

非負寄与を与える発生源探索によるケミカルマスバランス法

花石竜治 佐藤裕久

リセプターモデルの中の発生源寄与解析手法であるケミカルマスバランス法のうち、広く使用されている有効分散最小二乗法について、物理的に解釈不能な負の発生源寄与を与えない「非負寄与を与える発生源探索」機能を備えた計算コードを作成した。解析対象として本邦の底質に対する農薬不純物中のダイオキシン類の影響の問題を取り上げ、上記手法を適用したところ、併せてコードした「非負拘束有効分散最小二乗法」、従来の「負の寄与の発生源除去」と、「非負寄与を与える発生源探索」の決定係数の最適化での検索の3法が同等の結果を与えた。一方、「非負寄与を与える発生源探索」のうち、残差平方和最小の条件で探索したところ、他の「負の寄与の発生源除去」等の結果と異なる結果を与えた。

Key words : chemical mass balance, effective variance, non-negatively fitting, least squares

1. 緒言

ケミカルマスバランス法 (CMB) は、リセプターモデルにおいて、発生源寄与と推定を行う一法である。CMB 計算手法の中で、有効分散最小二乗法 (EVLSQ) はよく用いられ、発生源側の測定誤差と、環境側の測定誤差の両方を考慮できる点で有利である。この方法は、J. G. Watson らのグループにより進展¹⁾がなされ、計算ソフトの最新バージョンが現在では CMB8.2 として、米国環境保護局 (USEPA) の公式のアプリケーションとして頒布されている²⁾。

さて、EVLSQ 法は、他の方法と同様に、物理的に解釈不可能な、負の発生源寄与を算出することがある、という欠点を持っている。そのため、計算方法としては、CMB 計算の過程で負の寄与の発生源を逐次、除去していくという機能があり、この方法は CMB8 でも適用されている。

ところで、C. L. Lawson and R. J. Hanson の教科書³⁾に非負拘束最小二乗法 (NNLSQ) の記述がある。

D. Wang らは、NNLSQ 法を CMB に使用したと報告している⁴⁾。しかし、その報告では、測定値の誤差の考慮が環境データにはなされているが、発生源データにはなされていない。それ以後、不等式で拘束される最小二乗法の CMB 法への積極的応用例は報告されていない。

一方、本邦の統計数学者のグループが、ベイズ統計に基づいて、関数関係解析を行い、それによる CMB 法を開発した⁵⁻⁸⁾。これはリセプターモデルの

研究の範囲を広げる、将来的に有望な方法としての可能性を持つと考えられる。しかし、そこでも発生源データの誤差は考慮されていない。現代の精度管理主流の時代にあつて、環境データはもちろん、発生源データも測定誤差の考慮が必須であろう。また、コンピュータ技術の発達した現代においては複雑な計算も可能になっており、橋本らが指摘している⁹⁾ように、何らかの新しい CMB 計算手法が提案されるべきと思われる。

本論文では、従来の EVLSQ 法の範疇に入り、決して新規な方法ではないが、その中の BestFit という最適発生源検索機能を非負寄与発生源の選択に用いた計算例を紹介する。

2. 方法

2.1 アルゴリズム

原典の CMB8 コードは EVLSQ 法を使用しているが、これを市販の表計算ソフトであるマイクロソフト社の Excel の Visual Basic for Application (VBA) マクロに書き換えたものを元にコーディングした。この CMB8 の書き換え版は、著者と早狩によるもの¹⁰⁾で、CMB8J と名付けて無償頒布してきた¹¹⁾。今回は、この CMB8J の機能を拡張した。

発生源の取捨選択は、すでに公表している CMB8J における、最適発生源の選択の原理と同一である。すなわち、発生源の選択を 1、非選択を 0 として、二進数で選択状態を表し、これを十進数で表記して、

コード上で管理し、EVLSQ 計算を各選択状態について行い、結果の指標をソートし、事後評価で最も良好な発生源の組み合わせを得るものである。表 1 に発生源が 6 個あるときの例を示した。

表 1 発生源の取捨選択と二進数、十進数

発生源	A	B	C	D	E	F
選択・非選択	選択	非	選択	非	選択	選択
フラッグ	1	0	1	0	1	1
二進数表記	101011					
十進数表記	=2 ⁵ ×1+2 ⁴ ×0+2 ³ ×1+2 ² ×0+2 ¹ ×1+2 ⁰ ×1					
対応十進数	43					

この発生源選択状態の管理方法により、発生源の取捨選択状態を整数で管理できている。

さて、事後評価すべき指標であるが、3 つの指標を用意した。すなわち、決定係数 r^2 を最大にする非負寄与の発生源探索、残差平方和 χ^2 および自由度で重みを付けた χ^2 を最小にする非負寄与の発生源探索である。とくに χ^2 については、

$$\chi^2 = \sum_i \frac{(C_i - \sum_j a_{ij} S_j)^2}{\sigma_{ci}^2 + \sum_j \sigma_{aij}^2 S_j^2}$$

である。ここに、 a_{ij} は発生源 j の化学種 i の濃度、 C_i は環境データの化学種 i の濃度、 S_j は、発生源 j の寄与とする。 C_i の誤差の標準偏差を σ_{ci} とし、 a_{ij} の誤差の標準偏差を σ_{aij} と仮定する。

残差平方和 χ^2 をフィッティングの自由度（化学種数－発生源数）で除したものが、自由度で重みを付けた χ^2 である。

図 1 に、CMB8J の入力促進ダイアログを示した。

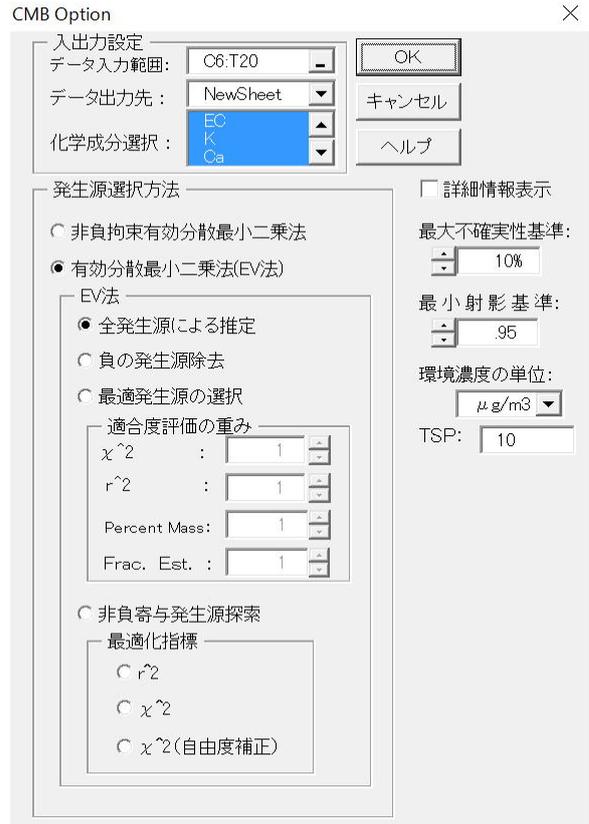


図 1 CMB8J の入力促進ダイアログ

2.2 非負拘束有効分散最小二乗法について

図 1 のダイアログでは、選択すべき計算手法として、「非負拘束有効分散最小二乗法」というものがある。これについて若干の説明をしておく。

緒言の項でも述べたように、最小二乗法には、非負拘束最小二乗法 (NNLSQ) という一法があり、古くから知られているところである。これを、EVLSQ 法と組み合わせたものが、「非負拘束有効分散最小二乗法」である。

EVLSQ 法は繰り返し計算によって発生源寄与を求め、その収束可能性は J. G. Watson らの研究¹⁾で証明されている。他方、CMB8J で採用している NNLSQ 法は、非負拘束法の中でも、回帰分析の用語で "active methods" という方法の一つであり、有限回の繰り返し計算で収束解が得られることが分かっている¹²⁾。

ここに紹介するのみに留めたのは、この方法が EVLSQ 法に付属している「負の寄与の発生源除去」法とは原理的に異なるものであるが、発生源プロファイルの中に深刻な多重共線性がある場合以外は、一般的に解析計算が同じ結果を与えることが多く、全くの新規かつ有力な計算法ではないという理由からである。

「非負拘束有効分散最小二乗法」のアルゴリズム

については、単純な NNLSQ 法と EVLSQ 法の組み合わせであるので、ここでは説明を割愛する。

2.3 CMB 計算における「寄与」について

CMB8J 以後のバージョンでの注意点が一つある。原典の CMB8 では、発生源データを mg/kg 単位、環境データを $\mu\text{g}/\text{m}^3$ 単位で記述し、寄与の次元は $(\mu\text{g}/\text{m}^3)/(\text{mg}/\text{kg})$ であった。しかし、CMB8J では、プロファイルフィッティングという観点での使用も考慮し、環境データが仮に大気中濃度であっても、TSP (全粉じん濃度) で除算し、環境データを mg/kg 単位に換算して、それにより CMB 計算の内部処理を行ってきた。CMB8J ver.2.0 からの最新版¹¹⁾では、とくに注意すべき点として、寄与の次元が $(\text{mg}/\text{kg})/(\text{mg}/\text{kg})$ 、すなわち無次元になっていることである。これにより、mass balance の観点から与えられる “unknown” としての寄与を整合性よく、ワークシートの中で扱えるようにしている。

2.4 ダイオキシン類問題への応用

ここでは適用事例として、ダイオキシン類問題への応用を取り上げる。ダイオキシン類の発生源としては、過去には、ばい煙なども問題視されていたが、規制法の整備により、この 15 年ほどで、格段に環境濃度が低下している。

ここでは、発生源側が過去の農薬で、環境データが海底底質のダイオキシン類問題の場合を取り上げる。ダイオキシン類問題への CMB 法の応用はいくつか既報^{5-6, 8-9)}がある。底質における高濃度の OCDD (八塩素化ジベンゾジオキシン) と OCDF (八塩素化ジベンゾフラン) の例では、燃焼などではなく、農薬不純物による可能性が大きいとも言われている。

農薬におけるダイオキシン類異性体のプロファイルは、文献値¹³⁾を用いた。2,3,7,8-塩素化体のポリ塩素化ジベンゾジオキシン (PCDD) とポリ塩素化ジベンゾフラン (PCDF) の 17 種類の農薬中の含有量を発生源プロファイルとして使用した。ターゲットとした環境データは、2008 年の東京湾の st.5 における底質中のダイオキシン類のデータであった。このデータは、東京都環境局のウェブサイト¹⁴⁾から引用した。

表 2 に、発生源データと、環境データの表を示した。誤差の標準偏差は記載していないが、それは文献¹³⁾に記載がなかったからであり、今回の計算では測定濃度の 5% と仮定した。

3. 結果および考察

3.1 6 法による結果の比較

表 3 に、EVLSQ 法の 2 法、NNLSQ-EVLSQ 法および非負寄与を与える発生源探索の 3 指標での計算結果を示した。

3.2 6 法の比較

全ての発生源による “ALL” モードでは、負の発生源寄与が現れている。また、「負の寄与の発生源除去」(DeIM と表記)、NNLSQ-EVLSQ 法 (NNSLQ と表記) および「非負寄与の発生源探索」(NN と表記) の r^2 を最大化する方法 (NN- r^2 と表記) で同等の結果が出ている。注目すべき差異は、これら 3 法の非負寄与を求める方法の結果と、NN- χ^2 法 (「非負寄与の発生源探索」で自由度補正をしないもの、および自由度補正をしたものの残差平方和最小の発生源探索) との間のものである。

「負の寄与の発生源除去」(DeIM) と「非負拘束有効分散最小二乗法」(NNLSQ-EVLSQ) は、この事例では同一結果を与えている。

両者のアルゴリズムの中で、理解しやすく明確に異なる点を挙げれば、「負の寄与の発生源除去」では不可逆的に発生源を除去していくという点である。

これらに加えて、非負の寄与を与えながら決定係数 r^2 を最大にする発生源の組み合わせが同一の結果を与えたことは、上記 3 つの解析法が、この事例の場合、適切な非負解を与えている証左とも考えられる。

一方、「非負寄与の発生源探索」である NN- χ^2 および NN- r^2 (自由度補正) が異なった非負解を与えた理由は、現在では不明であり、今後、多数のモデルを検討することにより、洞察を深化させていきたい。

4. 結語

「非負寄与の発生源探索」の計算アルゴリズムをコードして、CMB8J に追加して機能させた。

事例として、海底底質中のダイオキシン類問題への応用を試みた。その結果、「非負寄与の発生源探索」での決定係数の最適化が、従来の「負の寄与の発生源除去」法および非負拘束有効分散最小二乗法と同一の結果 (非負解) を与えた。一方、「非負寄与の発生源探索」法である NN- χ^2 および NN- r^2 (自由度補正) が異なった非負解を与えた。

表2 解析に用いたダイオキシン類のデータ

Species	Concentrations of emission sources (mg/kg)												Environment(ng/kg)
	PCP1967	PCP1970	PCP1971	PCPUK	CNP1978	CNP1983	CNP1986	CNP1987	CNP1989	NIP1967	TPN1973	TPN1993	Measured
	A	B	C	D	E	F	G	H	I	J	K	L	
	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.	Conc.
2378-TeCDD	0	0	0	0.0019	0.0051	0	0	0	0	0.000083	0	0	1.2
12378-PeCDD	0.0058	0	0	0.011	0.73	0.34	0.0019	0.00062	0.00027	0.000064	0.0000023	0.0000033	2.7
123478-HxCDD	0.0081	0.0022	0.0077	0.13	0.02	0.028	0	0.000023	0	0	0	0	4
123678-HxCDD	0.87	0.0046	0.12	0.0095	0.48	0.3	0.065	0.00054	0.0014	0.000027	0.0000051	0	9.9
123789-HxCDD	0.038	0	0.0046	0.008	0.17	0.13	0.024	0.00023	0.00052	0	0	0	7.9
1234678-HpCDD	58	0.29	3.1	0.38	0.13	0.075	0.0072	0.00019	0.00025	0.00004	0.00003	0.000034	200
OCDD	5800	34	12	1.8	0	0.0016	0.00061	0	0.000083	0.00014	0.00021	0.0002	3700
2378-TeCDF	0.015	0	0	0.0012	0.0048	0	0.0001	0.000015	0.000023	0.00033	0.00001	0.0000042	4.9
12378-PeCDF	0.0039	0.0038	0.0079	0.005	0.012	0	0	0.000038	0.000078	0.000054	0.000011	0.0000053	6.2
23478-PeCDF	0.0014	0.0018	0	0.0028	0.043	0	0	0.00004	0.000029	0.000012	0.000006	0.0000054	5.1
123478-HxCDF	0.2	0.11	0.091	0.011	0.002	0	0	0	0	0	0.00001	0.000006	11
123678-HxCDF	0.031	0.018	0.033	0.004	0.0098	0	0.0055	0.000027	0.00016	0	0.0000071	0.000006	7.7
123789-HxCDF	0	0	0.0026	0	0	0	0	0	0	0	0	0	0.7
234678-HxCDF	0.014	0.0085	0.012	0.0061	0.82	0.25	0.0028	0.00027	0.00028	0	0.0000055	0.0000066	9.7
1234678-HpCDF	4.6	0.64	0.96	0.052	0.018	0.0098	0.00042	0	0.000024	0	0.00014	0.000038	61
1234789-HpCDF	0.6	0.28	0.26	0.0053	0	0.0002	0.000024	0	0	0	0.0000075	0	8.5
OCDF	47	4.4	3	0.081	0	0.00063	0.00011	0	0	0	0.024	0.0027	170

表3 EVLSQ法での発生源寄与計算結果(寄与率%)

方法	A	B	C	D	E	F	G	H	I	J	K	L
ALL	0.0001	-0.0020	0.0066	0.0035	0.0007	-0.0011	0.0027	-0.7051	0.9837	1.2698	-13.2230	111.6641
DelM	0.0013		0.1382	0.0812	0.0010				20.4703	38.8539		40.4540
NNLSQ	0.0013	0.0000	0.1383	0.0812	0.0010	0.0000	0.0000	0.0000	20.5075	38.8578	0.0000	40.4129
NN-r ²	0.0013		0.1382	0.0812	0.0010				20.4703	38.8539		40.4540
NN-χ ²				8.7673						91.2327		
NN-χ ² (自由度補正)				8.7673						91.2327		

文献

1)Watson, J. G. et al.:The effective variance weighting for least squares calculations applied to the mass balance receptor model.*Atom. Environ.*,**18**,1347-1355(1984).
 2)https://www3.epa.gov/scram001/receptor_cmb.htm
 (2016年12月現在アクセス可能)
 3)Lawson, C. L. et al.:Solving Least Squares Problems.Prentice-Hall,New Jersey(1974).
 4) Wang, D. et al.:The use of constrained least-squares to solve the chemical mass balance problem.*Atom. Environ.*,**23**, 2143-2150(1989).
 5) 柏木宜久他:関数関係解析による Chemical Mass Balance.応用統計学,**31**, 59-74(2002).
 6) 鈴木貴博他:関数関係解析によるケミカルマスバランス法を用いたダイオキシン類の発生源寄与率推定法に関する検討.環境化学,**16**,437-448(2006).
 7)Kashiwagi, N.:Chemical mass balance when an unknown source exists.*Environmetrics*,**15**,777-796(2004).
 8) 柏木宜久他:環境汚染に対する未確認発生源の寄与率の推定. 統計数理,**54**,123-146(2006).
 9) 橋本俊次他:ケミカルマスバランス法による環境汚染物質に対する発生源寄与率の推定.ぶんせき,**144**-151(2014).
 10) 早狩進他:環境データ解析用表計算マクロの紹介と解析例(Ⅱ)—CMB法解析マクロ—. 大気環境学会誌, **36**, 39-45(2001).
 11)<http://rhanaiishi.world.coocan.jp/>

(2016年12月現在、アクセス可能)
 12)Chen, D.et al.:Nonnegativity constraints in numerical analysis. in: *Symposium on the Birth of Numerical Analysis*, Leuven Belgium(2007).
 13)Masunaga, S. et al.:Dioxin and dioxin-like PCB impurities in some Japanese agrochemical formulations.*Chemosphere*,**44**,873-885(2001).
 14)<http://www.kankyo.metro.tokyo.jp/chemical/chemical/dioxin/result/index.html>
 (2016年12月現在アクセス可能)