

Bootstrap 法による面源負荷量推定値の信頼区間

Mass flux estimate with confidence interval using bootstrap method

○多田 明夫*・田中丸 治哉**

○ Akio TADA*, Haruya TANAKAMARU**

1. はじめに 面源負荷量の推定において、これまで精度についての議論は十分になされてこなかった。この原因としては、水質濃度の母集団分布についての情報が十分に収集できない、すなわち水質データの不足がその根本的な理由であったものと思われる。しかしながら、例えば汚濁物質の流域での浄化や、流域への物質の投入・沈着量と流出量の比較におけるマスバランス研究において、この精度についての議論を避けて通ることは誤った結論を与えてしまう可能性がある。

望ましい面源負荷量への精度の与え方は、データ量が多ければより狭く、少なければより広くなるように、バイアス(偏り)の小さな最確値周りに信頼区間を設定することが適当なものであると思われる。この点については、LQ 式による推定値に対して信頼区間を与える方法が例えば多田ら(2006)により検討が行われている。しかしながら推定値に問題が以前残されていることも指摘されている。これは解析的な方法では、観測データ分布に対する LQ 式の適合の悪さと共に、残差の分布が正規分布や対数正規分布に仮託されることに起因している。これらのずれは最終的にバイアスとなって現れるため、推定された信頼区間が必ずしも期待した結果を与えない。

本報告では、データ分布に経験分布を想定しないリサンプリング法である bootstrap 法を利用して、面源負荷量の推定を行うこととした。

2. 使用データ 解析に用いたデータは、2004 年度に奈良県五條市山林流域(12.82ha)で観測された 15 分間隔の K⁺データを、10 分間隔に整理し直したものである。データには間断的に欠測期間が含まれるが、約 8 ヶ月弱の長さの水質・流量観測値が得られている。

3. 負荷量推定法 3.1 データ抽出法 bootstrap 法は観測されたデータ(標本)集団を小母集団と見なし、そこから再度リサンプリング(通常復元抽出)により、標本集団を再構成し、母集団分布や統計量の推定を行う方法である。このため、そもそもの標本集団が母集団と明かに異なる場合には、良い推定を与えない。また bootstrap 法自体は近似方法であることにも留意が必要である。

ここでは上記の点に鑑み、観測データを流量の大きさ別に、等しい標本数を含むように N 分割する。各 N 分割されたグループから、ランダム抽出によって n 個のデータを抽出する。こうすることで N × n 個の標本が抽出される。今回の解析では N = 10 とし、n = 5, 10, 100 の 3 ケースについて計算を行うこととした。山林流域では水質濃度の変動要因の主たるものが流量であるため、このような方法を取ることにした。

3.2 負荷量推定方法 負荷量の推定法には、バイアスが小さいといわれている次(1)式の Beale 比推定法(Beale, 1962)を採用した。(1)式の L_a は平均負荷量の推定量、 \bar{Q}_a は全流量データの平均流量、 \bar{l}_s 、 \bar{q}_s はそれぞれ標本データによる負荷量と流量の平均値である。 Sl_q 、 S^2_q

*神戸大学大学院自然科学研究科, **神戸大学農学部、*Graduate School of Science and Technology, Kobe university, **Faculty of Agriculture, Kobe University, キーワード：面源負荷量, bootstrap, 比推定

は(2)式で与えられる.

$$L_a = \bar{Q}_a \frac{\bar{l}_s}{q_s} \left(\frac{1 + \frac{1}{n} \frac{S_{lq}}{(\bar{l}_s)(q_s)}}{1 + \frac{1}{n} \frac{S_q^2}{(q_s)^2}} \right) \quad (1)$$

$$\begin{cases} S_{lq} = \frac{1}{n-1} \sum_{i=1}^n (l_i - \bar{l}_s)(q_i - \bar{q}_s) \\ S_q^2 = \frac{1}{n-1} \sum_{i=1}^n (q_i - \bar{q}_s)^2 \end{cases} \quad (2)$$

3.3 ブートストラップ推定法 以上の Beale 比を用いて, 期間中の全流出負荷量の推定を行うこととした. ブートストラップによる信頼区間はパーセンタイル法によるものとし, リサンプリングは 2000 回行った.

4. 結果と考察 推定の結果得られたヒストグラムの一例を **Fig.1** に示した. また中央値, 両側 95%信頼区間の上下限值, および真の負荷量の値を **Table 1** に示した. 表中には, 比較のため, 全母集団から全くランダムに同じデータ数を抽出して bootstrap 推定を行った結果も示して

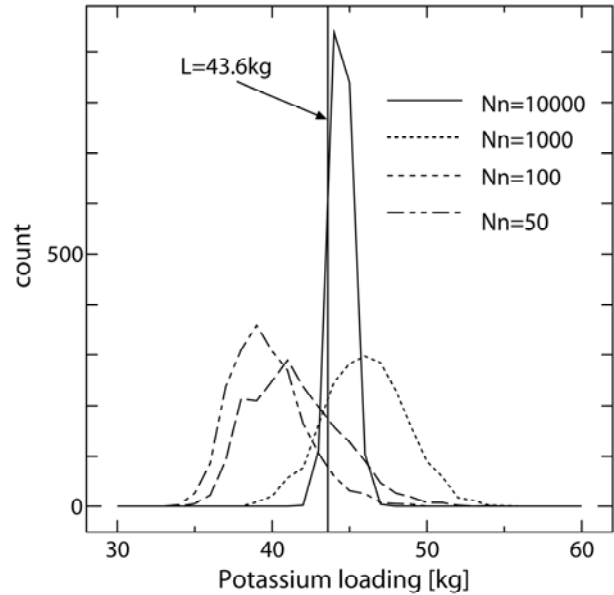


Fig.1 K⁺負荷量のブートストラップ分布

Table 1 Bootstrap 法による負荷量推定信頼区間(パーセンタイル法)

	K ⁺ (流量分布抽出)			K ⁺ (単純ランダム抽出)		
	1000	100	50	1000	100	50
真値	43.6			43.6		
上側 97.5%	50.5	47.3	44.6	59.9	62.2	116.6
中央値	45.5	40.7	38.9	45.4	41.0	57.0
下側 2.5%	40.5	36.4	35.3	35.6	26.0	27.5

いる. これらの結果からも明らかなように, べき乗型の LQ 式などによる推定などよりは bootstrap 法ではより適切な信頼区間を与えることが可能な上, 計算方法も簡便である. 今後は bootstrap 標本を作製する標本の抽出方法や(すなわちサンプリング計画), 様々な抽出標本データセットについて, 信頼区間の予測成否についてさらに検討を加える必要がある. また **Fig.1** から, 推定された負荷量は上側にテーリングした非対称分布をしており, BCa 法などのバイアス修正法がより有効であると思われる. いずれにせよ負荷量の推定にはまだ改良の余地が残されている.

【引用・参考文献】1) Beale, E.M.L. (1962) :Some uses of computers in operational research, Industrielle Organisation, 31, 51-52., 2) Efron, B. and Tibshirani, R. J. (1993) :An Introduction to the bootstrap, CRC Press, Boca Raton., 3) 多田明夫, 田中丸治哉, 畑 武志 (2006) : LQ 式による回帰の信頼区間について—山林小流域からの溶存イオン流出負荷量推定を例として—, 農土論集, 244, 197-206