確率的サンプリングを利用した打ち切りデータに基づく懸濁物質負荷量の区間推定
Interval estimation of suspended solids loads with censored data using stochastic sampling

○ 多田 明夫，田中丸治哉
○ Akio TADA and Haruya TANAKAMARU

１．はじめに

人為的影響の少ない山林渓流水の濁度や懸濁物質濃度は，低水時にはゼロを示し，洪水時・増水時にのみ値を示すことが多い。このようなデータは打ち切りデータ（censored data）と呼ばれる。打ち切りデータに基づいて流域からの総流出負荷量を推定する場合には多くの困難を伴う。本報告は，このようなケースにおける，流域からの総流出負荷量の適切な区間推定法について考察したものである。

２．使用データ

対象流域は奈良県五條市の山林流域 (12.82ha) である。対象項目は懸濁物質である。懸濁物質の濃度は，後方散乱式濁度計（KENEK，RM1000/RMT-3F）により 10 分間隔で連続観測されている濁度の値から換算した。換算数式は，現地での洪水時の調査に基づいて決定した。解析期間は 2009 年 2 月 1 日から 2010 年 12 月 31 日の約 2 年間とした（期間中の全 10 分間値は 100,576 個，うちゼロでない濃度値は 3,374 個）。

３．負荷量の区間推定方法

3.1 確率的サンプリング（SALT 法）

全データのうち，打ち切りでないデータは 3.4% に過ぎないため，通常の定期的な採水方法（定期サンプリング）では有効な観測値を得ることが困難である。また特定の洪水時のみに行われた調査データだけに基づいて負荷量を算定すると，大きく偏った推定値が得られることが予想される。このような場合，確率論に基づくサンプリングが有効であろう。本報告では，SALT（Selection At ListTime）法（Thomas，1985）を確率的サンプリング法として採用した。

SALT 法の考え方は以下のようである。今ある時間間隔での濃度（c_i）と流量（q_i）の値がわかっているとすると，全期間（総時間間隔数 N）に対する総流出負荷量 L は次式で与えられる。

$$L=\sum_i^N l_i \quad , \quad l_i = c_i q_i K \Delta t \tag{1}$$

ここで，l_i は瞬間負荷量，K は単位変換の係数，Δt は単位時間の長さ，である。

今ここで，(1) 式を次式のように書き換える。

$$L=\sum_i^N l_i = \frac{1}{N}\sum_i^N \left(\frac{L}{l_i}\right) l_i = \frac{1}{N}\sum_i^N \frac{l_i}{w_i} \tag{2}$$

(2) 式の w_i（$= l_i / L$）は負荷量 l_i の持つ重みである。観測された限られたデータ（標本）に基づいて総負荷量を推定する場合には，(2) 式最右辺をそのまま適用することはできない。このため，l_i に代わり，次式の量 x_i に基づいて w_i および負荷量を計算する。

$$x_i = \hat{c}_i q_i K \Delta t \quad , \quad \hat{c}_i = a q_i^b \tag{3}$$

ここで，$\hat{c}_i$ は標本集団から推定される濃度であり，a，b は定数である。q_i のべき乗型関数として $\hat{c}_i$ を推定するものとすると，各観測負荷量 l_i に対する重み w_i の推定値 p_i を次式で得ることができよう。

$$p_i = x_i / X \quad , \quad X = \sum_i^N x_i \tag{4}$$

SALT 法とは，調査流域での過去の観測資料などに基づいて (3) 式を決定し，(4) 式に従って実際にサンプリングを実施する方法である。具体的には，あらかじめ流域からの期間中の総負荷量 X とサンプル数 N を仮定しておき，モンテカルロ法により [1，X] の区間で N 個の乱数を発生させる。この乱数の値が，採水を行うべき時の x_i の累積値となっている。SALT 法を実施するためには水位計と連動した自動採水機，累積負荷量の推定値を計算・保存し，しかるべきタイミングでサンプリングを実施するためのコントローラーが必要となる。また実際に得られるサンプル数は，$N \times (L / X)$ となることに注意が必要である。

3.2 負荷量計算方法

負荷量の計算方法として，Thomas(1985) の方法，べき乗型 LQ 式による方法，打ち切りデータの回帰に用いられる Tobit モデルの 3 種類について検討した。

（所属） 神戸大学大学院農学研究科，Graduate School of Agricurtural Science, Kobe University
（キーワード） 濁度，懸濁物質，区間推定，確率的サンプリング，SALT 法，打ち切りデータ

Thomas の算出方法は次式の通りである。

$$\begin{cases} \hat{L} = \dfrac{1}{n}\sum_{i}^{h} r_i \dfrac{l_i}{p_i} = \left(\sum_{i}^{h} x_i\right)\dfrac{1}{n}\sum_{i}^{h} r_i \dfrac{l_i}{x_i} \\ n = \sum_{i}^{h} r_i \end{cases} \tag{5}$$

ここで，n は重複を含む総観測データ数，h は独立な観測データ数，r_i はあるデータに対する，SALT 法により重複抽出された回数，である。

べき乗型の回帰式による負荷量の計算は，次式の回帰式による負荷量の推定値の総和による。

$$\hat{l}_i = aq_i^b \tag{6}$$

ここで，a，b は定数であり，最小自乗法で求める。Tobit モデルによる負荷計算は，(6) 式の代わりに次式の Tobit モデルにより負荷量の推定を行うものである。

$$\hat{l}_i = \begin{cases} aq_i^b & \left(aq_i^b \geq l_{thd}\right) \\ 0 & \left(aq_i^b < l_{thd}\right) \end{cases} \tag{7}$$

ここで，l_{thd} は打ち切りデータの閾値であり，今回は観測された負荷量の最小値としている。

(6)・(7) 式の負荷量計算では，対数変換した標本の l_i と q_i の値に対して最小自乗法を適用しているため，総流出負荷量を過小推定していることが予想される。これを修正するためのバイアス修正因子（BCF）として，Duan(1983) の smearing 修正因子（8 式）と Ferguson（1986）の QMLE（9 式）について検討した。

$$BCF = \frac{1}{n}\sum_{i}^{n} \exp\left(\log l_{obi} - \log \hat{l}_i\right) \tag{8}$$

$$BCF = \exp\left(\frac{s^2}{2}\right) \quad , \quad s = \sqrt{\frac{\sum_{i}^{n}\left(\log l_{obi} - \log \hat{l}_i\right)}{n-2}} \tag{9}$$

ここで，l_{obi} は観測負荷量である。

3.3　信頼区間の構成方法

本報告では，信頼区間の構成方法として Bootstrap 法を採用した。Bootstrap 複製標本数は 2,000 とした。

４．結果および考察

SALT 法の適用にあたって，平均的な採水間隔がそれぞれ 1 月，2 週間，1 週間，1 日に相当するよう，標本抽出数 N を定めた。すなわち，N = 23，50，100，683 個について検討を行った。それぞれの N に対して，500 組の標本集団を抽出し，それぞれの標本集団に対して負荷量の区間推定を行った。この結果を Table 1 に示した。表中のカバー率とは，500 個の信頼区間のうち，すべての観測値を用いて (1) 式で計算される真の負荷量をそのうちに含むものの割合である。本報告では 95% 信頼区間について検討しているので，N の値に関わらず，一貫して 95% に近いカバー率を示すような信頼区間ほど良いということになる。

SALT 法では総流出負荷量 X の区間に対して，N 個の乱数を発生させて採水データを決定する。このため，x_i が大きなデータは重複して複数回(r_i 回)抽出される。またこうして決定された採水タイミングでも，値ゼロのデータが観測されることに注意が必要である。この数値についても表中に示している。

表より，もっとも良い区間推定を与えるのはべき乗型 LQ 式により区間推定を行った場合であり，Thomas の方法が次に良いこと，Tobit モデルによる推定は実用に耐えないことがわかる。またバイアス修正を考慮した場合，ほとんどの場合信頼区間の下限値が真値よりも大きくなり，やはり低いカバー率を与える結果となっていた。このような結果となった理由として，圧倒的な比率での打ち切りデータの存在と，モデルの残差分布が関係していると推察される。

引用・参考文献

1)Thomas, R. B. (1985) :Estimating total suspended sediment yield with probability sampling, Water Resour. Res., 21, 1381-1388.,2)Duan, N. (1983) :Smearing estimate: a nonparametric retransformation method, J. Am. Stat. Assoc., 78(383), 605-610., 3)Ferguson, R. I. (1986) :River loads underestimated by rating curves, Water Resour. Res., 22, 74-76.

Table 1　SALT 法による SS 負荷量の 95% 区間推定のカバー率 (%)
Coverage rate for 95% interval estimation of suspended solids loads using SALT sampling (%)

N*	N_{ind}*	N_{nz}*	*Thomas*	*power type LQ regression*			*Tobit model*		
				no BCF	*Smearing*	*QMLE*	*no BCF*	*Smearing*	*QMLE*
23	22.3 ± 0.4	10.1 ± 2.3	82.2	92.8	80.6	70.2	97.8	70.0	48.0
50	49.3 ± 0.9	27.2 ± 3.5	87.6	93.4	54.8	50.4	69.6	14.0	2.8
100	97.4 ± 1.6	43.8 ± 4.7	90.8	94.4	25.4	23.2	24.2	2.8	0.0
698	607.6 ± 9.3	307.3 ± 13.3	92.6	94.0	0.0	0.0	0.0	0.0	0.0

*N ; number of SALT samples, N_{ind} ; number of independent samples, N_{nz} ; number of non-zero samples