

## ★CMBK v2.1 について

CMBK v2.0 をバージョンアップしました。改善点は以下の通りです。

- (1) ベイズ型組成半因子モデルの残差表示に誤りがありましたので、修正しました。
- (2) 乱数が不足する事例がありましたので、1 回の計算で乱数ファイル全部を使用できるようにしました。これに伴い、使用乱数ファイルを指定する必要はなくなりました。
- (3) 環境測定値の相対誤差を物質毎に指定できるようにしました。相対誤差は測定値標準偏差の測定値平均値に対する百分率で指定します。指定場所は3行目です。相対誤差が全物質で同じなら、1列目のみ指定して下さい。物質毎に異なる場合は、1列目にゼロを入れ、2列目以降に各物質の相対誤差を指定して下さい。
- (4) 以上の変更に伴い、1行目で指定するのは、尤度計算の有無だけになりました。

以上です。

## ★CMBK v2 について

CMBK v2.0 は Chemical Mass Balance (CMB) のベイズ型組成回帰モデルとベイズ型組成半因子モデルを実行するための Windows 用ソフトウェアです。CMB は化学物質の環境測定値から発生源の寄与の程度を推定するための方法で、様々なモデルが提案されています。米国環境保護局は、測定誤差モデルの縮退版である Berkson モデル (1950, *JASA*, **45**, 164-180) に基づく方法を EV 法と名付け、そのソフトウェア (CMB8) を CMB の標準ソフトウェアとして整備し配布しています。CMB8 は使い勝手の良いソフトウェアですが、問題が生じる場合があります。それらの問題を緩和するため、CMBK v2.0 を整備しました。

ベイズ型組成回帰モデルは、CMB8 が抱える問題の内、a) 発生源濃度の推定値がしばしば負になる、b) その原因のひとつである多重共線性に対応できない、c) 発生源濃度の総和と環境濃度の総和が一致しない、といった問題を緩和します。そのため、濃度を組成に変換して非負で総和は1という制約条件を活用すると共に、多重共線性を緩和できるベイズ法を採用しました。ベイズ型組成回帰モデルを最初に提案したのは Billheimer (2001, *Environmetrics*, **12**, 451-467) です。彼は事前分布としてロジスティック正規分布を仮定しましたが、本ソフトでは、事後平均に尤度を重みにした重み付き平均という意味を持たせるため、事前分布として一様分布を仮定しました。また、組成回帰モデルの残差に正規分布を仮定しましたが、残差 (組成) の共分散を推定するため、濃度に対数正規分布あるいは打ち切り正規分布を仮定した2つのモデルを用意しました。

ベイズ型組成半因子モデルは、上記の CMB8 が抱える問題を緩和すると同時に、未知発生

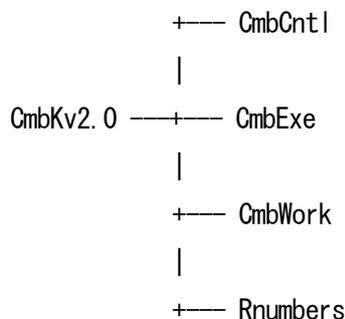
源に関する推論も可能にします。CMB では全ての発生源とその組成は既知と仮定されますが、実際には全ての発生源やその組成が既知とは限りません。そこで、組成が未知の発生源の存在を仮定し、その組成と寄与率を推定できるようにしました。未知発生源の存在を仮定すると、推定は一般に不定になりますが、ベイズ法を用いて不定問題を解決しました。ただし、未知発生源に関する推論の精度は環境測定値の数に依存します。詳細については、Kashiwagi (2004, *Environmetrics*, 15, 777-796)、あるいは

<http://www.ism.ac.jp/editsec/toukei/pdf/54-1-123.pdf>

を参照して下さい。

蛇足ながら、環境濃度の総和と発生源濃度の総和の差を未知発生源由来と解釈する向きがありますが、それは誤りです。差は誤差に過ぎません。実際、差は正にも負にもなり得るのです。詳細はCMB8のpercent massをご覧ください。

-----  
ディレクトリ構成



- CmbCntl にある filename.txt に解析するデータファイルのパス名を書き込みます。CmbCntl と filename.txt という名称、および CmbExe から見た相対位置は変更できません。
- CmbExe に実行ファイルがあります。
- CmbWork にデータファイルの書式および例題のデータファイルと出力結果があります。
- Rnumbers に計算に使用する乱数ファイルがあります。

-----  
利用方法

1. 解析データのためのディレクトリを適切な場所に適切な名前で作成して下さい。
2. そのディレクトリの下にデータファイルを作成して下さい。データの書式については CmbWork にある「データ書式.xlsx」と「Sample.csv」を参考にして下さい。
3. 作成したら、ファイルを CSV 形式で保管して下さい。
4. データファイルのパス名を CmbCntl にある filename.txt に書き込んで下さい。パスを省略した場合は「../CmbWork/」が設定されます。なお、計算結果はデータファイルと同じディレクトリに出力されます。
5. CmbExe にある実行ファイルを実行して下さい。実行ファイル名は次の規則に従っています。
  - Bp … ベイズ型組成回帰モデル、Sf … ベイズ型組成半因子モデル
  - Ln … 対数正規分布、Tn … 打ち切り正規分布

+

| 出力内容 |

+

出力ファイル名は「データファイル名+実行ファイル名+相対誤差.CSV」です。出力内容を簡単に説明します。

1. Log Likelihood は残差分布の選択に使用して下さい。データやモデルが異なる場合は比較しても意味がありません。
2. Variance Components の内、Delta は仮定した相対誤差に対する残差の倍率です。添付の sample は悪い例ですが、データの質が揃っている場合、仮定した相対誤差が 10% で Delta の平均が 3 なら、相対誤差を 30% に変更すると Delta の平均は約 1 になります。
3. Contribution Rates に発生源寄与率の推定値があります。各発生源の有意性を判断したい場合には、標準偏差を利用して下さい。寄与率の事後分布形が不明なので、正確ではありませんが、概ね標準偏差の 2 倍より平均値が大きければ、有意と考えていいでしょう。より正確な数値が必要なら、95%信頼区間の算出を検討します。
4. Regression にデータと寄与率の平均値を使った回帰の結果があります。残差を見て、適合の具合を判断して下さい。
5. ベイズ型組成半因子モデルの場合、Composition of Unknown に未知発生源の組成の推定値があります。ただし、推定精度は未知発生源の寄与率に依存します。未知発生源の寄与率が小さい場合、当然ながら組成に関する情報は得られませんので、組成の推定値はほぼ一様になります。また、環境測定値の数が少ない場合、未知発生源寄与率の推定値は小さくなる傾向があり、組成に関する情報も得られないようです。