

2S09

## Microsoft Excel を用いたケモメトリクス計算 (3)

## 一重回帰による定量法一

○吉村季織、茂谷明宏、高柳正夫

東京農工大学農学府高柳研究室

(〒183-5809 東京都府中市幸町 3-5-8)

## はじめに

実験的に得られた化学データを数学的・統計的手法により解析する「ケモメトリクス」が、近年盛んに用いられるようになってきた。しかしながら、現在日本においてケモメトリクスの教育は盛んではなく、教材や教法の開発が必要である。教育の場においては、計算結果が得られれば良いというのではなく、ケモメトリクス計算の概念や手順といった計算原理を学ぶことが、計算結果のより深い理解やケモメトリクスの発展のためにも重要であると考えられる。そこで、最も身近な計算ソフトウェアである Microsoft Excel (Excel) を用いたケモメトリクス計算法を開発してきた。第 1 報では行列計算と重回帰を扱った。重回帰は多変量解析の最も基本的な手法であり、定量への応用は近赤外スペクトルの解析に用しばしば用いられ、例えば、セメント中の塩化物イオンの定量[2]、ブルーベリーの糖度推定[3]、コンポスト中の可給態窒素定量[4]などがある。そこで、第 3 報である本報告では吸光スペクトルから定量モデルを作成することを目的とする。スペクトルのシミュレーション、モデル作成、定量のためのワークシートを作成した。モデル作成には吸光の基本法則である Lambert-Beer 則を基にした重回帰 (LBA と呼ぶ：一般には CLS と呼ばれる) と、一般的取り扱いの重回帰 (MLR) による定量モデルの作成を行った。

## スペクトルの生成

ある化学種の吸光度は濃度に比例するもの (Lambert-Beer 則に従っている) とし、化学種間の相互作用はないものとする。想定化学種 2 種、汚染化学種 1 種の混合系 5 試料を 21 個の波長で測定したときのスペクトルを生成するため、図 1 のようなワークシートを作成した。スペクトルにノイズも含ませられるようになっている。生成したスペクトルを図 2 に示す。

## 定量モデル

$M$  個の波長で測定した  $N$  個の校正試料のスペクトルをまとめて  $M \times N$  行列  $A_{cal}$  とする。この  $N$  個の試料中の  $L$  種の化学種の濃度をまとめて  $L \times N$  行列  $C_{cal}$  とする (本報告では、 $L=3, M=21, N=5$  である)。定量のモデルの

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		$\lambda$	1	2	3														
2		$R_1$	1	0.8	0.9														
3		$\lambda_{10}$	2994	3004	3000														
4		$W_1$	8	14	10														
5		$\omega_1$	0.043	0.014	0.028														
6																			
7																			
8	m	$\lambda_m$	$k_{m1}$	$k_{m2}$	$k_{m3}$														
9	1	2980	2E-04	2E-04	1E-05														
10	2	2982	0.002	9E-04	1E-04														
28	20	3018	1E-11	0.05	1E-04														
29	21	3020	2E-13	0.021	1E-05														

図 1 スペクトル生成用ワークシート

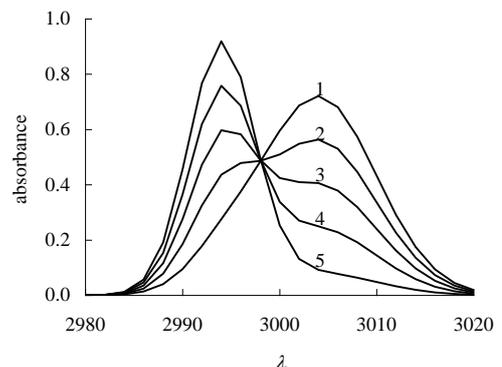


図 2 生成されたスペクトル

作成は、スペクトルから濃度を予測する式を作成することである。

### Lambert-Beer 則を基にした重回帰 : LBA

Lambert-Beer 則を基にすると、 $C_{cal}$  と  $A_{cal}$  の間には、 $L$  種の化学種の単位濃度のスペクトルをまとめた  $M \times L$  行列  $K$  を用いて

$$A_{cal} = KC_{cal} \quad (1)$$

という関係が成り立つ。重回帰により、 $K$  を次のように求める。

$$K = A_{cal} C_{cal}^T (C_{cal} C_{cal}^T)^{-1} = A_{cal} D \quad (2)$$

“T” は転置である。未知試料のスペクトル  $A_{un}$  から未知試料中の化学種の濃度  $C_{un}$  を求めるためには、再度重回帰を用いて、

$$C_{un} = (K^T K)^{-1} K^T A_{un} = JA_{un} \quad (3)$$

となる。これらの計算を Excel で行うためのワークシートが図 3 である。

### 一般的取り扱いの重回帰 : MLR

MLR ではスペクトルから濃度を直接予測する  $C = FA$  というモデルを立てる。スペクトルでは、一般に  $M > N$  のため  $(A_{cal} A_{cal}^T)^{-1}$  が求まらない。そこで、 $A_{cal}$  から適当な波長のデータを抽出した  $A_{cal}^e$  を用いたモデル、

$$C_{cal} = FA_{cal}^e \quad (4)$$

を立て、係数行列  $F$  を次のように求める。

$$F = C_{cal} A_{cal}^{eT} (A_{cal}^e A_{cal}^{eT})^{-1} = C_{cal} B \quad (5)$$

得られた  $F$  を用いて未知試料のスペクトル  $A_{un}$  から該当する波長のデータを取り出し  $A_{un}^e$  を作成し、 $C_{un}$  を予測する。

$$C_{un} = FA_{un}^e \quad (6)$$

図 4 には MLR によるモデルの作成と定量的ためのワークシートを示しておいた。

この Excel ファイルは、<http://www.tuat.ac.jp/~mt2459/chemom/files/chemom3.xls> にてダウンロード可能である。

### LBA と MLR の比較

汚染化学種やスペクトルにノイズがない場合、どちらの方法とも（コンピュータの精度の範囲で）完全なモデルが作成でき、濃度予測も完全である。しかし、汚染(第 3)化学種やノイズがある場合はモデルの完全性は崩れ、また濃度予測にも誤差が生じるようになった。汚染化学種には選択波長が適切であれば MLR の方が優れていた。また、ノイズには局所的な変動の影響を受けにくい LBA の方が優れていることが明らかになった。

### 参考文献

- [1] 吉村季織, 高柳正夫, J. Comput. Chem. Jpn., 7, 71 (2008).
- [2] 郡政人, 古川智紀, 上田隆雄, 水口裕之, セメント・コンクリート論文集, 61, 189 (2008).
- [3] 高井秀悦, 澤田譲, 神佑太, 実践教育, 22, 19 (2007).
- [4] Takayuki FUJIWARA and Keiichi MURAKAMI, Soil Sci. Plant Nutr., 53, 102 (2007).

図 3 LBA 用ワークシート  
①モデル作成、②定量

図 4 MLR 用ワークシート