予測誤差を用いた水文確率値の評価基準の研究

北海学園大学大学院工学研究科　○学生員　岡田拓巳
北海学園大学工学部社会環境工学科　フェロー　許士達広
帯広市役所　後藤翔生

1. はじめに

河川計画に用いられる水文確率値の算定については、一般には既に一定の理論的整理がなされていると考えられている。しかし現実にはどのような確率モデルが良いか、その算定方法はどの方式が良いか、それを何で評価するかについては、依然として明確な答えは得られていない。本研究では水文確率の最適値の算定に重要な根本的問題である「何をもって最適とするか」の判断基準について考察し、モデルの適合性と変動性の両方を評価する指標として、ジャックナイフ誤差分散を用いた予測誤差の適用を検討する。

2．予測誤差

水文確率モデルの評価指標は適合性を示すものと変動性を示すものに大別される。適合性はモデルに使用したデータと推定値の差を分布全体について見るものであり、具体的指標として誤差分散 se^2、決定係数、SLSC などが用いられている。変動性は観測データが追加されたとき計画値が大きく変動しないように、推定値の変動する度合いを推定点ごとに見るもので、モデルとなる関数ごとに積率法や最尤法による推定誤差分散の式が導出されているほか、ジャックナイフ法やブートストラップ法といったリサンプリング法が使用される。どちらを優先するかが問題であるがこの二つの評価指標は一般的に両立しない。データへの適合性が高いモデルはデータの変動に追随して変化し、変動が大きくなるためである。

必要なことは各種の確率分布と算定方法に対する評価を、適合性と変動性の両方を考えた指標で算出することである。回帰式においては予測誤差の概念があり、推定点に対する回帰推定の評価を適合性だけでなく変動性も加えた誤差で表すことができる。式１）の１項目が残差分散で適合性を、２，３項目は残差による変動誤差分散を表す。

$$S_T^2 = V(y-\hat{y}) = s_e^2 + s_e^2\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right] = s_e^2\left[1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right] \quad \cdots 1)$$

ただし, S_T^2 : 予測誤差分散　S_{xx} 偏差平方和　s_e^2 : 残差による誤差分散　x_0 : 推定点　n : データ数

点 x_0 の予測誤差分散を全点で平均すると 多変量　$\hat{y}_i = a_0 + a_1x_1 + a_2x_2 + \cdots + a_px_p$　のとき

$$\overline{S_T^2} = = \frac{n+p+1}{n-p-1} s_e^{2*} \quad \cdots 2) \qquad \text{ただし} \quad s_e^{2*} = S_e/n$$

Se：残差平方和

となる。2)式は最終予測誤差と呼ばれており、パラメータの数 k =p+1 に置き換え対数をとり n 倍すると、AIC と同じもの $(nlogs_e^{2*} + 2k)$ が得られる。すなわち AIC は適合性に加え分布全体の変動性をパラメータ数として含んでいる指標である。また確率値の精度は、特に外挿する場合分布全体でなく推定する点の精度で表す必要があるため、予測誤差分散が適合度と変動性を合わせた推定点の評価指標として有効である。

3．確率値の変動誤差の算出

確率推定値の場合横軸はデータではなく標準変量であるため、変動誤差分散は１）式とは別の式となり、２母数分布以外はかなり複雑になる。確率値の変動誤差分散の理論式は一般に積率法で以下のように与えられる。

$$S_T^2 = \frac{\mu_2}{n}\left\{1 + K\gamma_1 + \frac{K^2}{4}[\gamma_2 - 1] + \frac{\delta K}{\delta\gamma_1}[2\gamma_2 - 3\gamma_1^2 - 6 + K(\gamma_3 - 6\gamma_1\gamma_2/4 - 10\gamma_1/4)] + \left(\frac{\delta K}{\delta\gamma_1}\right)^2\left[\gamma_4 - 3\gamma_3\gamma_1 - 6\gamma_2 + \frac{9\gamma_1^2\gamma_2}{4} + \frac{35\gamma_1^2}{4} + 9\right]\right\} \quad \cdots 3)$$

$\gamma_1 = \mu_3/{\mu_2}^{3/2}$　$\gamma_2 = \mu_4/{\mu_2}^{2}$　$\gamma_3 = \mu_5/{\mu_2}^{5/2}$　$\gamma_4 = \mu_6/{\mu_2}^{3}$　μ_2～μ_6：2～6 次モーメント

$K = \dfrac{\varepsilon_p - \bar{\varepsilon}}{\sigma_\varepsilon}$　ε_p：確率値の標準変量　$\bar{\varepsilon}$：標準変量の平均　σ_ε：標準変量の標準偏差

グンベル分布について３)式でモーメントを全てグンベル分布の標準変量から算出した場合以下の式になる。

$$S_T{}^2 = \frac{\mu_2}{n}\left\{1 + 1.1396(\varepsilon_p - 0.5772)\frac{\sqrt{6}}{\pi} + 1.1(\varepsilon_p - 0.5772)^2\frac{6}{\pi^2}\right\} \quad \cdots 4)$$

３），４）式でμ_2は通常全変動であるが、予測誤差では残差変動$s_e{}^2$に直す必要がある。これら理論式の$S_T{}^2$はデータから推定される分布形状が理論分布に一致している場合の値であり、かつ最尤法と積率法など算出方法によって異なるため、データごとのモデルや手法による変動性の違いを厳密には比較することができない。これに対しジャックナイフ誤差分散はどのような分布関数や算出法によるものでも誤差分散の計算は同じである。このためここでは以下の５)式のように変動成分としてジャックナイフ誤差分散を用い、全変動を残差変動に直す補正を行った。これによりモデルの関数と算出方法の両方について、適合性と変動性を合わせた予測誤差により評価することができる。

$$S_T{}^2 = s_e{}^2 + V_J * s_e{}^2 / s_y{}^2 \quad \cdots 5)$$

V_j：ジャックナイフ誤差分散　$s_y{}^2$：全変動分散　$s_e{}^2$：残差分散

４．理論式とジャックナイフ法の変動誤差分散の比較

予測誤差の算出に５)式のジャックナイフ誤差分散を使う場合３)，４）の理論値とどの程度違いが生ずるか、理論式とジャックナイフ法の$S_T{}^2$値の比較を行う。全国 20 か所の日雨量年間最大値について計算を行うと、一般化極値分布で理論式のモーメントをデータから求めた場合、両者の比は 20 か所の平均で 0.767、グンベル分布でモーメントをデータから求めた場合 0.931、同じくモーメントを標準変量から求めた場合 0.912 となり、理論式とジャックナイフ法で大きな違いは生じない。このためジャックナイフ誤差分散を用いて５)式から予測値の誤差分散を求める。

５．確率分布への適用

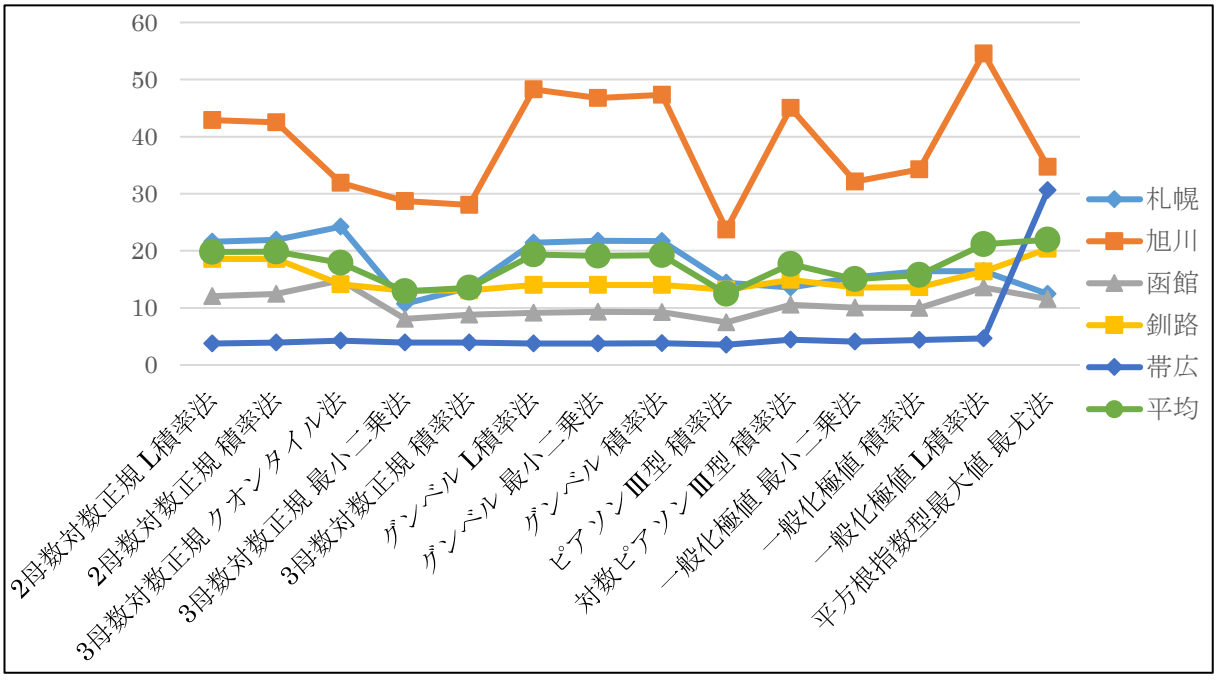

図-1 適合性と変動性を合した指標(予測誤差)

図は北海道内の５観測所の過去 100 年の既往年最大日降雨について、モデルと計算法の組み合わせ別に予測誤差を算出したものである。計算は SLSC やジャックナイフ誤差分散等の指標についても行い、公開ソフト水文統計ユーティリティーと比較し確認している。ピアソンⅢ型積率法、３母数対数正規最小二乗法(α＝0.5))が比較的良好となった。

参考文献　後藤翔生，岡田拓巳，許士達広：水文確率値の評価基準の研究　土木学会北海道支部 74 号　B-55　2018 年 1 月