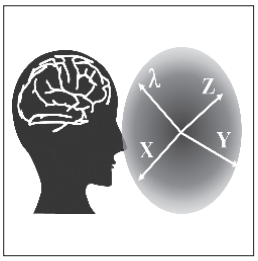


2段階多変量スペクトル分解法を用いた4次元ビッグデータ解析

Four-Dimensional Big Data Analysis Using Two-Step Multivariate Curve Resolution

星名　豊*
Yutaka Hoshina

上村　重明
Shigeaki Uemura

岡本　悠
Haruka Okamoto

久保　優吾
Yugo Kubo

材料の研究開発においては化学種の3次元的な分布の把握が多くの場合で重要であり、その情報を豊富に含む分析4次元ビッグデータの有効活用が今後重要な鍵となる。我々は「2段階多変量スペクトル分解法」という新しいデータ解析手法を開発し、4次元構造のデータを直感的に理解できる形で表現することに成功した。本手法は、2回の多変量スペクトル分解およびその間の「2値化」処理を行うことが特徴である。薄膜サンプルの飛行時間型二次イオン質量分析法によって得られたデータに本解析手法を適用し、複雑な3次元的微細構造を理解することができた。従来のデータ表現法と比べ、2段階多変量スペクトル分解法を用いることで4次元の分析データが明瞭に、かつ理解しやすくなることがわかった。

For research and development in material science, it is important to understand the three-dimensional (3D) distributions of chemical species in samples. The effective utilization of 4D big data that contain a lot of information about the 3D distributions is a key factor. This paper demonstrates a new 4D data analysis technique called “two-step multivariate curve resolution (MCR).” To obtain an intuitive expression of 4D data, we devised a process involving two iterations of MCR with digitization in between. The new technique was applied to the analysis of time-of-flight secondary ion mass spectrometry data derived from a thin-film sample to assist in the interpretation of complex 3D local microstructures. Compared to conventional methods of data presentation, two-step MCR was found to greatly facilitate the clarification and understanding of the 4D analysis data.

キーワード：多変量スペクトル分解法、教師なし機械学習、4次元、飛行時間型二次イオン質量分析法（ToF-SIMS）

1. 緒　　言

機器分析の技術進歩によって、材料の4次元構造データが容易に得られるようになった。例えば、微小領域の質量分析が可能な飛行時間型二次イオン質量分析法（ToF-SIMS）[※1]はスパッタ技術と組み合わせることで**図1**に示すような、空間の各点（*X, Y, Z*）における質量スペクトル（*m/z*）の組、という4次元データを得ることができる。このような4次元データは化学種の空間的な分布に関する大量かつ貴重な情報をもっているが、それらはほとんどの場合有効に活用されていない。それは4次元データが2次元平面上にグラフとして描画することができず、我々が直感的に解釈できないためである。

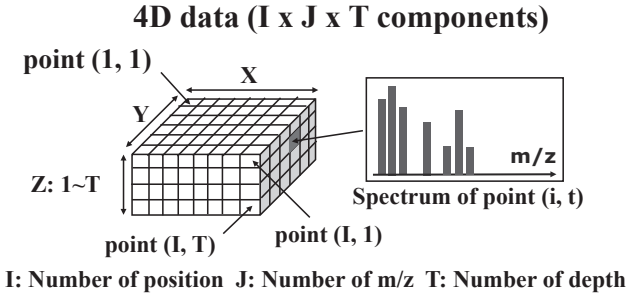


図1　ToF-SIMS分析で得られる4次元データ

4次元データを2次元平面上に表現するには何らかのデータの圧縮あるいは切り取りが必要である。単純なものとしては以下の3つの手法が用いられている。手法1：深さプロファイルのみを描く方法（**図2 (a)**）、手法2：いくつかの質量電荷比に関して3Dプロットする方法（**図2 (b)**）、手法3：ある平面に関する強度マップを描く方法（**図2 (c)**）である。

手法1は、4次元データを*XY*平面に関して平均化するため、*XY*面に関して不均一なサンプルにはふさわしくない。手法2では、異なる質量電荷比のイオン同士の関係を捉えるのが難しい。手法3では、ある特定の面の情報しか得られない。これらの手法にはそれぞれに弱点があり、4次元情報の全体を十分に表現できているとはいえない。

この問題を解決するため、我々は教師なし機械学習に基づき、上記3手法のように単純かつ恣意的な圧縮や切り取りをすることなく、4次元データの重要な特性を自動的に

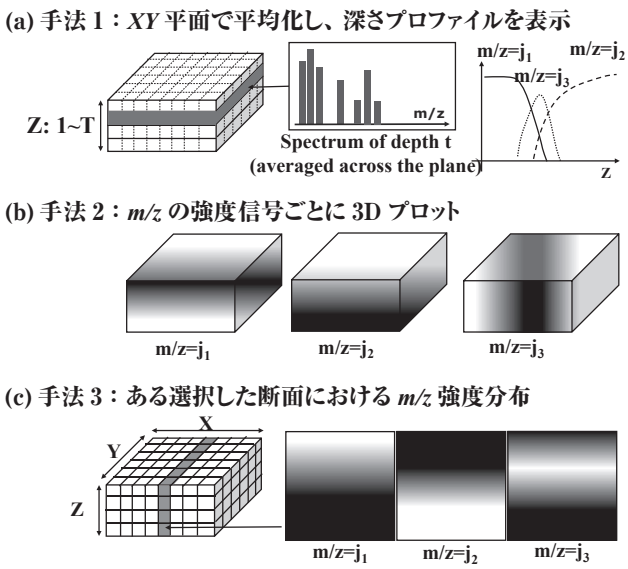


図2　4次元データの古典的な表示手法

抽出することができるような新たな手法を開発した。本論文は、以下、次のように構成されている。まず本手法の数学的手順に関して説明する。その後、本手法をToF-SIMS分析で得られたデータに適用し、試料中における化学種の空間的な分布を明確かつ直感的に表現できることを示す。

2.　新解析手法の数学的手順

2−1　2段階MCRの概要

大量のスペクトルデータを、元の重要な情報を損なわずに数学的に圧縮する技術にはいろいろなものがある[(1)]。その中で最も一般的なものは多変量スペクトル分解法(MCR※2)[(1),(2)]である。新たな解析手法はこのMCRを2回繰り返し、その間に適切な中間処理を行うものであるため「2段階MCR」と名付けた。

MCRは行列因子分解の一種であるため、はじめに今回取り扱う4次元データの行列表現を説明する。ToF-SIMS分析で得られる4次元データは以下のように表される。

$$D=\begin{pmatrix} d_{111} & \cdots & d_{1J1}\,d_{112} & \cdots & d_{1J2} & d_{11T} & \cdots & d_{1JT} \\ \vdots & \ddots & \vdots\ \vdots & \ddots & \vdots\ \cdots\ \vdots & \ddots & \vdots \\ d_{I11} & \cdots & d_{IJ1}\,d_{I12} & \cdots & d_{IJ2} & d_{I1T} & \cdots & d_{IJT} \end{pmatrix} \quad \cdots\cdots (1)$$

行列要素は面内位置（XおよびY。1～I）、質量電荷比（m/z。1～J）、深さ（Z。1～T）の3つの指標で表されている。つまりd_{ijt}は面内位置i、質量電荷比j、深さtに対応するデータである。MCRではXYの2次元を1次元にまとめて表現することが一般的である。元データは、ToF-SIMS分析の信号強度がポアソン統計に基づくことを利用したポアソン補正をかけて用いた。後に示す本論文の事例では、質量電荷比1～500におけるユニットマス圧縮を用いたためJは500である。

2段階MCRでは、**式 (1)** の4次元データを次のように近似する。

$$d_{ijt} \cong \sum_{k=1}^{K}\sum_{l=1}^{L} c_{ik}s_{jl}f_{tl}^{(k)} \quad \cdots\cdots (2)$$

ここでc_{ik}、s_{jl}、$f_{tl}^{(k)}$はそれぞれ「ユニット」kの面内分布、化学種lのスペクトル、化学種lのユニットkにおける深さプロファイルである。この「ユニット」とは、深さプロファイル（全ての質量電荷比に関するZ方向依存性）およびその面内方向重みづけの組に相当する。ここでKおよびLはそれぞれ、ユニットの数および化学種の数である。

これら2つの分解パラメータKとLこそが本手法の特徴であり、形式的には似た手法であるパラレルファクター解析(parallel factor analysis: PARAFAC)[(3)]では用いられていないものである。PARAFACはテンソル分解法の1種であり、**式 (1)** のデータを次のように近似する。

$$d_{ijt} \cong \sum_{l=1}^{L} c_{il}s_{jl}f_{tl} \quad \cdots\cdots (3)$$

ここでc_{il}、s_{jl}、f_{tl}はそれぞれ成分lの面内分布、スペクトル、深さプロファイルである。**式 (3)** のlは**式 (2)** のl同様、試料中の「化学種」の指標とみなすことができる。2段階MCR法は2つの独立したパラメータ（KとL）のおかげで、

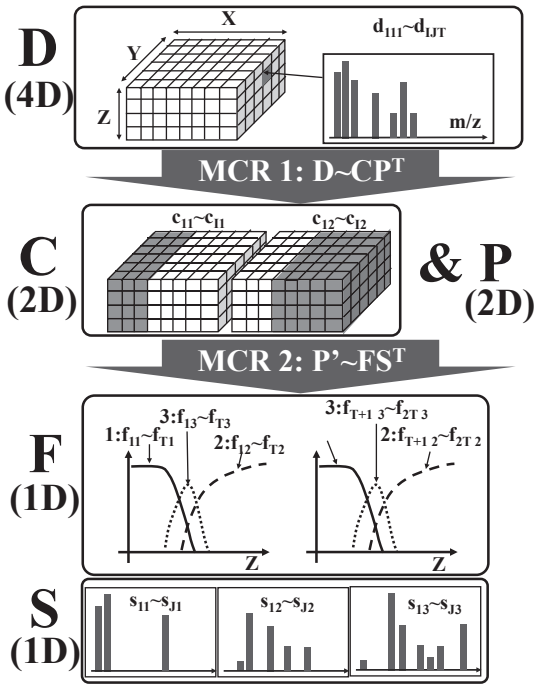


図3　2段階MCRの概要

げでPARAFACに比べ系を表現する自由度が高く、例えば試料中の複雑な化学種の分布を表現するのに有利である。

図3に2段階MCRの概要を示す。図中の“c_{11}”，“$f_{T+1\,3}$”，“s_{J2}”などの行列成分はそれぞれ後述の**式 (4)**、**(6)** のそれに対応する。ここでは簡単にするため、**式 (4)** の$K=2$、**式 (6)** の$L=3$の場合を示している。元のデータ行列Dから、集約された行列C, F, Sが2回のMCR計算により順番に得られる。行列分解は交互最小二乗法によって行う[2),3)]。最初に、ポアソン補正を施した**式 (1)** の行列Dを以下のように2つの小サイズの行列の積に分解する。

$$D \sim CP^T$$

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1K} \\ \vdots & \ddots & \vdots \\ c_{I1} & \cdots & c_{IK} \end{pmatrix}$$

$$P^T = \begin{pmatrix} p_{1\,1} & \cdots & p_{J\,1} & p_{J+1\,1} & \cdots & p_{J\times 2\,1} & p_{J\times(T-1)+1\,1} & \cdots & p_{J\times T\,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \ddots & \vdots \\ p_{1\,K} & \cdots & p_{J\,K} & p_{J+1\,K} & \cdots & p_{J\times 2\,K} & p_{J\times(T-1)+1\,K} & \cdots & p_{J\times T\,K} \end{pmatrix} \quad \cdots\cdots (4)$$

ここで行列Cは面内濃度分布、行列Pは全質量電荷比に関する深さプロファイルである。**式 (4)** の分解の後行列Cを2値化し、Pもそれに応じて修正する。この2値化の重要性および具体的な手順については後述する。

式 (4) の計算は多くの場合時間がかかり、また大量のメモリを消費する。そのためデータの間引きなどの処理がしばしば必要となる。

行列Pを追加のMCRによってさらに分解するが、その前にPを以下のように変形しておく。これは次のMCR処理によって化学種を抽出できるようにするためである。

$$P \rightarrow P' = \begin{pmatrix} p_{1\,1} & \cdots & p_{J\,1} \\ p_{J+1\,1} & \cdots & p_{J\times 2\,1} \\ \vdots & & \vdots \\ p_{J\times(T-1)+1\,1} & \cdots & p_{J\times T\,1} \\ \vdots & & \\ p_{1\,K} & \cdots & p_{J\,K} \\ p_{J+1\,K} & \cdots & p_{J\times 2\,K} \\ \vdots & & \vdots \\ p_{J\times(T-1)+1\,K} & \cdots & p_{J\times T\,K} \end{pmatrix} \quad \cdots\cdots (5)$$

P'のj列はj番目の質量電荷比（m/z）に対応することに注目されたい。この変形されたP'を以下のように、2つの小サイズの行列の積に分解する。

$$P' \sim \begin{pmatrix} f_{1\,1} & \cdots & f_{1\,L} \\ f_{2\,1} & \cdots & f_{2\,L} \\ \vdots & \vdots & \vdots \\ f_{T\,1} & \cdots & f_{T\,L} \\ \vdots & & \\ f_{T\times(K-1)+1\,1} & \cdots & f_{T\times(K-1)+1\,L} \\ f_{T\times(K-1)+2\,1} & \cdots & f_{T\times(K-1)+2\,L} \\ \vdots & \vdots & \vdots \\ f_{T\times K\,1} & \cdots & f_{T\times K\,L} \end{pmatrix} \begin{pmatrix} s_{11} & \cdots & s_{J1} \\ \vdots & \ddots & \vdots \\ s_{1L} & \cdots & s_{JL} \end{pmatrix} = FS^T \quad \cdots\cdots (6)$$

ここで行列Sは化学種のスペクトル、行列Fは化学種の深さプロファイルに相当する。Lは試料の本質部分を特徴づけるのに必要な化学種の数に相当する。

行列Fは全てのK個のユニットにおける深さプロファイルを含んでいるため、次のように表現できる。

$$F = \begin{pmatrix} F^{(1)} \\ F^{(2)} \\ \vdots \\ F^{(K)} \end{pmatrix} \quad F^{(k)} = \begin{pmatrix} f_{11}^{(k)} & \cdots & f_{1L}^{(k)} \\ f_{21}^{(k)} & \cdots & f_{1L}^{(k)} \\ \vdots & \vdots & \vdots \\ f_{T1}^{(k)} & \cdots & f_{TL}^{(k)} \end{pmatrix}, \quad \cdots\cdots\cdots (7)$$

ここでサブ行列$F^{(k)}$はユニットkにおける深さプロファイルに相当する。

上記の手続きによって、元の4次元データは面内濃度分布・化学種のスペクトル・化学種の深さプロファイルという3つの行列で近似表現された。近似表現に必要な領域および化学種の数（**式 (4)** のKおよび**式 (6)** のL）は後の事例でも示すように、多くの場合10以下であり、これによって4次元データを限られたスペースの平面上に表現することが可能となる。

本論文内では、2回の分解にMCRのアルゴリズムを用いた。他のアルゴリズム、例えば非負値行列因子分解[1)]を用いることもでき、類似の結果が得られる。MCRも、非負値行列因子分解も、近似による残差行列D-CP^Tの成分2乗和を最小化することを目的としているためである。

2−2　行列Cの2値化

MCRでは、行列Cは一般に各々の列において複数の非ゼロ値をもつ。これは異なる「ユニット」が互いにオーバーラップすることを意味する。単一のMCR解析ではこのことは単に「化学種の混合」に相当し、データを直感的に解釈する上での困難は生じない。しかし2段階MCRでは近似分解を2回行うため、面内方向と深さ方向で2段階の重ね合わせが生じることになり、データの直感的な解釈が極めて困難となる。

表現の簡潔さを向上させるため、行列Cを2値化し、異なるユニットの重ね合わせを禁止する。言い換えると、行列Cは全ての列に関して1つだけ「1」の値をもち、他はすべて「0」になるようにする。この2値化は例えばK-平均法によって行うことができる。対応する行列Pは通常の最小二乗法によって修正する。後の**図5**に示すように、2段階MCRの結果表示において、行列Cの面内分布は「1」か「0」のみで示される。

MCR自体は、成分の重ね合わせを許容する「ソフトクラスタリング」に分類される。一方で、上記で述べたMCRとその後の2値化を組み合わせたものは、重ね合わせを禁止する「ハードクラスタリング」とみなせる。2段階MCRでは、面内方向にはハードクラスタリング、深さ方向にはソフトクラスタリングを行うことで、4次元データの直感的な把握を可能にしている。

3. 2段階MCRによるデータ解析のBZY薄膜への適用事例

本手法の適用事例として、固体酸化物形燃料電池の電解質に用いる有望な材料である$BaZr_{1-x}Y_xO_{3-\delta}$（BZY）の解析について紹介する。ToF-SIMS分析を行ったテストサンプルは**図4 (a)** に示すようなBZY-A層（100 nm）/BZY-B層（>1 μm）/NiO+BZY基板という構造である。同図に矢印で示すようにNiがNiO+BZY基板からBZY層へ拡散することが知られており、その評価を実施した。

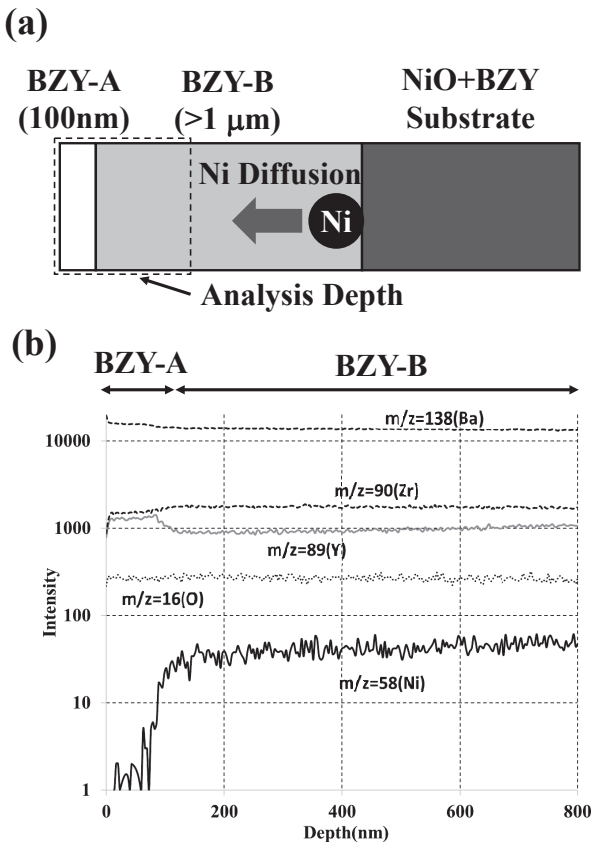


図4 （a）測定サンプルの断面構造
（b）主要なm/zに関する深さプロファイル

測定では、アルバック・ファイ㈱製造の「PHI nano TOF II」ToF-SIMS装置を用いた。30 kVのBi^{3++}イオンビームを1次イオン、2 kVのCs^+イオンビームをスパッタ用として用いた。スパッタ時間はSiO_2標準試料のスパッタレートを用いて深さに換算した。分析エリアは100 μm角とし、正イオンのデータを取得した。なお、Cs, Cs_2, Cs_3に相当するm/z＝133, 266, 399のデータは解析前に除外した。正イオンモードにおいてはCsの感度が極めて高く、これらの信号は2段階MCRの分解作業およびデータの解釈の妨げになるからである。

まずは先の**図2 (a)** で示した従来手法による解析結果を**図4 (b)** に示す。これは測定データを*XY*面に関して平均することで得られたものである。Ni（m/z＝58）はBZY-B層で確認されたが、*XY*方向の分布に関してはこの図からはわからない。

同じデータを2段階MCRで処理した結果を**図5**に示す。3つのマップは2値化された行列*C*の面内分布を示しており、白の部分がそのユニットに対応する。**図4 (b)** ではデータを*XY*面全域で平均化して表示したが、**図5**では*XY*面を3つの領域に分割して表示している。深さプロファイルは各ユニットにおける行列*F*に相当し、化学種A, B, Cの深さ方向の分布を示す。質量スペクトルは行列*S*に相当し、化学種A, B, Cの質量電荷比の内訳を示している。この例では、

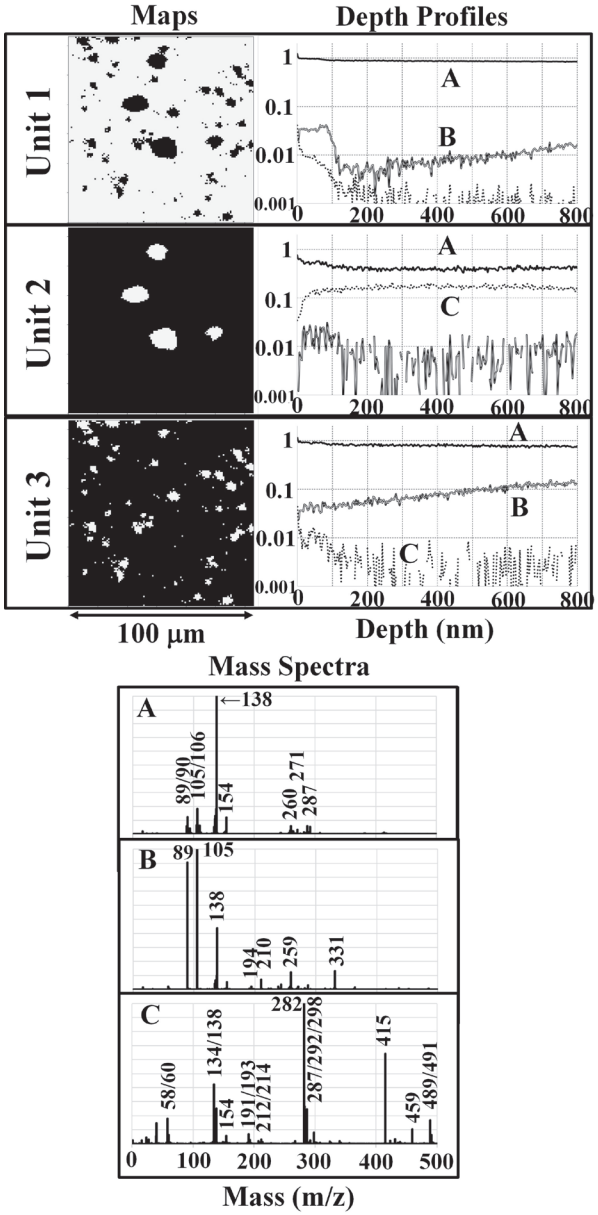


図5　BZYの測定データに関する2段階MCRの結果

3つの異なるユニットおよび化学種が自動的に抽出された。

ユニット1では、化学種Aの信号強度が深さに対してほぼ一定である。化学種Aは*m/z*＝89（Y），90（Zr），105（YO），138（Ba），154（BaO）をメインに含んでおり、これよりユニット1はBZYの骨格構造であることがわかる。

他の2つのユニットは試料の微細構造に対応する。ユニット2では、化学種Cの濃度がユニット1に比べ高い。化学種Cは*m/z*＝58（Ni），60（Ni），191（CsNi），193（CsNi），212（BaNiO），214（BaNiO），282（Cs2O），415（Cs3O），489（Cs3NiO2），491（Cs3NiO2）を含んでおり、これはNi原子が局所的なパスを通ってBZY層に拡散していることを示唆している。

ユニット3では、化学種Bの濃度がユニット1より高い。化学種Bは*m/z*＝89（Y），105（YO），194（Y_2O），210（Y_2O_2），259（$BaYO_2$），331（Y_3O_4）を含んでおり、これはY原子がBZY-B層中で凝集していることを示す。

図5の結果を、別の観点から眺めてみる。**図6**に代表的な質量電荷比（*m/z*）強度の3Dプロットを示す。*m/z*＝89、105の分布は互いに似ており、**図5**のユニット3に対応する。ここで注目すべきは**図6 (a)** のNiの分布である。今回の測定では、**図4 (b)** に示したようにNi（*m/z*＝58）の信号が他の元素の信号に比べ桁違いに弱く、これだけを見ても分布の特徴を掴むことは容易ではない。しかし、2段階MCRの結果（**図5**）よりNiの分布は強度の強いCs_2O（*m/z*＝282）のそれと同じであることが明らかとなっているため、**図6 (a)** の代わりに**図6 (e)** を見ることで間接的にNiの分布を把握することができる。

図6の6つの質量電荷比は単純に、**図5**のマススペクトルを見ることで選別した。通常、このような分布が似ている質量を見出すためには、今回の場合500個の全質量電荷比に関して3次元プロットしなければならず、これは至難の業である。2段階MCRを用いることで、もとの4次元ビッグデータから重要な3次元ユニットおよび化学種を自動的に抽出でき、材料の理解が強力にアシストされる。

このように、2段階MCRは従来見落としていた4次元データの特徴箇所の直感的な可視化を可能にしてくれるため、材料開発に有効活用することができる。

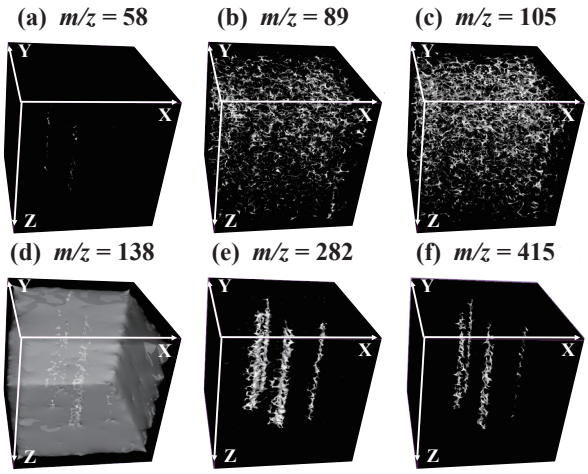


図6　BZY試料中の6つの*m/z*に関する強度分布。
このうち (d) のみが、Ba (*m/z*=138) の「空乏領域」を示すために半透明で表示されている。
この「空乏領域」は図5のUnit2に対応する。

4. 結　　言

「2段階MCR」という新たな4次元データ解析の手法を開発した。この手法をBZY薄膜のToF-SIMS分析データに適用したところ、NiおよびY原子の局所的な分布をシンプルに示すことができた。このような情報は従来の解析手法では得られないものである。

原理的には、2段階MCRはToF-SIMS分析データに限らずあらゆる4次元データ、例えばエネルギー分散型X線分析、蛍光X線分析、X線コンピュータ断層撮影法のデータなどにも適用可能である。この2段階MCRは幅広い分野のデータ解析に関する問題解決に貢献できると期待する。

用 語 集

※1　　ToF-SIMS

飛行時間型二次イオン質量分析法（Time-of-flight secondary ion mass spectrometry）。表面分析法の一種。試料に細く絞ったイオンビームを照射し試料表面からイオンを検出する。検出器までの飛行時間（Time-of-flight）によって質量分析にかける。

※2　　MCR

多変量スペクトル分解法（Multivariate curve resolution）。教師なし機械学習に基づく行列の因子分解法の一種であり、もとの高次元の行列をそれよりも次元の小さい2つの行列の積で近似する手法。

参 考 文 献

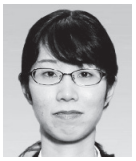
(1) N. Verbeeck, R. M. Caprioli, and R. Van de Plas, Mass Spectrometry Reviews, 2019 00, 1–47 (2019)
(2) S. Muto, T. Yoshida, and K. Tatsumi, Materials Transactions, 50 (5) 964 (2009)
(3) R. Bro, Chemometrics and Intelligent Laboratory Systems, 38 (2) 149-171 (1997)

執　筆　者

星名　豊*　：解析技術研究センター　主査
博士（工学）

上村　重明　：解析技術研究センター　主席
博士（理学）

岡本　悠　：解析技術研究センター



久保　優吾　：解析技術研究センター　主席
博士（工学）

*主執筆者