

環境汚染に対する未確認発生源の寄与率の推定

柏木 宣久¹・吉澤 正²・茨木 剛³・加藤 謙一⁴・橋本 俊次⁵・佐々木 裕子⁶

(受付 2005 年 8 月 31 日; 改訂 2006 年 3 月 24 日)

要 旨

環境管理にとって汚染発生源の同定は本質的な問題である。発生源を同定できなければ、如何なる発生源対策も立てようがないからである。発生源を同定するためリセプターモデルと呼ばれる様々な数理的方法が提案されている。しかしながら、確立した方法は存在しない。未確認発生源、すなわち組成が未知の発生源の存在を仮定すると、未知変数は識別不能になるからである。この識別問題を一般的に解決するのは困難である。ただし、ベイズ的方法を用いると、識別不能な未知変数の実用的な推定値を得られるようになる。本稿では、汚染物質の環境濃度に対する未確認発生源の寄与率を推定するためのベイズ的方法を詳解する。加えて、実際の環境汚染問題に対する適用結果を示す。

キーワード：ベイズ，ケミカルマスバランス，ダイオキシン類，環境管理，リセプターモデル。

1. はじめに

環境管理のための数理的方法のひとつにリセプターモデルがある。環境における汚染物質等の濃度に基づき発生源について推論するための方法で、粒子状物質による大気汚染の問題化に伴い、1970 年前後から米国で活発に議論されるようになった。当時、環境管理のための数理的方法として移流拡散モデルが既に知られており、環境に対する発生源の影響を評価するため、硫酸化物や窒素酸化物による大気汚染等で頻りに利用されていた。ただし、この方法を利用するためには、発生源の位置や発生源からの汚染物質の排出量といった発生源情報と汚染物質の環境中での挙動や環境媒体の移流に関する情報を予め特定しなければならない。ところが、粒子状物質の場合、発生源が多種多様かつ至る所に存在したため、発生源情報の特定さえ容易でなく、発生源について推論する必要が生じ、リセプターモデルが注目されるようになった。

粒子状物質と同様の状況にあるのが近年多くの関心を集めたダイオキシン類である。ダイオキシン類とは polychlorinated dibenzo-para-dioxins(PCDDs), polychlorinated dibenzofurans(PCDFs), および dioxin-like polychlorinated biphenyls(ダイオキシン様 PCBs)という 3 種類

¹ 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

² 千葉県環境研究センター：〒290-0046 千葉県市原市岩崎西 1-8-8

³ 新潟県保健環境科学研究所：〒950-2144 新潟県新潟市曾和 314-1

⁴ 宮城県保健環境センター：〒983-0836 宮城県仙台市宮城野区幸町 4-7-2

⁵ 国立環境研究所：〒305-8506 茨城県つくば市小野川 16-2

⁶ 東京都環境科学研究所：〒136-0075 東京都江東区新砂 1-7-5

の有機物の総称で、何れも 2 つのベンゼン環が結合した骨格構造と塩素基を持ち、塩素基の数や位置が異なる異性体が多数存在する。これらダイオキシン類の主な発生源として農業、PCB 製品、燃焼過程等が知られているが、農業は農地にくまなく撒布され、PCB 製品は依然として各地に散在し、燃焼過程はそれこそ至る所に存在する。そのため、ダイオキシン類の場合も、発生源情報の特定は容易でなく、リセプターモデルの利用が検討されている(例えば, Su and Christensen, 1997; Imamoglu and Christensen, 2002; 柏木 他, 2002)。ただし、リセプターモデルの利用にも、いくつか問題が存在する。

最も基本的な問題が発生源の同定である。ダイオキシン類の発生源は各発生源で観測された異性体の組成と照合し同定される。ただし、農業、PCB 製品、燃焼過程には極めて多くの種類があり、同一種類であっても異性体の組成は生成条件等により変化する。そのため、あらゆる発生源組成を網羅するのは現実には困難である。また、ダイオキシン類の骨格構造は分解されにくく、発生源から放出された後、構造を維持したまま様々な媒体を通じて移動し蓄積される。そのため、発生時期、場所が近年、近隣とは限らず、発生源候補を安易に絞れない。同時に、骨格構造は分解されにくい、塩素基が紫外線等の影響により置換されたり、異性体により揮発蒸散や生体濃縮の程度が異なるため、発生源から放出された後、異性体の組成は多かれ少なかれ変化する。こうした事情から、ダイオキシン類の場合、発生源の同定は容易でなく、既知でありながら同定できなかった発生源や既知の組成から変化した組成あるいは全く未知の組成を持つ発生源(以降、これらをまとめて未確認発生源と呼ぶ)の存在を常に疑わなければならない。

発生源の同定は粒子状物質の場合にも話題になり、因子分析に由来する発生源の同定法が断片的に議論されている(文献は、例えば、Hopke, 1995, 2003 を参照)。ただし、何れの方法も、解の一意性等の因子分析が抱える伝統的な問題を克服できず、主要な発生源を主観的に同定するための補助情報を提供するだけで、未確認発生源について推論できる段階には至っていない。そこで、Kashiwagi (2004) は、環境濃度に対する未確認発生源の寄与率を推定するためのベイジックの方法を提案し、未確認発生源について客観的に推論する方策を示した。本稿では、このベイジックの方法を詳解する。加えて、実際の環境問題に対する適用結果を示す。

先ず次節でリセプターモデルの概要について述べる。次いで第 3 節で提案のモデルを説明する。第 3.1 節で未確認発生源が存在する場合の観測モデルを仮定し、第 3.2 節で残差の共分散成分について述べる。第 3.3 節では仮定した観測モデルの識別可能性について議論し、識別可能性を確保するための事前分布を第 3.4 節で仮定する。第 4 節では推定法を説明し、第 5 節で計算法について述べる。第 6 節では提案の方法の性能をシミュレーションにより検証する。最後に第 7 節で提案の方法を実際の環境問題に適用する。

2. リセプターモデルの概要

本節では、未確認発生源の取り扱いに注目しつつ、リセプターモデルの概要について述べる。なお、習慣に従い、確率変数を大文字で表記する。

リセプターモデルの基本となるのが以下の質量釣合方程式である。

$$(2.1) \quad \psi_i = \sum_{j=1}^{n_j} \phi_{ij} = \sum_{j=1}^{n_j} \xi_{ij} \phi_j$$

ここに、 ψ_i は第 i 番目 ($i=1, \dots, n_i$) の物質の環境濃度、 ϕ_{ij} は第 j 番目 ($j=1, \dots, n_j \leq n_i$) の

発生源に由来する第 i 番目の物質の濃度を表し、そして

$$\phi_j = \sum_{i=1}^{n_i} \phi_{ij}, \quad \xi_{ij} = \phi_{ij} / \phi_j \quad \text{when } \phi_j > 0$$

である。

等式(2.1)は、環境濃度 ψ_i と環境における発生源組成 ξ_{ij} の観測値に基づく発生源寄与 ϕ_j の推定に主な関心がある時、Chemical Mass Balance (CMB) 方程式とも呼ばれ、リセプターモデルの中心的な方法である CMB を導く。ただし、環境で発生源組成を観測するのは現実には困難なため、CMB では ξ_{ij} の観測値として発生源で観測された発生源組成を利用できると仮定する。そして、多くの場合、以下の回帰モデルに対応する最小 2 乗法を用いて ϕ_j を推定する。

$$(2.2) \quad Y_i = \sum_{j=1}^{n_j} x_{ij} \phi_j + V_i$$

ここに、 Y_i は ψ_i の観測 ($\psi_i = E(Y_i)$)、 x_{ij} は ξ_{ij} の観測 X_{ij} の実現値 ($\xi_{ij} = E(X_{ij})$)、および V_i は残差を表す。最小 2 乗法を CMB に初めて適用したのは Friedlander (1973) である。Kowalczyk et al. (1978) は、測定精度が物質に依存する点に注目し、通常のリセプターモデルを利用した。Watson et al. (1984) は、発生源組成の変動を明示的に扱うため、測定誤差モデルについて議論し、測定誤差モデルの近似的な方法として知られる Berkson モデル (Berkson, 1950) に対応する有効分散法と呼ぶ重み付き最小 2 乗法の利用を提案した。この方法は米国環境保護局の標準的方法として現在でも利用されている。その他、関連する方法として、非負制約付き最小 2 乗法 (Wang and Hopke, 1989)、2 段階最小 2 乗法 (Ohtaki et al., 1997)、頑健推定法 (Wiens et al., 2001)、ベイズ型組成モデル (Billheimer, 2001)、および組成に基づく関数関係モデル (柏木 他, 2002) がある。ただし、モデル(2.2)から分かるように、これらの方法では全ての発生源は同定されていると仮定される。

既存の CMB で未確認発生源の存在を許容できるのは線形計画法 (例えば、Henry et al., 1984 を参照) だけである。この方法では

$$y_i + 3\hat{\sigma}_{y_i} \geq \sum_{j=1}^{n_j} x_{ij} \phi_j, \quad 0 \leq \phi_j \leq \sum_{i=1}^{n_i} y_i$$

という制約条件の下で費用関数 $\sum_j \phi_j$ が最大になるよう各 ϕ_j を推定する。ここに、 y_i は Y_i の実現値、および $\hat{\sigma}_{y_i}^2$ は Y_i の分散の推定値を表す。ただし、費用関数に実際的な根拠は無く、統計的な仮定も設定されないため、線形計画法では如何なる統計的推論もできない。

発生源が未知で環境濃度の系列 $\psi_i(t)$ ($t=1, \dots, n_t$) が観測されている時、発生源を同定するため、しばしば因子分析が用いられる。因子分析もリセプターモデルの一種で、モデルは

$$\tilde{Y}_i(t) = \sum_{j=1}^{n_j} \tilde{\xi}_{ij} \Phi_j(t) + V_i(t)$$

と書ける。ここに、 $\tilde{Y}_i(t)$ は $\psi_i(t)$ の標準化された観測、 $\tilde{\xi}_{ij}$ は発生源組成と関係する因子負荷量、 $\Phi_j(t)$ は発生源寄与と関係する因子得点、および $V_i(t)$ は残差を表す。因子数 n_j を特定すれば、通常の方法により未知変数 $\tilde{\xi}_{ij}$ と $\Phi_j(t)$ を形式的に推定できる。ただし、よく知られているように、推定値は少数の例外を除き一意に決まらない。加えて、形式的に得られる推定値は非現実的である。なぜなら、いくつかの推定値は必然的に負になるからである。そのため、因子分析で得られる結果は回転法の助けを借りながら主観的に解釈される。リセプターモデルのための回転法も提案されている (例えば、Hopke, 1995 を参照)。

因子分析の欠点を是正するため, Paatero and Tapper (1993) は Positive Matrix Factorization と呼ぶ方法を提案している. モデルは

$$Y_i(t) = \sum_{j=1}^{n_j} \xi_{ij} \phi_j(t) + V_i(t)$$

で与えられる. ここに, $Y_i(t)$ は $\psi_i(t)$ の観測 ($\psi_i(t) = E(Y_i(t))$), および $\phi_j(t)$ は第 t 番目の観測に対する第 j 番目の発生源の寄与を表す. 現実的な推定値を得るため, 未知変数 ξ_{ij} と $\phi_j(t)$ に対し自然な制約条件を仮定し, 制約付き逐次的交互最小 2 乗法を用いて推定する. ただし, この方法によっても推定値は少数の例外を除き一意に決まらない. 提案者は一意性の問題を実用上無視できると主張していたが, Park et al. (2002) が一意性を確保できる少数の例外を全て指摘し問題の存在を明確にした. Park et al. の論文は, 実質的に否定的な内容にもかかわらず, 掲載雑誌としては異例の討論付き論文として扱われている. 発生源の同定に対する関心の高さを窺わせる.

環境を管理するには発生源寄与を評価する必要がある, 発生源寄与を偏り無く推定するためには全ての発生源を同定しなければならない. しかし, 全ての発生源を同定するための手段は確立されていない. せめて未確認発生源の寄与を推定できれば偏りを是正できるが, 未確認発生源の組成と寄与を変数としてモデルに組み込めば, 因子分析と同様, 変数は少数の例外を除き識別不能になる. この識別問題を一般的に解決するのは困難である. そこで, Kashiwagi (2004) は, 実用的な推定値の獲得に焦点を絞り, 未確認発生源の寄与率を推定するためのベイズ的方法を提案した. 次節以降で提案の方法を説明する.

3. 提案のモデル

本節では, 未確認発生源の寄与率を推定するためのベイズモデルについて述べる. 仮定する未確認発生源の数は高々 1 とする. すなわち, 複数の未確認発生源の分解までは考慮しない. 分解を実現するには Park et al. (2002) のように未確認発生源に関する事前情報が必要になるからである. 方法の客観性を可能な限り確保するため, 自然な制約条件以外的事前情報は利用しない. また, ダイオキシン類を頻繁に測定するのは困難なため, 因子分析とは異なり, 系列データの存在を前提としない. ただし, 観測の繰り返しは推定精度の改善に有用である.

3.1 CMB モデル

同定された発生源が n_j ある傍ら, 未確認発生源の存在が疑われる時, 質量釣合方程式 (2.1) は以下のように書き換えられる.

$$(3.1) \quad \psi_i = \sum_{j=1}^{n_j} \xi_{ij} \phi_j + \omega_i \phi_{n_j+1}$$

ここに, ω_i は未確認発生源に由来する第 i 番目の物質の組成, および ϕ_{n_j+1} は未確認発生源の寄与を表す. 一般に, CMB モデルは (2.2) のように質量釣合方程式から直接導かれる. ただし, 環境における観測が組成の形で与えられる場合も稀ではない. また, 自然な制約条件を仮定するには発生源寄与を比率の形で扱う方が好ましい (Billheimer, 2001; 柏木 他, 2002). そこで (3.1) を更に以下のように書き換える.

$$(3.2) \quad \zeta_i = \sum_{j=1}^{n_j} \xi_{ij} \theta_j + \omega_i \left(1 - \sum_{j=1}^{n_j} \theta_j \right)$$

ここに,

$$\zeta_i = \psi_i / \psi, \quad \theta_j = \phi_j / \psi, \quad \psi = \sum_{i=1}^{n_i} \psi_i = \sum_{j=1}^{n_j+1} \phi_j > 0$$

である.そして,環境組成 ζ_i の観測の繰り返しが n_k あるとして,CMB モデルを以下のように仮定する.

$$(3.3) \quad Z_{ik} = \sum_{j=1}^{n_j} x_{ij} \theta_j + \omega_i \left(1 - \sum_{j=1}^{n_j} \theta_j \right) + V_{ik}, \quad i=1, \dots, n_i-1, \quad k=1, \dots, n_k$$

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

ここに, Z_{ik} は ζ_i の第 k 番目の観測 ($\zeta_i = E(Z_{ik})$), x_{ij} は ξ_{ij} の観測 X_{ij} の実現値 ($\xi_{ij} = E(X_{ij})$), および V_{ik} は平均 0 の正規分布に従う残差を表す. 環境における観測を組成の形で扱うため, 添字 i の範囲を n_i までではなく $n_i - 1$ までとしている. また, 同じ理由により, 残差 V_{ik} の共分散を定義しなければならないが, ここでは以下のように書いておく.

$$\text{cov}(V_{ik}, V_{i'k'}) = \begin{cases} \delta^2 \sigma_{ii'} & \text{if } k = k' \\ 0 & \text{otherwise} \end{cases}$$

ここに, δ^2 は未知の調整変数, および $\sigma_{ii'}$ は次節で述べる共分散成分を表す.

提案のモデルでは,問題を単純化するため,同定された発生源の組成 x_{ij} の不確実性を無視できると仮定した. 発生源組成の不確実性を扱うには関数関係(3.2)と共に以下の測定誤差モデルを仮定すればよい.

$$X_{ij} = \xi_{ij} + U_{ij}, \quad Z_{ik} = \zeta_i + V_{ik}$$

ここに, U_{ij} と V_{ik} は残差を表す.

3.2 共分散

観測の繰り返し数 n_k が十分に大きければ, Z_{ik} の標本共分散を計算でき,それを V_{ik} の共分散の推定値として利用できる.しかし,通常は $n_k < n_i$ であり,その場合,標本共分散行列は退化する.そこで,様々な確率分布を用いて共分散成分 $\sigma_{ii'}$ を計算し,最善の共分散成分を尤度に基づき選択する.本節では,2通りの共分散成分の計算法を提示する.

第1の方法が多項分布を使う方法である.今,環境組成 ζ_i を多項比率,および Z_{ik} を ζ_i の推定量と見なす.すると, Z_{ik} の共分散は直ちに

$$\text{cov}(Z_{ik}, Z_{i'k}) \propto \begin{cases} \zeta_i(1 - \zeta_i) & \text{if } i = i' \\ -\zeta_i \zeta_{i'} & \text{otherwise} \end{cases}$$

与えられる.これをそのまま $\sigma_{ii'}$ としてもよいが, ζ_i は未知のため計算手続きが複雑になる.そこで, ζ_i を \bar{z}_i で置き換え, $\sigma_{ii'}$ を

$$(3.4) \quad \sigma_{ii'} = \begin{cases} \bar{z}_i(1 - \bar{z}_i) & \text{if } i = i' \\ -\bar{z}_i \bar{z}_{i'} & \text{otherwise} \end{cases}$$

とする.ここに, $\bar{z}_i = \sum_{k'} z_{ik'} / n_k$ であり, z_{ik} は Z_{ik} の実現値を表す.

第2の方法は環境観測濃度 Y_{ik} の分散 $\sigma_{y_i}^2$ の値が指定された場合のための方法である.実際,CMBの適用例では,観測値 y_{ik} の平均値 $\bar{y}_i = \sum_{k'} y_{ik'} / n_k$ の一定割合という形で標準偏

差の値 $\hat{\sigma}_{y_i}$ を指定する場合が多い．そうした場合， Y_{ik} は独立に密度 $p(y_{ik}|\alpha_i, \beta_i)$ を持つ 2 母数非負確率分布に従うと仮定する．未知母数 α_i と β_i の推定値は例外を除き積率方程式

$$(3.5) \quad m_1(\alpha_i, \beta_i) = \bar{y}_i, \quad m_2(\alpha_i, \beta_i) = \hat{\sigma}_{y_i}^2$$

の解として得られる．ここに， $m_1(\alpha_i, \beta_i)$ と $m_2(\alpha_i, \beta_i)$ は各々 α_i と β_i の関数として表現された Y_{ik} の平均と分散を表す．一旦 α_i と β_i が推定されれば， $p(y_{ik}|\alpha_i, \beta_i)$ から Y_{ik} の乱数を生成するのは容易であり，生成された乱数を使い Z_{ik} の標本共分散 $\hat{\sigma}_{z_i z_i'}$ を計算できる．この標本共分散を用いて，

$$(3.6) \quad \sigma_{ii'} = \hat{\sigma}_{z_i z_i'}$$

とする．本稿では， $p(y_{ik}|\alpha_i, \beta_i)$ として打ち切り正規分布，対数正規分布，およびガンマ分布を採用する．

式(3.4)あるいは(3.6)で与えられる共分散成分の行列を $\sigma = \{\sigma_{ii'}\}_{i,i'=1,\dots,n_i-1}$ とする．この行列は \bar{z}_i あるいは標準偏差の値を平均値の一定割合という形で指定した場合の \bar{y}_i が一つでも 0 だと退化する．本稿では，退化を防ぐため，共分散成分の計算時にのみ \bar{z}_i あるいは \bar{y}_i の最小値を組成に換算して 0.0001 と定め，最小値を下回る場合には最小値を割り当てる．これは濃度が低い場合に観測精度が悪化するという事実にも対応している．

3.3 識別可能性

第 2 節で未確認発生源の組成と寄与を変数としてモデルに組み込めば変数は識別不能になると述べた．提案のモデル(3.3)には未確認発生源の組成と寄与率を変数として組み込まれている．本節では，変数の識別可能性について検討する．

モデル(3.3)は行列形式で以下のように書ける．

$$(3.7) \quad Z_k = x\theta + \omega(1 - \mathbf{1}_{n_j}\theta) + V_k, \quad V_k \sim \text{i.i.d.} N(\mathbf{0}_{n_i-1}, \delta^2\sigma), \quad k = 1, \dots, n_k$$

$$\mathbf{0}_{n_i-1} \leq \omega, \quad \mathbf{1}_{n_i-1}\omega \leq \mathbf{1}, \quad \mathbf{0}_{n_j} \leq \theta, \quad \mathbf{1}_{n_j}\theta < \mathbf{1}$$

ここに， $\mathbf{0}_n = (\underbrace{0, \dots, 0}_n)'$ ， $\mathbf{1}_n = (\underbrace{1, \dots, 1}_n)'$ ， $Z_k = (Z_{1k}, \dots, Z_{n_i-1,k})'$ ， $\omega = (\omega_1, \dots, \omega_{n_i-1})'$ ， $\theta = (\theta_1, \dots, \theta_{n_j})'$ ， $V_k = (V_{1k}, \dots, V_{n_i-1,k})'$ ，および x は x_{ij} を要素とする大きさ $n_i - 1 \times n_j$ の行列である．このモデルに対する尤度は以下のように書ける．

$$(3.8) \quad l(\omega, \theta, \delta) = (2\pi)^{-(n_i-1)n_k/2} \delta^{-(n_i-1)n_k} |\sigma|^{-n_k/2} \\ \times \exp \left[-\frac{1}{2\delta^2} \sum_{k=1}^{n_k} \{z_k - x\theta - \omega(1 - \mathbf{1}_{n_j}\theta)\}' \sigma^{-1} \{z_k - x\theta - \omega(1 - \mathbf{1}_{n_j}\theta)\} \right]$$

ここに， z_k は Z_k の実現値であり， $z = (z'_1, \dots, z'_{n_k})'$ としておく．これにより，対数尤度は

$$\log l(\omega, \theta, \delta) = -(n_i - 1)n_k \log \delta \\ - \frac{1}{2\delta^2} \sum_{k=1}^{n_k} \{z_k - x\theta - \omega(1 - \mathbf{1}_{n_j}\theta)\}' \sigma^{-1} \{z_k - x\theta - \omega(1 - \mathbf{1}_{n_j}\theta)\} + \text{const.}$$

と書け， ω と θ に関する尤度方程式は

$$(3.9) \quad \frac{\partial}{\partial \omega} \log l(\omega, \theta, \delta) \propto (1 - \mathbf{1}_{n_j}\theta) \sigma^{-1} \sum_{k=1}^{n_k} \{z_k - x\theta - \omega(1 - \mathbf{1}_{n_j}\theta)\} = \mathbf{0}_{n_i-1}$$

$$(3.10) \quad \frac{\partial}{\partial \theta} \log l(\omega, \theta, \delta) \propto (x - \omega \mathbf{1}_{n_j})' \sigma^{-1} \sum_{k=1}^{n_k} \{z_k - x\theta - \omega(1 - \mathbf{1}_{n_j}\theta)\} = 0_{n_j}$$

で与えられる．式(3.9)から

$$\sum_{k=1}^{n_k} \{z_k - x\theta - \omega(1 - \mathbf{1}_{n_j}\theta)\} = 0_{n_i-1}$$

が得られ，同時に，式(3.10)にも同じ因子が含まれているのを確認できる．従って，尤度方程式から得られる条件は

$$\omega = \sum_{k=1}^{n_k} (z_k - x\theta) / n_k(1 - \mathbf{1}_{n_j}\theta)$$

だけであり， θ については，

$$0_{n_j} \leq \theta, \quad \mathbf{1}_{n_j}\theta < 1, \quad \sum_{k=1}^{n_k} (z_k - x\theta) \geq 0_{n_i-1}$$

という自然な制約条件しか得られない．

3.4 事前分布

前節で確認したように， ω と θ の最尤推定値は一意に決まらない．それでも尤度(3.8)は依然として ω と θ の特定の値の当てはまりの良さを測るための有用な指標である．最尤推定が困難でも，尤度に基づき推定するのが望ましい．幸い，ベイズの定理を用いると，そうした推定を実現できる．そこで，ベイズの定理を用いるため，未知変数に事前分布を仮定する．仮定する分布は以下の密度を持つ一様分布である．

$$(3.11) \quad \begin{aligned} p(\omega) &= (n_i - 1)!, & 0_{n_i-1} \leq \omega, & \quad \mathbf{1}_{n_i-1}\omega \leq 1 \\ p(\theta) &= n_j!, & 0_{n_j} \leq \theta, & \quad \mathbf{1}_{n_j}\theta < 1 \\ p(\delta) &= 1/\hat{\delta}_{\max}, & 0 < \delta < \hat{\delta}_{\max} \end{aligned}$$

ここに， $\hat{\delta}_{\max}$ は十分に大きな値であり，本稿では，

$$Z_k = \mathbf{1}'_{n_i-1} \mu + V_k, \quad V_k \sim \text{i.i.d.} N(0_{n_i-1}, \delta_\mu^2 \sigma), \quad k = 1, \dots, n_k$$

を仮定した時に得られる最尤推定値 $\hat{\delta}_\mu^2$ を使い， $\hat{\delta}_{\max} = 10\hat{\delta}_\mu$ とする．事前分布として一様分布を仮定した効果については次節で述べる．

4. 推定

未知変数 ω ， θ ，および δ はベイズの定理を用いて推定でき，最善の共分散成分 $\sigma_{ii'}$ は(3.7)と(3.11)の双方に対する尤度を用いて選択できる．本節では，これらの手続きについて述べる．

モデル(3.7)と(3.11)を仮定する時，ベイズの定理により ω ， θ ，および δ の事後密度は以下のように与えられる．

$$(4.1) \quad p(\omega, \theta, \delta | z) = \frac{l(\omega, \theta, \delta) p(\omega) p(\theta) p(\delta)}{l(z)} \propto l(\omega, \theta, \delta)$$

ここに，

$$(4.2) \quad \begin{aligned} l(z) &= \int_{R(\omega, \theta, \delta)} l(\omega, \theta, \delta) p(\omega) p(\theta) p(\delta) d\omega d\theta d\delta \\ &= \frac{(n_i - 1)! n_j!}{\hat{\delta}_{\max}} \int_{R(\omega, \theta, \delta)} l(\omega, \theta, \delta) d\omega d\theta d\delta \end{aligned}$$

であり, $R(\omega, \theta, \delta)$ は ω, θ, δ の台を表す. 未知変数 ω, θ, δ を推定するには事後密度(4.1)に関する事後平均と事後分散を計算すればよい. ここで, $\lambda = (\lambda_1, \dots, \lambda_{n_i+n_j})' = (\omega', \theta', \delta)'$, $l(\lambda) = l(\omega, \theta, \delta)$, $p(\lambda|z) = p(\omega, \theta, \delta|z)$, および Λ_m を λ_m に対応する確率変数とすれば, Λ_m の事後平均は

$$(4.3) \quad E(\Lambda_m) = \int_{R(\lambda)} \lambda_m p(\lambda|z) d\lambda$$

と書け, 事後分散は

$$\text{var}(\Lambda_m) = E(\Lambda_m^2) - E^2(\Lambda_m)$$

により計算できる. ここに,

$$(4.4) \quad E(\Lambda_m^2) = \int_{R(\lambda)} \lambda_m^2 p(\lambda|z) d\lambda$$

である. 事前分布として一様分布を仮定したため, 式(4.1)に記述したように, 事後密度 $p(\lambda|z)$ は尤度 $l(\lambda)$ に比例する. そのため, 事後平均(4.3)は, 尤度 $l(\lambda)$ を重みに持つ λ_m の重み付き平均と解釈でき, 尤度を直接反映した推定値となる.

一方, 正規化定数(4.2)はモデル(3.7)と(3.11)の双方に対する尤度と考えられており(例えば, Akaike, 1980 を参照), ベイズ型尤度あるいは積分尤度とも呼ばれている. この尤度を用いれば, 最善の共分散成分を選択できる. すなわち, 第 3.2 節で述べた多項分布, 打ち切り正規分布, 対数正規分布, およびガンマ分布を仮定して計算される共分散成分の内, ベイズ型尤度が最大になる共分散成分を最善の共分散成分として選択する.

5. 計算

推定手続きは多重積分(4.2)(4.3), および(4.4)により構成されている. 本節では, これらの多重積分を計算するためのマルコフ連鎖モンテカルロ(例えば, Gilks et al., 1996 を参照)について説明する.

5.1 事後平均と事後分散

多重積分(4.3)と(4.4)は以下のモンテカルロ積分により近似的に計算できる.

$$\int_{R(\lambda)} \lambda_m p(\lambda|z) d\lambda \approx \frac{1}{n_r} \sum_{r=1}^{n_r} \lambda_m(r), \quad \int_{R(\lambda)} \lambda_m^2 p(\lambda|z) d\lambda \approx \frac{1}{n_r} \sum_{r=1}^{n_r} \lambda_m^2(r)$$

ここに, $\lambda_m(r)$ は $p(\lambda|z)$ から生成された Λ_m の第 r 番目の乱数, および n_r は乱数の数を表す. 乱数を生成するため, $p(\lambda|z)$ を定常分布とするマルコフ連鎖を single-component Metropolis-Hastings(SCMH) (Metropolis et al., 1953; Hastings, 1970)により構成する.

SCMH は主として 2 つの手続きから成り立っている. 第 1 が提案密度 $q_m(\lambda'_m(r+1)|\lambda_m(r), \lambda_{-m}(r))$ からの乱数 $\lambda'_m(r+1)$ の生成である. ここに, $\lambda_{-m}(r) = (\lambda_1(r+1), \dots, \lambda_{m-1}(r+1), \lambda_{m+1}(r), \dots, \lambda_{n_i+n_j}(r))'$ である. 第 2 が $\lambda_m(r+1)$ の候補である乱数 $\lambda'_m(r+1)$ の取捨選択である. その採択率は

$$(5.1) \quad \eta_m(r+1) = \min\left(1, \frac{q_m(\lambda_m(r)|\lambda'_m(r+1), \lambda_{-m}(r))p(\lambda'_m(r+1)|\lambda_{-m}(r), z)}{q_m(\lambda'_m(r+1)|\lambda_m(r), \lambda_{-m}(r))p(\lambda_m(r)|\lambda_{-m}(r), z)}\right)$$

により計算される．ここに， $p(\lambda'_m(r+1)|\lambda_{-m}(r), z)$ および $p(\lambda_m(r)|\lambda_{-m}(r), z)$ は

$$(5.2) \quad p(\lambda_m|\lambda_{-m}, z) = p(\lambda|z) \Big/ \int_{R(\lambda_m)} p(\lambda|z) d\lambda_m = l(\lambda) \Big/ \int_{R(\lambda_m)} l(\lambda) d\lambda_m$$

で定義される完全条件付き密度，および $\lambda_{-m} = (\lambda_1, \dots, \lambda_{m-1}, \lambda_{m+1}, \dots, \lambda_{n_i+n_j})'$ である．これらの手続きを初期値 $\lambda_m(0)$ から始めて交互に実行する．ただし， $\lambda_m(0)$ を $p(\lambda|z)$ から生成するのは困難なため，初期の $\lambda_m(r)$ は目的の乱数として使用しない．この過程を burn-inn と呼ぶ．以上が SCMH の概要であるが，方法の詳細は提案密度により決まる．

本稿では， $\lambda_m (m = 1, \dots, n_i + n_j - 1)$ の提案密度として Gibbs 法 (Geman and Geman, 1984) を採用する．すなわち，

$$q_m(\lambda'_m(r+1)|\lambda_m(r), \lambda_{-m}(r)) = p(\lambda'_m(r+1)|\lambda_{-m}(r), z)$$

と仮定する．この場合，採択率は常に 1 になる．完全条件付き密度は (3.8) と (5.2) から，分布の形で以下のように与えられる．

$$(5.3) \quad \Lambda_m \sim \begin{cases} TN \left(\frac{1}{\tau_{mm}} \left\{ \frac{(\bar{z} - \mathbf{x}\theta)' \tau_m}{(1 - \mathbf{1}_{n_j} \theta)} - \omega'_{-m} \tau_{-m} \right\}, \frac{\delta^2}{n_k \tau_{mm} (1 - \mathbf{1}_{n_j} \theta)^2} \right) \\ \quad \quad \quad 0 \leq \lambda_m \leq 1 - \mathbf{1}_{n_i-2} \omega_{-m}, \quad 1 \leq m \leq n_i - 1 \\ TN \left(\frac{\left\{ \bar{z} - \mathbf{x}_{-(m-n_i+1)} \theta_{-(m-n_i+1)} - \omega_{(1-\mathbf{1}_{n_j}-1)\theta_{-(m-n_i+1)}} \right\}' \sigma^{-1} (\mathbf{x}_{m-n_i+1} - \omega)}{(\mathbf{x}_{m-n_i+1} - \omega)' \sigma^{-1} (\mathbf{x}_{m-n_i+1} - \omega)} \right), \\ \quad \quad \quad \frac{\delta^2}{n_k (\mathbf{x}_{m-n_i+1} - \omega)' \sigma^{-1} (\mathbf{x}_{m-n_i+1} - \omega)} \\ \quad \quad \quad 0 \leq \lambda_m < 1 - \mathbf{1}_{n_j-1} \theta_{-(m-n_i+1)}, \quad n_i \leq m \leq n_i + n_j - 1 \end{cases}$$

ここに， $TN(\mu, \sigma^2)$ は正規分布 $N(\mu, \sigma^2)$ から派生した打ち切り正規分布， $\bar{z} = (\bar{z}_1, \dots, \bar{z}_{n_i-1})'$ ， $\omega_{-i} = (\omega_1, \dots, \omega_{i-1}, \omega_{i+1}, \dots, \omega_{n_i-1})'$ ， $\theta_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_{n_j})'$ ， τ_{ij} は σ^{-1} の第 (i, j) 番目の要素， τ_i は σ^{-1} の第 i 列， $\tau_{-i} = (\tau_{1i}, \dots, \tau_{i-1,i}, \tau_{i+1,i}, \dots, \tau_{n_i-1,i})'$ ， \mathbf{x}_j は \mathbf{x} の第 j 列，および \mathbf{x}_{-j} は \mathbf{x} の第 j 列を除いた行列を表す．完全条件付き密度 (5.3) の λ_{-m} に $\lambda_{-m}(r)$ を代入すれば提案密度が得られる．

一方， $\lambda_{n_i+n_j}$ については，乱数の生成を容易にするため，酔歩法を採用する．すなわち，

$$q_{n_i+n_j}(\lambda'_{n_i+n_j}(r+1)|\lambda_{n_i+n_j}(r), \lambda_{-(n_i+n_j)}(r)) = 1/2a$$

$$\lambda_{n_i+n_j}(r) - a < \lambda'_{n_i+n_j}(r+1) < \lambda_{n_i+n_j}(r) + a$$

と仮定する．この場合，採択率 (5.1) は

$$\eta_{n_i+n_j}(r+1) = \begin{cases} \min \left(1, \frac{l(\lambda_{-(n_i+n_j)}(r), \lambda'_{n_i+n_j}(r+1))}{l(\lambda_{-(n_i+n_j)}(r), \lambda_{n_i+n_j}(r))} \right), & 0 < \lambda'_{n_i+n_j}(r+1) < \hat{\delta}_{\max} \\ 0, & \text{otherwise} \end{cases}$$

と書き換えられる．帯域幅 $2a$ は実際の採択率が 0.45 以上 0.55 以下になるよう調整する．この調整過程は，高い精度を必要としないため，burn-inn としても利用する．

5.2 ベイズ型尤度

ベイズ型尤度 (4.2) は Ogata-Tanemura 法 (例えば，Ogata, 1989 を参照) により近似的に計算できる．今， $\varepsilon (0 \leq \varepsilon \leq 1)$ を尺度変数， $f(\lambda, \varepsilon)$ を $f(\lambda, 1) = l(\lambda)$ となる関数，および

$$p(\lambda|\varepsilon) = \frac{f(\lambda, \varepsilon)}{m(\varepsilon)}, \quad m(\varepsilon) = \int_{R(\lambda)} f(\lambda, \varepsilon) d\lambda$$

とし, $m(0)$ は既知とする. Ogata (1989) はベイズ型対数尤度を

$$\begin{aligned} \log \int_{R(\lambda)} l(\lambda) d\lambda &= \log m(0) + \int_0^1 \left[\int_{R(\lambda)} \left\{ \frac{\partial}{\partial \varepsilon} \log f(\lambda, \varepsilon) \right\} p(\lambda|\varepsilon) d\lambda \right] d\varepsilon \\ &= \log m(0) + \int_0^1 g(\varepsilon) d\varepsilon \end{aligned}$$

と書き, 最後の項の積分を台形公式

$$\int_0^1 g(\varepsilon) d\varepsilon \approx \frac{1}{2} \sum_{s=1}^{n_s} (\varepsilon_s - \varepsilon_{s-1}) \{g(\varepsilon_s) + g(\varepsilon_{s-1})\}$$

により計算するよう提案した. ここに, $0 = \varepsilon_0 < \varepsilon_1 < \dots < \varepsilon_{n_s} = 1$ であり, 各 $g(\varepsilon_s)$ はマルコフ連鎖モンテカルロにより計算する.

この方法を用いるため, 本稿では,

$$f(\lambda, \varepsilon) = \{l(\lambda)\}^{\varepsilon^b}$$

と定義する. ここに, b は台形公式による積分が原点付近で不安定になるのを防ぐための定数で, 本稿では $b=3$ を採用する. この時, $m(0) = \hat{\delta}_{\max} / (n_i - 1)! n_j!$ であり, $g(\varepsilon_s)$ のモンテカルロ積分は

$$\begin{aligned} g(\varepsilon_s) &= \int_{R(\lambda)} \left\{ \frac{\partial}{\partial \varepsilon_s} \log f(\lambda, \varepsilon_s) \right\} p(\lambda|\varepsilon_s) d\lambda \\ &\approx \frac{b\varepsilon_s^{b-1}}{n_r(n_i + n_j)} \sum_{r=0}^{n_r-1} \sum_{m=1}^{n_i+n_j} \log l(\lambda_m(r+1), \lambda_{-m}(r)) \end{aligned}$$

と書ける. 乱数 $\lambda_m(r+1)$ を生成するには, 完全条件付き密度(5.3)の δ^2 を δ^2/ε^b に置き換え, 採択率(5.4)を

$$\eta_{n_i+n_j}(r+1) = \begin{cases} \min\left(1, \frac{f(\lambda'_{n_i+n_j}(r+1), \lambda_{-(n_i+n_j)}(r), \varepsilon_s)}{f(\lambda_{n_i+n_j}(r), \lambda_{-(n_i+n_j)}(r), \varepsilon_s)}\right), & 0 < \lambda'_{n_i+n_j}(r+1) < \hat{\delta}_{\max} \\ 0, & \text{otherwise} \end{cases}$$

と書き換えれば, 前節で述べた SCMH を利用できる.

6. シミュレーション

本節では, 提案の方法の性能を調べるために企画したシミュレーションの結果について述べる. 計算に先立ち, 環境濃度の真値 ψ_i を PCB データを用いて定義した. 日本では, PCB 汚染が依然として深刻な問題だからである. 実際, 飯村 他(2002)は, 東京湾に生息する魚類やプランクトンに含まれるダイオキシン類の内, 毒性等量に換算して 70~92% がダイオキシン様 PCBs であると報告している. 柏木 他(2002)は, これらダイオキシン様 PCBs の 90% 以上が PCB 製品に由来していると推定した. 日本で最も広く利用された PCB 製品が KC300, KC400, KC500, および KC600 を主力とするカネクロールと呼ばれる製品である. 高菅 他(1995)は上記 4 つの KC 製品に含まれる PCBs の全異性体を測定している. 表 1 に主要異性体の組成を示す. 表の数値を尺度変換された数値と見なし, 環境濃度の真値 ψ_i ($i=1, \dots, 24$) を以下のように定義した.

$$\psi_i = .4 \times (\text{KC300})_i + .3 \times (\text{KC400})_i + .2 \times (\text{KC500})_i + .1 \times (\text{KC600})_i$$

表 1. カネクロール中 PCBs の主要異性体の組成(単位 %, 高菅 他(1995)から引用).

i	名称	KC300	KC400	KC500	KC600
1	#8/5	7.105	0.204	0.179	0.099
2	#18	7.752	2.656	0.262	0.096
3	#31	9.787	5.032	0.318	0.114
4	#28	10.611	3.722	0.296	0.118
5	#33/20	7.423	1.887	0.217	0.085
6	#52	2.245	5.811	3.483	0.214
7	#44	2.029	4.114	1.051	0.085
8	#70	3.301	7.868	2.391	0.164
9	#66	3.084	6.069	0.855	0.162
10	#95	0.519	2.541	6.164	1.922
11	#101	0.740	3.434	10.204	2.677
12	#110	1.514	4.300	10.256	1.278
13	#118	0.665	3.658	8.391	0.678
14	#105	0.349	2.234	2.952	0.240
15	#149	0.145	0.367	4.776	9.230
16	#153	0.193	0.313	5.296	10.585
17	#138	0.265	0.601	7.895	7.452
18	#187	0.000	0.016	0.304	6.623
19	#174	0.000	0.010	0.368	4.507
20	#180	0.022	0.039	0.714	10.032
21	#170	0.010	0.032	0.478	2.469
22	#199	0.000	0.000	0.013	1.764
23	#203/196	0.000	0.000	0.029	2.002
24	#194	0.000	0.000	0.022	1.362

表 2. 想定した 5 通りの場合 .

	KC300	KC400	KC500	KC600
Case 0	同定済み	同定済み	同定済み	同定済み
Case 1	未確認	同定済み	同定済み	同定済み
Case 2	同定済み	未確認	同定済み	同定済み
Case 3	同定済み	同定済み	未確認	同定済み
Case 4	同定済み	同定済み	同定済み	未確認

そして模擬データ y_{ik} ($k=1, \dots, n_k$) を

$$y_{ik} = \max\{\psi_i + e_{ik}, 0\}$$

により生成した．ここに, e_{ik} は平均 0, 分散 $(.1 \times \psi_i)^2$ の正規乱数である．すなわち, いわゆる 10% 誤差を真値に加えた．仮定した n_k は 1, 2, 5, および 10 であり, 試行回数は 10 である．これらの条件に基づき 40 のデータセットを作成した．各データセットに対しては, 表 2 に示すように, 同定された発生源が異なる 5 通りの場合を想定した．マルコフ連鎖モンテカルロで生成する乱数の数 n_r は, 帯域幅の調整時が 1 万, ベイズ型尤度の計算時が 25 万, および事後平均と事後分散の計算時が 400 万である．また, Ogata-Tanemura 法における n_s は 128 であり, 節点 ε_s は等間隔とした．そして共分散成分(3.6)の計算時には $\hat{\sigma}_{y_i}$ を \bar{y}_i の 10% と仮定した．

表 3. 10 回の試行で最善と判断された回数.

n_k	Case	多項分布	打ち切り正規分布	対数正規分布	ガンマ分布
1	0	8	0	1	1
	1	4	1	3	2
	2	5	1	2	2
	3	5	4	1	0
	4	10	0	0	0
2	0	1	5	2	2
	1	1	2	5	2
	2	1	2	3	4
	3	1	3	3	3
	4	3	3	3	1
5	0	3	5	0	2
	1	3	5	2	0
	2	3	4	1	2
	3	3	2	1	4
	4	3	3	1	3
10	0	2	6	0	2
	1	2	3	2	3
	2	2	4	1	3
	3	2	2	3	3
	4	2	3	2	3

表 3 に共分散成分の計算に用いた 4 つの確率分布の各々が最善と判断された回数を示す. 試行回数が小さいため, はっきりした傾向は見られないが, 観測の繰り返し数 n_k が 1 の時, 多項分布が優越している. 恐らく, 積率方程式 (3.5) により未知母数を許容精度で推定できなかったためと, 多項分布の場合には共分散成分の計算に 2 次モーメントの値 $\hat{\sigma}_{y_i}$ を必要としないためと考えられる. また, n_k が 5 以上の時, 打ち切り正規分布が僅かに優越している. 打ち切り正規乱数により誤差を生成しているため合理的な傾向であるが, 打ち切り正規分布, 対数正規分布, およびガンマ分布に対するベイズ型対数尤度の値に大きな差は無い.

図 1 から 4 に発生源寄与率の推定値を示す. 図中, 丸印は各試行毎の事後平均, 垂直の点線は各試行毎の事後分散に基づく $\pm 2\sigma$ の区間, および水平の実線は真の値を示す. 図 1, すなわち $n_k = 1$ の時, いくつかの場合に, 事後平均と真の値との乖離は非常に大きく, $\pm 2\sigma$ 区間もかなり長い. しかも, 第 2 と第 5 番目のデータセットについては, 5 つの Case 全てにおいて推定に失敗している. 実際, $\pm 2\sigma$ 区間は 1 を含んでいる. 未確認発生源の寄与率が 1 になると, モデル (3.3) は

$$Z_{ik} = \omega_i + V_{ik}$$

に縮退し, 自明な解しか得られない無意味なモデルとなる. ただし, こうした失敗がある一方で, 尤もらしい推定値が得られている場合も少なからず存在する. これは驚くべき結果と言える. なぜなら, n_k が 1 の時, モデル (3.3) に含まれる未知変数の数はデータの大きさより大きいからである.

観測の繰り返し数が 2 以上になると, 推定精度は目立って改善する. 実際, $n_k = 2$ の時でさえ, 図 2 から分かるように, 事後平均と真の値との乖離は顕著に小さくなり, $\pm 2\sigma$ 区間も顕著に短くなる. 図 1 で見られた明らかな推定の失敗も見当たらない. こうした傾向は n_k が大きくなるに従い明確になる. 実際, 図 3 と 4 に示すように, n_k が 5 以上になると, 事後平均は真の値にかなり近接する. 依然として明らかな偏りは見られるが, 提案の方法は n_k が 5 前後

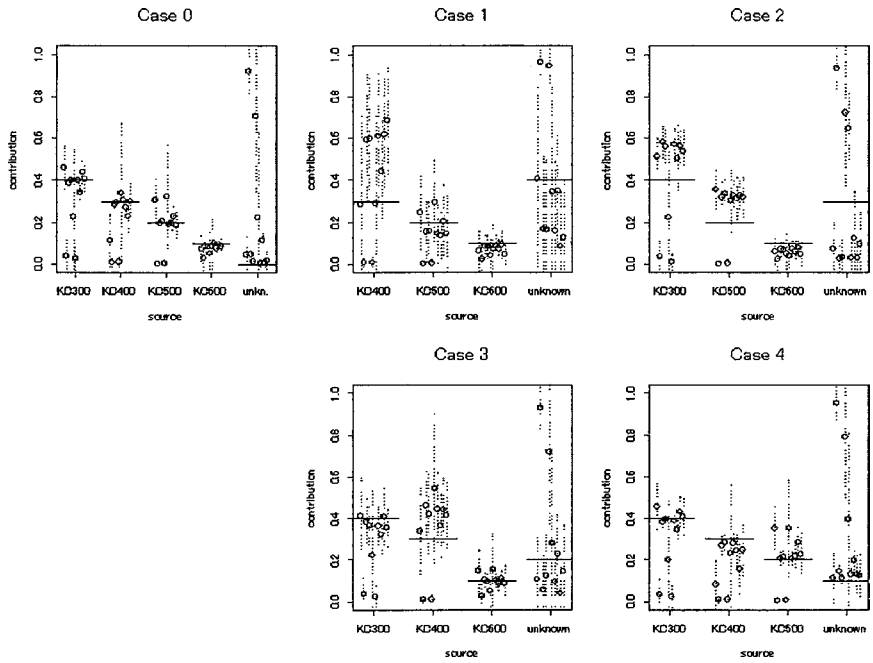


図 1. 発生源寄与率の推定値($n_k = 1$). 丸印は試行毎の事後平均, 垂直の点線は試行毎の事後分散に基づく $\pm 2\sigma$ の区間, および水平の実線は真の値を表す.

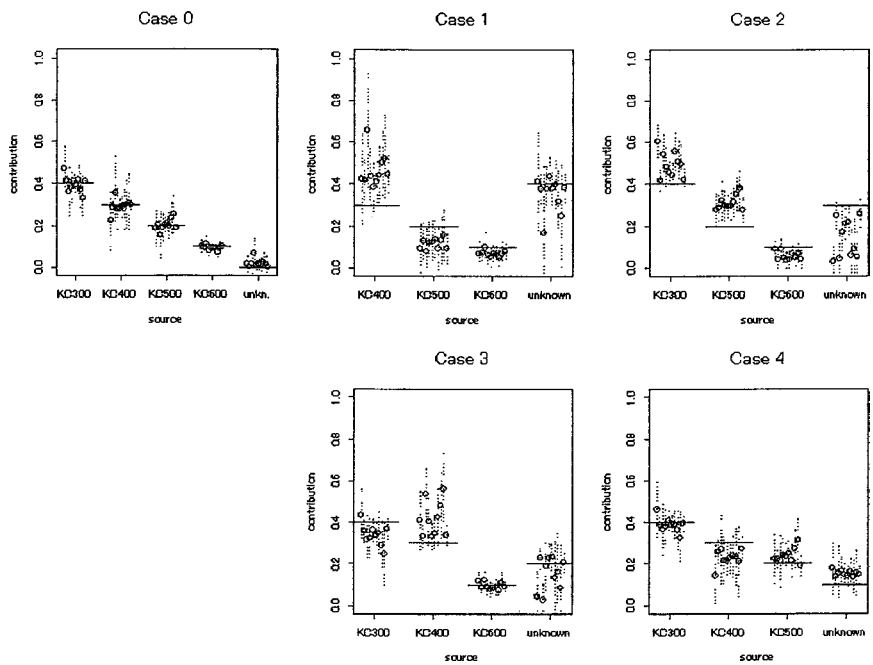


図 2. 発生源寄与率の推定値($n_k = 2$).

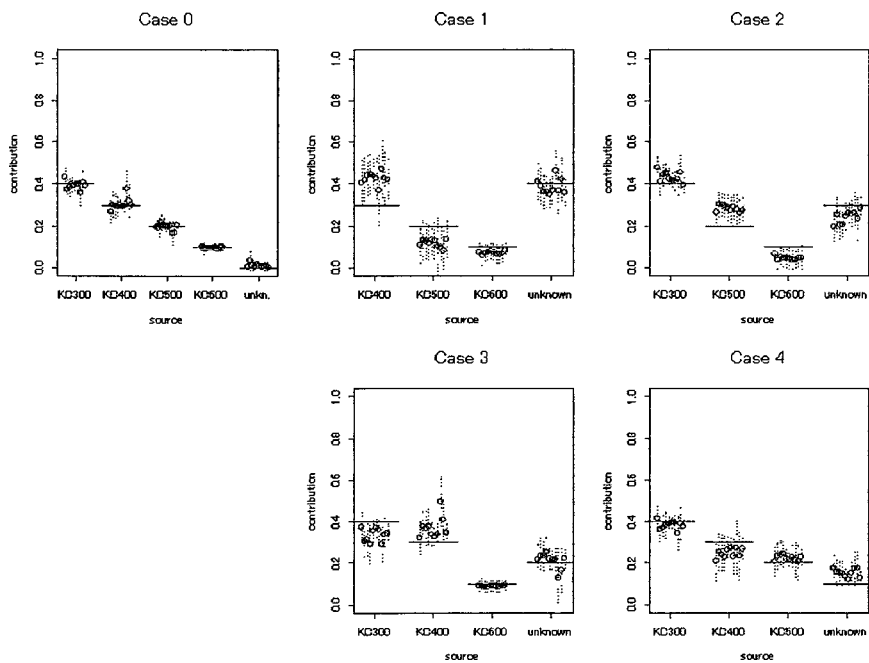


図 3. 発生源寄与率の推定値($n_k = 5$).

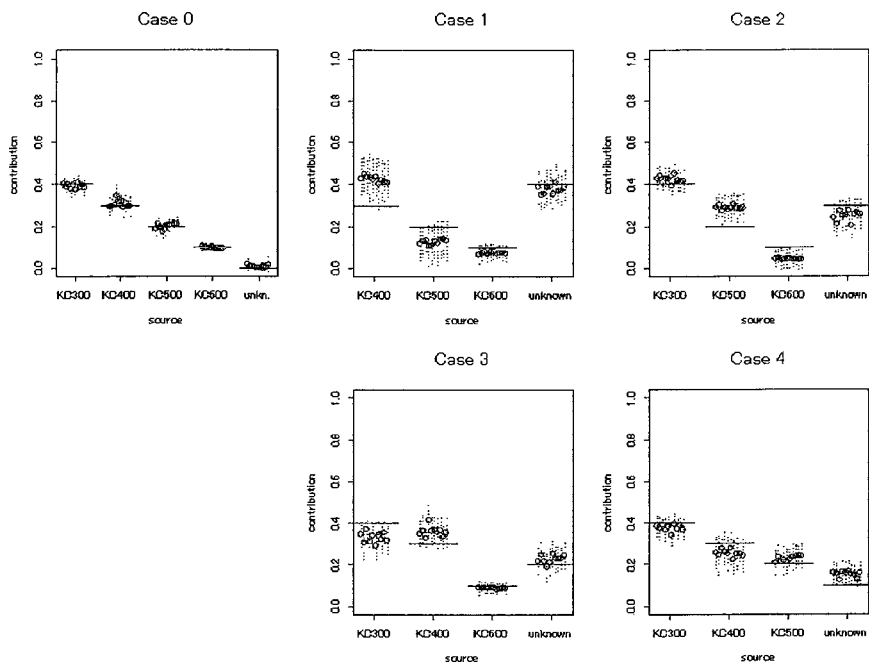


図 4. 発生源寄与率の推定値($n_k = 10$).

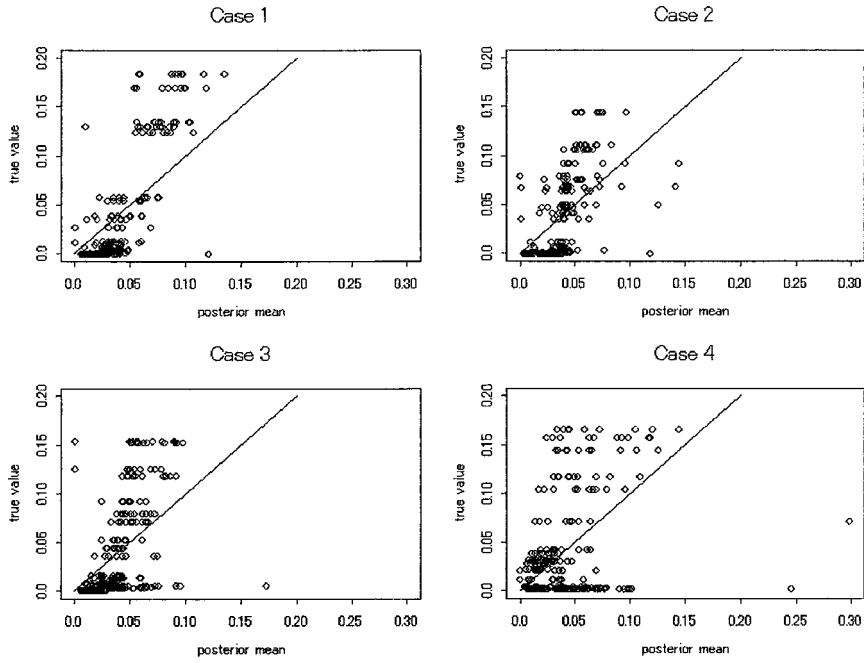


図 5. 未確認発生源の組成の事後平均と真の値の相関図($n_k = 1$). 図中の斜めの線は傾き 1 の直線を表す.

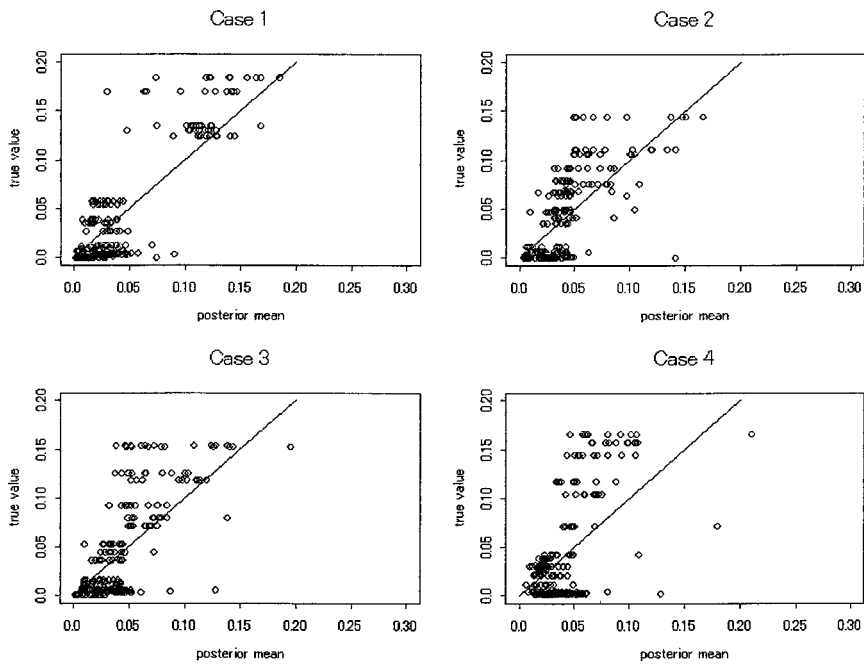


図 6. 未確認発生源の組成の事後平均と真の値の相関図($n_k = 2$).

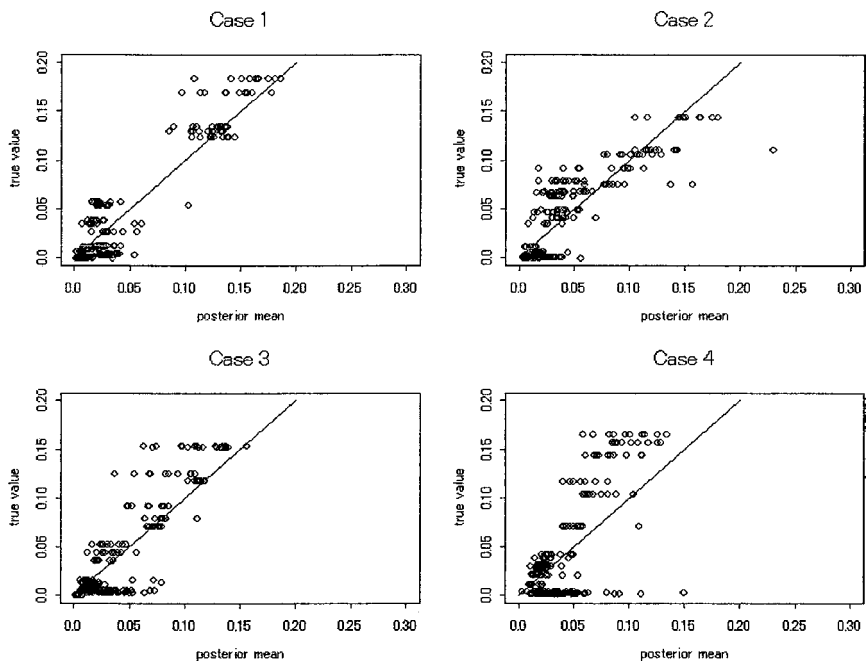


図 7. 未確認発生源の組成の事後平均と真の値の相関図($n_k = 5$).

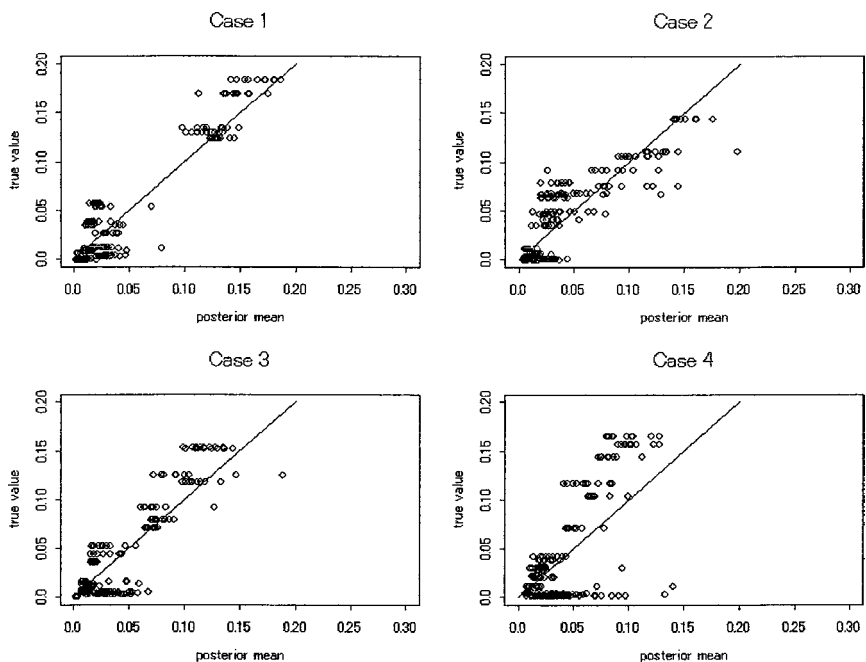


図 8. 未確認発生源の組成の事後平均と真の値の相関図($n_k = 10$).

より大きい時に十分実用的と考えられる。

提案の方法の目的は未確認発生源の寄与率の推定にあるが、付随して未確認発生源の組成も推定できるので、結果を示しておく。図 5 から 8 に Case 0 を除く未確認発生源の組成の事後平均と真の値の相関図を示す。横軸が事後平均、縦軸が真の値、および図中の斜めの線は傾き 1 の直線を表す。図 1 の結果から容易に想像できるように、 $n_k = 1$ の時の推定は不正確である。実際、図 5 に示すように、多くの丸印が斜めの線から遠く離れている。すなわち、事後平均と真の値の乖離は大きい。乖離の程度は Case 番号に伴い大きくなっている。これは未確認発生源に仮定した寄与率に原因がある。また、真の値が 0 に近い場合に事後平均が過大評価になり、真の値が 0 から離れるに従い事後平均が過小評価になる傾向がある。これは事前分布として一様分布を仮定したためである。ただし、事後平均が全くランダムでない点は注目に値する。改善の余地が残されていると考えられるからである。

発生源寄与率の場合と同様、未確認発生源の組成の推定は n_k が大きくなるに従い改善される。実際、図 6 から図 8 を見ると分かる通り、系統的な過大評価や過小評価の傾向は n_k の増大に伴い徐々に緩慢になる。特に $n_k = 10$ の時、Case 4 を除き丸印は斜めの線に十分近接している。Case 4 における推定が相対的に不正確なのは、未確認発生源の寄与率が小さいためである。実際、Case 4 で仮定した未確認発生源寄与率は 0.1 であり、測定誤差としていわゆる 10% 誤差を仮定した。また、寄与率が小さい場合、組成が尤度に与える影響は小さくなる。これらの点を考慮すると、Case 4 の結果は許容可能と言える。

7. 適用例

シミュレーションの結果、観測の繰り返し数 n_k がある程度大きければ、提案の方法により未確認発生源の寄与率を実用的に許容できる精度で推定できるのが分かった。そこで、 n_k がある程度大きい実際の事例を選び、未確認発生源の寄与率の推定を試みた。本節ではその結果について述べる。

7.1 海域底質

市原市が平成 12 年度に実施した行政調査により、市原港の底質から極めて高濃度のダイオキシン類が検出され、問題化した。その後、千葉県が実施した詳細な調査により、表層泥が広範囲にわたって汚染され、市原港開設時期にあたる昭和 35 年前後より後に形成された堆積層も表層に至るまで同様に汚染されているのが判明した。新たな継続的大量流入は認められないが、港内には既に相当量のダイオキシン類が堆積しており、汚染規模は国内最大級と言える(詳細については、例えば、吉澤 他, 2004 を参照)。港内の環境を修復するため、早急に対策を立てる必要がある。最大の問題は環境修復に膨大な費用が掛かると予想される点である。その費用負担を巡り、発生源の同定と寄与率の算定が求められている。

千葉県と市原市の調査結果によれば、底質のダイオキシン類濃度は、港の奥から中央部にかけて高く、港内に流入する河川では桁違いに低かった。そのため、発生源は港周辺に存在すると推察された。一方、港内底質のダイオキシン類の組成を見ると、7, 8 塩素の PCDDs と PCDFs が優越し、フェノール法で製造された農薬 pentachlorophenol (PCP) に含まれるダイオキシン類の組成に類似していた。港に隣接する事業所を対象にアンケート調査を実施したところ、ある事業所から過去にフェノール法で PCP を製造していたとの回答を得た。ただし、製造していた PCP や反応残渣、排水等の産業廃棄物に含まれていたダイオキシン類の組成は不明である。そこで、提案の方法と既知のダイオキシン類組成を用い、港内底質と PCP のダイオキシン類組成の類似性について検討した。

方法の適用に際し、千葉県が平成 12 年度に測定した結果の内、港中央近くに位置する隣接

した 5 地点の表層泥の測定値を環境組成とし、橋本(2004)が収集、分類した農薬中ダイオキシン類組成の内、PCP の典型組成 4 種類を発生源組成とした。表層泥の平均組成と PCP の典型組成を表 4 に示す。PCP1~3 がフェノール法により製造された PCP である。PCP 以外にも様々な発生源を想定できるが、それらを網羅するのは極めて困難なため、全部をまとめてひとつの未確認発生源と仮定した。

表 5 に発生源寄与率の推定結果を示す。未確認発生源の寄与率の事後平均は僅かに 0.00024 であり、PCP2 と PCP3 の寄与率の事後平均が合計で 0.999 となった。表層泥とフェノール法により製造された PCP のダイオキシン類組成は極めて類似しているという結果である。発生源を絞り込む大きな手掛かりと言える。ただし、表 4 を見ると分かるように、表層泥と仮定した PCP の組成には若干の乖離がある。乖離の程度は未確認発生源の存在を示唆する程大きく

表 4. 海域底質と PCP における PCDD/Fs の同族・異性体組成。

名称	表層泥(平均)	PCP1	PCP2	PCP3	PCP4
D1368	0.00150	0.00117	0.00301	0.00045	0.00182
D1379	0.00060	0.00060	0.00110	0.00021	0.00043
D2378	0.00001	0.00014	0.00000	0.00001	0.00011
D12378	0.00002	0.00082	0.00002	0.00008	0.00195
D123478	0.00006	0.00970	0.00012	0.00026	0.03347
D123678	0.00131	0.00279	0.00455	0.00038	0.00487
D123789	0.00009	0.00068	0.00011	0.00022	0.00517
D1234678	0.05743	0.08861	0.04989	0.01924	0.33468
OCDD	0.69954	0.46136	0.77151	0.93014	0.00000
F2378	0.00001	0.00009	0.00000	0.00000	0.00061
F12378	0.00031	0.00046	0.00004	0.00005	0.00280
F23478	0.00002	0.00021	0.00003	0.00002	0.00088
F123478	0.00076	0.00269	0.00095	0.00027	0.00280
F123678	0.00038	0.00094	0.00026	0.00002	0.00113
F123789	0.00001	0.00004	0.00002	0.00000	0.00000
F234678	0.00006	0.00074	0.00066	0.00002	0.00170
F1234678	0.01224	0.02944	0.01157	0.00212	0.02677
F1234789	0.00230	0.00410	0.00235	0.00054	0.00000
OCDF	0.10489	0.17345	0.08335	0.01955	0.00000
TeCDDs	0.00007	0.02723	0.00007	0.00088	0.28908
PeCDDs	0.00061	0.01330	0.00037	0.00079	0.16806
HxCDDs	0.00204	0.00855	0.00350	0.00211	0.02945
HpCDDs	0.02147	0.02418	0.02225	0.01012	0.03036
TeCDFs	0.00170	0.00039	0.00013	0.00017	0.00759
PeCDFs	0.00222	0.00603	0.00077	0.00040	0.00665
HxCDFs	0.02909	0.02814	0.00984	0.00141	0.01867
HpCDFs	0.06126	0.11413	0.03356	0.01054	0.03096

表 5. 海域底質における発生源寄与率の事後平均と標準偏差。

発生源	PCP1	PCP2	PCP3	PCP4	未確認
事後平均	0.00029	0.65117	0.34827	0.00003	0.00024
標準偏差	0.00030	0.04581	0.04587	0.00003	0.00033

ないという推定結果であるが、本件の発生源の同定は環境修復の責任を負うべき原因者の特定とも深く関連するため、慎重に推論する必要がある。そのため、現在、より詳細な解析を実施している。結果については稿を改めて述べる。

7.2 河川底質

新潟県内の河川の底質から比較的高濃度のダイオキシン類が検出された。近接する他の地点の濃度は低く、一過性の現象と考えられた。また、組成を見ると、8塩素のPCDD(OCDD)が約8割を占め、前節の場合と同様、主な発生源はPCPと推察された。実際、新潟県では過去

表 6. 河川底質, CNP および排ガスにおける PCDD/Fs の同族・異性体組成。

名称	底質(平均)	CNP1	CNP2	CNP3	CNP4	排ガス
D1368	0.00472	0.62967	0.50648	0.69298	0.75137	0.02181
D1379	0.00214	0.23303	0.17139	0.28896	0.22712	0.01341
D2378	0.00008	0.00009	0.00005	0.00000	0.00001	0.00171
D12378	0.00041	0.00180	0.00309	0.00010	0.00191	0.00302
D123478	0.00095	0.00008	0.00012	0.00000	0.00007	0.00305
D123678	0.00228	0.00154	0.00206	0.00055	0.00489	0.00425
D123789	0.00234	0.00059	0.00074	0.00021	0.00170	0.00380
D1234678	0.05583	0.00038	0.00061	0.00008	0.00181	0.01628
OCDD	0.80724	0.00000	0.00000	0.00000	0.00002	0.04245
F2378	0.00000	0.00007	0.00004	0.00000	0.00001	0.01167
F12378	0.00011	0.00009	0.00011	0.00001	0.00008	0.01312
F23478	0.00005	0.00011	0.00022	0.00000	0.00009	0.01467
F123478	0.00067	0.00001	0.00001	0.00000	0.00000	0.01370
F123678	0.00034	0.00005	0.00007	0.00002	0.00004	0.01321
F123789	0.00002	0.00000	0.00000	0.00000	0.00000	0.00080
F234678	0.00024	0.00139	0.00371	0.00004	0.00148	0.01234
F1234678	0.00421	0.00005	0.00010	0.00000	0.00000	0.03042
F1234789	0.00079	0.00000	0.00000	0.00000	0.00000	0.00298
OCDF	0.01338	0.00000	0.00000	0.00000	0.00000	0.02461
TeCDDs	0.00139	0.03837	0.03426	0.00413	0.00000	0.04334
PeCDDs	0.00437	0.06379	0.23375	0.00750	0.00772	0.04691
HxCDDs	0.01911	0.00265	0.00887	0.00020	0.00024	0.05745
HpCDDs	0.05782	0.00001	0.00004	0.00000	0.00003	0.01676
TeCDFs	0.00127	0.02175	0.02350	0.00514	0.00125	0.32665
PeCDFs	0.00246	0.00413	0.00987	0.00008	0.00011	0.17310
HxCDFs	0.00733	0.00034	0.00088	0.00000	0.00004	0.07642
HpCDFs	0.01044	0.00000	0.00001	0.00000	0.00000	0.01207

表 7. 河川底質における発生源寄与率の事後平均と標準偏差。

発生源	CNP1	CNP2	CNP3	CNP4	PCP1
事後平均	0.00152	0.00197	0.00148	0.00135	0.00118
標準偏差	0.00126	0.00160	0.00119	0.00109	0.00116
発生源	PCP2	PCP3	PCP4	排ガス	未確認
事後平均	0.00303	0.82257	0.00113	0.00049	0.16527
標準偏差	0.00297	0.02250	0.00067	0.00037	0.02170

表 8. 河川底質における未確認発生源の組成の事後平均と標準偏差 .

名称	D1368	D1379	D2378	D12378	D123478	D123678	D123789
事後平均	0.00525	0.00463	0.00051	0.00230	0.00487	0.01330	0.01470
標準偏差	0.00448	0.00186	0.00006	0.00028	0.00062	0.00161	0.00176
名称	D1234678	OCDD	F2378	F12378	F23478	F123478	F123678
事後平均	0.27175	0.10277	0.00004	0.00043	0.00021	0.00310	0.00215
標準偏差	0.03070	0.09290	0.00003	0.00006	0.00005	0.00038	0.00026
名称	F123789	F234678	F1234678	F1234789	OCDF	TeCDDs	PeCDDs
事後平均	0.00021	0.00136	0.01710	0.00255	0.00114	0.00186	0.02060
標準偏差	0.00004	0.00017	0.00217	0.00035	0.00110	0.00117	0.00332
名称	HxCDDs	HpCDDs	TeCDFs	PeCDFs	HxCDFs	HpCDFs	
事後平均	0.11778	0.33388	0.00615	0.01390	0.04180	0.01570	
標準偏差	0.01380	0.03730	0.00104	0.00170	0.00502	0.00355	

表 9. 大気と PCB 発生源におけるダイオキシン様 PCBs の異性体組成 .

名称	大気(平均)	KC300	KC400	KC500	KC600	排ガス 1	排ガス 2
#77	0.02137	0.30421	0.09378	0.00934	0.00114	0.28615	0.24760
#81	0.00162	0.02832	0.00000	0.01317	0.00000	0.00958	0.07079
#126	0.00135	0.00000	0.00509	0.00383	0.00178	0.00712	0.11641
#169	0.00016	0.00000	0.00000	0.00000	0.00057	0.00024	0.02533
#105	0.18445	0.17961	0.28687	0.16169	0.01296	0.08405	0.08940
#114	0.01639	0.03317	0.03459	0.01168	0.00000	0.23611	0.03195
#118	0.64133	0.34547	0.49310	0.47637	0.02921	0.23884	0.11842
#123	0.01314	0.01537	0.01973	0.01933	0.01731	0.00631	0.01820
#156	0.01689	0.02751	0.02079	0.06604	0.02094	0.01415	0.05659
#157	0.00302	0.00971	0.00403	0.01270	0.00000	0.00272	0.03184
#167	0.00792	0.03074	0.02695	0.11237	0.02173	0.00514	0.02744
#189	0.00058	0.00000	0.00000	0.00224	0.00541	0.00000	0.03402
#170	0.02584	0.00809	0.00679	0.04465	0.17460	0.03653	0.07588
#180	0.06594	0.01780	0.00827	0.06660	0.71435	0.07306	0.05615

表 10. 大気における発生源寄与率の事後平均と標準偏差 .

発生源	KC300	KC400	KC500	KC600	排ガス 1	排ガス 2	未確認
事後平均	0.01486	0.07544	0.02218	0.04760	0.01984	0.00229	0.81778
標準偏差	0.01153	0.04916	0.01507	0.02634	0.01506	0.00158	0.04967

表 11. 大気における未確認発生源の組成の事後平均と標準偏差 .

名称	#77	#81	#126	#169	#105	#114	#118
事後平均	0.00649	0.00085	0.00059	0.00011	0.19624	0.01132	0.69025
標準偏差	0.00508	0.00153	0.00124	0.00118	0.03066	0.00470	0.04180
名称	#123	#156	#157	#167	#189	#170	#180
事後平均	0.01317	0.01593	0.00288	0.00293	0.00030	0.02007	0.03886
標準偏差	0.00283	0.00344	0.00172	0.00223	0.00099	0.00667	0.02232

に水田に散布された農薬がダイオキシン類汚染の発生源になる場合が多い。

念のため、組成に基づく重回帰による CMB(柏木 他, 2002)を用いて発生源寄与率の推定を試みた。新潟県が実施した 6 回の平行測定により得られた底質の測定値を環境組成とし、橋本(2004)が収集、分類した農薬中ダイオキシン類の典型組成 8 種類、すなわち CNP1~4 および PCP1~4 と新潟県が測定した燃焼排ガス中ダイオキシン類組成の平均値を発生源組成とした。表 6 に PCP 以外の組成を示す。計算の結果、PCP3 の寄与率が 0.987 と推定され、残りを CNP2、排ガス、および PCP4 で分け合う形になった。残差を見ると、全体として許容できる程度の大きさではあったが、OCDD の残差がやや大きく、当てはまりに若干の疑問が生じた。

そこで、提案の方法を用いて発生源寄与率の推定を再度試みた。計算結果を表 7 および 8 に示す。表 7 を見ると、未確認発生源の寄与率の事後平均は 0.165 であり、仮定した発生源組成で説明できない割合が一定程度存在すると示唆している。一方、表 8 を見ると、7 塩素の PCDDs が未確認発生源組成の約 6 割を占め、その他では 6 および 8 塩素の PCDDs が優越し、全体として特殊な組成を構成している。その後の調査により判明した現場周辺の状況との関連が明らかになれば、発生源に関する新たな知見を得られる可能性がある。

7.3 大気

宮城県は大気中のダイオキシン類を年 4 回 8 地点で継続的に測定している。その結果によれば、平成 13 年度から 15 年度の間、何れの地点の濃度も環境基準を下回っていた。ただし、石巻市では夏期になるとダイオキシン様 PCBs の濃度が他の地点の数倍から数十倍高くなるという特異的な現象がみられた。そこで、宮城県は原因を究明するため平成 16 年夏期に石巻市の大気を 4 検体採取した。表 9 に、大気 4 検体中ダイオキシン様 PCBs の平均組成と共に、主要な発生源と考えられた PCB 製品 4 種類の組成および市内主要燃焼施設排ガスの典型組成 2 種類を示す。この表を一見しただけで #118 の環境組成と発生源組成が整合し得ないのが分かる。そこで、組成の不整合の程度について検討するため、提案の方法を適用した。

計算結果を表 10 および 11 に示す。表 10 を見ると、未確認発生源の寄与率の事後平均は 0.818 となり、提案の方法が期待通りに機能するのを改めて確認できた。ただし、未確認発生源が主要な発生源である可能性が極めて高くなり、問題は複雑化した。表 11 を見ると、#118 が未確認発生源組成の約 7 割を占めている。残念ながら #118 の組成がこれ程大きい発生源は現時点では確認されていない。また、PCBs を長期に保管あるいは高温下で使用しても組成は大きく変化しないと報告されている(野馬 他, 2004)。夏期に濃度が高くなる点を考慮すると何等かの媒体に含まれていた PCBs が気温上昇に伴い大気中に放出された可能性が高い。発生源について更なる調査検討が必要になった。

謝 辞

本研究は部分的に統計数理研究所共同研究プログラム(17-共研-2064)の援助を受けています。また、有益なコメントを下さった査読者に感謝申し上げます。

参 考 文 献

- Akaike, H.(1980) Likelihood and the Bayes procedure, *Trabajos de Estadística*, **31**, 143-166.
Berkson, J.(1950) Are there two regressions?, *Journal of the American Statistical Association*, **45**, 164-180.
Billheimer, D.(2001) Compositional receptor modeling, *Environmetrics*, **12**, 451-467.

- Friedlander, S. K. (1973). Chemical element balances and identification of air pollution sources, *Environmental Science and Technology*, **7**, 235–240.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, Boca Raton.
- 橋本俊次(2004). 統計学的手法によるダイオキシン類の発生源推定のための基礎検討, *環境化学*, **14**, 263–285.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.
- Henry, R. C., Lewis, C. W., Hopke, P. K. and Williamson, H. J. (1984). Review of receptor model fundamentals, *Atmospheric Environment*, **18**, 1507–1515.
- Hopke, P. K. (1995). The mixture resolution problem applied to airborne particle source apportionment, *Chemometrics in Environmental Chemistry*, 47–86, Springer-Verlag, Heidelberg.
- Hopke, P. K. (2003). Recent developments in receptor modeling, *Journal of Chemometrics*, **17**, 255–265.
- 飯村文成, 佐々木裕子, 津久井公昭, 吉岡秀俊, 東野和雄, 竹田宜人, 葛西孝司(2002). 東京湾における魚類のダイオキシン類, PCBs 汚染, *環境化学*, **12**, 343–352.
- Imamoglu, I. and Christensen, E. R. (2002). PCB sources, transformations, and contributions in recent Fox River, Wisconsin sediments determined from receptor modeling, *Water Research*, **36**, 3449–3462.
- Kashiwagi, N. (2004). Chemical mass balance when an unknown source exists, *Environmetrics*, **15**, 777–796.
- 柏木直久, 佐々木裕子, 飯村文成, 安藤晴夫(2002). 関数関係解析による Chemical Mass Balance, *応用統計学*, **31**, 59–74.
- Kowalczyk, G. S., Choquette, C. E. and Gordon, G. E. (1978). Chemical element balances and identification of air pollution sources in Washington, D. C., *Atmospheric Environment*, **12**, 1143–1153.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machine, *Journal of Chemical Physics*, **21**, 1087–1091.
- 野馬幸生, 石川 紫, 能勢和聡, 峯戸松勝秀, 滝上秀孝, 酒井伸一, 泉澤秀一, 鍋木儀郎(2004). 保管 PCB 廃棄物の PCBs およびダイオキシン類, *環境化学*, **14**, 501–518.
- Ogata, Y. (1989). A Monte Carlo method for high dimensional integration, *Numerische Mathematik*, **55**, 137–157.
- Ohtaki, M., Sato, M. and Nitta, H. (1997). Estimating source apportionment of particulate matters based on source profiles with fluctuations, *Environmetrics*, **8**, 341–350.
- Paatero, P. and Tapper, U. (1993). Analysis of different modes of factor analysis as least squares fit problems, *Chemometrics and Intelligent Laboratory Systems*, **18**, 183–194.
- Park, E. S., Spiegelman, C. H. and Henry, R. C. (2002). Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models (with discussion), *Environmetrics*, **13**, 775–809.
- Su, M. C. and Christensen, E. R. (1997). Apportionment of sources of polychlorinated dibenzo-p-dioxins and dibenzofurans by a chemical mass balance model, *Water Research*, **31**, 2935–2948.
- 高菅卓三, 井上 毅, 大井悦雅(1995). 各種クリーンアップ法と HRGC/HRMS を用いたポリ塩化ビフェニル(PCBs)の全異性体詳細分析方法, *環境化学*, **5**, 647–675.
- 吉澤 正, 強口英行, 石渡康尊, 半野勝正, 田中 崇, 依田彦太郎, 木村満男, 原 雄(2004). 市原港

におけるダイオキシン類汚染調査 II (鉛直分布), 全国環境研会誌, 29, 23-30.

- Wang, D. D. and Hopke, P. K. (1989). The use of constrained least-squares to solve the chemical mass balance problem, *Atmospheric Environment*, **23**, 2143-2150.
- Watson, J. G., Cooper, J. A. and Huntzicker, J. J. (1984). The effective variance weighting for least squares calculations applied to the mass balance receptor model, *Atmospheric Environment*, **18**, 1347-1355.
- Wiens, D., Florence, L. Z. and Hiltz, M. (2001). Robust estimation of chemical profiles of air-borne particulate matter, *Environmetrics*, **12**, 25-40.

Estimation of Contribution of Unidentified Sources to Environmental Contamination

Nobuhisa Kashiwagi¹, Tadashi Yoshizawa², Tsuyoshi Ibaraki³, Kenichi Kato⁴,
Shunji Hashimoto⁵ and Yuko Sasaki⁶

¹The Institute of Statistical Mathematics

²Chiba Environmental Research Institute

³Niigata Prefectural Institute of Public Health and Environmental Science

⁴Miyagi Prefectural Institute of Public Health and Environment

⁵National Institute for Environmental Studies

⁶Tokyo Metropolitan Research Institute of Environmental Protection

For environmental management, it is essential to identify the sources of contamination for countermeasures to be taken. Several mathematical methods called receptor models have been proposed to identify sources of environmental contamination. However, no method has been established because unknown quantities become unidentifiable when the existence of unidentified sources, i.e. sources whose compositions are unknown, is assumed. It is difficult to solve this identification problem generally. However, the Bayesian method sometimes enables us to obtain practical estimates of unknown quantities that are unidentifiable. This article explains in detail a Bayesian method for estimating the contribution rate of unidentified sources to ambient concentrations of contaminants. The results of application of the method to actual environmental contamination problems are also shown.