

最尤法による非負寄与を与えるケミカルマスバランス法

花石竜治

環境汚染における発生源寄与解析手法であるリセプターモデルのケミカルマスバランス (CMB) 法に、計算手法として最尤法を適用した例を取り上げる。最尤法は尤度関数を最大とする確率変数を求める方法である。最尤法の解法のうち、非負拘束を制約条件とした場合の非線形最適化解法の一つが「主双対内点法」である。本論文では、主双対内点法により CMB を解く機能を表計算ソフトのマクロとして作成した計算手法について紹介する。

Key words: Chemical mass balance, Primal-dual interior-point method, Non-linear optimization

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、発生源から環境の測定点へ、元素や難分解性化合物の化学種がその組成を保持しながら到達すると考え、発生源からの寄与の線形和 (重ね合わせ) によって環境測定点における化学種の濃度が決まるとし、これにより環境問題における発生源寄与を求める方法である。

CMB 計算を重ね合わせの逆と解釈すれば、最小二乗法が計算手法として選択されることは、想像に難くない。この手法のうち、発生源側の測定誤差および環境側の測定誤差の両方を考慮できる有効分散最小二乗法 (EV)¹⁾ は多用され、現在も米国環境保護局 (EPA) の公認解析手法²⁾ となっている。

今般のコンピュータ技術の進展により、より高度な計算を手軽に行うことができる時代となって、我が国の柏木らが新しい CMB 法を開発した。それらには、関数関係解析を利用した最尤法^{3,4)} とともに、CMB 計算の中で最も計算量の多い「ベイズ統計」による方法^{5,6)} も含まれる。これらの計算手法の開発は、従来法の最小二乗法が物理的に解釈不可能な負の寄与を与える可能性があること⁷⁾ が動機となっているようである。

筆者らが 2016 年度に公表した非負拘束有効分散最小二乗法⁸⁾ は、負の発生源寄与を与えない方法ではあるが、そのアルゴリズムは、EV の選択肢である「負の寄与の発生源除去」法とほぼ一致する解を与える。

本論文では、柏木らが提案した「最尤法」

(Maximum Likelihood Method) による CMB 計算手法を、近年の数値計算環境の改善で取り組みやすくなった最適化数学の知見を取り入れて改良し、従来法と同様に表計算ソフト上で完結するソフトウェアとしたので紹介する。

2. 理論

2.1 CMB 計算への最尤法の適用

CMB 計算の前提を次のとおりとする。

- ① 発生源からの寄与が線形和 (重ね合わせ) である
- ② 環境の化学種濃度の偶然誤差が正規分布をなす

ここで、化学種 i について、環境濃度が y_i 、発生源 j でのその化学種濃度が x_{ij} 、発生源寄与が g_j 、真の発生源濃度が ζ_{ij} であるとする。

いま環境濃度の偶然誤差の分散が実験的に求められており、誤差分散 $\sigma_{y_i}^2$ で正規分布をなすと仮定すれば、発生源プロファイルの不確かさを考慮しない場合 ($\zeta_{ij} = x_{ij}$)、尤度 $L(\mathbf{g})$ は

$$L(\mathbf{g}) \propto \prod_i \exp \left[-\frac{(y_i - \sum_j \zeta_{ij} g_j)^2}{2\sigma_{y_i}^2} \right] \quad \dots(1)$$

と表せる³⁾。この尤度を最大化する確率変数 g_j の真値を求めることが最尤法である。

2.2 最尤法と非線形最適化⁹⁾

尤度 $L(\mathbf{g})$ の最大化は、それらの対数尤度の最大化と等価であり、式 (2) で表す関数 $f(\mathbf{g})$ の最小化に帰着される。

$$f(\mathbf{g}) = \sum_i \frac{(y_i - \sum_j \zeta_{ij} g_j)^2}{2\sigma_{y_i}^2} \quad \dots(2)$$

さらに、不等号制約条件として、式 (3) で表すとおおり、寄与 g_j がすべて非負と仮定する。

$$\forall g_j \geq 0 \quad \dots(3)$$

この問題は、式 (2) が非線形であり、また不等式 (3) による制約条件が存在するので、「不等号制約条件がある場合の非線形最適化」に帰着される¹⁰⁾。

2.3 主双対内点法による解

「不等号制約条件がある場合の非線形最適化」の解法の一つである「主双対内点法」の適用に必要な数式を導出する。

ヘッセ行列の要素 $\nabla_{gg}^2 f(\mathbf{g})_{km}$ 、勾配ベクトルの要素 $\nabla_g f(\mathbf{g})_k$ は次式に示すとおおりとなる。

$$\nabla_{gg}^2 f(\mathbf{g})_{km} = \frac{\partial^2 f(\mathbf{g})}{\partial g_k \partial g_m} = \sum_i \frac{\zeta_{ik} \zeta_{im}}{\sigma_{y_i}^2}$$

$$\nabla_g f(\mathbf{g})_k = \frac{\partial f(\mathbf{g})}{\partial g_k} = - \sum_i \frac{(y_i - \sum_j \zeta_{ij} g_j) \zeta_{ik}}{\sigma_{y_i}^2}$$

行列・ベクトルの方程式は以下のとおおりである。ここで、 nso は発生源の数である。

$$\mathbf{B}_{l,g} = \begin{bmatrix} \frac{\partial^2 f(\mathbf{g})}{\partial g_1^2} & \dots & \frac{\partial^2 f(\mathbf{g})}{\partial g_1 \partial g_{nso}} \\ \vdots & \frac{\partial^2 f(\mathbf{g})}{\partial g_k \partial g_m} & \vdots \\ \frac{\partial^2 f(\mathbf{g})}{\partial g_{nso} \partial g_1} & \dots & \frac{\partial^2 f(\mathbf{g})}{\partial g_{nso}^2} \end{bmatrix}$$

$$\boldsymbol{\mu}_g^l = \text{diag}(\mu_{g,1}^l, \mu_{g,2}^l, \dots, \mu_{g,nso}^l)$$

$$= \begin{bmatrix} \mu_{g,1}^l & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mu_{g,nso}^l \end{bmatrix}$$

$$\mathbf{G}^l = \text{diag}(g_1^l, g_2^l, \dots, g_{nso}^l)$$

$$= \begin{bmatrix} g_1^l & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & g_{nso}^l \end{bmatrix}$$

解くべき連立方程式は式 (4) である¹⁰⁾⁻¹²⁾。

$$\begin{bmatrix} \mathbf{B}_{l,g} & -\mathbf{I} \\ \boldsymbol{\mu}_g^l & \mathbf{G}^l \end{bmatrix} \begin{bmatrix} \Delta \mathbf{G}^l \\ \Delta \boldsymbol{\mu}_g^l \end{bmatrix} = \begin{bmatrix} -\nabla_g f(\mathbf{g}) \\ \mathbf{W}\mathbf{1} - \mathbf{G}^l \boldsymbol{\mu}_g^l \end{bmatrix} \quad \dots(4)$$

ここで、 Δ を増分の記号として、

$$\Delta \mathbf{G}^l = [\Delta g_1^l, \Delta g_2^l, \dots, \Delta g_n^l]^T$$

$$\Delta \boldsymbol{\mu}_g^l = [\Delta \mu_{g,1}^l, \Delta \mu_{g,2}^l, \dots, \Delta \mu_{g,n}^l]^T$$

$$\mathbf{W} = \sum_j g_j^l \boldsymbol{\mu}_{g,j}^l$$

$$\mathbf{1} = [1, 1, \dots, 1]^T$$

$$\mathbf{G}^l \boldsymbol{\mu}_g^l = [g_1^l \mu_{g,1}^l, g_2^l \mu_{g,2}^l, \dots, g_{nso}^l \mu_{g,nso}^l]^T$$

また、上付きの T は転置を表し、 \mathbf{I} は単位行列である。上記により、解の増分 ($\Delta \mathbf{G}^l$) および計算過程のパラメータの増分 ($\Delta \boldsymbol{\mu}_g^l$) を求める。(4) 式の左辺の行列は正則であることが証明されている¹⁰⁾。

3. 計算方法

3.1 主双対内点法における連立一次方程式計算

方程式 (4) は、部分的ピボットニングを行うガウス消去法で解いた。

3.2 繰り返し計算法

方程式 (4) の解であるパラメータの増分を求め、既報¹³⁾の方法により、繰り返し計算を行って、

g_j^l 、 $\mu_{g,j}^l$ を更新した。

3.3 計算の初期値および収束条件

前項の繰り返し計算における発生源寄与の初期値は、 i 番目の環境濃度を y_i 、発生源プロファイルでの i, j 番目の要素を $\zeta_{ij} (> 0)$ としたとき、 j 番目の発生源寄与の初期値 g_j^l を次式で取った。

$$g_j^l = \min\left(\frac{y_i}{\zeta_{ij}}\right)$$

収束条件は、

$$DE = 10^{-2} \cdot \max(g_j^l)$$

として、繰り返しの10回毎について

$$\left| \sum_{k=l}^{l+10} \max(|\Delta g_j^{k+10}|) - \sum_{k=l}^{l+10} \max(|\Delta g_j^k|) \right| < DE$$

となることとした。

3.4 用いたCMBモデルと比較のための計算手法

米国EPAのCMBダウンロードサイト²⁾で公表されている発生源プロフィールと環境データを使用した(pack2, 1978年1月24日、微小粒子)。表1にテストデータを示した。

また、花石らが既報⁸⁾で解析例を報告した、東京湾底質のダイオキシン類をPCPなどの農薬不純

物由来と仮定する解析を今回の最尤法でも試行した。データセットは既報⁸⁾にあるのでここでは略す。

非負拘束法を含む有効分散最小二乗法(EV)の計算は、筆者が頒布しているCMBマクロ^{12),13)}により行った。方法は、(1)EV全発生源選択、(2)EV負の寄与を与える発生源除去法、(3)非負拘束EVおよび(4)最尤法を試みた。

4. 結果

4.1 大気中微小粒子(米国EPAテストデータ)の解析結果

表2に寄与率を示し、図1に寄与率の円グラフを描画した。なお、「(a)EV全発生源」では表2の寄与率が図1の数値と異なっているが、これは負の寄与を描画できないため、非負の寄与だけで描画しているためであり、間違いではない。

4.2 底質のダイオキシン類の解析結果

表3に寄与率を示し、図2に寄与率の円グラフを描画した。「(a)EV全発生源」については前項と同様である。

表1 テストデータとして用いた発生源プロフィールおよび環境濃度(米国EPAテストデータ)

成分	発生源 (mg/kg)																環境濃度(μg/m3)	
	MARINE		UDUST		AUTPB		RDOIL		KRAFT		ALPRO		STEEL		FERMN		実測値	標準偏差
NA	400000	40000	12500	3400	0	500	35000	17000	127000	7100	41000	17000	12600	4800	31000	3000	0.19	0.01
MG	48000	9000	13000	2700	0	5000	0	30000	6300	5300	28000	18000	65000	7100	0	40000	0	0.08
AL	0	0	88400	27100	11000	5000	5300	2400	2500	140	270000	80000	6500	1600	6400	1300	0.19	0.01
SI	0	0	223000	11000	8200	3000	9600	4800	1500	140	3400	3400	50000	2100	9900	3000	0.39	0.02
CL	400000	100000	0	100	30000	10000	0	1000	18000	4200	13300	8400	18500	7800	4200	1600	0.79	0.03
K	14000	2000	10300	600	720	290	2800	1000	15000	1400	2200	600	9200	700	105000	39000	0.19	0.01
CA	14000	2000	24400	4000	12500	5000	15800	6400	0	5000	3300	300	62000	5700	13000	1400	0.11	0.01
TI	0	0	8400	1200	0	1000	1100	380	60	20	400	300	2000	200	460	70	0.011	0.001
V	0	0	230	50	0	50	34400	7500	10	10	640	360	600	100	240	190	0.028	0.001
CR	0	0	450	170	0	100	470	150	2800	800	0	100	21000	14000	420	30	0.004	0.001
MN	0	0	1230	170	0	160	460	130	300	40	110	80	87000	9000	173000	9900	0.027	0.001
FE	0	0	60000	6000	21000	8000	29700	6100	12000	1100	4500	2900	320000	30000	21000	14000	0.227	0.008
NI	0	0	90	30	180	80	53600	12100	1300	140	1900	850	7000	700	0	50	0.043	0.002
CU	0	0	300	120	730	300	750	250	210	130	440	70	2800	300	360	40	0.033	0.002
ZN	0	0	1100	370	3500	1300	4000	1800	690	210	150	110	12000	1200	5800	2500	0.106	0.004
BR	2000	500	200	60	50000	17000	130	210	1300	600	370	540	0	1000	1600	2200	0.419	0.012
PB	0	0	3700	1500	200000	30000	1100	640	130	110	120	100	7600	700	450	160	0.94	0.02
OC	0	0	118000	43000	500000	100000	70000	62000	17000	2100	39000	39000	0	0	90000	9000	15.3	1.6
EC	0	0	18500	9100	38000	14000	31000	25000	2200	2300	23000	10000	0	0	15000	1500	1.8	0.4
SO4	100000	40000	4200	3100	13000	4000	481000	119000	400000	21000	44000	39000	25000	13000	42000	4000	6.9	0.3
NO3	0	0	0	100	9100	3000	6500	4400	0	3000	4100	1800	0	14000	57000	28000	2.9	0.2
																	TSP	43

表2 CMB計算で得られた寄与率(米国EPAテストデータ)

発生源	MARINE	UDUST	AUTPB	RDOIL	KRAFT	ALPRO	STEEL	FERMN
EV全発生源	-11.82	4.02	58.99	2.54	49.12	-1.53	-3.48	2.15
EV負の寄与発生源除去		4.42	77.98	6.32	10.59			0.69
非負拘束 EV	0.00	4.38	77.98	6.34	10.61	0.00	0.00	0.69
最尤法	5.61	17.57	65.14	9.61	0.25	0.00	0.13	1.69

表3 CMB計算で得られた寄与率(東京湾底質のダイオキシン類)

発生源	PCP1967	PCP1970	PCP1971	PCPUK	CNP1978	CNP1983	CNP1986	CNP1987	CNP1989	NIP1967	TPN1973	TPN1993
EV全発生源	0.00	0.00	0.01	0.00	0.00	0.00	0.00	-0.71	0.98	1.27	-13.22	111.66
EV負の寄与発生源除去	0.00		0.14	0.08	0.00				20.47	38.85		40.45
非負拘束 EV	0.00	0.00	0.14	0.08	0.00	0.00	0.00	0.00	20.51	38.86	0.00	40.41
最尤法	0.00	0.00	0.06	0.05	0.00	0.00	0.06	1.59	7.21	24.63	1.68	64.70

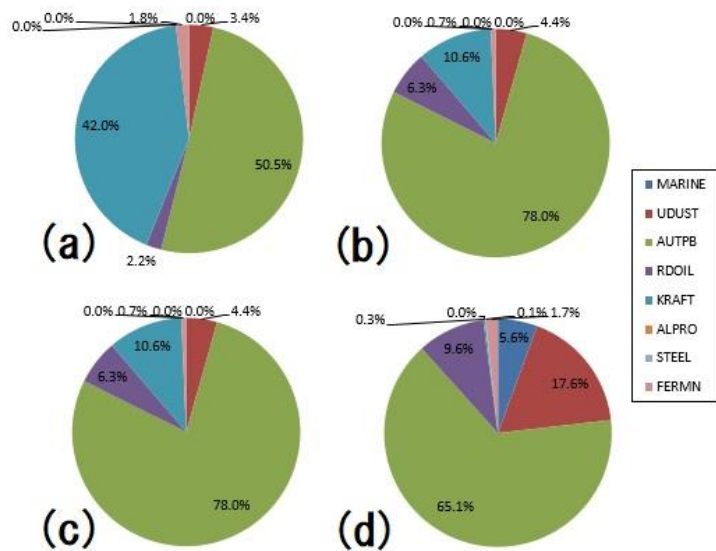


図1 米国 EPA テストデータについての CMB 計算により得られた発生源寄与率。(a) EV 全発生源選択、(b) EV 負の発生源寄与除去、(c) 非負拘束 EV 法、および (d) 最尤法。

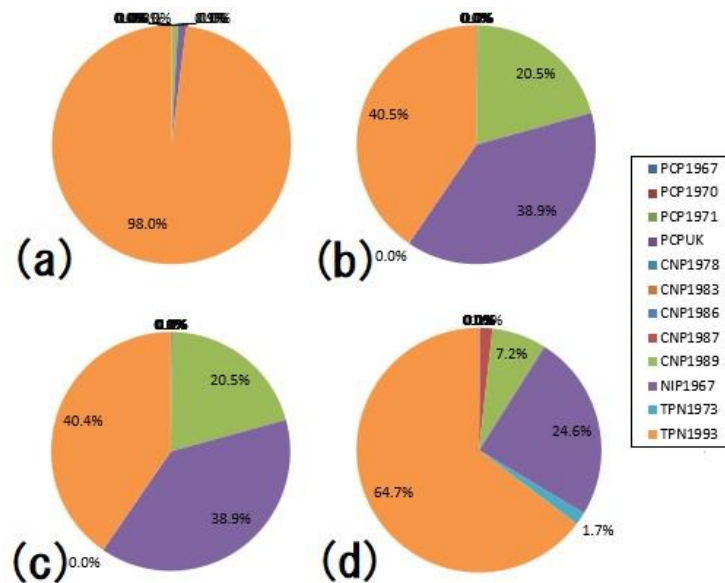


図2 東京湾底質 DXN 類についての CMB 計算により得られた発生源寄与率。(a) EV 全発生源選択、(b) EV 負の発生源寄与除去、(c) 非負拘束 EV 法、および (d) 最尤法。

5. 考察

5.1 多重共線性がある場合の有効分散最小二乗法 (EV) と最尤法での寄与の与え方について

今回用いたデータセット (米国 EPA テストデータおよび東京湾底質 DXN 類) では、全発生源を用いた有効分散最小二乗法 (EV) の結果で、負の寄与が現れており、その回避が必要であった。

これら 2 種類のデータで、「全発生源 EV」で負の寄与を与えた発生源の寄与が、「EV 負の寄与の発生源除去」および「非負拘束 EV」でゼロになっている。これはこれらの計算アルゴリズムが同じように負の寄与の発生源を除去するからと考えら

れる。

一方、最尤法は、米国 EPA テストデータにおいて、「全発生源 EV」で負の寄与を与えた発生源について、この方法の収束解として正の発生源寄与を与えた。この結果は、前述の二つの方法と際立って違う点であった。最尤法は「非線形最適化」を解く方法に置換されるが、その最適解は、単純な発生源選択 (全か、あるいは部分的に取捨選択するかどうか) がなされる場合に比べて、より真値に近い結果を与えられられた。

5.2 多重共線性により解として対をなす発生源

表2で、米国EPAテストデータについてのEV計算結果を最尤法計算結果と比較すると、KRAFTという発生源寄与が、前者で高く算出されたことが分かった。このKRAFTの寄与率は、負の発生源除去および非負拘束有効分散最小二乗法では、全発生源選択の場合よりも低く算出された。すなわち、KRAFTは多重共線性の影響を比較的強く受ける発生源の可能性があると考えられた。

一方、発生源MARINEは、「全発生源EV」で負の寄与率を与え、方法(b)および(c)で寄与がゼロ、(d)最尤法で正の寄与を与えた。最尤法で発生源KRAFTの寄与率がそれ以外の方法による結果と比べて、著しく低下していることを考え合わせれば、発生源KRAFTとMARINEは多重共線性があり、最小二乗法を行う線形代数の計算上、対をなして挙動すると考えられた。

6. 結語

最尤法によるCMB計算について、非線形最適化の一法である主双対内点法により数式を展開し、その解法を表計算ソフトのマクロに組み込み、表計算ソフトで計算が完結するようにした。これは、データの取扱いの上で取り組みやすい解析環境を与えたと考える。

データセットの一例を取り上げ、最尤法による発生源寄与の算出結果を従来法の有効分散最小二乗法によるものと比較した。

今回取り上げたデータセットはあくまでも一例に過ぎず、今後、この方法による発生源寄与推定結果を積み重ねることで、方法の長所および短所が明らかになると考える。

謝 辞

イサラ研究所 齊藤勝美氏には、本プログラムコードの開発に当たり、試用していただいた。ここに記して感謝申し上げます。

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
(2018年10月現在アクセス可能)

3) 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.

4) 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.

5) Kashiwagi, N.: Chemical mass balance when an unknown source exists. *Environmetrics*, **15**, 777-796, 2004.

6) 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.

7) 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.

8) 花石竜治, 他: 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. *青森県環境保健センター研究報告*, **27**, 70-73, 2016.

9) 東京大学教養学部統計学教室 編: 基礎統計学 III 自然科学の統計学. 東京大学出版会, 東京, 1992.

10) 山下信夫: 応用最適化シリーズ 非線形計画法. 朝倉書店, 東京, 2015.

11) 寒野美博, 他, 東京大学工学教程編纂委員会編: 基礎系数学 最適化と変分法. 丸善出版, 東京, 2016.

12) Boyd, S. et al.: *Convex Optimization*. Cambridge, UK, 2004.

13) <https://github.com/convexbrain/Totsu/blob/master/solver/src/PrimalDualIPM.cpp>

(2018年10月現在、アクセス可能)

14) <http://rhanaiishi.world.coocan.jp/>

(2018年10月現在、アクセス可能)

15) 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

A chemical mass balance method giving nonnegative source contribution by the maximum likelihood method

Ryuji Hanaishi

The author takes up an example where the maximum likelihood method is applied to the chemical mass balance (CMB) method of the receptor model which is a source apportionment analysis method for environmental pollution. The maximum likelihood method is a method of finding random variables that maximize the likelihood function. Of the calculation methods of the maximum likelihood method, the "primal-dual interior-point method" is one of nonlinear optimization solutions subjected to nonnegative constraints. In this paper, the author introduces the calculation method which solves the CMB by a macro of the spreadsheet software using the primal-dual interior-point method.

Key words: Chemical mass balance, Primal-dual interior-point method, Non-linear optimization