

風味分析におけるマーカー探索ワークフロー

Workflow for searching maker compounds in Flavor analysis by GCxGC-TOFMS

樺島文恵、土屋文彦

Fumie Kabashima and Fumihiko TSUCHIYA

LECO ジャパン合同会社質量分析営業部

〒105-0014 東京都港区芝 2-13-4

TEL 03-6891-5800

E-mail : fumie_kabashima@leco.co.jp

1 はじめに

サンプルの性状、特性を決定づけるカギとなる成分「マーカー」を探索する方法論はクロマトグラフ質量分析の技術的な進歩とともに、疾患マーカーなど臨床分野で近年急速に発展してきました。ここで使用される手法とは、分析の本質が探索である以上、あらかじめ想定された分析上の設定（ターゲット分析）は極力排除され、先入観の無い網羅的な同定（ノンターゲット分析）が求められるべきであると考えられます。（図1参照）

食品の分野においても近年、匂いや味のキャラクター化に対するニーズが、品質管理や製品開発の分野で高まっています。特に価値観の多様化が進む現代の日本においては、食品に求められる要素が単なる安全性にとどまらず、高い嗜好性、栄養面における付加価値などと複雑化してきたのは当然と言えます。こうした様々な消費者のニーズに答えるためには“美味しさ”、“健康性”を客観的な数値データとして表現し、その性質をコントロールする必要があります。つまり味のデザインや成分の機能性を明らかにする事を目的とするのであれば、その食品の機能や特性の本体を明確にするために、その性質に寄与する化合物を多種多様な共存成分の中から探し出す必要があります。このような成分の探索には、分析の網羅性の高さと“カギ”となる化合物を膨大なデータから抽出し決定するために多変量解析など統計的な手法を用いた情報科学的なテクニックが必要となります。本発表では、昨年本研究会で報告したマーカー探索の内容を方法論的な立場から具体的に示し、その有効性について考察します。

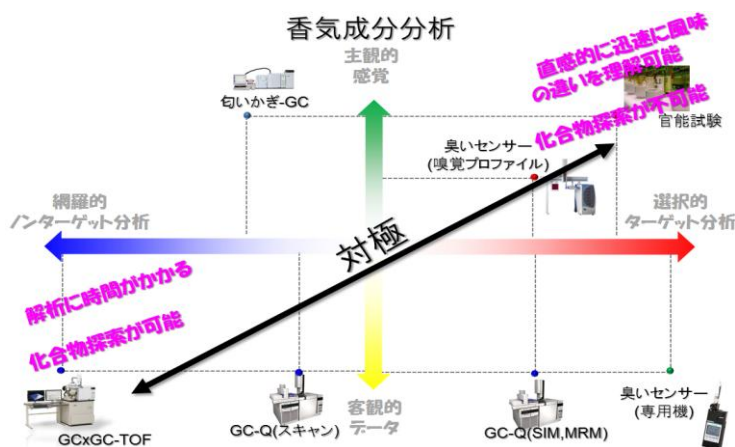


図1 GCxGC と官能試験の位置関係

2 背景

ガスクロマトグラフ質量分析装置(以下、GC-MS)は揮発性化合物の分析において定量性と分離能に優れた装置として他の分離分析技術と比べても多くの優位性を有しています。更に飛行時間型質量分析装置との組み合わせは、四重極型質量分析装置の原理的な欠点であるスペクトル取得速度の問題を解消することが可能であり、近年、定性分野において GC-MS を用いた香り成分などのマーカー探索の潜在能力を大きく向上させました。これは高速化が進む四重極型質量分析装置においても、網羅的二次元 GC-MS テクニックである GC×GC で得られる 1 ピーク(約 0.1 秒幅)に、デコンボリューションによる正確なピーク抽出に必要な 10~20 ポイント以上を与えることは困難である事実を考えれば、その価値を容易に再確認する事ができます。

探索を行う上で、ポイントとなる点を挙げるとすれば、①高い網羅性、②ピーク検出の自動化、③データの軽量化、④システマティックな情報処理技術が挙げられます。①の高い網羅性に対しては、GC×GC によるピークの多次元展開が最も有効です。この技術は通常の一次元ガスクロマトグラム上で分離が困難な共溶出するピークを、分離特性が異なる二つのカラムにより二次元展開する事で分離していく技術であり、類似した構造を有する多種の化合物が混在する複雑な系では同定されるピーク数が一次元分離に比べ 5 倍から 10 倍増加する事が期待できます。②の自動化では、近年ノンターゲット分析で一般化しつつあるデコンボリューションが欠かすことができません。ピーク分離のアルゴリズムに従い、共通の保持時間、ピーク形状と異なる mass 値により、電子イオン化法により開裂したフラグメントイオンを時間軸及び m/z 軸で再構成し、ピーク及びクロマトグラムを自動的に単離していくデコンボリューションは、香り分析で得られる数百から数千のピークの解析の効率化を可能にします。③のデータの軽量化は、複数のデータを同時に処理する統計解析において PC への負担と処理速度に直結する問題です。特に精密質量型の飛行時間型質量分析装置では、一つの GC×GC データで十数ギガバイトに達するため、統計解析に耐えるデータ数を処理するのが困難です。この場合、全ての mass データを整数に近似する事で、データサイズを 2 桁減らすことが可能です。④のシステマティックな情報処理技術は、ピークのアライメントから、ANOVA や T 検定などによるデータの評価、多変量解析と言った探索の流れをスムーズに行う事のできる支援ソフトウェアの存在が重要です。以上挙げた 4 つの条件を全て満たすシステムとして、LECO 社の Pegasus 4D® GCxGC-TOFMS システム^{1,2)} が知られています。演者らはこの GC-TOFMS を用いて、数々の食品サンプルのキャラクタリゼーションを行っており、サンプルの性状に寄与する化合物の探索は同一のワークフローにより行われています。そのプロセスについて説明します。

3 網羅的分析への戦略

1 ノンターゲット分析と GC-TOFMS

前述の通りフレーバー分析において対象とするサンプルは非常に多種多様の揮発成分を有する食品であるため、分析手法における網羅性の高さがポイントとなります。この二つの条件を可能にする分析装置としては、LECO 社の PEGASUS® 4D GCxGC TOFMS^{1,2,3} が有効です。その理由は本機のデータ取得速度(最大 500 Hz)が GC×GC 分離(0.1 秒のピーク幅)に完全に対応し、ピーク検出アルゴリズムであるデコンボリューションに必要な 1 ピーク当たり 20 ポイント以上のデータポイント(マススペクトル)を得る事を可能にする点にあります。このシステムではデータポイントの十分な密度(高い時間分解能)が確保できれば、デコンボリューション処理により、

ピーク分離、さらには同定まではほぼ自動的に行われます。これらの機能により、醸造系のサンプルでは数千のピークを検出する事が期待できます。しかし探索の目的は特徴に寄与するマーカを絞り込むことであり、ここまででは目的は達成しておりません(図 2 参照)。

マーカ化合物の探索 第一段階

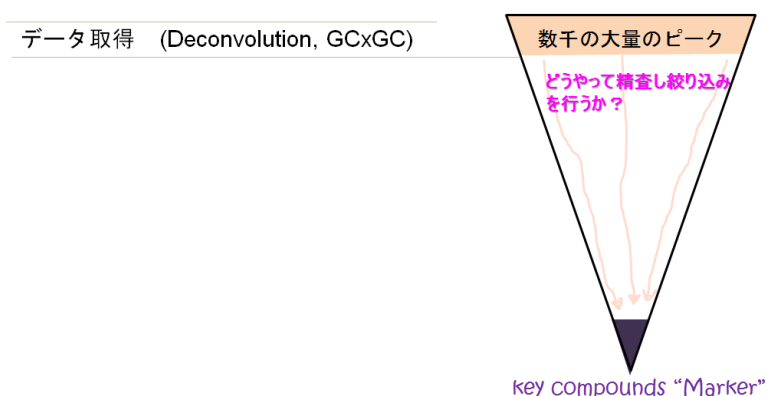


図 2 網羅的ピーク同定とマーカの絞り込みの位置関係

探索及び成分比較を行う場合、この装置により得られる膨大な化合物情報は、人間の処理能力を大きく超えてしまい直接利用する事が極めて困難ですが、その問題を解消するために統計的データ処理機能(LECO 社 **Statistical Compare** ソフトウェア)が導入されています。これは、実験回数ごとに作成されるピークリストを編集統合するための自動ピークアライメントとフィッシャーレシオ値⁴算出の二つの機能で構成されます。

2 フィッシャーレシオ値の利用

フィッシャーレシオ値は検体間で明確な有意差があり、繰り返し再現性の高い信頼できるピーク情報を、ピークリスト上の膨大なデータから抽出するための判断基準として用いられます。その仕組みは算出法にある、基本的な算出法は **ANOVA** 法と類似していますが、一言で言い表すと、ある化合物についての「検体間における分散値と各検体内での再現性に関する分散値の比」です (下式及び図 2 参照)。

$$\text{Fisher ratio} = \text{検体間分散値} / \text{検体内分散値}^{3)}$$

つまり、**Fisher Ratio** 値の算出で、分母に相当する成分は繰り返し分析の再現性を示し、再現性が高いほど分散値は低くなり、その結果 **Fisher Ratio** 値を引き上げます。また、分子に相当する成分は検体間での有意差を反映し、ある化合物 **A** に検体間で顕著な差がみられた場合、その差が大きいほど化合物 **A** の **Fisher Ratio** 値を引き上げます。そのため **Fisher Ratio** 値の高い順番に同定された化合物を並び替えると、サンプルの特徴に寄与率の高い化合物が上位に来る可能性が高くなります。この場合、再現性と検体間有意差の両方が良好な結果である場合は **Fisher Ratio** 値が高くなる事を意味し、再現性が悪い、または検体間で差がみられない化合物は効果の相殺により **Fisher Ratio** 値は順位を下げることになります。このような再現性が悪い、または検体間で存在量に差がみられない化合物は、検体間を比較する上では不要と考えられノイズ的悪影響をも

たらず可能性があるため、多変量解析で差がみられない一因となります。Fisher Ratio 値によるピーク情報の順位付けは、検体間比較を行うために、より差を明確化するための重要なポイントとなります(図 3 参照)。

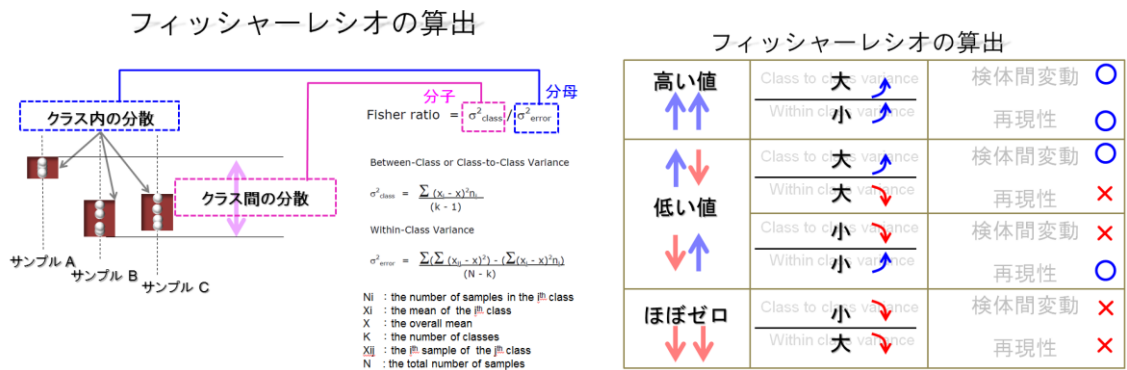


図 3 フィッシャーレシオの算出 と 数値の意味

フィッシャーレシオ値によりデータを精査する事で、網羅的に検出された際に数千あったピークは数百まで精査されます(図 4 参照)。多変量解析を用いてうまく解析できないケースではこの段階の処理を行っていない場合が多く、演者らは本ワークフローで最も重要なステップとして認識しています。

マーカ化合物の探索 第二段階

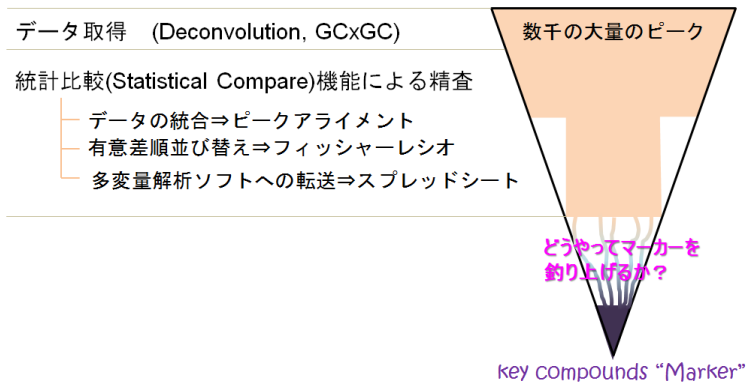


図 4 フィッシャーレシオによる精査後のピーク数と
マーカの絞り込みとの位置関係

3 多変量解析の利用

フィッシャーレシオ値により有意差の高い化合物を絞りこむことができれば、主成分分析やクラスター解析において明確な検体間の差を確認する事ができます。各解析がどのようにマーカ探索に役立つかについて述べます。

1 主成分分析(PCA) スコアプロット

最も基本的な検体間比較ツールであり、容易に検体間の差を確認する事ができます。図中の PCA-A は、フィッシャーレシオの低い値のものを含めた解析結果を PCA-B はフィッシャーレシオ値 10 以上について解析した結果をそれぞれ示します。図のように、フィッシャーレシオ値で

精査を行った方が各検体の繰り返し間の収束率が高く、検体間の差がより明確に表現される事が知られています。

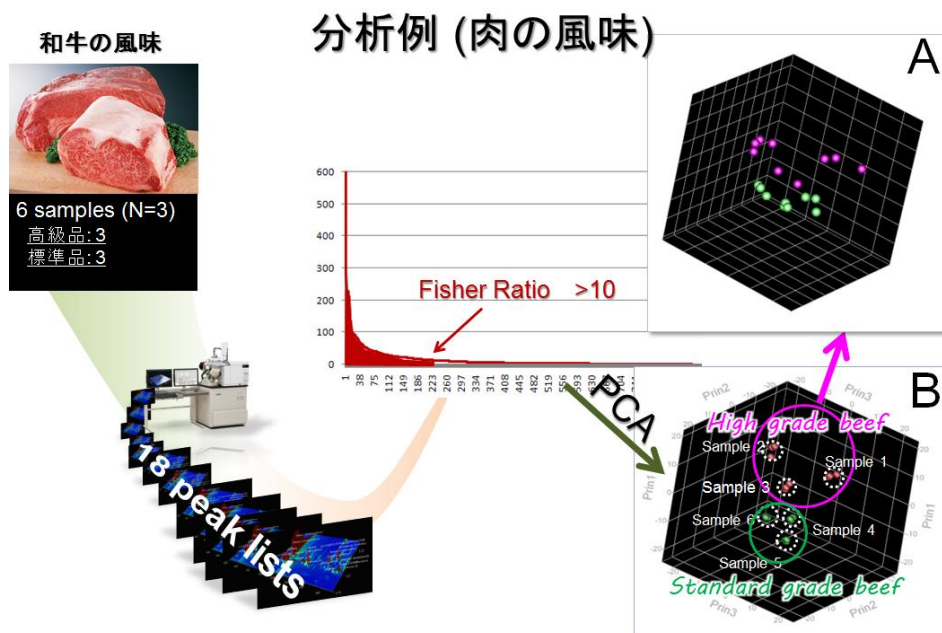


図5 肉の風味分析

2 主成分分析(PCA) ローディングプロット

スコアプロットと同様に検体間の差を表現するツールですが、化合物に対して位置関係を表現するため、官能試験のスコア値と化合物の面積値などを同一の主成分軸に展開する事が可能で、互いの距離情報から相関係数を導く事が可能です。探索をする上で、検体の特性(臭い、物性など)に関するスコア値を MS データ(面積値)と比較する事重要なステップです。その作業はローディングプロットにおける主成分軸上で視覚的に確認する事ができます(図-6 参照)。詳細は講演にて説明します。

ビール風味の特徴(GC-MS&官能試験)

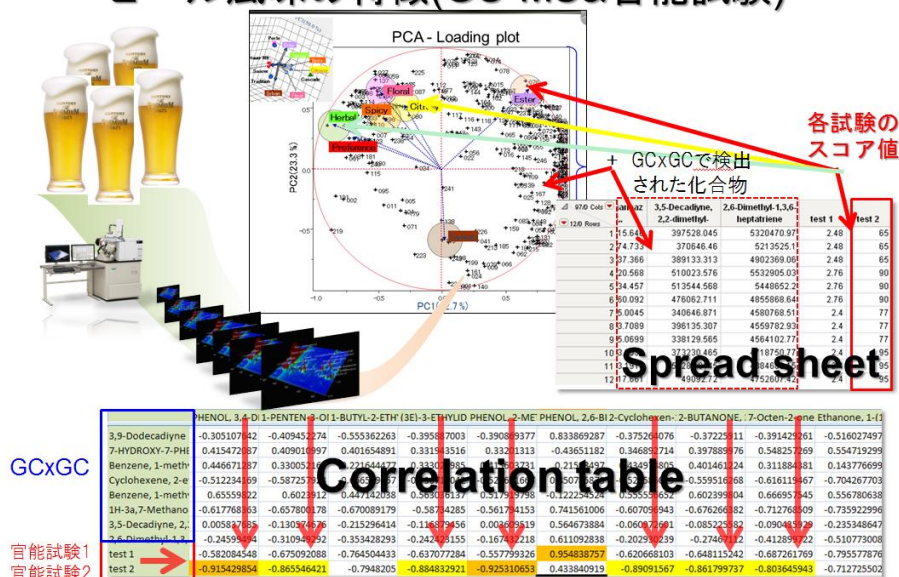


図6 ローディングプロットと、官能試験とのフィッティング

3 クラスター解析

クラスター解析は、系統樹を用いて検体の性質と化合物の共通性を分類し表現するツールです。サンプルの共通性は互いの変数間の相関係数により導かれたクラスターにより示されます(図 7 参照)。

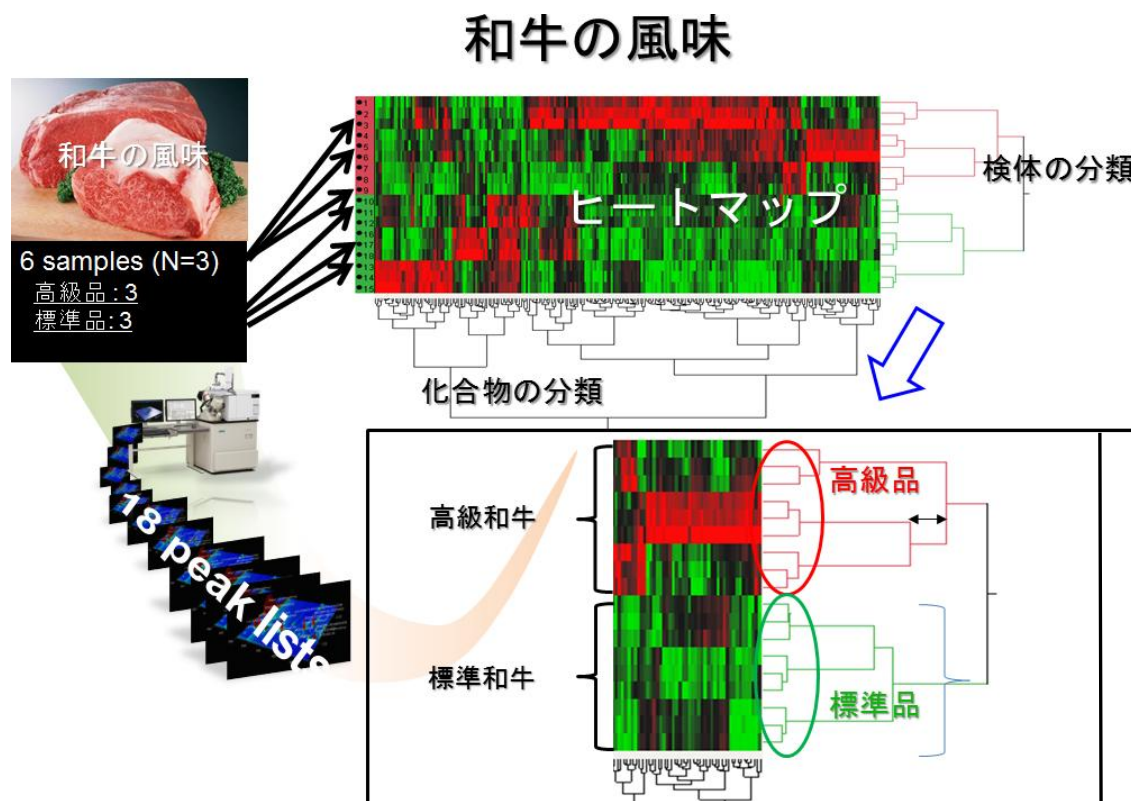


図 7 クラスター解析の有効性

多変量解析の結果は、挙動を共にする変数(面積値、官能試験スコア値) 同士の相関係数などの情報から関連性を知ることができます。これらの多変量解析ツールを用いる事で、全く異なる実験である機器分析結果と官能試験との結果の結び付けが行われ、寄与率の高い香気成分が導かれます。このステップを通じマーカー候補が絞り込まれます(図 8 および 9 参照)。

風味の特徴と関連の高い香気化合物

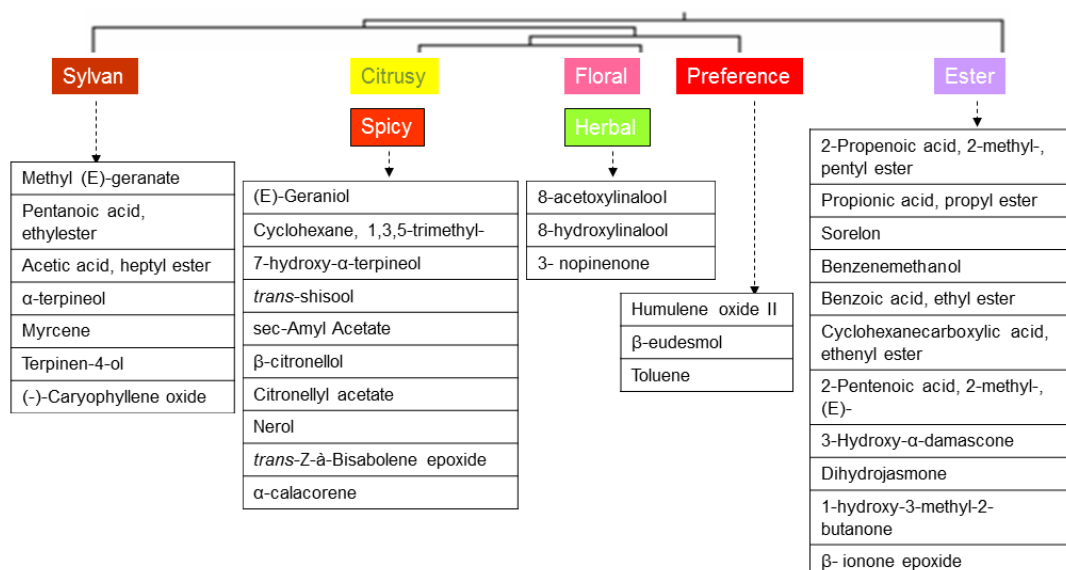


図 8 ビールの分析におけるサンプルの特徴に寄与率の高い化合物の例



マーカー化合物の探索

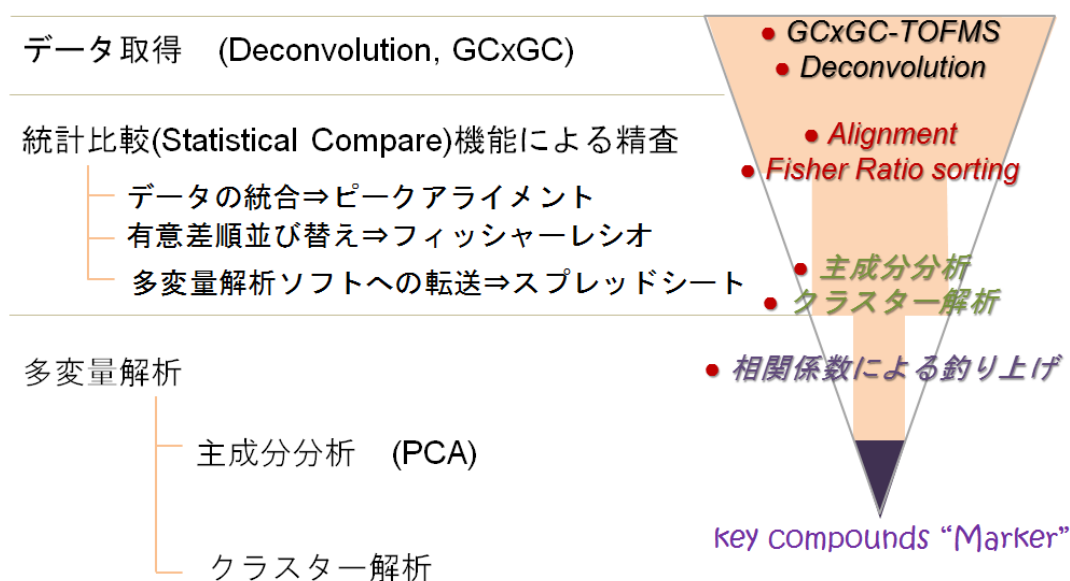


図 9 多変量解析とマーカー探索

3 おわりに

今回示したノンターゲット分析のデータ処理方法は、特に GC×GC のような膨大な情報を有するデータにおいて検体間の比較及び特徴に寄与する化合物の抽出を行う上で効果的であると考えられます。しかしながら、この結果で得られた化合物はノンターゲット分析である以上あくまで候補を探したに過ぎず、最終的な化合物の確認は標準品をもってして行う必要があります。

多くて数十本のピークを解析していた数十年前の装置とは異なり、最近の分析機器の性能の著しい進歩は非常に膨大な情報を我々にもたらす一方、しばしば人間の処理能力を超越した作業を要求する事は、一つ一つピークを確認するという分析化学の基本を困難なものにしています。そのため、確認できるレベルまで情報科学的な手法でピーク数を減らす作業は必須となってきます。複雑な香気性化合物の網羅的分析では、非常に多くのピークを対象としますが、ターゲット分析では得られない予期しない発見をノンターゲット分析では期待できるため、挑戦する価値は十分にあると考えられます。本分析法はデータの精査という点で大きな役割を担うと考えられ、読者の方々の日々の研究の手助けになれば幸いです。

4 参考文献

- 1 土屋文彦、矢島敏行:月刊フードケミカル 1月号,2009
- 2 土屋文彦 :月刊フードケミカル 6月号,2012
- 3 Pierce, K. et., al, *Anal. Chem.*, **78** (14), 5068 -5075 (2006)