多変量回帰分析法による数種多環芳香族炭化水素の同時蛍光定量

山口茂六*¹、熊谷 哲*²、西岡 洋*²

*¹姫路短期大学 〒670 姫路市新在家本町1-1-1 2

*²姫路工業大学工学部 〒671-22 姫路市書写2 1 6 7

yamasan@hjc.himeji-tech.ac.jp, kumagai@esci.eng.himeji-tech.ac.jp, nishioka@esci.eng.himeji-tech.ac.jp

(Received : July 2, 1995 ; Accepted for publication : August 30, 1996)

5種類のPAH（ベンゾ[a]ピレン、ペリレン、フルオランテン、ピレン、およびクリセン）の濃度を $10^{-2} \sim 10^{-1} \mu\text{g cm}^{-3}$ の範囲で種々変化させた *n*-ヘキサン溶液の蛍光スペクトル（ E_x : 305 nm, E_m : 330~520 nm）をアスキー形式のデータセットに変換した後マイクロコンピュータに読み込みPLS (Partial Least Squares) 回帰分析法を用いて回帰モデルを構築した。別に作成した濃度既知の合成試料溶液の蛍光スペクトルに回帰モデルを適用し、5種類のPAHについて計算して得られた予測値と良い一致を見た。

Key Words : Polycyclic Aromatic Hydrocarbon, Simultaneous Determination, Multivariate Regression Analysis, Fluorometry

I. 緒 言

光吸収や蛍光を用いた分析では共存する物質の吸収や蛍光の存在が定性や定量を大きく妨害し、発色や発光のための反応試薬が高感度であるにもかかわらず、共存物質の妨害が測定法の実用化を阻んできた例は数多く存在する。試料中の全ての物質が既知であり、スペクトルの非直線性をもたらす干渉がなければ、スペクトルに対するそれぞれの寄与を求め、これまでの線形重回帰分析法により濃度を推定できる。しかしながら、一般的には共存する物質が全て既知であることは期待できないし、互いに影響しあってスペクトルの直線性に影響を及ぼす。重回帰分析の欠点を克服するために主成分回帰分析法やPLS (Partial least squares) 法が提唱されている¹⁻³⁾。これらの方法は説明変数に主成分分析で得られた主成分スコアを用いて最小自乗法によりモデルを構築する手法である。とくに S. Wold 等によって提唱されたPLS法¹⁾は説明変数に加えて潜在変数を導入することにより説明変数の数がサンプル数を上回る場合でも検量モデルを構築することが出来る。逆行列演算を含まないために共線性の問題が生じない。潜在変数を通して説明変数の情報を順次使うので、PLSモデルの自由度を変えながら予測性能を検討することが出来るなどの利点があり期待されている⁴⁾。自動車などの排気ガスと共に排出される多環芳香族炭化水素 (PAH) は癌の誘因物質として知られ、大気中の浮遊粉塵中に存在し我々の

健康に影響を与えている。その種類は多く、極めて微量のため古くは薄層クロマトグラフ等によりそれぞれのPAHに単離後蛍光分析する方法⁵⁾が知られ、近年では分光蛍光検出器付属の高速液体クロマトグラフ法⁶⁾等が用いられている。これはPAHの多くが特有の蛍光を発し、その蛍光スペクトルは比較的鋭いピークを持っているが、多くは重なっているため分離操作を必要とするからである。これはスペクトルデータのうち、最大発光波長のみを使い、大部分のスペクトル情報を捨てていたことに起因している。最近のパーソナルコンピュータの急速な普及と劇的な性能向上は得られたスペクトルデータの全てを用いる統計解析を実験室レベルで可能ならしめた。そこで著者等は選択性に問題があるが高感度である蛍光分析法とPLS法を組み合わせた分析法を共存する数種のPAHの同時蛍光定量に適用し、本法による予測性能を調べた。

II. 確認法 (Validation)

回帰分析の目的は独立に測定されたデータとサンプルのいくつかの属性との間の関係をモデル化することである。本研究では合成試料溶液の蛍光スペクトルと数種のPAHの濃度の関係をPLS法を用いて検量モデルを作成することにより(1)式の関係が得られる。

$$\hat{y}_v = x_v \hat{\beta} \quad (1)$$

ここで $\hat{\beta}$ は検量モデル、 x_v は独立変数で蛍光強度、そして $\hat{y}_v$ は従属変数でPAHの濃度である。

$$\varepsilon_v = y_v - \hat{y}_v \quad (2)$$

(2)式は検量モデルから得られた予測値 $\hat{y}_v$ と真の値 y_v との関係を示したもので ε_v は予測誤差を表している。適当な検量モデルを得る手法の一つとしてはPRESS (Prediction Error Sum of Squares) が最小になる $\hat{\beta}$ の条件を求める方法である。

(3)式に予測誤差 ε_v とPRESSの関係を示した。

$$PRESS = \sum_{v=1}^{n_v} \varepsilon_v^2 \quad (3)$$

ここで n_v は検量モデル作成のために用いた試料溶液の数を示している。(3)式を一般化してSEP (Standard Error of Prediction) を定義する。

$$SEP = \left[\frac{PRESS}{n_v} \right]^{1/2} \quad (4)$$

検量モデルの因子数 k の増加とともにPRESSは減少するが多くなりすぎるとオーバーフィットとなり、予測誤差の原因ともなるので因子数も考慮したSEC (Standard Error of Calibration) が提案されている。

$$SEC = \left[\frac{PRESS}{n_v - k - 1} \right]^{1/2} \quad (5)$$

この因子数 k はスコア、負荷、個有値から E.R.Malinowski⁷⁾ による I D N法やFテストによって計算されるが経験的に両者の平均を採用している。

S E PやS E Cによって得られた値は検量モデル作成に用いたデータセットについて計算して得られた値で、検量モデルとは別のデータセットについて検討したものではない。従って検量モデルの有効性に若干の問題が残っている。しかし、有効性を確認するためのデータセットが常に用意できるとは限らない。その解決策としてクロスバリデーション法が提案されている。これはデータセット中のサンプルを1つ除いた検量モデルを作成し、除いたサンプルの予測を行い誤差を記録する。この手法を全てのサンプルについて繰り返し、得られた誤差について P R E S Sを計算する。これを $SEC V$ (Standard Error of Cross-validation)と定義する。これは検量モデル中にないサンプルを用いて計算したことになるのでより信頼性が高いと考えられている。本研究では5種類のPAHについてそれぞれS E P、S E Cおよび $SEC V$ を計算しその平均値 ($\overline{SEP}$, $\overline{SEC}$, $\overline{SEC V}$) を検量モデル作成の目安とした。尚、P L S法の数学的背景については多数の論文⁸⁻¹²⁾ や成書¹³⁻¹⁷⁾ が出版されているので参照されたい。

III. 実 験

III. 1 試 薬

和光純薬製のベンゾ[a]ピレン(B[a]p)、ペリレン(Pery)、フルオランテン(Fluo)、ピレン(Py)およびクリセン(Chry)の約10mgを精秤し、無蛍光性の n -ヘキサン100mlに溶解して保存溶液とし、冷暗所に保存した。使用に際してはさらに1 μ g/mlに希釈したものをを用いた。

III. 2 合成試料溶液

10mlのメスフラスコを用いてベンゾ[a]ピレン、ペリレン、フルオランテン、ピレンおよびクリセンの濃度を0.01~0.1 μ g/mlの範囲で種々変化させて29種類の合成試料溶液を調製した。

III. 3 測定装置

Table 1 Experimental conditions of fluorometry

Ex λ (励起波長)	305nm
Em λ (蛍光波長)	330~520nm
PMT Gain	Medium
励起バンド幅	10nm
蛍光側バンド幅	10nm
データの取り込み間隔	1.0nm

調製した試料溶液の蛍光スペクトルの測定にはスペクトル補正をした日本分光工業社製分光蛍光光度計F P-777型 (A S C I I変換プログラム付属)を使用した。定量条件

を Table 1 に示す。

Ⅲ. 4 トレーニングセット

Ⅲ. 2 で調製した 29 種類の試料溶液の蛍光スペクトルを Table 1 の条件で測定し、ブランク補正後 A S C I I 形式に変換してロータスに読み込む。蛍光スペクトル部 191 カラムに 5 種類の P A H の濃度項を加え 29×196 のデータマトリックスを構築した。スペクトルデータ項および試料濃度項はそれぞれ独立変数および従属変数として定義した。このデータマトリックスを基にして試料数、データ取り込み間隔など種々条件を変えて構築したトレーニングセットを Table 2 に示した。このうち試料番号 1 ~ 21 を検量モデル作成に用い、試料番号 22 ~ 29 を検量モデルの有効性の判定に用いた。

Table 2 Training sets

No.	Training Set	Number of Samples	Interval (nm)	Independent Variables	Dependent Variables
1	D1	29	1	191	5
2	D2	29	2	96	5
3	D4	29	4	48	5
4	D8	29	8	24	5
5	D20	29	20	10	5
6	D30	29	30	7	5

Ⅲ. 5 データ解析

データの P L S 1^{10} 解析については、データを中央平均化 (Mean Centering) した後 InfoMetrix 社製の Pirouette を用い I B M 社製 P S / V 2 4 1 1 型 (16 MB ラム付属) 卓上型コンピュータ上でおこなった。

Ⅳ. 結果と考察

Ⅳ. 1 蛍光スペクトル

5 種類の P A H ($0.1 \mu\text{g}/\text{ml}$) の蛍光スペクトルを Table 1 の条件で測定した。結果を Fig. 1 の (1) ~ (5) に示した。それぞれの P A H は特徴のあるスペクトルを持つが重なっている。通常これらの P A H を定量する場合、薄相クロマトグラフ法やガスクロマトグラフ法によりあらかじめ分離後個別に定量する方法が採用されている。スペクトル (6) はトレーニングセットに用いた合成試料溶液の蛍光スペクトルの一例である。共存している 5 種類の P A H はお互いに干渉しているので 5 種類の P A H を定量するには線形重回帰分析法よりも P L S 法が有利である。

Ⅳ. 2 試料数と $\overline{SEC}$, $\overline{SECV}$

良好な検量モデルを作成する上でトレーニングセットの大きさが問題になる。トレーニングセットが小さいと未知の試料溶液の特徴を十分説明できないことは明白である。一方

Fig.1 Fluorescence spectra of PAHs.

(1) B[a]p: $10^{-1} \mu \text{g cm}^{-3}$; (2) Pery: $10^{-1} \mu \text{g cm}^{-3}$;
 (3) Fluo: $10^{-1} \mu \text{g cm}^{-3}$; (4) Py: $10^{-1} \mu \text{g cm}^{-3}$;
 (5) Chry: $10^{-1} \mu \text{g cm}^{-3}$; (6) B[a]p: $9 \times 10^{-2} \mu \text{g cm}^{-3}$,
 Pery: $10^{-2} \mu \text{g cm}^{-3}$, Fluo: $9 \times 10^{-2} \mu \text{g cm}^{-3}$, Py: $4 \times 10^{-2} \mu \text{g cm}^{-3}$, Chry: $7 \times 10^{-2} \mu \text{g cm}^{-3}$; Solvent:
 n-Hexane; Ex $\lambda = 305 \text{ nm}$, Em $\lambda = 330 \sim 520 \text{ nm}$.

Fig.2 Relationship between $\overline{\text{SEC}}$ and the sample number of training set.

(1):D1, (2):D2, (3):D4, (4):D30

大きすぎるトレーニングセットは膨大な計算による誤差の集積が予測精度に影響することも考慮に入れなければならない。そこで Table 2 の全てのトレーニングセットを用いて検量モデルに用いられる試料の数が SEC と SEC V におよぼす影響を調べた。Fig. 2 および Fig. 3 に結果の一部を示した。横軸はトレーニングセットを構成するために用いた試料溶液の数である。縦軸はトレーニングセットを用いて 5 種類の PAH について計算して得られたそれぞれの PRESS を SEC および SEC V に換算した後平均したものである。図中の (1)、(2)、(3) および (4) はそれぞれ D1、D2、D4 および D30 を表している。検討したトレーニングセットは何れもよく似た傾向を示すが D30 の場合、高い SEC 値を示している。これは 5 種類の PAH の内ベンゾ[a]ピレン、ペリレンおよびフルオランテンの PRESS が高いことが原因であっ

Fig.3 Relationship between $\overline{\text{SEC V}}$ and the sample number of training set.

(1):D1, (2):D2, (3):D4, (4):D30

た。これはこれらのPAHの特徴を示す蛍光を十分に捕らえきれなかったためである。トレーニングセットの内D1～D8はほぼ同じ傾向を示し、D20はこれらとD30の中間的な傾向を示した。検討した範囲で $\overline{SEC}$ 値は試料数の増加に伴って僅かであるが減少傾向にあるが $\overline{SECV}$ 値は試料数が12以上でほぼ一定の値を示している。

IV. 3 検量モデル

IV. 2の検討でD1～D8のトレーニングセットを用いて構築される検量モデルはSECおよびSECVの何れの条件においてもトレーニングセットに用いられる試料数が12以上で同程度の予測性能を持つことが予想されるがSECおよびSECV間の予測性能の比較はできなかった。そこでTable 2のトレーニングセットでSECおよびSECVの検量モデルの構築に用いられなかった4個のデータを選び、5種類のPAHの濃度項を削除して未知試料とし、Fig. 2およびFig. 3の条件でそれぞれ構築した検量モデルを用いて5種類のPAHの濃度を予測した。Fig. 4はD4についての検討結果を示したものである。(1)および

Fig. 4 Comparison of calibration model.

Trainingset: D4, (1): $\overline{SEC}$, (2): $\overline{SEP}$ (SEC), (3): $\overline{SECV}$, (4): $\overline{SEP}$ (SECV)

(2)はそれぞれFig. 2およびFig. 3におけるD4の $\overline{SEC}$ および $\overline{SECV}$ である。また(3)および(4)はそれぞれSECおよびSECVに対応する検量モデルによる $\overline{SEP}$ を示している。結果は $\overline{SECV}$ とSECによる検量モデルによる $\overline{SEP}$ (SECV)が検討したトレーニングセットの試料数の範囲で極めてよく似た傾向を示し、SECの場合と比較してもSECVの予測性能が高いことを示している。これは検討したD1、D2、D8についてもほぼ同様の傾向であった。

IV. 4 予測性能

合成した試料溶液の内、検量モデルに用いなかった4試料溶液より濃度項を除いたデータにIV. 3で構築したSECVの条件による検量モデルを適用し、予測性能を検討した。結果はD1～D4について、モデルに使用した試料数が12～20では同程度の良好な予測性能を示し、試料数の増加による影響は見られなかった。Table 3はトレーニングセットD4、サンプル数14の検量モデル(LvM14D4)を用いた場合の予測値および実際の値を示したものである。5種類のPAHについて予測残差およびPRESSの最大はそれぞれ $+0.39 (\times 10^{-2} \mu\text{g cm}^{-3}, \text{Chry})$ および $0.252 (\times 10^{-2} \mu\text{g cm}^{-3}, \text{Chry})$ であった。

以上のことから、共存する数種類のPAHをPLS法と蛍光分析法を組み合わせることによって簡便かつ精度良く同時定量する可能性が開かれたものと考えられる。今後は同時定量

が可能なPAHの種類や濃度範囲などの諸条件を検討していきたいと考えている。また、これまで共存物質の妨害のために実用化を阻まれたその他の分析方法についても本法の適用を検討していきたい。

Table 3 Simultaneous five components analysis of PAHs by means of PLS

Sample no.		B[a]p	Pery	Fluo	Py	Chry
1	Actual conc.	4.00	7.00	7.00	10.00	10.00
	PLS, ($10^{-2}\mu\text{g cm}^{-3}$)	3.97	6.82	6.80	9.74	9.72
2	Actual conc.	2.00	10.00	6.00	5.00	4.00
	PLS, ($10^{-2}\mu\text{g cm}^{-3}$)	1.98	10.24	6.13	5.02	3.87
3	Actual conc.	8.00	8.00	8.00	6.00	3.00
	PLS, ($10^{-2}\mu\text{g cm}^{-3}$)	8.20	8.22	8.14	6.00	2.92
4	Actual conc.	2.00	3.00	2.00	9.00	9.00
	PLS, ($10^{-2}\mu\text{g cm}^{-3}$)	1.92	2.84	1.87	9.18	9.39
	SEP ($\times 10^{-2}\mu\text{g cm}^{-3}$)	0.109	0.202	0.156	0.157	0.252

Average SEP of five components: $0.175(\times 10^{-2}\mu\text{g cm}^{-3})$; Model (LvM14D4)

VI. 文 献

- 1) W. J. Dunn, III, S. Wold, J. Edlund, S. Hellberg, and J. Gasteiger, *Quant. Struct. Act. Relat.*, 3, 131 (1984).
- 2) K. R. Beebe, and B. R. Kowalski, *Anal. Chem.* 59, 1007A (1987).
- 3) H. Martens and T. Næs, "*Multivariate Calibration*", John Wiley & Sons, p.9(1989).
- 4) 宮下芳勝, 佐々木慎一, "ケモメトリックスー化学パターン認識と多変量解析", 共立出版, p.55-56 (1995).
- 5) 松下秀鶴, *大気汚染研究*, 10, 723 (1976).
- 6) K. Ogan, E. Katz, W. Skavin, *Anal. Chem.*, 51, 1315 (1979).
- 7) E. R. Malinowski, "*Factor Analysis in Chemistry*", 2nd ed., John Wiley & Sons, p.98 (1991)
- 8) P. Geladi, and B. R. Kowalski, *Anal. Chim. Acta.*, 185, 1 (1986).
- 9) A. Lober, L. E. Wangen, and B. R. Kowalski, *J. Chemom.*, 1, 19 (1987).
- 10) R. Manne, *Chemom. Intell. Lab. Syst.* 2, 187 (1987).
- 11) D. M. Haaland, and E. V. Thomas, *Anal. Chem.*, 60, 1193 (1988).
- 12) D. W. Osten, *J. Chemometrics*, 2, 39 (1988).
- 13) W. H. Press, B. P. Flanny, S. A. Teukolsky, and W. T. Vetterling, "*Numerical Recipes*", Cambridge University Press, p.52 (1986).
- 14) H. Wold, "*Partial Least Squares*", in S. Katz and N. L. Johnson, ed., *Encyclopedia of Statistic Sciences*, Vol. 6, Wiley, New York, p.581 (1985).
- 15) S. Weisberg, "*Applied Linear Regression*", Chapter 5, John Wiley & Sons, New

York (1982).

- 16) R.D.Cook, and S.Weisverg, "*Residuals and Influence in Regression*", Chapter2
John Wiley & Sons, New York (1982).
- 17) 相島鐵郎, "ケモメトリックス", 3, 4, 7 章, 丸善 (1992).

Simultaneous Fluorometric Determination of Polycyclic Aromatic Hydrocarbons by Multivariate Regression Analysis

Shigeroku YAMAGUCHI*¹, Tetsu KUMAGAI*² and Hiroshi NISHIOKA*²

*¹Himeji College of Hyogo, *²Himeji Institute of Technology

A method for the simultaneous determination of polycyclic aromatic hydrocarbons (PAHs) was developed by means of fluorometric analysis combined with partial least squares regression (PLS) analysis. Fluorescence spectra ($E_x=305\text{nm}$, $E_m=330-520\text{nm}$) of twenty one synthetic solutions, which contain five sorts of PAH($10^{-2}-10^{-1} \mu\text{g cm}^{-3}$) in n-hexane, were converted into ASCII data set. These were imported into a microcomputer and modified to make a regression model by employing PLS software. Newly prepared mixtures of all five PAHs (Benzo[a]pyrene, Perylene, Fluoranthene, Pyrene and Chrysene) were analyzed simultaneously using this model. The results predicted were in good agreement with the real data.