

## 米国の12流域の水質データによる流出負荷量不偏推定法 IR法の検証

Validation of the importance resampling method for unbiased load estimation based on daily water quality data from 12 watersheds in the United States

○川村 純, 多田明夫, 田中丸治哉

○Jun KAWAMURA, Akio TADA, Haruya TANAKAMARU

**1. はじめに** 効果的な水質汚濁対策を講じるためには面源から発生する汚濁負荷量を正確に推定する必要がある。この負荷量の不偏推定量を得るためには、負荷量の期待値の大きさに比例した確率 (PPS) で抽出された標本が必要である<sup>1)</sup>。このため IR (Importance Resampling) 法を用いて、既に収集されたデータから PPS 標本を再抽出して負荷量を推定する方法が開発された。これまでこの IR 法による不偏推定と区間推定の妥当性は山林小流域 (奈良県五條市, 12.14ha) の高頻度水質データ (Cl, K, Na, SS の4項目) を用いて確認されていた。本報告では、より流域面積が大きく土地利用も複雑で人間活動の影響を強く受ける流域からの栄養塩などの複数の水質項目に対して、IR 法による負荷量の不偏推定が可能であるかを検証した。

**2. 解析データと方法** **2.1 解析データ** 検証データとして、米 Heidelberg 大学 National Center of Water Quality Research (NCWQR) が Web<sup>2)</sup>上で公開する、米エリー湖周辺の12流域からの水質8項目 (TP, SS, SRP, NO<sub>3</sub>+NO<sub>2</sub><sup>-</sup>, TKN, Cl<sup>-</sup>, SO<sub>4</sub><sup>2-</sup>, SiO<sub>2</sub>) について、8年間の日単位データを用いた。Table 1 に12流域の流域面積と土地利用割合を示した。また流量データには、米地質調査所 (USGS) が公開している日平均流量データ<sup>3)</sup>を用いた。**2.2 解析方法** IR法では、AD 検定量<sup>4)</sup>A<sup>2</sup>の値がある基準値以下となるような標本集団を、PPS 標本集団として元の標本集団から再抽出する。これまでの山林小流域の検討では、非超過確率  $\gamma=5\%$  に相当する一様乱数列に対する A<sup>2</sup>の値を基準とすると適切な負荷量の推定が可能であることが分かっている。この基準が他流域でも妥当であるか確認するため、今回  $\gamma=1, 5, 10\%$  について米国のデータを用いて負荷量推定を行った。負荷量推定では、各流域で単年度毎8年分の流出負荷量について評価した。具体的には年度毎に52個の標本を時間軸上でランダム抽出し、これをもとに IR 法で負荷量推定を行った。負荷量の点推定量の偏りについては、 $pBIAS = (E[\hat{L}] - L_{true})/L_{true}$  ( $\hat{L}$ ; 負荷推定量,  $L_{true}$ ; 真の値) が  $\pm 10\%$ 以内でおおむね不偏と判断した。区間推定については95%信頼区間を bootstrap-t 法で構成し、被覆確率 (信頼区間内に真値が含まれる割合) が90%以上であればおおむね適切に信頼区間が構成されているものとした。また、再抽出された PPS 標本数  $n_{pps}$  が4以下となり信頼に足る負荷量の区間推定ができない場合は推定不能としている。

Table 1 Watershed summary  
(出典<sup>2)</sup>: Basic station information)

| 流域名       | 流域面積(km <sup>2</sup> ) | 土地利用(%) |    |    |
|-----------|------------------------|---------|----|----|
|           |                        | 農地      | 森林 | 都市 |
| Chickasaw | 43                     | 88      | 3  | 9  |
| Cuyahoga  | 1,830                  | 21      | 34 | 40 |
| Grand     | 1,774                  | 40      | 50 | 1  |
| Honey     | 308                    | 83      | 10 | 7  |
| Lost      | 11                     | 86      | 8  | 4  |
| Maumee    | 16,388                 | 80      | 6  | 11 |
| Muskingum | 19,215                 | 42      | 43 | 12 |
| Raisin    | 2,698                  | 68      | 11 | 11 |
| Rock      | 90                     | 80      | 11 | 9  |
| Scioto    | 9,965                  | 70      | 11 | 17 |
| Sandusky  | 3,239                  | 82      | 9  | 8  |
| Vermilion | 679                    | 73      | 25 | 1  |

**3. 結果および考察** **3.1 点推定量の不偏性の評価** Fig. 1 に各  $\gamma$  に対する  $pBIAS$  と被覆確率の

(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) 流出負荷量, 不偏推定, IR 法, LQ 式, Lake Erie

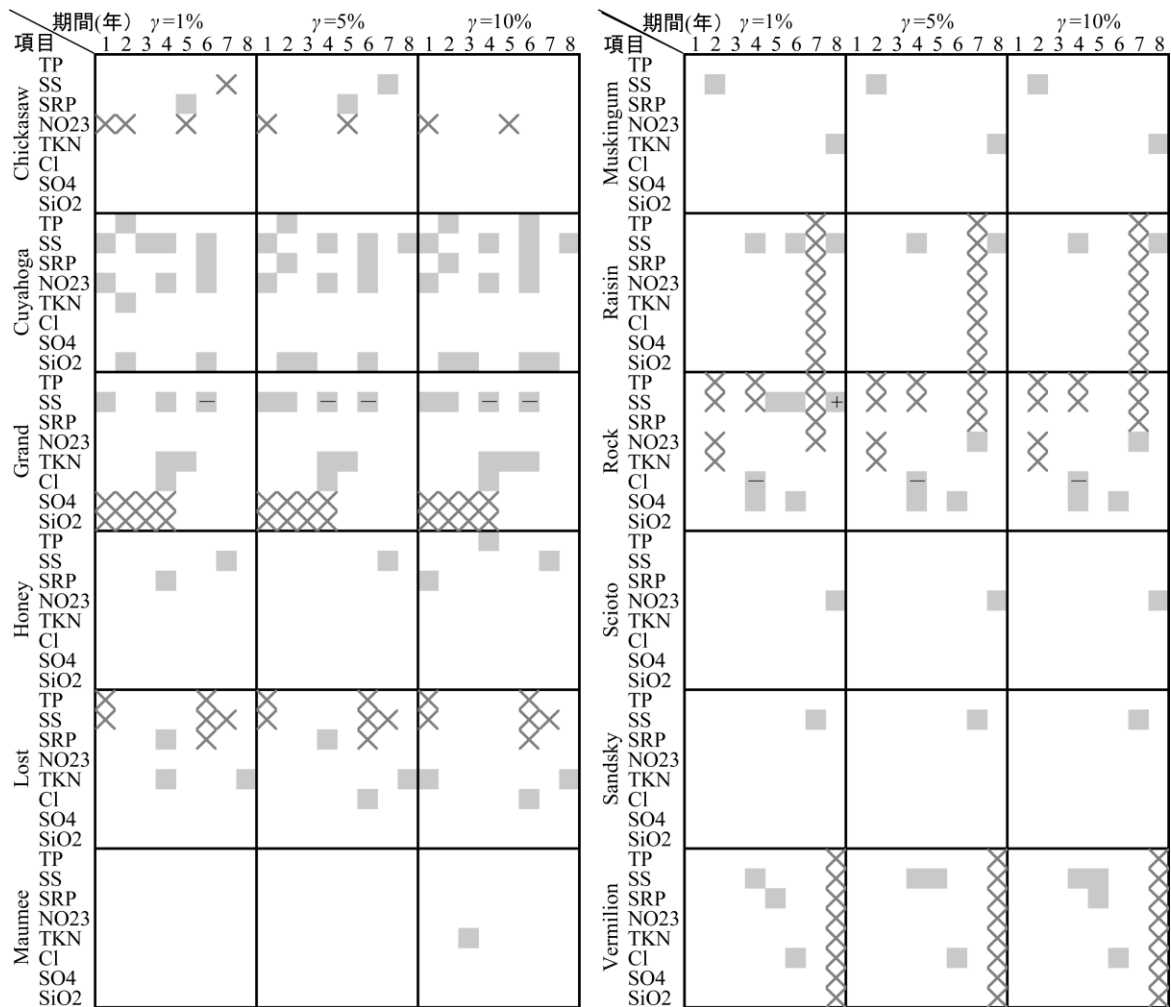


Fig. 1 各 $\gamma$ での $pBIAS$ と被覆確率

Fig. 1  $pBIAS$  values and coverage rates of CIs for each  $\gamma$  value.

95% Coverage rate  $pBIAS$  (注1)  $NO_{23}: NO_3^- + NO_2^-$   
 ■  $\pm 5\% \sim$  +  $+10\% \sim$  (注2)  $SO_4: SO_4^{2-}$   
 X 推定不能 -  $\sim -10\%$  (注3)  $SiO_2: SiO_2$

計算結果を示した。この図から、 $|pBIAS|$  はほとんどの場合で 10% 以下であった。SS では  $\gamma$  に関わらず最大で 2 割程度の偏りが見られた。これらは  $n_{pps}$  が小さく回帰残差が大きい場合であった。3.2 被覆確率の評価と採用すべき妥当な  $\gamma$  の検討 被覆確率が妥当な範囲となるデータセットの数は  $\gamma = 1\%$  で最も多いものの、 $\gamma$  が小さいために PPS 標本が減少してしまった。一方  $\gamma = 5\%$  と  $10\%$  の間には有意な差は見られていない。以上より、米国の 12 流域からのデータセットに対して今回検討した  $\gamma$  の中では、 $\gamma = 10\%$  が最も妥当な基準と考えられた。

4. 終わりに 本検討により、米国の 12 流域に対しても IR 法により不偏推定が可能であることが示された。また、再抽出の基準となる  $A^2$  値は非超過確率  $\gamma = 10\%$  に対する値でよいことも分かった。ただし、山林流域の  $K^+$  の 10 分間隔高頻度データでは  $\gamma = 5\%$  を要求していたため、頑強な推定結果を得るためには  $\gamma$  の値として 5% を採用すべきと考える。

参考文献 1) 多田明夫, 田中丸治哉: 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, 2017 2) NCWQR: <https://ncwqr.org/monitoring/data/>, 2019/10/1 閲覧 3) USGS: <https://waterdata.usgs.gov/nwis/sw>, 2019/1/5 閲覧 4) M. A. Stephen: EDF Statistic for goodness of fit some comparison, *J. Am. Stat. Assoc.*, 730-737, 1974