

期別の水文確率値に関する研究

A study for the variation of hydrological probability according to the difference of period

北海学園大学社会環境工学科

○学生員 佐藤慎哉 (Shinya Sato)

北海学園大学社会環境工学科

学生員 岡田太一 (Taichi Okada)

北海学園大学社会環境工学科

正員 許士達広 (Tatuhiko Kyoshi)

1. まえがき

ダム貯水容量の期別配分は、一般的に制限水位方式と称して洪水期と非洪水期の2つに期間を分けるか、年間を通じて治水容量が一定なオールサーチャージ方式がとられている。既存ダムの下流に環境用水の放流が無いダムが多く存在し、放流が望まれるが、ダムの治水容量は決められており、現在の容量配分の範囲では放流できない。

一方、非洪水期に洪水調節のための容量が無く、5～6月や10～11月に降雨や融雪による大きな出水があった時に危険なダムがある。近年ダムの弾力管理や、事前放流の検討、但し書き操作の見直しなど、ダムの治水や利水さらにダムの安全のための機能を高めようという試みが行なわれているが、効果は限定的である。

もしも貯水池の治水容量配分をより期別に細分化すれば、治水容量と利水確保容量との関係で、所定の治水および利水の安全度を維持しながら運用に余裕が生じることが考えられ、近年行われている貯水位の弾力的管理や事前放流によらなくとも、利水と治水の効果を高めることが可能となる。一般に制限水位方式の洪水期は7～9月で設定されているが、降雨の実績は7月に比較的少なく、10月はかなり多い。ここでは実態に合わせて細分化した期別の治水容量を定めるために、降雨や流量データを期別に区分する手法について基礎的な検討を行う。

2. 期間確率の必要性

洪水調節容量を機別に細分化して定めるためには、水文データの期別区分の考え方が必要である。例えば9月1日といった特定の日を考え、その日の100年確率の降雨に対応するためのダムの治水容量は、どれだけ空けておくべきかを考える。9月1日のみの降雨データから確率を取った場合では、過去に大きな雨が観測されていない場合が多く、逆に9月5日や10日に大きな雨が降った記録があれば、同様の雨が9月1日にも降る可能性が大きいと、安全性の面から不十分である。降雨確率値を求めるには、どの期間の範囲まで同じ規模の現象が起こりうるかを知る必要があり、その期間を区分して降雨最大値に対応しなくてはならない。

3. 期間区分・統合の手法

一定年数の日雨量データがあるとして、それらを10日あるいは20日といった期間に分け、それぞれの期間における各年の日雨量最大値を抽出する。このとき隣接する期間が同程度の降雨と見なされる場合には、隣接する各期間がそれぞれ20日であれば40日に統合し、統合した後の各期間について降雨確率を求める。

期間の区分・統合の判断は、期別に分けた過去の年最大日雨量のデータ群が、互いに同じ母集団か別の母集団かを判別することで行うことが出来る。そのためには期別の降雨量（期別の年最大日雨量）のデータ群について差の検定を行なう。

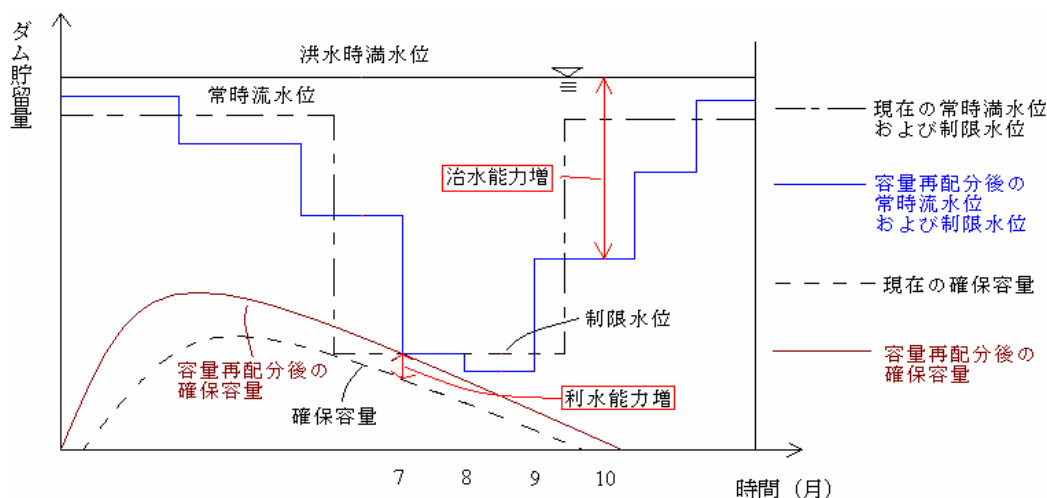


図-1 貯水池容量再配分の効果

主要な水文確率値は以下のように表わされる。

$$\begin{aligned} \hat{y}_p &= \bar{y} + \sigma_y \varepsilon \\ \hat{y}_p &: \text{非超過確率 } p \text{ の水文確率値} \\ \bar{y} &: \text{水文データ } y \text{ の平均} \\ \sigma_y &: \text{データ } y \text{ の標準偏差} \\ \varepsilon &: \text{確率に対応する標準変量} \end{aligned}$$

この式に示されるように水文確率値を求める観点から母集団を区別する要素は、平均と標準偏差（分散）であることが分かる。このため、期間毎のデータ群（標本）の区分・統合のためには、データ群の間の母平均の差の検定と母分散の差の検定を行なうのが一つの方法である。

上記の2つの期別データの違いの検定の問題として以下のようなことが考えられる。

① t - 検定や F - 検定は2つの分布が正規分布をしている（あるいはデータ数が多く、平均値が正規分布とみなされる）という前提で行なわれる。どちらかが正規分布でない場合はノンパラメトリック検定を行なうが、t、F 検定とは検出力に差があるため、同等の比較とはいえない。

② 上記の検定は水文データ x に対して $y = 1 - n_x$ とおいた対数正規分布が仮定されるが、実際は最適となる確率分布は比較する期間によって異なることが考えられる。厳密には、例えば一般化極値分布で最適確率値を求める期間と、対数ピアソンⅢ型で確率値を求める期間の同一性を、対数正規分布で仮定して行なうのは矛盾がある。

4. 確率値の差の検定方法

上記の問題に対応するための新たな方法として各期間で確率値を計算して、確率値の分散を利用した確率値の差の検定を行なう。

1) 確率値の差の検定

二つの期間のデータによる確率値が x_{p1}, x_{p2} でありそれぞれのデータ数が n_1, n_2 誤差分散が S_{T1}^2, S_{T2}^2 で表されるとき、確率値の差の検定は

$$t = \frac{x_{p1} - x_{p2}}{\sqrt{S_{T1}^2 + S_{T2}^2}}$$

の t 検定で行なわれる。自由度 f は $S_{T1}^2 = S_{T2}^2$ のとき

$$f = n_1 + n_2 - 4$$

$S_{T1}^2 \neq S_{T2}^2$ のとき

$$f = \frac{(S_{T1}^2 + S_{T2}^2)^2}{\left(\frac{S_{T1}^4}{n_1 - 2} + \frac{S_{T2}^4}{n_2 - 2} \right)}$$

であり、確率値の誤差分散が等しいかどうかは F 検定で判断する。

$$F = \frac{S_{T1}^2}{S_{T2}^2} \quad \text{自由度 } f = n_1 + n_2 - 4$$

確率値の誤差分散として用いられるものに①水文デー

タと適用した確率分布モデルの間の誤差分散、②ブートストラップ法などのリサンプリング法による繰り返し計算から求まる分散③低要する確率分布の性状により導かれる確率値の母分散の3つがある。ここでは確率地を求める二つの期間の母集団を区分することが目的であるので、③の方法を用いる。

2) 推定誤差 S_T の基本式

誤差分散 $\text{Var}(\hat{x}_p)$ を S_T^2 と表す。一般的に、再現期間 T 年の確率値 x_T 誤差分散は積率の推定値を用いて以下のように導かれる。

もし確率値 x_T が、分布における1次、2次、3次の積率と再現期間 T 年を用いて、以下のように表されるとする。

$$x_T = f(m_1', m_2, m_3, T)$$

ただし、 m_1' は原点周りの積率、 m_2 は平均値周りの積率である。このとき、確率値 x_T の誤差分散 S_T^2 は、分散 Var 、共分散 cov を用いて、

$$\begin{aligned} S_T^2 &= \left(\frac{\partial x_T}{\partial m_1'} \right)^2 \text{var} m_1' + \left(\frac{\partial x_T}{\partial m_2} \right)^2 \text{var} m_2 + \left(\frac{\partial x_T}{\partial m_3} \right)^2 \text{var} m_3 + 2 \frac{\partial x_T}{\partial m_1'} \cdot \frac{\partial x_T}{\partial m_2} \text{cov}(m_1', m_2) \\ &\quad + 2 \frac{\partial x_T}{\partial m_1'} \cdot \frac{\partial x_T}{\partial m_3} \text{cov}(m_1', m_3) + 2 \frac{\partial x_T}{\partial m_2} \cdot \frac{\partial x_T}{\partial m_3} \text{cov}(m_2, m_3) \end{aligned}$$

となる。実際には確率値 x_T は積率を用いて、

$$x_T = m_1' + K \sqrt{m_2}$$

と表される。ここで、 K はひずみ係数 γ の関数である。

これを解いてここで、既往の研究により、母集団のモーメントを用いて置き換え、さらにひずみ度、尖り度などを用いて単純化する。この結果もしも係数 K が、ひずみ係数に依存しない確率分布では

$$S_T = \delta \sqrt{\frac{\sigma^2}{n}} \quad \begin{array}{l} \sigma: \text{分布の分散} \\ n: \text{データ数} \end{array}$$

となる。ただし係数 δ は以下のように表される。

$$\delta = \left\{ 1 + \gamma_1 K + (\gamma_2 - 1) \frac{K^2}{4} \right\}^{1/2}$$

例えば、正規分布においては、

$\mu_3 = 0, \mu_4 = 3\sigma^2, \mu_2 = \sigma^2$ 、(σ : 母標準偏差) より $\gamma_1 = \mu_3 / \mu_2^{3/2} = 0$ 、 $\gamma_2 = \mu_4 / \mu_2^2 = 3\sigma^2 / \sigma^2 = 3$ なることから、係数 δ は

$$\delta = \left\{ 1 + \frac{K^2}{2} \right\}^{1/2}$$

となる。ここで、係数 K は、正規分布の標準変数 Z_p であり、また、 μ_2 を標本分散 S_x^2 とすると、正規分布における確率値 x_p の誤差分散は、以下になる。

$$S_T^2 = \frac{S_x^2}{n} \left(1 + \frac{Z_p^2}{2} \right)$$

3) ゲンベル分布の推定誤差の算定方法

ゲンベル分布において、非超過確率 P に対する水文学

確率値の推定誤差分散 S_T^2 は、積率法により決定した尺度母数と標準変量の2つのパラメータをもとに、漸近的に以下の式から概算することができる。

$$S_T^2 = \frac{\alpha^2(1.11+0.52y+0.61y^2)}{n}$$

ここで、 n : データ年数、 α : 積率法によるグンベル分布の尺度母数 a y : グンベル分布の標準変量

$$G_p = -\ln[-\ln(p)]$$

4) 分布の差の検定と確率値の差の検定の関係
確率値は 一般的に

$$\hat{x}_p = \bar{x} + \sigma_x \varepsilon_p$$

確率値の分散は

$$S_T^2 = \delta^2 \frac{\sigma^2}{n}$$

で表されるため2つの確率分布の種類が同じ場合の分布の差の検定は

$$t = \frac{x_{p1} - x_{p2}}{\sqrt{S_{T1}^2 + S_{T2}^2}} = \frac{(\bar{x}_1 + \sigma_1 \varepsilon_{p1}) - (\bar{x}_2 + \sigma_2 \varepsilon_{p2})}{\sqrt{\delta^2 \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)}} = \frac{(\bar{x}_1 - \bar{x}_2) + (\sigma_1 - \sigma_2) \varepsilon_p}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} * \delta}$$

式のうち左の部分は分布の平均値の差の検定と同じ式であり、確率値の差の検定は分布の差の検定の要素を含むことが分かる。

5、降雨区分の計算例

道内の雨量観測所を対象に、降雨量の期別区分を行なった。一雨降雨の単位としては1日、2日、3日が考えられるが、河川整備計画においては3日降雨をとる例が多いことから、3日を基本とした。またそれに対応して降雨確率をとる最小期間幅としては5日、10日、15日、20日を対象に、2母数対数生起確率とグンベル分布の当てはまり具合を検討したが、期間幅が短いほど対数正規分布のひずみ度が大きくなり妥当な値が求まらず、またグンベル分布のSLSCも大きいことから、20日間を最小単位とした。

表は旭川の50年間のデータについて4節に示した方法で年間の降雨を20日間ごとと期間に区分し、それぞれの期間の確率値が同一と見なせるか、t検定、及びF検定により統合区分の検討を行なった例である。表の○印は確率値が同一と見なせるもの、棄却と書いたものが同一と見なせないものである。また網掛けの部分はF検定で分散が等しくないと判断されウェルチ検定を行なった部分である。○印が隣接する場合に限り統一するものとする、2~4期間の単位で統合するのが適当である区間が存在することが分かる。

このようにして統合された区間についてそれぞれ年間最大値を取り、グンベル法で100年確率値を③周知したものを示したのが下図である。多くの期間は、複数の区分に重複しているが、治水計画の安全上重ね合わせたうちの大きい方の値に対応するのが適当である

旭川	3/2~3/21	3/22~4/10	4/11~4/30	5/1~5/20	5/21~6/9	6/10~6/29	6/30~7/19	7/20~8/8	8/9~8/28	8/29~9/17	9/18~10/7	10/8~10/27	10/28~11/16	11/17~12/6	12/7~12/26
3/2~3/21		○	○	棄却	棄却	棄却	棄却	棄却	棄却	棄却	棄却	棄却	棄却	棄却	棄却
3/22~4/10			○	○	○	棄却	棄却	棄却	棄却	棄却	棄却	棄却	○	○	○
4/11~4/30				○	○	棄却	棄却	棄却	棄却	棄却	棄却	棄却	○	○	○
5/1~5/20					○	○	棄却	棄却	棄却	棄却	棄却	棄却	○	○	棄却
5/21~6/9						○	棄却	棄却	棄却	棄却	棄却	棄却	○	○	○
6/10~6/29							○	棄却	棄却	棄却	棄却	○	○	○	棄却
6/30~7/19								棄却	棄却	棄却	○	○	棄却	棄却	棄却
7/20~8/8									○	○	棄却	棄却	棄却	棄却	棄却
8/9~8/28										○	棄却	棄却	棄却	棄却	棄却
8/29~9/17											棄却	棄却	棄却	棄却	棄却
9/18~10/7												○	棄却	棄却	棄却
10/8~10/27													棄却	棄却	棄却
10/28~11/16														○	○
11/17~12/6															○
12/7~12/26															

表-1 期間の組み合わせによるt検定とF検定

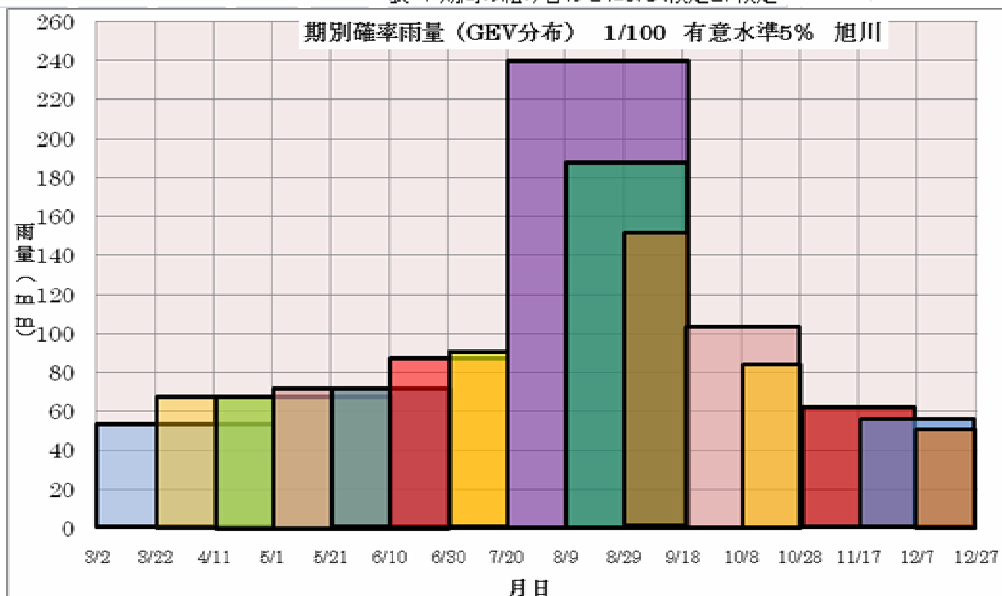


図-2 期別確率値

このようにして、期別に降雨確率値を求め、流出解析及び放流計算を行って期別の必要な治水容量を求め、利水容量との再配分により、より高度な貯水池運用が期待される。

6. AICによる区分

t検定やF検定には有意水準が設定され、通常5%であるが、10%や1%がとられることもある。有意水準次第で、二つの期間を統一するかどうかが変わるため、そのあり方が議論になる可能性がある。これを解消する方法として情報量基準 AIC の利用が考えられる。AIC は、データと仮定したモデルとの適合度を知る指標であり、

$$AIC = n \log 2\pi + n \log \hat{\sigma}^2 + n + 2k$$

$\hat{\sigma}^2$: 残差分散 n : データ数

k : パラメータ数 (確率分布の母数+1) で表される。

AIC はその絶対値は問題ではなく、仮定したモデルごとに得られる AIC の差に意味がある。AIC の差が 1~2 程度以上であればその差は有意と考えられ、AIC の値が小さいほうが良いモデルあるいは良い仮説である。

前節の問題を AIC で表現すると①確率値も分散も等しくない。 $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$ ②平均も分散も等しい $(\hat{\mu}, \hat{\sigma}^2)$ ③分散のみ等しい $(\hat{\mu}_1, \hat{\mu}_2, \sigma^2)$ ④平均のみ等しい $(\hat{\mu}, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$ の4つについて計算を行い、②の場合のみ分布が同一とみなすことにすればよい。ただし

$\hat{\mu}_1$: データ n_1 による確率値

$\hat{\mu}_2$: データ n_2 による確率値

$\hat{\sigma}_1^2$: データ n_1 による確率値の誤差分散

$\hat{\sigma}_2^2$: データ n_2 による確率値の誤差分散

でそれぞれがデータから計算で求められる推定値である。

①の場合は

$$AIC1 = n_1 \log 2\pi + n_1 \log \hat{\sigma}_1^2 + n_1 + 2k_1$$

$$AIC2 = n_2 \log 2\pi + n_2 \log \hat{\sigma}_2^2 + n_2 + 2k_2$$

$$AIC = AIC1 + AIC2$$

$$= n \log 2\pi + n_1 \log \hat{\sigma}_1^2 + n_2 \log \hat{\sigma}_2^2 + n + 2k \quad (k=4)$$

②の場合は

$$\hat{\mu} = \frac{n_1 \hat{\mu}_1 + n_2 \hat{\mu}_2}{n_1 + n_2}$$

$$\hat{\sigma}^2 = \frac{n_1}{n} \left\{ \hat{\sigma}_1^2 + (\hat{\mu}_1 - \hat{\mu})^2 \right\} + \frac{n_2}{n} \left\{ \hat{\sigma}_2^2 + (\hat{\mu}_2 - \hat{\mu})^2 \right\}$$

$$AIC = n \log 2\pi + n \log \hat{\sigma}^2 + n + 2k \quad (k=2)$$

③の場合は

$$\hat{\sigma}^2 = \frac{n_1}{n} \hat{\sigma}_1^2 + \frac{n_2}{n} \hat{\sigma}_2^2$$

$$AIC = n \log 2\pi + n \log \hat{\sigma}^2 + n + 2k \quad (k=3)$$

④の場合は

$$\mu^3 - A\mu^2 + B\mu - C = 0$$

$$\text{ただし } D_1 = \frac{n_1}{n} (\hat{\sigma}_2^2 + \hat{\mu}_2^2) \quad D_2 = \frac{n_2}{n} (\hat{\sigma}_1^2 + \hat{\mu}_1^2)$$

$$\text{として } A = \left(1 + \frac{n_1}{n}\right) \hat{\mu}_2 + \left(1 + \frac{n_2}{n}\right) \hat{\mu}_1$$

$$B = 2\hat{\mu}_1 \hat{\mu}_2 + D_1 + D_2 \quad C = \hat{\mu}_1 D_1 + \hat{\mu}_2 D_2$$

3 次方程式の解は実根が 1 つだけの場合はそれが採用され 3 つ有る場合は正でかつ対数尤度が最大にするものを求める。この解は μ の最尤推定量 $\tilde{\mu}$ である。

σ_1 と σ_2 の最尤推定量は

$$\tilde{\sigma}_1^2 = (\tilde{\mu} - \hat{\mu}_1)^2 + \hat{\sigma}_1^2 \quad \tilde{\sigma}_2^2 = (\tilde{\mu} - \hat{\mu}_2)^2 + \hat{\sigma}_2^2$$

AIC は以下ようになる。

$$AIC = n \log 2\pi + n_1 \log \tilde{\sigma}_1^2 + n_2 \log \tilde{\sigma}_2^2 + n + 2k \quad (k=3)$$

上記の検討を、先に示した 5 節と同じ期間区分で検討した。紙面の都合で一部を示すが、このようにほとんどの期間が統合されない結果となっている。統計の分野で提案されている新しい方法の適用を図ったものであるが、利用法についてさらに検討が必要である。

表一 AIC による期間の平均と分散の比較

平均と分散	期間組み合わせ	3/22~4/10	4/11~4/30
全て等しくない	3/2~3/21	566.9	562.8
	3/22~4/10		594.4
全て等しい	3/2~3/21	1807.5	1807.5
	3/22~4/10		1851.7
分散等しい	3/2~3/21	569.8	564.5
	3/22~4/10		592.4
平均等しい	3/2~3/21	646.9	654.2
	3/22~4/10		593.4
一番低い AIC	3/2~3/21	566.9	562.8
	3/22~4/10		592.4

7. 終わりに

貯水池運用の高度化に向けて期間区分方法の検討を行った。確率値及びその分散を求めて区分する方法は、データそのものの平均や分散の差を検定する方法の要素を含み、結果も比較的に類似してくることが予想される。AIC については、さらに検討が必要である。このほかノンパラメトリック検定を主体とする方法もあり、今後実用的で理解されやすい方法を選定していきたい。

参考文献

- 1) G.W.Kite FREQUENCY AND RISK ANALYSES IN HYDROLOGY by WATER RESOURCES PUBLICATIONS pp.33-35、pp.69-77、pp.97、pp.105-113、pp.123-127
- 2) 坂井慶行、石黒真木夫、北川原四郎：情報量統計学 P42 - 64、p107-122、p127-133
- 3) David R. Maidment HANDBOOK OF HYDROLOGY McGraw - Hill, Inc Chapter 18 pp.18.29-18.33