

三次元形式の分析データに対応した多変量解析ソフト 3D SpectAlyze® の紹介と応用例

Multivariate Analysis Software Corresponding to Analysis Data of Three Dimensional Form
Introduction and Application Example of 3D SpectAlyze®

佐藤 拓哉^{*1}, 伊藤 千晶^{*1}, 藤宮 仁^{*1}, 堀込 純^{*2}, 坂元 秀之^{*2}

1. 背景

品質管理の現場では、分析装置から出力された連続的・経時的な測定データをもとに、試料の定量的定性的な評価が行われます。正確な情報を得るために、反復計測や複数条件下での測定が必要となります。測定結果である多変量(多次元)分析データには、成分や量など多くの情報が内在しています。データの特徴によるグループ分けや因果関係の調査のためには、多変量解析によるデータの特徴の把握が不可欠です。

近年、分析装置の高性能・高集積化もあって、多変量解析ソフトウェアの需要が高まっています。ただ、既存の汎用ソフトウェアには様々な手法による解析機能が搭載されていますが、利用者にとっては手法の選択が最初のハードルとなります。

そこで、筆者らは、多次元の分析データに特化した多変量解析ソフトウェア『3D SpectAlyze®』を開発しました。処理する対象を分光蛍光光度計やHPLCの測定データに特化することで、データの取り込みから解析までをシンプルな操作性でユーザーが直観的に利用できることを目指しました。

今回は、3D SpectAlyze®の機能と利用例を紹介いたします。

2. 多変量解析ソフトウェア3D SpectAlyze®について

3D SpectAlyze®は、分光蛍光光度計より得られた三次元蛍光スペクトル(蛍光指紋)や、HPLCのクロマトグラム、またダイオードアレイ検出器(DAD)から得られる三次元クロマトグラム等のデータファイルを取り込み、データの加工やグラフ表示、各種統計解析(多変量解析)を行います。解析結果は画像データあるいはCSVファイルとして出力することができます(図1および図2)。

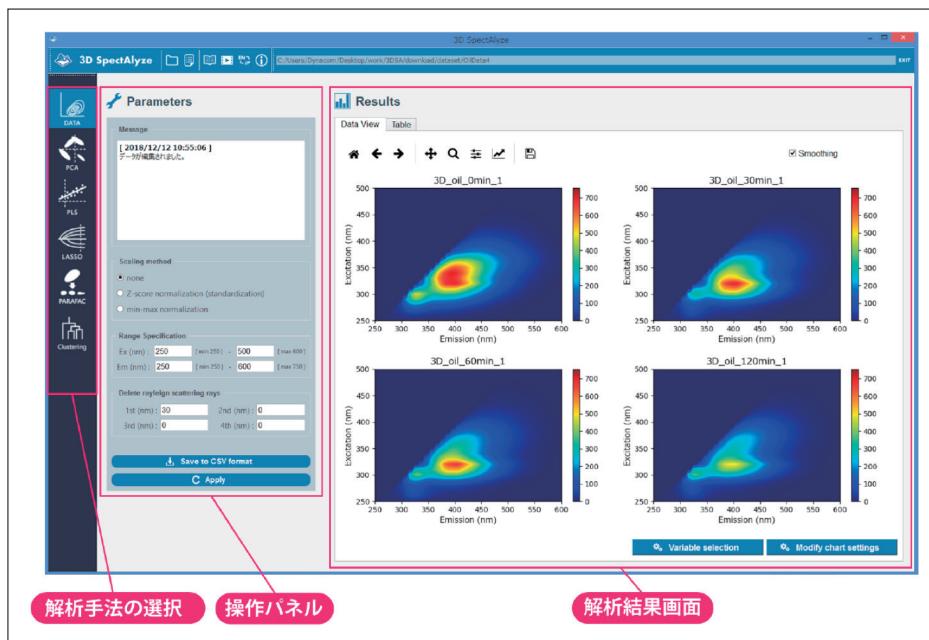


図1 3D SpectAlyze® 画面構成

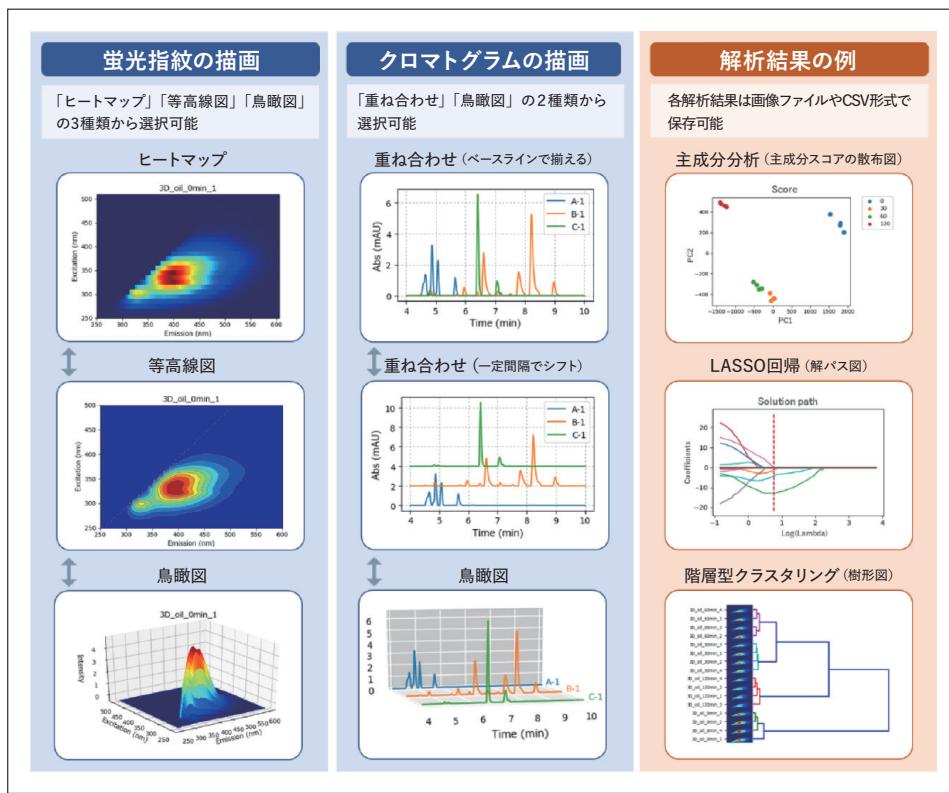


図2 3D SpectAlyze® 出力結果

搭載している主な機能は以下の通りです。

2-1. データ前処理

散乱光の除去・不要領域の除去・正規化など、分光蛍光光度計やHPLCの測定データに特化した前処理機能を搭載しています。前処理後のデータはすぐに各種統計解析画面に反映されます。

2-2. グラフ表示・視覚化

ヒートマップ・等高線図・鳥瞰図・クロマトグラムなど各種グラフ表示機能を搭載しています。

2-3. 多変量解析（教師なし）

(1) 主成分分析：分析データの特徴でグループ分けや産地判別に応用

主成分分析は、情報の損失を最小に抑えつつ、データを縮約する分析手法です。

例えば蛍光指紋の場合、1試料につき数千～数十万もの計測点（蛍光強度データ）を含んでおり、この状態では、試料間の比較が困難です。主成分分析で、項目数を2つあるいは3つにまで縮約し、散布図として描画することで、視覚的に試料間の関係性を把握できます。例えば、産地が既知／不明の試料の計測データを合わせて主成分分析を行うことで、不明試料の産地を判別することが可能です。

(2) PARAFAC：混合試料の成分分離に応用

蛍光物質が混在している試料を蛍光光度計で測定した場合、複数の蛍光物質由来のピークが重なり合った状態で観測されます。PARAFAC (Parallel Factor Analysis) はこの重なり合ったピークを試料中の成分ごとの蛍光データとして分離する手法です。分離後の各ピーク強度を比較することで、試料中の成分と含有量を推測することができます。

(3) 階層的クラスタリング：蛍光指紋の特徴でグループ分類

階層的クラスタリングは、集団の中から性質が似ている試料の分類をする手法です。2試料間で最も距離が近い（=似ている）ものからグループ化し、最終的に1つのグループになるまでグループ同士を連結していきます。連結の過程を樹形図（デンドログラム）で可視化します。

本製品では、デンドログラム上のサンプル名の位置に合わせて、蛍光指紋画像やクロマトグラムを描画するため、グループの特徴が一目でわかります。

2-4. 多変量解析（教師あり）

(1) PLS回帰：予測モデル構築、未知データの予測値算出

全ての測定データを用いて、試料の特徴を説明する項目（種類・濃度など）を予測するモデルを構築します。PLS回帰（Partial Least Squares regression）では、多次元データが持つ情報を少ない次元に圧縮してから予測式を推定するため、蛍光指紋のように測定する波長数が多く、波長間に相関があるようなデータに強いことが特徴です。

(2) LASSO回帰：寄与率の高い波長の選定、予測モデル構築、未知データの予測値算出

LASSO（Least Absolute Shrinkage and Selection Operator）では、試料の特徴を説明する項目（種類・濃度など）と関連の高い特徴点を多次元の測定値の中から自動で選定し、予測モデルを構築します。一般的に、測定したデータを学習し過ぎた（過学習）予測モデルは、未知データへの適応力（予測性能）が低下します。LASSO回帰は、これを回避するため、Cross Validationによって予測性能を確認しながら、予測への寄与率が高い波長の選択を行い、予測モデルを構築できるリーズナブルな手法です。

3. 多変量解析例の紹介

3D SpectAlyze[®]を用いた解析例として、主成分分析・PARAFAC解析による食用油の加熱変化、LASSO解析による蛍光塗料の混合比率の推定を以下に紹介します。

3-1. 食用油の加熱変化における蛍光指紋の主成分分析例

F-7100形分光蛍光光度計を用いて、加熱時間を0, 30, 60, 120分と変化させた際の食用油の三次元蛍光スペクトルを取得しました。この三次元蛍光スペクトルが図1に相当します。加熱時間の変化に伴い、蛍光指紋形状が変化していることが分かります。この変化を3D SpectAlyze[®]の主成分分析法にて解析しました。結果を図3に示します。主成分分析を行うことにより、加熱時間に伴う試料の変化を捉えることができました。図3(a)に示すスコアプロットより、0分から30分にかけて、PC1とPC2の成分の減少が生じ、その後、30分から60分の変化は少なく、60分から120分にかけてPC1の減少とPC2の上昇に伴う変化が確認されました。このとき、PC1の成分は図3(b)に示すPC1のローディングプロットより確認することができました。PC1のローディングプロットは図3(d)に示す三次元蛍光スペクトルの蛍光強度分布と一致していることから、PC1は全体の蛍光強度に相当すると考えられます。つまり、加熱時間の経過と共に、全体の蛍光強度が減少していることを反映しています。一方、PC2は、励起波長320 nm, 蛍光波長400 nm付近の値が負の値となっています。PC2の変化として、0分から30分にかけてPC2のスコアが減少していることからこの成分が増加し、60分から120分にかけてPC2のスコアが上昇していることから、この成分が減少していることが分かります。

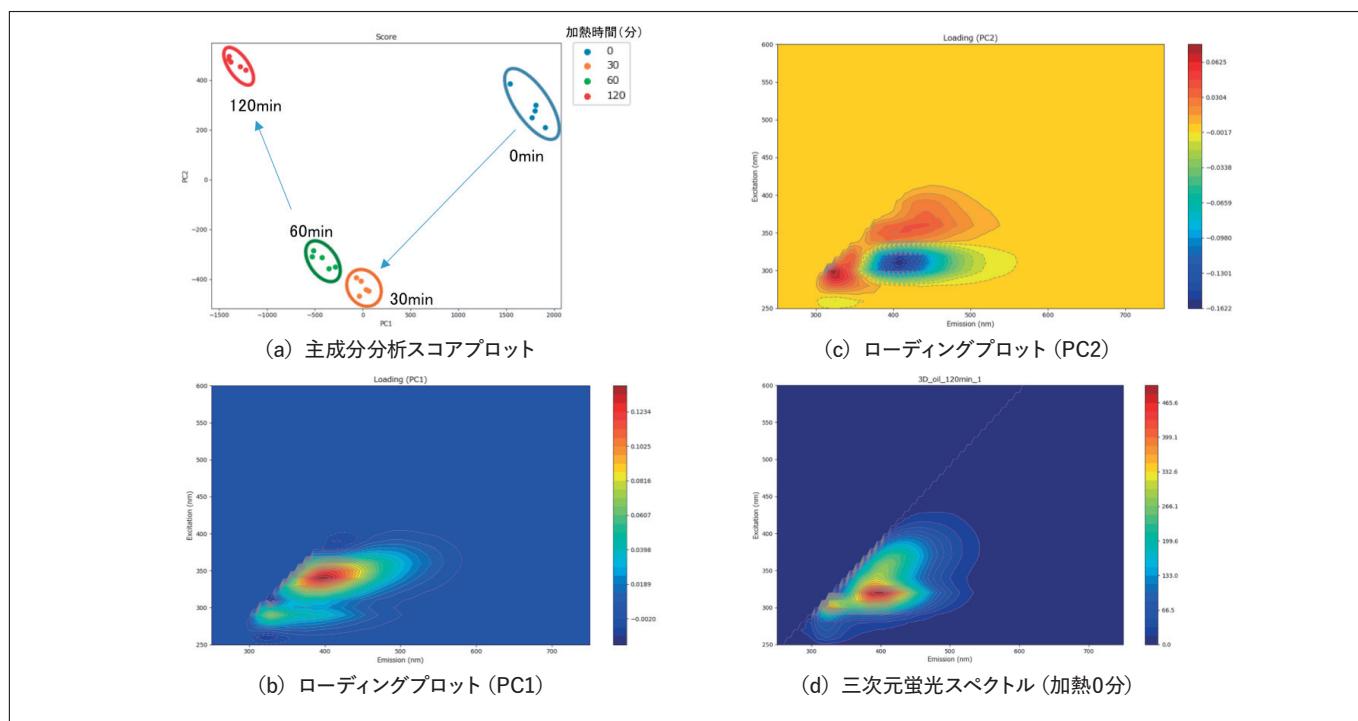


図3 主成分分析結果

このように主成分分析にて算出したスコアを散布図として描画することで、加熱時間による食用油の変化を視覚的にとらえることが可能となります。

3-2. 食用油の加熱変化における蛍光指紋のPARAFAC解析

PARAFAC解析は、蛍光指紋の重なり合ったピークを試料中の成分ごとの蛍光データとして分離する手法です。分離後の各ピーク強度を比較することで、試料中の成分や含有量を推測することができます。食用油の蛍光指紋のPARAFAC解析を図4に示します。3つの成分に分離することができました。文献¹⁾より、Component 1と3は、酸化生成物由来、Component 2はビタミンE由来と推測されました。PARAFAC解析で算出された各成分のスコア値を比較することで、成分の相対的な変化を確認することが可能となります。

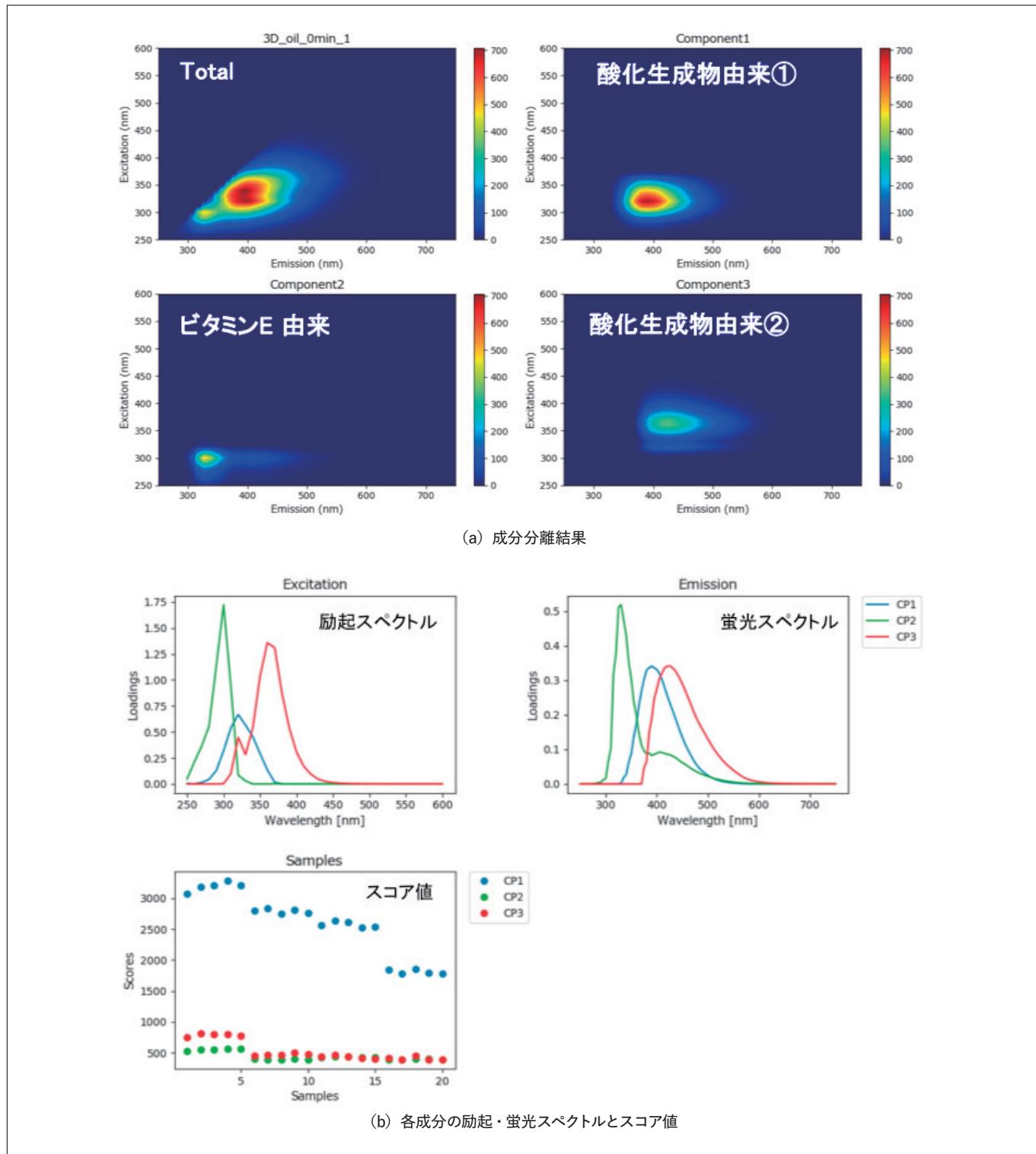


図4 PARAFAC解析結果

3-3. LASSO 解析による蛍光塗料の混合比率の推定

混合比率を変えた2種類の蛍光塗料(黄色:0%~100%, 赤色:100%~0%, それぞれ合計100%となるように混合)の蛍光指紋を測定し、LASSO 法にて混合比率を推定しました。図5に混合比率を変えた際の蛍光指紋を示します。0%は、黄色塗料が0%, 赤色塗料が100%であり、励起波長550 nm, 蛍光波長600 nm 付近に赤色塗料由来の蛍光指紋が確認されます。混合比率を変え、100%は、黄色塗料が100%, 赤色塗料が0%であり、励起波長450 nm, 荧光波長515 nm 付近に黄色塗料由来の蛍光指紋が確認されます。

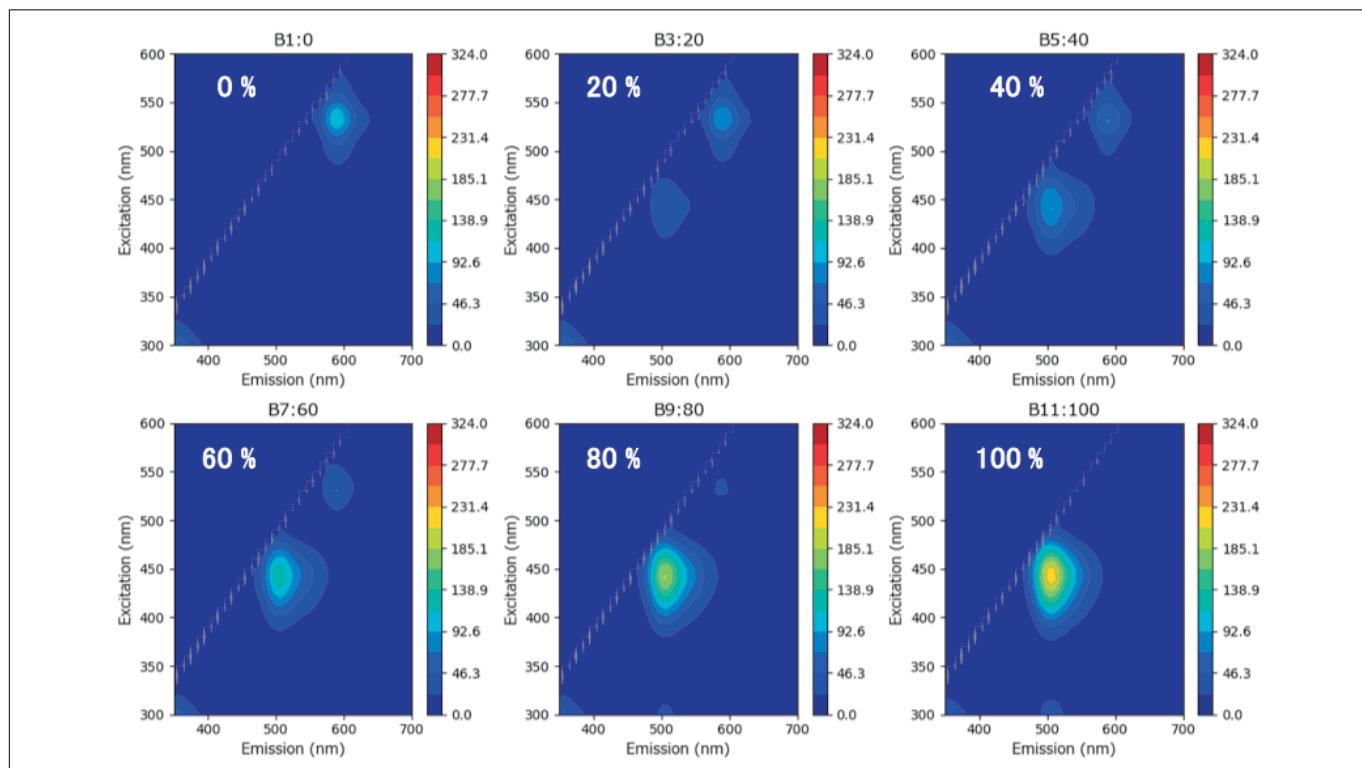


図5 混合比率を変えた蛍光塗料の三次元蛍光スペクトル

続いて、LASSO 解析にて混合比率を推定するため、各試料5回測定、合計55データのうち、3回分をモデル作成のための Calibrationデータ、2回分をモデル検証のための Validationデータに分類しました(図6 参照)。

LASSO では、試料の特徴を説明する項目(種類・濃度など)と関連の高い特徴点(波長)を多次元の測定値の中から自動で選定し、予測モデルを構築します。今回、蛍光指紋の波長数は約7,100波長ありました。LASSO 解析で混合比率を推定するために有効な波長として75波長が抽出されました。3D SpectAlyze®のLASSO 解析では、三次元蛍光スペクトルと同様に励起波長・蛍光波長・抽出された波長の係数を表示することができる、解析に有効な波長を視覚的にとらえることに有効な表示方法を備えています。

抽出された各波長に対して係数が導き出され、モデル式から混合比率を推定します。モデルを作成した既知データでの検証では、既知の値と算出した値の相関性は $R^2=0.9997$ 、モデル未使用のデータでの検証では、 $R^2=0.9992$ と良好な結果が得られました。

通常、定量値を得る際には、1つの波長に対する検量線から算出しますが、LASSO 解析や PLS 解析を用いることで、多波長の検量線を作成し、より精度の高い目的値を得ることができます。

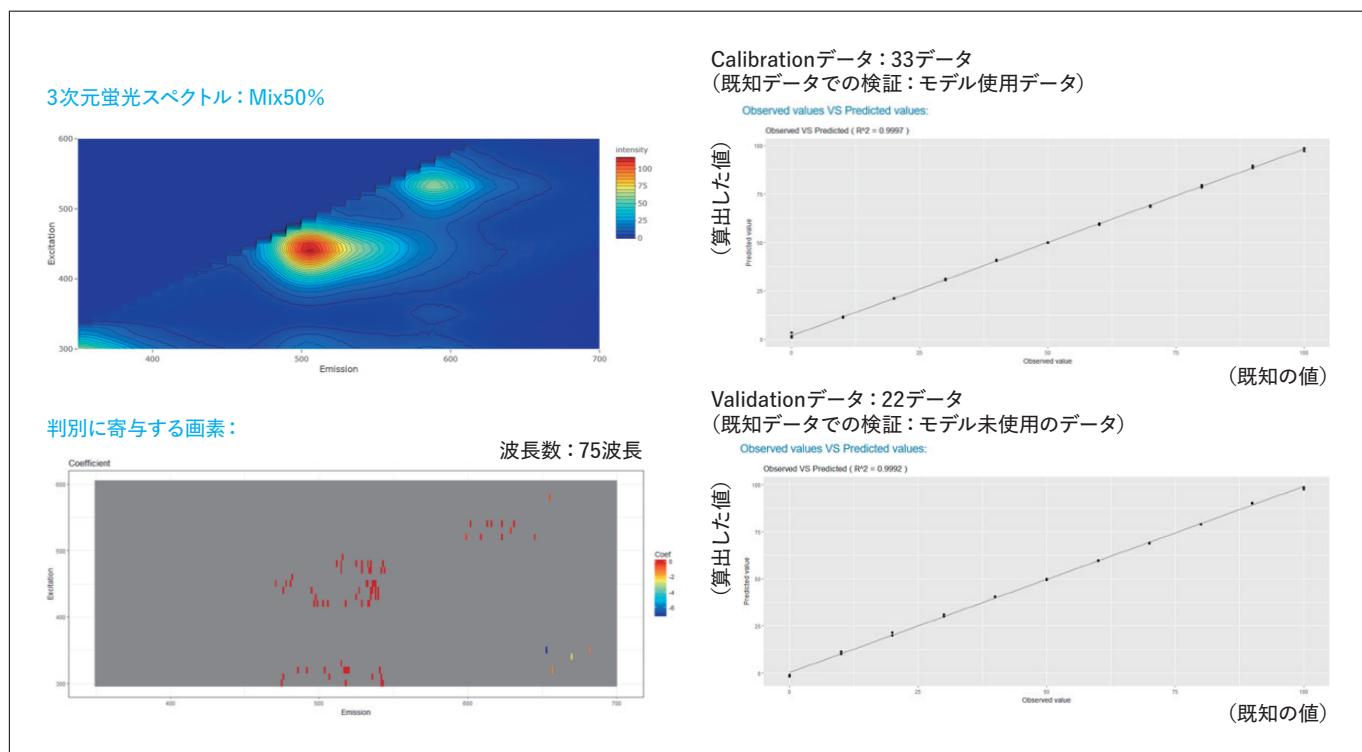


図6 LASSO法による蛍光塗料の混合比率の推定

4. まとめ

分析装置の性能は日々向上し、短時間で多くの情報が得られるようになりました。その反面、データ解析に長い時間を必要としますが、多変量解析は膨大な量のデータを短時間に明瞭にできます。3D SpectAlyze[®]は多変量解析をより簡便に扱うことができる目的としたソフトウェアで、今後幅広い分野で使われる多変量解析をより身近に扱うことができるツールです。

参考文献

- 1) J. Christensen, L. Nørgaard, R. Bro, and S. B. Engelsen, *Chemical reviews*, **106** (6), 1979 (2006).

1. “3D SpectAlyze”は、株式会社ダイナコムの日本国における登録商標です。

著者紹介

¹佐藤 拓哉, 伊藤 千晶, 藤宮 仁
(株)ダイナコム

²堀込 純, 坂元 秀之
(株)日立ハイテクサイエンス