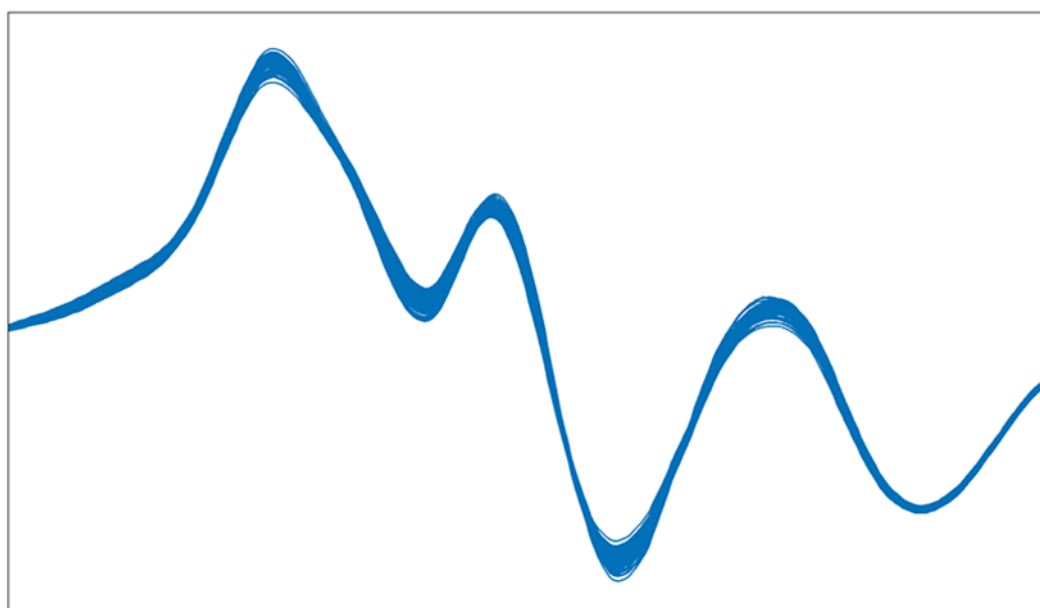


OMNIS NIR 理論



ハンドブック

8.0600.8101JA / v10 / 2026-04-08



Metrohm AG
Ionenstrasse
CH-9100 Herisau
Switzerland
+41 71 353 85 85
info@metrohm.com
www.metrohm.com

OMNIS NIR 理論

ハンドブック

8.0600.8101JA / v10 /
2026-04-08

本文書は、著作権法で保護されています。本文書の無断複写・転載を禁じます。

この文書はオリジナル文書です。

本文書は細心の注意を払い作成されていますが、それでも、誤りが含まれている場合があります。お気づきの点がございましたら、上記の宛先までご連絡ください。

免責条項

不適切な保管または使用などに起因する故障に対し、Metrohm は一切の保証の責任を負わないものとします。独断による製品の変更(改造または拡張など)の場合も、それに起因する損傷およびその結果に対し、メーカーはいかなる責任も負いません。Metrohm の製品文書の取扱説明書および注意には厳密に従ってください。そうでない場合、Metrohm はいかなる責任も負わないものとします。

目次

1	概要	1
1.1	導入	1
1.2	コンセプトの枠組み	1
1.3	本文書について	2
1.4	より詳しい情報	2
2	スペクトル	3
2.1	光と物質の相互作用	3
2.2	数学的基礎	6
2.2.1	Beer-Lambert の法則	6
2.2.2	線形回帰	6
2.3	光がスペクトルに変換される方法	8
3	装置のセットアップ	12
3.1	波長のキャリブレーション	12
3.2	リファレンスの標準化	13
3.2.1	OMNIS NIR Analyzer	14
3.2.2	2060 NIR	15
3.3	装置の性能テスト	24
3.3.1	外部 装置の性能テスト (OMNIS NIR Analyzer)	27
4	モデル開発	29
4.1	物理的サンプル	30
4.2	主成分分析(PCA)	33
4.3	データ前処理	37
4.3.1	データ前処理	37
4.3.2	波長範囲	45
4.3.3	スペクトルアウトライヤー	46
4.3.4	基準値アウトライヤー (定量化)	53
4.3.5	データセット分割	54
4.4	定量化	56
4.4.1	PLS 回帰	56
4.4.2	定量化モデルの検証	58
4.4.3	OMNIS Model Developer (OMD)	65
4.4.4	スロープ/y インターセプト補正	67
4.5	識別と検証	69
4.5.1	サポート Vector マシン(SVM)	69
4.5.2	サンプルの製品属性の予測	72
4.5.3	識別モデルの検証	74

4.6	適格性評価	75
4.6.1	適格性確認モデルの計算	75
4.6.2	適格性確認モデルの検証	76
5	予測	78
5.1	定量化	78
5.1.1	アウトライヤーと結果のモニタリング	78
5.2	識別と検証	80
5.3	適格性評価	81
6	補遺	82
6.1	線形回帰の例	82
6.2	PCA アルゴリズム	86
6.3	PLS アルゴリズム	88
6.4	ホテリングの T^2 と Q 残差	89
6.5	スペクトルアウトライヤー - アルゴリズム	90
6.6	基準値アウトライヤー - アルゴリズム	92

1 概要

1.1 導入

近赤外分光法（NIR 分光法）は、サンプルの幅広いスペクトルに適している、非破壊で、迅速で、試薬不要の分析方法です。複数のパラメータを同時に分析し、材料の物理的プロパティと化学的プロパティの両方を測定できます。これには特に、分析対象物の濃度、密度、粒径、固有粘度が含まれています。

加えて、NIR 分光法 を用いて不明なサンプルを識別 (OMNIS Software 4.0 以降) することも、サンプルを検証 (OMNIS Software 4.2 以上) することもできます。

サンプルを非破壊でリモート測定できる機能は、品質管理とプロセス監視において重要な意味を持ちます。

このハンドブックでは、OMNIS Software に実装されている近赤外スペクトルの記録、処理、分析の手法に関する技術とアルゴリズムについて説明しています。第 2 章では、測定信号の吸収スペクトルへの変換方法を簡単に説明します。第 3 章では、装置のキャリブレーション方法について説明します。第 4 章では、対象パラメータ（定量化）または生成物の関連性（識別）をよそ予測できるモデルの開発について説明します。第 5 章では、未知物の予測について説明します。第 6 章は、さまざまなアルゴリズムの説明を含む補遺です。

1.2 コンセプトの枠組み

提示された方法は、次の枠組みに適合します。

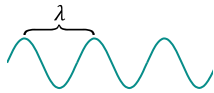
1. **キャリブレーション、標準化、性能テスト**
装置により記録された吸収スペクトルの伝達性と信頼性の確保。
2. **モデル開発**
モデルは、定量パラメータの予測またはサンプルの識別のために開発されます。
開発は、既知の対象パラメータまたは既知の製品。
3. **サンプル分析**
分析したサンプルからスペクトルが記録されます。スペクトルを基にして、定量化モデルが定量予測を行うか、または識別モデルがサンプルを識別もしくは検証します。
4. **モニタリング**
モデルおよび装置をモニタリングすることにより、システムがさらなる分析に適していることを確認します。

2 スペクトル

2.1 光と物質の相互作用

スペクトロメーターは、サンプルが光とどのように相互作用するかを測定します。光は、さまざまな程度で吸収または散乱することができます。相互作用は、光のプロパティ、特に波長、および材料のプロパティ、特に分子構造に依存します。

波長



光は電磁放射です。光は、振動する電場と磁場の波として空間を移動します。波は空間と時間の中で伝播します。したがって、波はその波長 λ （例えばナノメートル = 10^{-9} m）と周波数(Hz)によって特徴づけられています。

波長は、波の周波数と反比例しています。高い周波数（1秒あたりの振動が多い）の波は短い波長を持ち、あるいはその逆も同様です。この関係に従って、波は波長(nm)または周波数(Hz)で記述できます。

光は光子と呼ばれる離散量子単位でエネルギーを交換できます。各光子のエネルギー E は、その周波数 f または波長 λ に依存しています。

$$E = hf = h \frac{c}{\lambda}$$

ただし、 h はプランク係数、 c は光速です。

図1は、電磁放射のそれぞれの範囲を示します。

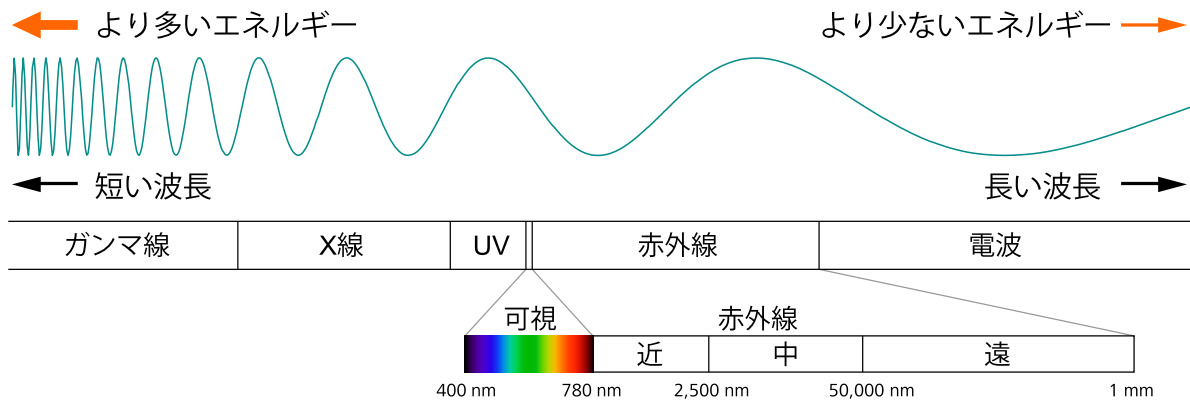


図 1 電磁放射の範囲。近赤外線範囲(NIR)は可視光の範囲の隣にあります。NIR の波長は 780 nm ~ 2,500 nm です。

放射源

電磁放射のそれぞれの範囲には、それぞれの放射源があります。NIR 範囲では、その源は**熱放射**です。赤外線カメラは、例えば人体は周囲より温度が高いので、人体を検出します。

強度

電磁波の振幅が光の強度を決定します。振幅が高いほど、強度は高くなります。可視光の場合、強度は輝度として認識されます。

光と物質の相互作用

以下に、分子による光の吸収プロセスを説明します。

光と物質が相互作用する方法は、電磁放射の範囲によって異なります。例えば、可視光のエネルギーが分子に伝達されると、分子内の電子が低いエネルギー準位から高いエネルギー準位に変化します（電子遷移）。赤外範囲では**振動遷移**が発生します。化学結合、官能基、分子は伸縮振動、変形振動、回転振動など、さまざまな方法で振動します。

分子は離散的な振動状態しかとることができません。環境温度ではほとんどの分子は基本振動状態（レベル 0）にあります。基本の振動状態から励起状態への線維は、次のスキームに従って名付けられています。

振動遷移	名前
$i \rightarrow j$	
$0 \rightarrow 1$	基本的な遷移
$0 \rightarrow 2$	第 1 倍音遷移
$0 \rightarrow 3$	第 2 倍音遷移

高い振動状態は、高いエネルギー準位に対応します。状態 i から励起状態 j に遷移する場合、分子は特定の遷移エネルギー ΔE_{ij} を吸収する必要があります。

光は光は、 $E = hf$ の部分でエネルギー変換できます。ただし、 f は光の周波数です。光子エネルギー hf が遷移エネルギー ΔE_{ij} と同じである場合、光の吸収が行われます。

許容される振動状態は、特に結合強度と関係する原子の質量に依存します。そのため、特定の種類の結合は、特徴的な遷移エネルギーまたは吸収される波長に関係づけることができます。

その場合、光が吸収されるためには、さらなる条件が満たされる必要があります。振動遷移は、分子の電気双極子モーメントが変化するように、電荷分布をシフトさせる必要があります。エネルギー吸収の確率は、該当する化学結合に沿った双極子モーメントの変化の大きさに依存しています。

振動遷移は、極性および無極性の両方の分子および官能基の双極子モーメントの変化を引き起こすことができます。N₂ のような同核分子は赤外を吸収しません。

励起振動状態の持続時間は限られています。分子が低い振動状態に戻ると、エネルギーは熱に変換されます。

NIR のスペクトル範囲

基本遷移に対応する波長は中赤外線範囲にあります。近赤外範囲には、倍音遷移と結合帯域が含まれています。図2は、さまざまな分子と官能基により吸収される波長帯域を示します。

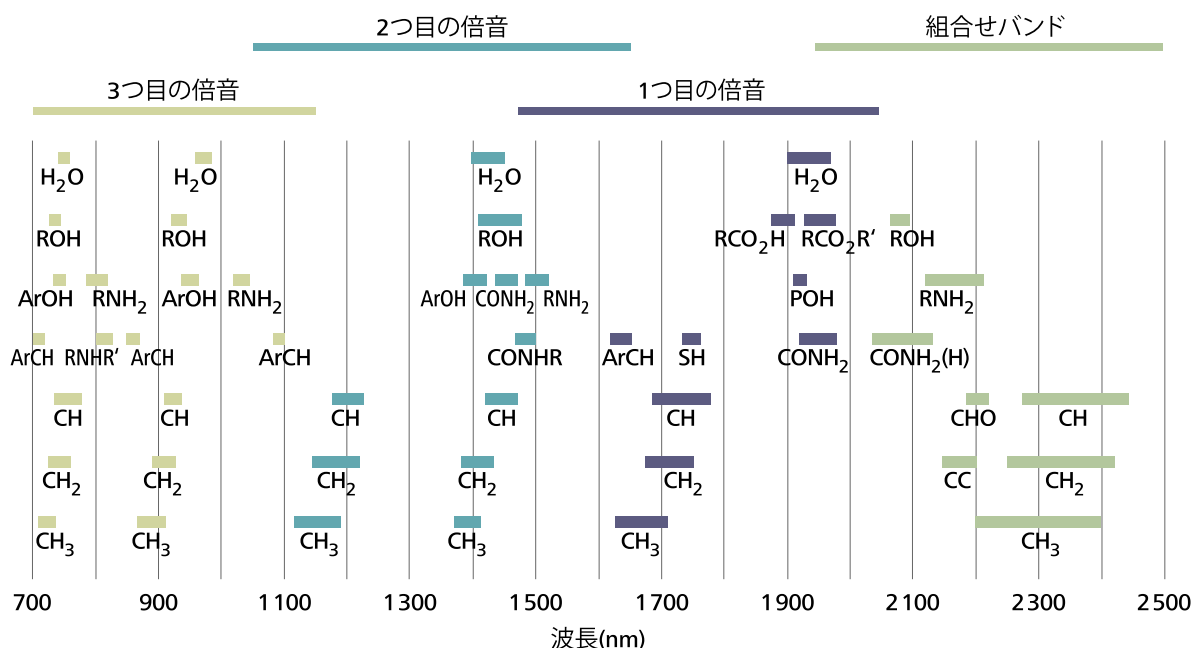


図 2 NIR 吸収帯域

基本遷移は最も確率の高い遷移であり、最も頻繁に発生します。倍音遷移の確率は低いです。そのため、基本遷移は倍音遷移よりも多くの光を吸収します。一般的に、吸収は倍音ごとに減少します。そのため、倍音は吸収が高い分子に対応します。

2つ以上の基本振動は、基本振動の結合周波数に対応する単一の周波数の光で同時に励起できます。対応する吸収帯域は、**結合バンド**と呼ばれます。結合バンドは NIR 範囲、つまり 1,900~2,500 nm です。

2.2 数学的基礎

2.2.1 Beer-Lambert の法則

Beer-Lambert の法則は、均質なサンプルによる光の吸収が、サンプル内の吸収物質のプロパティにどのように依存しているかを説明します。

$$A = \varepsilon \cdot c \cdot l$$

ただし、 A は吸光度、 ε は吸収体のモル吸光係数 (L/mol/cm)、 c は吸収体の濃度 (mol/L)、 l はサンプルの層厚 (cm)に対応します。

モル吸光係数 ε は、物質がどれだけ吸収するかを示す定数です。モル吸光係数は、波長 i と物質 j に固有の値です。混合物の総吸光度は、混合物に含まれているすべての物質の吸光度の合計です。

$$A_i = \sum_{j=1}^N \varepsilon_{ij} c_j l$$

ただし、 A_i は波長 i の場合の吸光度、 N は混合物内の物質数、 ε_{ij} は波長 i および物質 j の場合のモル吸光係数、 c_j は物質 j の濃度です。

Beer-Lambert の法則は、吸光度と濃度との間の線形関係、および吸光度とモル吸光係数との間の線形関係を前提としています。この線形関係は多くの状況に当てはまります。

Beer-Lambert の法則に基づき、分光吸収測定では以下を検出できます。

- 吸収体の濃度変動。
これは頻繁に利用されます。
- モル吸光係数に影響を及ぼす因子の変動。
溶媒の温度、粘度、pH 値、誘電率は、モル吸光係数に影響を与えます。場合によっては、これを分光測定に利用できます。

散乱効果は、Beer-Lambert の法則とは関係ありません。散乱効果は、粒径変動を認識する場合などに使用できる場合があります。

2.2.2 線形回帰

1つの波長

最もシンプルな場合では、混合物には1つの吸収体のみが含まれています。Beer-Lambert の法則に従うと、特定波長において吸光度は吸収体の濃度に比例します。

図1では、各点は吸収体の既知の濃度 (x軸) と、測定された吸光度 (y軸) のサンプルが示されています。線形回帰の結果は、回帰曲線 **A** となります。

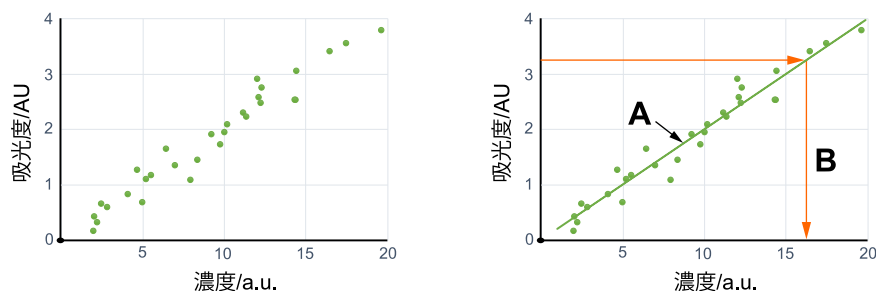


図 3 吸光度と濃度との関係

吸収体の濃度が未知であるサンプルに対して、その濃度は次のようにして測定できます。

1. 指定された波長における吸光度の測定。
2. 濃度(B)を決定するために回帰直線を使用。
回帰曲線は、吸収体の濃度などの対象パラメータを予測するための簡単な定量化モデルです。

この手順は、混合物が異なる濃度の吸収体を含んでいる場合には失敗します。

複数の波長

単一の波長で吸収を測定する代わりに、複数の波長で吸収測定を行い、スペクトルを生成することができます。上記同様に、線形回帰を用いてスペクトルと対象パラメータとの間の関係を決定することができます。複数の波長の場合、重線形回帰(MLR)が必要になります。

混合物に濃度が異なる複数の吸収体が含まれている場合であっても、重線形回帰によって対象パラメータを予測できることが証明されています (82 ページ, 6.1 章を参照)。

ただし、重線形回帰の場合には波長の数よりも多くのサンプル数が必要になります。PCA や PLS などの他のメソッドも分光予測に使用できます。

2.3 光がスペクトルに変換される方法

スペクトロメーター（または分光器）は光源と検出器ユニットで構成されています。光源は広範囲のスペクトルの光、つまり多色光を放射します。光はサンプルと相互作用します。次に、スペクトロメーターは残りの光を波長の関数として記録します。

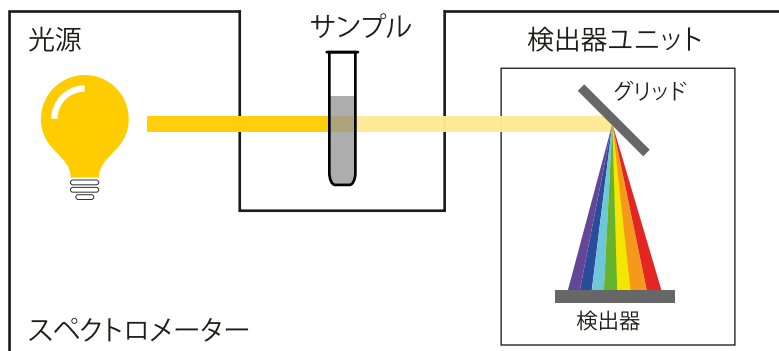


図 4 光源と検出器ユニットを備えるスペクトロメーター。

スペクトロメーターでは、光は格子を使って個々の波長を分割します。検出器は、各波長がラインセンサ内にある異なる要素、あるいは画素に当たる光を測定します。

スキャンはすべての画素にわたる測定です。各画素は、光の強度に比例する光電信号を生成します。信号は画素に対してプロットすることができます。

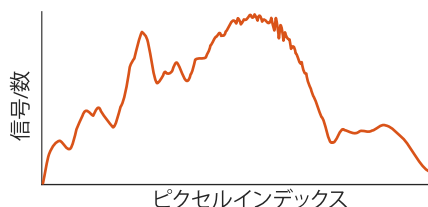


図 5 画素の関数としての検出器信号のスペクトル。

積分時間

積分時間は、検出器が光を集める時間です。積分時間が長くなると信号は増えます。

積分時間が長すぎると、検出器の飽和の原因となり、情報を失う場合があります。積分時間が短すぎると信号が少なくなり、それによって信号ノイズ比が低下します。

自動積分時間により最適な露光時間を提供します。つまり、飽和することなく最適な信号ノイズ比が得られます。各サンプルスキャンや基準スキャンの前に、複数の測定を行います。積分時間は、最大強度の信号が約 90%の既知の範囲に達するように調整します。

積分時間による違いは、さらなる計算において考慮されます。

- **OMNIS NIR Analyzer**

積分時間は常に自動的に調整されます。

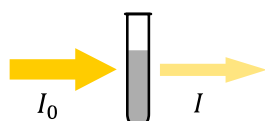
- **2060 NIR**

積分時間を手動または自動で調整できます。

手動での積分時間により、繰り返しの測定や類似の測定の場合に、測定時間を短縮できます。積分時間が長すぎることにより、検出器が飽和することを避けるため、積分時間は十分に余裕をもって調整してください (45 ページ、「飽和」を参照)。

吸光度

分光測定では、どれだけの光をサンプルが吸収または散乱するかを測定します。サンプルは光線にさらされます。検出器は、光源から放射された光と、サンプルとの交換作用後に残った光を測定します。



吸光度 A は、サンプルと光線との交換作用前の光強度 (I_0) と、サンプルと光線との交換作用後の光強度 (I) との間の比率の常用対数として定義されます。

$$A = \log_{10} \frac{I_0}{I}$$

吸光度 1 とは 10% の光がサンプルを通過することを意味し、吸光度 2 とは 1% の光が光がサンプルを通過することを意味します。

吸光度に単位はありません。

基準スキャンとサンプルスキャン

上の計算式により、吸収スペクトルを計算するには 2 つのスキャンが必要です。スキャンは、各画素に対して光電信号 S を測定します：

- **基準スキャン** は、サンプルと光線との交換作用前の信号 S_0 を測定します。
- **サンプルスキャン** は、サンプルと光線との交換作用後の信号 S を測定します。

測定された画素の光電信号は、画素面にわたる光強度の平均値に比例します。したがって、 $S_0/S = I_0/I$ となります。これにより、光電信号を吸光度の計算に用いることができます。

$$A = \log_{10} \frac{I_0}{I} = \log_{10} \frac{S_0}{S}$$

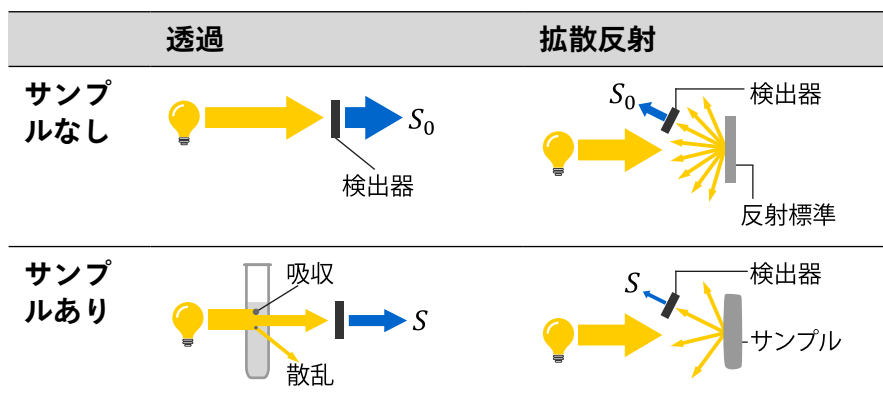
透過と反射

透過モード では、サンプルを通過する光が測定されます。 S_0 はサンプルなしで測定されます。 S は、サンプルを通過した光から測定されます。

反射モード では、サンプルから反射した光が測定されます。基準として、サンプルの代わりに反射標準が使用されます。反射標準は理想的に 100% の光を反射します。反射した光の一部は検出器に送ら



れ、信号 S_0 が得られます。信号 S は同じ方法で測定されますが、光を反射したサンプルで得られます。



計算された吸光度 A は、検出器に達していないすべての光を示します。そのため、 A はサンプルにより吸収された光だけでなく、以下も含んでいます：

- 検出器から散乱して検出器に到達しなかった光。
- 検出器に誤って散乱した光。

吸収スペクトル

吸収スペクトルは、基準スキャン（信号 S_0 ）およびサンプルスキャン（信号 S に基づいて、上の計算式から計算されます。

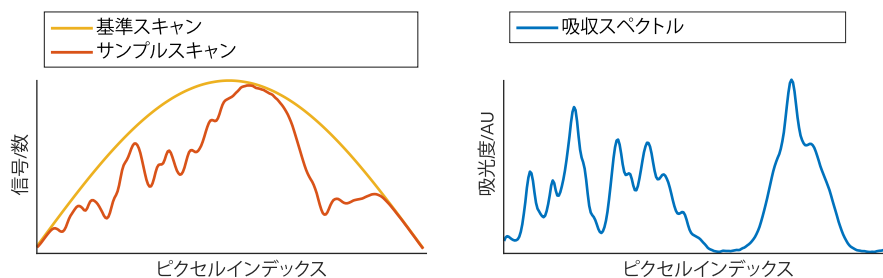


図 6 画素インデックスの関数としての基準スキャンおよびサンプルスキャン（左）と、計算された吸収スペクトル（右）。

上記の計算は、基準スキャンとサンプルスキャンが同じ光路、または類似した光学的特性を有する光路を使用していることを前提としています。プロセス環境では、多段基準アプローチを使用します（15 ページ、3.2.2 章を参照）。

画素から波長

画素スケールは波長スケールに変換されます。装置は各画素に正確な波長を割り当てます。例：

画素 6 → 波長 1,009.4 nm

各画素に対する正しい波長は、波長のキャリブレーションにより測定されます（12 ページ、3.1 章を参照）。

波長スケールの変換

スペクトルは標準波長スケールの補間により変換されます：

1,000.0 nm、1,000.5 nm、1,001.0 nm、...

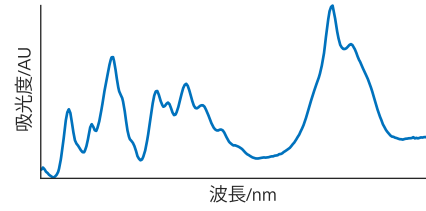


図 7 波長スケールの吸収スペクトル。

3 装置のセットアップ

次のステップでは、異なる時間、製品ファミリー **OMNIS NIR Analyzer** の同じ装置または別の装置、あるいはタイプ **2060 NIR** の別の装置で記録されたかに関係なく、特定のサンプル測定セットアップにより特定のサンプルに対して同一のスペクトルが得られるようにします。

スペクトルの x 軸と y 軸の両方を両方を考慮する必要があります。

- **x 軸**：波長のキャリブレーション (12 ページ, 3.1 章を参照)
- **y 軸**：リファレンスの標準化 (13 ページ, 3.2 章を参照)

さらに、装置の性能が要件を満たしていることを確認する必要があります。つまり：

- 装置でスペクトルを記録する前に、**装置の性能テスト**が成功している必要があります。(24 ページ, 3.3 章を参照)。

標準

波長のキャリブレーションと装置の性能テストに対して、装置は内部の計量学的に追跡可能な**波長基準**を使用します。

反射モードを使用する場合、それについて、装置タイプにより異なりますが、リファレンスの標準化と装置の性能テストに対して**反射標準**が必要です。

- **OMNIS NIR Analyzer**：内部反射標準
- **2060 NIR**：外部反射標準

3.1 波長のキャリブレーション

波長のキャリブレーションは波長値、つまりスペクトルの x 軸を正規化します。光センサアレイの各画素を波長に割り当てます。

波長のキャリブレーションは、内部の計量学的に追跡可能な波長基準を使用します。波長基準は、定義されたピークと既知のピーク位置を有する吸収スペクトルを備えています。

コマンド **CAL WL** は次のステップを実行します。

1. 波長基準の吸収スペクトルは、内部基準光路を用いて画素スケール上に記録されます。
2. 記録されたスペクトルでは、ピーク位置がサブ画素精度で識別されます。
3. 画素スケール上の測定されたピーク位置と、波長基準の正規化されたピーク位置を用いて、多項式回帰が行われます。
4. 回帰多項式は各画素にそれぞれの波長を割り当てます。

回帰係数は装置上に保存されます。

- **OMNIS NIR Analyzer** : 各サンプル表示に対して回帰係数のセットが装置に保存されます。つまりつまり、OMNIS NIR Analyzer Liquid/Solid は波長のキャリブレーションと検証を両方のファンクションユニットに対して実行する必要があることを意味しています。
- **2060 NIR** : 回帰係数は装置固有です。すべてのセットに対して同じセットを使用します。

波長のキャリブレーションの検証

波長のキャリブレーションを実行した後に、これを検証する必要があります。コマンド **VAL WL** は次のステップを実行します。

1. 波長基準の吸収スペクトルは記録されます。
2. 波長の検証 :
 - a. 記録されたスペクトルでは、波長スケールのピーク位置が識別されます。
 - b. 測定されたピーク位置と既知のピーク位置との間の波長残差が計算されます。
 - c. テストに合格するには、各ピークに対して波長残差は許容誤差内である必要があります。
3. 帯域の検証 :
 - a. 記録されたスペクトルではピークワイズが測定されます。
 - b. 測定されたピークワイズと既知のピークワイズとの間の帯域残差が計算されます。
 - c. テストに合格するには、各ピークに対して帯域残差は許容誤差内である必要があります。
4. 上記すべての残差が許容誤差内である場合、検証の総ステータスは完了します。

装置でスペクトルを記録する前に、検証が成功している必要があります。

3.2 リファレンスの標準化

リファレンスの標準化は吸光度の値、つまりスペクトルの y 軸を正規化します。

吸光度の測定

サンプルの吸光度 A を計算するには、信号 S_0 (基準スキャン) および S (サンプルスキャン) が必要です (8 ページ, 2.3 章を参照) :

$$A = \log_{10} \frac{S_0}{S}$$

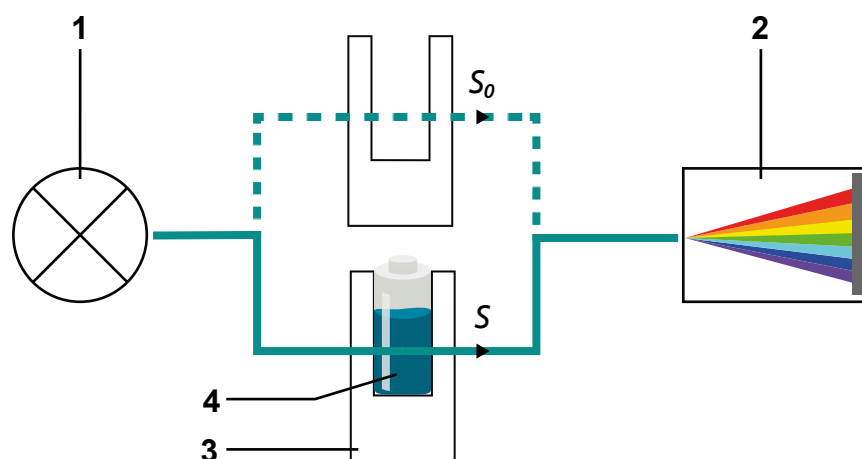


図 8 透過モードにおける光路 (液体のサンプル提示の例として)。

図 8 では、光源の光(1)はサンプルホルダー(3)を通じて検出器(2)に達しています。

基準信号 S_0 はサンプルなしで測定され、信号 S はサンプル(4)ありで測定されます。そうでない場合、両方の光路に対する光学的特性は同一であり、両方の信号に対して同じ割合分、蒸着を生じさせます。これによって上記の計算式の結果が変わることはありません。

方程式は、 S を基準である S_0 に関連付けます。両方の信号は同様に重要です。2つの信号の一つの偏差があると、別の別の吸光度の値が発生し、最終的には別のスペクトルが発生します。

S と S_0 の両方は、装置と環境条件の変動により影響を受けます。この影響が互いに打ち消すようにするため、両方の信号を短時間で測定する必要があります。

この原理の実装は装置タイプにより異なります：

- **OMNIS NIR Analyzer**
サンプルの吸収スペクトルは、信号 S と S_0 から計算されます (14 ページ, 3.2.1 章を参照)。
- **2060 NIR**
プロセス環境では、 S と S_0 の測定に対して同じ光路を使用することは現実的ではありません。そのため、さらなる対策が必要です (15 ページ, 3.2.2 章を参照)。

3.2.1 OMNIS NIR Analyzer

リファレンスの標準化は、信号 S_0 および S の測定と、吸光度 A の計算によって行われます。

サンプルのスペクトル記録

i ファンクションユニットでスペクトルを記録する前に、装置の性能テストがファンクションユニットに対して成功していなければなりません (24 ページ, 3.3 章を参照)。

1. サンプルは、サンプル提示で準備されていなければなりません。
2. 吸収スペクトルは、それぞれ最後に記録された標準スペクトル S_0 で計算されます。 S_0 の値を含めるために、コマンド **MEAS REF SPEC** を実行できます。
固形物のサンプル提示を使用するたするために、装置は自動的に反射標準を光路に追加します。この反射標準では、信号 S_0 の補正は必要ありません。
3. コマンド **MEAS SPEC** によりサンプルは測定されます。これにより信号 S が生成されます。
4. ソフトウェアは A 、つまりサンプルの吸光度を計算します:

$$A = \log_{10} \frac{S_0}{S}$$

その際、 S_0 は基準光路で測定された信号と、 S はサンプルで測定された信号と一致します。

3.2.2 2060 NIR

タイプ **2060 NIR** の装置は外部リファレンスの標準化を必要とします。

外部リファレンスの標準化

信号 S (サンプルあり) および信号 S_0 (サンプルなし) を同じ光学プロパティの光路で繰り返し測定することは、時間がかかり、エラーが発生します。

そのため、2つのさらなる光路が導入されます (15 ページ、[図9](#) を参照)。

- 装置内の**内部基準**。内部基準光路は、簡単な方法で測定できる信号 S_{ref} を提供します。
- 光ファイバが**キャリブレーション装置**と接続されている別の外部光路。この光路は信号 S_{fiber} を提供します。

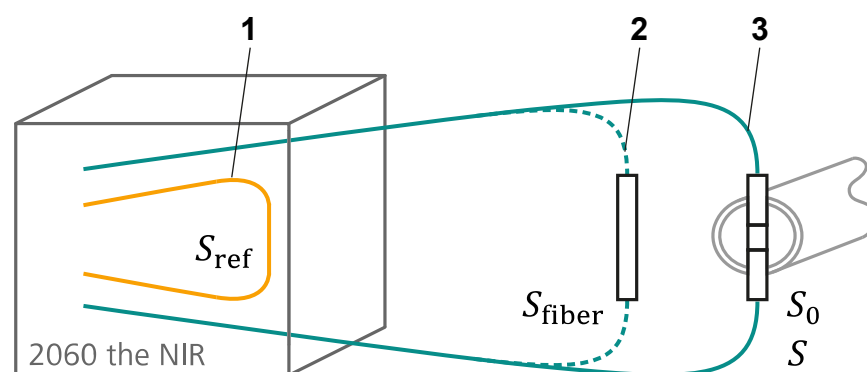


図 9 透過モードでの光路例：内部基準光路(1)、キャリブレーション装置と接続された外部光ファイバ(2)、サンプルあり/なしでプローブと接続された外部光ファイバ。(3) 光路2 および3 は、単に別に接続されているだけで、同じ光ファイバを示しています。

キャリブレーション装置は光ファイバーを固定し、それによって基準光路(2)を形成します。透過モードでは空気は基準として機能し、100%の光を伝送します。反射モードでは、キャリブレーション装置は反射標準も記録します。まず、100%の光を反射する理想的な反射標準から始めます。

サンプルの吸光度Aは、信号S₀ およびSから計算されます。2つの追加信号S_{ref} およびS_{fiber} が分母と分子に追加されますが、結果は変わりません。

$$A = \log_{10} \frac{S_0}{S} = \log_{10} \left(\frac{S_{ref}}{S} \cdot \frac{S_{fiber}}{S_{ref}} \cdot \frac{S_0}{S_{fiber}} \right)$$

この方程式は次のように変換できます。

$$A = \log_{10} \left(\frac{S_{ref}}{S} \right) - \log_{10} \left(\frac{S_{ref}}{S_{fiber}} \right) - \log_{10} \left(\frac{S_{fiber}}{S_0} \right)$$

3つの項は吸光度の値を示し、次のように示すことができます。

$$A = A_{total} - A_{fiber} - A_{window}$$

図10は、信号S_{ref}、S_{fiber}、S₀ およびSの測定方法を示します。

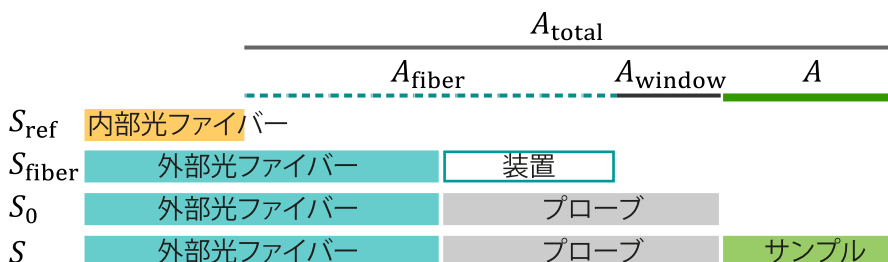


図10 外部リファレンスの標準化

A_{total} は、外部光ファイバー、プローブおよび内部光ファイバーに関するサンプルの吸光度です。

A_{fiber} は、外部光ファイバーと、内部光ファイバーに関するキャリブレーション装置の吸光度です。

A_{window} は、キャリブレーション装置を差し引いた差し引いたプローブの吸光度です。

環境変動の除去

サンプルの吸光度Aを測定するために、3つの吸光度の値を測定します。3つの値から、Aは上記方程式により計算されます。

$$A = A_{total} - A_{fiber} - A_{window}$$

これにより、それぞれの時点での3つの吸光度の値を測定できます。この方法により、3つの各測定に対して装置の変動や環境条件の変動を簡単に排除できます。

- A_{total} は各サンプル測定で測定されます。その際、変動を除去するために、 S_{ref} および S を短時間のうちに測定する必要があります。
- グラスファイバーの補正スペクトル A_{fiber} は、算出されることは比較的まれです。その際、変動を除去するために、 S_{ref} および S_{fiber} を短時間のうちに測定する必要があります。
- ウィンドウ補正スペクトル A_{window} は、算出されることは比較的まれです。一般的には取り付け後1回のみです。その際、変動を除去するために、 S_{fiber} および S_0 を短時間のうちに測定する必要があります。

ウィンドウ補正はいつ必要ですか？

キャリブレーション装置がプローブの光学プロパティを完全に再現する場合、 A_{window} は0になります。この場合、ウィンドウ補正をスキップできます。

透過モードでは一般的にウィンドウ補正が必要です。反射モードではウィンドウ補正は不要です。ただし、次の表に示されている例外があります。

測定モード	プローブ	リファレンスの標準化
透過	透過ペア	ガラスファイバー + ウィンドウ
	透過プローブ	
	単繊維の半透過プローブ	
反射	反射プローブ	ガラスファイバー
	MicroBundle の半透過プローブ	ガラスファイバー + ウィンドウ

ウィンドウ補正が必要であるかを決定するには、キャリブレーション装置とプローブとの組み合わせを個々に確認する必要があります。キャリブレーション装置がプローブの光学プロパティを完全に再現しない場合、ウィンドウ補正が必要です。

チャンネル

タイプ **2060 NIR** の装置は複数のチャンネルを提供します。各チャンネルは、別の光ファイバー構成とプローブ構成と接続できます。そのため、リファレンスの標準化は各チャンネルに対して別々に実施する必要があります。

すべてのチャンネルは、同じ内部基準光路、つまり同じ信号 S_{ref} を使用します。マルチプレクサは、内部基準とそれぞれの測定チャンネルを切り替えます。

グラスファイバー補正の実行

初回セットアップ後や、チャンネルの光ファイバー設定を変更する場合、グラスファイバー補正を実行してください。ランプの交換や環境条件の極端な変化によっても、標準化が再度推奨される場合があります。

この方法では基準材料を使用します。

- 反射モードでは基準材料は反射標準です。理想的でない反射標準 (99 %など) から始めます。反射標準は既知の正規化された吸収スペクトル A_{nominal} を備えています。
- 透過モードでは空気は基準として機能します。正規化された吸収スペクトルは、空気が光を吸収しないと想定されるので、ゼロライン($A_{\text{nominal}} = 0$)です。

図11 は次のアプローチを示しています。

1. 外部光ファイバーは、キャリブレーション装置に接続されている必要があります。
反射モードでは、キャリブレーション装置は反射標準と組み合わせられます。
2. インターフェース**グラスファイバー**のコマンド **REF STD** は次のスキャンを実行します。
 - a. 内部基準スキャンは S_{ref} の値を提供します。
 - b. 外部スキャンは、外部光ファイバー、キャリブレーション装置および基準材料を測定します。これにより信号 S_{raw} が生成されます。
3. ソフトウェアは A_{raw} (**測定された生スペクトル**)を計算します。

$$A_{\text{raw}} = \log_{10} \frac{S_{\text{ref}}}{S_{\text{raw}}}$$

ただし、 A_{raw} は、外部光ファイバーの吸光度、キャリブレーション装置および内部光路に関する基準材料に対応します。

4. 正規化された基準材料の吸収スペクトル A_{nominal} は、ソフトウェアでは**標準スペクトル**として示されます。標準スペクトルを A_{raw} から差し引くと、 A_{fiber} を得られます。

$$A_{\text{fiber}} = A_{\text{raw}} - A_{\text{nominal}}$$

ただし、 A_{fiber} は、光ファイバーの吸光度および内部光路に関するキャリブレーション装置に対応します。

注記：透過モードでは $A_{\text{nominal}} = 0$ 、 $A_{\text{fiber}} = A_{\text{raw}}$ です。

A_{fiber} はグラスファイバー**補正スペクトル**を示しています。

5. A_{fiber} は、各チャンネルに対して再度グラスファイバー補正が実行されるまで変化しません。

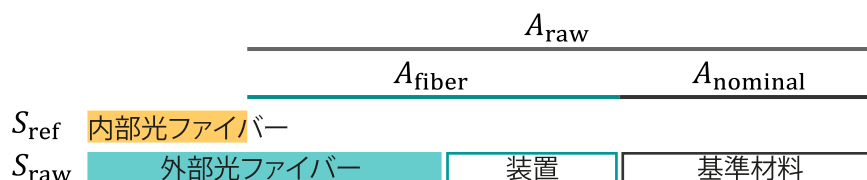


図 11 グラスファイバー補正

グラスファイバー補正の検証

グラスファイバー補正は、同じ測定パラメータと同じキャリブレーション装置で検証してください。

この方法では基準材料を使用します。

- 反射モードでは基準材料は反射標準です。理想的でない反射標準 (99 %など) から始めます。反射標準は既知の正規化された吸収スペクトル $A_{nominal}$ を備えています。
- 透過モードでは空気は基準として機能します。正規化された吸収スペクトルは、空気が光を吸収しないと想定されるので、ゼロライン ($A_{nominal} = 0$) です。

図 12 は、検証の残差の測定方法を示します。

1. 外部光ファイバーは、キャリブレーション装置に接続されている必要があります。
反射モードでは、キャリブレーション装置は反射標準と組み合わせられます。
2. インターフェース **グラスファイバー** のコマンド **VAL REF STD** は次のスキャンを実行します。
 - a. 内部基準スキャンは S_{ref} の値を提供します。
 - b. 各チャンネルの外部スキャンは、外部光ファイバー、キャリブレーション装置および基準材料を測定します。これにより信号 S_{raw} が生成されます。
3. ソフトウェアは A_{raw} (**測定された生スペクトル**) を計算します。

$$A_{raw} = \log_{10} \frac{S_{ref}}{S_{raw}}$$

4. A_{raw} はグラスファイバーの補正スペクトルにより補正され、光ファイバーおよびキャリブレーション装置の吸光度を削除します。

$$A_{corrected} = A_{raw} - A_{fiber}$$

$A_{corrected}$ は、ソフトウェアで**測定され、補正されたスペクトル**として表示されます。

5. 理想的には $A_{corrected}$ が**標準スペクトル** $A_{nominal}$ と一致しなければなりません。両者の差は**検証の残差**として計算されます。

$$A_{residual} = A_{corrected} - A_{nominal}$$

注記：透過モードでは $A_{nominal} = 0$ 、 $A_{residual} = A_{corrected}$ です。

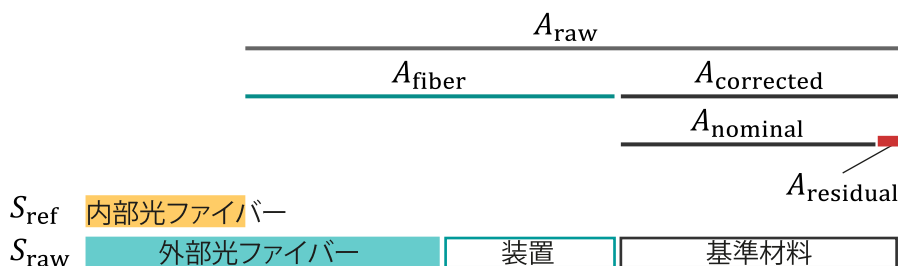


図 12 グラスファイバー補正の検証のための残差

検証の残差確認するために、波長範囲を複数のセグメントに分割します。各セグメントに対して、波長にわたる残差の二乗平均値により **RMS 残差**（単位：mAU）が求められます。

$$A_{RMS} = \sqrt{\frac{\sum_{i=1}^f (A_{residual_i})^2}{f}}$$

ただし、 f はセグメントの波長の数であり、 $A_{residual_i}$ は波長 i の残差です。

各セグメントは、 A_{RMS} に対する事前に定義された許容誤差を満たしていなければなりません。すべてのセグメントが許容誤差を満たしている場合、全体の検証は成功です。

装置で各チャンネルのスペクトルを記録する前に、検証が成功している必要があります。

ウィンドウ補正の実行

ウィンドウ補正が必要である場合、初回セットアップ後や、プローブ設定またはチャンネルの光ファイバー設定を変更する場合に必ず、グラスファイバー補正を実行してください。汚れなどによりプローブを変更する場合も、新たに標準化することが推奨される場合があります。

この方法では基準材料を使用します。

- 反射モードでは基準材料は反射標準です。理想的でない反射標準（99%など）から始めます。反射標準は既知の正規化された吸収スペクトル $A_{nominal}$ を備えています。
- 透過モードでは空気は基準として機能します。正規化された吸収スペクトルは、空気が光を吸収しないと想定されるので、ゼロライン ($A_{nominal} = 0$) です。

図 13 は次のアプローチを示しています。

1. A_{fiber} の現在の値を得るため、上記に示されたようにグラスファイバー補正を実行してください。
重要： グラスファイバー補正はウィンドウ補正前に行ってください。

2. 外部光ファイバーは、サンプルなしでプローブに接続してください。必要に応じてサンプルの代わりに反射標準を適用します。
3. インターフェース **ウィンドウ** のコマンド **REF STD** は次のスキャンを実行します。
 - a. 内部基準スキャンは S_{ref} の値を提供します。
 - b. 各チャンネルの外部スキャンは、外部光ファイバー、プローブおよび基準材料を測定します。これにより信号 S_{probe} が生成されます。
4. 内部光路に関するきゅ吸光度 A_{probe} は次のとおりです。

$$A_{probe} = \log_{10} \frac{S_{ref}}{S_{probe}}$$

5. ソフトウェアは A_{raw} (**測定された生スペクトル**) を計算します。
 $A_{raw} = A_{probe} - A_{fiber}$
 ただし、 A_{raw} は、プローブの吸光度、キャリブレーション装置に関する基準材料に対応します。
6. 正規化された基準材料の吸収スペクトル $A_{nominal}$ は、ソフトウェアでは **標準スペクトル** として示されます。標準スペクトルを A_{raw} から差し引くと、 A_{window} を得られます。

$$A_{window} = A_{raw} - A_{nominal}$$

ただし、 A_{window} は、キャリブレーション装置に関するプローブの吸光度です。

注記：透過モードでは $A_{nominal} = 0$ 、 $A_{window} = A_{raw}$ です。

A_{window} はウィンドウ **補正スペクトル** を示します。

7. A_{window} は、各チャンネルに対して再度ウィンドウ補正が実行されるまで変化しません。

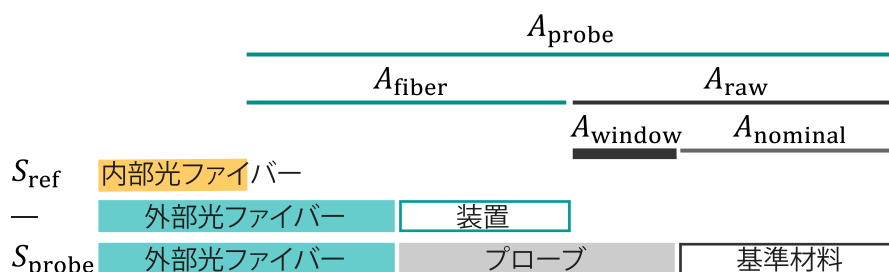


図 13 ウィンドウ補正

ウィンドウ補正の検証

ウィンドウ補正は、同じ測定パラメータと同じキャリブレーション装置で検証してください。

この方法では基準材料を使用します。

- 反射モードでは基準材料は反射標準です。理想的でない反射標準 (99 % など) から始めます。反射標準は既知の正規化された吸収スペクトル $A_{nominal}$ を備えています。

- 透過モードでは空気は基準として機能します。正規化された吸収スペクトルは、空気が光を吸収しないと想定されるので、ゼロライン($A_{nominal} = 0$)です。

図14は、検証の残差の測定方法を示します。

- 外部光ファイバーは、サンプルなしでプローブに接続してください。必要に応じてサンプルの代わりに反射標準を適用します。
- インターフェース**ウィンドウ**のコマンド **VAL REF STD** は次のスキャンを実行します。
 - 内部基準スキャンは S_{ref} の値を提供します。
 - 各チャンネルの外部スキャンは、外部光ファイバー、プローブおよび基準材料を測定します。これにより信号 S_{probe} が生成されます。
- 内部光路に関するきゅ吸光度 A_{probe} は次のとおりです。

$$A_{probe} = \log_{10} \frac{S_{ref}}{S_{probe}}$$

- ソフトウェアは A_{raw} (**測定された生スペクトル**)を計算します：

$$A_{raw} = A_{probe} - A_{fiber}$$
- ウィンドウ補正スペクトルを A_{raw} から差し引くことによって、キャリブレーション装置とプローブとの間の吸光度差が除去されます：

$$A_{corrected} = A_{raw} - A_{window}$$
 $A_{corrected}$ は、ソフトウェアで**測定され、補正されたスペクトル**として表示されます。
- 理想的には $A_{corrected}$ が**標準スペクトル** $A_{nominal}$ と一致しなければなりません。両者の差は**検証の残差**として計算されます。

$$A_{residual} = A_{corrected} - A_{nominal}$$

注記：透過モードでは $A_{nominal} = 0$ 、 $A_{residual} = A_{corrected}$ です。

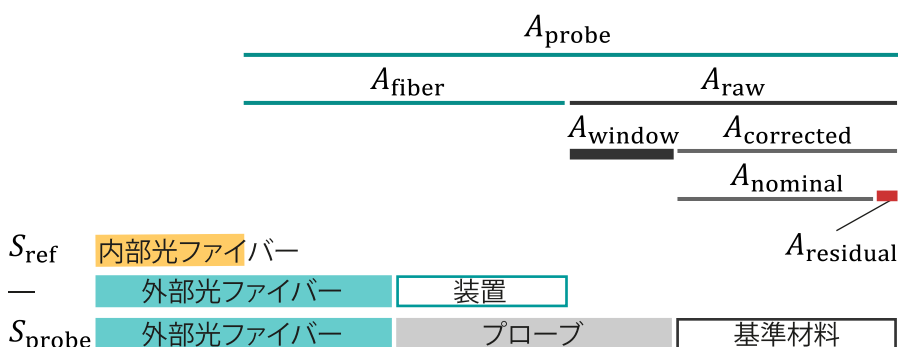


図 14 ウィンドウ補正の検証のための残差

検証の残差確認するために、波長範囲を複数のセグメントに分割します。各セグメントに対して、波長にわたる残差の二乗平均値により **RMS ノイズ** (単位：mAU) が求められます。

$$A_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^f (A_{\text{residual}_i})^2}{f}}$$

ただし、 f はセグメントの波長の数であり、 A_{residual_i} は波長 i の残差です。

各セグメントは、 A_{RMS} に対する事前に定義された許容誤差を満たしていなければなりません。すべてのセグメントが許容誤差を満たしている場合、全体の検証は成功です。

サンプルのスペクトル記録

i 装置でスペクトルを記録する前に、装置の性能テストが各チャンネルで成功している必要があります (24 ページ, 3.3 章を参照)。

サンプルのスペクトルを記録する手順は、[図 15](#) に示されています。

1. 外部光ファイバーは、プローブに接続されている必要があります。サンプルがなければなりません。
2. 吸収スペクトルは、それぞれ最後に記録された標準スペクトル S_{ref} で計算されます。 S_{ref} の値を得るために、コマンド **MEAS REF SPEC** を実行できます。
3. コマンド **MEAS SPEC** は、プローブと光ファイバーを含むサンプルを測定します。これにより信号 S が生成されます。
4. ソフトウェアは、 A_{total} において、内部光路を基準としてプローブおよび光ファイバー光学系を含むサンプルの吸光度を計算します:

$$A_{\text{total}} = \log_{10} \frac{S_{\text{ref}}}{S}$$

5. 続いて、サンプルの吸光度は上記に示されているように、各チャンネルのガラスファイバー補正スペクトル A_{fiber} およびウィンドウ補正スペクトル A_{window} を用いて計算されます:

$$A = A_{\text{total}} - A_{\text{fiber}} - A_{\text{window}}$$

A はサンプルのスペクトルを示します。

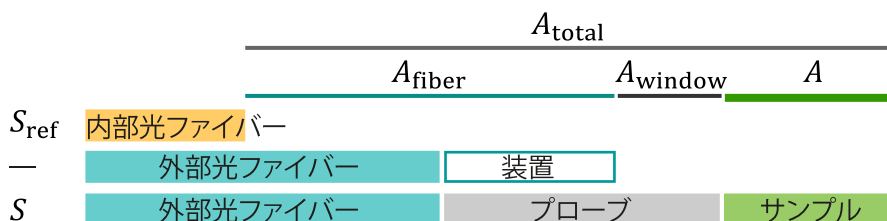


図 15 サンプルのスペクトル記録

3.3 装置の性能テスト

装置の性能テストは、内部光路および外部光路で実行できます。

- **OMNIS NIR Analyzer**

- **内部装置の性能テスト** (必須): 内部テストは、各サンプル提示の基準光路を使用します。テストは波長と信号ノイズを確認します。
テスト前に、各サンプル提示の波長のキャリブレーションが成功し、検証されなければなりません。
装置で各サンプル提示のスペクトルを記録できるようにする前に、内部テストが正常に完了している必要があります。
- **外部装置の性能テスト** (オプション): 外部テストは、USP <856>、Ph.Eur 2.2.40、JP 2.27 などに準じた検証をサポートします。波長、信号ノイズ、測光直線性を確認します (27 ページ, 3.3.1 章を参照)。

- **2060 NIR**

テストは、内部光路または外部光路を使用できます。テストは波長と信号ノイズを確認します。
テスト前に各チャンネルの波長のキャリブレーションと外部基準標準化が成功し、検証されなければなりません。
装置で各チャンネルのスペクトルを記録する前に、装置の性能テストが成功している必要があります。事前定義された許容誤差を順守する必要があります。許容誤差は、装置固有のデータで指定される各チャンネルの光ファイバー構成に依存します (測定モード、ファイバーのタイプ、ファイバーの長さ)。

波長テスト

波長テストでは、波長正確性と波長精度を確認します。その際、定義されたピークと既知のピーク位置を有する吸収スペクトルを備えた波長基準が使用されます:

- **内部:** 内部の計量学的に追跡可能な波長基準の吸収スペクトルは、内部光路によって複数回測定されます:

$$A_{WL,x} = \log_{10} \left(\frac{S_{ref}}{S_{ref,WL,x}} \right)$$

ここでは、 $A_{WL,x}$ は x 回目の測定での内部波長基準の吸収スペクトル、 S_{ref} は内部基準光路で測定された標準スペクトル、 $S_{ref,WL,x}$ は内部波長基準使用による参照パスで測定された x 回目の測定のスペクトルに相当します。

- **外部:** 製品ファミリー **OMNIS NIR Analyzer** の装置用: 外部波長テストはオプションです (27 ページ, 「外部波長テスト」を参照)。

- **外部**、タイプ **2060 NIR** の装置用:
 外部の光ファイバーは、透過時にはキャリブレーション装置に、
 反射時には反射標準に接続されている必要があります。
 内部の、計量学的に追跡可能な波長基準の吸収スペクトルは、外
 部光路によって複数回測定されます：

$$A_{WL,x} = \log_{10} \left(\frac{S_{ref}}{S_{fiber,WL,x}} \right) - A_{fiber}$$

ここでは、 $A_{WL,x}$ は x 回目の測定での内部波長基準の吸収スペクトル、 S_{ref} は内部基準光路で測定された標準スペクトル、 $S_{fiber,WL,x}$ は、各チャンネルの光路で測定された x 回目の測定のスペクトルに相当しますが、その際、ファイバーはキャリブレーション装置または反射標準に接続されています。内部波長基準は光路に挿入されており、 A_{fiber} は、リファレンスの標準化から得られたグラスファイバー補正スペクトルに相当します。

$A_{WL,x}$ は反射標準の吸光度も含みますが、次の計算には無関係です。理想的には、 $A_{WL,x}$ のピーク位置は、波長基準の正規化されたピーク位置と同一です。

波長正確性と波長精度は以下のように確認が行われます：

1. スペクトル $S_{ref,WL,x}$ もしくは $S_{fiber,WL,x}$ を 10 回測定し、その結果、10 の吸収スペクトル $A_{WL,x}$ が得られます。
2. 吸収スペクトルではピーク位置が同定されます。
3. 各ピーク位置に対して、次の統計は 10 の吸収スペクトルから計算されます：
 - a. 平均値 (単位: nm)
 - b. 標準偏差 (単位: nm)
4. **精度**：各ピークについて、中間ピーク位置と正規化されたピーク位置の差が事前定義された許容誤差内でなければなりません。
 ➡ **注記**：正規化されたピーク位置はテストにより若干変動する可能性があります。これは、ピーク位置が温度補正されるためです。同様の温度補正は、波長のキャリブレーションの場合に発生します。これは、特定の温度での測定がすべての装置で同等の結果を確実に提供することに役立ちます。
5. **正確性**：各ピークについて、標準偏差は事前定義された許容誤差内でなければなりません。
6. すべてのピークの許容誤差が満たされている場合、波長テストの総ステータスは成功です。

ノイズテスト

信号ノイズは内部または外部で確認できます：

- **内部**：ノイズは、同一の光路における基準測定の吸光度を基準として、内部光路の吸光度から複数回算出されます：

$$A_{\text{noise},x} = \log_{10} \left(\frac{S_{\text{ref}}}{S_{\text{ref},x}} \right)$$

ここでは、 $A_{\text{noise},x}$ は x 回目の測定のノイズスペクトル、 S_{ref} は内部基準光路で測定された標準スペクトル、 $S_{\text{ref},x}$ は同じパス上で測定された x 回目の測定のスペクトルに相当します。

- **外部**、製品ファミリー **OMNIS NIR Analyzer** の装置用：外部ノイズテストはオプションです (27 ページ、「外部ノイズテスト」を参照)。

- **外部**、タイプ **2060 NIR** の装置用：

外部の光ファイバーは、透過時にはキャリブレーション装置に、反射時には反射標準に接続されている必要があります。

ノイズは、測定された吸収スペクトルと公称吸収スペクトルの差分として複数回測定されます：

$$A_{\text{noise},x} = \log_{10} \left(\frac{S_{\text{ref}}}{S_{\text{fiber},x}} \right) - A_{\text{fiber}} - A_{\text{nominal}}$$

ここでは、 $A_{\text{noise},x}$ は x 回目の測定のノイズスペクトル、 S_{ref} は内部基準光路で測定された標準スペクトル、 $S_{\text{fiber},x}$ は、各チャンネルの光路で測定された x 回目の測定のスペクトルに相当しますが、その際、ファイバーは、それぞれキャリブレーション装置および反射標準に接続されています。 A_{fiber} は、リファレンスの標準化から得られたガラスファイバー補正スペクトルであり、 A_{nominal} は、反射標準の公称吸収スペクトルに相当します。

注記：透過モードでは $A_{\text{nominal}} = 0$ です。

理想的には $A_{\text{noise}} = 0$ です。

ノイズテストでは以下の手順を実行します：

1. スペクトル $S_{\text{ref},x}$ および $S_{\text{fiber},x}$ を 10 回測定し、その結果、10 のノイズスペクトル $A_{\text{noise},x}$ が得られます。
2. ノイズスペクトルは様々な波長セグメントに分割されます。
3. 各ノイズスペクトルと各セグメントに対し、3 つの数量が計算されます：
 - a. 測光ノイズ (単位：mAU)
 - b. ピーク ツー ピークノイズ (単位：mAU)
 - c. ノイズのベースラインバイアス (単位：mAU)
4. 各セグメントの 3 つの数量それぞれについて、10 のノイズスペクトルの平均値が計算されます。
5. すべての平均値が事前定義された許容誤差内にある場合、ノイズの総ステータスは成功です。

3.3.1 外部装置の性能テスト (OMNIS NIR Analyzer)

製品ファミリー **OMNIS NIR Analyzer** の装置は、USP <856>、Ph.Eur 2.2.40、JP 2.27 などの薬局方に準じて検証できます (OMNIS Software のバージョン 4.4 以降)。このテストには外部の計量的に追跡可能な参照基準が必要となります。参照基準には、基準装置を使って環境温度で測定した、個別の公称吸収スペクトルがあります。

外部波長テスト

波長正確性と波長精度は以下のように確認が行われます：

1. 外部波長基準 (透過または反射) をポジションに配置する必要があります。
2. 外部波長基準の 10 の吸収スペクトルが記録されます：

$$A_{WL,x} = \log_{10} \left(\frac{S_{ref}}{S_{WL,x}} \right)$$

ここで、 $A_{WL,x}$ は x 回目の測定における外部波長基準の吸収スペクトル、 S_{ref} は内部基準光路で測定された標準スペクトル、そして $S_{WL,x}$ は外部波長基準によって測定された x 回目の測定のスペクトルに相当します。

3. 10 の吸収スペクトル $A_{WL,x}$ ではピーク位置が同定されます。
4. 各ピーク位置に対して、次の統計は記録されたスペクトルから計算されます：
 - a. 平均値 (単位: nm)
 - b. 標準偏差 (単位: nm)
5. **精度**：各ピークについて、中間ピーク位置と正規化されたピーク位置の差が事前定義された許容誤差内でなければなりません。
6. **正確性**：各ピークについて、標準偏差は事前定義された許容誤差内でなければなりません。
7. すべてのピークの許容誤差が満たされている場合、波長テストの総ステータスは成功です。

外部ノイズテスト

信号ノイズは、低い光子ストリーム (ローフラックス) で 1 回、高い光子ストリーム (ハイフラックス) で 1 回確認します：

1. 低フラックステストまたは高フラックステスト用の外部波長基準 (透過または反射) をポジションに配置する必要があります。

2. 測定された吸収スペクトルと参照基準の公称吸収スペクトルとの差分として 10 のノイズスペクトルが記録されます：

$$A_{\text{noise},x} = \log_{10} \left(\frac{S_{\text{ref}}}{S_{\text{ND},x}} \right) - A_{\text{nominal}}$$

ここでは、 $A_{\text{noise},x}$ は x 回目の測定のノイズスペクトル、 S_{ref} は内部基準光路で測定された標準スペクトル、 $S_{\text{ND},x}$ は外部参照基準で測定された x 回目の測定のスペクトル、 A_{nominal} は参照基準の公称スペクトルに相当します。

3. 10 のノイズスペクトルは様々な波長セグメントに分割されます。
4. 各ノイズスペクトルと各セグメントに対し、3 つの数量が計算されます：
 - a. 測光ノイズ（単位：mAU）
 - b. ピーク ツー ピークノイズ（単位：mAU）
 - c. ノイズのベースラインバイアス（単位：mAU）
5. 各セグメントの 3 つのそれぞれの数量に対し、記録されたノイズスペクトルで平均値が計算されます。
6. すべての平均値が事前定義された許容誤差内にある場合、ノイズの総ステータスは成功です。

測光直線性

このテストの目的は、全波長範囲で反射率（または透過率）と測定された吸光度の線形関係を検出することです：

1. 異なる反射率（または透過率）を有する 5 つの参照基準の吸収スペクトルが記録されます。
2. 反射率（または透過率）と測定された吸光度の線形関係が、線形回帰によって複数の波長で確保されます。
3. すべての回帰直線のスロープおよび y インターセプトが事前定義された許容誤差内にある場合、テストの総ステータスは成功です。

4 モデル開発

以下のモデルの種類に区別されます：

- **定量化モデル**は、記録したサンプルのスペクトルの対象パラメータ（例えば水分）の依存性を記述します。
- **識別モデル**（OMNIS Software のバージョン 4.0 以降）は、それぞれの製品（例えばそれぞれの種類のコーヒー豆）の記録されたスペクトルに従ってサンプルを分類します。
製品は、特定の化学物質、または特定の物理的プロパティ（例えば粒径）を備える特定の化学物質を示します。

未知のサンプルの分析に対して、サンプルのスペクトルが記録されます。アプリケーションに応じてスペクトルは次のように使用されます。

- 定量化：スペクトルに基づいて、定量化モデルは、例えばサンプルの水分に対する予測を作成します。
- 識別：スペクトルを基に、識別モデルは、例えばアラビカコーヒーとしてなど、サンプルを識別します。
- 検証（OMNIS Software のバージョン 4.2 以降）：スペクトルを基に、識別モデルは、例えばサンプルがアラビカコーヒーであるかなどを検証します。

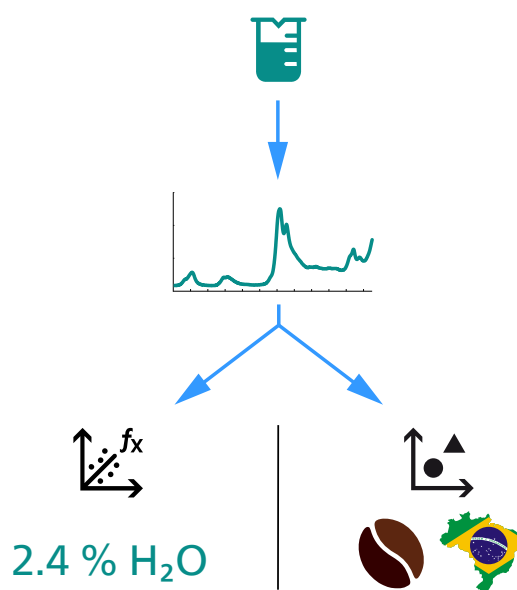


図 16 定量化（左下）および識別（右下）。

モデル開発とサンプル分析には次の手順が含まれています。

1. サンプルング

物理的にサンプルが採取され、処理されます。

- a. 各サンプルについて、1つのスペクトルが記録されます。
- b. 定量化では、対象パラメータ（例えば水分）の基準値が基準メソッド（例えば滴定）で測定されます。基準測定は精度があり、正確である必要があります。
- c. 識別では、サンプルの製品属性が既知でなければなりません。

2. モデル開発

モデル開発は、次の手順を含む反復プロセスで行われます。

- a. 校正セット、検証セットおよびアウトライヤーデータセットのデータセットの分割。
- b. 適切なデータ前処理のアプリケーションとスペクトルの波長範囲。
- c. 校正セットを基にしたモデルの計算。
- d. モデルの検証により、モデルが要件を満たしていることを確認します。検証は、まず第一に、モデルの開発では使用されなかった検証セットに依存します。
定量化では、モデルは検証セットのスペクトルの対象パラメータを予測します。続いて、算出値は既知の基準値と比較されます。
識別ではモデルはスペクトルをそれぞれの製品に割り当てます。予測された製品属性は、それぞれの実際の製品属性と比較されます。

3. モニタリング

モデルのモニタリングにより、時間の経過とともに予測能力が低下しないことが保証されます。プロセスやサンプルを変更した場合には再検証が必要です。

4.1 物理的サンプル

物理的サンプルの収集と分析から始まります。適切なサンプルングは、堅牢なモデル開発にとっての前提条件です。考慮すべき点はいくつかあります。

変動範囲

サンプルには、一般的に将来予想されるサンプル変動を含める必要があります。すべての化学成分の濃度と粒径は、少なくとも予測される変動範囲をカバーする必要があります。

サンプルは、適切な条件バリエーションと適切な時間をカバーする必要があります。プロセス変動、季節変動、環境条件などのすべての変動を考慮する必要があります。

サンプルは、変動範囲全体にわたって均一に分割される必要があります。定量化では、これには基準値の範囲が含まれます。例えば、

基準値の範囲が 1%～10%である場合、サンプルは 1%～10%で均等に分布する必要があります。

校正サンプルおよび検証サンプル

一般的に 2 つのサンプルセットを使用します。

- 校正セット：モデル開発に使用するサンプル。
- 検証セット：モデル検証に使用するサンプル。

2 つのセットは予想される変動をカバーしている必要があります。モデルの堅牢性を確認するには、検証セットに対して独立したサンプルを収集することが望ましいです。それぞれのオペレータ、それぞれのサプライヤまたは必要に応じてそれぞれの装置などのシナリオを考慮する必要があります。

キャリブレーションと検証用に独立したサンプルを収集することが現実的でない場合、利用可能なサンプルのスペクトルを校正セットと検証セットに分割することも可能です。セットの独立性を保証するために、自動分配アルゴリズムを使用してください。

モデルが開発されると、アウトライヤーの検出を確認するために、明示的なアウトライヤーサンプルの使用を考慮できます。

すべてのサンプルは同じ方法で取り扱う必要があります。スペクトルを記録する場合には、同じハードウェア構成と同じ測定パラメータを使用した同じメソッドを使用する必要があります。定量化の場合には、同じ測定パラメータを使用した同じ基準メソッドを使用する必要があります。

サンプル数

条件、化学成分、粒径のより多くの変動をカバーする必要があるほど、より多くのサンプルが必要です。

定量化：統計分析が適切に機能するには、少なくとも約 50 個のサンプルが必要で、校正セットおよび検証セットにはそれぞれ少なくとも 20～25 個のサンプルが含まれている必要があります。

識別と検証：各製品に対し、サンプルが予想されるバリエーションをカバーしている必要があります。製品は異なるサンプル数を有することができ、最低数は 3 です。

少なくとも 10～20 個のサンプル(バリエーションの数に応じて)により、最初のモデルは検証セットなしで開発できます。交差検証(定量化モデル用)または内部検証(識別モデル用)が適切なモデルを作成できることを示唆している場合、最終モデルを開発するために、引き続き校正サンプルおよび検証サンプルを収集する必要があります。

複製

特定の条件または基準値の範囲のサンプルが非常に少ない場合があります。これを補うために、これらのサンプルを複製したくなる場

- **基準値アウトライヤー（定量化）**

定量化の場合、個々の基準値が異常を示し、後で基準値アウトライヤーとして認識される場合があります（53 ページ、4.3.4 章を参照）。

OMNIS Model Developer (OMD)は、ASTM D8321-22 に準拠したアウトライヤー検出に基づいて、該当するサンプルをアウトライヤーとして検出します（65 ページ、4.4.3 章を参照）。

4.2 主成分分析(PCA)

校正サンプルの分光データには、多数の変数（波長）が含まれています。変数は高く相互に関連しています。そのため、データの冗長性が高くなります。この種のデータを取り扱うためには、PCA や PLS などの潜在変数を使用します。

主成分分析（PCA、英語で *principal component analysis*）は、基準値を考慮することなくスペクトルに焦点を当てています。

i OMNIS Software は、モデル開発中、自動データセット分割とスペクトルのアウトライヤー検出に対して PCA を使用します。

準備手順

次の準備手順が必要です。

1. **データ前処理**：OMNIS Software は、指定されたデータ処理をスペクトルに適用します（37 ページ、4.3.1 章を参照）。
2. **波長範囲**：OMNIS Software は、指定された波長選択をスペクトルに適用します（45 ページ、4.3.2 章を参照）。
3. **平均センタリング**：各波長に対して、平均吸光度値が計算され、各スペクトルの各波長から差し引きます。

1 番目の主成分

準備手順の後、PCA はスペクトルデータ内の情報を再配置し、関連データをノイズから分離します。この目的のために、PCA は波長変数を新しい変数空間、つまり**主成分（PC、英語で Principal Components）**に変換します。

PCA は、さまざまな波長から関連情報をわずか数個の主成分に変換します。簡単な例によってこのコンセプトを説明するために、波長変数が数千ではなく、2つだけあり、これら2つの変数が1つの主成分に低減されると仮定します。

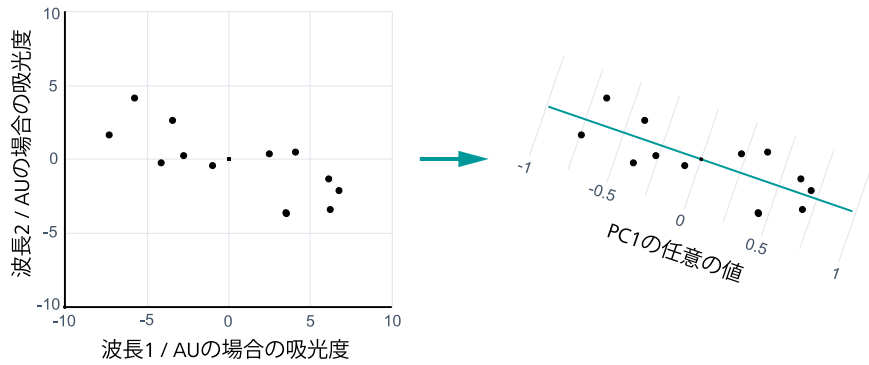


図 17 2次元波長空間内のスペクトルを示す点 (左)。1次元主成分空間における同じ点 (右)。

左図17において、横軸と縦軸が元の2変数の波長空間を形成しています。したがって、各点は2つの波長のみのスペクトルを示します。すべての波長値の平均値はゼロ点を形成します。

右では、最大の分散を説明するデータの方向が主成分 PC1 を形成します。この例では、PC1 が主成分空間の唯一の変数です。その結果、元の2つの変数は1つに低減します。

スコアと残差

図18は、スペクトル*i*を特徴づける量を示しています。

- 主成分空間で測定した、中央からの距離 s_i 。主成分が1つのみの例では、 s_i は PC1 の方向で測定されます。距離 s_i は、スペクトル i のスコアと呼ばれます。
- 主成分空間からスペクトルのオフセット e_i 。距離 e_i は、スペクトル i の残差と呼ばれます。

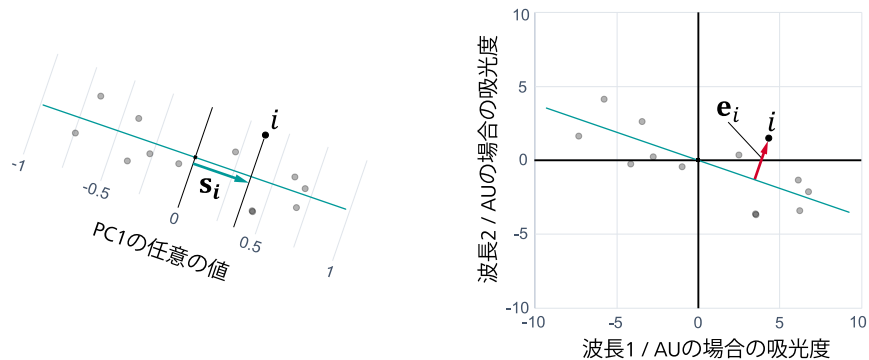


図 18 スコア (左) および残差 (右) のスペクトル*i*。

i スコア s_i は主成分空間で測定されます。残差 e_i は元の波長空間で測定されます。

複数の主成分への変換

一般的に、分光データの適切な説明には複数の主成分を必要とします。

図19には3つの元の変数 x_1 、 x_2 、 x_3 があります。各点は3つの波長のスペクトルを示します。

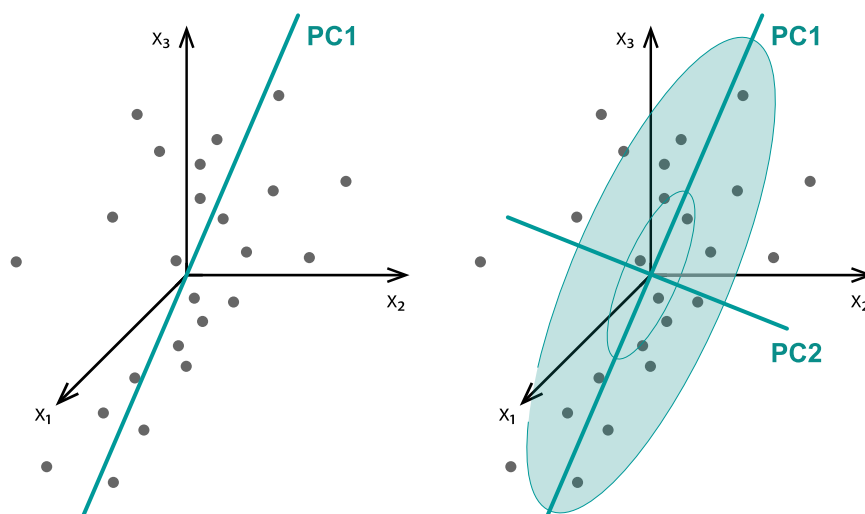


図 19 3つの元の変数は、1つの主成分（左）または2つの主成分（右）に低減されます。PC1 およびPC2 は2次元主成分空間を形成します。

1番目の主成分 PC1 は、やはり最大の分散を説明するデータの方法です。

2番目の主成分 PC2 は、残りの最大分散を説明するデータの方法です。同じことが後の主成分にも適用され、それぞれが残りの最大分散を示します。そのため、最初のいくつかの主成分はデータの分散の大部分を占めますが、その他の主成分には主にノイズが含まれるため、無視できます。このようにして変数の数を減らすことができます。

PCAの主な特徴は、すべての主成分が互いに**直交**（直角）であるということです。そのため、スコアには相互関係はありません。

マハラノビス距離

上記に示されたように、スペクトル i のスコアは主成分空間で測定される一方、残差は元の波長空間で測定されます。

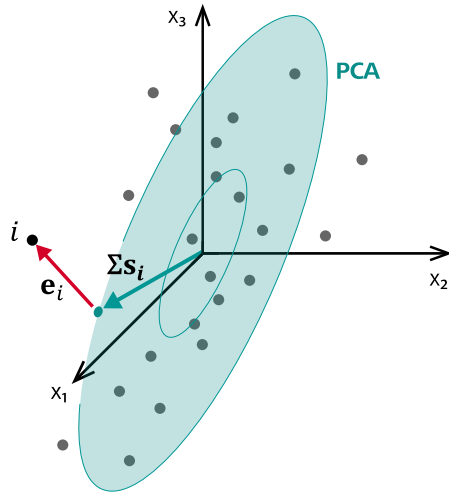


図 20 スペクトル*i*のスコアと残差。緑の点は点*i*(スペクトル*i*を示す)の主成分空間への正投影です。

図 20 において、スコアベクトル Σs_i は、PCA モデルの中心から、主成分空間へのスペクトルの正投影までの絶対距離（ユークリッド距離）を示します。



例ではスペクトルのユークリッド距離は、PC2 方向よりも PC1 方向でさらに離れています。伝播は**分散**として測定できます。PC1 の分散は、PC2 の分散より大きいです。

正規化されたスコアベクトル s_i は、正規化された距離、つまり**マハラノビス距離**を示します。マハラノビス距離は、それぞれの主成分方向のそれぞれの分散を考慮します。各方向には同じ重み付けが含まれています。そのため、小さい分散の方向の小さなユークリッド距離は、より大きい分散の方向の大きなユークリッド距離と同じくらい重要です。

複数の波長のスペクトルの変換

同じコンセプトが、さまざまな波長変数のスペクトルを主成分に変換する場合にも当てはまります。図 21 では、各スペクトルはカーブ（左）と点（右）で示されます。

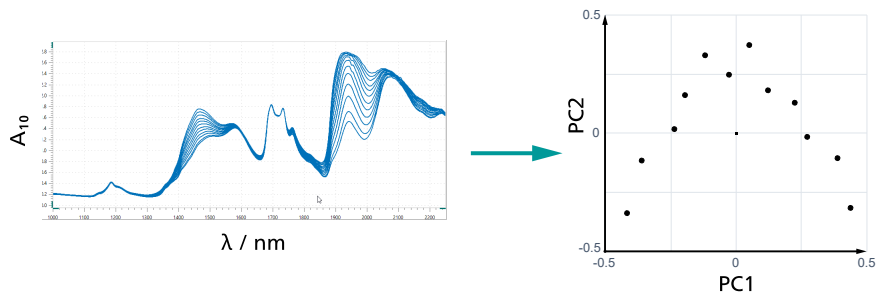


図 21 スペクトルデータの主成分空間への変換。右ページのスコアは任意の単位で示されます。

右図は、最初の 2 つの主成分である PC1 および PC2 を示します。同じ方法で、続く主成分 PC3、PC4 などを示すことができます。

PCA モデルは、固定数の主成分を使用します。主成分が増えるほど、モデルが説明する関連のスペクトル変動が多くなります。同時に、モデルは多くの無関係なスペクトル変動（ノイズ）も補足します。バランスの取れた妥協が必要になります。

i OMNIS Software が主成分分析を実行すると、主成分の数は、説明された分散が少なくとも 95 % となるように選択されます。

PCA アルゴリズム

元データを主成分空間に変換する方法は複数あります。OMNIS Software は特異点分解を実行します (86 ページ, 6.2 章を参照)。

4.3 データ前処理

4.3.1 データ前処理

分光モデルは、吸光度値と対象パラメータ (定量化) または製品属性 (識別、検証) との間の関係に基づいています。スペクトルの**パラメータ設定**により、スペクトルがこの関係を適切に表現できるようになります。その目的は、有用な情報を失うことなく無関係な分散を削除することです。アーチファクトおよび非線形性を修正します。パラメータ設定が正しく行われると、モデルの精度と堅牢性が向上し、予測の繰り返し性と再現性も向上します。

パラメータ設定は、校正セット、検証セット、アウトライヤーデータセットに加え、モデルで分析される今後の未知のサンプルにも適用されます。

パラメータ設定の始めの手順は、**データ前処理**です。データ前処理は指定された順序で行われます。パラメータ設定の 2 番目の手順では、関連する波長範囲を定義できます (45 ページ, 4.3.2 章を参照)。

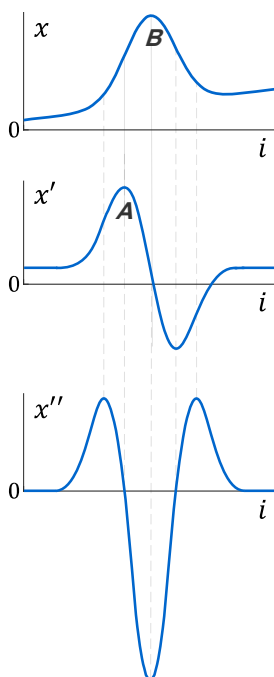
ノイズ低減

スペクトルは、信号に関するさまざまな種類のランダムな変動を含む可能性があります。この例では、検出器および装置の電子回路によって発生する高周波の伊豆屋、スキャン測定中の装置のドリフトによって発生する低周波ノイズがあります。

スペクトロメーターは一連の単一測定で平均化されたスペクトルを提供します。それにより、高周波ノイズは顕著に低減されます。さらにノイズを低減するにはスムージングフィルターにより実現できます。このフィルターは、ノイズは高周波であり、信号は低周波であるという考えに基づいています。これは、信号を隣接する吸光度値により近似し、平均化によりノイズを低減します。

4.3.1.1 導関数

i 導関数は、Gap-Segment フィルターまたは Savitzky-Golay フィルターにより実行できます。



スペクトルの導関数は、各点におけるカーブのスロープまたはスロープを示します。スロープは出力スペクトルの変化です。

スペクトルでは、 x_i は波長 i の場合の吸光度です。一次導関数 x'_i は波長 i の場合のスペクトルのスロープを示します。出力スペクトルが最も急である場合、一次導関数は最大値を持ちます(A)。出力スペクトルがピーク(B)を有する場合、一次導関数は 0 です。

一次導関数はベースラインとオフセットを除去し、ベースラインの傾きをベースラインのオフセットに変換します。

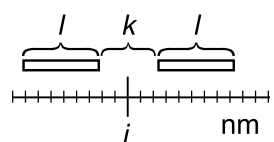
二次導関数 x''_i は、波長 i の場合の一次導関数のスロープに対応します。元のスペクトル(B)における正のピークは負のピークとなり、その逆も同様です。

二次導関数は、元のスペクトルからベースラインのオフセットとベースラインの傾きを除去します。

元のスペクトルに大量のノイズが含まれている場合には注意が必要です。各導関数により、信号ノイズ比が大幅に悪化します。この理由により、Gap-Segment フィルターや Savitzky-Golay フィルターでは、導関数をスムージング関数と組み合わせます。

4.3.1.2 Gap-Segment

Gap-Segment フィルターはスペクトルをスムージング。オプションでは、Gap-Segment フィルターは一次導関数または二次導関数を実行します。計算は、導関数を使用したかどうかにかかわらず依存しています。



- **微分階数 0**：各波長 i に対して、Gap-Segment フィルターはセグメントサイズ l (10 nm など) の 2 つのセグメントの平均値を計算します。2 つのセグメントはサイズ k (5 nm など) の距離により分割されます。
- **微分階数 1**：一次導関数に対して、2 つのセグメントの平均値は個別に計算されます。続いて、2 つの平均値の差が形成されます。
- **微分階数 2**：二次導関数は、同様の方法で一次導関数から計算できます。

スペクトルの始めと終わりには、スペクトルの外にあるセグメント波長のゼロ値を使用して、 $l + k/2$ 波長が計算されます。

スペクトルの始めと終わりには、スペクトルの外にあるセグメント波長に対してゼロ値が使用されます。

スムージングにより、ピークにわずかなシフトやバイアスが伴う可能性があります。

パラメータ設定

より強力なスムージングは、次により実現されます。

- 低次導関数
- 大きなセグメントサイズ
- セグメント間隔

i 過度なスムージングにより、関連の分散の損失をもたらし、モデルの予測能力を低下させます。

4.3.1.3 Savitzky-Golay

Gap-Segment フィルターのように、Savitzky-Golay フィルターはスペクトルをスムージング氏、オプションで一次導関数または二次導関数を実行します。ただし、Savitzky-Golay フィルターは別のスムージングメソッドを使用します。

各波長 i に対して、Savitzky-Golay フィルターはそれぞれの波長の範囲内で低次の多項式を当てはめます。波長 i の多項式の値は、スムージングされた値です。導関数を作成する場合には、導関数の値を使用します。

隣接する値の加重合計により、すべてが一度に計算されます。

$$x_i = \sum_{j=-k/2}^{k/2} c_j x_{i+j}$$

ただし、 k はフィルター幅、 c_j は微分階数、多項式次数およびフィルター幅に依存し、表中で検索できる畳み込み係数、 x_{i+j} は波長 $i+j$ の場合の出力スペクトルの吸光度値です。

スペクトルの始めと終わりには、スペクトルの外にあるフィルター波長に対して外挿値を使用します（水平外挿）。

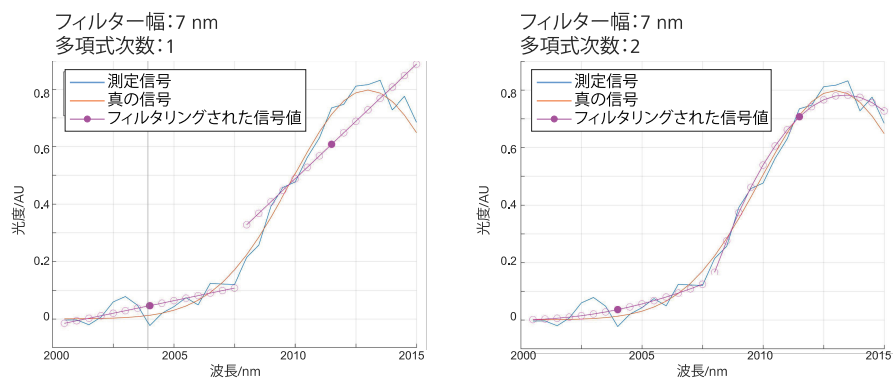


図 23 それぞれの多項式次数を備える Savitzky-Golay フィルタリング

図 23 は Savitzky-Golay フィルタリングを示します。フィルター幅は 7 nm です。波長 2,004 nm と 2,011.5 nm の多項式が示されています

す。新しい値は塗り潰された点、または導関数を使用する場合には、その導関数です。その他のすべての波長は、同じ方法で取り扱われます。

フィルター幅は、各多項式が適合する波長範囲を定めます。畳み込みは、吸光度値の影響がそれぞれの波長の両側に向かって減少するように重み付けされます。

パラメータ設定

より強力なスムージングは、次により実現されます。

- 低次導関数
- 大きいフィルター幅
- 低い多項式次数

i 過度なスムージングにより、関連の分散の損失をもたらし、モデルの予測能力を低下させます。

4.3.1.4 SNV – Standard Normal Variate (標準正規変量)

SNV は単一のスペクトルを分散 1 および平均値 0 に正規化します。SNV は指定された波長範囲内の各波長 i の吸光度値 x_i を次のように正規化します。

$$x_i = \frac{x_i - m}{s}$$

ただし、 m は平均値、 s は指定された波長範囲内のすべての吸光度値の標準偏差です。

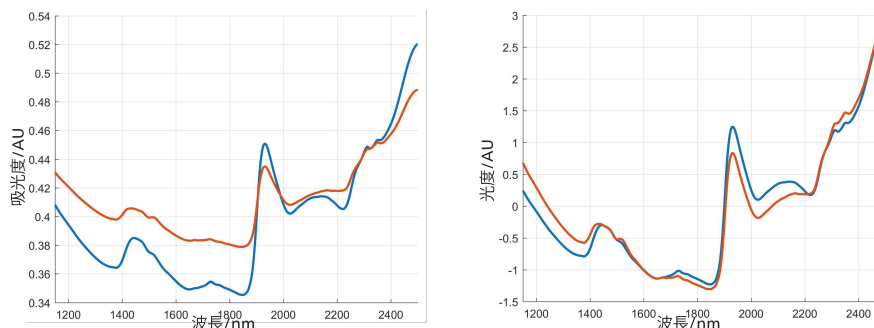


図 24 吸収スペクトル (左) および SNV で取り扱われたスペクトル (右)。

正規化によりスペクトル間の分散が除去されます。分散が、重要なプロパティ、例えば粒状または粉末のサンプル、あるいは濁った媒体における層厚の違いに起因する場合、これは理にかなっていません。

注記：SNV の後に導関数が適用される場合、除去された一部の分散が再び出現する場合があります。そのため、導関数は SNV の前に適用する必要があります。SNV が例外的に導関数の前に実行する必要があります場合、detrend、SNV、導関数の順序を考慮できます。ただし、

通常の順序は導関数、SNV、detrend です (44 ページ, 4.3.1.7 章を参照) .

パラメータ

▪ **波長範囲**

アーティファクトが特定の波長範囲に影響を及ぼす場合 (例えば飽和や強いノイズ)、これらの範囲を除外できます。

平均値と標準偏差を計算する場合、定義された波長範囲のみが考慮されます。続けて行われる正規化は、定義された波長範囲でも中間範囲でも実行されます。スペクトルの始端で除外された波長については、隣接する開始波長の正規化値が適用されます。スペクトルの終端で除外された波長については、隣接する終了波長の正規化値が適用されます。

必要に応じて、除外された波長をモデルの計算でも除外することができます (45 ページ, 4.3.2 章を参照)。

注記: OMNIS Software のバージョン 4.6 以降では、複数の波長を定義できます。

4.3.1.5 detrend

detrend は、最小二乗法を使用して、二次多項式をスペクトル全体に当てはめます。続いて、detrend はスペクトルから多項式を差し引きます。

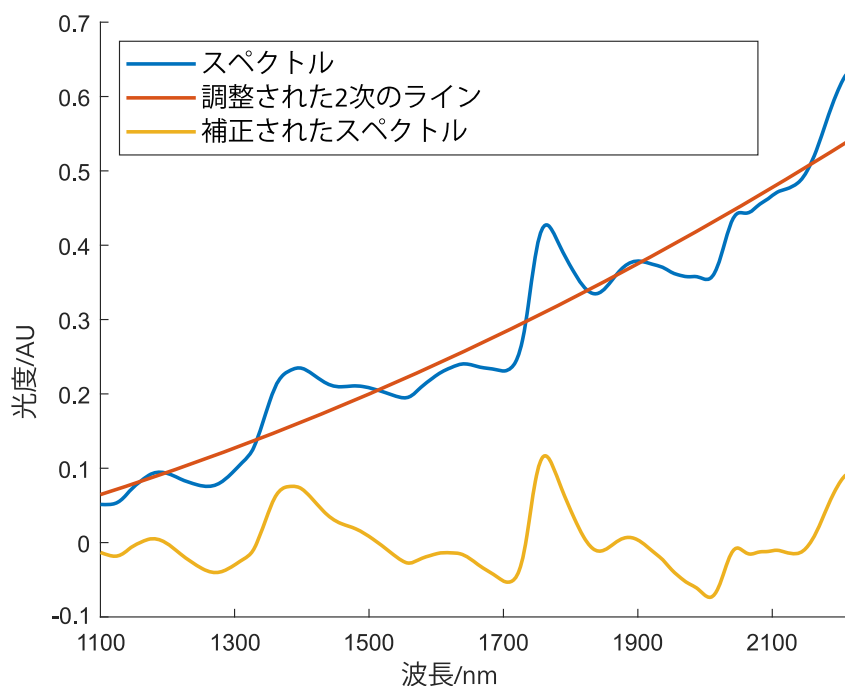


図 25 detrend は青色のスペクトルを黄色のスペクトルに変換します。

detrend は、波長依存の散乱効果を、ベースラインの二次の傾きまで低減します。

上図は、トレンドが支配的であるスペクトル（青）を示しています。この支配的なトレンドがすべてのスペクトルで同様である場合、detrend は好ましく機能できます。他の場合には、detrend は有用なバリエーションを削除する傾向があります。このような場合には、導関数は良い選択となる場合があります。

各スペクトルに個別の多項式が適合されるので、さらに厄介な分散が発生する可能性があります。通常、SNV は detrend の前に適用されます。これにより、多項式の係数のより堅牢な推測が可能になります。

パラメータ

波長範囲

アーティファクトが特定の波長範囲に影響を及ぼす場合（例えば飽和や強いノイズ）、これらの範囲を除外できます。

多項式は、定義された波長範囲のすべての強度値に調整されます。続いて、定義されているすべての波長範囲でスペクトルの多項式が減算されます。除外された波長に対しては、強度値はゼロに設定されます。

必要に応じて、除外された波長をモデルの計算でも除外することができます（45 ページ、4.3.2 章を参照）。

注記: OMNIS Software のバージョン 4.6 以降では、複数の波長を定義できます。

4.3.1.6 データ前処理の概要

前処理	目的	正の効果	負の効果
Gap-Segment	スムージング 微分階数が低い、セグメントサイズが大きい、あるいはセグメント間隔が大きいほど、強力なスムージングが行われます。	<ul style="list-style-type: none"> 高周波ノイズが低減されます。 	<ul style="list-style-type: none"> 過度にスムージングを行うと、関連の分散の損失につながります。
Gap-Segment 付きの導関数	ベースライン補正	<ul style="list-style-type: none"> 一次導関数：ベースラインのオフセットの除去。 二次導関数：ベースラインのオフセットとベースラインの傾きを除去。 	<ul style="list-style-type: none"> ノイズを増幅します。 スペクトルの外観を変更します。

前処理	目的	正の効果	負の効果
Savitzky-Golay	スムージング 微分階数が低い、フィルター幅が大きい、あるいは多項式次数が低いほど、強力なスムージングが行われます。	<ul style="list-style-type: none"> 高周波ノイズが低減されます。 	<ul style="list-style-type: none"> 過度にスムージングを行うと、関連の分散の損失につながります。
Savitzky-Golay の導関数	ベースライン補正	<ul style="list-style-type: none"> 一次導関数：ベースラインのオフセットの除去。 二次導関数：ベースラインのオフセットとベースラインの傾きを除去。 	<ul style="list-style-type: none"> ノイズを増幅します。 スペクトルの外観を変更します。
SNV – Standard Normal Variate (標準正規変量)	散乱補正*	<ul style="list-style-type: none"> ベースラインのオフセットの除去。 	<ul style="list-style-type: none"> 関連の分散は、必要に応じて除去されません。
detrend	散乱補正*	<ul style="list-style-type: none"> ベースラインのオフセットの除去。 ベースラインの傾きとベースラインの二次の傾きを除去します。 	<ul style="list-style-type: none"> 関連の分散は、必要に応じて除去されません。 無関係な分散となる可能性があります。

*注記：アーティファクト（飽和や強いノイズなど）のある波長範囲は、除去する必要があります。

4.3.1.7 複数のデータ処理手順の場合の順序

複数のデータ処理手順を使用する場合、順序が重要になる場合があります。基本的なルールは次のとおりです。

i Gap-Segment または Savitzky-Golay は SNV の前に使用し、SNV は detrend の前に使用することが好ましいです。

一次導関数と SNV を使用する例：一次導関数はベースラインの傾きをベースラインのオフセットに変換します。後続の SNV はこのオフセットを除去します。逆の順序の場合、SNV はベースラインの傾きを変更しません。後続の一次導関数は、これをベースラインのオフセットに変換します。オフセットは残ります。

二次導関数および SNV を使用する例：ベースラインのオフセットとベースラインの傾きは、いずれの場合も除去されます。ただし、二次導関数と SNV を正しい順序で使用することにより、二次のベース

ラインの傾きを除去できます。二次導関数は、ベースラインの二次の傾きをベースラインのオフセットに変換します。後続の SNV はこのオフセットを除去します。逆の順序の場合、SNV はベースラインの二次の傾きを変更しません。後続の二次導関数は、これをベースラインのオフセットに変換します。オフセットは残ります。

4.3.2 波長範囲

データ前処理後 (37 ページ, 4.3.1 章を参照) にパラメータ設定の 2 番目の手順を行います：波長範囲の選択により、目的に適していない範囲を除外できます。特に、ノイズが多い、あるいは飽和している波長範囲は、後続の計算に影響を及ぼす可能性があり、除外する必要があります。

ノイズ

ノイズは、少量の光しか検出器に達しない場合、高い吸光度値において発生します。次の図はノイズが発生する範囲を示します。

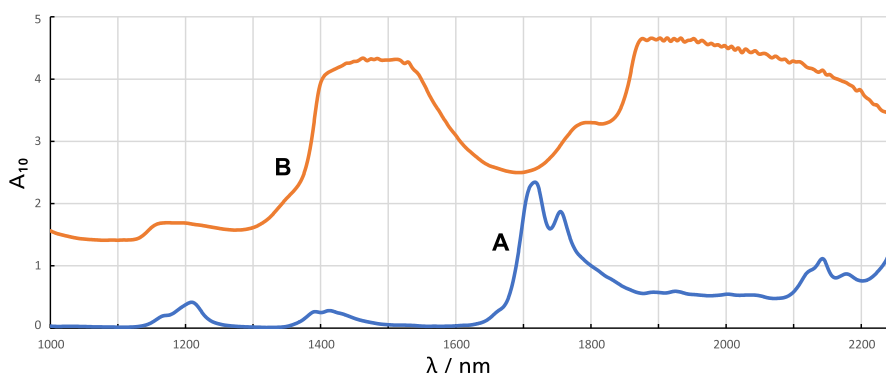


図 26 ノイズが発生した波長範囲の例。

スペクトル A には通常のピーク形状があります。スペクトル B には 2 つのノイズが発生する範囲があります：1,400～1,550 nm と 1,870 nm 以上です。ノイズが発生する範囲では強くノイズが発生しており、ベル曲線やその組み合わせには似ていません。

飽和

大きな光量が検出器に到達する、つまり吸光度値が小さい場合、検出器が飽和します。

- **OMNIS NIR Analyzer**
積分時間は常に自動的に調整されます。これにより飽和を回避し、ノイズを最初源に抑えます (8 ページ, 「積分時間」を参照)。
- **2060 NIR**
自動積分時間が有効である場合、飽和は発生しません。
手動積分時間が有効である場合、積分時間が長すぎて検出器が飽和する場合があります。飽和した範囲は小さい吸光度値で発生しますが、視覚的に検出することは必ずしも容易ではありません。そのため、手動積分時間は十分な余裕をもって設定してください (8 ページ, 「積分時間」を参照)。

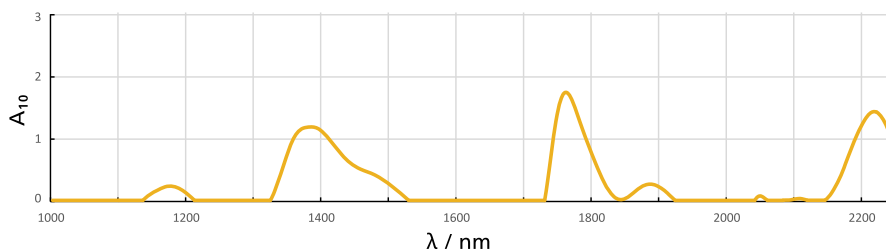


図 27 飽和した波長範囲の例。

その他の理由

波長範囲を含めたり除外する理由は他にもあります。対象パラメータと各吸収帯域の知識に基づいて選択できます (3 ページ, 2.1 章を参照)。ただし、前処理の種類に応じて、関連する情報が他の波長範囲にシフトする可能性があるということを考慮してください。

データ前処理時にスペクトルの始めと終わりに異常が挿入された場合、該当する波長を除去できます。

化学成分の変動または環境条件の変動は、特定の波長範囲に影響を及ぼす可能性があります。この波長範囲を除去することにより、モデルの堅牢性を向上できます。

i 適切に形成された波長範囲を除外する場合には注意してください。情報が含まれていないように見える範囲は、実際には隠された重要な情報が含まれている可能性があります。これらは、アウトライヤーの検出や干渉する吸収帯域を処理する場合に有効です。実際のところ、干渉する吸収帯域は、多変量測定を実行する主な理由です (82 ページ, 6.1 章を参照)。

4.3.3 スペクトルアウトライヤー

他のほとんどのスペクトルと異なるスペクトルは、スペクトルアウトライヤーと呼ばれます。

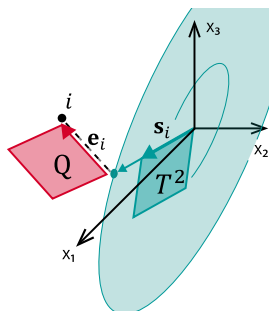
アウトライヤーを注意深く確認する必要があります。例えば、汚染されたサンプルや測定エラーが原因である場合、アウトライヤーはモデルを歪ませるおそれがあります。この場合、モデルを計算するときにアウトライヤーを考慮しないでください。

一方、アウトライヤーは他のスペクトルでは十分にカバーされないプロパティを示す場合があります。この場合、アウトライヤーは実際にモデルを改善します。アウトライヤーが有効なサンプルであると思われる場合、校正サンプルが変動範囲にわたって均等に分布しているかどうかを確認してください。

アウトライヤーを検出するための適切な尺度は、ホテリングの T^2 と Q 残差です。

ホテリングの T^2 と Q 残差

分光データを主成分空間に変換する場合、スペクトルはスコアと残差により特徴づけることができます (33 ページ, 4.2 章を参照)。同様のことが、潜在変数の空間への変換にも当てはまります (56 ページ, 4.4.1 章を参照)。



例：3次元波長空間(x_1, x_2, x_3)は2次元空間(緑色)に変換されます。点 i はスペクトル i を示し、3次元空間から2次元空間に投影されます。そこから次のものが生成されます。

- マハラノビス距離を示す2次元空間内のスコアベクトル Σs_i またはその正規化されたスコアベクトル s_i 。
- 3次元空間内の残差 e_i 。

s_i および e_i から次の量を導き出すことができます (89 ページ, 6.4 章を参照)。

- **ホテリングの T^2** または単に T^2 は、2次マハラノビス距離の2乗、つまりモデル中心から主成分空間または潜在変数空間へのスペクトルの正射影までの正規化された距離の2乗です。スペクトルのすべてのスコアが平均値と等しい場合、 T^2 は0であり、スペクトルはモデルの中心点にあります。モデルは中心点付近で最もよく合います。中心点から離れると、モデルの適合が悪くなる可能性があります。 T^2 値は高いです。 T^2 値が高い場合には、極端なスペクトルを示します。例えば、化学成分の極端な組成を含むサンプルを示します。
- **Q 残差** は残差の2乗です。つまり、スペクトルから主成分空間または潜在変数の空間までの直交距離の2乗です。 Q 残差は、モデルでは説明できない変動を示します。 Q 残差が高いときには、例えば測定したサンプルが別の物質を含んでいる場合に、スペクトルがモデルに適合していない可能性があることを示しています。

スペクトルアウトライヤーの検出

スペクトルアウトライヤーの検出は、母集団から外れたスペクトルを識別します。

- **ホテリングの T^2 アウトライヤー**、別名レバレッジアウトライヤー (英語で Leverage outlier) : T^2 が高いとは、主成分空間 (PCA) または潜在変数空間へのスペクトル (PLS) の投影が、モデルの中心点から遠く離れているということを意味します。
- **Q 残差-アウトライヤー** : Q 残差が高いとは、スペクトルがモデルによって十分に記述されていないということを意味します。

図28は、さまざまなビューにおける複数のスペクトルを示します。

- **左影響プロット** : Q 残差は、モデルによって説明されない変動を説明しますが、ホテリングの T^2 はモデル自体内の変動を説明します。
破線は、指定された有意水準の**臨界値**または**限界値**を示します (90 ページ, 6.5 章を参照)。
有意水準が高いほど、限界値は低くなり、より多くの点が限界値の外側に位置する可能性があります。
- **右** : 例として潜在変数で 2 次元空間に変換された 3 つの変数 x_1 、 x_2 、 x_3 を持つ元の空間の例。
点 A から D について、平面までの直交距離 (破線) と潜在変数の空間内のモデル化された点 (緑色の点) が示されます。

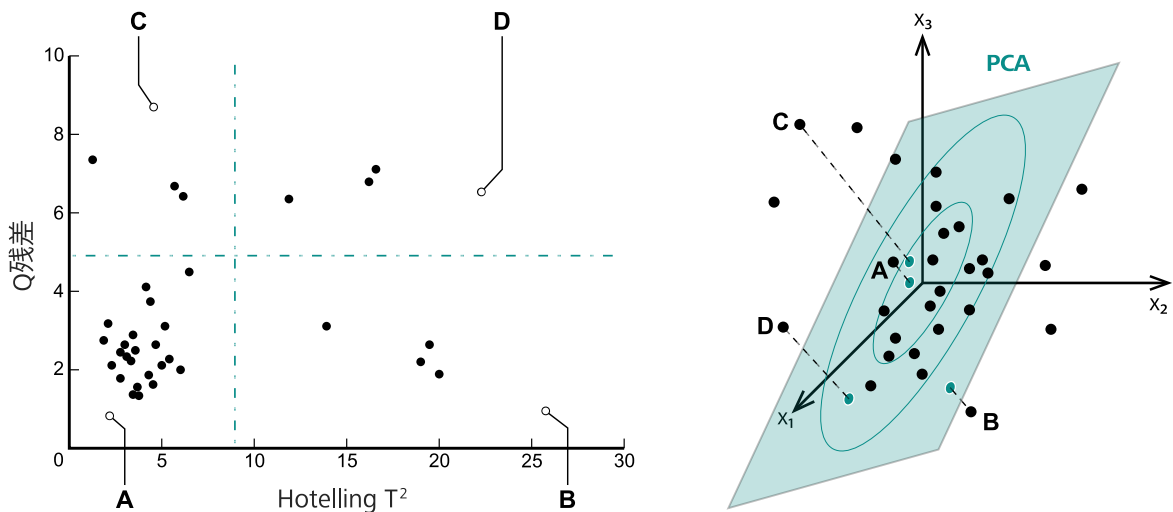


図 28 影響プロット (左)、元の空間および潜在変数の空間 (右)。各スペクトルは、左図の点と右図の点で表示されます。

両方のビューで、異なるプロパティを備える 4 個の点が強調表示されています。

- スペクトル A は低いスコアと低い残差を持っています。モデル中心点の近くにあり、モデルにより分かりやすく説明されます。
- スペクトル B はホテリングの T^2 アウトライヤーです。中心点から外れていますが、モデルによりよく説明されます。
- スペクトル C は Q 残差-アウトライヤーです。中心点から外れており、モデルによりよく説明されていません。



- スペクトル D はホテリングの T^2 アウトライヤーと Q 残差-アウトライヤーの両方です。中心点から外れており、モデルにより部分的に説明されます。

影響プロットは、さまざまなスペクトルがモデルにどのように影響を及ぼすかを示します。潜在変数のすべては中心点を通過するので、中心点付近のスペクトル（スペクトル A など）は、潜在変数の方向を変えるチャンスがほとんどありません。これにはレバレッジはありません。中心点からの距離が遠いほどレバレッジは大きくなり、モデルに影響を与える可能性が高くなります。実際にモデルをその方向に引くことができるスペクトルもある（スペクトル B）一方、ある程度しかできない（スペクトル D）、あるいは全くできない（スペクトル D）場合もあります。

すべてのスペクトルに基づくモデルと比較して、スペクトル B を含まないモデルの計算は、スペクトル D がいないものよりも大きく変更される可能性が高く、スペクトル C がいないものよりもさらに大きく変更される可能性があります。スペクトル B は、良くも悪くも定量化モデルに大きな影響を及ぼす可能性があります。影響プロットの右下象限にある潜在的アウトライヤーを削除するかどうかの決定には特別の注意が必要です。

理想的には、モデルは多数のスペクトルの分散を捉える必要があります。モデルが少数のスペクトルによってのみ特徴づけられるということは望ましいことではありません。上図では、複数のスペクトルが中心点から大きく離れており、他のほとんどのスペクトルからも大きく離れています。これは疑わしいです。モデルはいくつかのスペクトルの影響を受けます。これは確認する必要がある潜在的アウトライヤーです。また、サンプルが変動範囲全体にわたって均一に分割されていることを確認する必要があります。

PCA および PLS 影響プロット

PCA 影響プロットはスペクトルにのみ依存します。PLS 影響プロットは、スペクトルと基準値に依存します。

次の表は、それぞれの設定が PCA 影響プロットおよび PLS 影響プロットにどのように作用するかを示しています。

	PCA 影響プロット	PLS 影響プロット
スペクトル	基礎となる PCA モデルは、校正セット、検証セットおよびアウトライヤーデータセットのすべてのスペクトルに基づいています。	基礎となる PLS モデルは、校正セットのスペクトルに基づいています。 この PLS モデルに基づいて、3 つすべてのデータセットのスペクトルの T^2 および Q 残差値が計算され、プロットで表示されます。

	PCA 影響プロット	PLS 影響プロット
パラメータ設定	<p>選択したデータ前処理と波長範囲を考慮します。</p> <p>注記: アウトライヤー検出は PCA に基づいており、ユーザー設定と OMNIS Software のバージョン (47 ページ, 「スペクトルアウトライヤーの検出」を参照) に応じたパラメータ化を考慮します。</p>	<p>選択したデータ前処理と波長範囲を考慮します。</p> <p>注記: 予測におけるアウトライヤーの評価は PLS に基づいており、データ前処理と波長範囲を考慮します。</p>
変数の数	<p>少なくとも 95% の説明された分散に達する主成分の数を使用します。</p>	<p>実際に選択された潜在変数の数を使用します。</p>
有意水準と臨界値	<p>臨界値の計算と可視化のために現在選択された有意水準を使用します (破線)。</p> <p>識別: 最後に実施されたデータセット分割がアウトライヤー検出なしで行われた場合、影響プロットは有意水準 5% を使用します。</p> <p>注記: 有意水準を高くすることにより、臨界値は低くなり、それによってモデル開発時にアウトライヤーは増加します。</p>	<p>臨界値の計算と可視化のために現在選択された有意水準を使用します (破線)。</p> <p>注記: 有意水準を高くすることにより、臨界値は低くなり、それによって予測時にアウトライヤーは増加します。</p>
基準値 (定量化)	<p>基準値は PCA モデルに影響を及ぼしません。</p> <p>ただし、各スペクトルには関連する基準値アウトライヤーがある可能性があり、アウトライヤーとしてマークされる可能性があります。</p>	<p>基準値は PLS モデルに影響を及ぼすため、PLS 影響プロットにも影響を及ぼします。</p> <p>さらに、各スペクトルには関連する基準値アウトライヤーがある可能性があり、アウトライヤーとしてマークされる可能性があります。</p>

アウトライヤーの分析

潜在的アウトライヤーを分析する場合、次の要因を考慮する必要があります。

- ホテリングの F^2 アウトライヤーは、他のサンプルと比べて化学成分の組成が極端であるサンプルを示します。
- 例えば、Q 残差-アウトライヤーは、サンプルが汚染されていたり、スペクトル記録時のエラーを示している可能性があります。

潜在的なアウトライヤーは、慎重に確認する必要があります。真のアウトライヤーはスペクトルのリストから削除する必要があります。有効なサンプルは保持する必要があります。データセットが再

度分割されると、アウト라이어検出により、最初の実行では見つからなかった潜在的アウト라이어が見つかる場合があります。考えられる理由は、新しい PCA モデルでは 95 % の説明される分散を実現するために必要な主成分が少なくなっているためです。新しく見つかったアウト라이어を有効なサンプルであることが判明した場合には、それを保持する必要があります。この場合、アウト라이어検出なしで自動分割を繰り返すことができます。

4.3.3.2 スコアプロット

スコアプロットの基礎となるのは、PCA モデルまたは PLS モデルです：

- **定量化:** スコアプロット (OMNIS Software のバージョン 3.0 以降) は、**PLS** (56 ページ, 4.4.1 章を参照) を基にしています。
- **識別:** スコアプロット (OMNIS Software のバージョン 4.3 以降) は、**PCA** (33 ページ, 4.2 章を参照) を基にしています。

各スペクトルには、主成分または潜在変数ごとにスコア値があります。スコアプロットでは、各スペクトルが点で表されます。x 軸は例えば 1 番目の潜在変数のスコアを示し、y 軸は例えば 2 番目の潜在変数のスコアを示します。同様に、潜在変数の各ペアを表示できます。

各波長変数の吸光度値は平均センタリングされているので、各潜在変数のスコアも平均センタリングされています。スコアプロットを中心点の近くにある点(0/0)は、表示されている 2 つの潜在変数に関連する平均スペクトルを示します。示された 2 つの潜在変数に関して、互いに近い点は類似のスペクトルを示し、互いに離れている点は異なるスペクトルを示します。

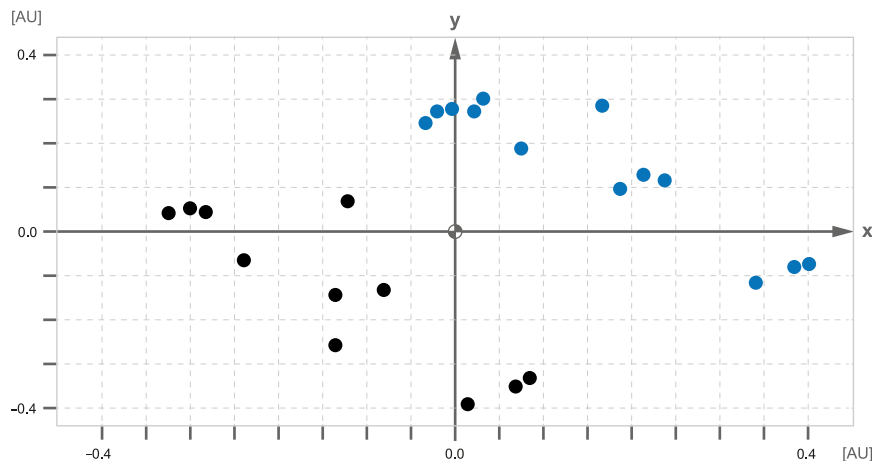


図 29 潜在変数1 (x 軸) および潜在変数2 (y 軸) のスコアプロット。AU = 任意の単位。

図 29 は、異なる条件で測定された 2 つのデータセットの例を示します。スコアは正規化され、各潜在変数は同じ重み付けを受け取ります。

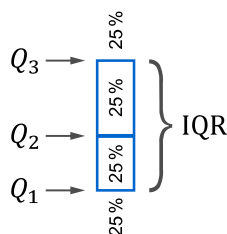
i スペクトルのすべての主成分または潜在変数のスコアをシングル値 (ホテリングの T^2) に組み合わせることができ、これが影響プロットの x 軸に表示されます。

4.3.4 基準値アウトライヤー (定量化)

定量化モデルの場合には、スペクトルアウトライヤーに加えて基準値アウトライヤーも決定されます。基準値アウトライヤーは基準値で異常を示します。

一般的に、基準値アウトライヤーは誤って伝送された数値で、14.3 の代わりに 143、51.9 の代わりに 15.9 などです。アウトライヤー検出は、経験的なアプローチに基づいてそのような伝送エラーや転写エラーを特定します。明らかなエラーのみが、さらに調べるためにマークされます。

ボックスプロット



基準値アウトライヤーは、ボックスプロットに基づくメソッドを利用して決定されます。**ボックスプロット**は基準値を昇順に並び替えます。四分位数はデータセットを4つの部分に分割します。各部分は基準値の25%を含みます。

最初の四分位数 Q_1 は25%の最低値を残りから分離します。 Q_2 は中央値であり、50%の最低値をのこりから残りから分離します。3つ目の四分位数 Q_3 は75%の最低値を残りから分離します。縦の長方形はデータの中央の50%、四分位数範囲 (IQR、英語で interquartile range) を示します。



IQR ボックスから外れる一定量のデータは潜在的なアウトライヤーとみなされ、小さな縁で示すことができます。アウトライヤーの上限と下限は多くの場合、IQR の1.5倍で定義されます。

$$[Q_1 - 1.5 \text{ IQR}; Q_3 + 1.5 \text{ IQR}]$$

ここで、 Q_1 は第1四分位数、 Q_3 は第3四分位数、IQR は四分位数範囲 ($Q_3 - Q_1$) です。

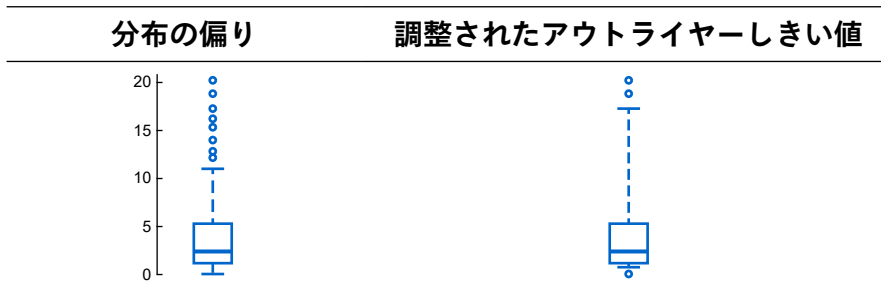
ボックスプロットを完全にするには、ボックスの上下のひげが、潜在的なアウトライヤーとしてマークされていない最も遠い点まで到達します。

偏った分布の調整

通常ボックスプロットは、データがほぼ対称的に分布していることを前提とします。分布が偏っている場合には通常、多くの通常の基準値を潜在的なアウトライヤーとしてマークされます。これにより、OMNIS Software は分布の偏りを考慮した調整されたバージョンのボックスプロットを使用します。

次の例では、分布は高い基準値に向かってシフトされます。それにより、高い値を持つ8個のアウトライヤーが得られます。アウトラ

イヤーしきい値を調整すると、そのうちの2個のみが残ります。一方、低い値を持つ新しいアウトライヤーが表示されます。



調整に関する説明は補遺をご覧ください (92 ページ, 6.6 章を参照)。

4.3.5 データセット分割

データセットは、スペクトルと基準値 (定量化) またはスペクトルとそれに関連する製品名 (識別、検証) で構成されます。データセットは、モデルの開発と検証のために使用されます。そのため、データセットを校正セット、検証セットおよびアウトライヤーデータセットに分割する必要があります。分割は手動または自動で行うことができます。

十分なスペクトルが利用できると仮定して、例えば全データの 20-30% を検証セットとして使用できます。

自動分配アルゴリズム

自動データセット分割により、校正セットと検証セットが母集団を示し、互いに独立していることを保証します。その目標は、PCA 空間のほぼ同じ範囲をカバーし、類似の統計的プロパティを有する 2 つのセットにデータを分割することです。使用されるアルゴリズムは、R. D. Snee, *Validation of Regression Models: Methods and Examples*, Technometrics Band 19, No. 4 (Nov. 1977), P 415-428 の Duplex アルゴリズムを若干修正したものです。

i 複製

データセットには複製があってはなりません。このアルゴリズムは、複製や疑複製を削除したり、特定のサンプルの複製を同じデータセットに追加することを強制することはありません。

1. パラメータ設定は次のように考慮されます:
 - a. OMNIS Software のバージョン 4.2 以降: パラメータ設定 (データ前処理と波長選択) を適用するかをユーザーが決定します。
パラメータ設定に後から加えた変更は、データセット分割に影響を与えません。
 - b. OMNIS Software のバージョン 3.3 ~ OMNIS Software のバージョン 4.1 まで: データ前処理を考慮するかをユーザーが決定します。波長選択、およびデータ前処理に後から加えた変更は、データセット分割に影響を与えません。
 - c. OMNIS Software のバージョン 3.2 以前: データ前処理は、アウトライヤー検出時に設定されたとおりに考慮されます。波長選択、およびデータ前処理に後から加えた変更は、データセット分割に影響を与えません。
2. この分割は、すべての平均値中心スペクトルの PCA モデルに基づいています。主成分の数は、説明された分散が少なくとも 95 % となるように選択されます。
3. PCA スコアから、考えられるすべてのスペクトルペア間の距離が計算されます。
4. 互いに最も離れたスペクトルの 2 つが校正セットに割り当てられます。
5. 残ったスペクトルのうち、互いに最も離れた 2 つのスペクトルが検証セットに割り当てられます。
6. 残ったスペクトルのうち、既に校正セットに含まれているスペクトルから最も離れているスペクトルが、校正セットに割り当てられます。
7. 残ったスペクトルのうち、既に検証セットに含まれているスペクトルから最も離れているスペクトルが、検証セットに割り当てられます。
8. この交換は、データセットの一つが意図したサイズに達するまで継続されます。残りのスペクトルは別のデータセットに割り当てられます。

4.4 定量化

4.4.1 PLS 回帰

i OMNIS Software は、定量化モデルを計算するために PLS 回帰を使用します。

PCA 同様に、**部分最小二乗回帰 (PLS 回帰、英語で *partial least squares regression*)** は分光データをより少ない変数に減らします。ただし：

- PCA の主成分は、PLS では**潜在変数**と呼ばれています。通所、潜在変数と主成分の方向は同様ですが、同じではありません。
- PLS 回帰は、スペクトルに加えて**基準値**も考慮します。そのため、基準値との十分な相関を実現するために必要な潜在変数は少なくなり、結果としてノイズが少なくなります。

PLS は、広範で冗長性の高い分光データから、基礎となる**潜在変数**を抽出します。潜在変数は直接測定されないため、**隠れ変数**とも呼ばれます。潜在変数は、データ変動を可能な限り説明し、同時に基準値を適切にモデル化します。

準備手順

PLS 回帰の準備手順は PCA の場合同様です。

1. **パラメータ設定**：OMNIS Software は、指定されたデータ前処理および指定された波長選択をスペクトルに適用します
2. **平均センタリング**：各波長に対して、平均吸光度値が計算され、各スペクトルの各波長から差し引きます。基準値も平均センタリングされます。

潜在変数への変換

準備手順の後、PLS 回帰はスペクトルデータを、基準値を考慮して潜在変数空間に変換します。図 30 は最初の 2 つの潜在変数 LV1 および LV2 を示します。同じ方法で、潜在変数 LV3、LV4 などを示すことができます。

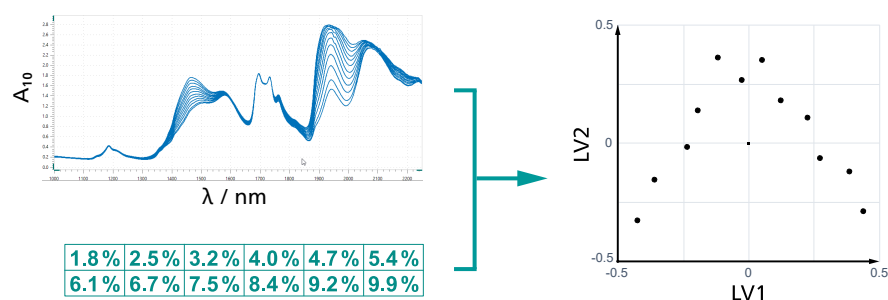
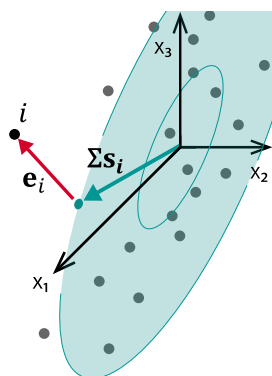


図 30 スペクトルと基準値の潜在変数空間への変換。右ページのスコアは任意の単位で示されます。

最初の潜在変数 LV1 は、スペクトルデータの分散を最も良く説明し、基準値との相関も最も高くなります。続くすべての潜在変数 LV2、LV3 などは残りの分散を最も良く説明し、基準値との相関も最も高くなります。そのため、最初の潜在変数は大部分の分散を説明し、相関は最大化される一方、その他の潜在変数には主にノイズが含まれるため、無視できます。

スコアと残差



PLS は PCA 同様のサイズを有しています。

- **スコア**：スコアは潜在変数空間で測定されます。サンプル i を潜在変数の各方向に直交投影すると、中心点からのユークリッド距離を示すスコアベクトル Σs_i を得られます。
マハラノビス距離 s_i は、各方向に同じ重み付けを与える正規化されたスコアベクトルです。
- **残差**：残差ベクトル e_i は、サンプル i と元の波長空間において測定された潜在変数空間との間のオフセットです。

PLS アルゴリズム

PLS アルゴリズムは、スペクトルと基準値との間の共分散を最大化します (88 ページ, 6.3 章を参照)。

4.4.1.1 潜在変数の数

定量化モデルの潜在変数の数の選択は、モデルの予測能力の基礎となります。潜在変数の数が少なすぎると、関連のスペクトルの変数を得られません。これは**過少適合**と呼ばれ、予測精度が低くなります。

潜在変数の数が多すぎると、校正サンプルのモデル化が良すぎます。モデルは無関係のスペクトル変動（ノイズ）も補足します。これは**過剰適合**と呼ばれ、未知のサンプルの予測が変動し、不安定で、精度が低くなります。

潜在変数の最適な数を見つけるために、次の目標間のバランスを実現する必要があります。

- SEP は最小値に十分近づいています。
検証セットなしで SECV を使用します。
- 定量化モデルでは、できる限り少ない潜在変数を使用してください。疑わしい場合には、少ない方の数を使用してください。
- 検証セットの相関プロットは最適値に近づいています。理想的には、スロープが 1 に近く、Y インターセプトが 0 に近く、データポイントの広がりが最小限であることが必要です。
検証セットなしで、交差検証の値を使用します (58 ページ, 「交差検証」を参照)。

交差検証では、次のいずれかの方法で複数ラウンドの手順を使用します。

- **一個抜き (Leave-One-Out)**

各ラウンドでは、一個抜き交差検証(LOO-CV)により 1 つのサンプルがリセットされ、残りのサンプルからモデルが作成されます。次に、このモデルはリセットされたサンプルの対象パラメータを予測します。この予測はサンプルの推定値として機能します。このサイクルは、各サンプルが一度リセットされるまで続きます。

- **k-fold**

交差検証方法 k-fold では、校正セットができる限り同じサイズの k ブロックに分割されます。

各ラウンドでは、1 つのブロックがリセットされ、残りのブロックからモデルが作成されます。次に、このモデルはリセットされたモデルの対象パラメータを予測します。

このサイクルは、各ブロックが一度リセットされるまで続きます。

ブロックの選択は、さまざまな方法で行うことができます。

- **Fixed Blocks (DUPLEX)** (OMNIS Software のバージョン 3.2 以降) : ブロックは Duplex アルゴリズムに基づいて再現可能に選択されます。各予測は、各サンプルに対する直接推定値として機能します。
- **random** : ブロックはランダムに選択されます。上記に説明された手順を複数回繰り返します。校正セットは毎回異なる方法で k ブロックに分割されます。最後に各サンプルに対して複数の推定値があります。その平均値は、各サンプルに対する推定値として機能します。

一般的に一個抜き (Leave-One-Out) が好ましいです。校正セットが大きい場合、交差検証方法 k-fold を使用し、計算時間を節約できます。 k の典型的な値は 5 です。

4.4.2.1 相関プロット

相関プロットは、基準値と算出値との間の相関を表示します。これにより、定量化モデルの評価を一目で確認できます。

算出値は次のように決定できます。

- 検証セットおよびアウトライヤーデータセット : 定量化モデルによる予測
- 校正セット : 交差検証の推定値

相関プロットでは、各サンプルは点により示されます。サンプルの算出値は x 軸、基準値は y 軸に示されます。回帰直線は、変数間のシステム関係を示します。回帰直線は、1 のスロープと 0 の y インターセプトを有し、すべての点が直線上にあることが理想的です。これは、各サンプルでは算出値は基準値に対応することを意味します。



理想的な場合からの逸脱に基づき、システムエラーとランダムエラーを区別できます。回帰直線の位置はシステムエラーを示します。回帰直線からの点の距離はランダムエラーを示します。

次の相関プロット **A** は優れた相関を示します。別のプロットはさまざまな種類のエラーを示し、それについては以下で説明します。

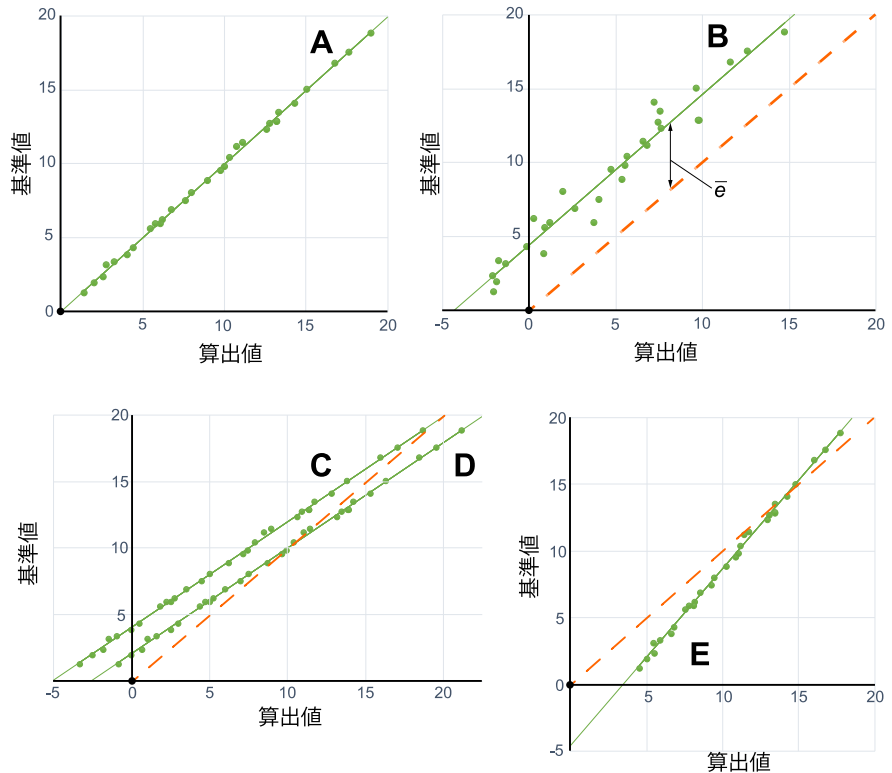


図 32 相関プロット。各点はサンプルを示し、その基準値および算出値を示します。破線は理想的な 45° 直線を示します。

システムエラー

システムエラーは、常に発生し、特定のアプリケーションで繰り返し発生するエラーです。システムエラーは修正できます。回帰直線のバイアス \bar{e} およびスロープ b により定量化できます。

$$y = b\hat{y} + \bar{e}$$

スロープが 1 でバイアスが 0 である場合、システムエラーはありません。

バイアスは、基準値と算出値との平均エラーです。

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) = \bar{y} - \bar{\hat{y}}$$

ただし、 n はサンプル数、 e_i は、 i 番目のサンプルエラー、 y_i は i 番目のサンプルの基準値、 \hat{y}_i は i 番目のサンプルの算出値、 \bar{y} は基準値の平均値、 $\bar{\hat{y}}$ は算出値の平均値です。

直線 **B** および **C** は正のバイアスを有し、直線 **E** は負のバイアスを有しています。

符号が反対のエラーは打ち消し合っています。そのため、直線 **D** のバイアスはほぼ0になります。

回帰直線のスロープは次のようになります。

$$b = \frac{s_{\hat{y}y}}{s_{\hat{y}}^2}$$

ただし、 $s_{\hat{y}y}$ は基準値と算出値との共分散、 $s_{\hat{y}}^2$ は算出値の分散です。

スロープはプロパティに依存するエラーとしてみなすことができます。

- $b > 1$ (直線 **E**) : 算出値が大きいほど、バイアスに寄与するエラーは大きくなります (正に)。
- $b < 1$ (直線 **C** および **D**) : 算出値が大きいほど、バイアスに寄与するエラーは小さくなります (負に)。
- $b = 1$ (直線 **A** および **B**) : バイアスに寄与するエラーは一定です。

y 軸と回帰直線の **y インターセプト**は $\bar{y} - b\bar{\hat{y}}$ です。

i スロープと y インターセプトは、基準値を従属変数 (y 軸)、算出値を独立変数 (x 軸) として計算されます。

ランダムエラー

すべての点が回帰直線上にある場合、ランダムエラーはありません。点が散在しているほど、ランダムエラーが大きくなります。

相関プロット **B** では、ランダムエラーは他の相関プロットより大きくなります。

エラーの種類を表示

上図の直線は、次のエラーの種類を示します。

直線	システムエラー			ランダムエラー
	バイアス	スロープ	y インターセプト	
A	~ 0	~ 1	~ 0	小
B	> 0	~ 1	> 0	大
C	> 0	< 1	> 0	小



直線	システムエラー			ランダムエラー
	バイアス	スロープ	y インターセプト	
D	~ 0	< 1	> 0	小
E	< 0	> 1	< 0	小

4.4.2.2 性能指数

性能指数は、基準値と予測値の一致を示します。予測値は定量化モデルにより決定されます。

R² - 決定係数

決定係数 R²（英語で coefficient of determination）は定量化モデルの適合度を測定します。特定のデータセットに対して、これは、定量化モデルにより説明される基準値の変動割合です。

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

ただし、SS_{reg} は回帰平方和（予測値の分散、つまり説明された分散）、SS_{tot} は平方和の合計（基準値分散）、SS_{res} は残差平方和（残差分散、つまり説明されていない分散）、y_i は i 番目のサンプルの基準値、ŷ_i は i 番目のサンプルの予測値、ȳ は基準値の平均値です。

R²値は 1 の分数です。R²が 1 というのは、予測値が基準値と完全に一致していることを意味します。R²が 0.9 というのは、90 % の基準値分散が予測値により説明され、10 % が説明されないということを示します。

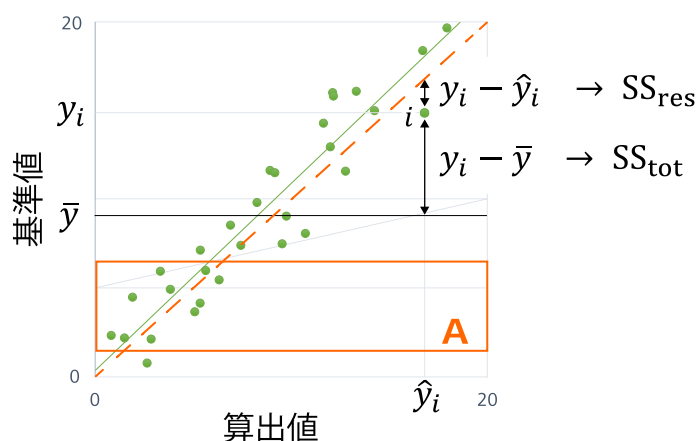


図 33 R²の計算のためのコンポーネント。破線は 45° 直線です。上図はサンプル i と PLS 回帰の残差の相関プロットを示します。残差変動は SS_{res} に含まれ、基準値変動は SS_{tot} に含まれます。

i R²値が大きくても、利用可能な定量化モデルや正確な予測が保証されるわけではありません。R²の大きさは、基準値変動に直接依存しています。

基準値範囲が小さい（範囲 A）の回帰では、残差変動はほぼ同じですが、基準値変動は小さくなります。結果として発生する R² 値は小さくなります。

R²が大きい理由は、基準値範囲が非現実的に大きいことが考えられます。一方で、例えば製造プロセスからのデータからの値の範囲が制限されている場合、その結果として R²値は小さくなります。予測能力を評価するために、標準エラーを使用する必要があります。

R²の絶対値は注意する必要があります。さらに意味があるのは、追加の潜在変数ごとの変化の程度です（57 ページ、4.4.1.1 章を参照）。

計算のために使用される値により、それぞれの R²が得られます。

- R²C（OMNIS Software では表示されません）：校正セットにおけるスペクトルの予測値で計算されます。
- R²CV：校正セットにおけるスペクトルの交差検証の予測値で計算されます（58 ページ、「交差検証」を参照）。
- R²P：検証セットにおけるスペクトルの予測値で計算されます。

OMNIS Software は、計算するためにピアソンサンプル相関係数の 2 乗を使用します $r_{y,\hat{y}}$ 。

交差検証の決定係数：

$$R^2CV = r_{y,\hat{y}_{cv}}^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_{cv_i} - \bar{\hat{y}}_{cv}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \cdot \sum_{i=1}^n (\hat{y}_{cv_i} - \bar{\hat{y}}_{cv})^2}$$

ただし、 y_i は i 番目のサンプルの基準値、 \bar{y} は基準値の平均値、 \hat{y}_{cv_i} は i 番目のサンプルの交差検証の推定値、 $\bar{\hat{y}}_{cv}$ は交差検証の推定値の平均値、 n は校正セットにおけるサンプル数です。校正セットの各サンプルには、正確に 1 つの交差検証の推定値があることに注意してください。

予測の決定係数：

$$R^2P = r_{v,\hat{v}}^2 = \frac{(\sum_{i=1}^v (v_i - \bar{v})(\hat{v}_i - \bar{\hat{v}}))^2}{\sum_{i=1}^v (v_i - \bar{v})^2 \cdot \sum_{i=1}^v (\hat{v}_i - \bar{\hat{v}})^2}$$

ただし、 v_i は i 番目の検証サンプルの基準値、 \bar{v} は基準値の平均値、 \hat{v}_i は i 番目の検証サンプルの予測値、 $\bar{\hat{v}}$ は予測値の平均値、 v は検証サンプル数です。

SEC - 校正の標準エラー

校正の標準エラー(SEC)は校正セットに基づいています。SECは、理論上の最高の予測精度の推定値とみなせます。SECは、部分最小二乗回帰(PLS)の残差の標準偏差です。

$$SEC = \sqrt{\frac{\mathbf{e}^t \mathbf{e}}{n - k - 1}}$$

ただし、 \mathbf{e} はモデルにより示されない校正セットにすべての基準値変動を含む残差ベクトル、 n は校正サンプル数、 k は潜在変数の数です。分母 $n-k-1$ は残差ベクトル \mathbf{e} の自由度の数です。

言い換えると、SECは校正セットのサンプルに対する基準値と予測値との間の差の標準偏差です。

$$SEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}}$$

ただし、 y_i は*i*番目の校正サンプルの基準値、 \hat{y}_i は*i*番目の校正サンプルの予測値、 n は校正サンプル数、 k は潜在変数の数です。

SECはRMSECと呼ばれる場合もあります。SECはランダムエラーとシステムエラー（スロープとバイアス）を含みます。

SECV - 交差検証の標準エラー

交差検証の標準エラー(SECV)は校正セットに基づいています。SECVは校正セットと交差検証方法に基づいて予測精度を推定します (58ページ、「交差検証」を参照)。SECVは初期モデル評価または潜在変数の最適数を決定するために使用できます。

SECVは校正セットのサンプルに対する基準値と交差検証の予測値との間の差の標準偏差です。

$$SECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{cv_i})^2}{n}}$$

ただし、 y_i は*i*番目のサンプルの基準値、 \hat{y}_{cv_i} は*i*番目のサンプルの交差検証の推定値、 n は校正セットにおけるサンプル数です。

i SECVは、すべてのエラー、つまりランダムエラーおよびシステムエラー（スロープとバイアス）を含みます。

別のアプローチでは、バイアス補正されたSECVと、RMSECVと呼ばれる補正されていないSECVに対して別の値を使用します。バイアスが低い場合、この値は類似しています。

SEP - 予測の標準エラー

予測の標準エラー(SEP)は検証セットに基づいています。そのため、SEPは予測精度の最も現実的な推定値を提供します。

SEPは検証セットのサンプルに対する基準値と予測値との間の差の標準偏差です。

$$SEP = \sqrt{\frac{\sum_{i=1}^v (v_i - \hat{v}_i)^2}{v}}$$

ただし、 v_i は*i*番目の検証セットの基準値、 \hat{v}_i は*i*番目の検証サンプルの予測値、 v は検証サンプルの数です。

i SEPは、すべてのエラー、つまりランダムエラーおよびシステムエラー（スロープとバイアス）を含みます。

別のアプローチでは、バイアス補正されたSEPと、RMSEPと呼ばれる補正されていないSEPに対して別の値を使用します。バイアスが低い場合、この値は類似しています。

性能指数の解釈

性能指数は推定値です。例：SEP値は利用可能なサンプルに基づく標準偏差の推定値であり、独自の標準偏差も有しています。検証サンプルの数が多いほど、推定値の信頼性が高まります。

SEPはSECVおよびSECと同等でなければなりません。差が大きすぎると、過剰適合を示す可能性があります（57 ページ、4.4.1.1 章を参照）常識的な規則として、その差は20%を超えてはなりません。

標準エラーは、基準メソッドの実験室標準エラー(SEL)と比較して考慮する必要もあります。NIRメソッドの精度は、SEPがSELよりも1.4~2.0倍高い場合に許容できます。要件を満たしていれば、より大きなSEPも引き続き受け入れられます。

基準メソッドのSELを下回るSECまたはSECVは過剰適合を示します。

SELとの上記関係では、SELが各サンプルの繰り返し基準測定の正確な回数に基づいていることを前提としています（32 ページ、「基準メソッド（定量化）」を参照）。

4.4.3 OMNIS Model Developer (OMD)

定量化モデルの開発は困難で時間がかかり、一定レベルの専門知識を必要とします。OMNIS Model Developer (OMD、OMNIS Softwareのバージョン4.0以降)は開発を自動化し、最適化された定量化モデルを提供します。



機能

適切なサンプリングは前提条件です (30 ページ, 4.1 章を参照)。スペクトルと、それに対応する基準値からなるデータセットは、OMD の入力として機能します。

OMD は、データ前処理を考慮することなく、5 %の有意水準と補遺に記載されているアルゴリズムでスペクトルアウトライヤーを検出します (90 ページ, 「モデル開発時のスペクトルアウトライヤーの検出」を参照)。

データセットの分割は、アウトライヤー検出後に残るスペクトルの数に依存します。

スペクトルの数	交差検証方法	検証セット
> 99	k-fold (5 ブロック、Duplex)	25%のスペクトル
30~99	k-fold (5 ブロック、Duplex)	—
< 30	一個抜き (Leave-One-Out)	—

分割後、ASTM D8321-22 に準拠した校正セット内の追加のアウトライヤーが決定されます。

モデルは、さまざまなキー数値に基づいて評価および分類が行われます。OMD は過剰適合のリスクと過少適合のリスクとの間のバランスを目指して、データ前処理、波長選択、潜在変数の数を最適化します。

結果

OMD の結果は、予測力の順に並べられたモデルのリストです。予測力は、モデルの複雑さ、性能指数、データセットのサイズに基づいて計算されます。

リストは色分けされており、好ましいモデルの選択を容易にします。

- 緑色：優れた予測力。
サンプル数が十分に大きい場合、モデルは同じタイプの未知のすべてのサンプルについて適切に機能します。性能指数により、将来のエラーの信頼性の高い推定を可能にします。
- 黄色：中程度の予測力。
サンプル数が十分に大きい場合、モデルは適切に機能することが期待されます。性能指数は将来のサンプルに対して楽観的過ぎる可能性があります。個別の検証をお勧めします。
- 赤色：不十分な予測力。
モデルには重大な欠陥があります。使用しないでください。

同じ色のモデルは、低い推定予測エラーと少数の潜在変数との間のバランスを優先する情報基準に従って分類されます。

パラメータ設定の最適化

モデル全体を自動的に作成する代わりに、パラメータ設定の最適化のみ行うこともできます。現在の設定（データセット分割、交差検証方法など）は変わらずそのままですが、最適化には影響を与えません。

4.4.4 スロープ/γ インターセプト補正

スロープ/γ インターセプト補正により、定量化モデルを使用する場合に、システムエラー（バイアス、スロープ）を修正できます。

校正セット内のシステムエラーの考えられる原因は次のとおりです。

- 定量化モデルにおけるシステムエラー。既知のアウトライヤーまたは不十分なサンプル数など。
- 分光測定方法におけるシステムエラー。
- 基準測定方法におけるシステムエラー。

検証セットのシステムエラーが発生した場合、他の原因が考えられます。

- 分光測定方法の変更、例えば装置。
- 基準測定方法の変更、例えば新しい実験室、新しい装備。
- 取り扱い、保管、輸送などにおけるサンプルの変更。

バイアス補正と、さらにはスロープ/γ インターセプト補正は注意して適用する必要があります。

システムエラーが有意でない場合、補正を適用しないでください。エラーが有意である場合、徹底的に調べる必要があります。可能であればエラーの原因を取り除く必要があります。正当な理由でエラーを取り除けない場合、バイアス補正またはスロープ/γ インターセプト補正を適用できます。

信頼できるバイアス推定値を得るためには、20 以上のサンプルが必要です。スロープの信頼できる推定値を得るには、30 以上のサンプルが必要です。

バイアス補正

次の相関プロットはバイアス補正を示します。元の回帰直線 **F** のスロープは変化しないままです。補正後（回帰直線 **G**）、正のエラート負のエラーは相殺されます。

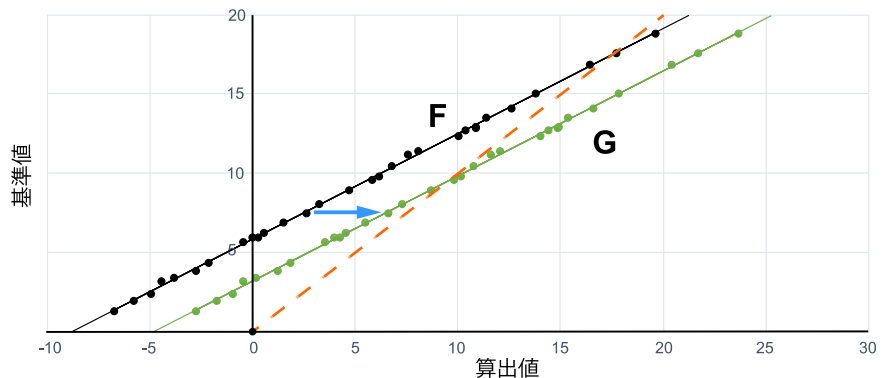


図 34 バイアス補正

スロープ/y インターセプト補正

次の相関プロットはスロープ/y インターセプト補正を示します。元の回帰直線 **H** は、スロープおよび y インターセプトについて補正されます。その結果、スロープとバイアスの両方が補正されます（回帰直線 **K**）。

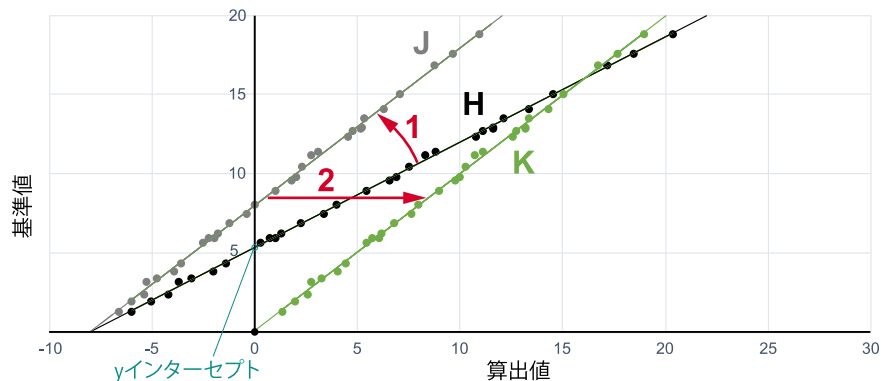


図 35 スロープ/y インターセプト補正

SEP

補正データセットのサンプルがスロープ/y インターセプト補正の基礎となります。これらのサンプルを基礎として OMNIS Software は以下の予測の標準エラー (SEP) を計算します。分母はそれぞれ対応する自由度を考慮します:

補正のタイプ	SEP
未補正	$SEP = \sqrt{\frac{\sum_{i=1}^n (v_i - \hat{v}_i)^2}{n}}$
バイアス補正	$SEP = \sqrt{\frac{\sum_{i=1}^n (v_i - \hat{v}_i)^2}{n-1}}$
スロープ/γ インターセプト補正	$SEP = \sqrt{\frac{\sum_{i=1}^n (v_i - \hat{v}_i)^2}{n-2}}$

この時、 v_i は補正データセットの i サンプルの基準値、 \hat{v}_i は補正データの i サンプルの予想値、 n は補正データのサンプル数に相当します。

i SEP は、すべてのエラー、つまりランダムエラーおよびシステムエラー（スロープとバイアス）を含みます。

4.5 識別と検証

4.5.1 サポート Vector マシン(SVM)

識別モデル（OMNIS Software のバージョン 4.0 以降）は、それぞれの製品間の分類にサポート Vector マシン(SVM)を使用します。サポート Vector マシンはモニタリングされた機械学習アルゴリズムです。マシンは、校正サンプルに基づいて新しいサンプルを製品に割り当てるために学習します。

i 便宜上、2 つの製品間の分類を以下に説明します。このコンセプトは拡張可能で、最終モデルは任意の数の製品間で分類されます。

線形分類

図36 (左) は 2 つの変数を有する入力データを示します。

i 便宜上、パラメータ設定されたスペクトルは、2 次元変数空間で表示されます。各点はスペクトルを示し、色は製品属性を表示します。

製品は線形に分離可能です。SVM アルゴリズムは、製品間に超平面を作成します（右図）。

i 超平面は、3 次元空間の平面を任意の次元の空間に一般化したものです。超平面の次元は、その周囲の空間の次元よりも一つ小さくなります。2 次元空間の超平面は直線になります。

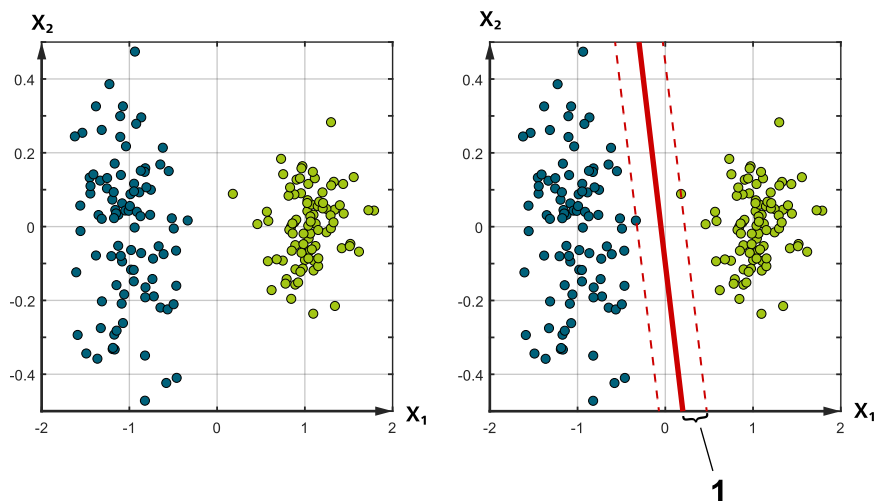


図 36 入力データ (左) およびSVM から作成された超平面 (右の赤い実線)。その値は任意単位で与えられます。

SVM アルゴリズムは、超平面と各辺の最も近い点との間のマージン(1)を最大化します。新しいスペクトルを同じ空間にマッピングし、超平面のどちら側に該当するかに応じて製品に割り当てることができます。

超平面を定義するために、SVM アルゴリズムは反対側の製品の点に最も近い点のみを考慮します。この点またはベクトルは、超平面の形成をサポートし、サポートベクトルと呼ばれます。

アウトライヤーなどによって点が線形に分離できない場合でも、線形分類する超平面を決定できます。この場合、最適化アルゴリズムは超平面から各側へのサポートベクトルまでの間隔を増やすことと、すべての点が超平面の正しい側にあることの保証との間の妥協点を見出します。正規化パラメータは妥協点を制御し、それによって超平面の最終的なポジションを制御します。

非線形分類

図37 (左) では、製品は非線形に分離可能です。製品を分離するには非線形分類が必要です。

線形または非線形のカーネル関数は、データを高次元の特徴空間に変換します。変換は、特徴空間内のデータが超平面により線形に分離できるように実行されます。

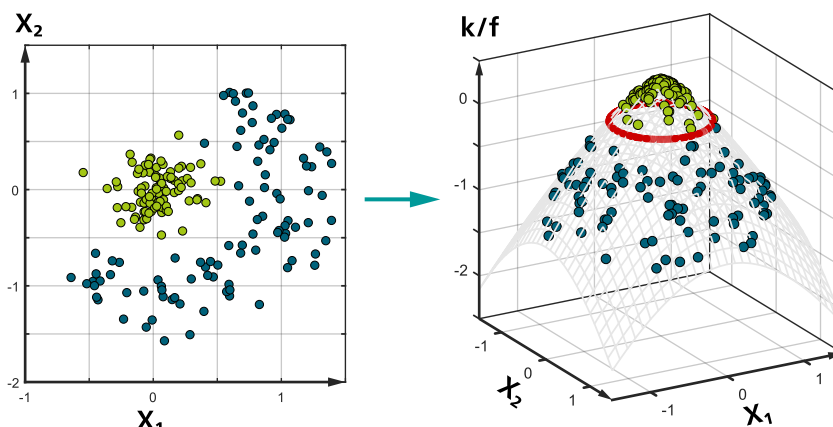


図 37 非線形分離可能な製品 (左)。追加の次元 k/f (カーネル特徴) により、分離が容易になります (右)。その値は任意単位で与えられます。

図では、データを 2 次元空間から 3 次元空間に変換しています。一方の製品の点は、元の平面よりも引き上げられる一方、別の製品の点は下に引き下げられます。製品間の線形超平面は、元の平面に非常に類似している 2 次元平面です。2 次元平面でみると、判定限界は非線形の線であり、この場合には赤い円の線です。

ここでも超平面は分類子として機能します。新しいサンプルは、そのスペクトルが超平面のどちらの側に該当するかに応じて製品に割り当てることができます。

OMNIS Software は、半径方向の基底関数を備えたカーネルを使用し、データを特徴空間に変換します。カーネルは、非線形性の程度を制御するスケールパラメータを使用します。

パラメータ選択

超平面のポジションを制御する正則化パラメータと、非線形性を制御するスケールパラメータに対して適切な値を選択する必要があります。

2 つのパラメータはサポート Vector マシンの汎用能力に影響します。つまり、校正スペクトルから新しく未知のスペクトルにどの程度うまく一般化できるかに影響します。例えば、スケールパラメータが高度な非線形性を可能にする場合、超平面は校正スペクトルに過剰に適合する可能性があります (過大調整)。例えば、スケールパラメータがわずかな非線形性しか実現しない場合、超平面は校正スペクトルに十分に適合しない可能性があります (過少適合)。

適切に一般化するには **グリッド検索** を使用します。このアルゴリズムは、可能な限り少ない試行で最適なパラメータの組み合わせを見つけます:

1. 事前定義されたパラメータの組み合わせのセットを使用します。

2. 選択された一部のパラメータの組み合わせについて、SVM はさまざまな製品の校正スペクトルを最大限の精度で区別する分類規則を学習します。
3. 各分類規則について、交差検証に従って製品割り当てがどのように成功するかを推定します。
4. 交差検証の結果に基づき、さらなるパラメータの組み合わせが選択されます。
SVM は該当する分類規則を再学習し、交差検証方法は分類精度を推定します。
複数回繰り返した後、最終的なパラメータの組み合わせが特定されます。
5. 最終的なパラメータの組み合わせおよびすべての検証スペクトルを用いて、SVM は最終的な分類規則を評価します。

確率

識別モデルは、分類規則を基に、特定のサンプルが特定の製品に属しているかどうかの確率を計算します。確率は、製品の母集団までの距離とモデルパラメータに依存します。

こうして計算された確率で、サンプルの製品への割り当てをコントロールできます (72 ページ, 「[サンプルの指定 \(OMNIS Software のバージョン 4.4 以降\)](#)」を参照)。

誤った正の識別を最小限に抑えるため、OMNIS Software のバージョン 4.4 以降ではさらに、各製品に対する適格性確認モデルのトレーニングが行われます (75 ページ, 4.6 章を参照)。

4.5.2 サンプルの製品属性の予測

特定のサンプル識別のため、識別モデルは製品ごとに確率を用意しています (72 ページ, 「[確率](#)」を参照)。

i 各確率は、0% から 100% の個別の値です。製品の値の合計は 100% になりません。
値は相対的なため、異なる製品を比較することができます。

サンプルの指定 (OMNIS Software のバージョン 4.4 以降)

評価は、調整可能な**確率しきい値**と個々の製品の適格性評価を使用して行われます:

1. 確率が確率しきい値を超える各製品に対し、対応する適格性確認モデルによるサンプルの適格性評価が実行されます。適格性評価に失敗した場合は、対応する製品の確率がゼロに設定されます。

2. 手順 1 の、修正された確率による評価:
- a. 確率が確率しきい値より上にない場合、識別は失敗します (識別ステータス**未識別**)。
 - b. 確率が 1 つだけ確率しきい値より上にある場合、サンプルは正常に識別され、対応する製品に割り当てられます (識別ステータス**識別済**)。
 - c. 複数の確率が確率しきい値より上にある場合、予測は多義的であり、識別は失敗します (識別ステータス**多義的**)。

テーブル 1 は、異なる確率しきい値を有する評価の例を示しています。この例では、製品 A と C の適格性評価はどちらも成功しています。

テーブル 1 異なる確率しきい値の評価例 (OMNIS Software のバージョン 4.4)

確率	確率しきい値	適格性評価	識別結果
製品 A: 87%	90%	–	未識別
製品 B: 71%	80%	製品 A: 成功 → 87%	製品 A
製品 C: 68%	70%	製品 A: 成功 → 87%	製品 A
製品 D: 30%		製品 B: 失敗 → 0%	
	60%	製品 A: 成功 → 87%	多義的
		製品 B: 失敗 → 0%	
		製品 C: 成功 → 68%	

サンプルの指定 (OMNIS Software のバージョン 4.0 ~ 4.3)

サンプルの確率は、調整可能な**確率しきい値**を使って評価されています:

- 確率が確率しきい値より上にない場合、識別は失敗します (識別ステータス**未識別**)。
- 確率が 1 つだけ確率しきい値より上にある場合、サンプルは正常に識別され、対応する製品に割り当てられます (識別ステータス**識別済**)。
- 複数の確率が確率しきい値より上にある場合、予測は多義的であり、識別は失敗します (識別ステータス**多義的**)。

テーブル 2 は、異なる確率しきい値を有する評価の例を示しています。

テーブル 2 異なる確率しきい値の評価例 (OMNIS Software のバージョン 4.0 ~ 4.3)

確率	確率しきい値	識別結果
製品 A: 87%	90%	未識別
製品 B: 72%	80%	製品 A
製品 C: 68%	70%	多義的

4.5.3 識別モデルの検証

識別モデルは通常、次のように検証されます。

1. 限られたサンプル数から開始して、検証セットなしでモデルは開発され、検証セット内のサンプルを用いてテストされます。
2. 十分なサンプル数がある場合、データセットを校正セットおよび検証セットに分割できます。検証セットのサンプルは、モデルの開発に使用されません。
3. 最後に、外部検証セットのサンプルは別の日に、場合によっては別の人により、別の装置で収集・測定されます。

各サンプルに関して、予測された製品属性は、各実際の製品属性と比較されます。一致する場合、予測は正しい（成功）、そうでない場合は間違っている（失敗）です。

検証

OMNIS Software は、モデルがどの程度機能するかを示す識別モデルの次の量を示します。理想的には、すべての指数は 100% です。

成功率% (合計) は精度、つまりモデルの全体的な正確性を測定します。このパーセンテージは、「モデルはサンプルのうち何個を正しく識別できますか？」という質問に答えます。

$$\text{成功率\% (合計)} = \frac{\text{正しい分類}}{\text{すべての分類}}$$

i 成功率% (合計) は各サンプルに等しい重みを与えます。そのため、より多くのサンプルの製品は、より少ないサンプルの製品よりもパーセントに大きな影響を及ぼします。

各製品に対しても、同様の成功率%の数値を得られます。

識別の改善

次のアクションはモデルの改善に役立ちます。

- 確率しきい値の調整**
 - 多数の予測が多義的であったり、多数の 0.0% 確率が生じる場合、確率しきい値を上げることができます。
 - 確率しきい値に達していないために多数のサンプルが識別されない場合、確率しきい値を下げるすることができます。
- パラメータ設定の調整**

適切なデータ前処理および波長選択を決定します。

■ モデル階層の使用

モデル階層により、識別モデルの階層構造化を可能にします。
例：4つの異なる製品を含む識別モデルでは、類似の製品であるフルクトースとグルコースを簡単に区別できません。フルクトースとグルコースを製品グループ「糖」にまとめると、モデルは糖と2つの別の製品に区別することができます。サンプルを糖として識別すると、別のモデルはフルクトースとグルコースとの間の区別を処理します。このモデルは専門的であるため、類似した製品をより簡単に区別できます。

識別されないサンプル

サンプルが誤って識別されない場合:

- サンプルとサンプル処理に異常がないか確認してください。
- 確率しきい値を確認し、必要に応じて下げます。
- データ前処理と波長選択を確認してください。
- 識別されないサンプルのスペクトルをスコアプロットで確認します:
 - スペクトルがまだモデルに含まれていない場合: 各製品の検証セットにスペクトルを追加します。
 - スコアプロットでは、校正セットのスペクトルのスコアと確認するスペクトルのスコアを比較します。

サンプルがアウトライヤーではないが、校正セットでそのバリエーションが過少提示されている場合、それに応じて校正セットを拡充する必要があります。

4.6 適格性評価

適格性確認モデル (OMNIS Software のバージョン 4.4 以降) はサンプルのグループと他のサンプルを区別します。このモデルは、例えば、使用できるサンプル (正のサンプル) と使用できないサンプル (負のサンプル) を区別する場合などに適しています。

4.6.1 適格性確認モデルの計算

適格性確認モデルの計算は、識別モデルの場合と同じように行われます (69 ページ, 4.5.1 章を参照)。ただし、適格性評価のための校正セットには 1 種類のサンプル (正のサンプル) のみが含まれます。

サポート Vector マシン (SVM) は、入力データを高次元空間に変形します。正則化パラメータは超平面のポジションを定め、スケーリングパラメータは非線形性の程度を決定します。グリッドサーチは適切な射影を算出します。この射影の判定限界は、適格性確認モデルの基礎を形成します。

4.6.2 適格性確認モデルの検証

手順

適格性確認モデルは、通常、段階的に開発・検証されます。この時、サンプル数は徐々に引き上げられます：

1. 正の検証セット：
 - a. 正のサンプルが制限されている場合、正の検証セットは当面の間作成されません。
 - b. 十分な数の正のサンプルが使用可能になると、自動データセット分割を使用して正の検証セットが作成されます。

負の検証セット：

- a. 収集された負のサンプルは、負の検証セットに割り当てられます。
 - b. 負のスペクトルの自動算出を使用して、スペクトルのアウトライヤーを検出し、負の検証セットに割り当てることができます (90 ページ, 6.5 章を参照)。
2. その後、正と負の検証セットのサンプルは別の日に、場合によっては別の人により、別の装置で収集・測定されます。

i 正と負の検証セットのサンプルは、モデルの計算には使用されません。

検証

各サンプルに対し、適格性確認モデルは結果を算出します (正または負)。校正セットおよび正の検証セットのサンプルに対しては、正の結果が予測されます。負の検証セットのサンプルに対しては負の結果が予測されます。結果が予測と一致している場合、予測は正しく (= 成功)、そうでない場合は誤り (= 失敗) です。

OMNIS Software は、適格性確認モデルに対して以下の値を表示します：

成功率% (合計) はモデルの全体的な正確性を測定します。

$$\text{成功率\% (合計)} = \frac{\text{正確な予測}}{\text{すべての予測}}$$

各データセットに対しても、同様の成功率% の数値を使用できます。すべての指数が 100% であれば理想的です。

適格性評価の改善

パラメータ設定 (データ前処理と波長選択) をより適切に行うことで、適格性確認モデルを改善できます。

認証されないサンプル

サンプルが誤って認証されない場合：

- サンプルとサンプル処理に異常がないか確認してください。
- データ前処理と波長選択を確認してください。
- 認証されないサンプルのスペクトルをスコアプロットで確認します:
 - スペクトルがまだモデルに含まれていない場合: 正の検証セットにスペクトルを追加します。
 - スコアプロットでは、校正セットのスペクトルのスコアと確認するスペクトルのスコアを比較します。

サンプルがアウトライヤーではないが、校正セットでそのバリエーションが過少提示されている場合、それに応じて校正セットを拡充する必要があります。

5 予測

5.1 定量化

未知の対象パラメータを含むサンプルを予測する場合の手順は次のとおりです。

1. サンプルのスペクトルを記録します。
2. 定量化モデルは、校正セットのスペクトルと同じデータ前処理と波長範囲を適用します。
3. 結果のスペクトルに従って、モデルは対象パラメータを予測します。

i 定量化モデルを計算する場合、校正スペクトルおよび基準値は平均中心となりました。これはもちろん、上記の手順で考慮されません。

5.1.1 アウト라이어と結果のモニタリング

定量対象パラメータの予測では、様々な種類のアウト라이어を検出し、結果をモニタリングすることができます：

アウト라이어/モニタリング		原因 (例)
スペクトルアウト라이어	ホテリングの T^2	化学コンポーネントの濃度が校正サンプルの濃度範囲外にあります。
	Q 残差	校正サンプルには存在しない構成成分がサンプルに含まれています。
Nearest Neighbor アウト라이어 (オプション、OMNIS Software のバージョン 4.2 以降)		サンプルが、校正サンプルのこの組み合わせには存在しないサンプルバリエーションを示しています。
結果のモニタリング (オプション)		対象パラメータの結果値が、ユーザーが選択した承認基準の範囲外にあります。

スペクトルアウト라이어

スペクトルアウト라이어は次のように算出されます：

1. ソフトウェアは、定量化モデル (PLS モデル) に基づいて、スペクトルのホテリングの T^2 および Q 残差の値を計算します。

2. スペクトルの T^2 または Q 残差値が、モデルが算出した、対応する限界値よりも大きい場合、サンプルは適用モデルに関してアウトライヤーとしてマーキングされます (91 ページ, 「予測の場合のアウトライヤー評価 (定量化)」を参照)。

i T^2 および Q 残差値は、OMNIS Software で変数として使用できません。これらは、定量化モデルの PLS 影響プロットの値と比較することができます。破線は限界値を示します。

Nearest Neighbor アウトライヤー

(OMNIS Software のバージョン 4.2 以降)

校正サンプルが、考えられるサンプルバリエーションのすべての組み合わせを網羅しているのが理想的です。実際には、一部の組み合わせは頻繁に出現し、その他はまったく現れません。そのため、潜在変数の領域では校正サンプルは不均一に分散しています。一部の範囲には多くの校正サンプルがありますが、その間には隙間があります。

不明なサンプルのスペクトルが校正サンプルの間の隙間に落ちると、予測結果は無効、または不正確になることがあります。このような事例を検出するため、不明なサンプル i からそれぞれの校正サンプル u までの距離 D を算出します:

$$D = \sqrt{(\mathbf{s}_i - \mathbf{s}_u)^t (\mathbf{s}_i - \mathbf{s}_u)}$$

ここで、 \mathbf{s}_i は不明なサンプル i のスコアに、 \mathbf{s}_u は校正サンプル u のスコアに相当します。これらのスコアは正規化されており、直交しています。

最小の距離は最も近い校正サンプルまでの距離で、**Nearest Neighbor Distance** (NND) と呼ばれています:

NND 値が特定の NND 限界値を超えると、その不明なサンプルは Nearest Neighbor アウトライヤーと呼ばれます。

NND 制限値は以下の手順で測定されます:

1. それぞれの校正サンプルに対して NND 値が測定されます。この値は、残りの校正サンプルのうち最も近いものとの距離に相当します。
2. すべての校正サンプルの最大 NND 値が NND 限界値です。

OMNIS Software では、不明なサンプルおよび NND 限界値の NND 値を変数として使用できます。

結果のモニタリング

OMNIS Software では、結果のモニタリングを使用し、予測結果の範囲に警告範囲や制限範囲を定義することができます。オプションで、定義した限界を超えた時に作動する措置を定義することができます。

5.2 識別と検証

サンプルの識別または検証の手順は次の通りに行います:

1. サンプルスペクトルを記録します。
2. 識別モデルは、校正セットのスペクトルと同じデータ前処理と波長選択を適用します。
3. 識別モデルはスペクトルを評価します (72 ページ, 4.5.2 章を参照)。
4. 識別モデルがモデル階層の一部であり、他の識別モデルが特定された製品とリンクされている場合、そのモデルが実行されます。一つまたは複数の定量化モデルが特定された製品とリンクされている場合、それらのモデルが実行されます。
5. 識別結果または検証結果は、モデル階層で、場合によっては定量化結果でも表示されます。

識別ステータス

- 識別済
識別は成功しました。
- 多義的
複数の製品が確率しきい値を超えています。識別は失敗しました。
- 未識別
製品は確率しきい値を超えていません。識別は失敗しました。

検証ステータス

- 成功
サンプルは正常に識別され、結果は予測された製品と一致しています。
- 失敗
検証は失敗しました。

5.3 適格性評価

サンプルの適格性評価を行う場合の手順は次の通りです:

1. サンプルスペクトルを記録します。
2. 適格性確認モデルは、校正セットのスペクトルと同じデータ前処理と波長選択を適用します。
3. 結果のスペクトルを基に、モデルはサンプルを認証します。
4. 適格性確認結果が表示されます。

適格性評価ステータス

- 成功
- 失敗

6 補遺

6.1 線形回帰の例

単変量線形回帰

最もシンプルな場合では、混合物には1つの吸収体のみが含まれており、スペクトルは1つのピークのみを持っています。吸収体の濃度が異なるサンプルには、それぞれの吸光度値のピークがあります(6 ページ, 2.2.1 章を参照)。

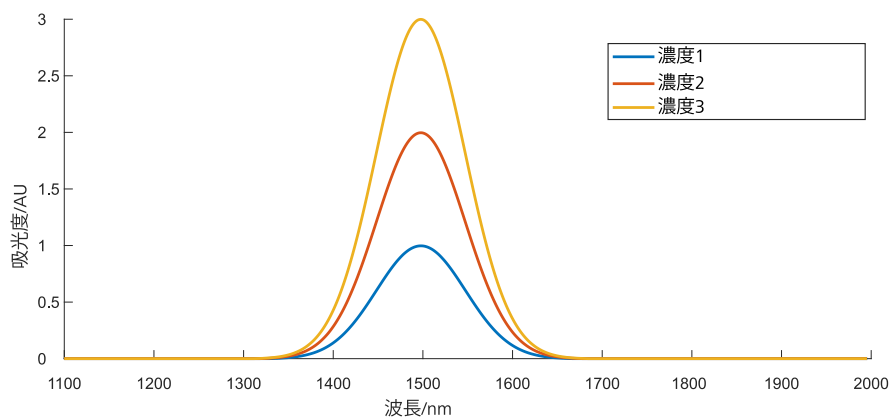


図 38 1,500 nm にピークを持つサンプルのモデル化されたデータ。

1,500 nm で測定された3つの吸光度値を、吸収体の濃度に対してプロットできます。

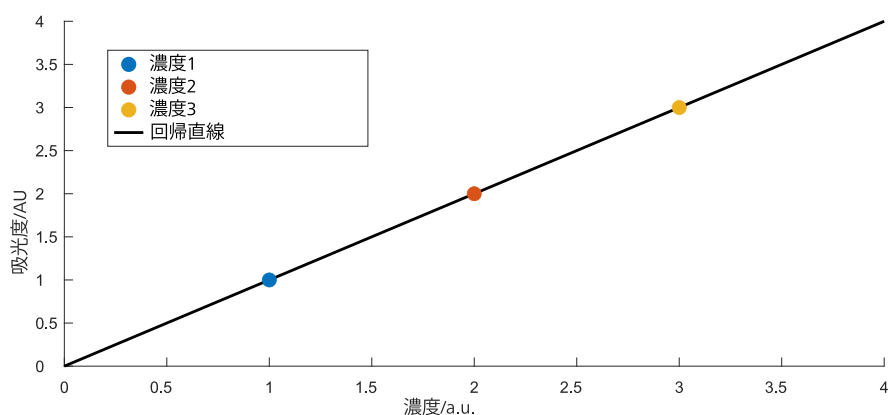


図 39 吸光度と濃度との関係。

Beer-Lambert の法則に従うと、吸光度値と濃度との間の関係は線形です。3つのサンプルのセットにより線形回帰を実行すると、次のような回帰直線を得られます。

1つの変数（1つの波長）しかないため、これは単変量線形回帰です。回帰は定量化モデルとして使用できます。未知の濃度のサンプルの場合、吸光度 A は1,500 nmで測定されます。回帰直線から、対応する吸収体の濃度 c を得られます。

$$c = bA$$

係数 b は一定であり、回帰直線のスロープと同一です。

すべてのサンプルには、同じモル吸光係数を持つ同じ吸収体が含まれている必要があるということに注意してください。さらに、すべての吸収測定は同一の層厚で実行しなければなりません。

多変量線形回帰

実際の混合物には1つ以上の吸収体が含まれています。記録されたスペクトルは、すべての吸収体スペクトルの合計です（6ページ、2.2.1章を参照）。

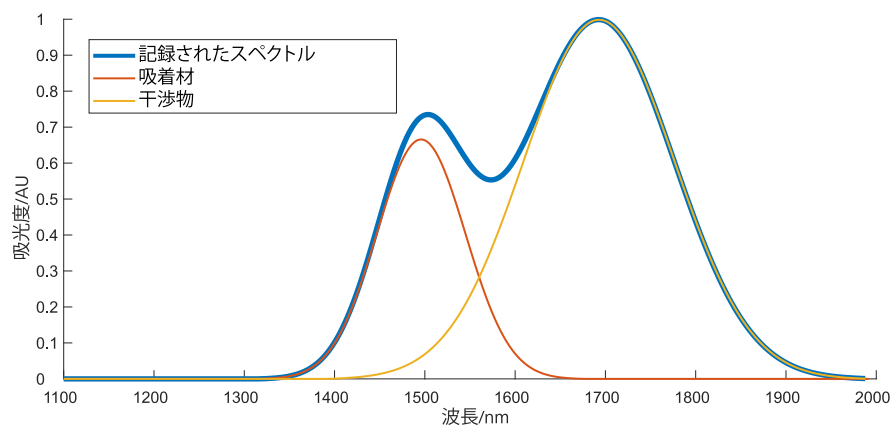


図 40 2つのコンポーネントを含むモデル化されたデータ。吸収体（赤の線）は定量化されなければなりません。

記録されたスペクトル（青の線）は、純粋な吸収体スペクトルと、重複する干渉スペクトルの合計です。1,500 nmでは、測定された吸光度値は、吸収体の吸光度だけでなく、干渉体の吸光度も含んでいます。対象パラメータは、1,500 nmでの1回の測定では定量化できません。また、干渉体が存在したかどうか、また測定が信頼できるかどうかを知ることもできません。

2つの波長を例えば1,500 nmおよび1,700 nmで測定するとどうなるでしょうか？波長1で測定された吸光度、 A_1 は、純粋な吸収体の信号 A_1^a （インデックス a =吸収体）および純粋な干渉体の信号 A_1^f （インデックス f =干渉体）の合計です。同じことを、波長2で測定された吸光度、 A_2 にも適用します。

$$A_1 = A_1^a + A_1^f = \varepsilon_1^a c_a + \varepsilon_1^f c_f$$

$$A_2 = A_2^a + A_2^f = \varepsilon_2^a c_a + \varepsilon_2^f c_f$$

ただし、 ε_1^a および ε_1^f は吸収体または干渉体に対する波長 1 の場合のモル吸光係数、 c_a および c_f は吸収体または干渉体の濃度です。

上記方程式では、層厚 l は Beer-Lambert の法則から除外されています。これにより、後の代数が少し簡単になります。当然のことながら、すべてのサンプルについて層厚は同一でなければなりません。そのため、方程式中の吸光度は 1cm あたりの吸光度です。

方程式は行列で次のように示すことができます。

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} \varepsilon_1^a & \varepsilon_1^f \\ \varepsilon_2^a & \varepsilon_2^f \end{bmatrix} \begin{bmatrix} c_a \\ c_f \end{bmatrix}$$

したがって：

$$\begin{bmatrix} c_a \\ c_f \end{bmatrix} = \begin{bmatrix} \varepsilon_1^a & \varepsilon_1^f \\ \varepsilon_2^a & \varepsilon_2^f \end{bmatrix}^{-1} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$$

吸収体の濃度に対する解を得られます。

$$c = c_a = \frac{\varepsilon_2^f}{\varepsilon_1^a \varepsilon_2^f - \varepsilon_1^f \varepsilon_2^a} A_1 + \frac{-\varepsilon_1^f}{\varepsilon_1^a \varepsilon_2^f - \varepsilon_1^f \varepsilon_2^a} A_2$$

これにより、干渉物質が存在する場合であっても、2つの波長で吸光度を測定し、各吸光度に定数を掛けることにより、吸収体の濃度を計算できます。

その定数はモル吸光係数を指し、表で調べることができます。ただし、現実では決してそのようなことは起こりません。その代わりに、校正ステップと、PLS 回帰などの多変量線形回帰による線形方程式を解くことにより決定されます。そのため、定数は回帰係数 b_1 および b_2 と呼ばれます。

$$c = b_1 A_1 + b_2 A_2$$

2 個以上の吸収体

上記に示されたように、吸収体が 1 つの場合、1 つの波長での吸光度を測定するだけで十分です。吸収体が 2 つの場合、2 つの波長での吸光度を測定するだけで十分です。

これは一般化できます。吸収体がより多くある場合、それぞれの波長 i でより多くの吸光度値 A_i を必要とします。これは依然として線形関係です。

$$c = b_1 A_1 + b_2 A_2 + \dots + b_n A_n$$

定量化モデルの開発

上記方程式が未知のサンプルの濃度を予測する前に、係数 b_1 、 b_2 などを決定しなければなりません。そのためには校正手順が必要です。対象パラメータの濃度が異なる複数のサンプルを測定します。

後で PCA および PLS に対して使用される用語に従って、 c は y に、 A は x に置換できます。各校正サンプルに対して書き出された上記方程式は次のようになります。

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,f} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,f} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,f} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

ただし、 n はサンプル数、 f は波長の数、 y_1 は基準メソッド（滴定など）で測定されたサンプル 1 の基準値、 $x_{1,1}$ は波長 1 のサンプル 1 の測定された吸光度、 $\{x_{1,1} \dots x_{1,f}\}$ は f の波長で測定されたサンプル 1 のスペクトルです。 $b_1 \dots b_n$ は回帰係数、 $e_1 \dots e_n$ は回帰係数が測定データをどの程度モデル化したかを示すエラー項目です。

よりコンパクトな行列形式では次のようになります。

$$\mathbf{y} = \mathbf{X}^t \mathbf{p} + \mathbf{e}$$

\mathbf{X} は行列 $f \times n$ として定義されます。 \mathbf{X}^t は転送行列 \mathbf{X} つまり、行と列が交換され、上記 $n \times f$ 行列が得られます。予測ベクトル \mathbf{p} は上記回帰係数 \mathbf{b} です。

予測ベクトル \mathbf{p} により、新しいサンプルに対して、そのスペクトル \mathbf{x} を基にしてたいしょ対象パラメータの予測を実現します。予測値 \hat{y} は次の次のとおりです。

$$\hat{y} = \mathbf{x}^t \mathbf{p}$$

多変量線形回帰の機能は、最小のエラー項目を生成する回帰係数を決定することです。ただし、重線形回帰(MLR)では、波長の数よりも多くの校正サンプル数が必要になります。もう一つの障害は、変数間の高い相関です。

他のメソッドも分光予測に使用できます。PCA によりデータ量を大幅に削減し、相関を完全に排除できます。PLS 回帰により、さらにサンプルの基準値を考慮します。

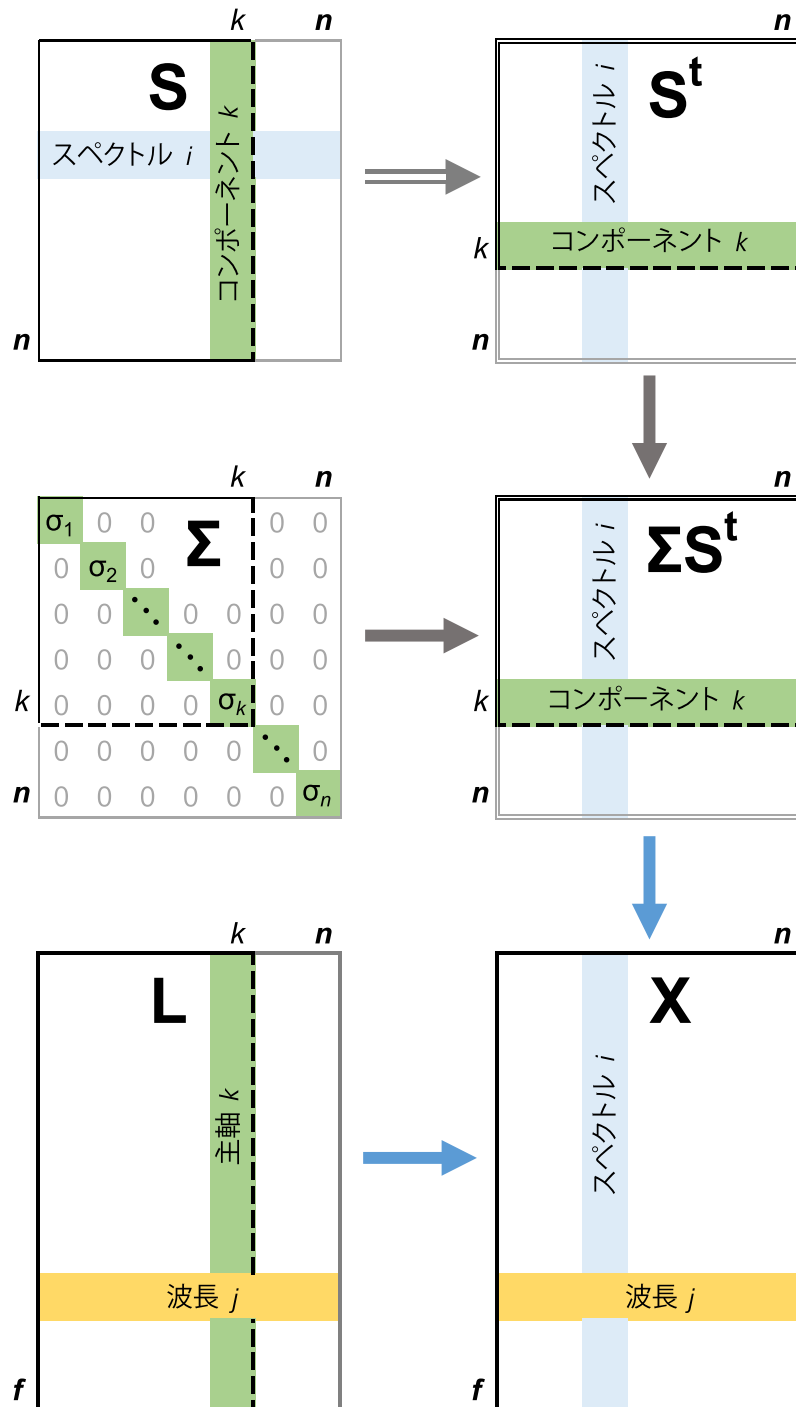


図 41 グラフ形式の特異値分解の式。破線は、 k の主成分を含むモデルに対する短縮行列を示しています。 $n-k$ 個の省略された主成分の情報は、残差行列に流れます (行列 $f \times n$ 、図示されない)。

残差行列

PCA モデルは計算された n 個の主成分のうちの最初のいくつかのみを使用します。図41では n 個の主成分のうち最初の k 個です。元のデータ \mathbf{X} は、モデルにより記述されたデータと、モデルにより記述されていないデータに分けることができます。

$$\mathbf{X} = \mathbf{L}_a \Sigma_a \mathbf{S}_a^t + \mathbf{E}$$

ただし、 \mathbf{X} は元のスペクトルデータ ($f \times n$ 行列)、 \mathbf{L}_a ($f \times k$ 行列) は \mathbf{L} のうちの k 列、 Σ_a ($k \times k$ 対角行列) は k の特異値、 \mathbf{S}_a ($n \times k$ 行列、 k の主成分を含む) は \mathbf{S} のうちの k 列、 \mathbf{E} は残差行列 ($f \times n$) であって、すべてのスペクトル変動が \mathbf{X} に含まれ、モデルにより記述できないものです。

通常、 $k \ll n \ll f$ です。例： $k=3$ 主成分、 $n=100$ サンプルおよび $f=2,500$ 波長。

残差行列 \mathbf{E} の各列 \mathbf{e}_i は、スペクトル i の PCA 空間、つまり残差に対する対角距離を示します。モデルが多くの主成分を使用するほど、残差は小さくなります。

6.3 PLS アルゴリズム

部分最小二乗回帰 (PLS 回帰、英語で *partial least squares regression*) は、定量化モデルの計算に使用されます (56 ページ、4.4.1 章を参照)。

PLS は次の 2 つのデータブロックを含みます。

- パラメータ設定され、平均中心の行列 \mathbf{X} (スペクトル)。
- 平均中心 \mathbf{y} ベクトル (基準値)。

PLS は行列 \mathbf{X} を 2 つの行列に分解します。

$$\mathbf{X} = \mathbf{L}\mathbf{S}^t + \mathbf{Z}$$

ただし、 \mathbf{X} ($f \times n$ 行列、 f の波長と n のサンプルを含む) は前処理された平均中心のスペクトル、 \mathbf{L} はローディング ($f \times k$ 行列、 k の潜在変数を含む)、 \mathbf{S} はスコア ($n \times k$ 行列)、 \mathbf{Z} は残余行列 ($f \times n$ 行列) であって、すべてのスペクトル変動は \mathbf{X} に含まれ、モデルにより記述できないものです。

PCA ではスコア行列 \mathbf{S} が \mathbf{X} の分散を説明しますが、PLS ではスコア行列 \mathbf{S} が \mathbf{X} と \mathbf{y} との間の共分散を説明します。PLS はスコアにより説明される共分散を最大化します。これは、スコアが \mathbf{X} の分散を最もよく説明するだけでなく、基準値と限りなく高い相関を有することを意味します。

\mathbf{X} と \mathbf{y} との間の共分散を最大化するため、PLS アルゴリズムは \mathbf{X} と \mathbf{y} との間でデータを交換します。したがって、 \mathbf{X} と \mathbf{y} はた単一の統合

されたシステムに統合されます。このとき、スコア \mathbf{S} を基準値 \mathbf{y} に対して回帰し、回帰係数 \mathbf{b} を得ます。

$$\mathbf{y} = \mathbf{S}\mathbf{b} + \mathbf{e}$$

ただし、 \mathbf{e} は、モデルによって説明できない \mathbf{y} のすべての基準値変動が含まれている残差ベクトルです。

予測

回帰係数 \mathbf{b} から予測ベクトル \mathbf{p} を決定します。新しいサンプルの対象パラメータ \hat{y} の予測には、予測ベクトル \mathbf{p} および前処理されて平均中心のスペクトル \mathbf{x} が使用されます。

$$\hat{y} = \mathbf{x}^t \mathbf{p}$$

i OMNIS Software は、SIMPLS アルゴリズムと基準値の単一セット (PLS-1) により PLS を実装します。

6.4 ホテリングの T^2 と Q 残差

ホテリングの T^2 と Q 残差は、PCA または PLS モデルにおけるスペクトルを特徴づけます。特に、考えられるアウトライヤーの識別に役立ちます (47 ページ、「ホテリングの T^2 と Q 残差」を参照)。

ホテリングの T^2

マハラノビス距離は、モデルの中心点からスペクトルの偏差の大きさの尺度です。距離は正規化されます。各主成分または潜在変数には同じ重みが含まれています。

スペクトルまたはスコアが正規分布していると仮定すると、マハラノビス距離の 2 乗、 MD^2 はホテリングの T^2 分布に従います。

$$MD^2 \sim T^2$$

スペクトル i のマハラノビス距離の 2 乗は、最初の k 個の主成分または潜在変数の正規化されたスコアの二乗和です。

$$MD_i^2 = \mathbf{s}_i \mathbf{s}_i^t = \sum_{a=1}^k s_{i,a}^2$$

ただし、 \mathbf{s}_i は短縮化されたスコア行列 \mathbf{S} の i 番目の行、 $s_{i,a}$ スペクトル i および主成分 (または潜在変数) a の正規化されたスコア、 k は使用する主成分または潜在変数の数です。

MD^2 は s シンプルに T^2 と呼ぶことができます。 $T^2 = 0$ を含むスペクトルは平均中心モデルの中心点に投影され、つまりスコアは中心にあります。超平面へのスペクトルの投影が中心点から離れるほど、 T^2 値は高くなります。モデルは中心点付近で最もよく合います。中心点から離れると、モデルの適合が悪くなる可能性があります。

アウトライヤーを検出する仮説検定には例えば 5 %の**有意水準**を設定できます (90 ページ, 6.5 章を参照)。

Q 残差

スペクトル i の Q 残差は、PCA または PLS モデルからスペクトルまでの二乗距離です。

$$Q_i = \mathbf{e}_i^t \mathbf{e}_i = \sum_{j=1}^f e_{i,j}^2$$

ただし、 \mathbf{e}_i は残差行列 \mathbf{E} の i 番目の列、 e_{ij} はスペクトル i の残差、 f は波長の数です。

Q 残差は、モデルでは説明できない変動を示します。Q 残差が高いときには、例えば測定したサンプルが別の物質を含んでいる場合に、スペクトルがモデルと一致していない可能性があることを示しています。

アウトライヤーを検出する仮説検定には例えば 5 %の**有意水準**を設定できます (90 ページ, 6.5 章を参照)。

6.5 スペクトルアウトライヤー - アルゴリズム

スペクトルアウトライヤーの検出は、母集団から外れたスペクトルを特定します (47 ページ, 「[スペクトルアウトライヤーの検出](#)」を参照)。アルゴリズムは、テストされたスペクトルのホテリングの T^2 または Q 残差の値がランダムな変動またはシステムバリエーションの結果であるかを評価します。

モデル開発時のスペクトルアウトライヤーの検出

1. パラメータ設定は次のように考慮されます:
 - a. OMNIS Software のバージョン 4.2 以降: パラメータ設定 (データ前処理と波長選択) を適用するかをユーザーが決定します。
パラメータ設定に後から加えた変更は、データセット分割に影響を与えません。
 - b. OMNIS Software のバージョン 3.3 ~ OMNIS Software のバージョン 4.1 まで: データ前処理を考慮するかをユーザーが決定します。波長選択、およびデータ前処理に後から加えた変更は、データセット分割に影響を与えません。
 - c. OMNIS Software のバージョン 3.2 以前: データ前処理は、アウトライヤー検出時に設定されたとおりに考慮されません。波長選択、およびデータ前処理に後から加えた変更は、データセット分割に影響を与えません。

2. スペクトルアウトライヤーの検出は、スペクトルのリスト中のすべての平均値中心スペクトルの PCA モデルに基づいています (33 ページ, 4.2 章を参照)。テストするスペクトルも同様に PCA モデルで記録されます。主成分の数は、説明された分散が少なくとも 95 % となるように選択されます。

3. **ホテリングの T^2 アウトライヤー :**

ホテリングの T^2 値 (89 ページ, 「ホテリングの T^2 」を参照) に基づき、仮説検定は H. Hotellings, *The Generalization of Student's Ratio*, *The Annals of Mathematical Statistics* Band 2, Nr. 3 (Aug. 1931), S. 360–378 に従って実行されます。

- a. 帰無仮説は、調査対象のスペクトルの T^2 値が PCA モデルの T^2 値と一致することです。帰無仮説が真の場合、 T^2 はホテリングの T^2 分布に従います。分布はスケーリングされた F 分布として示されます。

$$T^2 \sim \frac{k(n-1)}{n-k} F_{k,n-k}$$

ここで、 k は主成分の数、 n はスペクトルの数、 $F_{k,n-k}$ パラメータ k および $n-k$ を含む F 分布です。

- b. 設定可能な有意水準は、帰無仮説が真である場合に、それを棄却する確率を制御します (タイプ I 過誤と呼ばれます)。標準値は 5 % です。
- c. この分布と有意水準に基づいて、 T^2 の限界値が計算されます。
- d. スペクトルの T^2 値が限界値よりも大きい場合、帰無仮説は棄却され、スペクトルは潜在的なアウトライヤーとしてマークされます。

4. **Q 残差-アウトライヤー**

Q 残差 (90 ページ, 「Q 残差」を参照) に基づき、仮説検定は、J. E. Jackson and G. S. Mudholkar, *Control Procedures for Residuals Associated With Principal Component Analysis*, *Technometrics* Band 21, Nr. 3 (Aug. 1979), S. 341–349 に従って実行されます。

- a. 帰無仮説は、調査対象のスペクトルの Q 残差が PCA モデルの Q 残差と一致することです。帰無仮説が真の場合、ほぼ正規分布した Q 残差テスト平均化を作成できます。
- b. 設定可能な有意水準は、帰無仮説が真である場合に、それを棄却する確率を制御します (タイプ I 過誤と呼ばれます)。標準値は 5 % です。
- c. このテスト平均化と有意水準に基づいて、Q の限界値が計算されます。
- d. スペクトルの Q 残差値が限界値よりも大きい場合、帰無仮説は棄却され、スペクトルは潜在的なアウトライヤーとしてマークされます。

予測の場合のアウトライヤー評価 (定量化)

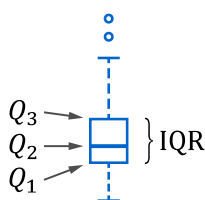
定量化の場合、予測中にアウトライヤー評価を利用できます。

1. 準備手順：
 - a. 定量化モデルは上記同様のアルゴリズムを使用します。ただし、基本は、校正セットのすべての平均中心スペクトルを含む定量化モデル (PLS モデル) です。データ前処理、波長範囲、潜在変数の数は、定量化モデルで指定されたとおりに考慮されます。
 - b. 指定された有意水準に基づき、定量化モデルは T^2 および Q の限界値を計算します (PLS 影響プロットの破線)。限界値は定量化モデルに保存されます。
2. 予測時に、スペクトルの T^2 および Q 値は定量化モデルにより計算されます。
3. スペクトルの T^2 または Q 値が各限界値よりも大きい場合、サンプルは使用する定量化モデルに関して潜在的アウト라이어とみなされます。

6.6 基準値アウト라이어 - アルゴリズム

ボックスプロットにより、基準値のアウト라이어を検出できます (53 ページ, 4.3.4 章を参照)。

分布の歪みを考慮するために、アウト라이어しきい値は次の計算で調整されます。



Medcouple (MC) は基準値の歪みを測定します。計算は、ボックスプロットの中央値、 Q_2 から始めます。関数は、基準値の上半分と下半分から、すべての可能なペア (英語で *couples w*) を使って計算されます。結果の中央値は Medcouple です。

$$MC = \text{med}_{y_i \leq Q_2 \leq y_j} \frac{(y_j - Q_2) - (Q_2 - y_i)}{y_j - y_i}$$

ただし、 Q_2 はボックスプロットのセンターラインを定義する定義する第 2 四分位数、 y_i, y_j は基準値のペアです。

Medcouple は常に $-1 \sim 1$ です。対称分布の場合、 $MC = 0$ です。 $MC > 0$ の偏った分布ではより高い基準値に偏り、 $MC < 0$ ではより低い基準値に偏ります。

調整されたアウト라이어値の計算は、分布がどちらの側にシフトしたかに依存しています。

$$\begin{aligned} MC \geq 0: & [Q_1 - 1.5 e^{-4MC} \text{ IQR}; Q_3 + 1.5 e^{3MC} \text{ IQR}] \\ MC < 0: & [Q_1 - 1.5 e^{-3MC} \text{ IQR}; Q_3 + 1.5 e^{4MC} \text{ IQR}] \end{aligned}$$

対称分布 ($MC = 0$) の場合、限界値とボックスとの間の距離は 1.5 IQR です。

M. Hubert und E. Vandervieren, *An adjusted boxplot for skewed distributions*, Computational Statistics & Data Analysis Band 52, Nr. 12

.....

(Aug. 2008), S. 5186–5201 によって経験的に証明されているように、指数関数により、それぞれの歪みを持つさまざまな分布で、正確かつ堅牢なアウトライヤー検出が可能になります。

マーキングされているアウトライヤーの予想パーセントは約 1 % で、正規分布の標準ボックスプロットのパーセントと非常に似ています。注記：このパーセントは有意水準から独立しています。