

## サンプリング法の負荷量推定への影響 Effects of Sampling Strategy on Load Estimation

○多田 明夫\*, 栗原 周平\*\*, 田中丸治哉\*

○Akio TADA\*, Syuhei KURIHARA\*\* and Haruya TANAKAMARU\*

### 1. はじめに

著者らはこれまで奈良県五条市の山林小流域を対象に、溶存態項目 ( $K^+$ ,  $Na^+$ ,  $Cl^-$ ) や懸濁物質 (濁度より推定) の高頻度水質濃度・流量データを用いて、集水域からの総流出負荷量の区間推定方法について検討してきた。これら水質項目は流量変動に対する濃度変動も様々であるため、複数項目に共通して適切な区間推定量を与える手法が見つければ、集水域からの総流出負荷量推定法の構築に繋がると考えている。なお著者らの前提とする条件は、流量や雨量などの水文資料は十分高頻度に観測されているが、水質濃度はもっと低い頻度でしか観測されていないという状況である。

総流出負荷量の区間推定法は、現地にて、いつ水質試料を観測するかという①サンプリング方法、総流出負荷量をどう計算するかについての②負荷量算出法、および③信頼区間の構成方法からなる。総流出負荷量の推定量の偏りは①～③の全てで発生する。②・③の過程で発生する偏りについて、著者らはこれまで様々な検討を加えてきたが、①のサンプリング時の偏りの影響が大きくなることがわかってきた。

本報告では、総流出負荷量の推定量の偏りを減少させるため、重点的サンプリング (加重サンプリング, importance sampling) について検討したので報告する。

### 2. 重点的サンプリング

水質試料のサンプリング方法の代表的なものに、等間隔 (定期) サンプリング, 層別サンプリング, 確率論的サンプリング, 降雨出水イベントなどを考慮した opportunistic な不定期のサンプリング, およびこれらを組み合わせたものがある。層別サンプリングは推定量の分散や偏りを小さくするものの、層分割の取扱いが難しく、普及しているとはいえない。また等間隔サンプリングと洪水時調査の組み合わせについても確たる指針があるわけではなく、問題が残されている。

重点的サンプリングは確率論的なサンプリングに該

当する。重点的サンプリングは、モンテカルロシミュレーションで用いられる精度を高めるための古典的な手法である。この方法の基本的な考え方は、ある推定量に対する影響の大きさに応じて、よりサンプルが抽出され易くしてやるというものである。今時刻  $t$  における流域からの瞬間負荷量の観測値が  $l(t)$  であったとすれば、期間  $[t_1, t_2]$  中の総流出負荷量の値  $L$  は

$$L = \int_{t_1}^{t_2} l(t) dt \quad (1)$$

と書ける。通常我々は採水時刻  $t$  について一様な等間隔を考えているので、結局は次式で計算を行っている。

$$L \cong K \Delta t \sum_{i=1}^{t_2} f(q_i), \quad \hat{l}(t) = f(q_i) \quad (2)$$

ここで、 $q_i$  は時刻  $t$  における瞬間流量であり、 $f(q_i)$  は  $q_i$  による  $l(t)$  の推定量を与える関数、 $K$  は単位換算係数、 $\Delta t$  は採水間隔である。(2) 式は総流出負荷量の推定に対して非効率である。なぜならば、総負荷量に大きく寄与する出水時のデータは、時間軸上では少ない割合でしか発生しないためである。このため  $\Delta t$  が大きくなると急激に総負荷量推定量の偏りが増大する。

重点的サンプリングでは等時間間隔のような時間的に一様な確率密度分布の代わりに、 $f(q_i)$  と  $|l(t)|$  が比例するような確率密度分布  $g(t)$  を考え、それに従ってサンプリングを行う。それにより、推定量への影響の大きなデータはより抽出されやすくなる。この結果、等間隔サンプリングを想定した (2) 式よりはずっと偏りの小さな総流出負荷量の推定量を得ることができる。

$$L \cong \int_{t_1}^{t_2} \frac{l(t)}{g(t)} g(t) dt \quad (3)$$

問題は、どのようにして確率密度分布  $g(t)$  を決定するか、というこである。また区間  $[t_1, t_2]$  で  $g(t)$  の積分が 1 になる必要がある。我々は既に  $f(q_i)$  の候補として、ベキ乗型の LQ 式、 $f(q_i) = a q_i^b$  ( $a, b$  は定数) など、多くの流量・負荷量間の経験推定式を知ってい

(所属) \* 神戸大学大学院農学研究科, Graduate School of Agricultural Science, Kobe University

\*\* 日本工営株式会社, Nippon Koei

(キーワード) 重点的サンプリング, サンプリング戦略, 総流出負荷量, 区間推定

る。これを用いれば、 $g(t)$  を次式で表すことができる。

$$g(t) = \frac{f(q_t)}{\int_{q_1}^{q_2} f(q_t) dt} \quad (4)$$

忘れてならないのは、事前に  $g(t)$  を推定するための予備調査データが必要となる点である。また (3)・(4) 式の確率密度  $g(t)$  には対数変換等の変数変換により生じる偏りが相殺されるというメリットもある。

実際に (3)・(4) 式に基づいて現地で水質試料をサンプリングしようとする方法が、Thomas(1985) が提案した SALT 法<sup>1)</sup> である。Thomas は論文で、 $\Sigma f(q_i)$  と  $L$  が一致しない場合の対処法も提案している (Thomas はその論文で  $f(q_i)$  の推定式を具体的に与えていないのである)。

### 3. サンプリング法の違いが区間推定に与える影響

本報告では、2009～2011 年度に奈良県五條市の山林小流域 (12.82ha) において観測された 10 分間値の流量データと、15 分間隔の水質濃度観測値から線形補間されて得られた 10 分間値の水質濃度データを解析に供した。この 10 分間値の全流量・水質濃度データを本報告では母集団と見なし、10 分間値の水質濃度と流量データの積である瞬間流出負荷量の総積算値を真の総流出負荷量と仮定する。なお欠測期間が項目により異なるため、総データ数は  $K^+$  で 83,605 個、 $Cl^-$  で 81,957 個、 $Na^+$  で 71,304 個である。それぞれの項目について、等間隔サンプリングと重点的サンプリングに基づいて標本集団を抽出し、総流出負荷量の区間推定を行った。信頼区間内に真値を含む確率 (被覆確率) が設定した信頼水準 (今回は 95%) に近いかどうかで、区間推定の良否を判断している。

等間隔サンプリングでは 1 日、3 日、7 日間隔を設定した。この方法ではそれぞれの採水間隔内に含まれる 10 分間値の個数だけの独立した標本集団が得られる。重点的サンプリングでは 1 日、3 日、7 日間隔の等間隔サンプリングに相当するデータ数を水質項目毎に設定した。重点的サンプリングでは 500 組の標本データセットを発生させて、被覆確率を評価した。負荷量の計算および重点的サンプリングの設計に用いたレーティングカーブは USGS Loadestimator の model 6<sup>2)</sup> に、対数変換に起因する偏りを修正する QMLE バイアス修正項を加味したモデルである。

$$\ln \hat{l}(t) = \beta_0 + \beta_1 \ln q_t^* + \beta_2 (\ln q_t^*)^2 + \beta_3 \sin(2\pi t^*) + \beta_4 \cos(2\pi t^*) \quad (5)$$

ここで、 $\hat{l}(t)$  は瞬間流出負荷量の推定値、 $q_t^*$  は基準化された瞬間流量、 $t^*$  は基準化された時間、 $\beta_0 \sim \beta_4$  は偏回帰係数である。

なお、1 日、3 日、7 日間隔の等間隔サンプリングのデータセットの内、良い区間推定を与えた標本集団、区間推定の結果が悪かった標本集団からそれぞれ抽出間隔毎に 1 組ずつ選択し、それらを用いて重点的サンプリングを設計した。なお重点的サンプリングの設計に用いたデータ数と評価のために抽出するデータ数は一致させている。信頼区間はブートストラップ法によるものである。結果を Table 1～3 に示す。表から、等間隔サンプリングでは区間推定は良好でないが、重点的サンプリングは一貫した良好な区間推定を与えていることがわかる。

### 4. おわりに

今後の課題として、どの程度のデータ量があれば、あるいはどのようなデータであれば、偏りが小さくできるかを判別する方法の検討があろう。

#### 参考・引用文献

- 1) Thomas, R. B. (1985); Estimating Total Suspended Sediment Yield With Probability Sampling, Water Resour. Res. 21(9), 1381-1388. 2) Runkel, R., Crawford, C. G. and Cohn, T. A. (2004) : Load Estimator (LOADEST): A Fortran Program for Estimating Constituent Loads in Streams and Rivers, U.S. Geol. Surv. Techniques Methods, Book 4, Chap.A5.

**Table 1** 95% 信頼区間の被覆確率 (%) 等間隔サンプリング  
Cove rage rate of 95% C.I. - Periodical Sampling

interval	1d	3d	7d
$K^+$	81.3	89.6	88.4
$Cl^-$	100	93.8	91.6
$Na^+$	96.5	89.6	88.7

**Table 2** 95% 信頼区間の被覆確率 (%) 重点サンプリング (A)  
Cove rage rate of 95% C.I. - Importance Sampling (A)

eq.interval	1d	3d	7d
$K^+$	91.4	91.4	93.8
$Cl^-$	95.2	95.2	92.0
$Na^+$	94.6	91.6	93.4

eq.interval; equivalent interval based on the number of samples

**Table 3** 95% 信頼区間の被覆確率 (%) 重点サンプリング (B)  
Cove rage rate of 95% C.I. - Importance Sampling (B)

eq.interval	1d	3d	7d
$K^+$	94.6	94.4	93.2
$Cl^-$	94.2	94.4	94.4
$Na^+$	94.6	95.2	95.0

eq.interval; equivalent interval based on the number of samples