analytical chemistry

Article

pubs.acs.org/ac

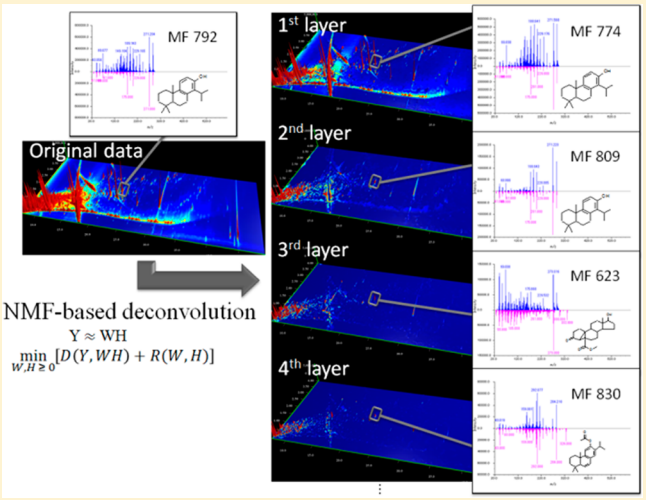
Global Spectral Deconvolution Based on Non-Negative Matrix Factorization in GC × GC−HRTOFMS

Yasuyuki Zushi,* Shunji Hashimoto, and Kiyoshi Tanabe

Center for Environmental Measurement and Analysis, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan

Supporting Information

ABSTRACT: A global spectral deconvolution, based on non-negative matrix factorization (NMF) in comprehensive two-dimensional gas chromatography high-resolution time-of-flight mass spectrometry, was developed. We evaluated the ability of various instrumental parameters and NMF settings to derive high-performance detection in nontarget screening using a sediment sample. To evaluate the performance of the process, a NIST library search was used to identify the deconvoluted spectra. Differences of the instrumental scan rates (25 and 50 Hz) in deconvolution were evaluated and results show that a high scan rate enhanced the number of compounds detected in the sediment sample. A higher mass resolution in the range of 1 000 to 10 000 and a higher m/z precision in the deconvolution were needed to obtain an accurate mass database. After removal of multiple duplicate hits, which occurred in batch processes of NIST library search on the deconvolution result, 62 unique assignable spectra with a match factor ≥900 were obtained in the deconvoluted chromatogram from the sediment sample, including 54 spectra that were refined by the deconvolution. This method will help to detect and build up well-resolved reference spectra from various complex mixtures and will accelerate nontarget screening.



The identification and quantification of substances in complex mixtures have been important issues in forensics, food and flavor analysis, development and chemical management of industrial and medical products, and environmental research for many years. A recently developed, powerful analytical instrument, the comprehensive two-dimensional gas chromatograph−high-resolution time-of-flight mass spectrometer (GC × GC−HRTOFMS), has great potential to separate complex mixtures into their two capillary columns with different polarities and detect separated components with a high mass resolution.[1] This advanced analytical technique is also able to quantify each compound[2] and to characterize chromatographic peaks;[3,4] therefore, this technology has the potential to be used widely in various analytical fields. However, the instrument detects all ions within a certain mass range in the matrix-enriched samples during full scan mode analysis, resulting in unresolved peaks from noise and coeluted components, even in the two-dimensional chromatogram.[5,6] Extracting specific compounds from the total ion chromatogram based on target accurate mass records allows us to achieve target screening, even in unresolved chromatograms.[7−9] However, this process is not similarly beneficial in nontarget screening that tries to identify unknown chromatographic peaks. This is because the mass spectra that overlap because of noise or coeluted components inhibit the identification of the original compounds from such disturbed data acquired in the nontarget screening. Furthermore, the situation will be more problematic for automating the data screening in samples of complex mixtures.

One way to solve this drawback in nontarget screening is to apply signal deconvolution methods. To date, several studies have used deconvolution based on multivariate analysis for GC−MS,[10] and LC−diode array detection.[11,12] Several peak model-based deconvolution methods have also been developed.[13,14] There are many types of deconvolution, such as intensity threshold cutoff signals, baseline correction for noise subtraction, and resolution of overlapping peaks. In this article, deconvolution is defined as a method for resolving overlapping peaks that reconstructs several separate components from signals of mixed components, including noise.

A tensor factorization method called parallel factor analysis (PARAFAC),[15] which solves three-way arrays of acquired data (i.e., mass spectra for each data point of a chromatographic line

Received: October 15, 2014
Accepted: January 8, 2015
Published: January 8, 2015

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

with different retention in GC1) by alternating least-squares, has been applied in GC × GC−MS to obtain resolved mass spectra.[16] One of the drawbacks of the tensor factorization is that it requires consistent variation in the focus component in different arrays. Retention time (RT) shifts of each peak in GC × GC, which are caused by correlation of the oven temperature control by a single oven, affect the spectral deconvolution results. This issue can be avoided by being very careful with the RT shift by PARAFAC2.[17] Another drawback of the tensor factorization PARAFAC is that the outputs consist of a mix of positive and negative values as a result of a simple iterative calculation process. This requires an additional step, such as simple baseline correction subtraction, to avoid negative values.[16] Spectrum bias by subtraction is a concern for small peaks. There are several methods for general factorization (but not tensor factorization), such as non-negative matrix factorization (NMF)[18] and independent component analyses (ICAs),[19] that apply non-negative constraints in their iterative calculation processes (not all of ICA methods include the non-negative constraints). ICA was mostly developed in acoustics to separate sources of sound.[20,21] As ICA has developed, several functions including non-negative constraint have been added to ICA, resulting in the diffusion of several software applications or program codes to perform ICA, such as, among others, fast ICA algorithms, JADE algorithms, and extended Infomax algorithms.[22,23] NMF has been developed in image analysis to extract common features in pictures, such as eyes, nose, and ears in human faces.[18,24] NMF is built on the intuitive concept that signals in the real world are basically the addition of positive values. Therefore, it is becoming more popular and feasible to realistically separate overlapping signals in various study fields.[25−27] The release of the NMF application for the free statistical software R[28] has also increased its popularity.

To date, there have been various studies that apply NMF to GC−MS data.[29,30] Several of them are to do with the development of tensor factorization with non-negative constraints, so-called three-dimensional non-negative matrix approximation.[30,31] This has not been applied to GC × GC−MS data such as the study applying PARAFAC,[15] but it has been applied to several already-analyzed GC−MS data sets to construct three-way arrays (MS spectra, RTs, analytical samples).[31] These studies focused on just one or a few intense mixed signals to introduce the deconvolution methods; therefore, their performance in various patterns of signal mixing is still unclear.

The automation of deconvolution so that it can be applied to an entire chromatogram, namely, global spectral deconvolution method, also remains a challenge. The Automated Mass Spectral Deconvolution and Identification System (AMDIS), which was developed for deconvolution of GC−MS mass spectra, identifies the shape of each peak in mixed signals by individually calculating the derivation of each recorded ion.[13] AMDIS includes a function to select chromatogram peaks and then automatically deconvolutes all the selected peaks (i.e., peak model-based deconvolution). The peak selection method and approach for calculating the derivation are not directly applicable to two-dimensional peaks. Hoggard et al. challenged the application of automated peak deconvolution in GC × GC−MS using PARAFAC.[32] In their study, peak selection was not applied, but regular rectangular divisions of the entire chromatogram were used. The use of regular rectangles includes several resolved peaks. Thus, the issue of needing a calculation process to decide the appropriate number of factors (the number of compounds in the region) remains in the application and results in increased calculation cost. They also discussed that the RT shift needs to be modified.

The development of high-resolution mass spectrometry (HRMS) reinforces the detection power of analytes both in target and nontarget analysis because of their high-resolution mass spectrum analysis. However, to date, the deconvolution method based on multivariate analysis for HRMS has not been studied. One of the reasons for this is that HRMS has been developed recently; therefore, the HRMS peak deconvolution technique is not yet well-known. Another reason is that the accurate mass obtained by HRMS will dramatically increase the size of the measured data and factorization of the data matrix. It leads to an unrealistic calculation burden for the deconvolution of HRMS data in older laptop or notebook computers. Therefore, there have been no studies on these methods for HRMS data.

In this study, we developed a spectral deconvolution method based on NMF for two-dimensional chromatograms of accurate masses (i.e., GC × GC−HRTOFMS). We developed an automated process for the deconvolution of the entire chromatogram (global spectral deconvolution) in free software and evaluated its performance in various instrumental mass resolutions, ion scan rates, and also in various NMF calculation scenarios. The research goal was to provide a practical solution for performing NMF-based spectral deconvolutions in real-life samples of complex mixtures by GC × GC−HRTOFMS and demonstrate their ability to enhance the qualitative detection of nontarget analytes. This procedure is expected to enhance the performance of nontarget screening. In addition, it will be useful to store the accurate mass spectra from actual samples of complex mixtures, as it will enhance postscreening of the target compound in a later process.

METHODS

Experimental Data. A 10 mL aliquot of extract in toluene from sea sediment (5 g dry weight), collected in a small Japanese harbor on the Pacific Ocean coast, was analyzed using a 7890GC (Agilent Technologies, Santa Clara, CA) with a

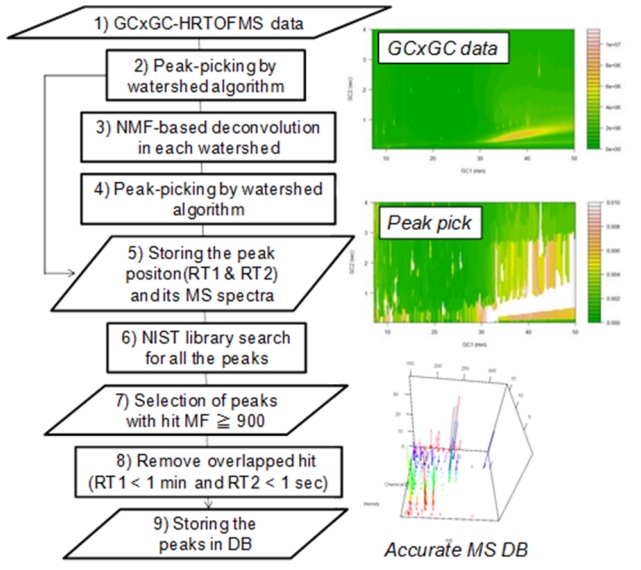


Figure 1. Workflow of NMF-based global spectral deconvolution and spectral storage.

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

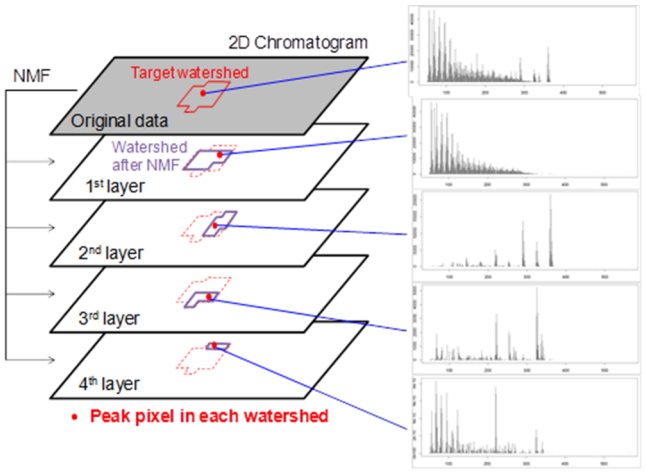


Figure 2. Schematic of layer-style output of NMF-based deconvolution. Each layer format corresponds to the original GC × GC 2D chromatogram with accurate mass spectra in each data point (pixel). The peak top of each generated peak within the original target watershed is picked as a potential list of compound spectra, forwarding the following steps of the workflow shown in Figure 1.

KT2006 GC × GC system (Zoex, Houston, TX) coupled with a JMS-T100GCV 4G-N HRTOFMS (JEOL, Tokyo, Japan). The first GC column was an InertCap 5MS/Sil (45 m × 0.25 mm i.d., 0.1 μm film thickness, GL Sciences, Tokyo, Japan), and the second GC column was a SGE BPX-50 (1 m × 0.1 mm i.d., 0.1 μm film thickness, Sigma-Aldrich (St. Louis, MO). The injection volume was 1 μL in splitless mode at 246 kPa (70 °C), with He as the carrier gas. The GC oven temperature was held at 70 °C for 1 min, ramped to 180 °C at a rate of 50 °C min^{-1}, then ramped to 230 °C at 3 °C min^{-1}, then ramped again to 300 °C at 5 °C min^{-1}, and then was held at 300 °C for 16.133 min; the total time for this process was 50 min. The modulation period during the analysis was 4 s (releasing, 0.25 s) to split peaks in the first dimension of the GC. The ionization voltage for electron impact (EI) was set at 45 eV for the HRTOFMS detector. The ionization voltage generally does not change the fragment patterns from the typical setting of 70 eV.[33] The voltage of the microchannel plate detector was set at 2000 V. To evaluate the effect of changing the parameters on the spectral deconvolution, the sample was analyzed six times using a mass range between 35 and 600 m/z, different mass resolutions of 1 000, 5 000, and 10 000 by manual instrumental parameter tuning, and different scan rates of 25 and 50 Hz. Measurement data of a matrix-rich sample of air-dried lake sediment CRM from Lake Ontario (WMS-01, Wellington Laboratories) were processed at a mass resolution of 5 000 and a scan rate of 25 Hz to allow partial evaluation of the spectral deconvolution. The details of the analytical conditions are found elsewhere.[8]

Theory of NMF. The NMF estimates the basis matrix W ($n \times r$ non-negative matrices) and the coefficient matrix H ($r \times p$ non-negative matrices) for the original matrix Y ($n \times p$ non-negative matrix) so that the sum of the loss function D (Y, WH) and the regularization function R (W, H) are minimized.[28] For practical purposes, the factorization rank r is $r \ll \min(n, p)$.

$$Y \approx WH \tag{1}$$

$$\min_{W,H\geq 0}[D(Y, WH) + R(W, H)] \tag{2}$$

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.

The value of D is calculated by the Frobenius distance, generally known as the Euclidean distance or Kullback–Leibler (KL) divergence.

Frobenius distance:

$$\text{D: Y, WH} \rightarrow \frac{Tr(Y(WH)^t)}{2} = \frac{1}{2}\sum_{kn}(y_{kn} - x_{kn})^2 \tag{3}$$

KL divergence:

$$\text{D: Y, WH} \rightarrow KL(Y||WH) = \sum_{kn} y_{kn} \log \frac{y_{kn}}{x_{kn}} - y_{kn} + x_{kn} \tag{4}$$

where x and y are elements of the matrices X and Y, respectively. The distance between Y and WH symmetrically increases when $x > y$ and also when $x < y$ in eq 3. However, the distance (or more accurately, the divergence) largely increases when $x < y$ in eq 4. This means there are penalties when $WH > Y$ in eq 4, which results in different values for the decomposed components in eqs 3 and 4. The matrix elements of Y and WH are used for iterative calculation, and W and H are updated with non-negative constraints so that the distance (divergence) is decreased according to the following equations of 5–8.

NMF algorithm in Frobenius distance:

$$H_{km} \leftarrow H_{km}\frac{\sum_n Y_{kn}W_{mn}}{\sum_n (WH)_{kn}W_{mn}} \tag{5}$$

$$W_{mn} \leftarrow W_{mn}\frac{\sum_k Y_{kn}H_{km}}{\sum_k (WH)_{kn}H_{km}} \tag{6}$$

NMF algorithm in KL divergence:

$$H_{km} \leftarrow H_{km}\frac{\sum_n \frac{Y_{kn}W_{mn}}{(WH)_{kn}}}{\sum_n W_{mn}} \tag{7}$$

$$W_{mn} \leftarrow W_{mn}\frac{\sum_k \frac{Y_{kn}H_{km}}{(WH)_{kn}}}{\sum_k H_{km}} \tag{8}$$

There are several seeding methods that can be used to initialize the W and H matrices. One of these is the random seeding method in which a seed is drawn from a uniform distribution. This seeding method sometimes has difficulty in converging the calculation. We used the non-negative double singular value decomposition seeding method (NNDSVD), which completed the iterative process more quickly. We also used ICA as a seeding method, which uses only the positive part of the result. The above-mentioned algorithms and seeding methods are available in the NMF package for R.[28]

Implementation of NMF for GC × GC–HRTOFMS. The NMF is implemented in an in-house R code that deconvolutes all the peaks in the GC × GC 2D chromatogram. Version 3.0.2 of R (64-bit version)[34] was used in this study.

We first completed the processing step to define the peak and its region before deconvolution in the code program. This peak selection process is done using an inverse-watershed algorithm, which was proposed as the peak selection method for the two-dimensional chromatogram and was also

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829–1838

implemented in GC Image version 2.2b.[35] This was followed by a NIST11 library search (main library with 212 961 spectra) to identify the compounds for all of the deconvoluted peaks and the stored RTs in GC1 and GC2 in the peak list. The same process was carried out for the original chromatogram. When evaluating the results, a library match factor (MF) ≥ 900 was considered as a clean peak, once library hits overlapped in close proximity (GC1 < 1 min, GC2 < 1 s) were removed to avoid multiple counts of peaks with the same origin and isomers with similar structures. In the removal of the multiple counts, the peak with the higher MF was kept and included in a list. The selected spectra list was manually checked point-by-point as a final inspection to rank the confidence of the identification, making partial reference to the level system concept in high-resolution MS/MS.[36] The level system concept suggested in the MS/MS is arranged for the GC-HRMS as follows: Level 5 indicates a few spectra for a compound in the library or a detected peak to assign; level 4 indicates that there is excessive noise from the instrument or sample matrix (or loss of important spectra in the assignment); level 3 is assignable but there is the possibility that similar structured compounds will remain; level 2 is assignable as unique spectra; and level 1 is confirmed by the standard. A lower unique level indicates reliable spectra for the assignment. The entire workflow is shown in Figure 1.

The NMF was carried out iteratively for all the selected peaks using several mass precision settings (0.05, 0.1, and 1) that changed the size of the data matrix in NMF, so that the performance of the method could be compared. The studied instrumental mass resolution is 10 000 (Δ0.03 at m/z 300). Given the high separating ability in GC × GC, the number of overlapping peaks was assumed to be up to 4. We considered that this was a suitable number of factors to be deconvoluted in NMF. In practice, the number of factors was set at 5, assuming an additional factor for the noise component. The deconvoluted spectra were stored in a newly generated chromatogram, called a layer, in descending order of the total ion intensity in the top pixel of the deconvoluted peak. The four primary output layers were evaluated in this study. Figure 2 shows the schematic of the layer-style output of the NMF-based deconvolution. NNSVD was used as the baseline seeding method to generate initial matrices of W and H for the spectral deconvolution. Frobenius was chosen as the baseline NMF algorithm method for approximate calculation of the matrices. This combination of methods reduces the calculation cost compared with other combinations; therefore, it was used and evaluated in this study unless otherwise specified. The validity of these parameter settings, including the NMF algorithm, seeding method, and number of factors, was evaluated and is discussed in the Results and Discussion section. The typical computational times to deconvolute all of the peaks in the sediment sample (data size, 170 MB; resolution, 10 000; scan rate, 50 Hz) and generate a layer were 1, 2.5, and 3 h for m/z precisions of 1, 0.1, and 0.05, respectively, by a 64-bit computer (CPU, Core i7−2760QM 5160, quad-core of 2.4 GHz; physical memory, 32 GB RAM with 1333 MHz).

Evaluation of Spectrum Similarity. The deconvoluted spectra were evaluated by a NIST MS search 2.0 with the "similarity" setting, a "simple" match factor, and a "default" presearch. According to the guidelines built into the NIST software, 900 or greater is an excellent match, 800−900 is a good match, 700−800 is a fair match, and less than 600 is a very poor match.

When the spectrum similarity, rather than the reference spectrum in the NIST library, was performed to make comparisons between different spectra, the MF and the reverse match factor (RMF) were calculated by the Euclidean dot product (eq 9), in an in-house program within R. This calculation procedure in the in-house program corresponded with the above-mentioned settings in the NIST MS search.

$$\mathrm{MF} = \left(\frac{m\sqrt{A_\mathrm{d}}}{\sum m\sqrt{A_\mathrm{d}}}\right)\cdot\left(\frac{m\sqrt{A_\mathrm{u}}}{\sum m\sqrt{A_\mathrm{u}}}\right) \times 1000 \tag{9}$$

where A_u is the normalized intensity of the m/z of the target spectrum and A_d is its counterpart spectrum. RMF is calculated by the same equation but the m/z signals that do not exist in A_d are removed from A_u. The in-house program is built so that it can calculate the dot product of the accurate mass spectrum to the desired precision.

RESULTS AND DISCUSSION

Overview of the Deconvoluted Chromatogram. The GC × GC−HRTOFMS measurement data (MS resolution, 10 000; scan rate, 50 Hz) of the crude sediment extract were

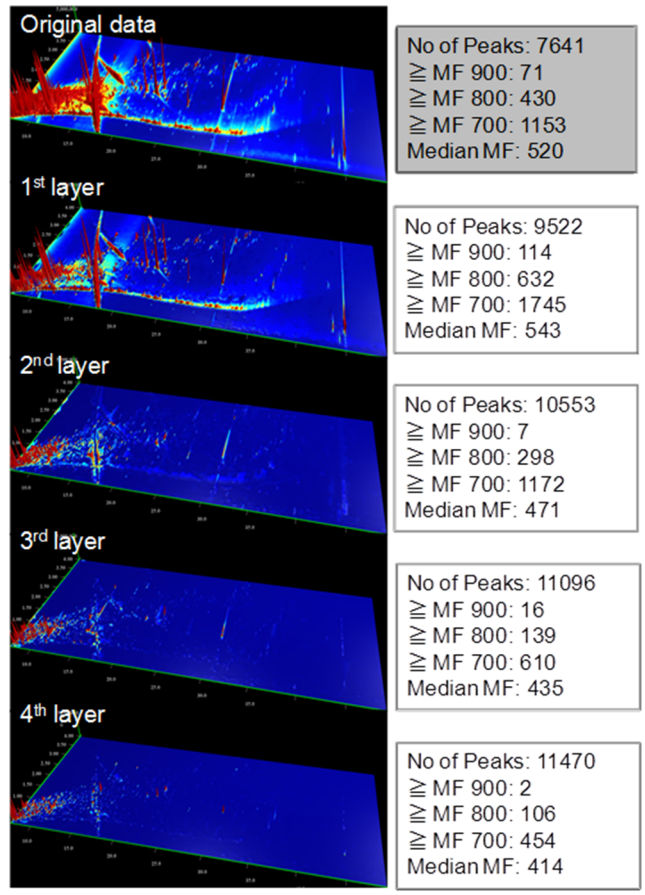


Figure 3. Original chromatogram and generated chromatogram layer after NMF-based spectral deconvolution. MF indicates the match factor of the detected peaks by NIST MS search. The peaks in the original data and each generated layer were picked by an inverse-watershed algorithm implemented in GC Image. Peak assignment is done by NIST MS search. Duplicated assigned peaks in close proximity and between different layers were not removed in this summary to provide an unconsolidated result by NMF-based spectral deconvolution.

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

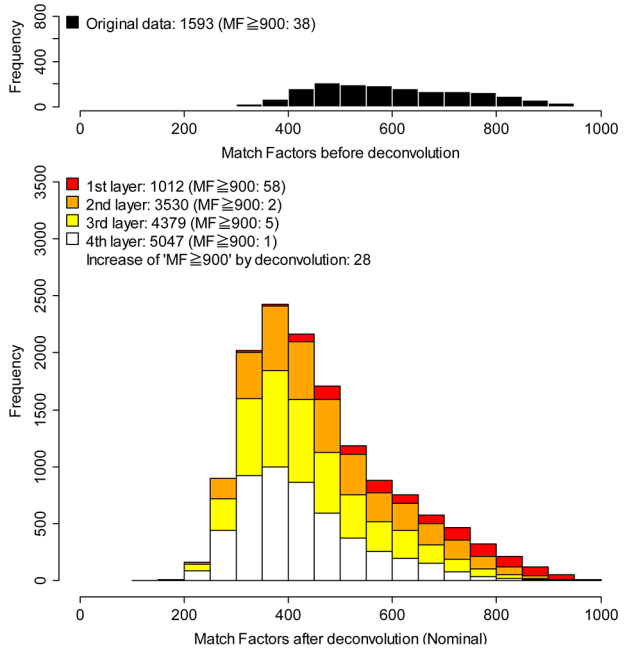


Figure 4. Histograms of match factor (MF) of the peaks in the GC × GC 2D chromatogram. The result of the sediment sample measurement with a MS resolution of 10 000 and a scan rate of 50 Hz after removing duplication. The NMF was conducted with a m/z precision of 1. Duplications in close proximity (GC1 < 1 min, GC2 < 1 s) were removed in the original data (upper). In addition to this, duplications among the layers were removed in deconvoluted chromatograms (lower). When the duplication was found, only the peak with a higher MF was retained and counted.

deconvoluted with a m/z precision of 1 (i.e., nominal) and the individual summaries of each deconvoluted layer are shown in Figure 3. To allow comparison among the layers, duplicated peaks were not removed from Figure 3 (see Figure S-1 in the Supporting Information for the fifth layer). The number of peaks in the deconvoluted layer increased (9 522, 10 553, 11 096, and 11 470 for the first, second, third, and fourth layer, respectively) relative to the number of peaks in the original data (7 641). The trend became more noticeable as the layer became increasingly deep. This is because there was more noise associated with the shape of the deconvoluted peaks in the deeper layers as a result of the statistical processing, so the peaks were detected separately by the inverse-watershed method. With the exception of the first layer, the median MF and the number of peaks with MF ≥ 900, MF ≥ 800, and MF ≥ 700 basically decreased. This means that the deeper layers contained statistical noise and there was a tendency for trivial information to accumulate in the analysis. However, they still contained a spectrum that could potentially be recognized, so it was still meaningful to use a screening procedure for these deeper layers. This suggests that rapid automated screening of several layers should be accompanied by the deconvolution process, as implemented in this study.

Because multiple duplicate hits in close proximity among layers or separately identified peaks close together in the same layer (as mentioned above) disturb the evaluation of the effects of deconvolution in nontarget screening, the duplicate hits were removed and the assigned spectra were counted (Figure 4). As a result of deconvolution, the total number of peaks with MF ≥ 900 increased by 28, which was 1.7 times greater than the number of peaks before deconvolution.

On the basis of the selection criteria, there were many duplicate peaks close together, even within the original data itself (the number of peaks decreased from 7 641 to 1 593). In the histogram of the deconvoluted chromatogram (Figure 4), the number of peaks with a low MF was seen to expand, and these peaks were not removed by the process of duplicate removal. This expansion was considered to be statistical noise which is the result of the deconvolution. From the perspective of statistical comparison, the issue of multiple comparisons still remains, though it has not given rise to problems of practical importance. For final determination of the obtained candidate, confirmation by a reference standard is desirable. The method provided in this study should be recognized as to extract potentially assignable spectra toward final determination.

Differences in Various Measurement Conditions and m/z Precision in Deconvolution. The MS resolution, measurement scan rate, and m/z precision in the NMF-based deconvolution were evaluated to find the optimal conditions for compound identification (after removing the duplication). The number of detections with MF ≥ 900 in all conditions after deconvolution is 1.7−3.1 times greater than with no deconvolution (Table S-1 in the Supporting Information). Changes in the MS resolution in the measurement and the m/z precision in the deconvolution did not change the significance result for MF ≥ 900, MF ≥ 800, and MF ≥ 700. Increases in the scan rate enhanced the detection power for the same samples regardless of how the deconvolution was executed. The rate of increase in the number of peaks with MF ≥ 900 by deconvolution was slightly more pronounced for 25 Hz (2.2−3.1 times) than for 50 Hz (1.7−2.2 times). However, the rate of increase in the number of peaks with MF ≥ 700 and 800 was almost the same or just slightly more for 50 Hz. This suggests that there were no significant differences between 25 and 50 Hz. The number of peaks increased by 50 Hz recording, indicating that the high picture resolution of the chromatogram enhanced its ability to recognize slightly different peak elution times. This property will help enhance the detection power in the chromatographic analysis. Recording with a higher scan rate is recommended to find a large number of identifiable peaks; however, if the scan rate is increased too much, the instrumental detection limit and data processing speed will be affected. Histograms showing detailed deconvolution results in a range of conditions are provided in the Supporting Information (Figures S-2−S-8). Although the intensity threshold cutoff (3 000) in the original data attempted to enhance the detection power of compounds (Figure S-8 in the Supporting Information), there was an increase in accidental matches,[33] which is caused by a spectrum with a few numbers in the library, or in a target spectrum, such as in a subtracted spectrum. Although the threshold cutoff at 1 000 showed similar trends as described above, there was no difference between the intensity cutoff points of 500 and 100 and the no-threshold cutoff for this analytical condition. The intensity value strongly depends on the voltage of the detector or other instrumental settings, but a small reduction in the threshold (500 in the current study) has an advantage. The slight reduction in the threshold suppresses the computational calculation cost.

The advantage of using a high m/z precision in the deconvolution was highlighted from the perspective of spectral storage with accurate mass. In the deconvolution, the m/z values need to be rounded to a certain precision to make the input data matrix for the NMF. In practice, accurate mass

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

Table 1. Non-Target Compound List (MF ≥ 900) from the Sediment Sample by GC × GC−HRTOFMS with the Global Spectral Deconvolution[a]

no.	compound name	classification	RT I (min)	RT II (s)	formula	MF	RMF	unique level	CAS [NIST no.]	detected layer	new detection or MF update	layers of duplicate hit found in close proximity	total number of duplicate hit [in different RTs]
1	Tetradecane, 2,6,10-trimethyl-	Alkane	10.08	0.26	C17H36	927	938	3	14905-56-7	Original data		Original data	1
2	Gleenol	Terpene	10.21	0.50	C15H26O	912	956	2	[374181]	Original data			0
3	Heptadecane, 9-octyl-	Alkane	13.48	0.46	C25H52	924	929	3	7225-64-1	Original data		3, 4	2 [1]
4	9-Hexadecenoic acid	Carboxylic acid	15.01	1.01	C16H30O2	911	921	3	2091-29-4	Original data			0
5	Eicosane, 7-hexyl-	Alkane	16.08	0.59	C26H54	901	906	3	55333-99-8	Original data			0
6	Cyclic octaatomic sulfur	Inorganic chemical	17.61	0.33	S8	907	944	2	10544-50-0	Original data		Original data	3 [2]
7	1-Phenanthrenecarboxaldehyde, 7-ethenyl-1,2,3,4,4a,4b,5,6,7,9,10,10a-dodecahydro-1,4a,7-trimethyl-, [1R-(1α,4aβ,4bα,7β,10aα)]-	Terpene	20.55	1.77	C20H30O	916	926	3	472-39-9	Original data		1, 2	4
8	Tetratriacontane	Alkane	32.55	0.77	C34H70	910	946	3	14167-59-0	Original data			0
9	Decane	Alkane	7.21	1.09	C10H22	911	945	3	124-18-5	1	New	1	3 [1]
10	1,6-Dioxacyclododecane-7,12-dione	Alkyl ester	9.88	1.31	C10H16O4	927	934	2	777-95-7	1	New		0
11	Cyclooctasiloxane, hexadecamethyl-	Siloxane	10.35	0.15	C16H48O8Si8	910	945	2	556-68-3	1	New		0
12	Eicosane, 10-methyl-	Alkane	12.55	0.42	C21H44	943	946	3	54833-23-7	1	New	3	4 [1]
13	Benzenesulfonamide, N-butyl-	Aromatic compound	12.68	1.68	C10H15NO2S	941	961	2	3622-84-2	1	Updated	Original data	1
14	i-Propyl 12-methyl-tridecanoate	Alkyl ester	12.88	0.59	C17H34O2	900	942	3	[336604]	1	New		0
15	2-Pentadecanone, 6,10,14-trimethyl-	Alkyl ketone	13.21	0.66	C18H36O	938	978	3	502-69-2	1	New		0
16	Heneicosane	Alkane	13.48	0.46	C21H44	970	971	3	629-94-7	1	New	1, 2	2 [5]
17	Hexadecanoic acid, methyl ester	Alkyl ester	14.61	0.79	C17H34O2	904	944	3	112-39-0	1	New	2	1
18	1-Dodecanol, 2-hexyl-	Alkyl alcohol	15.01	0.53	C18H38O	908	921	3	110225-00-8	1	New		0
19	Palmitoleic acid	Carboxylic acid	15.01	1.01	C16H30O2	964	985	3	373-49-9	1	New	1	2
20	Kaur-16-ene, (8a,13a)-	Terpene	17.28	1.29	C20H32	947	969	3	20070-61-5	1	Updated	Original data	1
21	1-Naphthalenepropanol, α-ethenyldecahydro-α,5,5,8a-tetramethyl-2-methylene-, [1S-[1α(S*),4aβ,8aα]]-	Terpene	17.68	1.36	C20H34O	966	974	2	596-85-0	1	Updated	Original data	1
22	Methyl stearate	Alkyl ester	18.75	0.98	C19H38O2	946	967	3	112-61-8	1	Updated	Original data	1
23	Octadecanoic acid	Carboxylic acid	19.55	1.16	C18H36O2	954	965	3	57-11-4	1	Updated	Original data, 1, 2	5
24	Decanedioic acid, dibutyl ester	Alkyl ester	19.68	1.36	C18H34O4	924	947	3	109-43-3	1	New		0
25	Hexadecanamide	Alkyl amide	19.95	1.68	C16H33NO	923	945	3	629-54-9	1	New	2	2
26	Hexadecanoic acid, butyl ester	Alkyl ester	20.01	0.98	C20H40O2	947	950	3	111-06-8	1	Updated	Original data	1
27	11H-Benzo[b]fluorene	PAH	21.68	2.86	C17H12	904	917	3	243-17-4	1	New	2	1
28	9-Octadecenamide, (Z)-	Alkyl amide	23.81	1.66	C18H35NO	928	947	3	301-02-0	1	Updated	Original data	1
29	Octadecanamide	Alkyl amide	24.28	1.49	C18H37NO	936	953	3	124-26-5	1	Updated	Original data, 2	4

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

Table 1. continued

no.	compound name	classification	RT I (min)	RT II (s)	formula	MF	RMF	unique level	CAS [NIST no.]	detected layer	new detection or MF update	layers of duplicate hit found in close proximity	total number of duplicate hit [in different RTs]
30	Hexadecane	Alkane	24.68	0.74	C16H34	928	935	3	544-76-3	1	New		0
31	Tetracosane	Alkane	26.35	0.70	C24H50	959	971	3	646-31-1	1	New	1, 2	22 [9]
32	13-Docosenamide, (Z)-	Alkyl amide	30.95	1.92	C22H43NO	909	931	3	112-84-5	1	New	1	4
33	Dibenzo[def,mno]chrysene	PAH	37.95	2.91	C22H12	932	932	3	191-26-4	1	New		0
34	1,5-Heptadien-3-yne	Alkyne	5.81	3.56	C7H8	911	911	3	3511-27-1	2	New	2	1
35	2-Methyl-Z-4-tetradecene	Alkene	8.01	0.11	C15H30	903	907	3	[130783]	2	New		0
36	Cyclohexanemethanol, 4-ethenyl-a,a,4-trimethyl-3-(1-methylethenyl)-, [1R-(1a,3a,4a)]-	Terpene	9.88	0.44	C15H26O	913	969	3	639-99-6	2	Updated	Original data, 1, 2	3
37	Heptadecane	Alkane	12.35	0.35	C17H36	910	931	3	629-78-7	2	New		0 [1]
38	Benz[a]azulene	Aromatic compound	13.01	1.68	C14H10	913	945	3	246-02-6	2	New	2	1
39	Eicosane	Alkane	13.15	0.39	C20H42	964	969	3	112-95-8	2	Updated	Original data, 1, 2, 3	20 [11]
40	Pentadecanoic acid	Carboxylic acid	13.48	0.90	C15H30O2	953	961	3	1002-84-2	2	Updated	Original data, 1, 2	8
41	Phenanthrene, 1-methyl-	PAH	15.41	2.14	C15H12	938	962	3	832-69-9	2	New	1, 2	3
42	Kaur-15-ene, (5a,9a,10a)-	Terpene	16.28	1.12	C20H32	951	967	3	511-85-3	2	Updated	Original data, 1, 2, 4	6
43	Propanoic acid, 3-mercapto-, dodecyl ester	Alkyl ester	16.48	1.29	C15H30O2S	910	929	3	6380-71-8	2	New	2	3
44	Octadecane, 1-(ethenyloxy)-	Alkyl ether	18.48	0.92	C20H40O	919	931	3	930-02-9	2	New		0
45	Hexadecane, 1-(ethenyloxy)-	Alkyl ether	18.68	0.92	C18H36O	916	933	3	822-28-6	2	New	2	2
46	2(3H)-Furanone, dihydro-5-tetradecyl-	Alkyl ester	22.88	1.51	C18H34O2	901	904	3	502-26-1	2	New		0
47	Hexanedioic acid, bis(2-ethylhexyl) ester	Alkyl ether	24.28	1.09	C22H42O4	960	982	3	103-23-1	2	Updated	Original data, 1, 2, 3	6
48	Chrysene	PAH	26.21	3.00	C18H12	947	950	3	218-01-9	2	New	2, 4	9
49	2-methylhexacosane	Alkane	26.35	0.74	C27H56	920	948	3	[376727]	2	New		0 [12]
50	9H-Tribenzo[a,c,e)cycloheptene	Aromatic compound	28.35	2.91	C19H14	901	928	3	213-10-5	2	New		0
51	2-methyloctacosane	Alkane	29.88	0.83	C29H60	935	950	3	[376728]	2	New		0 [7]
52	Perylene, 3-methyl-	Aromatic compound	33.81	3.50	C21H14	904	922	3	[155255]	2	New		0
53	Propanoic acid, 3,3′-thiobis-, didodecyl ester	Alkyl ester	44.95	0.74	C30H58O4S	948	950	2	123-28-4	2	New	1, 2, 3	22
54	Nonane	Alkane	7.41	2.32	C9H20	904	932	3	111-84-2	3	New	3	3
55	Tetradecanoic acid	Alkyl ester	11.95	0.66	C14H28O2	932	959	3	544-63-8	3	New	3	2
56	Cyclononasiloxane, octadecamethyl-	Siloxane	12.35	0.26	C18H54O9Si9	936	963	2	556-71-8	3	Updated	Original data, 1, 2	3
57	Cyclodecasiloxane, eicosamethyl-	Siloxane	14.81	0.31	C20H60O10Si10	916	926	2	18772-36-6	3	New		0
58	1-Heptatriacotanol	Alkyl alcohol	15.15	1.01	C37H76O	912	920	3	105794-58-9	3	New	3	1
59	Diphenyl sulfone	Aromatic compound	15.15	2.91	C12H10O2S	913	952	2	127-63-9	3	New	2, 3	2
60	Thunbergol	Terpene	17.48	1.29	C20H34O	901	955	3	25269-17-4	3	New		0

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

Table 1. continued

no.	compound name	classification	RT I (min)	RT II (s)	formula	MF	RMF	unique level	CAS [NIST no.]	detected layer	new detection or MF update	layers of duplicate hit found in close proximity	total number of duplicate hit [in different RTs]
61	trans-13-Docosenamide	Alkyl amide	31.01	1.53	C22H43NO	912	913	3	10436-09-6	3	Updated	Original data, 3	3
62	Heptadecane, 2,6,10,15-tetramethyl-	Alkane	17.68	0.63	C21H44	905	926	3	54833-48-6	4	New	4	2 [1]

[a]Mass resolution, 10 000; scan rate, 50 Hz; m/z precision in deconvolution, 1, NMF algorithm, "Frobeniuse"; seeding method, ICA. The numbers of multiple same hits in different RTs are shown in brackets, but they are not listed. "New detection" indicates the spectrum was newly assigned with MF ≥ 900 or MF was updated by the deconvolution from the original data. The lower unique level indicates reliable spectra for assignment. Level 5 indicates a few spectra for a compound in the library or a detected peak to assign; level 4 indicates that there is excessive noise from the instrument or sample matrix (or loss of important spectra in the assignment); level 3 is assignable but there is the possibility that similar structured compounds will remain; level 2 is assignable as unique spectra; and level 1 is confirmed by the standard. Only the assigned peaks with unique levels lower than 3 are listed.

values are retrievable, even after deconvolution with an m/z precision of 1, by tracking the original mass value before rounding. This function has been built into the program code that was developed in this study. However, when the accurate masses are within the same range as a nominal mass, their rounded m/z is mixed (summed) and the result is irreversible (e.g., both m/z 100.1 and m/z 100.2 are recognized as 100 when the m/z precision is 1 and cannot be restored). In the program, the smaller mass is systematically chosen for forced restoration. This often occurs in analyzed samples, resulting in a decrease in the accuracy of the spectrum collected after low-precision deconvolution from the actual sample (Table S-2 and Figure S-9 in the Supporting Information). This issue will be addressed from the perspective of algorithm development in a future study. When the m/z precision of the deconvolution was 0.1, no great deterioration of mass was observed in the sample, even when the MF between the reference spectra (accurate mass measurement) and the deconvoluted mass with a precision of 0.1 was calculated with a precision calculation at 0.05. The instrumental mass resolution and the possibility of detecting several masses in a narrow mass range are correlated, thus the appropriate m/z precision in the deconvolution needs to be carefully considered. Even when analyzing the sediment sample that contained enriched matrices with the instrument (mass resolution of 10 000), it was difficult to find interference mass spectra within the margin of m/z 0.1; therefore, in practice a m/z value of 0.1 is recommended when spectral storage with accurate mass and nontarget screening are considered. The m/z precision should be greater than 0.1 when the resolution of the detector and the computer performance are much higher. A m/z precision value of 1 in the deconvolution is only acceptable for nontarget screening, not including the step of spectral storage.

Influence of the NMF Parameter Settings in Spectral Deconvolution. The spectra obtained by different NMF algorithms (Frobenius and KL) were compared (Table S-3 in the Supporting Information). Comparison of the spectra of the largest 50 peaks shows that there was no clear difference between the methods. However, when the peaks were chosen randomly, the deconvoluted spectra were not closely matched below layer 2. This indicates that there was some difficulty in extracting the second (third and fourth) intense component(s) of the weaker peaks by the deconvolution, and the process may have been affected by the relatively intense peaks, or perhaps there were no clear spectra in the original data. Primary peaks were unchangeably obtained in both algorithms. Manual inspection of the spectra and compounds assigned by NIST11 with MF ≥ 900 in both NMF algorithms confirmed that there were no significant differences in meaningful spectra in their lists (63 and 64 spectra were listed as meaningful spectra by Frobenius and KL, respectively). This indicates that the differences in the NMF algorithms do not really influence the results of nontarget screening and spectral storage. The different NMF algorithms are compared in Tables S-4 and S-5 in the Supporting Information.

NMF using three to eight factors was evaluated to determine the difference resulting from the number of factors. As with the NMF algorithms, there were no significant differences due to the number of factors for primary peaks when between three and eight factors were used (Tables S-6 and S-7 in the Supporting Information).

The NNDSVD, ICA, and random seeding methods were compared. There was a large difference in the deconvoluted

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

spectra even for the largest 50 peaks in layer 1 (Table S-9 in the Supporting Information). As described in the earlier Methods section, NNDSVD is not a randomization method but rather is based on singular value decomposition, and ICA uses the positive part of the result as an initial seed. The random seeding method, which draws values from a uniform distribution, does not provide a convergent output and causes a calculation break in some cases. In practice, there are similar issues with the ICA calculation process, but they occur less frequently than with the random method. By manually inspecting the spectra and the compounds assigned by NIST11 with MF $\geq$ 900 for both the NNDSVD and ICA seeding methods, the number of meaningful spectra was confirmed to be greater using the ICA method (without considering removal of duplicate hits at different RTs, 63 and 102 spectra were found by NNDSVD and ICA seeding methods, respectively). Therefore, the ICA method shows higher performance for nontarget screening in this study provided there are no calculation breaks. However, using NNDSVD, which is a deterministic method, as an initial seeding solves the equation faster. Therefore, it works for sequential data analysis and obtains a reproducible output. The choice of the seeding method will depend on the objective of the analysis.

Comparison of the Performances among NMF and PARAFACs. Performance of the NMF-based deconvolution was compared with other PARAFAC-based deconvolution methods, whose basis was applied for GC × GC with low mass resolution TOFMS in previous studies.[16] As discussed in the introduction, the PARAFAC in the TOFMS required several steps to avoid having negative values in the output in their study. The recently available PARAFAC program code in Matlab has non-negative constraints in the algorithm, which could be considered as a non-negative tensor factorization. We applied NMF, PARAFAC, and PARAFAC with non-negative constraint (NPARAFAC) to a mixed peak containing hexachlorinated biphenyl (HxCB) in the CRM sediment extract (Figure S-9 in the Supporting Information) and compared their performance. The tensor factorization PARAFAC has a longer calculation time than NMF (calculated in R), even in Matlab, which generally supports faster calculation than R, the time taken for NPARAFAC in Matlab was more than 500 times compared with NMF in R (Table S-10 in the Supporting Information). The spectra deconvoluted by normal PARAFAC contained both positive and negative values resulting in lower MF than the other two methods (Figures S-10−S-12 in the Supporting Information). Examination of the extraction of HxCB mass spectrum shows that both NMF and NPARAFAC provide deconvoluted spectra of high MF, though that provided by NMF is slightly higher (MF 927 in NMF and 915 in NPARAFAC). The results of the obtained mass spectra and the required calculation cost for each method show that NMF is one of the best choices for global deconvolution.

NMF-Based Deconvolution in Practice. Considering the above discussion, we have chosen to use the ICA seeding method, the Frobenius NMF algorithm, and a factor number of 5 in NMF, together with a mass resolution of 10 000 and a scan rate of 50 Hz; this was considered appropriate for this sample to perform nontarget screening for high efficiency in detection. To save calculation time, a m/z precision of 1 is better for spectral assignment; however, a m/z precision of 0.1 is recommended for spectral storage on an accurate mass basis. The results of nontarget screening that included the deconvolution process based on the above conditions, but with an m/z precision of 1, are shown in Table 1. In the original measurement, 36 spectra including the same hits in different RTs were assigned MF $\geq$ 900. When the deconvoluted spectra list was included, their 28 spectra were updated with higher MF spectra (including multiple hits with different RTs). In total, 98 spectra were either newly added or updated by the deconvolution of the original list. After removing multiple hits in different RTs (extracting a unique compound name in the list), 62 unique assignable spectra, including 54 additional (39 new and 15 updated) spectra, were obtained from the original data by the deconvolution process (Table 1). Several assignable peaks in original data were lost from the list obtained from the deconvoluted data, possibly because of a spectral change that resulted in a slight deterioration of the MF relative to the original in some cases. Although there was only a slight degree of MF deterioration for the peaks for which the original MF value was high, losses from the list were complemented by including the hit list from the original data that was applied in this study. Table 1 shows that the detection power of qualitative nontarget screening is improved by the application of NMF-based global spectral deconvolution. Although the peak shapes of the obtained spectra are not ideally reconstructed by the deconvolution because of, for example, the number of chromatographic data points (scan rate), the detected position of the obtained peaks in the watershed, and interference from unseparated spectra with relatively high intensity, global deconvolution is effective for qualitative nontarget screening and spectral storage. The peak shape issue will be complemented by consequent use of automatic peak extraction as a target screening using the obtained RTs and mass spectra. The method has been suggested in previous reports as TSEN.[8]

From the perspective of spectral storage, the obtained spectra from actual samples of complex mixtures are storable in the database with accurate mass as reference spectra (a m/z precision of 0.1 is recommended with a mass resolution of 10 000). An example of the database obtained in this study is provided as an Excel file in the Supporting Information. A formatted database is obtainable via the software introduced in the Supporting Information. Table S-2 in the Supporting Information shows that the accurate mass database will have a high potential for exclusive assignment of compounds. In this context, using a higher mass resolution with a more precise m/z setting in deconvolution (e.g., mass resolution of 100 000 with m/z precision of 0.01) will allow us to construct a more accurate database that will then result in more exclusive assignment of the compounds. However, there are several challenges, such as the development of higher resolution MS with higher scan rates and faster computational processing during the deconvolution. Recent developments in MS and computers will help overcome these challenges. Furthermore, using a high-speed computer language in the processing, such as Matlab and C++, has the potential to reduce the calculation time.

Storage of spectra, which are assigned with reference data (library), obtained from actual samples will be beneficial for building up knowledge of chemicals in samples and constructing a spectral database. As mentioned above, the NMF-based deconvolution accelerated the construction of the database with high resolution. After stored assigned spectra are confirmed by chemical standards, they can be used for target screening, which is performed by a previously developed

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838

automatic processing technique TSEN,[8] even in different samples. The accumulated measurement data will form the basis of the measurement-based accurate mass database. Nontarget screening with NMF-based deconvolution will provide an overview and more information about the samples of interest, learning from actual samples.

CONCLUSION

In this study, we developed an NMF-based deconvolution for GC × GC−HRTOFMS measurement and evaluated its performance in detail using a sediment sample. We found that the mass resolution does not enhance the deconvolution performance but that a high scan rate increases the number of assignable spectra. Deconvolution with nominal mass precision for GC × GC−HRTOFMS successfully increased the number of assignable spectra, which therefore demonstrates its potential to accelerate nontarget screening in GC × GC−HRTOFMS. This study demonstrates that applying deconvolution to a high-resolution mass allows us to collect and construct a spectral database on an accurate mass basis from actual samples of complex mixtures. Selecting an appropriate level of m/z precision in the deconvolution is critical so that the measurement-based database can be constructed, as complex mixtures contain a lot of noise in narrow m/z ranges. The accurate mass database will be important for highly accurate assignment of compounds. The development of a measurement-based database from actual sample analysis is also expected to accelerate target screening after the obtained database list is confirmed by chemical reference standard. Although there is still an issue with calculation cost, implementation of high-speed processing in a language such as Matlab and C++, and further development of laptop computers, has the potential to reduce the calculation time. The global spectral deconvolution in GC × GC−HRTOFMS will accelerate the process of nontarget screening and be beneficial for the consequent process of target screening in complex mixtures.

ASSOCIATED CONTENT

Supporting Information

Figures S-1−S-12 and Tables S-1−S-10 (pdf file) and a database obtained in this study (Excel file). This material is available free of charge via the Internet at http://pubs.acs.org. The program source code to perform NMF-based deconvolution on the GC × GC−HRTOFMS data are available via the Internet at http://www.nies.go.jp/analysis/downloads.html.

AUTHOR INFORMATION

Corresponding Author

*Phone: +81-29-850-2914. E-mail: yasuyuki.zushi@gmail.com.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the Uchem at Eawag (Zurich, Switzerland) and the Environmental Chemistry Modeling Laboratory in EPFL (Lausanne, Switzerland) for giving us valuable comments on the study. This study was supported by JSPS Research Fellowships for Young Scientists (Grant No. 249089) and JSPS KAKENHI (Grant No. 26241026).

REFERENCES

(1) Ochiai, N.; Ieda, T.; Sasamoto, K.; Fushimi, A.; Hasegawa, S.; Tanabe, K.; Kobayashi, S. *J. Chromatogr., A* **2007**, *1150*, 13−20.
(2) Shunji, H.; Yoshikatsu, T.; Akihiro, F.; Hiroyasu, I.; Kiyoshi, T.; Yasuyuki, S.; Masa-aki, U.; Akihiko, K.; Kazuo, T.; Hideyuki, O.; Katsunori, A. *J. Chromatogr., A* **2008**, *1178*, 187−198.
(3) Reichenbach, S. E.; Tian, X.; Tao, Q.; Ledford, E. B., Jr; Wu, Z.; Fiehn, O. *Talanta* **2011**, *83*, 1279−1288.
(4) Nabi, D.; Gros, J.; Dimitriou-Christidis, P.; Arey, J. S. *Environ. Sci. Technol.* **2014**, *48*, 6814−6826.
(5) Hashimoto, S.; Takazawa, Y.; Fushimi, A.; Tanabe, K.; Shibata, Y.; Ieda, T.; Ochiai, N.; Kanda, H.; Ohura, T.; Tao, Q.; Reichenbach, S. E. *J. Chromatogr., A* **2011**, *1218*, 3799−3810.
(6) Hashimoto, S.; Zushi, Y.; Fushimi, A.; Takazawa, Y.; Tanabe, K.; Shibata, Y. *J. Chromatogr., A* **2013**, *1282*, 183−189.
(7) Ieda, T.; Ochiai, N.; Miyawaki, T.; Ohura, T.; Horii, Y. *J. Chromatogr., A* **2011**, *1218*, 3224−3232.
(8) Zushi, Y.; Hashimoto, S.; Fushimi, A.; Takazawa, Y.; Tanabe, K.; Shibata, Y. *Anal. Chim. Acta* **2013**, *778*, 54−62.
(9) Zushi, Y.; Hashimoto, S.; Tamada, M.; Masunaga, S.; Kanai, Y.; Tanabe, K. *J. Chromatogr., A* **2014**, *1338*, 117−126.
(10) Hu, C.-D.; Liang, Y.-Z.; Guo, F.-Q.; Li, X.-R.; Wang, W.-P. *Molecules* **2010**, *15*, 3683−3693.
(11) Liang, Y. Z.; Kvalheim, O. M.; Keller, H. R.; Massart, D. L.; Kiechle, P.; Erni, F. *Anal. Chem.* **1992**, *64*, 946−953.
(12) Manne, R.; Shen, H.; Liang, Y. *Chemometr. Intell. Lab.* **1999**, *45*, 171−176.
(13) Stein, S. E. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770−781.
(14) Wei, X.; Shi, X.; Kim, S.; Patrick, J. S.; Binkley, J.; Kong, M.; McClain, C.; Zhang, X. *Anal. Chem.* **2014**, *86*, 2156−2165.
(15) Harshman, R. A. *UCLA Work. Pap. Phonet.* **1970**, *16*, 84.
(16) Hoggard, J. C.; Synovec, R. E. *Anal. Chem.* **2007**, *79*, 1611−1619.
(17) Skov, T.; Hoggard, J. C.; Bro, R.; Synovec, R. E. *J. Chromatogr., A* **2009**, *1216*, 4020−4029.
(18) Lee, D. D.; Seung, H. S. *Nature* **1999**, *401*, 788−791.
(19) Comon, P. *Signal Process.* **1994**, *36*, 287−314.
(20) Kothandaraman, M.; Pachaiyappan, A. *Aust. J. Basic Appl. Sci.* **2013**, *7*, 108−113.
(21) Hyvärinen, A. *Philos. Trans. R. Soc. London, A: Math., Phys. Eng. Sci.* **2013**, *371*, 20110534.
(22) Vosough, M. *Anal. Chim. Acta* **2007**, *598*, 219−226.
(23) Langlois, D.; Chartier, S.; Gosselin, D. *Tutorials Quant. Methods Psychol.* **2010**, *6*, 31−38.
(24) Hoyer, O. P. *J. Mach. Learn. Res.* **2004**, *5*, 1457−1469.
(25) Pauca, V. P.; Piper, J.; Plemmons, R. J. *Linear Algebra Appl.* **2006**, *416*, 29−47.
(26) Drakakis, K.; Rickard, S.; de Frein, R.; Cichock, A. *Int. Math. Forum* **2008**, *3*, 1853−1870.
(27) Yu, S.; Zhang, Y.; Liu, W.; Zhao, N.; Xiao, X.; Yin, G. *J. Chemometr.* **2011**, *25*, 586−591.
(28) Gaujoux, R.; Seoighe, C. *BMC Bioinf.* **2010**, *11*, 367.
(29) Gao, H.-T.; Li, T.-H.; Chen, K.; Li, W.-G.; Bi, X. *Talanta* **2005**, *66*, 65−73.
(30) Gao, H. T.; Dai, D. M.; Li, T. H. *Chin. Chem. Lett.* **2007**, *18*, 495−498.
(31) Sun, J.; Li, T.; Cong, P.; Xiong, W.; Tang, S.; Zhu, L. *Talanta* **2010**, *83*, 541−548.
(32) Hoggard, J. C.; Siegler, W. C.; Synovec, R. E. *J. Chemometr.* **2009**, *23*, 421−431.
(33) Stein, S. *Anal. Chem.* **2012**, *84*, 7274−7282.
(34) R Development Core Team. http://www.rproject.org/.
(35) Reichenbach, S. E.; Ni, M.; Kottapalli, V.; Visvanathan, A. *Chemometr. Intell. Lab.* **2004**, *71*, 107−120.
(36) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 2097−2098.

DOI: 10.1021/ac5038544
Anal. Chem. 2015, 87, 1829−1838