

# 近赤外分光法による非侵襲血糖値測定と多変量解析法

尾崎 幸洋・森田 成昭

## Non-Invasive Measurement of Blood Glucose by Near-Infrared Spectroscopy and Multivariate Analysis

Yukihiro OZAKI and Shigeaki MORITA

Recently, studies on non-invasive measurements of blood glucose have become active. Because of the development of spectrometers and detectors, it has become possible to measure *in vivo* high quality near-infrared (NIR) spectra of human skin. One must use multivariate analysis methods to develop calibration models for predicting blood glucose concentration from the obtained spectra. In this article, we first outline importance of multivariate analysis and the relation between spectra and multivariate analysis. Then, we report the calibration models for blood glucose concentration built by partial least squares (PLS) regression combined with new wavelength selection methods proposed recently by our group.

**Key words:** near-infrared spectroscopy, blood glucose, partial least squares (PLS), non-invasive measurement, multivariate analysis

本稿では、非侵襲で測定したヒトの皮膚の近赤外スペクトルから血糖値を予測するのに、多変量解析法がなぜ重要かについて解説する。

### 1. なぜ多変量解析<sup>1)</sup>が必要か

図1の破線は非侵襲で測定した皮膚の近赤外スペクトルである<sup>2)</sup>。この図には、参照のために粉末のグルコースのスペクトル（実線）も示してある。*in vivo*で測定した皮膚のスペクトルは水のスペクトルにきわめてよく似ており、それから直接、血糖値に関する情報を得ることは不可能である。皮膚のような生体組織の近赤外スペクトルを解析する場合、大きく分けて3つの問題がある<sup>3)</sup>。1つは「水」という妨害物質の存在である。水の強いバンドをいかに避けるかが重要な問題となる。2番目の問題は、近赤外スペクトルでは多くの倍音、結合音が重なり合うため、多重共線性が強くなるということである。第3の問題は、生体組織内で起こる光の多重散乱の問題である。このような問題の解決に多変量解析は役に立つ。

血糖値の検量モデル作成などに用いられる多変量解析法は、PLS (partial least squares) 回帰分析法とよばれるものである。この方法は、最も一般的な多変量解析法である主成分回帰分析法をさらに発展させたものである<sup>1)</sup>。

### 2. 新しいスペクトル領域選択

さて、図1のようなスペクトルに多変量解析を適用する場合、検討しておくべき重要なことがある。それは、スペクトル領域の選択である。図1に示された領域すべてを無目的に用いるより、一般には特定の領域を選んだほうがよい。最近、筆者らは波数領域選択に関して、moving window partial least squares regression (MWPLSR) 法<sup>4)</sup>, changeable size moving window partial least squares (CSMWPLS) 法<sup>5)</sup>, searching combination moving window partial least squares (SCMWPLS) 法<sup>5)</sup>という3つの方法を提案した。これらは密接に関係するものである。その中で、最も基礎となるのが MWPLSR 法である。

#### 2.1 MWPLSR 法

MWPLSR 法の目的は、すぐれた PLS モデルを構築するために有益な情報を含む領域 (informative region) を選び出すことにある (図2)<sup>4)</sup>。そのため MWPLSR 法では、*i*番目のデータポイントから始まり、(*i*+*h*-1)番目のデータポイントで終了するウィンドウ ( $X_i$ ) を設定する (*h*はウィンドウの大きさ)<sup>4)</sup>。ウィンドウはスペクトルの全領域を動く。全スペクトル領域には  $n-h+1$  個 ( $n$  はデータポイントの数) のウィンドウがあり、それぞれのウィンドウ

関西学院大学大学院理工学研究科化学専攻、近赤外環境モニタリングシステム研究センター (〒669-1337 三田市学園2-1)  
E-mail: ozaki@ksc.kwansei.ac.jp

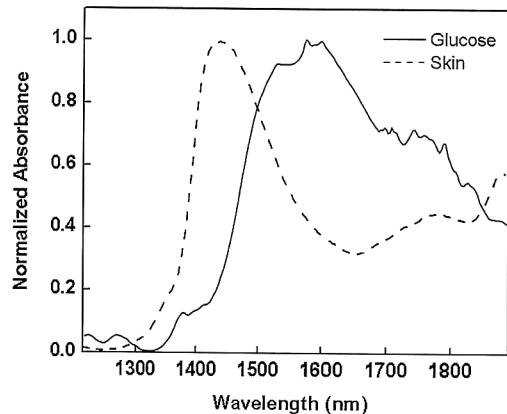


図1 非侵襲で測定した皮膚の近赤外拡散反射スペクトルと粉末のグルコースの近赤外拡散反射スペクトル（文献2より）。

はもとのスペクトル  $X$  の部分領域に対応する。おのおのの部分領域に対して、潜在変数の数を1から  $k$  まで変えてPLSモデルを構築し、残差の二乗和（sums of squared residues; SSR）を計算する（図2）。すべての部分領域に対して SSR の計算が終わったあと、ウィンドウの位置の関数として SSR をプロットする。

残差スペクトルの例を図2に示す。残差スペクトルは2つの重要な情報を含んでいる。1つはinformative regionに関するものであり、もう1つは潜在変数の数の見積もりに関するものである。もちろん、残差の小さい領域ほど情報を多く含んでいることになる。図2では、a, b領域をinformative regionとみなすことができる。特定の潜在変数に対して1つの残差スペクトルが得られるので、図2では多くの残差スペクトルが重なって表示されている。SSRの大きさから、最適潜在変数の数を各informative regionごとに見積もることができる。MWPLSRの残差スペクトルは、それがスペクトルのような形をしている。また、informative regionは、それがバンドが重なった形（1本のバンドのこともある）をしている（図2）。したがって、informative regionの範囲を決めるのは容易である。しかし、こうして選ばれたinformative regionが真に最善の予測結果を与えるとは限らない。それは、informative regionの中によりよい予測結果を与える副領域が隠されている可能性があるからである。informative regionの中からさらに最善の領域を選ぶ方法として考え出されたのがCSMWPLS法である<sup>5)</sup>。また、MWPLSR法により1つ以上のinformative regionが見いだされることが多い。その場合、すぐれたPLSモデルを構築するためにinformative regionの組み合わせが必要となることがある。しかし、単にinformative regionを組み合わせただけで最善の結果が得られるとは限らない。筆者らは、この最善の組み合わ

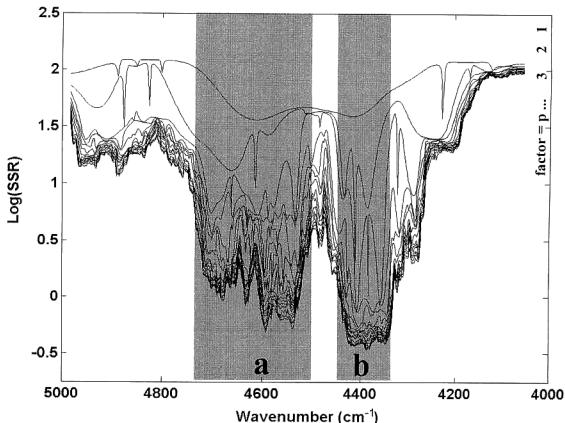
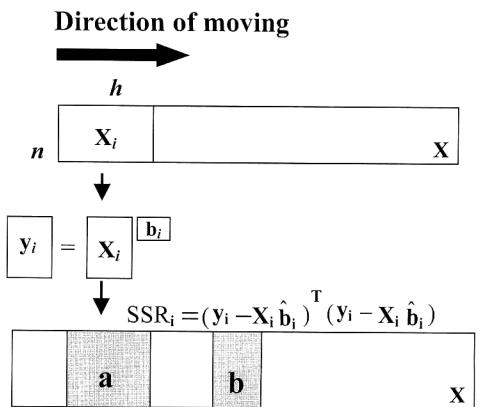


図2 MWPLSR法とMWPLSR法によって計算された残差スペクトルの例。

せをさがす方法としてSCMWPLS法を提案した<sup>5)</sup>。

## 2.2 CSMWPLS法

CSMWPLS法の目的は、MWPLSRで選ばれたinformative regionの中からさらに最適化された領域を選び出すことである。そのために、CSMWPLSでは、与えられたinformative region ( $p$ 個のデータポイントからなるとする)に対してウィンドウの大きさ  $w$  を1から  $p$  まで変化させる。ウィンドウは最初のデータポイントから  $(p-w+1)$  番目のデータポイントまで動き、おのおののウィンドウの大きさに対してすべての可能なサブウィンドウ（sub-window）を集め。たとえば  $w=1$  のとき、ウィンドウを最初の点から最後の点まで走らせ、大きさ1のすべての可能なサブウィンドウを集め。  $w$  が1以外の場合も同様である。CSMWPLSでは、informative regionの範囲内で考えられるすべてのスペクトル領域（副領域）を考慮する。そして、すべてのウィンドウに対し、選択された潜在変数を用いてPLSモデルが構築され、そのRMSEC（root mean square error of calibration）が計算される。最も小さいRMSECを与える副領域を最適化されたスペクトル領域と考える。

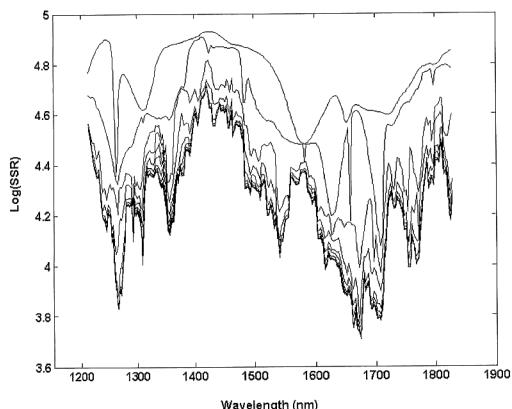


図3 48個の皮膚の近赤外拡散反射スペクトルに対してMWPLSR法を適用して得られた残差スペクトル(文献2より).

### 2.3 SCMWPLS 法

MWPLSRによって得られたおののinformative regionは、CSMWPLSによって選ばれる最適化された副領域を含んでいる。これらの副領域を単に組み合わせても最良のPLSモデルが得られるとは限らない。SCMWPLS法の目的は、informative regionの最適な組み合わせをさがすことがある<sup>5)</sup>。すべてのinformative regionの副領域のすべての可能な組み合わせを考慮すれば(exhaustive search)，最適な組み合わせが得られるはずである。しかし、それには大変な時間を要する。たとえば，a, b, cの3つのinformative regionがあり、それぞれに40, 50, 60個の副領域があったとすると、 $1.9133 \times 10^9$ 個の組み合わせの数があることになる。それが、SCMWPLS法を用いると、組み合わせの数を3925まで減らすことができる。SCMWPLSの最初のステップは、1番目のinformative regionに対してCSMWPLSを適用し、最小のRMSECをもつ最適化された副領域をさがすことである。この領域を基礎領域(base region)とよぶ。第2のステップは、基礎領域と2番目のinformative regionから選ばれた副領域の1つとの組み合わせに対してCSMWPLSを適用し、PLSモデルを構築してRMSECを計算することである。これによって新しい基礎領域を選び出す。次のステップでは、その基礎領域と第3のinformative regionとの組み合わせにより、さらに新しい基礎領域をさがしだす。このようにして、最後のinformative regionとの組み合わせまで、次々と新しい基礎領域をさがしだし、最後に残ったものを最適化された組み合わせと考える。

### 3. MWPLSR, CSMWPLS, SCMWPLS 法を用いた血糖値の検量

Kasemsumranら<sup>2)</sup>は、MWPLSR, CSMWPLS, SCMWPLS法を非侵襲で測定した皮膚の近赤外スペクトル

表1 全領域、MWPLSRによって選ばれた領域、MWPLSRによって選ばれた領域を直接組み合わせた領域、SCMWPLSによって選ばれた領域に対する血糖値の予測結果(文献2より)。

方法	スペクトル領域(nm)	潜在変数の数	相関係数	RMSEV(mg/dl)
全領域	1212~1889	7	0.8936	20.1977
MWPLSR	1228~1323	6	0.8519	24.2398
MWPLSR	1574~1736	4	0.9091	18.3642
MWPLSR	1739~1800	4	0.8302	24.8947
MWPLSR	1228~1323, 1574~1736, 1739~1800	6	0.8840	20.9073
MWPLSR	1228~1323, 1574~1736	6	0.8984	19.4118
MWPLSR	1574~1736, 1739~1800	5	0.9060	18.7775
SCMWPLS	1616~1733	4	0.9205	17.1924

に適用し、血糖値を予測するPLSモデルを構築した。近赤外のスペクトルの測定は、健常者のグルコース負荷試験中に48回、5分ごとに行われた。実験およびデータ解析の詳細は文献<sup>8)</sup>を参照のこと。

図3は、皮膚の近赤外スペクトルに対してMWPLSRを適用して得られた15個の残差スペクトルである<sup>2)</sup>。図3から、3つのinformative region, 1228~1323, 1574~1736, 1739~1800 nmを見つけることができる。1574~1736, 1739~1800 nmの領域はそれぞれ、グルコースのOHとCH伸縮振動の第一倍音が観測される領域である。これら3つの領域にSCMWPLS法が適用された。血糖値の予測は、全領域、MWPLSRによって選ばれた領域、MWPLSRによって選ばれた領域の直接的組み合わせによる領域、SCMWPLSによって選ばれた領域に対して行われた。その結果を表1にまとめる。

MWPLSRによって選ばれた3つのinformative regionのうち、1574~1736 nmの領域は、全領域よりも明らかによい結果を与えた(表1)。これら3つの領域の直接的組み合わせは、必ずしもよりよい結果を与えるなかった。しかしながら、SCMWPLSにより選択された1616~1733 nmの領域は、潜在変数の数、相関係数、RMSEV(root mean square error of validation)のいずれの点においても、他の領域に比べはっきりとよい結果を与えた。

### 文 献

- 尾崎幸洋ほか：化学者のための多変量解析－ケモメトリックス入門(講談社サイエンティフィク, 2002).
- S. Kasemsumran, et al.: Chemom. Intel. Lab. Syst. (in press)
- 尾崎幸洋ほか編著：近赤外分光法(学会出版センター, 1997).
- J. H. Jiang, et al.: Anal. Chem., 74 (2002) 3555-3565.
- Y. P. Du, et al.: Anal. Chem. Acta, 50 (2003) 183-191.

(2004年3月18日受理)