

## 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

ケミカルマスバランス(CMB)法は、レセプターモデルに基づく環境汚染解析の一手法である。柏木(2004)が確立した CMB 法のアルゴリズムはベイズ統計手法 (以下、「ベイズ法」と言う。)を採用している。本研究では、この数値計算法を改良して、表計算ソフトのマクロとして記述した。ベイズ法の特徴は、モデル式に未知発生源の寄与と組成を組み入れて評価できることである。また、ベイズ法は非負の発生源寄与のみを与え、さらに、組成モデルを用いているため、従来法よりも、物理的に解釈が容易な発生源寄与を与える。発生源寄与および未知発生源の寄与と組成を仮定してシミュレーションを行ったところ、これらが概ね良好に再現できた。また、実測データへの応用では、他の CMB 法よりも、環境汚染の観測点の実測値を正確に再現できた。以上から、このベイズ法が発生源解析に有用であることが分かった。

Key words: Chemical mass balance, Bayesian statistics, Markov chain Monte Carlo method, Unknown source

### 1. はじめに

環境汚染の削減のためには、寄与の大きい発生源を特定し、そこからの排出量を削減することが、もっとも効果の高い手法である。このため、採取した環境試料の化学分析の結果を解析し、発生源を推定することが、環境行政にとって重要な業務となる。

現代の発生源解析には、大きく分けて、Chemical Mass Balance(ケミカルマスバランス、CMB)法と、Positive Matrix Factorization(PMF)法<sup>1)</sup>の二つの流れがある。

その両方が仮定している前提は、レセプターモデルであり、発生源からの汚染物質が、環境汚染の観測点まで、化学種として組成を変えずにその線形和をなして到達するという、重ね合わせの原理(Superposition principle)が成り立つことである。よって、レセプターモデルの適用に適した化学種は、金属元素や難分解性有機化合物のように、移送過程で変質せず、かつ指標となるものである。

CMB 法では、発生源からの汚染を、既知の発生源プロファイルとし、環境汚染の観測点の化学種濃度から、発生源の寄与を推定する。PMF 法では、環境汚染の観測点の化学種濃度の複数のデータセットから、因子分析の考えにより、発生源のプロファイルおよび寄与の推定を行う。

本報告で取り上げる CMB 法は、1970 年代に提唱され<sup>2)</sup>、以来、半世紀にわたって、計算方法やア

ルゴリズムに改良が加えられている<sup>3),4)</sup>。この中でもっとも多く流布しているアプリケーションは、米国環境保護庁(USEPA)の CMB8<sup>5),6)</sup>である。CMB8 は、環境汚染の観測点の化学種濃度の誤差と、発生源プロファイルの化学種濃度の誤差の両方を考慮する「有効分散最小二乗法」(EVLSQ)を用いている。

さて、最小二乗法(LSQ)計算の欠点は、発生源プロファイルに組成の類似した、すなわち多重共線性の強い組み合わせを設定した場合、負の発生源寄与を与える可能性があることである。この負の寄与は物理的に意味がないため、これまでの CMB 法の開発研究においては、負の寄与を回避することに力点が置かれてきた。

USEPA の CMB8<sup>5),6)</sup>には、「負の寄与を与える発生源除去」機能があるが、発生源の取捨選択には恣意が入る可能性があるため、本質的に非負の寄与を与える方法が望まれてきた。

このような中で、著者ほか<sup>7)</sup>は、非負寄与を与える CMB 法として、従来の EVLSQ 法を、非負条件で解く非負拘束最小二乗法(NN-EVLSQ)を提案した。さらに著者<sup>8)</sup>は、発生源寄与についての最尤法を、非線形最適化法の一法である主双対内点法で解く方法(MLM)を提案した。

非負の発生源寄与を与える CMB 法のもう一つの流れは、ベイズ法の応用である。ベイズ法は、例えば CMB 法においては、寄与などのパラメー

タを確率変数と見なし、この確率分布をベイズ更新と呼ばれる方法で求め、パラメータの期待値などを算出する。Billheimer<sup>9)</sup> は、比率モデルを採用してこの方法を CMB 法に応用し、続いて、本邦の柏木<sup>4)</sup> および柏木ほか<sup>10)</sup> が計算手法を確立した。ベイズ法の応用は、その後、Park ほか<sup>11)</sup>、Keats ほか<sup>12)</sup>、および Massoudieh ほか<sup>13)</sup> により、それぞれ研究がなされた。

柏木らのベイズ法 CMB の特徴は、未知発生源の存在を仮定して、その寄与と組成を求められることである。柏木らは、この計算法を用いてスタンドアロンで動作する実行可能アプリケーション<sup>14)</sup> を開発した。

一方、早狩および著者は、USEPA の CMB8 を、データの取り扱いで多用されている表計算ソフトのマイクロソフト社 Excel の VBA マクロに移植して公開してきた経緯<sup>15)</sup> がある。利用者は、このマクロ機能を用い、独自の入出力設定を行わず、データを含んだ Excel ワークシートから直接 CMB 計算を行うことができる。

そこで著者は、今回、柏木らの CMB 法について、まず、数値計算を高速化するための計算方法を検討した。その上で、計算に時間がかかるマルコフ連鎖モンテカルロ(MCMC)計算を行う C 言語で記述した DLL を作成し、これを Excel VBA でコードした主プログラムから呼び出すマクロとした。著者らの CMB マクロは、これまで最小二乗法などの従来の計算法での解析が可能であったが、今回、ベイズ法も使用可能にした。本報告では、その計算手法の検討<sup>16)</sup> および試用結果を述べる。

## 2. 理論

### 2.1 CMB モデル式

柏木ら<sup>4)</sup> により、CMB モデル式を比率モデル<sup>9),17)</sup> 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、

$\zeta_i$  : 環境汚染の観測点の化学種濃度 (単位:g/g)

$\xi_{ij}$  : 発生源プロファイルの行列 (単位:g/g)

$\theta_j$  : 寄与 (無次元)

$\omega_i$  : 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。ここで、柏木<sup>4)</sup> 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、 $n_i - 1$  個となり、モデル式には  $n_i - 1$  個の化学種を既述すればよい。

しかし、本研究では、独立な  $n'_i$  個の化学種に着目し、もう 1 個の化学種の存在を考えて、濃度の総和が 1 g/g となるように考える。すなわち、 $n'_i$  個の化学種の濃度測定による完全な組成把握を前提とせず、 $n'_i + 1$  個の化学種濃度の和が 1 g/g になることを想定する。数式で表すと、

$$n_i = n'_i + 1$$

である。そのため、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、環境汚染の観測点の化学種濃度  $\zeta_i$  と発生源プロファイル濃度測定値  $\xi_{ij}$  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  $\theta_j$  および未知発生源の組成  $\omega_i$  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。

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_{n_i-1} \end{pmatrix}$$

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

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{n_j} \end{pmatrix}$$

$$\boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_{n_i-1} \end{pmatrix}$$

本研究の取り扱いでは、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、環境汚染の観測点の化学種濃度測定誤差だけを考慮する。

## 2.2 分散共分散行列

柏木らの扱い<sup>4)</sup>では、分散共分散行列を設定するために、多項分布を用いる方法や、環境汚染の観測点の化学種濃度測定値が切断正規分布、対数正規分布、ガンマ分布の3種の確率分布のいずれかに従うと仮定し、それらの確率分布のモーメント方程式から平均と分散を求め、確率分布に従う乱数を生成させ、これにより標本共分散を求める方法を採用している。

本研究では、単純に、分散共分散行列  $\boldsymbol{\sigma}$  が対角的と仮定する。すなわち、環境汚染の観測点における化学種  $i$  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  $\sigma_i^2$  として、

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

とする。

## 2.3 尤度

(1)式のモデル式が与える尤度は、柏木ら<sup>4)</sup>により記述されており、分散共分散行列を(2)式に取った場合、記述し直すと、次式(3)のとおりとなる。

$$l(\boldsymbol{\omega}, \boldsymbol{\theta}) = (2\pi)^{-(n_i-1)/2} |\boldsymbol{\sigma}|^{-1/2} \cdot \exp \left[ -\frac{1}{2} \left\{ \mathbf{Z} - \mathbf{x}\boldsymbol{\theta} - \boldsymbol{\omega} (1 - \mathbf{1}_{n_j}\boldsymbol{\theta}) \right\}^t \boldsymbol{\sigma}^{-1} \left\{ \mathbf{Z} - \mathbf{x}\boldsymbol{\theta} - \boldsymbol{\omega} (1 - \mathbf{1}_{n_j}\boldsymbol{\theta}) \right\} \right] \quad (3)$$

ここで、

$$\mathbf{1}_n = \overbrace{(1, \dots, 1)}^n$$

である。また、 $t$  は転置を表す。

## 2.4 MCMC 法の適用

### (1) 確率変数としてのパラメータの期待値と分散の計算<sup>4)</sup>

計算上のパラメータセット  $\boldsymbol{\lambda}$  (ここでは、 $\theta_j$  および  $\omega_i$ ) をなすパラメータ  $\lambda_m$  に対応する確率変数を  $\Lambda_m$  と表記する。

いま、MCMC 法により、乱数を初期値として、事前分布から事後分布への更新がベイズの定理により  $n_r$  回行われたとすると、確率変数  $\Lambda_m$  の期待値  $E(\Lambda_m)$  および分散  $\text{var}(\Lambda_m)$  は、モンテカルロ積分により次式(4)で与えられる。

$$\left. \begin{aligned} E(\Lambda_m) &\cong \frac{1}{n_r} \sum_{r=1}^{n_r} \lambda_m(r) \\ \text{var}(\Lambda_m) &\cong \frac{1}{n_r} \sum_{r=1}^{n_r} \lambda_m^2(r) - E^2(\Lambda_m) \end{aligned} \right\} \quad (4)$$

ここで、 $\lambda_m(r)$  は、 $r$  回目の更新で得られた、計算上のパラメータ  $\lambda_m$  を示す。

### (2) 柏木の方法による MCMC 法

ベイズ法の計算の原理は、パラメータの初期値として乱数を与えて出発し、(3)式で与えられる尤度により、事前分布を事後分布に更新するベイズの定理を利用して、MCMC 法によりパラメータを更新していくことである。この更新では、ギブズサンプラーを用いる場合、マルコフ連鎖における遷移核<sup>18)</sup>としての完全条件付き密度を使用する。

柏木は、その切断正規分布、すなわちパラメータが確率変数として従う完全条件付き密度のパラメータの表式を導出した<sup>4), 10)</sup>。彼は、single-component Metropolis-Hasting (SCMH) アルゴリズムで採択率 1 としたが、これはギブズサンプラーを用いたためである<sup>18)</sup>。

本研究では、パラメータの初期値として区間 [0,1] の一様乱数を用い、柏木ら<sup>4), 10)</sup>が記述した確率変数  $\Lambda_m$  が従う切断正規分布の平均値および分散の式から、その切断正規分布に従う乱数を生成させ、この数を更新していく。これを単一連鎖法と呼ぶ。

表 1 に、柏木らの、発生源寄与  $\theta_m$  および未知発生源の組成  $\omega_m$  が従う切断正規分布のパラメ

ータを示した。ここで、期待値  $\mu_t$ 、分散  $\sigma_t^2$  であり、 $a$ 、 $b$  はそれぞれその分布の下限および上限である。

単一連鎖法における更新回で、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、パラメータセットのすべて完了し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。

表 1 発生源寄与  $\theta_m$  および未知発生源の組成  $\omega_m$  についての確率密度を与える切断正規分布のパラメータの表式

$\omega_m$	$\mu_t$	$\frac{1}{\tau_{mm}} \left\{ \frac{(Z - x\theta)^t \tau_m}{(1 - \mathbf{1}_{n_j} \theta)} - \omega_{-m}^t \tau_{-m} \right\}$
$\sigma_t^2$		$\frac{1}{\tau_{mm} (1 - \mathbf{1}_{n_j} \theta)^2}$
$a$		0
$b$		$1 - \mathbf{1}_{n_{i-2}} \omega_{-m}$
$\theta_m$	$\mu_t$	$\frac{\{Z - x_m \theta_{-m} - \omega (1 - \mathbf{1}_{n_{j-1}} \theta_{-m})\}^t \sigma^{-1}(x_m - \omega)}{\{x_m - \omega\}^t \sigma^{-1}(x_m - \omega)}$
$\sigma_t^2$		$\frac{1}{\{x_m - \omega\}^t \sigma^{-1}(x_m - \omega)}$
$a$		0
$b$		$1 - \mathbf{1}_{n_{j-1}} \theta_{-m}$

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。

表1に示したように、発生源寄与  $\theta_m$  の統計分布の上限を設定することで、これまでのCMB法とは異なり、発生源寄与を過剰評価しないようにしている。

### (3) 乱数の処理

切断正規分布に従う乱数  $x_t$  を求める方法は、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、連鎖の前段階で得られたパラメータ）、 $\phi$  を標準正規分布の累積分布関数として、次式(5)による。

$$x_t = \phi^{-1}[\phi(\alpha) + u\{\phi(\beta) - \phi(\alpha)\}]\sigma_t + \mu_t \quad (5)$$

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

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

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を、台形公式を用いて数値的に積分して行う。

また、その逆関数  $\phi^{-1}$  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<sup>19)</sup>を用いて計算する。

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、累積分布関数の値  $p$  について、

$$p = \phi(x)$$

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

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

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

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

が成り立つ。

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

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

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

$$c_0 = 2.515517$$

$$c_1 = 0.802853$$

$$c_2 = 0.010328$$

$$d_1 = 1.432788$$

$$d_2 = 0.189269$$

$$d_3 = 0.001308$$

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

と表せる<sup>19)</sup>。 $p_c > 0.5$  の場合、 $p'_c = 1 - p_c$  により計算できる。

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表される。

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

#### (4) MCMC 計算における数値処理

本研究では、前項の累積分布関数の逆関数の近似計算の安定性を確保するため、切断正規分布の範囲  $[a, b]$  に入らない出力パラメータ  $x_t < a$  あるいは  $x_t > b$  は、それぞれ、その下限  $a$  あるいは上限  $b$  の値に置き換えて、MCMC 計算を進める。

#### (5) モンテカルロ積分の数値計算

以上の方法で、初期値の一様乱数としてのパラメータセットから開始し、切断正規分布により新たなパラメータセットを生成させ、次の更新回で、このパラメータセットから切断正規分布に従うパラメータセットを新たに得る。単一連鎖法により、マルコフ連鎖が定常分布に達するまで更新を行う。

また、異なった初期乱数のセット ( $n_M$  回) について単一連鎖法を行い、パラメータの確率分布を

求め、期待値および分散を得る。これを多重連鎖法と呼び、これにより、初期値に依存しない確率変数としてのパラメータセットの期待値および分散が得られる。

なお、単一連鎖法において、初期の更新回数では、乱数初期値の影響が大きいため、モンテカルロ積分を求める際に、最初の一定回数 ( $n_0$  回) の結果を無視する。このことを burn-in と呼ぶ<sup>4),10),18)</sup>。

以上から、(4)式で表したモンテカルロ積分は、 $r_M$  回目の多重連鎖法における過程を示して、乱数値を  $\lambda_m(r_M, r)$  と表記し、次の(6)式で書き換えられる。

$$\left. \begin{aligned} E(\Lambda_m) &\cong \frac{1}{n_M(n_r - n_0)} \sum_{r_M=1}^{n_M} \sum_{r=n_0+1}^{n_r} \lambda_m(r_M, r), \\ \text{var}(\Lambda_m) &\cong \frac{1}{n_M(n_r - n_0)} \sum_{r_M=1}^{n_M} \sum_{r=n_0+1}^{n_r} \lambda_m^2(r_M, r) - E^2(\Lambda_m) \end{aligned} \right\} \quad (6)$$

以下に、MCMC 計算のアルゴリズムを記述する。

1) 確率変数のパラメータ  $\omega_{m'} (m' = 1, \dots, n_i - 1)$ 、 $\theta_{m''} (m'' = 1, \dots, n_j)$  の初期値を、区間  $[0, 1]$  の一様乱数に取る。いま、

$$m = \begin{cases} m' & (m' = 1, \dots, n_i - 1) \\ m'' + n_i - 1 & (m'' = 1, \dots, n_j) \end{cases}$$

とすると、 $m = 1, \dots, n_i + n_j - 1$  について、 $\lambda_m(1)$  は、範囲  $[0, 1]$  の一様乱数である。

2) MCMC 計算のループ

①  $r = 1$  とする。

②  $m = 1, \dots, n_i + n_j - 1$  について、以下の③、④の計算を行う。

③  $\lambda_m(r)$  から、切断正規分布の上限  $b$ 、期待値  $\mu_t$ 、分散  $\sigma_t^2$  を求める。

④ パラメータの初期値を入力値  $u = \lambda_m(r)$  として、③で規定される切断正規分布 (表 1) に従う乱数の出力値  $x_t$  を(5)式を用いて求める。これにより、 $\lambda_m(r+1) = x_t$  とする。

⑤  $r \leftarrow r+1$  ( $r \leq n_r$ ) として、②から計算を続ける。

### 2.5 寄与、未知発生源の組成および再現される環境汚染の観測点の化学種濃度の誤差

寄与  $\theta_j$  および未知発生源の組成  $\omega_i$  の誤差分散は(6)式に含まれる。

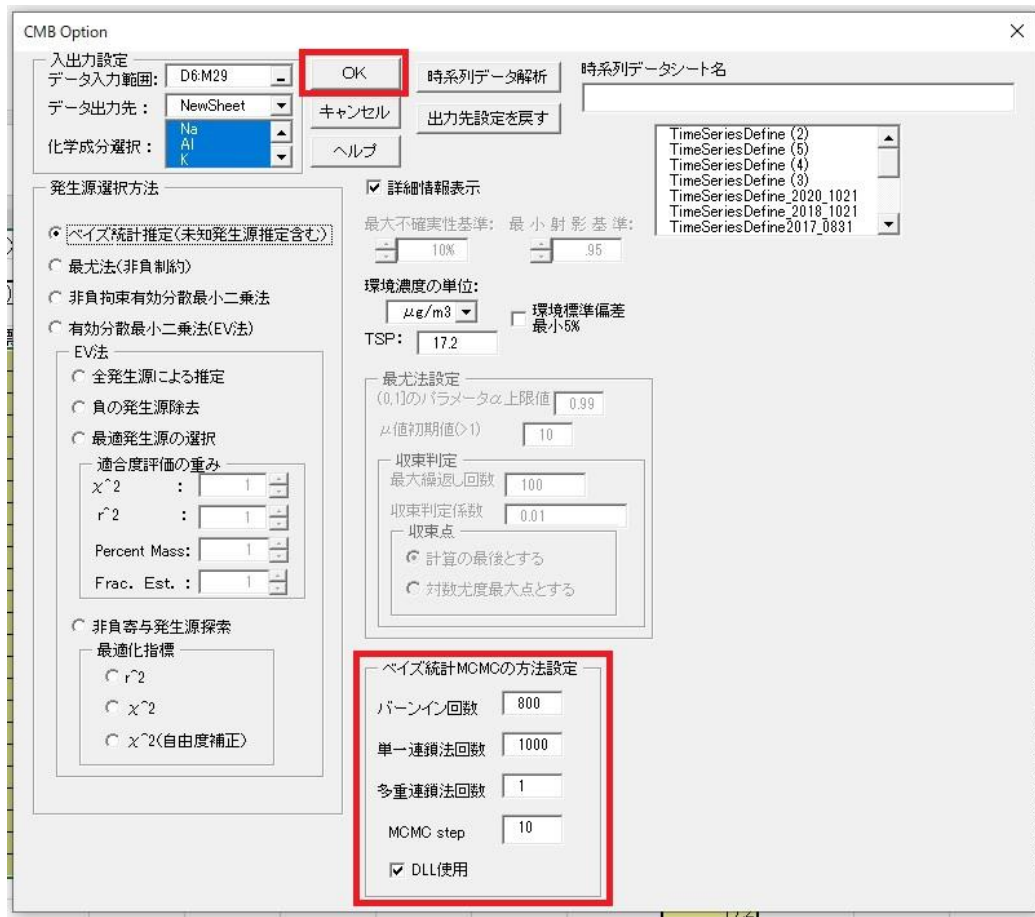


図 1 CMB マクロで「ベイズ統計推定 (未知発生源推定含む)」を選択したときのダイアログ。

再現される環境汚染の観測点の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、これらの積が環境汚染の観測点の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、この項の誤差分散が求まる。

以上から、環境汚染の観測点の化学種濃度の誤差分散は次式で表せる。

$$\text{var}(\zeta_i) = \sum_{j=1}^{n_j} \xi_{ij}^2 \text{var}(\theta_j) + \left\{ \frac{\text{var}(\omega_i)}{\omega_i^2} + \frac{\sum_{j=1}^{n_j} \text{var}(\theta_j)}{(1 - \sum_{j=1}^{n_j} \theta_j)^2} \right\} \left\{ \omega_i \left( 1 - \sum_{j=1}^{n_j} \theta_j \right) \right\}^2 \quad (7)$$

環境汚染の観測点の化学種濃度の再現率(C/M)は、その実測値に対する、計算された濃度の比である。その誤差分散は、誤差分散  $\sigma_i^2$  および CMB により再現された濃度についての(7)式の誤差分散から、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。

### 3. CMB マクロについて

図 1 に、CMB マクロを起動し、「ベイズ統計推定 (未知発生源推定含む)」を選択したときに現れる、利用者へ入力を促すダイアログを示す。

図 1 の時系列データ解析については、本報文中では扱わないこととし、稿を新たに記述したい。

使用者は MCMC 計算条件を、中央下の「ベイズ統計 MCMC の方法設定」で設定できる。既定値で DLL (C 言語で記述し、VBA から呼び出して使用)

を使用する設定になっている。図 1 に示した既定値の設定では、新しいシートに計算結果を記述するようになっており、OK を押すと CMB 計算が開始する。

#### 4. 計算法によるシミュレーション

##### 4.1 シミュレーションの目的

ここでは、発生源プロファイルと寄与を仮定して、本研究の CMB 法が、未知発生源が存在する場合としない場合とで、発生源寄与などを正しく再現できるかどうか、吟味する。

##### 4.2 用いた発生源プロファイル

環境省が公表した PM2.5 の発生源解析に用いる発生源プロファイル<sup>20)</sup> から、相互の相関係数の 2 乗が 0.6 を超えないものを、概ね線形独立性があると考え、これらを基礎となる発生源プロファイルとした。なお、この線形独立性は、環境汚染の観測点の化学種濃度には依存せず、発生源プロファイルだけに存在するものである。表 2 に発生源プロファイルを示す。

表 2 発生源プロファイル (単位 : mg/kg)

元素	A	B	C	D
Na	12500	13600	10000	120000
Al	61100	9990	2100	4200
K	12700	13200	850	200000
Ca	55200	45100	850	11000
Sc	13.3	1.32	0.09	0.46
Ti	5050	1000	740	900
V	108	125	6380	27
Cr	279	3160	210	850
Mn	1060	22000	120	330
Fe	53100	157000	4600	6100
Co	21.7	44	31	21
Ni	80.3	2900	3970	0
Cu	268	3700	1	3600
As	11.3	103.6	23	150
Se	1.43	51.1	48	0
Rb	55.6	76.8	2	260
Mo	8.55	0	0	0
Sb	13	90	6.9	952
Cs	3.59	3	0.1	12
Ba	499	500	1000	390
La	31.3	9.75	40	7.7
Ce	59.9	68.7	70	170
Sm	5.93	0.212	7.6	0.46
W	11.6	47.4	20	100

選択した発生源は、A (道路粉じん)、B (鉄鋼工業)、C (石油燃焼)、D (廃棄物焼却) であった。

計算では、Na、Al、K、Ca、Sc、Ti、V、Cr、Mn、Fe、Co、Ni、Cu、As、Se、Rb、Mo、Sb、Cs、Ba、La、Ce、Sm、W の 24 元素を項目とした。

##### 4.3 シミュレーションの方法

(1)式の比率モデルにおける無次元の発生源寄与を、表 3 のとおり、(a)未知発生源の寄与がない場合、および、(b) 未知発生源が存在する場合に設定した。この表から、(1)式のパラメータは、 $n_i - 1 = 24$ 、 $n_j = 4$  となる。

表 2 に示した発生源プロファイルから、行に元素、列に発生源を配した、単位系 g/g の要素を持つ行列を  $\mathbf{x}$  とする。また、表 3 の寄与設定値のベクトルを  $\boldsymbol{\theta}$  とする。

シミュレーションでは、未知発生源の  $m$  番目の化学種の組成  $\omega_m$  を、表 2 の発生源プロファイルから、

$$\omega_m = 0.5 \times \min(x_{m,1}, x_{m,2}, x_{m,3}, x_{m,4})$$

と設定した。これを単位系 g/g で要素を持つベクトルが、未知発生源組成ベクトル  $\boldsymbol{\omega}$  である。

ここで、未知発生源組成の設定には任意性があるが、本研究のように、発生源プロファイルの列方向で最小値を取るとは、このプロファイル行列に含まれる列ベクトルのすべてに直交するベクトルを選ぶ可能性につながる。よって、未知発生源の組成ベクトルを、発生源プロファイルとの多重共線性が弱いものとして選択したと解釈できる。

さて、環境汚染の観測点の化学種濃度  $\mathbf{Z}$  に誤差が含まれない場合、(1)式の CMB の方程式は次式で表される。

$$\mathbf{Z} = \mathbf{x}\boldsymbol{\theta} + \boldsymbol{\omega}(1 - \mathbf{1}_4\boldsymbol{\theta}) \quad (8)$$

本来ならば正規乱数を使用すべきであるが、ここでは簡便に、範囲  $[-0.5, 0.5]$  の一様乱数を要素を持つベクトルを  $\mathbf{R}$  とする。(8)式の CMB 方程式にこれを取り入れる際の精度を表す無次元の因子を  $f$  とし、環境汚染の観測点の化学種濃度のバラツキを考慮し、(8)式のベクトルの  $m$  行の要素を次の(9)式で表す。

$$Z_m = \{(\mathbf{x}\boldsymbol{\theta})_m + \omega_m(1 - \mathbf{1}_4\boldsymbol{\theta})\}(1 + fR_m) \quad (9)$$

ここで、添え字  $m$  は  $m$  行の要素を示す。

また、環境汚染の観測点の化学種濃度のバラツキの標準偏差  $\sigma_m$  を

$$\sigma_m = f \cdot \{(\mathbf{x}\boldsymbol{\theta})_m + \omega_m(1 - \mathbf{1}_4\boldsymbol{\theta})\}$$

と設定して、分散共分散行列の(2)式を構成する。

MCMC 計算は、burn-in 回数を  $n_0 = 800$  回、単一連鎖法回数を  $n_r = 1000$  回、多重連鎖法回数を  $n_M = 10$  回とした。

また、マルコフ連鎖においては、注目する計算回数の計算結果が直前の計算結果に大きく影響される。これを回避するため、単一連鎖法のマルコフ連鎖で一定の間隔 ( $n_{step}$  回) を空けて計算結果を採取して、モンテカルロ積分に供した<sup>21)</sup>。今回は、 $n_{step} = 10$  回とした。

#### 4.4 シミュレーションの結果と考察

表 3 に、 $f = 0.01$  の場合の、(a) 未知発生源の寄与がない場合、および (b) 未知発生源が存在する場合の発生源寄与の再現値、およびその比を示した。また、表 4 に、(b) の場合の組成の再現値を示した。

表 3 仮定した発生源寄与とシミュレーションによる再現値、およびその割合

##### (a) 未知発生源寄与がない場合

発生源	発生源寄与 (無次元)		
	設定値 ( $\alpha$ )	再現値 ( $\beta$ )	再現率 ( $\beta/\alpha \times 100\%$ )
A	0.82	0.825	103
B	0.1	0.0987	98.7
C	0.075	0.0713	95.1
D	0.005	0.00422	84.4

##### (b) 未知発生源の寄与がある場合

発生源	発生源寄与 (無次元)		
	設定値 ( $\alpha$ )	再現値 ( $\beta$ )	再現率 ( $\beta/\alpha \times 100\%$ )
A	0.42	0.406	96.7
B	0.1	0.101	101
C	0.075	0.0778	111
D	0.005	0.0172	344
未知	0.4	0.398	99.5

表 3 から、未知発生源の寄与がない場合 (a) には、寄与の再現率が 100% に近かった。ただし、未知発生源の寄与がある場合 (b) では、ない場合と比べると寄与推定の真度が下がり、特に寄与の小

さい発生源でその傾向が顕著であった。このことは、寄与が小さい発生源の推定において、CMB 計算で与えられる最大の寄与の、最小の寄与に対する比として測られる寄与のダイナミックレンジ (動的範囲) が大きい場合に、数値計算の真度が良好に保持されるかどうかの吟味が必要なことを示唆している。

表 4 未知発生源の存在を仮定した場合のその組成の再現値、および再現率

元素	設定値 ( $\alpha$ mg/kg)	再現値 ( $\beta$ mg/kg)	再現率 ( $\beta/\alpha \times 100\%$ )
Na	50000	45700	91.3
Al	10500	10900	103
K	4250	0	0
Ca	4250	4180	98.4
Sc	0.45	0	0
Ti	3700	3000	81.1
V	135	0	0
Cr	1050	806	76.7
Mn	600	0	0
Fe	23000	21900	95.1
Co	105	90.2	85.9
Ni	0	0	-
Cu	5	0	0
As	56.5	41.8	73.9
Se	0	0	-
Rb	10	0	0
Mo	0	0	-
Sb	34.5	0	0
Cs	0.5	0	0
Ba	1950	1660	85.3
La	38.5	31.3	81.3
Ce	299.5	259	86.4
Sm	1.06	0.285	26.8
W	58	46.7	80.5

表 4 から、未知発生源の寄与が存在する場合、その組成 (濃度) の設定値が 10 mg/kg 以上の元素は、ほとんどについて高い再現率が得られた。

以上のシミュレーションから、本研究で取り上げた CMB 法が、未知発生源の有無にかかわらず、その寄与および組成を概ね良好な再現率で推定できることが示された。

## 5. 実測データへの応用

### 5.1 環境汚染の観測点でのデータ

本研究では、青森県弘前市の文京小学校でサンプリングされた PM2.5 の成分分析結果を用いた。調査は、2019 年 5 月 8 日に 24 時間行った。表 5 に元素状および有機炭素を除いた、イオン成分と金属元素についての分析結果を、環境汚染の観測



点の濃度として示す。二次生成粒子に由来すると考えられる  $\text{NH}_4^+$ 、 $\text{NO}_3^-$ 、 $\text{Cl}^-$  および  $\text{SO}_4^{2-}$  を含むイオン成分の濃度は、CMB 解析には用いなかったが、発生源寄与を表示する表 6 には取り入れている。

CMB 計算に使用する測定値の標準偏差は、定量下限の 10 分の 1 として設定した。

## 5.2 CMB 法の適用と発生源プロフィール

CMB 発生源解析は、比較のため、本論文で記述したベイズ法(BAYES)に加え、早狩および著者が開発してきたマクロに含まれる次の四つの CMB 法で行った。すなわち、EVLSQ 法（有効分散最小二乗法）で負の寄与を与える発生源の除去を行った方法(EVLSQ-EN)、および行わなかった方法(EVLSQ-ALL)<sup>5), 15)</sup>、非負拘束最小二乗法(NN-EVLSQ)<sup>7)</sup>、非線形最適化法を用いた最尤法(MLM)<sup>8)</sup>で行った。

CMB 解析用の発生源プロフィールは、すでに報告されているもの<sup>16)</sup>を用いた。これらは、環境省<sup>20)</sup>、および東京都<sup>22)</sup>によるものである。表 5 にこれらのプロフィールを示した。

ベイズ法における MCMC 計算は、burn-in 回数を  $n_0 = 800$  回、単一連鎖法回数を  $n_r = 1000$  回、多重連鎖法回数を  $n_M = 10$  回、採用する間隔を  $n_{step} = 10$  回とした。

## 5.3 結果および考察

表 6 に、適用した各方法での発生源寄与としての CMB 計算結果を示す。また、図 2 の円グラフは、表 6 のうち、負の寄与を与えた EVLSQ-ALL を除き、結果から計算した寄与の比を示している。ここで、未知発生源寄与は、1 から各発生源寄与およびイオン成分の寄与の和を減算して推定した。各寄与の和が 1 を超えた場合、その未知発生源は寄与を持たないと判断した。

EVLSQ-ALL の方法を用いると、表 6 に示すように、負の発生源寄与が得られた。それゆえに、この PM<sub>2.5</sub> のデータ解析では、非負の寄与を与えるほかの CMB 法の利用が期待された。

表 6 および図 2 から、適用した CMB 法により異なった発生源寄与が得られており、このことは、解析対象とした PM<sub>2.5</sub> のデータセットの発生源プロフィール間の線形独立性が必ずしも完全でない（多重共線性が存在する）ために、数値計算が比

較的困難なことを示唆した。このような場合、多重共線性の影響を緩和できるベイズ法が本領を発揮すると考えられる。

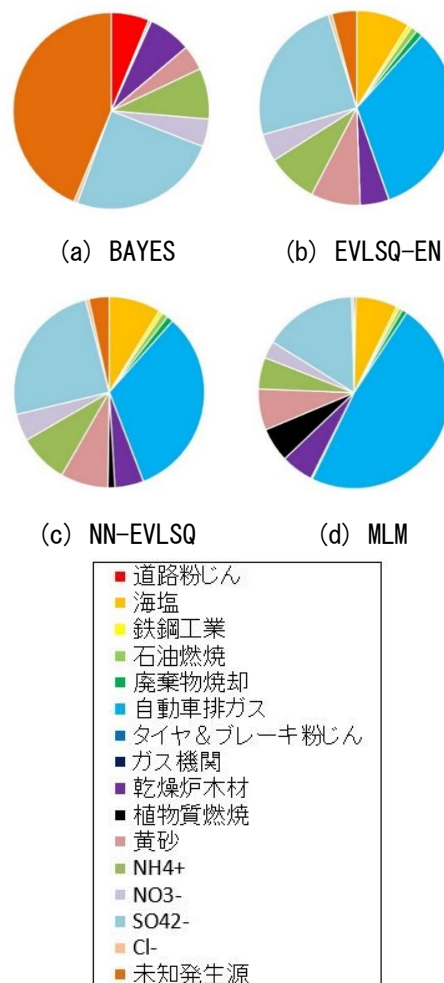


図 2 4 種類の CMB 計算で得られた発生源寄与。

表 6 に示した CMB 計算結果のうち、ベイズ法を用いた場合は、解析結果に比較的大きな未知発生源の正の寄与を与えるという、他と違った特徴を持っていた。

柏木のベイズ法以外の、LSQ などの従来法では、環境汚染の観測点での化学種濃度を質量濃度で除した濃度（単位：g/g）に換算し（以下、「換算濃度」と言う。）、また、未知発生源の存在を仮定せずに CMB 計算を行う。この換算濃度の考えは比率モデルの考えと同じである。このとき、未知発生源の寄与は、仮定した発生源に割り振られた寄与の合計を 1 から減算した値を用いる。

一方、柏木によるベイズ法について (1)式から、CMB モデルを(10)式のとおり表す。すなわち、発生源の寄与を  $\theta_j$  ( $1 \leq j \leq n_j$ )、未知発生源の組成

を  $\omega_i$  ( $1 \leq i \leq n_i - 1$ ) とすると、環境汚染の観測点における化学種濃度  $Z_i$  は、その誤差  $\epsilon_i$  が 0 を期待値とする正規分布をなすとして、

$$Z_i = \sum_{j=1}^{n_j} x_{ij}\theta_j + \omega_i \left( 1 - \sum_{j=1}^{n_j} \theta_j \right) + \epsilon_i \quad (10)$$

となる。ここで、 $x_{ij}$  は発生源プロフィールの行列要素である。

他方、従来法の CMB モデル式は、得られた寄与を  $\theta'_j$  ( $1 \leq j \leq n_j$ ) とすると、次の(11)式で表される。

$$Z_i = \sum_{j=1}^{n_j} x_{ij}\theta'_j + \epsilon'_i \quad (11)$$

ここで、 $\epsilon'_i$  は誤差項である。いま(10)(11)式を比較すると、

$$\sum_{j=1}^{n_j} x_{ij}\theta_j + \omega_i \left( 1 - \sum_{j=1}^{n_j} \theta_j \right) + \epsilon_i = \sum_{j=1}^{n_j} x_{ij}\theta'_j + \epsilon'_i$$

すなわち、

$$\sum_{j=1}^{n_j} x_{ij}(\theta'_j - \theta_j) = \omega_i \left( 1 - \sum_{j=1}^{n_j} \theta_j \right) + \epsilon_i - \epsilon'_i$$

発生源プロフィールおよび寄与、未知発生源組成を行列およびベクトルで表記すると、

$$\mathbf{x}(\boldsymbol{\theta}' - \boldsymbol{\theta}) = \boldsymbol{\omega} \left( 1 - \mathbf{1}_{n_j} \boldsymbol{\theta} \right) + \mathbf{E}$$

ここで、誤差のベクトルを  $\mathbf{E}$  とし、その要素を  $\epsilon_i - \epsilon'_i$  とした。

行列  $\mathbf{x}$  の逆行列は一般的には存在しない。しかし、特異値分解を使用すると、近似的な一般逆行列を求められる。すなわち、直交行列  $\mathbf{U}$ 、 $\mathbf{V}$  を用いると、特異値分解は、

$$\mathbf{x} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^t$$

と表せる。ここで、特異値を  $\{s_1, \dots, s_{n_j}\}$  とし、

$$\boldsymbol{\Sigma} = \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_{n_j} \\ & & & 0 \end{pmatrix}$$

である。次に、

$$\boldsymbol{\Sigma}' = \begin{pmatrix} s_1^{-1} & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & s_{n_j}^{-1} \end{pmatrix}$$

とし、直交行列の性質  $\mathbf{U}^{-1} = \mathbf{U}^t$ 、 $\mathbf{V}^{-1} = \mathbf{V}^t$  を用いると、一般逆行列  $\mathbf{x}'$  は、

$$\mathbf{x}' = \mathbf{V}\boldsymbol{\Sigma}'\mathbf{U}^t$$

となり、近似的に(12)式が成り立つ。

$$\boldsymbol{\theta}' \cong \boldsymbol{\theta} + \mathbf{x}'\boldsymbol{\omega} \left( 1 - \mathbf{1}_{n_j} \boldsymbol{\theta} \right) + \mathbf{x}'\mathbf{E} \quad (12)$$

(12)式から、従来法で得られる既知発生源の寄与  $\boldsymbol{\theta}'$  は、未知発生源寄与を仮定するベイズ法で得られる既知発生源寄与  $\boldsymbol{\theta}$  に、未知発生源由来の濃度の関数となる項  $\mathbf{x}'\boldsymbol{\omega} \left( 1 - \mathbf{1}_{n_j} \boldsymbol{\theta} \right)$  および誤差に関与する項を加えたものとなる事が分かる。すなわち、従来法は、既知発生源に対して、未知発生源の影響を含んだ寄与を与える。

以上から、現状では、未知発生源の寄与とその組成の推定は、柏木のベイズ法によってのみ可能であると言える。この方法は、範囲  $[0,1]$  に入る発生源寄与を与える。発生源寄与がこの範囲であることは、環境汚染の観測点における換算濃度が、発生源における濃度の 0 倍以上 1 倍以下であることを意味し、発生源から放出された化学種が希釈こそされ、濃縮されずに、環境汚染の観測点に到達するという物理的に合理的な図式である。それゆえ、柏木の方法は、物理的に容易に解釈できる発生源解析結果を与える。なお、従来法では、発生源寄与に上限を設けておらず、対象データによっては、非負であっても 1 を超える発生源寄与が

得られることがある。本研究でも表 6 に示した EVLSQ-ALL の方法で、発生源「自動車排ガス」に

対して、1 を超える寄与が得られている。このような場合、換算濃度が発生源濃度を上回るため、

表 5 環境汚染の観測点の化学種濃度と発生源プロファイル。環境汚染の観測点の化学種濃度における”Sd”は、測定値の誤差の標準偏差を意味する。

発生源プロファイル(mg/kg)												環境汚染の観測点の濃度	
元素	道路 粉じん	海塩	鉄鋼 工業	石油 燃焼	廃棄物 焼却	自動車 排ガス	タイヤ& ブレーキ	ガス機 関	乾燥炉 木材	植物質 燃焼	黄砂	測定値( $\mu\text{g}/\text{m}^3$ )	Sd( $\mu\text{g}/\text{m}^3$ )
Na	12500	304000	13600	10000	120000	76.4	9200	7200	100000	6500	72000	0.457	0.0014
Al	61100	0.29	9990	2100	4200	1570	49000	6100	300	63000	75300	0.0974	0.00098
K	12700	11000	13200	850	200000	197	8060	3300	72000	370	22100	0.0963	0.0002
Ca	55200	11700	45100	850	11000	1460	12000	79000	7600	420	30400	0.073	0.0013
Sc	13.3	0.0012	1.32	0.09	0.46	0.119	2.3	0	0	0	15.7	0.00002	2.2E-06
Ti	5050	0.029	1000	740	900	146	3000	0	0	0	4150	0.0059	0.00067
V	108	0.058	125	6380	27	7.25	110	0	3.8	0	111	0.00107	5.4E-06
Cr	279	0.0015	3160	210	850	11.6	340	720	75	0	97.5	0.00031	0.000048
Mn	1060	0.058	22000	120	330	19.3	1410	540	84	10	826	0.00308	6.6E-06
Fe	53100	0.29	157000	4600	6100	989	76600	32000	1200	100	45100	0.077	0.0014
Co	21.7	0.015	44	31	21	1.47	20	1000	1.2	0	20.1	3.03E-05	1.4E-07
Ni	80.3	0.015	2900	3970	0	9.89	260	0	29	0	54.9	0.0004	0.00013
Cu	268	0.017	3700	1	3600	113	8700	0	110	0	113	0.000742	8.1E-06
As	11.3	0.029	103.6	23	150	3.69	100	0	25	0	22.4	0.00195	2.5E-06
Se	1.43	0.12	51.1	48	0	1.67	20	0	0.20	0	3.15	0.000673	2.2E-06
Rb	55.6	3.5	76.8	2	260	0.49	34	0	130	0	112	0.000321	7.1E-07
Mo	8.55	0	0	0	0	5.91	512	0	0	0	11.9	0.000177	1.2E-06
Sb	13	0.014	90	6.9	952	19.6	4200	0	0	0	14.8	0.000304	1.2E-06
Cs	3.59	0.029	3	0.1	12	0.0745	3.7	0	0	0	10.5	3.62E-05	2.8E-07
Ba	499	0.86	500	1000	390	98.9	18800	0	94	0	609	0.00116	4.8E-06
La	31.3	0.009	9.75	40	7.7	0.341	7.5	0	0.11	0	62.1	5.64E-05	5.2E-07
Ce	59.9	0.012	68.7	70	170	1.08	24	0	0.25	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	8.05	7.6E-06	5.2E-07
W	11.6	0.003	47.4	20	100	0.686	9.6	140	0	0	3.38	0.000044	1.8E-06
											NH <sub>4</sub> <sup>+</sup>	1.08	
											NO <sub>3</sub> <sup>-</sup>	0.59	
											SO <sub>4</sub> <sup>2-</sup>	3.19	
											Cl <sup>-</sup>	0.078	
											質量濃度	12.9	

表 6 5種類のCMB発生源解析により得られた発生源寄与（無次元）。

	道路 粉じん	海塩	鉄鋼 工業	石油 燃焼	廃棄物 焼却	自動車排 ガス	タイヤ& ブレーキ	ガス機関	乾燥炉 木材	植物質 燃焼	黄砂	NH <sub>4</sub> <sup>+</sup>	NO <sub>3</sub> <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	Cl <sup>-</sup>	未知 発 生 源
BAYES	0.0627	0.0000	0.0000	0.0000	0.0034	0.0000	0.0013	0.0000	0.0709	0.0000	0.0414	0.0837	0.0457	0.2473	0.0060	0.4376
EVLSQ-ALL	-0.1783	0.0968	0.0107	0.0107	0.0073	1.3032	-0.0041	-0.0006	0.0450	0.0257	0.1963	0.0837	0.0457	0.2473	0.0060	0.0000
EVLSQ-EN	0.0000	0.0882	0.0070	0.0102	0.0108	0.3303	0.0000	0.0000	0.0481	0.0000	0.0817	0.0837	0.0457	0.2473	0.0060	0.0410
NN-EVLSQ	0.0000	0.0877	0.0070	0.0103	0.0107	0.3260	0.0001	0.0000	0.0487	0.0118	0.0807	0.0837	0.0457	0.2473	0.0060	0.0342
MLM	0.0048	0.1105	0.0085	0.0087	0.0133	0.7529	0.0020	0.0014	0.0860	0.0906	0.1091	0.0837	0.0457	0.2473	0.0060	0.0000

表 7 環境汚染の観測点の濃度、ベイズ法で得られた濃度、およびその比の比較、並びに未知発生源の推定組成。環境汚染の観測点の濃度における”Sd”は、測定誤差の標準偏差を示す。

元素	環境汚染の観測点の濃度 ( $\mu\text{g}/\text{m}^3$ )						未知発生源(mg/kg)
	実測値(M)		推定値(C)		C/M(%)		推定値
	実測値	Sd	推定値	Sd	計算値	Sd	推定値
Na	0.457	0.0014	0.451	<1E-14	98.6	0.3	31900
Al	0.0974	0.00098	0.0909	<1E-14	93.4	0.9	0
K	0.0963	0.0002	0.0963	<1E-14	100.0	0.2	0
Ca	0.073	0.0013	0.0685	<1E-14	93.9	1.7	0
Sc	0.00002	2.2E-06	1.92E-05	<1E-14	96.0	10.6	0
Ti	0.0059	0.000067	0.00639	<1E-14	108.3	1.2	0
V	0.00107	5.4E-06	0.00102	<1E-14	95.0	0.5	82
Cr	0.00031	0.000048	0.000389	<1E-14	125.6	19.4	0
Mn	0.00308	6.6E-06	0.00302	<1E-14	97.9	0.2	151
Fe	0.077	0.0014	0.0697	<1E-14	90.5	1.6	0
Co	3.03E-05	1.4E-07	3.06E-05	<1E-14	101.0	0.5	0
Ni	0.0004	0.00013	0.000126	<1E-14	31.4	10.2	0
Cu	0.000742	8.1E-06	0.000680	<1E-14	91.7	1.0	0
As	0.00195	2.5E-06	0.00193	<1E-14	98.8	0.1	177
Se	0.000673	2.2E-06	0.000651	<1E-14	96.8	0.3	61
Rb	0.000321	7.1E-07	0.000314	<1E-14	97.8	0.2	8
Mo	0.000177	1.2E-06	0.000165	<1E-14	93.3	0.6	14
Sb	0.000304	1.2E-06	0.000292	<1E-14	96.1	0.4	15
Cs	3.62E-05	2.8E-07	3.34E-05	<1E-14	92.3	0.7	2
Ba	0.00116	4.8E-06	0.00115	<1E-14	99.5	0.4	0
La	5.64E-05	5.2E-07	5.90E-05	<1E-14	104.7	1.0	0
Ce	0.000107	1.2E-06	0.000114	<1E-14	106.7	1.2	0
Sm	7.6E-06	5.2E-07	9.12E-06	<1E-14	120.0	8.2	0
W	0.000044	1.8E-06	2.56E-05	<1E-14	58.2	2.4	1

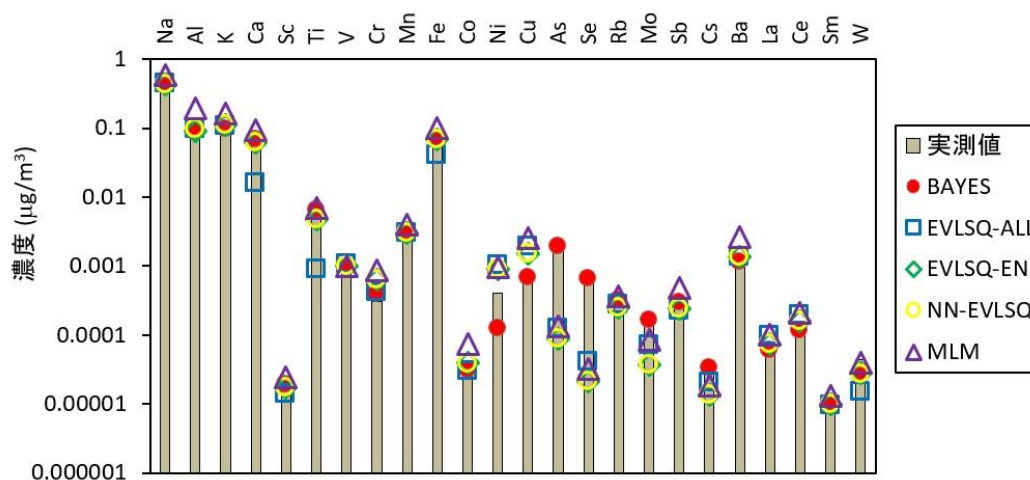


図 3 5 種類の CMB 法で計算された濃度と、実測の環境汚染の観測点の濃度の比較。

物理的解釈が困難となる。

図 3 では、5 種類の CMB 法により得られた、環境汚染の観測点の濃度推定結果を比較した。表 7 は、ベイズ CMB 法で計算された環境汚染の観測点の濃度実測値と推定値、それらの実測値に対する比、およびこの方法で得られた未知発生源の組

成を示す。

ベイズ法は、CMB 計算における環境汚染の観測点の濃度推定がほかの LSQ (最小二乗法) や MLM 法よりも極めて正確であるという利点を持つことが分かる。すなわち、表 7 と図 3 に示したように、Cu から Mo の元素の環境汚染の観測点での低濃度

が、ベイズ法では非常に良く再現された。また、この方法により、未知発生源の組成も与えられ、表7に示したように、この例では、Naに富んだ組成を推定した。

## 6. 結論

柏木のベイズ法を表計算ソフトであるマイクロソフト社 Excel のマクロに移植し、計算法を変更した。この方法は、組成モデルにより、また未知発生源の存在を明確に仮定し、その寄与と組成を推定できるという長所がある。

本研究で作成したマクロにより試みた CMB 計算で以下の3点が明らかになった。

- 1) シミュレーションにより、発生源の寄与のみならず、未知発生源の組成までを再現できた。
- 2) 本研究で取り上げたデータの限りでは、ベイズ法は、環境汚染の観測点の試料に含まれるいくつかの元素の低濃度測定値を、計算で非常に正確に再現した。
- 3) 作成したマクロにおいて MCMC 計算が少ない繰り返し回数で収束した。例えば、本研究の例では、1000回の繰り返し計算で十分な収束を与えた。収束が速い理由は、切断正規分布としての完全条件付き密度の計算過程において数値計算安定性を保持しているためと考えられる。

以上のとおり、本研究において、発生源解析に有効な CMB マクロを作成できた。このマクロにより、より正確な発生源解析が可能となったと考える。当センターの内海ほか<sup>23)</sup>は、このマクロを利用して、PM<sub>2.5</sub>成分分析結果解析を行った。

## 文 献

- 1) 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.
- 2) Friedlander, S. K.: Chemical element balances and identification of air pollution sources. *Environmental Science and Technology*, **7**, 235-240, 1973.
- 3) Hopke, P. K.: Recent developments in receptor modeling. *Journal of Chemometrics*, **17**, 255-265, 2003.
- 4) Kashiwagi, N.: Chemical mass balance when an unknown source exists. *Environmetrics*, **15**, 777-796, 2004.
- 5) Watson, J. G. et al.: The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atmospheric Environment*, **18**,

1347-1355, 1984.

- 6) [https://www3.epa.gov/scram001/receptor\\_cmb.htm](https://www3.epa.gov/scram001/receptor_cmb.htm) (2021.12.24 アクセス可能)
- 7) 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.
- 8) 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.
- 9) Billheimer, D.: Compositional receptor modeling. *Environmetrics*, **12**, 451-467, 2001.
- 10) 柏木宣久, 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.
- 11) Park, E. S. et al.: Multivariate receptor models and model uncertainty. *Chemometrics and Intelligent Laboratory Systems*, **60**, 49-67, 2002.
- 12) Keats, A. et al.: Bayesian treatment of a chemical mass balance receptor model with multiplicative error structure. *Atmospheric Environment*, **43**, 510-519, 2009.
- 13) Massoudieh, A. et al.: Bayesian Chemical Mass Balance Method for Surface Water Contaminant Source Apportionment, *Journal of Environmental Engineering*, **139**, 250-260, 2013.
- 14) [https://ismrepo.ism.ac.jp/?action=repository\\_uri&item\\_id=32744](https://ismrepo.ism.ac.jp/?action=repository_uri&item_id=32744) (2021.12.24 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.
- 16) 花石竜治: 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. 全国環境研会誌, **46**, 132-137, 2021.
- 17) 柏木宣久, 他: 関数関係解析による Chemical Mass Balance. 応用統計学, **31**, 59-74, 2002.
- 18) 中妻照雄: 入門 ベイズ統計学, 朝倉書店, 東京, 2007.
- 19) Zelen, M. et al.: Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. Abramowitz, M. et al. eds., Dover, New York, 1965.
- 20) 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. ムラタ計測機サービス株式会社, 2009.
- 21) 板野泰之, 他: 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.

22) [https://www.kankyo.metro.tokyo.lg.jp/air/air\\_pollution/torikumi/pm2\\_5/pm2\\_5.html](https://www.kankyo.metro.tokyo.lg.jp/air/air_pollution/torikumi/pm2_5/pm2_5.html)  
(2021.12.24 アクセス可能)

23) 内海宣俊, 他: ベイズ統計手法を用いた CMB 法による微小粒子状物質(PM2.5)成分の発生源解析. 青森県環境保健センター年報, **32**,123-144,2021.

## Development of a spreadsheet application macro for chemical mass balance calculation using Bayesian statistical technique

Ryuji Hanaishi

Chemical mass balance (CMB) method is one approach for analyzing environmental pollution based on receptor models. The algorithm for CMB method established by Kashiwagi (2004) adopted Bayesian statistical technique (hereinafter referred to as Bayesian method). In this study, the author improved the numerical calculation method and created a spreadsheet application macro. A feature of the Bayesian method is that contribution and composition of unknown sources can be included in the model formula for assessment. In addition, the Bayesian method gives nonnegative source contribution only and allows physically easier interpretation of source contribution than conventional approaches because it uses composition models. Simulations assuming source contribution and contribution and composition of unknown sources resulted in a generally successful reproduction of these values. In addition, the improved method could more correctly reproduce the observed values at the environmental pollution observation spots than other CMB methods. Thus, usefulness of the improved Bayesian method for source analyses was demonstrated.

Key words: Chemical mass balance, Bayesian statistics, Markov chain Monte Carlo method, Unknown source