

<報 文>

ベイズ統計手法によりケミカルマスバランス法計算を行う 表計算ソフトマクロの作成*

花石 竜治**

キーワード ①ケミカルマスバランス法 ②ベイズ統計 ③マルコフ連鎖モンテカルロ法 ④未知発生源

要 旨

環境汚染における発生源寄与解析手法であるレセプターモデルのケミカルマスバランス法のうち、ベイズ統計手法をマルコフ連鎖モンテカルロ (MCMC) 法と組み合わせた柏木らの既報 (2006) のアルゴリズムを、表計算ソフトのマクロに移植した。移植に当たって数値計算法を改良した。この方法では、モデル式および尤度に明確に未知発生源の寄与と組成を組み入れており、これらを評価でき、また物理的に意味のない負の発生源寄与を理論的に与えない。さらに、この方法は比率モデルを用いているため、従来よりも、物理的に解釈がより容易な発生源寄与を与える。計算結果から、本研究の方法は発生源解析に有用であることが示された。

1. はじめに

環境汚染解析において、ケミカルマスバランス (Chemical Mass Balance, CMB) 法は、Positive Matrix Factorization (PMF) 法と並んでよく用いられている¹⁾。

CMB法では、発生源サイドの化学種濃度である発生源プロファイルと、環境サイドの化学種濃度から、数理解析により、発生源寄与を求める。

この計算ソフトとして最もよく用いられてきたのが、米国環境保護庁 (USEPA) のCMB8²⁾である。このソフトウェアは、計算法として、発生源サイドと環境サイドの双方の観測誤差を考慮する「有効分散最小二乗法 (EVLSQ)」³⁾を採用している。

最小二乗法を用いるCMB法では、発生源プロファイルに、組成比の類似した発生源がある場合、これらの発生源に対して、負の寄与を与えることがあり、これまでのCMB計算法開発においては、この負の寄与を回避することに重きが置かれてきた^{4),5)}。この中で、我が国の柏木らの「ベイズ統計手法によるCMB法」⁶⁾は、最小二乗法などの従来法とは原理的に異なる特徴を持ち、また、未知発生源の存在を仮定し、解析によりその寄与と組成を求めることができる。

柏木らのCMBソフトはスタンドアロン⁷⁾であるが、早狩および著者は、USEPAのCMB8を、表計算ソフトとして

用いられているマイクロソフト社エクセルのマクロ機能に移植した経緯がある⁸⁾。そこで、本研究では、柏木らのベイズ統計手法によるCMB計算法を、数値計算法を改めつつ、エクセルマクロに移植し、表計算ソフト上であたかもその固有の機能のようにCMB計算ができるようにしたので報告する。

2. マルコフ連鎖モンテカルロ (MCMC) 法と柏木らのベイズ統計手法の概略

以下にMCMC法⁹⁾について概略を示す。

モンテカルロ法とは、乱数を用いて数理モデル計算を行う方法の総称であり、その論拠を「大数の法則」に置く。これは、R個の乱数を用いたモデル計算結果で、Rを無限大に取った場合の標本平均値が、その期待値に等しくなるというものである。

マルコフ連鎖とは、過去-現在-未来の、時系列の確率過程において、「未来」が「現在」だけに依存し、「過去」の記憶がないものを指す。

その連鎖において、時系列をつなぐ因子、マルコフ連鎖においては「現在」から「未来」を関係づける因子を推移核という。多数回の推移核の乗算で、確率分布がそれ以上変化しなくなった場合、その確率分布を定常分布という。

*A spreadsheet application macro for calculation of a chemical mass balance by means of Bayesian statistical technique

**Ryuji HANAISHI (青森県環境保健センター) Aomori Prefecture Public Health and Environment Center

柏木らのCMB研究⁶⁾の要点は、CMB法にベイズ統計を導入した場合の確率変数の母数としてのパラメータについての尤度（ゆうど、もっともらしさを示す式）から、ギブズサンプラーと呼ばれる推移核を採用して、条件付き確率分布の式である切斷（打ち切り）正規分布のパラメータの数式を導出したことであった。この数式によりMCMC法の計算が可能になり、これを繰り返すことにより、CMBパラメータの期待値などを求める方法が確立された。

3. 理論

3.1 CMBモデル式

柏木らによるCMBモデル式⁶⁾を比率モデル¹⁰⁾で記述すると次式のとおりとなる。

$$\zeta_i = \sum_{j=1}^{n_j} \xi_{ij}\theta_j + \omega_i \left(1 - \sum_{j=1}^{n_j} \theta_j \right) \quad (1)$$

条件として、

$$\omega_i \geq 0, \sum_{i=1}^{n_i-1} \omega_i \leq 1, \theta_j \geq 0, \sum_{j=1}^{n_j} \theta_j < 1$$

ここで、

ζ_i : 環境サイド化学種濃度（単位:g/g）

ξ_{ij} : 発生源プロファイルの行列要素（単位:g/g）

θ_j : 寄与（無次元）

ω_i : 未知発生源の組成（単位:g/g）

n_i : 化学種の数に1を加えた数

n_j : 発生源の数

i : 化学種を示す添え字

j : 発生源を示す添え字

である。

柏木らの原典⁶⁾では、比率モデルにおいて、 n_i 個の化学種の濃度の和が 1 g/g となることを想定している。よって、独立な化学種濃度は、 $n_i - 1$ 個となる。

一方、本研究では、 n_i' 個の観測可能な化学種に着目し、この個数の化学種が独立な濃度を持つと考え、完全な組成把握を前提としなかった。よって柏木らの記述の化学種数について、 $n_i = n_i' + 1$ 個となることから、柏木らの取り扱いの n_i を化学種の数+1と読み替える。

以下で、環境サイド化学種濃度 ζ_i と発生源プロファイル濃度測定値 ξ_{ij} の実現値を、それぞれ Z_i , x_{ij} とし、これらのベクトルおよび行列表現を \mathbf{Z} , \mathbf{x} とする。また、寄与 θ_j および未知発生源の組成 ω_i のベクトル表現をそれぞれ $\boldsymbol{\theta}$, $\boldsymbol{\omega}$ とする。

これらのベクトルおよび行列を要素で記述すると以下のとおりである。

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_{n_i-1} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_{11} & \cdots & x_{1,n_j} \\ \vdots & \ddots & \vdots \\ x_{n_i-1,1} & \cdots & x_{n_i-1,n_j} \end{pmatrix}$$

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{n_j} \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_{n_i-1} \end{pmatrix}$$

この取り扱いでは、発生源プロファイル行列の誤差を考慮せず、 $\xi_{ij} = x_{ij}$ とし、環境サイド化学種濃度測定誤差だけを考慮に入れる。

3.2 分散共分散行列

本研究では、分散共分散行列 $\boldsymbol{\sigma}$ が対角的と仮定した。すなわち、環境サイドにおける化学種 i の濃度の測定値 Z_i （単位:g/g）の不確かさとして与えられる分散を σ_i^2 として、

$$\boldsymbol{\sigma} = \text{diag}(\sigma_i^2) = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{n_i-1}^2 \end{pmatrix} \quad (2)$$

とした。

3.3 MCMC法の適用

3.3.1 柏木らの方法によるMCMC計算法

パラメータの乱数初期値として区間 [0,1] の一様乱数を用い、柏木らに記述された確率変数が従う切斷正規分布の平均値、分散および範囲（表1）から、その切斷正規分布に従う乱数を生成させ、この数を単一連鎖法で更新していく。表1において、平均値 μ_t , 分散 σ_t^2 であり、 a , b をその分布の下限および上限としている。

表1 発生源寄与 θ_j および未知発生源の組成 ω_i についての確率密度を与える切斷正規分布のパラメータの表式

θ_m	μ_t	$\frac{\{\mathbf{Z} - \mathbf{x}_m \boldsymbol{\theta}_m - \boldsymbol{\omega} (1 - \mathbf{1}_{n_j-1} \boldsymbol{\theta}_m)\}^t \boldsymbol{\sigma}^{-1} (\mathbf{x}_m - \boldsymbol{\omega})}{\{\mathbf{x}_m - \boldsymbol{\omega}\}^t \boldsymbol{\sigma}^{-1} (\mathbf{x}_m - \boldsymbol{\omega})}$
	σ_t^2	$\frac{1}{\{\mathbf{x}_m - \boldsymbol{\omega}\}^t \boldsymbol{\sigma}^{-1} (\mathbf{x}_m - \boldsymbol{\omega})}$
	a	0
	b	$1 - \mathbf{1}_{n_j-1} \boldsymbol{\theta}_m$
ω_m	μ_t	$\frac{1}{\tau_{mm}} \left\{ \frac{(\mathbf{Z} - \mathbf{x} \boldsymbol{\theta})^t \boldsymbol{\tau}_m - \boldsymbol{\omega}^t \boldsymbol{\tau}_m}{(1 - \mathbf{1}_{n_j} \boldsymbol{\theta})} \right\}$
	σ_t^2	$\frac{1}{\tau_{mm} (1 - \mathbf{1}_{n_j} \boldsymbol{\theta})^2}$
	a	0
	b	$1 - \mathbf{1}_{n_i-2} \boldsymbol{\omega}_m$

計算では、単一連鎖法によりそれぞれのパラメータを1個ずつ更新していくが、更新されるパラメータの添え字が表1における m である。1回の更新回ですべてのパラメータを1個ずつ更新して、それが終わった後に次の更新回に進む。これにより、乱数から出発したパラメータセットについて、マルコフ連鎖させ、定常分布を得る。

ここで、

$$\omega_{-i} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_{i-1} \\ \omega_{i+1} \\ \vdots \\ \omega_{n_i-1} \end{pmatrix}, \theta_{-j} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{j-1} \\ \theta_{j+1} \\ \vdots \\ \theta_{n_j} \end{pmatrix}, \tau_{ij} = (\sigma^{-1})_{ij}$$

$$\tau_i = \begin{pmatrix} \tau_{1i} \\ \vdots \\ \tau_{n_i-1,i} \end{pmatrix}, \tau_{-i} = \begin{pmatrix} \tau_{1i} \\ \vdots \\ \tau_{i-1,i} \\ \tau_{i+1,i} \\ \vdots \\ \tau_{n_i-1,i} \end{pmatrix}, x_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{n_i-1,j} \end{pmatrix}$$

$$x_{-j} = \begin{pmatrix} x_{11} & \cdots & x_{1,j-1} & x_{1,j+1} & \cdots & x_{1,n_j} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n_i-1,1} & \cdots & x_{n_i-1,j-1} & x_{n_i-1,j+1} & \cdots & x_{n_i-1,n_j} \end{pmatrix}$$

である。

3.3.2 乱数の処理

MCMC計算法における乱数の一般的な数値処理方法を記述する。

切断正規分布に従う乱数 x_t を求める方法は以下のとおりである。表1のパラメータの切断正規分布によって、 u を入力するパラメータ（単一連鎖法計算の初期値の場合は一様乱数、それ以降の場合は、連鎖の前段階の x_t ）、 ϕ を標準正規分布の累積分布関数として、次式によった。

$$x_t = \phi^{-1}[\phi(\alpha) + u\{\phi(\beta) - \phi(\alpha)\}]\sigma_t + \mu_t$$

$$\alpha = \frac{a - \mu_t}{\sigma_t}$$

$$\beta = \frac{b - \mu_t}{\sigma_t}$$

累積分布関数 $\phi(\alpha)$ の算出には、計算コードをすべてエクセルVBAで記述した場合はそのワークシート関数を用い、C言語で記述し、DLLに搭載した関数を用いる場合は、区間 $[-5\sigma_t + \mu_t, \alpha]$ を100等分して、一次元正規分布確率密度関数について、数値的に積分を求めた。

また、その逆関数 ϕ^{-1} は、次で定義される相補的累積分布関数の逆関数の有理近似公式¹¹⁾を用いて計算した。

まず、累積分布関数を $\phi(x)$ 、相補的累積分布関数を $\phi_c(x)$ とすると、

$$p = \phi(x)$$

$$\phi(x) + \phi_c(x) = 1$$

$$x = \phi^{-1}(p) = \phi_c^{-1}(1-p)$$

$$p = \phi(x) = \phi_c(-x)$$

$$x = \phi^{-1}(p) = -\phi_c^{-1}(p)$$

が成り立つ。

相補的累積分布関数 $p = \phi_c(x_p)$ について、 $0 < p \leq 0.5$ のとき、その逆関数 ϕ_c^{-1} は、近似的に有理関数で、

$$x_p = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \epsilon(p)$$

$$t = \sqrt{\log\left(\frac{1}{p^2}\right)}$$

$$c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328$$

$$d_1 = 1.432788, d_2 = 0.189269, d_3 = 0.001308$$

$$|\epsilon(p)| < 4.5 \times 10^{-4}$$

と表せて、 $p > 0.5$ のときは、 $p' = 1 - p$ から計算できる。

以上から、累積分布関数の逆関数は、相補的累積分布関数の逆関数を用いて次式で定義される。

$$x_t = \phi^{-1}(p) = \begin{cases} -\phi_c^{-1}(p) & 0 < p \leq 0.5 \\ \phi_c^{-1}(1-p) & 0.5 < p < 1 \end{cases}$$

3.3.3 累積分布関数の逆関数演算における数値処理と多重連鎖法による計算

本研究では、累積分布関数の逆関数計算の安定性を確保するため、切断正規分布の厳密な範囲 $[a, b]$ から、計算機イプシロンとして与える $\epsilon = 1 \times 10^{-10}$ だけ範囲の両端を狭め、出力パラメータの範囲を $[a + \epsilon, b - \epsilon]$ として、この範囲に入らないパラメータは、それぞれ、その下限 $a + \epsilon$ あるいは上限 $b - \epsilon$ の値とした。

また、単一連鎖法により定常分布を得る計算を、異なった乱数初期値から行う多重連鎖法と組み合わせ、乱数初期値に依存しない、確率変数としてのパラメータの期待値を取得した。

なお、単一連鎖法においては、初期の更新回数では、乱数初期値の影響が大きいため、確率変数の期待値を求める際に、最初の数十から数千回の更新結果を無視する。これを burn-in と呼ぶ。

3.4 計算で得られる寄与などの誤差

いま、確率変数として扱うパラメータ θ, ω を一般的に Λ_m と記述し、乱数初期値を出発点として、 r

回のMCMC計算を行い、 n_0 回までのburn-inの後で得られたパラメータを $\lambda_m(r)$ とすると、モンテカルロ積分の定義から、期待値および誤差分散として、次の(3)式を得る。

$$\left. \begin{aligned} E(\Lambda_m) &\cong \frac{1}{n_r - n_0} \sum_{r=n_0+1}^{n_r} \lambda_m(r) \\ \text{var}(\Lambda_m) &\cong \frac{1}{n_r - n_0} \sum_{r=n_0+1}^{n_r} \lambda_m^2(r) - E^2(\Lambda_m) \end{aligned} \right\} \quad (3)$$

以下に、本研究における発生源寄与の誤差推定を記述する。

再現される環境サイド化学種濃度の誤差は、発生源プロファイルが確率変数ではなく、寄与が確率変数であるから、単純に発生源プロファイルの重みを、その2乗として乗じ、絶対誤差の伝播により計算する。さらに、未知発生源の組成が確率変数で、その寄与は既知の発生源寄与としての確率変数の和を1から減じた数であり、これらの積が環境サイド化学種濃度に加わる。確率変数の和の誤差は絶対誤差の伝播で、確率変数と確率変数の積の誤差は相対誤差の伝播によることから、この項の誤差分散を求めることができる。

以上から、環境サイド化学種濃度の誤差分散は次式で表せる。

$$\begin{aligned} \text{var}(\zeta_i) &= \sum_{j=1}^{n_j} \xi_{ij}^2 \text{var}(\theta_j) \\ &+ \left\{ \frac{\text{var}(\omega_i)}{\omega_i^2} + \frac{\sum_{j=1}^{n_j} \text{var}(\theta_j)}{\left(1 - \sum_{j=1}^{n_j} \theta_j\right)^2} \right\} \left\{ \omega_i \left(1 - \sum_{j=1}^{n_j} \theta_j\right) \right\}^2 \end{aligned} \quad (4)$$

環境サイド化学種濃度の再現率(C/M)は、実測値に対する、計算された環境サイド化学種濃度の比である。実測値の誤差分散 σ_i^2 およびCMBにより再現された濃度についての(4)式の誤差分散から、再現率の誤差分散が相対誤差の伝播により計算できる。すなわち、

$$\text{var}\{(C/M)_i\} = \left\{ \frac{\text{var}(\zeta_i)}{\zeta_i^2} + \frac{\sigma_i^2}{Z_i^2} \right\} (C/M)_i^2$$

である。

4. PM2.5解析例

表2に青森県弘前市文京小学校で2019年5月8日に採取されたPM2.5の無機成分のうち、金属元素についてのCMB解析についての発生源プロファイル、環境サイドデータおよびCMB解析結果を示した。この場合の発生源プロファイルは、環境省¹²⁾および東京都¹³⁾の調査集計結果を引用した。

表2で、発生源プロファイルはmg/kg単位で、測定結果は $\mu\text{g}/\text{m}^3$ で記述しているが、計算の内部コードで、双方g/g単位に換算した。また、表2における環境サイド化学種濃度の標準偏差は、測定値の不確かさとしての定量下限を10で除した濃度により推定した。

CMB計算は、以下の5手法により試行した。それらは「ベイズ統計手法」、「EVLSQ 全発生源(有効分散最小二乗法)^{2),3)}」、「EVLSQ 負の寄与の発生源除去^{2),3)}」、「非負拘束EVLSQ⁴⁾」および「非線形最適化法⁵⁾」である。

ベイズ統計手法のCMB計算条件は、burn-in回数を800回、単一連鎖法回数を1000回、多重連鎖法回数を10回とした。なお、柏木ら⁶⁾は、単一連鎖法の計算回数を400万回取っているが、本研究の方法において表2のデータに応用する限りでは、その回数を1000回と取って、十分な精度(区間[0,1]の寄与に対して、 1×10^{-7} 未満の標準偏差)が得られた。

また、板野ほか¹⁴⁾が記述しているように、マルコフ連鎖では、更新された連鎖の乱数が、その直前の乱数に影響されるため、10回に1回の乱数データのみを、確率変数の期待値を求める際に使用した。

なお、CMB計算には、アンモニウムイオン、硝酸イオン、硫酸イオンおよび塩化物イオンは加えなかったが、二次生成粒子の影響を考慮するため、寄与として、結果の表に含めている。

この計算結果の表では、「ベイズ統計手法」以外は、発生源寄与の和を取り(上述のアンモニウムイオンほか4種の寄与を含む)、1から減じた値を未知発生源寄与とした。なお、発生源寄与の和が1以上の場合には、未知発生源寄与は0とした。

5. 考察

5.1 発生源寄与の与え方

今回検討した計算手法について、表2から、全発生源を用いた「EVLSQ 全発生源」では負の寄与が得られており、これを回避すべき改良した計算法の適用が期待された。すなわち、物理的に意味のある発生源寄与の解を求めるためには、負の寄与を与えないベイズ統計手法あるいは負の寄与の発生源を除去するEVLSQ法、非負拘束EVLSQ法、非線形計画化法の適用が必要となった。

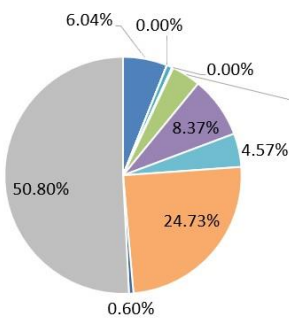
表2では、上記のベイズ統計手法以外で、かつ「EVLSQ 負の寄与の発生源除去」以外の方法では、未知発生源の寄与は0になった。これは、これらの方法で、高い寄与を与える発生源が存在し、寄与の和が1を超えるからである。これらの手法では、発生源寄与の下限とし

表2 (a) 青森県内のPM2.5測定結果解析の発生源プロフィールおよび環境サイド濃度

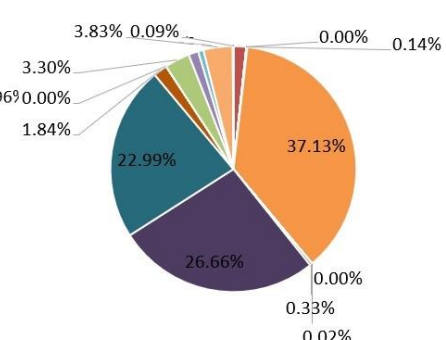
	発生源(mg/kg)															環境	
	道路粉塵	海塩粒子	鉄鋼工業	石油燃焼	廃棄物焼却	自動車排気	タイヤ&ブレーキ粉塵	ガス機関	乾燥炉木材	建設機械油圧ショベル	船舶補助発電ディーゼル	ディーゼル機関	ディーゼル車未規制車	植物質燃焼	黄砂	(μg/m ³)	標準偏差
	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値	実測値		
Na	12500	304000	13600	10000	120000	76.4	9200	7220	102188	0	339	740	122	6550	7230	0.457	0.0014
Al	61100	0.29	9990	2100	4200	1570	49000	6137	309	10000	268	0	0	63200	75300	0.0974	0.00098
K	12700	11000	13200	850	200000	197	8060	3321	71563	528	179	429	94	370	22100	0.0963	0.0002
Ca	55200	11700	45100	850	11000	1460	12000	79422	7594	0	607	11429	725	415	30400	0.073	0.0013
Sc	13.3	0.0012	1.32	0.09	0.46	0.119	2.3	0	0	0	0	0	0	0	16	0.00002	2.2E-06
Ti	5050	0.029	1000	740	900	14.6	3000	0	0	0	0	0	0	0	4150	0.0059	0.000067
V	108	0.058	125	6380	27	7.25	110	0	4	25	84	89	0	0	111	0.00107	5.4E-06
Cr	279	0.0015	3160	210	850	11.6	340	722	75	87	45	11	522	0	98	0.00031	0.000048
Mn	1060	0.058	22000	120	330	19.3	1410	542	84	377	0	5	0	10	826	0.00308	6.6E-06
Fe	53100	0.29	157000	4600	6100	989	76600	32491	1156	0	0	314	2672	100	45100	0.077	0.0014
Co	21.7	0.015	44	31	21	1.47	20	1011	1	0	0	1	8	0	20	3.03E-05	1.4E-07
Ni	80.3	0.015	2900	3970	0	9.89	260	0	29	0	0	37	0	0	55	0.0004	0.00013
Cu	268	0.017	3700	1	3600	113	8700	0	106	0	0	5	0	0	113	0.000742	8.1E-06
As	11.3	0.029	103.6	23	150	3.69	100	0	25	0	0	0	0	0	22	0.00195	2.5E-06
Se	1.43	0.12	51.1	48	0	1.67	20	0	0	0	0	0	0	0	3	0.000673	2.2E-06
Rb	55.6	3.5	76.8	2	260	0.49	34	0	128	0	0	0	0	0	112	0.000321	7.1E-07
Mo	8.55	0	0	0	0	5.91	512	0	0	0	0	0	0	0	12	0.000177	1.2E-06
Sb	13	0.014	90	6.9	952	19.6	4200	0	0	0	0	0	0	0	15	0.000304	1.2E-06
Cs	3.59	0.029	3	0.1	12	0.0745	3.7	0	0	0	0	0	0	0	11	3.62E-05	2.8E-07
Ba	499	0.86	500	1000	390	98.9	18800	0	94	0	0	2	0	0	609	0.00116	4.8E-06
La	31.3	0.009	9.75	40	7.7	0.341	7.5	0	0	0	0	0	0	0	62	5.64E-05	5.2E-07
Ce	59.9	0.012	68.7	70	170	1.08	24	0	0	0	0	0	0	0	108	0.000107	1.2E-06
Sm	5.93	0.001	0.212	7.6	0.46	0.0304	0.43	0	0	0	0	0	0	0	8	7.6E-06	5.2E-07
W	11.6	0.003	47.4	20	100	0.686	9.6	137	0	0	0	0	0	0	3	0.000044	1.8E-06
															NH ₄	1.08	0.002
															NO ₃	0.59	0.013
															SO ₄	3.19	0.0022
															Cl	0.078	0.0017
															質量濃度	12.9	

(b) CMB計算結果 (数値は寄与 (無次元))

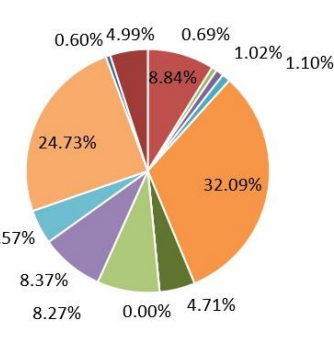
CMB解析手法	道路粉塵	海塩粒子	鉄鋼工業	石油燃焼	廃棄物焼却	自動車排気	タイヤ&ブレーキ粉塵	ガス機関	乾燥炉木材	建設機械油圧ショベル	船舶補助発電ディーゼル	ディーゼル機関	ディーゼル車未規制車	植物質燃焼	黄砂	NH ₄	NO ₃	SO ₄	Cl	未知発生源
ベイズ統計手法	0.060	0.000	0.000	0.000	0.007	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.084	0.046	0.247	0.006	0.508
EVLSQ 全発生源	-0.155	0.109	-0.021	-0.018	0.009	2.400	-0.009	0.001	0.021	1.723	1.486	0.119	-0.214	-0.315	0.213	0.084	0.046	0.247	0.006	0.000
EVLSQ 負の寄与の発生源除去	0.088	0.007	0.010	0.011	0.321	0.047	0.047	0.000	0.047	0.660	0.625	0.401	0.000	0.001	0.113	0.084	0.046	0.247	0.006	0.050
非負拘束EVLSQ	0.000	0.088	0.006	0.000	0.010	0.406	0.000	0.000	0.047	0.063	0.781	0.000	0.000	0.000	0.088	0.084	0.046	0.247	0.006	0.000
非線形最適化法	0.011	0.114	0.000	0.002	0.014	1.038	0.001	0.002	0.079	0.660	0.625	0.401	0.000	0.001	0.113	0.084	0.046	0.247	0.006	0.000



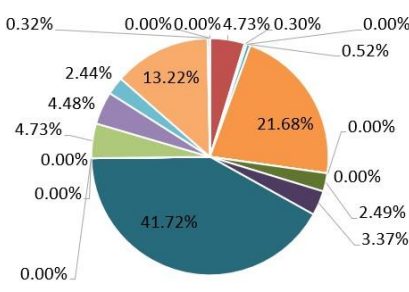
(a) ベイズ統計手法



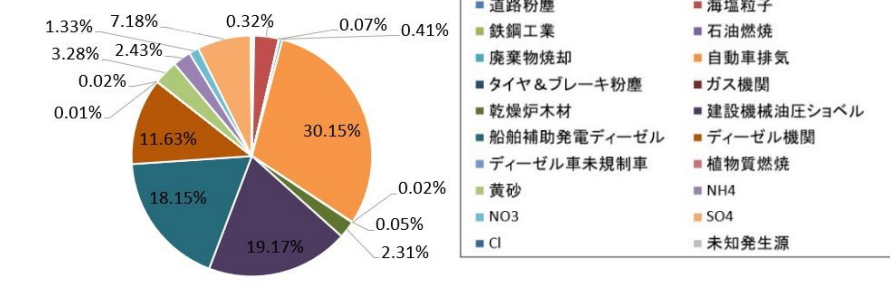
(b) EVLSQ全発生源



(c) EVLSQ負の寄与の発生源除去



(d) 非負拘束EVLSQ



(e) 非線形最適化法

図1 青森県内のPM2.5測定結果 (表2) の各種CMB計算法での計算結果比較。円グラフは寄与率。

て0を取るが、その上限は制限していない。一方、ベイズ統計手法では、寄与の和に上限 (未知発生源寄与を加えて和が1) を設けて、未知発生源の寄与と組成の推定を組

み入れている。このことが、上述の結果の違いを与えたと考えられた。

比率モデルでは、発生源寄与が区間[0,1]に入り、発生

源から希釈こそされ、濃縮されずに環境サイドに到達するという合理的な物理的解釈が可能である。また、発生源サイドから環境サイドへの質量均衡を考慮するためには、柏木らの方法のように、未知発生源の存在を考慮して、発生源寄与に上限を設けるべきである。これを考慮できる点で、柏木らの手法は、優れているといえる。

図1では、負の寄与の発生源を除去するEVLSQ法、非負拘束EVLSQ法および非線形計画化法の3法では、円グラフの傾向が類似した。一方、ベイズ統計手法では、それ以外の方法の結果とは異なった傾向であった。この原因は、ベイズ統計手法が他の方法と比較し、考え方がまったく異なる計算を行っているからと考えられた。

5.2 C言語によるDLLに搭載した関数を用いた場合の有利な点

表2のベイズ統計手法の計算におけるMCMC計算の所要時間は、C言語で記述したDLLに関数を搭載し、これをエクセルVBAから呼び出す方法では2.74秒、すべてエクセルVBAでコードした場合には169秒を要した。よって、C言語DLLの使用により、計算時間を60倍以上高速化することが可能であった。

6. 結論

ベイズ統計手法によるCMB法は、次の点で長所がある。

- 1) 比率モデルによる計算方法であり、物理的に明解な意味を持つ発生源寄与を与える。
- 2) その発生源寄与は負になることはない。また、寄与の上限についても設定している。
- 3) 未知発生源の存在を明確に考慮したモデルによっており、その寄与および組成がCMB計算で与えられる。
- 4) 多数回の繰り返し計算を行うが、C言語のDLL関数を呼び出すことで、その計算を高速化できる。

今回の数値計算手法の検討により、これまでのMCMC法と比較して、非常に高速に計算結果が得られた。よって、この方法は時系列データのような大量のデータ処理に耐えうる方法と考えられた。

以上から、本研究のベイズ統計手法によるCMB法が有用であることが示された。本研究で開発したエクセルマクロは、解析・集計で得られたデータ表から直接CMB計算を行える有利さを備えている。

なお、現在、青森県環境保健センターでは、本研究で開発したマクロの時系列データへの応用を検討中である。

7. 引用文献

- 1) 橋本俊次, 柏木宣久: 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
- 2) USEPA CHEMICAL MASS BALANCE MODEL: EPA-CMB8.2
https://www3.epa.gov/scram001/receptor_cmb.htm
(2021.7.26 アクセス)
- 3) Watson J. G., Cooper J., Huntzicker J.: The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atom. Environ.*, **18**, 1347-1355, 1984
- 4) 花石竜治, 佐藤裕久: 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
- 5) 花石竜治: 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. 青森県環境保健センター年報, **29**, 53-58, 2018
- 6) 柏木宣久, 吉澤 正, 茨木 剛, 加藤謙一, 橋本俊次, 佐々木裕子: 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) 統計数理研究所: 学術研究リポジトリ,
https://ismrepo.ism.ac.jp/?action=repository_uri&item_id=32744 (2021.7.26 アクセス)
- 8) 早狩 進, 花石竜治: 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
- 9) 中妻照雄: 入門 ベイズ統計学, 朝倉書店, 東京, 2007
- 10) 柏木宣久, 佐々木裕子, 飯村文成, 安藤晴夫: 関数関係解析によるChemical Mass Balance. 応用統計学, **31**, 59-74, 2002
- 11) Zelen M., Severo N. C.: Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. Abramowitz M., Stegun I. A. Eds., Dover, New York, 1965
- 12) 平成20年度環境省請負業務結果報告書 平成20年度微小粒子状物質等実測調査 データ解析(発生源寄与濃度の推計) 報告書, ムラタ計測器サービス株式会社, 2009
- 13) 東京都環境局: 微小粒子状物質(PM2.5)対策について,
https://www.kankyo.metro.tokyo.lg.jp/air/air_pollution/torikumi/pm2_5/pm2_5.html
(2021.7.26 アクセス)
- 14) 板野泰之, 高倉耕一: ベイズ統計手法による都市大気オゾンの日内変動と季節変動の分離評価. 大気環境学会誌, **46**, 179-186, 2011