

LQ 式の回帰残差特性が流出負荷推定に与える影響 Effect of properties of regression residuals on load estimation

○橋本航, 多田明夫, 田中丸治哉

○Wataru HASHIMOTO, Akio TADA, and Haruya TANAKAMARU

1. はじめに 農地や市街地, 山林などの面源から発生する物質の流出量 (流出負荷量) は, 流域の水質汚濁管理上重要な数値である. 近年, Importance Sampling 法(以下 IS 法) とべき乗型 LQ 式を用いて, 不偏な流出負荷量の点推定と, 適切な区間推定が可能となった. この手法は, 瞬間流出負荷量の推定量の大きさに比例した確率でのサンプリング (以下, 本調査)によるデータに基づき推定を行う. このために, IS 法では本調査に先立ち, 瞬間流出負荷量の推定を行うためのデータ収集(以下, 予備調査) も必要となる. このように IS 法は, 二段階のサンプリングを要求する推定法である.

この IS 法を用いれば, 従来よりも偏りの小さな流出負荷量の点推定と適切な区間推定が実現されるが, 実際の水質項目の中には IS 法を用いても, 95%信頼区間の被覆確率が 90%程度にとどまり, 区間推定結果が十分ではないものも見られた. そこで本報告では, IS 法による区間推定をこのように悪化させている要因について検討することとした. 具体的には, 観測データの LQ 式による回帰残差特性が, 区間推定にどのような影響を与えるかについて, 模擬データを用いて検証した.

2. 推定量を偏らせる要因 一般に, 時刻 t における瞬間流出負荷量 $l(t)$ と流量 $q(t)$ の間には, 対数線形モデル (べき乗型モデル, LQ 式, $\ln l(t) = a_0 + a_1 \ln q(t) + e(t)$, a_0, a_1 は回帰係数, $e(t)$ は回帰残差) が想定される. 古典的な線形回帰モデルでは, 回帰残差 $e(t)$ に次の 5 つの仮定, ①系統誤差がなく偶然誤差であること, ②正規分布に従うこと, ③期待値が 0 であること, ④均一分散であること, ⑤独立であること, を置いている. この 5 つの仮定を十分に満足していれば, IS 法とべき乗型 LQ 式により流出負荷量を適切に推定できることが判っている. しかし実際のデータはこの仮定を十分に満たさないため, 推定結果が悪化していると考えられた. このため, 本報告ではデータがこれら 5 条件を満足しないときに, 流出負荷推定量がどのように偏るのか, 模擬データを用いて検証した. 具体的には, 回帰残差 $e(t)$ が, (a) 正規分布に従う場合, (b) 系統誤差のみを含む場合, (c) 不均一分散のみを含む場合, (d) 系統誤差と不均一分散を含む場合について検討した. なお本報告では, 条件⑤については検討していない.

3. 模擬データの作成法 上記の(a)~(d)の特性を有する模擬データの作成方法は次のとおりである. まず全観測データに対し LQ 式の残差 $e(t)$ と残差分散 s^2 を求める. 次に全観測データを流量の大きさ順に昇順に並べ替え, これを n グループの等個数の標本集団に分割する ($n=500$). そしてグループ毎に残差 $e(t)$ の平均 m_i , 分散 s_i^2 ($i=1, \dots, n$) を求める. これらの値を用いて, (a) の場合には $e(t)$ が $N(0, s^2)$ に, (b) の場合にはグループ毎に $e(t)$ が $N(m_i, s^2)$ に, (c) の場合にはグループ毎に $e(t)$ が $N(0, s_i^2)$ に, (d) の場合にはグループ毎に $e(t)$ が $N(m_i, s_i^2)$ に従うよう疑似乱数を用いて生成し, これを $\hat{l}(t) = \exp(a_0 + a_1 \ln q_i)$ に加え模擬データとし

(所属) 神戸大学農学研究科 Graduate School of Agricultural Science, Kobe University
(キーワード) 流出負荷量, 区間推定, Importance Sampling, 回帰残差

た。模擬データの作成に用いた観測データは奈良県五條市の山林流域で得られた 10 分間隔の水質濃度と流量データであり、対象水質項目は区間推定結果が良好でなかった K^+ とした。解析対象期間は 2009 年 5 月 12 日~2011 年 4 月 28 日である。(欠測を除いた期間中のデータ数は 83,605 個で、これを母集団とした。)

4. 区間推定法とその評価方法 IS 法による流出負荷量の区間推定は母集団からの 24 個の標本をランダム抽出する予備調査と 40 個の標本を同じく母集団から IS 法により抽出する本調査の組合せで行い、95%信頼区間(両側危険率 2.5%)を構成した。なお、信頼区間構成法として、GCI 法²⁾と LAND の H 統計量³⁾を採用した。また、上記の方法で標本集団をモンテカルロシミュレーションにより 50,000 回抽出して信頼区間をそれぞれ構成している。区間推定結果については、真値 L_{true} よりも上側信頼限界が小さい場合に過小推定、信頼区間内に L_{true} を含んだ場合に被覆確率、 L_{true} よりも下側信頼限界が大きい場合に過大推定として、それぞれの割合を算出し、評価した。

5. 結果と考察 Table 1 に模擬データおよび観測データに対する区間推定の結果を示した。Fig.1 には 50,000 個の点推定量 \hat{L} の分布を示した。Fig.1 より \hat{L} の中央値が、(b)と(d)ではわずかに過小側に偏っていることが認められる。また Table 1 から、(c)では信頼区間の偏りや被覆確率の低下がほとんど認められないが、(b)と(d)では信頼区間の過小側への偏りと被覆確率の低下が認められ、この傾向は K^+ の実測データに対するものと同じであった。以上より、実測データを用いて算出した流出負荷量の信頼区間の偏りと被覆確率の低下には、LQ 式による回帰残差にみられる系統誤差による影響がより大きいものと考えられた。

6. おわりに 先に述べた回帰残差の仮定①~④からの逸脱は、どれも推定量に偏りを与えるものであるが、今回の検討により K^+ の流出負荷量の区間推定結果の悪化の原因の一つが系統誤差であることが示された。区間推定を改良するためには、系統誤差の影響を小さくしていく必要がある。このためには、WRTDS⁴⁾などの、重みつき回帰手法を用いた LQ 式により、系統誤差の軽減をはかる必要がある。

謝辞 本研究は科学研究費補助金(No.23580333, No.15K07646)の助成を受けて実施したものである。

参考文献 1) 多田明夫・田中丸治哉 (2015) : 平成 27 年度農業農村工学会大会講演会要旨集, pp.512-513 2) Krishnamoorthy, K and Mathew, T. (2003) : *Journal of Statistical Planning and Inference*, 115, pp.103-121 3) Land, E.L. (1975) : *Selected Tables in Mathematical Statistics Vol.III*, pp.385-419 4) Hirsch, R. M and Moyer, et al.(2010) : *Journal of the American Water Resources Association* 46 (5) , pp.857-880.

Table 1 模擬データによる区間推定の結果
Result of interval estimation

evaluation	Simulated data				Observed data
	(a)	(b)	(c)	(d)	
U	2.5	4.2	2.6	4.4	7.8
S	94.9	93.4	94.8	93.2	89.9
O	2.5	2.4	2.6	2.4	2.3

U: under estimation (true value > upper confidence limit)
S: true value is included in C.I
O: over estimation (true value < lower confidence limit)

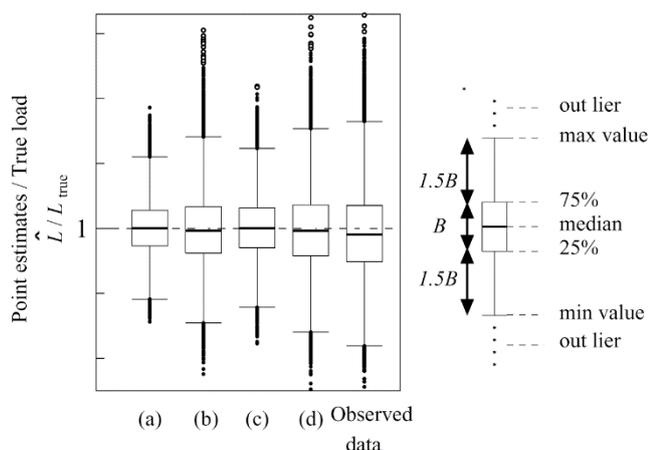


Fig.1 流出負荷量の点推定量の分布
Result of point estimation