

## 多変量統計解析を利用したスペクトル・イメージの解析： 利点と問題点

### Application of Multivariate Statistical Analysis to Spectrum-Imaging Datasets: Benefits and Disadvantages

渡辺 万三志<sup>a</sup>, 石塚 和夫<sup>b</sup>

Masashi Watanabe and Kazuo Ishizuka

<sup>a</sup>Department of Materials Science and Engineering, Lehigh University

<sup>b</sup>HREM Research Inc.

**要旨** 多変量統計解析 (MSA: multivariate statistical analysis) は、最近の分析電子顕微鏡 (AEM: analytical electron microscope) で測定可能となった巨大なスペクトルイメージ (SI: spectrum imaging) を効率よく解析する上で必須の解析法の一つである。本稿では、まず、MSA 法の中で最も一般的な主成分解析 (PCA: principal component analysis) をベースに、その原理と利点を解説する。次に、PCA 法を用いることで導入される恐れがあるアーティファクトについて言及する。最後に、PCA 法によって導入されるアーティファクトを抑えるために、筆者らが開発した LocalPCA 法を紹介する。

**キーワード**：スペクトル・イメージング法、多変量統計解析法、主成分解析法、検出限界

#### 1. はじめに

近年のハード・ソフト両方面における飛躍的な発展により、電子顕微鏡による材料評価において、莫大な量のデータを比較的短時間で測定できるようになった<sup>1)</sup>。特に、電子エネルギー損失分光 (EELS) やエネルギー分散型 X 線分光 (XEDS) といった分光法では、1 秒当たり数千スペクトルを計測できる状況になっており、完全なスペクトルを 2 次元の各点において記録するスペクトル・イメージング (Spectrum Imaging: SI) 法もより一般的に適用できるようになった。この SI 法は、測定領域の情報だけでなく、エネルギー方向の情報も有しているため、(1) 元素の情報も測定後に解析でき、取りこぼしが少なくなる、(2) 複数の異なるデータ処理を適用することができ、その比較も可能である、(3) 通常のスเปクトル処理の適用も可能といった利点がある<sup>2,3)</sup>。

一方で、SI 法によって得られる情報量は非常に多く、(たとえば、128 × 128 画素で 1024 チャンネルの SI データを測定した場合、1024 枚の像あるいは 16,384 本のスペクトルが含まれる)、注意深くデータを調べたとしても、重要な情報を見逃しかねない。また、完全なスペクトルを保存するため、原理的にはバックグラウンド削除やピーク分離等の通常のスぺクトル処理が可能であるが、測定ノイズの寄与が著し

い場合、スペクトル処理によって不本意なアーティファクトを導入してしまう危険性すらある。材料評価において、より多くの情報を得ることは望ましいことではあるが、その取扱いは決して容易になったとは言えない。筆者らは、統計学の分野で利用されている多変量統計解析法 (Multivariate statistical analysis: MSA) を SI データの解析に応用することで、統計的に顕著な特徴を自動的に抽出し、測定時のランダムノイズを除去したデータを再構成する手法を開発した<sup>4)</sup>。本特集は、解析の補助となるシミュレーションが主題であるが、本稿では、解析補助としての一つとして、MSA 法の原理とその適用例を紹介し、MSA 法適用時の注意点を言及する。最後に、MSA 適用によって起こり得るアーティファクトを最小限に抑えるために、最近、筆者らが開発した LocalPCA 法も紹介する。

#### 2. 多変量統計解析 (multivariate statistical analysis: MSA) 法の原理

多変量統計解析 (MSA) 法とは、内在している変数が未知で、比較的大きなデータを統計的に取り扱う一連の手法を指し、中でも主成分解析 (Principal Component Analysis: PCA) 法は最も一般的で、より高度な MSA 法の第 1 ステップとしても利用される<sup>5,6)</sup>。PCA 法の基本概念は、データから抽出した最小限の変数で、もともとのデータを表現することである。抽出した最小限の変数こそが、データ中で繰り返し現れる主な情報、すなわち主成分である。図 1 (a) に PCA

<sup>a</sup> 5. E. Packer Ave., Bethlehem, PA18015, USA.

E-mail: masashi.watanabe@lehigh.edu

2015 年 2 月 10 日受付

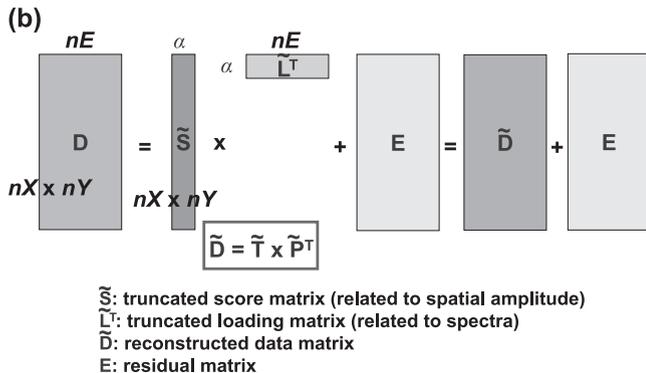
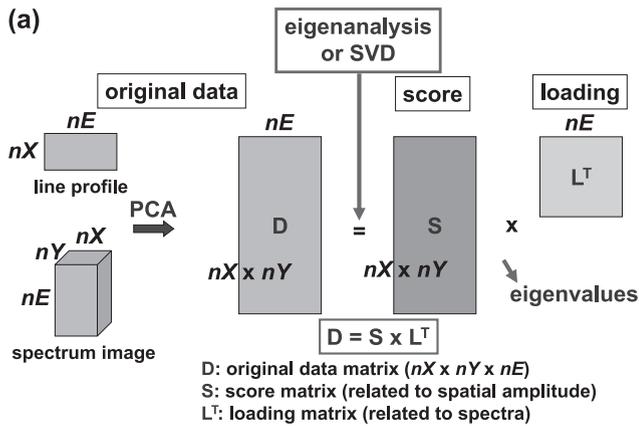


図1 PCA法の適用 (a) および PCA 法によるデータ再構成 (b) の概念図。

法によるデータ処理を模式的に示す。ライン分析あるいは SI 法によって得られるデータは、1次元あるいは2次元の空間の各測定点におけるスペクトルの集合である。このデータは、位置 (x, y) とエネルギーチャンネル (E) のデータ行列  $D_{(x,y), E}$  に展開でき、3次元のデータキューブであってもいわゆる 2way 問題として取り扱うことができる。このデータ行列  $D_{(x,y), E}$  に、前処理としてデータ測定時の計測ノイズを示す Poisson 分布による“重み”をかける (weighting あるいは scaling<sup>7)</sup>)。この Poisson-scaling 処理を行わなければ、以降の PCA 処理の際に、強度が高い特性 X 線ピークや EELS 低エネルギー損失側のバックグラウンドでのわずかな変動を示す情報が主要な成分として表れ、本来、データを表現するはずの主成分が隠蔽される危険性がある。Poisson-scaling に関しては、Keenan と Kotula による論文<sup>8)</sup>を参照されたい。前処理の後、データ行列  $D_{(x,y), E}$  は、固有値分解あるいは特異値分解によって、以下の行列に分解できる。

$$D_{(x,y), E} = S_{(x,y), n} \times L_{(E, n)}^T$$

ここで、 $L_{(E, n)}$  および  $S_{(x,y), n}$  は、それぞれ loading 行列および score 行列と呼ばれ、上付き文字 T は、loading 行列  $L_{(E, n)}$  が転置行列であることを示している。PCA 法では顕著な特徴は成分 (component) と呼ばれ、この分解によって、loading 行列 L の各列には、データ中に含まれる特徴的なスペクトル成分が格納される。また、score 行列は、対応するスペク

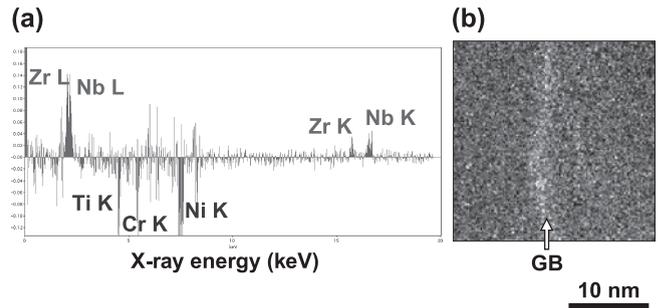


図2 Ni 基超耐熱合金粒界近傍で測定した X 線 SI データに、PCA 法を適用した例。(a) PCA 法で得られた loading スペクトルの一つ、(b) 対応する Score 像。Zr と Nb の粒界共偏析を示している<sup>9)</sup> (Cambridge Journals の再掲載許可を得た)。

トル成分の空間に対する増幅の情報を有している。

一例として、収差補正 STEM によって、Ni 基超耐熱合金の X 線マッピングを行い、上述の PCA 法により解析を行った結果を図2に示す<sup>9)</sup>。図2 (a) は、得られた loading スペクトルの一つで、Zr や Nb からの特性 X 線が正に、母相を構成する Ni, Cr および Ti は負に偏帰している。正あるいは負に偏帰した情報は、それらの信号が増幅あるいは減少したことを表しており、それが顕著に現れているのが図2 (b) の明るい部分である。この位置には結晶粒界が存在するため、図2が示しているのは、結晶粒界における Zr と Nb の共偏析である。なお、本材料では、Zr の公称組成は、わずかに 0.04wt% であり、通常であれば、定量分析はおろか検出すら試みられない量である。Ni 基超耐熱合金における Zr 粒界偏析の報告は皆無であるが、収差補正 STEM を用いて SI データを測定し、MSA 法を適用することで、その粒界偏析が明らかになった。定量解析の結果、Zr の偏析量は約 2–3 atoms/nm<sup>2</sup> であることが分かった。これは 0.12–0.17 単原子層に相当する量である。このように、従来の処理では見逃してしまう可能性が高いような着目していない情報も、MSA 法を用いれば自動的に抽出できるのである。

### 3. MSA 法を利用した微弱信号の検出

上述のように、データ行列の分解によって、特徴的な成分を抽出できる。解析的には、データ行列の分解は、行列の階数 n まで行われるため、n 種類の成分が得られることになるが、PCA 法では、最小限の成分数で元のデータ行列を表現するため、成分数の確定は極めて重要である。データの表現に必要な最小限の主成分 (principal component) とそれ以外の成分とに分離するには、データ行列  $D_{(x,y), E}$  の固有値を対応する成分に対してプロットした、いわゆる Scree プロットを用いるのが一般的である。scree プロットとは、各成分がデータ中で繰り返される頻度を示すヒストグラムと考えてよく、データ中に現れる頻度が高ければその成分はデータを表現するのに必要な主成分であり、逆に頻度が低ければ、その成分は繰り返し表れない情報、すなわちランダムノイズ成分として主成分から分離できる。図3 (a) に、Scree プロットの

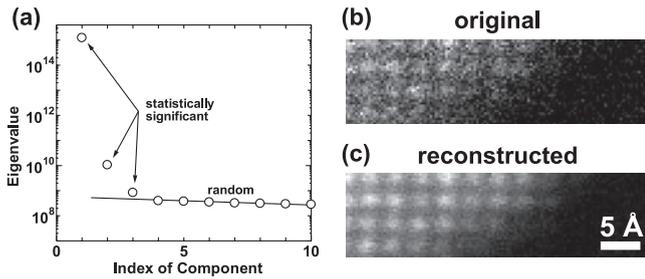


図3 (a) PCA法によって得られたScreeプロットの1例。(b)、(c) PCA法によってノイズ除去を行った例。

一例を示す。本例では、第3成分までが主成分と判断できる。ただし、Screeプロットによる判別は、あくまでも目安として使用すべきであり、最終的な判断は、各成分のloadingとscoreを吟味しなければならない。なお、通常のSIデータでは、主成分数 $\alpha$ は $n$ よりも十分に小さい( $\alpha < n$ )。

主成分数 $\alpha$ が決まれば、図1(b)に示すように、ランダムノイズ成分を分離した主成分のみで新たなデータ行列を再構成し、ランダムノイズを軽減することができる。図3(b)および(c)は、Ni基超耐熱合金中の母相/析出相の界面近傍でEELSによって測定した原子カラムTiマップであり、(b)はオリジナル、(c)はPCAによってノイズ軽減したものである。ランダムノイズが効果的に軽減されていることが分かる。各測定点での計測時間が限られているSIデータであっても、PCA法ではデータ全体をまとめて評価するため、ノイズ軽減は非常に有効である。特に、ランダムノイズに隠蔽されがちな微弱な信号は、ノイズ軽減処理によって明らかにできる場合もある。また、データビニング(binning)によって信号の計測強度を改善する方法とは異なり、データの空間分解能やエネルギー分解能を損なうこともない。ただし、主成分数の評価が不十分だと、データの再構成に主成分までも削除してしまいかねないため、主成分とランダムノイズ成分の判別は、慎重に行わなければならない。

#### 4. PCA法の適用限界

PCA法を初めとするMSAによるSIデータの解析は、非常に有効である。一方で、ノイズが非常に高い場合には、MSAによってアーティファクトが導入される危険性があるという指摘もある<sup>10</sup>。ここでは、PCA法が安全に適用できるノイズレベルに関して、シミュレートしたSIデータを用いて調べ、適用限界の指標を検討する。

薄膜試料からのX線スペクトルは、NIST Desktop Spectrum Analyzer (DTSA)を用いてシミュレートした。この際、試料は、膜厚40 nmのCu-Mn合金とし、装置および実験条件は、加速電圧200 kV、ビーム電流0.2 nA、取り込み立体角0.5 sr、および測定時間0.2 sとした。この実験条件は、AEMにおける典型的なSI取り込み条件である。シミュレートしたX線スペクトルは、図4(a)に示すように、ランダムなPoissonノイズを加え、10×10ピクセルのSIデータを構

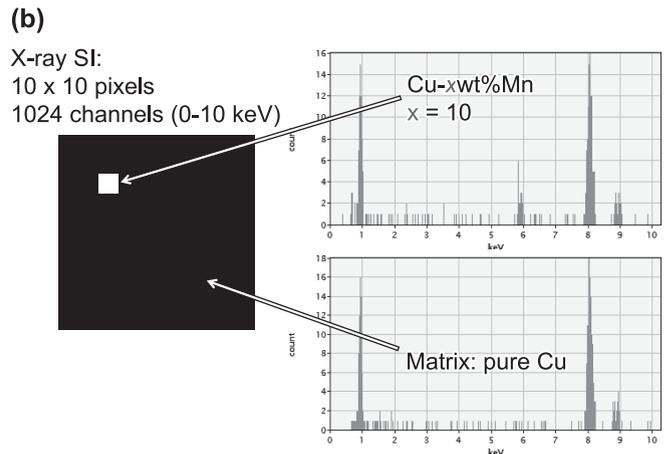
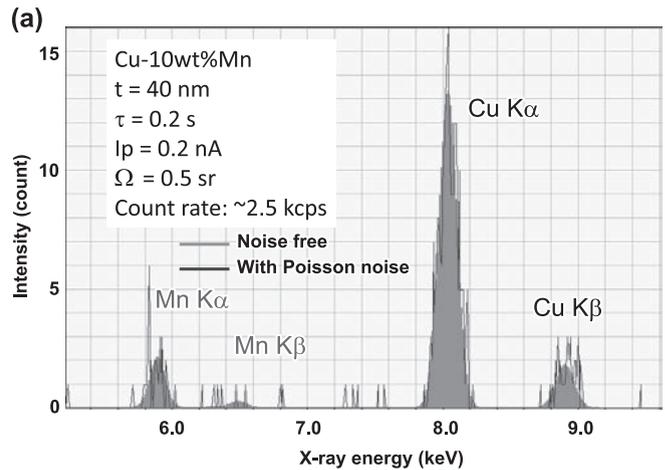


図4 (a) シミュレーションで得られたCu-10wt%MnのX線スペクトル。実線のスペクトルは、Poissonノイズを付加したもの。(b) シミュレーションで得られたX線スペクトルを元に構築したSIデータの模式図。

築した。この際、図4(b)に示すように、1点だけ、Mnの組成を2、5、10および20wt%と変え、残りの99ピクセルは純Cuとした。これは、100ピクセルから成るSIにおいて、1点のわずかな組成の変動をPCA法で検出できるかどうかを確認するための非常に厳しい設定であり、SI中の変数は、原理的には母相と組成変動箇所の二つとなる。なお、Mnの組成が10wt%以上であれば、Mn K $\alpha$ ピークは、ノイズの変動よりも大きく、2wt%ではノイズ変動よりも小さい。5wt%の場合、Mn K $\alpha$ ピークの大きさは、ノイズ変動と同レベルである。

シミュレートしたSIデータにPCA法を適用した結果、100ピクセル中の1点の組成変動であっても、信号の大きさがノイズレベルを上回っている10および20wt%の場合、Mn組成変動を示す二つ目の成分を検出できた。しかし、Mn K $\alpha$ ピーク変動がノイズレベルかそれ以下の場合、PCAを適用しても検出はできなかった。PCA法は、統計的見地から、データを表現する主成分かどうかを判別するため、真の信号変動がノイズ程度であれば、それはランダムノイズと見なされてしまうのである。

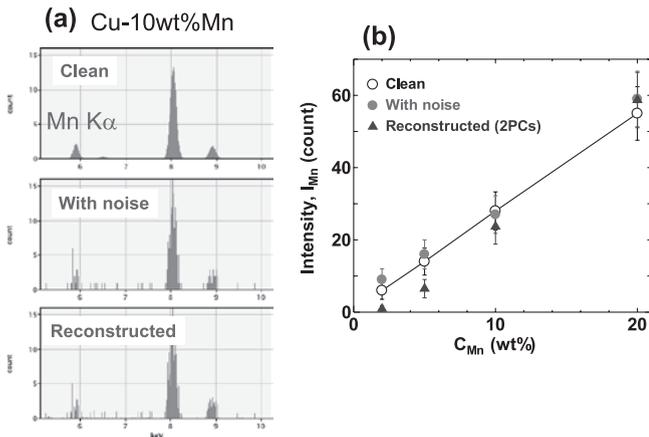


図5 (a) Cu-10wt%MnからのX線スペクトルを比較したもので、上からノイズを含まないもの、Poissonノイズを付加したもの、Poissonノイズを付加したのち、PCA法によって再構成したSIデータから抽出したものを示す。(b) 各組成を有するSIデータにおいて、PCA法による再構成後のMn K $\alpha$ 線の強度と元の強度との比較。

図5(a)は、Cu-10wt%MnからのX線スペクトルを比較したもので、上からノイズを含まないもの、Poissonノイズを付加したもの、Poissonノイズを付加したのち、PCA法によって再構成したSIデータから抽出したものである。PCA法によって明確に検出できる場合、ノイズ軽減を行い、再構築したデータであっても、ピーク強度に変化は見られない。図5(b)は、各組成を有するSIデータにおいて、PCA法による再構成後のMn K $\alpha$ 線の強度を元のスペクトルの強度と比較したものである。なお、再構成は、成分数2で行った。PCAで信号強度の変動が確認できた10wt%以上の場合、Mn K $\alpha$ 線の強度は、元のデータでの強度と一致している。しかし、信号の変動が、ノイズレベルになれば、再構成後に明確な強度低下がみられる。すなわち、PCAによる再構築で、真の信号強度の一部もノイズとともに除去されてしまったのである。

以上のように、SI中での信号の変動がランダムノイズレベルあるいはそれ以下であれば、PCAによる検出が難しいどころか、ノイズを低減した再構成時に信号強度も損なわれる可能性がある。特に、信号強度の低下は、定量分析を行う上で、致命的なアーティファクトである。したがって、PCAによるノイズ軽減を行ったSIを元に定量分析を行う場合、着目している成分がランダムノイズに近いものは、PCAによるノイズ軽減によって、信号強度も損なわれるということを理解しておくべきである。

### 5. PCA法の適用限界を改善するには

PCA法の検出感度には限界があり、アーティファクト導入の危険性もある。PCA法の検出感度を改善するには、(1)元のSIデータにおいて、ランダムノイズ成分を低減する、あるいは(2)真の信号変動を強調するといった方法が考えられる。前者のアプローチとしては、測定時間を延長するあるいは入射ビーム電流を増加することで測定する信号量を改

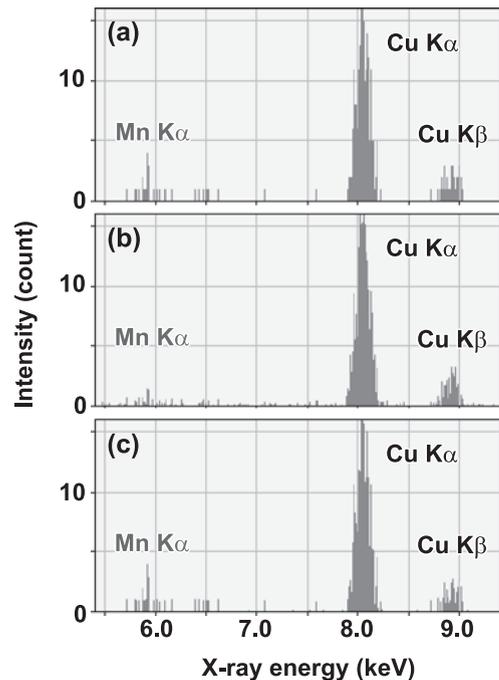


図6 Cu-5wt%Mnからのスペクトルを比較したもので、上から、(a) Poissonノイズを付加したもの、(b) 元および(c) 分割したSIデータにPCA法を適用し再構成したSIデータから抽出したものを示す。

善することで、本質的にランダムノイズを低減することができる。この場合、実験条件の変更が必須であり、再度データを測定しなければならない。後者の方法としては、より小さなセグメントに分割したSIデータにPCAを適用し、微弱な信号変動を検出する方法が考えられる。

上述のシミュレート実験では、100ピクセル中の1点だけの信号変動という非常に厳しい条件であったが、例えばこれが10ピクセル中の1点の信号変動であれば、相対的に信号変動が起きているデータ点が増加することになり、変動そのものを強調できる場合もある。上述の例で検出できなかった5wt%Mnからのスペクトルを含むSIデータを4分割して、25ピクセルのSIデータとし、PCA法を適用すると、信号変動がランダムノイズとほぼ同程度であった信号変動が検出できるようになった。

図6は、Cu-5wt%Mnからのスペクトルを比較したものであり、上からPoissonノイズを付加したもの、元および分割したSIデータにPCA法を適用し再構成したSIデータから抽出したものを示している。元のデータにPCAを適用すると、Mn K $\alpha$ 線の強度は低下するが、分割したデータの場合、Mn K $\alpha$ 線の強度変化はほとんど確認できない。実際、分割し、PCAによって再構築したデータからMn K $\alpha$ 線の強度を測定すると、 $15.7 \pm 4.0$  カウントとなり、ノイズ軽減する前の強度である  $16.0 \pm 4.0$  カウントと一致することが分かった。なお、元のデータに通常のPCAによるノイズ軽減を行うとMn K $\alpha$ 線の強度は、 $6.5 \pm 2.5$  カウントまで低下する(図5b参照)。データ中の真の信号変動がランダムノイズと同等で

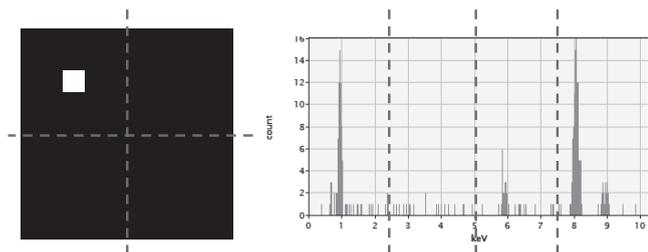


図7 LocalPCA法の概念図. (a) 位置分割 (Spatial LocalPCA) および (b) エネルギーチャンネル分割 (Spectral LocalPCA).

あれば、SI データをより小さなセグメントに分割し、相対的に信号変動のデータ点を増やすことで、PCA 法の適用限界を改善することができるのである。

筆者らは、この概念を発展させ、PCA 法の適用限界を改善した LocalPCA 法を開発した。これは、元の SI データを PCA 適用時に、より小さなセグメントに分割し、PCA の適用感度を改善する方法である。分割は、位置 (Spatial) あるいはエネルギーチャンネル (Spectral) で行うことができ、特にエネルギーチャンネルによる分割方法は、データ中での強度が極端に変動する EELS SI データに効果的である。

## 6. 終わりに

本稿では、今回の特集の一環として、PCA 法を主体とした MSA の原理と SI データへの応用を紹介させていただいた。PCA 法を利用した SI データの解析は非常に有効であるが、理解が不十分なままでの適用は、思いもよらぬアーティファクトを導入してしまう危険性を伴っている。データを再構築する際には、各成分をきちんと吟味し、顕著な主成分を判別しなければならない。また、SI に含まれる本質的な特徴であっても、変動の振幅がライダムノイズのそれと同程度ならば、PCA では抽出できない場合もある。本稿では、従来の PCA 法の感度を改善するために、新たに開発した LocalPCA についても言及した。他の効果的な手法やソフトウェアと同

様に、効果が高いものは、使い方によっては、薬にも毒にもなりうるのである。本稿が、PCA 法使用時の一助となれば、開発者の一人として、幸いである。

なお、本稿で紹介した PCA および LocalPCA を含む MSA プログラムは、Gatan Digital Micrograph 上のプラグインとして、現在、HREM Research より商品化されている。興味がある方は、HREM Research のウェブページ<sup>11)</sup>も参照されたい。

## 謝 辞

本研究の一部は、米国 National Science Foundation (DMR-0804528, DMR-1040229) の研究補助金を受けて実施したものである。なお、本稿中の図2は Cambridge Journals の再掲載許可を得た。最後に、本研究を遂行するに当たり、David Williams 教授 (現 Ohio State Univ.) および Christopher Kiely 教授 (Lehigh Univ.) の協力を得た。敬意を表する。

## 文 献

- 1) たとえば Watanabe, M.: *Microscopy*, **62**, 217–241 (2013)
- 2) Jeanguillaume, C. and Colliex, C.: *Ultramicrosc.*, **28**, 252–257 (1989)
- 3) Hunt, J.A. and Williams, D.B.: *Ultramicrosc.*, **38**, 47–73 (1991)
- 4) Watanabe, M., Okunishi, E. and Ishizuka, K.: *Microscopy and Analysis*, **23**, Issue 7, 5–7 (2009)
- 5) Jolliffe, I.T.: *Principal Component Analysis*, 2nd ed., Springer, New York (2002)
- 6) Malinowski, E.R.: *Factor Analysis in Chemistry*, 3rd ed., Wiley, New York (2002)
- 7) Cochran, R.N. and Home, F.H.: *Anal. Chem.*, **49**, 846–853 (1977)
- 8) Keenan, M.R. and Kotula, P.G.: *Surf. Interface Anal.*, **36**, 203–212 (2004)
- 9) Watanabe, M., Ackland, D.W., Burrows, A., Kiely, C.J., Williams, D.B., Krivanek, O.L., Dellby, N., Murfitt, M.F. and Szilagy, Z.: *Microsc. Microanal.*, **12**, 515–526 (2003)
- 10) Lichtert, S. and Verbeeck, J.: *Ultramicrosc.*, **125**, 35–42 (2013)
- 11) <http://www.hremresearch.com/Eng/plugin/MSAEng.html>