

## 定期モニタリングデータに基づく負荷量不偏推定法の改良 Improvement of the unbiased load estimation based on periodic samples

○多田 明夫\*・田中丸治哉\*

○Akio TADA\* and Haruya TANAKAMARU\*

1. はじめに 1 年間などの長期間に河川を流下する負荷量の不偏推定のため、著者らはべき乗型 LQ 式による負荷量計算法 (rating curve method, RCM) と重点的サンプリング (Importance sampling) を組み合わせた推定法 (RCM using IS 法) を開発した<sup>1)</sup>。この方法は、高頻度の流量観測値と瞬間負荷量の推定量 $\hat{l}$ に比例した確率でのサンプリング (sampling probability proportional to size; PPS sampling) 水質試料を必要とする<sup>2)</sup>。一方、実際の現場ではそのようなサンプリングの実施は容易でなく、定期的な水質モニタリングが実施されている。我々は定期データからでも、PPS を実現するようなデータをリサンプリングすることで、概ね不偏な負荷推定を実現する手法 (Importance resampling 法, IR 法) も開発した<sup>3)</sup>。この手法では、PPS である水質データが負荷量推定量の積算軸上で一様分布をとることに基づき、一様分布の適合度指標である Anderson-Darling (AD) 検定量<sup>4)</sup>の値  $A^2$  がある水準 (非超過確率  $p\%$  以下) 以下となるようにデータをリサンプリングする。これまで我々は奈良県五條市での高頻度水質観測データを元に、この  $p\%$  の値を  $5\%$  としてきたが、この  $5\%$  の値は LQ 式回帰の古典的仮定 (対数空間上での残差が、平均ゼロの正規分布で、系列相関も不均一分散も無く、互いに独立) の元で、現地での流量観測値を用いて決定されていた。しかしながら、この  $5\%$  の汎用性 (他流域への適用性能) の担保と、低いリサンプリング効率 (元の定期標本集団から負荷量推定に最低限必要な標本数である 4 個以上の PPS 水質データを抽出できないケースが相当数発生すること) の改良が課題であった。本発表では、これらの問題の改良法について報告する。

2. 現行の IR 法のアルゴリズムとその改良 現行の IR 法のアルゴリズムでは、非一様分布数列の  $A^2$  値がより大きな値をとることから、相対積算負荷推定量 ( $\sum_i \hat{l}_i / \sum_i \hat{l}_i$ ,  $N$  は観測期間の全単位時間数) 軸  $(0, 1]$  上で  $n$  個の標本値の数列から、数列の  $A^2$  値を最も大きくしているデータ 1 つを除去し、これ処理を数列の  $A^2$  値が所定の水準 (非超過確率  $p\%$  の AD 検定量  $A^2_{p\%}$ ) 以下となるまで繰り返す。現行の IR 法よりもリサンプリング標本数を増大させるため、初期標本集団から復元抽出で作成した標本サイズ  $n$  の複製標本集団を  $n$  組作成し、これらに現行と同様の IR 法を適用し、最大の標本サイズのリサンプリング PPS 標本集団を採択することとした。これにより、リサンプリング後に 4 個以上の標本数を得られる確率が向上した上、特に  $n$  が小さな時に、事後の標本数を増大させることができた。

3. IS 法による負荷量の不偏性と  $p\%$  値の関係について 次に、データの一様乱数化の判断基準である AD 検定量の非超過確率の値  $p\%$  をどの様に定めるべきか検討する。今、定期調査データを、LQ 式 ( $\hat{l}_i = \alpha q_i^\beta$ ,  $q_i$ : 瞬間流量,  $\alpha \cdot \beta$ ; パラメータ) の  $\beta$  値が 0 である PPS 標本集団と見なす。また、実際の母集団の LQ 式の指数パラメータを  $\beta_m$  とする。このとき、

(所属) \*神戸大学大学院農学研究科, Graduate school of agricultural science, Kobe university

(キーワード) 定期調査データ, 面源, 負荷量, 不偏推定, 総排出負荷量

以下のことを理論的に説明しうる。(1) 流量が対数正規分布でかつ LQ 式回帰の古典的仮定が成立すれば、どのような $\beta$ に基づく PPS 標本であっても、RCM using IS 法は不偏推定量を与える<sup>1)</sup>。(2) 流量は対数正規分布に従わないが、回帰の古典的仮定が成立するとき、 $\beta$ の値が $\beta_m$ に近い時にのみ、不偏推定量が得られる。(3) 流量が対数正規分布に従わず、回帰の古典的仮定も成立しないとき、 $\beta$ の値が $\beta_m$ に近くても推定量は偏る(ただし上の2条件が近似的に成立すれば、推定量の偏りは小さい)。我々は手元にある数少ない定期調査データを元に $p\%$ の値を定めるにあたり、(2)の条件の元で検討することとした。なお不偏推定量(あるいは概ね95%の被覆確率を有する中央95%信頼区間)を与えるような $p$ の値は、a)  $\beta$ と $\beta_m$ の差の大きさ、b) 標本集団から推定される指数パラメータ $\beta'$ の不確かさに依存する。それゆえに $p$ の値は水質項目に依存するが、a)・b)の条件が同じでも、流量データが対数正規分布から外れるほど、より厳しい条件(小さな $p$ の値)が必要となる。

**4. 結果および考察** 以上より、 $p$ の値は流量観測値と LQ 式の回帰の古典的仮定を満足する仮想データ(LQ 式のパラメータ $\alpha \cdot \beta$ 、残差分散と正規乱数から生成される負荷量データ)を基に、定期調査データ集団毎に決定されるべきである。実際には定期調査データから計算されるパラメータ $\alpha \cdot \beta$ と残差分散の推定量から仮想母集団データを生成し、そこから Monte Carlo 法により抽出した定期サンプリングデータから種々の $p$ 値に対する抽出 PPS 標本集団を用いて、信頼区間の被覆確率を計算することができる。このようにして $p$ 値と95%信頼区間の被覆確率の関係を調べ、最終的に95%の被覆確率の実現に必要とされる $p$ 値を決定すればよい。この方法を奈良県五條市の小山林流域で観測された高頻度水質観測データに適用した結果、月1回の定期調査データ(約2年のデータ長で25個程度)であれば K で40%、Cl と Na で50%、SS で70%の $p$ 値に対する $A^2_{p\%}$ を IR 法の基準値として採用すれば良いことがわかった(これらの $p$ 値は平均で、実際には元の定期データセット毎に異なる値をとる)。この基準値に基づいて95%信頼区間の被覆確率を同じ条件で求めたところ、K で93%、Cl と Na で94~95%、SS で92%となった。被覆確率の想定値(95%)と実際の値との差異は、LQ 式の回帰の古典的仮定の満足の程度に依存する(ただし、系統誤差や不均一分散が偏りを増大させる一方で系列相関は偏りを小さく止める効果がある)。以上が、水質項目毎・データセット毎に $p$ 値を効率的に定める方法である。

**5. おわりに** これまで IR 法のリサンプリングの基準となる AD 検定量の非超過確率の値 $p\%$ の決定根拠について十分に考察できていなかったが、これが流量分布の対数正規性に対する不一致の程度により大きく決定されること、そしてそれはモンテカルロシミュレーションで推定可能であることがわかり、計算時間は増大するものの従来よりも効率の良い推定が可能となった。実際のデータでは LQ 式の回帰残差が回帰分析の古典的前提条件を満足せず、推定量が偏るため被覆確率が低下するが、この改良は小標本サイズの水質データからは不可能である。この点に関して、手法の頑強性の検証が今後必要となろう。

**謝辞** 本研究は科学研究費補助金(No.15K07646)の助成を受けて実施したものである。

**参考・引用文献** 1)Tada, A. and Tanakamaru, H. (submitted); Confidence intervals for riverine loads based on a rating curve method, 2) 多田明夫・栗原周平・田中丸治哉(2013), サンプリング法の負荷量推定への影響, 平成25年農業農村工学会大会講演会, 3) 多田明夫・田中丸治哉(2017), 流出負荷量の不偏推定法の一般化, 平成29年農業農村工学会大会講演会, 4)Stephens, M. A. (1986); Goodness-of-fit techniques. Series Statistics: Textbooks and Monographs, 68.