analyzing GC–MS data. The model peak method first attempts to find a peak to be deconvoluted which fulfills peak criteria, such as peak height and sharpness for each ion chromatogram. The determined peaks for each ion are combined and reconstructed as deconvoluted peaks accompanied by mass spectra. The AMDIS and ADAP-GC are user-friendly software for deconvoluting GC–MS data whose algorithms based on the model peak method have been disclosed. The software AnalyzerPro and ChromaTOF bundled with the vendor’s instrument allow deconvolution; however, those algorithms are not disclosed. An R package, which is a script-based tool for deconvolution, based on MCR with an approximation method has been developed. Analytical tools with GUI for deconvolution based on the MCR have been developed, but it requires programming shareware MATLAB for operating with the GUIs. A web-based free software that is intuitive and easy to use for most of the analysts are not publicly available to the best of the author’s knowledge. Lack of such a user-friendly tool hampers the wide use of deconvolution based on approximation methods, such as MCR-ALS, ICA, and NMF, making it difficult to clarify their performances from various aspects. This includes the evaluation of cases occurring in practical situations, such as the analysis of environmental mixtures.

The objective of this study was to provide an easy-to-use deconvolution tool based on the MCR approach for GC–MS data. The spectral deconvolution tool based on non-negative matrix factorization (NMF), which calculates outputs using an approximation method, was implemented as a free web platform, namely, GC Mixture Touch, clarifying the effects of the parameters required for the deconvolution. The Mixture Touch is expected to enhance the opportunity to apply the developed deconvolution for various cases existing in our surroundings. The GC Mixture Touch equipped with NMF-based deconvolution was applied to the actual sample of road dust with spiked chemical standards. The performance was evaluated with respect to compound identification of the spiked chemicals in the environmental mixture, comparing that of AMDIS as a benchmark.

**MATERIALS AND METHODS**

Sample Preparation for Performance Evaluation. Road dust that was supplied as a certified reference material for quantitative analysis of elements (BCR-723, Sigma-Aldrich, Saint Louis, MO) was used as an environmental sample for the performance evaluation of NMF-based deconvolution equipped in the GC Mixture Touch. One gram of road dust was sonicated with 10 mL of hexane for 15 min. The hexane phase was collected after centrifugation of the sample. This process was repeated three times. Thereafter, the collected extract was concentrated with a rotary evaporator and subsequently purged with a gentle stream of nitrogen gas. Hexane (100 μL) was added to the sample, and a fraction of 25 μL was supplied for the instrumental analysis as the original road dust sample. A mixture of 12 chemical standards (Programmed Test Mix, Supelco, Bellefonte, PA) was added to another 25 μL hexane fraction. The constituent was analyzed as a chemical-spiked road dust sample.

**Instrumental Analysis.** The sample was analyzed through GC–MS (7890B, Agilent Technologies, Santa Clara, CA, MS; 7000A Triple Quadrupole, Agilent Technologies). The GC column was an Rxi-1ms (30 m × 0.25 mm I.D., 0.25 μm film thickness, Restek, Bellefonte, PA). One microlitre of the sample was injected in splitless mode. Helium gas was used as the carrier gas at a flow rate of 1 mL min⁻¹. The GC oven temperature was ramped from 90 to 300 °C at a rate of 6 °C min⁻¹. The ionization voltage for electron impact was set at 70 eV. The delta electron multiplier voltage was set to 700 V. The scan rate was set at 18 Hz with a range of m/z 14–300.

**GC Mixture Touch.** GC Mixture Touch was developed by programming language R to analyze the GC date on the web. The Mixture Touch series was launched initially from the version for the analysis of GC × GC data. These are developed as free web platforms. Mixture Touch involves data for demonstration; in addition, it allows visualization and analysis of the user’s own data from “DATA LOAD” in the left-hand sidebar of the website. The GC data format should be the common data format (CDF). Other file types can usually be converted to CDF using a function within the vendor’s software associated with the instrument. For a quick trial of the data loading using “LocalFile”, the GC data can be downloaded via the help icon beside the “DATA LOAD”. Each step or function tab is equipped with a help icon that explains the process of the step/function and provides further related references if available. The GC Mixture Touch allows visualization of a mass spectrum by pointing to any position on a chromatographic curve. When “pixel” as the pointing type is selected, a mass spectrum with its retention time of clicked point in a chromatogram is depicted. Meanwhile, “peaktop” as the pointing type, which allows the recognition of peak features from the extremal point of the chromatographic curve, visualizes a mass spectrum of the nearest peaktop from the clicked point. When a point of the chromatogram is selected with “peaktop” mode, NMF-based deconvolution is automatically performed on the selected peak. The details of the deconvolution are described in the subsequent section. The function of spectral search on the selected chromatogram including the deconvoluted peak using a spectral database is implemented in the GC Mixture Touch. The spectral similarity was calculated based on cosine similarity, which is a normalized dot product of two vectors of a spectrum.

**NMF-Based Deconvolution in GC Mixture Touch.** The NMF theory and computational method are found elsewhere. Briefly, in NMF, the basis matrix W and the coefficient matrix H for the original matrix Y were calculated such that the sum of the loss function D(Y, WH) and the regularization function R(W, H) was minimized, as expressed in eqs 1 and 2 as follows

\[ Y \approx WH \]  
\[ \min_{W,H \geq 0} [D(Y, WH) + R(W, H)] \]
The matrix elements of $Y$ and $WH$ were iteratively calculated and updated using non-negative constraints such that the metric of $D$ decreased. The loss function $D$ measures the quality of the approximation in eq 1. The regularization function $R$ is an optional function in NMF, defined to enforce the properties of $W$ and $H$, such as smoothness or sparsity.

$D$ is calculated by the Frobenius distance, as shown in eq 3, known as the Euclidean distance, or Kullback–Leibler (KL) divergence, as expressed in eq 4 as follows:

Frobenius distance:

$$D: Y, WH \rightarrow \frac{Tr(Y(WH)^T)}{2} = \frac{1}{2} \sum_{kn} (y_{kn} - x_{kn})^2$$

(3)

KL divergence:

$$D: Y, WH \rightarrow KL(Y\|WH) = \sum_{kn} y_{kn} \log \frac{y_{kn}}{x_{kn}} - y_{kn} + x_{kn}$$

(4)

where $x$ and $y$ are elements of matrices $X$ and $Y$, respectively. The matrix elements of $Y$ and $WH$ are used for iterative calculation, and $W$ and $H$ are updated with non-negative constraints such that the distance (divergence) is decreased according to the following eqs 5–8:

NMF algorithm in Frobenius distance:

$$H_{kn} \leftarrow H_{kn} \frac{\sum_n Y_{kn} W_{mn}}{\sum_n (WH)_{kn} W_{mn}}$$

(5)

$$W_{mn} \leftarrow W_{mn} \frac{\sum_k Y_{kn} H_{km}}{\sum_k (WH)_{km} H_{km}}$$

(6)

NMF algorithm in KL divergence:

$$H_{kn} \leftarrow H_{kn} \frac{\sum_n (WH)_{kn}}{\sum_n Y_{kn} W_{mn}}$$

(7)

$$W_{mn} \leftarrow W_{mn} \frac{\sum_k H_{km} (WH)_{kn}}{\sum_k H_{km} Y_{kn}}$$

(8)

In the first step of NMF, the $W$ and $H$ matrices are initialized using seeding methods. One of them is the random seeding method in which a seed is drawn from a uniform distribution. This seeding method sometimes experiences difficulty in converging the calculation. The non-negative double singular value decomposition seeding method (NNDSVD), which completes the iterative process more quickly than other seeding methods, is mainly used. The ICA, which uses only the positive part of the result, is also available as a seeding method.

The measured spectra were used as the original matrix $Y$, and then the matrix was decomposed by the NMF. The normalized spectral pattern of each rank is produced as the basis matrix $W$, and the intensity of the spectrum in each rank of each data point for reconstructing the original spectra is produced as the coefficient matrix $H$.

**RESULTS AND DISCUSSION**

Test Compounds for NMF-Based Deconvolution. A road dust sample with chemical standards was supplied for the performance evaluation of NMF-based deconvolution (Figure 1). The peaks of the four compounds of all of the spiked chemical standards were outside the range for the instrumental setting applied in this study. Three of the remaining eight compounds were aliphatics, which have several structurally similar compounds that eventually exhibit similar mass spectral patterns. These types of compounds were not advisable for the evaluation based on spectral library search. Therefore, these were also eliminated from the evaluation. Finally, the five test compounds in the chemical standards were focused for the evaluation. An overview of the compounds, including retention time and peak height in the chromatogram, is presented in Table 1.

**Overview of NMF Condition.** GC Mixture Touch equipped with NMF provides two different NMF algorithms (Frobenius or KL) and three different matrix initializations (NNDSVD, random, or ICA). Setting of the number of NMF ranks that indicates the number of assumed overlapping peaks is discussed in the subsequent section. All six combinations were performed and evaluated for the five test compounds with and without chromatographic smoothing (Table S1). Briefly, mass spectra of all of the test compounds were successfully extracted as ranks 1 or 2 spectra from chromatographic peaks in the chemically spiked road dust sample with any of the conditions, regardless of the smoothing condition. The difference between the algorithms did not have a remarkable influence on the results of compound identification after deconvolution. For matrix initialization, random and ICA methods relatively frequently failed to extract the correct mass spectra. Both methods are based on random seeding that provides different results in each calculation. For the different results obtained, there are possibilities of inadequate convergence within the set number of iterations and NMF run of 2000 and 1 in default setting, respectively. The quality of the data for calculation would also affect achieving the convergence. A large number of iterations and multiple NMF run, that is, high cost of computation time, would be required for the convergence.
depending on the data quality. The initialization method of NNDSVD provided superior performance in spectral extraction with the default setting; therefore, NNDSVD was selected for further practice. Chromatographic smoothing expands the width of peaks compared to that of the peaks in original chromatogram, because the chromatographic intensities are averaged out. Therefore, chromatographic smoothing was expected to have a positive effect on spectral extraction owing to the enhancement of the number of data points in the original chromatogram in the NMF calculation. However, no clear effects were observed. Meanwhile, the smoothing provided positive effects on obtaining the desirable shape of the chromatographic peak. Details of the effects of smoothing are discussed in the subsequent section. The coeluted original compounds were successfully separated from the spiked chemical by the NMF and identified through library searches for the peaks of three (2-ethylhexanoic acid, 2,6-dimethylaniline, and methyl dodecanoate) out of the five test compounds. Mass spectral patterns of the chromatographic regions for the remaining two peaks were originally ambiguous and difficult to identify through the spectral library search. **Effects of Smoothing and NMF Algorithm.** The smoothing of the chromatographic curve based on the moving-average method was performed to explore the advisable smoothing level, such as the n-day moving average (Table S2). According to the level of smoothing up to 50, window sizes, that is, time periods of chromatograms, for deconvolution were approximately 3−30 times increased depending on the characteristics of the chromatograms. However, a clear increasing trend of the match factors was not observed through smoothing. Meanwhile, the desirable shapes of the deconvoluted peaks were obtained through smoothing with a level higher than or equal to 5 (Figure S1). This is because a chromatographic peak with a higher level of noise signal tends to be partially out of range of the peak picking algorithm, which uses an extremal point for peak recognition, when the smoothing process is not applied (Figure S1a,b). Regarding the differences of the NMF algorithm, the KL algorithm

| retention time (min) | n-nonanal | 2,6-dimethylphenol | 2-ethylhexanoic acid | 2,6-dimethylaniline | methyl dodecanoate |
|---------------------|-----------|--------------------|---------------------|-------------------|-------------------|
| concentration in the chemical-spiked road dust sample (μg mL⁻¹) | 4.24 | 4.77 | 4.91 | 5.55 | 12.13 |
| I. Peak height in the road dust sampleb | 7.1 × 10⁶ | 1.7 × 10⁷ | 2.4 × 10⁴ | 3.7 × 10⁰ | 5.6 × 10⁰ |
| II. Peak height in the chemical-spiked road dust sample | 6.4 × 10⁶ | 2.3 × 10⁷ | 9.6 × 10⁵ | 1.1 × 10⁷ | 3.8 × 10⁷ |
| ratio (II/I) | 0.90 | 1.32 | 4.04 | 3.06 | 6.80 |
| remarks | with overlapped peak | with overlapped peak | no remarkable overlapped peak | no remarkable overlapped peak | no remarkable overlapped peak |

The intensities of the chromatographic peaks were calculated based on the total ion chromatogram (Figure 1).bPeaks eluted at the same positions as the respective chemicals were alternatively used for the intensity calculation.

![Figure 2. Result of NMF-based deconvolution (number of rank: 5, algorithm: Frobenius, and initialization: NNDSVD) of the chromatographic peaks including 2,6-dimethylphenol (12.8 μg mL⁻¹).](https://dx.doi.org/10.1021/acsomega.0c04982)

ACS Omega 2021, 6, 2742−2748
provided a smoother curve than the Frobenius algorithm (Figure S1). The evaluation function in the KL algorithm conceptually consists of a logarithmic transformation of the Frobenius algorithm. This form is possibly advantageous for convergence of the values to smaller fluctuations. Meanwhile, the differences in the NMF algorithm for each smoothing level were not observed in the spectral pattern, as mentioned above. **Effects of the Number of Ranks.** The number of ranks is always controversial when factorization is used. However, the number of ranks, in other words, the number of overlapping compounds, would be within a limited range; therefore, it is possible to narrow the hypothesis in the studied case. Generally, a chromatographic peak obtained by GC−MS contains at most a few compounds because of its high separation performance; therefore, it is not necessary to set the rank to a higher value. As shown in Table S3, the results of the spectral search were not clearly changed when the rank number was set to higher than or equal to 5. Although the optimal value would be slightly different depending on the setting of each system, the rank number would not strongly influence the deconvolution of the GC−MS peaks in this case, as shown. Therefore, rank number 5 was applied for further practice in this study. Nevertheless, in the case of a highly complex mixture, which would have a larger number of overlapping compounds, a higher number of rank would be required. **Deconvolution with GC Mixture Touch in Practice.** Details of NMF-based deconvolution in the GC Mixture Touch are shown through the practice of 2,6-dimethylphenol (Figure 2). The following parameters were selected for the deconvolution: the level of smoothing to be S, the number of ranks to be 5, NMF algorithm to be Frobenius, and initialization method to be NNDSVD. The total ion chromatographic peaks that include 2,6-dimethylphenol showed a bimodal peak, as shown in Figure S2. The compound eluted at a very close region with the peak assigned as 2-tridecanol through a spectral search in the original road dust sample. The peak height of the test compound was 3 times lower than that of the original compound on the extracted ion basis. In this case, the height of the TIC peak in the eluting period increased by 30% through spiking the test compound (12.8 μg mL−1 as a final concentration). The deconvoluted peak of rank 2 was successfully assigned as the test compound, whereas that of rank 1 was successfully assigned as the originally identified compound (Figure S3). The peak of 10 times diluted test compound (Figure S4) was not successfully identified through the deconvolution owing to the obtainment of the fair peak through the measurement (Figure S5). In this study, the test compound peaks were able to preferably separate from overlapping peaks whose intensities were 3 times higher than the test compound; however, this was not possible for the case with further differences. Several situations, such as the relative intensity of overlapping peaks, the number of neighboring peaks, and their spectral patterns, might affect the success of deconvolution. To completely clarify these issues, further intensive exploration that includes a quantitative aspect is advisable. **Comparisons between GC Mixture Touch and AMDIS.** Deconvolutions for the remaining four test compounds were performed using a GC Mixture Touch with the same settings, except for the level of smoothing. For the purpose of performance comparisons between the GC Mixture Touch and AMDIS that do not equip the same smoothing function, smoothing was not performed in either software. Only one deconvoluted peak was obtained using AMDIS for each test compound, although the GC Mixture Touch provided deconvoluted peaks with the numbers depending on the rank number setting (Table 2). AMDIS failed to identify the three compounds whose peak ratios of peak heights were the lowest among the five test compounds. Meanwhile, the GC Mixture Touch successfully identified all five test compounds. Furthermore, the GC Mixture Touch provides deconvoluted spectra other than the first rank spectrum, for which the features are generally not available in AMDIS. These findings show that NMF-based deconvolution equipped in the GC Mixture Touch achieves a high performance in extracting peaks with lower concentrations compared with the case using AMDIS, in addition to the ability to provide information on the other overlapping peaks.

### CONCLUSIONS

GC Mixture Touch equipped with NMF-based deconvolution was developed and applied to the environmental sample of road dust. The recommended parameter settings for the deconvolution were clarified through the study as follows: the extent of smoothing of the chromatogram as higher than S, the number of ranks as 5, and the NMF algorithm as a Frobenius algorithm for faster calculation, but a KL algorithm for a smoother curve in the chromatogram. All of the test compounds were correctly identified using GC Mixture Touch, outperforming AMDIS with respect to compound identification. GC Mixture Touch is provided a smoother curve than the Frobenius algorithm (Figure S1). The evaluation function in the KL algorithm conceptually consists of a logarithmic transformation of the Frobenius algorithm. This form is possibly advantageous for convergence of the values to smaller fluctuations. Meanwhile, the differences in the NMF algorithm for each smoothing level were not observed in the spectral pattern, as mentioned above. **Effects of the Number of Ranks.** The number of ranks is always controversial when factorization is used. However, the number of ranks, in other words, the number of overlapping compounds, would be within a limited range; therefore, it is possible to narrow the hypothesis in the studied case. Generally, a chromatographic peak obtained by GC−MS contains at most a few compounds because of its high separation performance; therefore, it is not necessary to set the rank to a higher value. As shown in Table S3, the results of the spectral search were not clearly changed when the rank number was set to higher than or equal to 5. Although the optimal value would be slightly different depending on the setting of each system, the rank number would not strongly influence the deconvolution of the GC−MS peaks in this case, as shown. Therefore, rank number 5 was applied for further practice in this study. Nevertheless, in the case of a highly complex mixture, which would have a larger number of overlapping compounds, a higher number of rank would be required. **Deconvolution with GC Mixture Touch in Practice.** Details of NMF-based deconvolution in the GC Mixture Touch are shown through the practice of 2,6-dimethylphenol (Figure 2). The following parameters were selected for the deconvolution: the level of smoothing to be S, the number of ranks to be 5, NMF algorithm to be Frobenius, and initialization method to be NNDSVD. The total ion chromatographic peaks that include 2,6-dimethylphenol showed a bimodal peak, as shown in Figure S2. The compound eluted at a very close region with the peak assigned as 2-tridecanol through a spectral search in the original road dust sample. The peak height of the test compound was 3 times lower than that of the original compound on the extracted ion basis. In this case, the height of the TIC peak in the eluting period increased by 30% through spiking the test compound (12.8 μg mL−1 as a final concentration). The deconvoluted peak of rank 2 was successfully assigned as the test compound, whereas that of rank 1 was successfully assigned as the originally identified compound (Figure S3). The peak of 10 times diluted test compound (Figure S4) was not successfully identified through the deconvolution owing to the obtainment of the fair peak through the measurement (Figure S5). In this study, the test compound peaks were able to preferably separate from overlapping peaks whose intensities were 3 times higher than the test compound; however, this was not possible for the case with further differences. Several situations, such as the relative intensity of overlapping peaks, the number of neighboring peaks, and their spectral patterns, might affect the success of deconvolution. To completely clarify these issues, further intensive exploration that includes a quantitative aspect is advisable. **Comparisons between GC Mixture Touch and AMDIS.** Deconvolutions for the remaining four test compounds were performed using a GC Mixture Touch with the same settings, except for the level of smoothing. For the purpose of performance comparisons between the GC Mixture Touch and AMDIS that do not equip the same smoothing function, smoothing was not performed in either software. Only one deconvoluted peak was obtained using AMDIS for each test compound, although the GC Mixture Touch provided deconvoluted peaks with the numbers depending on the rank number setting (Table 2). AMDIS failed to identify the three compounds whose peak ratios of peak heights were the lowest among the five test compounds. Meanwhile, the GC Mixture Touch successfully identified all five test compounds. Furthermore, the GC Mixture Touch provides deconvoluted spectra other than the first rank spectrum, for which the features are generally not available in AMDIS. These findings show that NMF-based deconvolution equipped in the GC Mixture Touch achieves a high performance in extracting peaks with lower concentrations compared with the case using AMDIS, in addition to the ability to provide information on the other overlapping peaks.

### CONCLUSIONS

GC Mixture Touch equipped with NMF-based deconvolution was developed and applied to the environmental sample of road dust. The recommended parameter settings for the deconvolution were clarified through the study as follows: the extent of smoothing of the chromatogram as higher than S, the number of ranks as 5, and the NMF algorithm as a Frobenius algorithm for faster calculation, but a KL algorithm for a smoother curve in the chromatogram. All of the test compounds were correctly identified using GC Mixture Touch, outperforming AMDIS with respect to compound identification. GC Mixture Touch is
easy to use on the web even for users without programming skills. This is expected to enhance the application of NMF-based deconvolution, and it should prove helpful for finding the compounds hidden in complex mixtures, such as environmental contaminants, cigarette smoke, biomarkers, foods, and intermediate products of materials including drugs, that are difficult to find using conventional approaches. In future, the wider use of the software, which would include analysis of data by various GC–MS systems for various sample compositions, is expected to clarify its performance and challenges in the next stage from various aspects.

**ASSOCIATED CONTENT**

*Supporting Information*

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c04982.

Results of NMF-based deconvolution with various NMF algorithms and initialization methods for chemicals spiked in the road dust sample (Table S1); influence of the window size for NMF-based deconvolution (Table S2); deconvoluted peaks using Frobenius or KL algorithms with several levels of smoothing (Figure S1); influence of the number of factors on NMF-based deconvolution (Table S3); chromatographic peaks including 2,6-dimethylphenol (12.8 µg/mL) in the road dust (Figure S2); spectrum search of original and deconvoluted spectra of the chromatographic peaks including 2,6-dimethylphenol (12.8 µg/mL) (Figure S3); chromatographic peaks including 2,6-dimethylphenol (1.28 µg/mL) in the road dust sample (Figure S4); and result of NMF-based deconvolution of the chromatographic peaks including 2,6-dimethylphenol (1.28 µg/mL) (Figure S5) (PDF)

**AUTHOR INFORMATION**

*Corresponding Author*

Yasuuki Zushi — Research Institute of Science and Sustainability, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaragi 305-8506, Japan; orcid.org/0000-0001-8062-1592; Phone: +81-29-861-2970; Email: zushi.yasuuki@aist.go.jp

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acsomega.0c04982

**Notes**

The author declares no competing financial interest.

GC Mixture Touch is freely available on the web at http://www.mixture-platform.net/GC_Mixture_Touch.

**ACKNOWLEDGMENTS**

This study was supported by Scientific Research (A) (grant no. 17H00796) and Scientific Research (B) (grant no. 19H04297). The author thanks Dr. Yuning Qiu and Dr. Irwin Kurland (Stable Isotope and Metabolomics Core Facility, Albert Einstein College of Medicine) for offering GC–MS data for demonstration of the GC Mixture Touch and for their valuable comments on it.

**REFERENCES**

(1) Frysinger, G. S.; Gaines, R. B. Separation and identification of petroleum biomarkers by comprehensive two-dimensional gas chromatography. *J. Sep. Sci.* 2001, 24, 87–96.

(2) Gröger, T.; Welthagen, W.; Mitschke, S.; Schäffer, M.; Zimmermann, R. Application of comprehensive two-dimensional gas chromatography mass spectrometry and different types of data analysis for the investigation of cigarette particulate matter. *J. Sep. Sci.* 2008, 31, 3366–3374.

(3) Zushi, Y.; Hashimoto, S.; Tanabe, K. Nontarget approach for environmental monitoring by GC×GC-HRTOFMS in the Tokyo Bay basin. *Chemosphere* 2016, 156, 398–406.

(4) Wallace, R. M. ANALYSIS OF ABSORPTION SPECTRA OF MULTICOMPONENT SYSTEMS1. *J. Phys. Chem. A* 1960, 64, 899–901.

(5) Maeder, M. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.* 1987, 59, 527–530.

(6) Jiang, J.-H.; Liang, Y.; Ozaki, Y. Principles and methodologies in self-modeling curve resolution. *Chemom. Intell. Lab. Syst.* 2004, 71, 1–12.

(7) Stein, S. E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.* 1999, 10, 770–781.

(8) Colby, B. N. Spectral deconvolution for overlapping GC/MS components. *J. Am. Soc. Mass Spectrom.* 1992, 3, 558–562.

(9) Wei, X.; Shi, X.; Kim, S.; Patrick, J. S.; Binkley, J.; Kong, M.; McClain, C.; Zhang, X. Data Dependent Peak Model Based Spectrum Deconvolution for Analysis of High Resolution LC-MS Data. *Anal. Chem.* 2014, 86, 2156–2165.

(10) Dromey, R. G.; Stelfk, M. J.; Rindfleisch, T. C.; Duffield, A. M. Extraction of mass spectra free of background and neighboring component contributions from gas chromatography/mass spectrometry data. *Anal. Chem.* 1976, 48, 1368–1375.

(11) Smirnov, A.; Jia, W.; Walker, D. L.; Jones, D. P.; Du, X. ADAP-GC 3.2: Graphical Software Tool for Efficient Spectral Deconvolution of Gas Chromatography—High Resolution Mass Spectrometry Metabolomics Data. *J. Proteome Res.* 2018, 17, 470–478.

(12) Tauler, R.; Casassas, E. Spectroscopic resolution of macro-molecular complexes using factor analysis: Cu(II)–polycyline system. *Chemom. Intell. Lab. Syst.* 1992, 14, 305–317.

(13) Toft, J. Evolutionary rank analysis applied to multidetectorial chromatographic structures. *Chemom. Intell. Lab. Syst.* 1995, 29, 189–212.

(14) Smirnov, A.; Qiu, Y.; Jia, W.; Walker, D. L.; Jones, D. P.; Du, X. ADAP-GC 4.0: Application of Clustering-Assisted Multivariate Curve Resolution to Spectral Deconvolution of Gas Chromatography—Mass Spectrometry Metabolomics Data. *Anal. Chem.* 2019, 91, 9069–9077.

(15) Gomperline, P. J. Prior estimates of the elution profiles of the pure components in overlapped liquid chromatography peaks using target factor analysis. *J. Chem. Inf. Model.* 1984, 24, 206–212.

(16) Vandeginste, B. G. M.; Derks, W.; Kateman, G. Multi-component self-modelling curve resolution in high-performance liquid chromatography by iterative target transformation analysis. *Anal. Chim. Acta* 1985, 173, 253–264.

(17) Lacey, R. F. Deconvolution of overlapping chromatographic peaks. *Anal. Chem.* 1986, 58, 1440–1441.

(18) van Stokkum, I. H. M.; Muller, K. M.; Mihaleva, V. V. Global analysis of multiple gas chromatography–mass spectrometry (GC/MS) data sets: A method for resolution of co-eluting components with comparison to MCR-ALS. *Chemom. Intell. Lab. Syst.* 2009, 95, 150–163.

(19) Jellem, R. H.; Krishnan, S.; Hendriks, M. M. W. B.; Mulvijik, B.; Vogels, J. T. W. E. Deconvolution using signal segmentation. *Chemom. Intell. Lab. Syst.* 2010, 104, 132–139.

(20) Domingo-Almenara, X.; Brezmes, J.; Vinaixa, M.; Samino, S.; Ramirez, N.; Ramon-Krauel, M.; Lerin, C.; Diaz, M.; Ibáñez, L.; Correig, X.; Perera-Lluna, A.; Yanes, O. eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* 2016, 88, 9821–9829.

(21) Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 1994, 5, 111–126.
(22) Kothandaraman, M.; Pachaiyappan, A. Comparison of Independent Component Analysis techniques for Acoustic Echo Cancellation during Double Talk senario. Aust. J. Basic Appl. Sci. 2013, 7, 108−113.

(23) Hyvärinen, A. Independent component analysis: recent advances. Philos. Trans. R. Soc., A 2013, 371, No. 20110534.

(24) Lee, D. D.; Seung, H. S. Learning the parts of objects by non-negative matrix factorization. Nature 1999, 401, 788−791.

(25) Gaujoux, R.; Seoighe, C. A flexible R package for nonnegative matrix factorization. BMC Bioinf. 2010, 11, 367.

(26) Gao, H.-T.; Li, T.-H.; Chen, K.; Li, W.-G.; Bi, X. Overlapping spectra resolution using non-negative matrix factorization. Talanta 2005, 66, 65−73.

(27) Sun, J.; Li, T.; Cong, P.; Xiong, W.; Tang, S.; Zhu, L. Direct decomposition of three-way arrays using a non-negative approximation. Talanta 2010, 83, 541−548.

(28) Kameoka, H.; Nobutaka, O.; Kunio, K.; Shigeki, S. In Complex NMF: A New Sparse Representation for Acoustic Signals, International Conference on Acoustics, Speech, and Signal Processing, 19−24 April, 2009; pp 3437−3440.

(29) Zushi, Y.; Hashimoto, S.; Tanabe, K. Global Spectral Deconvolution Based on Non-Negative Matrix Factorization in GC×GC−HRTOFMS. Anal. Chem. 2015, 87, 1829−1838.

(30) Zushi, Y.; Hashimoto, S. Direct Classification of GC×GC-Analyzed Complex Mixtures Using Non-Negative Matrix Factorization-Based Feature Extraction. Anal. Chem. 2018, 90, 3819−3825.

(31) Jaumot, J.; Tauler, R. MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution. Chemom. Intell. Lab. Syst. 2010, 103, 96−107.

(32) Jaumot, J.; de Juan, A.; Tauler, R. MCR-ALS GUI 2.0: New features and applications. Chemom. Intell. Lab. Syst. 2015, 140, 1−12.

(33) Zhang, Y.-Y.; Zhang, Q.; Zhang, Y.-M.; Wang, W.-W.; Zhang, L.; Yu, Y.-J.; Bai, C.-C.; Guo, J.-Z.; Fu, H.-Y.; She, Y. A comprehensive automatic data analysis strategy for gas chromatography-mass spectrometry based untargeted metabolomics. J. Chromatogr. A 2020, 1616, No. 460787.

(34) Zushi, Y.; Hanari, N.; Nabi, D.; Lin, B.-L. Mixture Touch: A Web Platform for the Evaluation of Complex Chemical Mixtures. ACS Omega 2020, 5, 8121−8126.

(35) MS-DIAL, FAQ. http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/index3.html (accessed Nov 26, 2020).

(36) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat. Methods 2015, 12, 523.

**NOTE ADDED AFTER ASAP PUBLICATION**

This paper was published ASAP on January 19, 2021, with errors in the captions of Figures 1 and 2. The corrected version was reposted on January 20, 2021.