

極値分布の母数推定法の比較評価

宝 馨・高棹 琢馬・清水 章

COMPARISON OF PARAMETER ESTIMATION METHODS FOR EXTREME VALUE DISTRIBUTIONS

By *Kaoru TAKARA, Takuma TAKASAO and Akira SHIMIZU*

Synopsis

This paper compares parameter estimation methods for the Gumbel distribution and the generalized extreme-value (GEV) distribution by using the Monte Carlo simulation technique with changing sample size $N=10, 20, \dots, 1000$.

The methods considered for the Gumbel distribution are the methods of moments (MoM), maximum likelihood estimation (MLE), probability weighted moments (PWM), least-squares (LS) with the Weibull or the Hazen plotting formula, and maximum entropy principle (PME). In terms of the root mean square error (RMSE) of the quantile estimates, the MLE is the best among these. The PWM gives the least biased estimation, followed by the MoM, PME and MLE. The difference of these is small. For the GEV distribution, the MoM, MLE and PWM are compared. In terms of both the RMSE and bias the PWM is the best.

1. はじめに

水工計画・設計の基本量を定めるには、当該の水文データを収集し、そのデータに水文学的および確率統計学的観点から一定の評価を加えた後、その水文学量の所定の再現期間に対応する値を求めるという手順を踏む。この手順を水文頻度解析と呼ぶ。

本論文では、水文頻度解析においてしばしば用いられる極値分布 (Gumbel 分布と一般化極値分布) の種々の母数推定法の優劣を数値実験により明らかにしようとする。特に、確率水文学量の推定値の偏倚と誤差に着目し、標本の大きさ (データ数) とも関連づけて検討する。

2. 母数推定法評価の手順と規準

以下の手順により母数推定法の優劣を評価する。

- Step 1) 母集団を想定する。すなわち、母集団の確率分布と母数を仮定する。
- Step 2) 母集団から大きさ N の標本を抽出する。すなわち、仮定した分布に従う乱数を N 個発生させる。
- Step 3) 発生させた標本に対して、各々の推定法により母数を推定し、確率水文学量を推定する。
- Step 4) Step 2 と Step 3 を M 回繰り返す。各々の推定法による母数と確率水文学量の推定値の平均値と標準偏差 (推定誤差) と平均平方誤差 (MSE, mean square error) を算定する。

Step 5) MSE の値が小さい (すなわち, 真値からの偏りが小さく, かつ, 推定誤差が小さい) 推定値を与える母数推定法が良いと判定する。

ここで, 平均平方誤差 (MSE) について説明を加えておく。MSE は推定量の偏倚と推定誤差を併せて評価できる指標である。ある統計量 θ の推定量 $\hat{\theta}$ に関する MSE は次式で与えられる。

$$\begin{aligned} \text{MSE} &= E[(\hat{\theta} - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2 + (E[\hat{\theta}] - \theta)^2] \\ &= \text{Var}[\hat{\theta}] + \{\text{Bias}[\hat{\theta}]\}^2 \dots\dots\dots (1) \end{aligned}$$

ここに, $E[\cdot]$ は期待値操作を, $\text{Var}[\cdot]$, $\text{Bias}[\cdot]$ はそれぞれ推定量 $\hat{\theta}$ の分散と偏りを表す。すなわち, MSE は, 推定量 $\hat{\theta}$ のばらつき具合と偏りの大きさの和で表されるものであり, この値が小さいほど良い推定であると言える。以下では, MSE の平方根をとったもの (平方根平均二乗誤差, root mean square error, RMSE) を用いる。100 年確率水文量 x_{100} の推定値, 推定誤差と RMSE の関係を図示したのが Fig. 1 で, 図中 $Sd[\cdot]$ は $(\text{Var}[\cdot])^{1/2}$ を表す。

上の作業により, 想定した分布の種々の母数推定法の中でどれが良いのかが明らかになる。また, 標本の大きさ N を変化させて Step 1 から Step 5 の手順を行うことにより, N の大きさによって, 最適な母数推定法が異なるのか, N の増加に伴って母数と確率水文量の推定誤差がどのように推移するのかが明

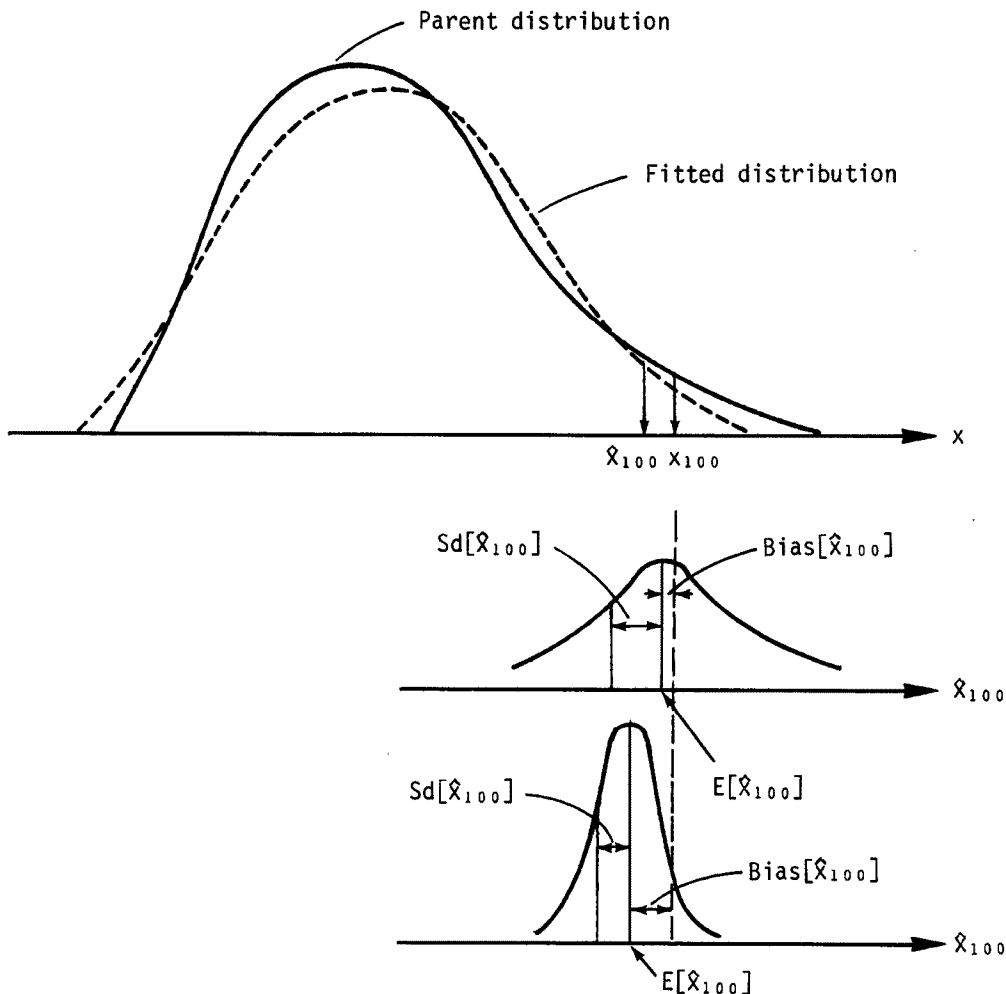


Fig. 1. Explanation of the root mean square error. (RMSE)

らかになる。

以下では、Gumbel 分布、一般化極値分布についてこの手順を適用し、母数推定法の良否を判定する。想定する母分布には、大阪の年最大日雨量の 92 年間（1889 - 1980 年）のデータ¹⁾および St. Marys 川の年最大日流量の 59 年間（1916 - 1974 年）のデータ²⁾に、上の二つの分布を最尤法であてはめたときの母数推定値に近いものを母数として採用した（Table 1）。

なお、標本の大きさ $N = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500, 1000$ とし、それぞれに対し、発生させる標本の個数 $M = 5000$ とした。

Table 1. Parameter values assumed

Distributions	Parameters	Rain-fall	Flood
Gumbel	α	0.04	2.50×10^{-4}
	u	77.0	1.20×10^4
GEV	x_0	75.0	1.20×10^4
	a	20.0	3.70×10^3
	k	-0.1	-0.05

3. Gumbel 分布の母数推定法の比較評価

3.1 Gumbel 分布

ある母集団から得られた 1 組の標本（1 群のデータ）の中の最小値および最大値（いわゆる極値）がどのような分布に従うかという問題を取り扱うのが極値統計論であり、1920 年代後半に Fréchet や Fisher と Tippett によって理論的に考察された。母集団分布に指数タイプ、Cauchy タイプ、有限タイプのもを仮定した場合に導かれる分布を、それぞれ、極値に対する第 I 種、第 II 種、第 III 種漸近分布と呼ぶ³⁾。

1940 年から 1950 年代にかけて、最大値に関する第 I 種漸近分布が年最大洪水流量により適合性を得ることが Gumbel によって確かめられたので、この分布は Gumbel 分布と呼ばれるようになった（この分布を最初に誘導した Fisher と Tippett の名をとって FT-I 分布、あるいはその式形から二重指数分布と呼ばれることもある）。

Gumbel 分布の確率密度関数は次式で与えられる。

$$f(x) = \alpha \exp[-\alpha(x-u) - \exp\{-\alpha(x-u)\}] \dots\dots\dots (2)$$

Gumbel 分布の母数推定法を比較評価した研究には、次のようなものがある。Lowery and Nask (1970)⁴⁾ は、積率法、回帰法（Weibull 公式を用いた最小二乗法）、Gumbel 法、最尤法の 4 方法を比較し、積率法がよいとしている。最尤法を否定しているのではないが、当時としては計算が煩雑なため最尤法の採用には積極的ではない。Lettenmaier and Burges (1982)⁵⁾ は、Gumbel が提唱し慣用されてきた手法の欠点を指摘し、さらに、積率法と最尤法の比較を行った。積率法では不偏分散より標本分散を用いる方がよく、また、小標本 ($N = 30$) でターンピリオドの大きい再現確率統計量（確率水量）の推定には、積率法より最尤法の方が若干よい結果を示すとしている。Raynal and Salas (1986)⁶⁾ は、積率法、最尤法、Weibull 公式を用いた最小二乗法、PWM 法、MIR (mode and interquartile range) 法、BLOCOS (best linear combination of order statistics) 法を比較検討した。その結果、最尤法が最も優れているが母数推定値に偏倚があり、偏倚の小さい推定法の中では、データ数が 20 以下のとき BLOCOS 法、さらにデータ数が多いとき PWM 法が良いことを示した。また、Weibull 公式を用いた最小二乗法、MIR 法は Gumbel 分布の母数推定に用いるべきでないとしている。Phien (1987)⁷⁾ は、積率法、最尤法、最大エントロピー法、PWM 法の比較を行っている。最尤法の優位性を示しているが、偏倚がやや大きいこと、PWM 法の偏倚は小さいが RMSE は大きく、最大エントロピー法が最尤法に次いで良いことを示した。Jain and Singh (1987)⁸⁾ は、米国における 104 ~ 2,500 km² の面積をもつ 55 流域の年洪水データに、七つの方法（積率法、PWM 法、混合積率 (mixed moments) 法、最尤法、不完全平均 (incomplete mean) 法、最大エントロピー法、最小二乗法) で Gumbel 分布をあてはめた。それらのうち最尤法が最も良く、続いて最大エントロピー法、PWM 法、積率法の順に良いが、それらの差はあ

まり大きくないとしている。

3.2 Gumbel 分布の母数推定法

本論文では、次の母数推定法を比較評価する。

- 最尤法
- 最小二乗法
- 積率法
- PWM 法
- 最大エントロピー法

これらの手法の概要を以下に示す。

(1) 最尤法

N 個のデータ x_1, x_2, \dots, x_N が与えられたとき、尤度関数 $L(\theta)$ は次式で定義される。

$$L(\theta) = \prod_{i=1}^N f(x_i; \theta)$$

ここに、 $f(x; \theta)$ は確率密度関数であり、 θ は母数のベクトルである。最尤法は $L(\hat{\theta}) > L(\theta)$ なる $\hat{\theta}$ を母数推定値とする方法である。通常、尤度関数をそのまま用いるのではなく、その対数をとった対数尤度

$$\log L(\theta) = \sum_{i=1}^N \log f(x_i; \theta)$$

の形式で用いられる。最尤法は、対数尤度を θ で偏微分して 0 とおいた連立方程式

$$\frac{\partial}{\partial \theta} \log L(\theta) = 0$$

を解くことになる。Gumbel 分布の場合、(2) 式に含まれる母数 a, u は

$$\frac{1}{a} = \frac{1}{N} \sum_{i=1}^N x_i - \frac{\sum_{i=1}^N x_i \exp(-ax_i)}{\sum_{i=1}^N \exp(-ax_i)}$$

$$u = -\frac{1}{a} \log \left\{ \frac{1}{N} \sum_{i=1}^N \exp(-ax_i) \right\}$$

という陰形式で与えられる⁹⁾ので、反復解法によってこれを数値的に解けばよい。ここでは、対数尤度を目的関数として、FACOM SSL II の改訂準ニュートン法による多変数関数の極小化プログラム DMINF 1¹⁰⁾を用いる。

(2) 最小二乗法

N 個のデータが与えられたとき、その値の小さい順に並べ替えた順序統計量を x_1, x_2, \dots, x_N とする。いま、 i 番目の順序統計量 x_i に対する非超過確率

$$P_i = F(x_i) = \int_{-\infty}^{x_i} f(x) dx = \exp[-\exp\{-a(x_i - u)\}] \dots \dots \dots (3)$$

をプロット・ポジション公式

$$P_i = \frac{i - \omega}{N + 1 - 2\omega} \dots \dots \dots (4)$$

で与えることにより、標準変量に関して最小二乗法を適用する¹¹⁾。(3) 式で $F(x_i)$ は累積分布関数、(4) 式で $\omega = 0$ のとき Weibull 公式、 $\omega = 0.5$ のとき Hazen 公式である。

(4) 式の P_i に対応する標準変量 s_i は、

$$s_i = F^{-1}(P_i) = -\log\{-\log P_i\} \dots \dots \dots (5)$$

一方、順序統計量 x_i の標準変量を s_i^* とすると、

$$s_i^* = \alpha(x_i - u) \dots\dots\dots (6)$$

であり、

$$\sum_{i=1}^N (s_i - s_i^*)^2 \rightarrow \min$$

となるような α, u を求める。ここでも、最尤法と同様 DMINF 1 を用いて数値的に解くことにする。また、用いるプロットング・ポジション公式は Weibull 公式と Hazen 公式の二つとし、両者の優劣をも検討する。

(3) 積率法

Gumbel 分布は 2 母数の分布であるから、2 次までの積率（平均値 μ と分散 σ^2 ）を用いることにより母数が推定できる。すなわち、以下の関係を用いる。

$$\mu = u + \frac{\gamma_E}{\alpha} = u + \frac{0.5772}{\alpha} \quad (\gamma_E: \text{Euler定数})$$

$$\sigma^2 = \frac{\pi^2}{6\alpha^2} = \frac{1.6449}{\alpha^2}$$

標本平均、標本分散（不偏分散）をこれらの式に代入して母数を求める。

(4) PWM 法

PWM 法は、1979 年に Greenwood ら¹²⁾により提案されたもので、積率法の一つと考えるとよい。通常の積率は、変量 x を原点まわり（あるいは平均値まわり）に累乗するが、PWM (probability weighted moments, 確率加重積率) は累積分布関数 $F(x)$ の累乗で加重した積率であり、次式で定義される。

$$M_{l,j,k} = E[X^l F^j (1-F)^k] \dots\dots\dots (7)$$

いま、累積分布関数 $F = F(x)$ の逆関数が陽形式 $x = x(F)$ で表記可能のとき、 $M_{l,j,k}$ は

$$M_{l,j,k} = \int_0^1 \{x(F)\}^l F^j (1-F)^k dF \dots\dots\dots (8)$$

となる。ここに、 l, j, k は実数である。 $j=k=0$ で l が非負の整数であるとき、 $M_{l,0,0}$ は原点まわりの l 次積率となる。すなわち、

$$M_{l,0,0} = \int_0^1 \{x(F)\}^l dF \dots\dots\dots (9)$$

j, k が非負の整数であるとき、 $M_{l,j,0}$ と $M_{l,0,k}$ とは次に示す関係がある。

$$M_{l,j,0} = \sum_{k=0}^j \binom{j}{k} (-1)^k M_{l,0,k} \dots\dots\dots (10)$$

$$M_{l,0,k} = \sum_{j=0}^k \binom{k}{j} (-1)^j M_{l,j,0} \dots\dots\dots (11)$$

通常、変量 x の次数 $l=1$ として、(10)、(11) 式のいずれかを用いる。ここでは、 $k=0$ すなわち $M_{l,j,0}$ を用いることにし、

$$M_j = M_{1,j,0}$$

と表記する。

次に、与えられたデータから標本 PWM を推定する方法を示す。もし j, k が非負の実数であるとき、ある母集団から抽出した $k+j+1$ 個のデータの第 $j+1$ 位順序統計量の l 次積率の期待値 $E[X_{(j+1, k-j, j)}^l]$ は $M_{l,j,k}$ に比例する。すなわち、

$$E[X_{j+1, k+j+1}^l] = \frac{M_{l,j,k}}{B(j+1, k+1)}$$

ここに、 $B(\cdot, \cdot)$ はベータ関数である。いま、 $l=1, k=0$ のとき

$$M_{1,j,0} = B(j+1, 1)E[X_{j+1, j+1}]$$

となる。 N 個のデータから $j+1$ 個 ($N > j+1$) のデータを抽出するとき、その最大値が x_i となる確率は $i-1 C_j / N C_{j+1}$ であるので、 $N \rightarrow \infty$ のある母集団から抽出した $j+1$ 個のデータの最大値の期待値 $E[X_{j+1, j+1}]$ は

$$E[X_{j+1, j+1}] = \lim_{N \rightarrow \infty} \sum_i x_i \frac{i-1 C_j}{N C_{j+1}}$$

となり、標本 PWM \hat{M}_j は

$$\hat{M}_j = \frac{1}{N} \sum_{i=1}^N x_i \frac{(i-1)(i-2)\cdots(i-j)}{(N-1)(N-2)\cdots(N-j)} \dots\dots\dots (12)$$

となる。PWM 法はこの \hat{M}_j を分布モデルの PWM と等しくおいて母数を推定する方法である。

Gumbel 分布の PWM と母数はそれぞれ次のようになる¹²⁾。

$$M_j = \frac{u}{1+j} + \frac{\ln(1+j) + \gamma_E}{\alpha(1+j)} \dots\dots\dots (13)$$

$$\alpha = \frac{\ln 2}{2M_1 - M_0} \dots\dots\dots (14)$$

$$u = M_0 - \frac{\gamma_E}{\alpha} \dots\dots\dots (15)$$

ここに、 j は非負の整数である。(12) 式の標本 PWM をこれらの式に代入することにより母数 α, u の推定値を得る。

Landwehr ら¹³⁾ は、Monte Carlo シミュレーションにより、この PWM 法と最尤法、積率法の比較を行っている。Gumbel 分布に従う乱数を発生させて、それが系列相関をもたない場合ともつ場合について検討し、前者の場合 PWM 法が母数、再現確率統計量の不偏推定値を与え、後者の場合でも他の二つの推定法よりも小さい偏倚をもつ推定量を与えることを示した。最尤法はどちらの場合に対しても母数 α と再現確率統計量の最小分散推定を与え、PWM 法は母数 u に対して最小分散推定を与える。積率法による再現確率統計量の分散は、小標本でリターンピリオドの大きい部分についてのみ PWM 法より小さい。全体として最尤法が良いが、小標本で系列相関があり非超過確率が 0.5 より小さいときには PWM 法が良いことを示した。

(5) 最大エントロピー法

エントロピーは、Boltzman が "a measure of our degree of ignorance as to its true state" と定義して、Shannon が情報理論に応用した概念である。高棹・池淵 (1969)¹⁴⁾ は、水文流出系 (長期流出) の解析にエントロピーの概念を導入し、さらに、降水観測がもたらす情報量¹⁵⁾、降雨・流出変換特性の解析¹⁶⁾にも適用した。

Sonuga (1972)¹⁷⁾ は、最大エントロピー原理から最大エントロピー分布なる確率密度関数を決定する方法を提案した。この手法は、標本から得られる母集団に関する情報は積極的に取り入れ、それ以外ではできるだけ確率変数がランダムに発生する (すなわちエントロピーを最大にする) 確率密度関数を求めるものであり、その後、寒川によって詳細に検討されている^{18, 19)}。

エントロピー $H(x)$ は次式で定義される。

$$H(x) = - \int p(x) \ln p(x) dx$$

ここに、 $p(x)$ は確率密度関数であり、

$$\int p(x) dx = 1$$

である。ある任意関数 $g_r(x)$ が与えられたとき、

$$\int g_r(x) p(x) dx = E[g_r(x)], r=1, 2, \dots, m \dots\dots\dots (16)$$

である。この制約条件のもとで $H(x)$ を最大にする $p(x)$ が最大エントロピー分布である。

いま、 $g_r(x)$ ($-\infty < x < \infty$) が

$$g_1(x) = 1$$

$$g_2(x) = \alpha(x-u)$$

$$g_3(x) = \exp\{-\alpha(x-u)\}$$

のとき、(16) 式を

$$\int_{-\infty}^{\infty} p(x) dx = 1 \dots\dots\dots (17)$$

$$\int_{-\infty}^{\infty} \alpha(x-u) p(x) dx = E[\alpha(x-u)] = \gamma_E \text{ (Euler の定数, } 0.5772\dots) \dots\dots\dots (18)$$

$$\int_{-\infty}^{\infty} \exp\{-\alpha(x-u)\} p(x) dx = E[\exp\{-\alpha(x-u)\}] = 1 \dots\dots\dots (19)$$

とすると、この制約条件のもとで $H(x)$ を最大にする $p(x)$ は

$$p(x) = \alpha \exp[-\alpha(x-u) - \exp\{-\alpha(x-u)\}] \dots\dots\dots (20)$$

すなわち、Gumbel 分布の確率密度関数となる²⁰⁾。このように、ある関数 $g_r(x)$ が与えられたとき、最大エントロピー分布は既存の確率分布に一致することが知られている¹⁶⁾。この $g_r(x)$ に含まれるパラメタ、すなわち母数を推定する方法が最大エントロピー法である。最大エントロピー法を Gumbel 分布に適用したときの母数および再現確率統計量の推定誤差については Phien²¹⁾ によって検討されている。

N 個のデータ x_1, x_2, \dots, x_N が与えられたとき、(18)、(19) 式は、

$$\frac{1}{N} \sum_{i=1}^N s_i = \gamma_E \dots\dots\dots (21)$$

$$\frac{1}{N} \sum_{i=1}^N \exp(-s_i) = 1 \dots\dots\dots (22)$$

となる。ここに、 $s_i = \alpha(x_i - u)$ である。(21)、(22) 式より α に関する方程式

$$\frac{1}{N} \sum_{i=1}^N \exp\{-\alpha(x_i - \bar{x})\} = \exp(\gamma_E) \dots\dots\dots (23)$$

が得られる。ここに、 \bar{x} は x の標本平均である。これを解いて、最大エントロピー法による α の推定値 $\hat{\alpha}$ が得られる。 u の推定値は次式で与えられる。

$$\hat{u} = \bar{x} - \frac{\gamma_E}{\hat{\alpha}}$$

本論文では、(23) 式を Newton 法により解くことにする。

3.3 結果と考察

想定する母数 α, u は Table 1 に示した通りである。

(1) 極値降水量の場合

Table 2 は、標本の大きさ N を 10 から 1000 まで変えたときに、各母数推定法によって得られる 100 年確率降水量 (真値は 192.0 mm) の推定値の平均、真値からの偏倚 (BIAS)、標準偏差、RMSE を示したものである。Table 3 は RMSE について、Table 4 は BIAS について、小さい順に各推定法に順位を付けたものである。RMSE は N の大きさにかかわらず、最尤法、最大エントロピー法、PWM 法の順に良い。BIAS を見ると、 $N \leq 20$ のとき積率法、PWM 法の順に良く、 $30 \leq N$ のとき PWM 法、積率法、最大エントロピー法の順に良い。Table 2 を見ると、PWM 法の BIAS が極めて小さいこと、また、最尤法も順位は 4 位であるが上位のものとは比べてあまり大きな差がないことがわかる。

積率法は RMSE、BIAS のどちらについても最小二乗法よりも良いと言える。積率法は、RMSE が常に 4 位で、BIAS は $N \geq 30$ のとき第 2 位、 $N \leq 20$ のとき第 1 位である。Gumbel 分布の場合、積率法で母数を推定するとき 2 次までの標本積率 (平均と分散) を用いる。標本平均は不偏統計量であり、標本分散も偏倚の大きさは理論的に導かれているので、偏倚を補正した不偏分散がよく用いられる。そのため、積率法は最小二乗法より良い結果を与えると考えられる。

(2) 極値流量の場合

Tables 5, 6, 7 は、極値流量に関する同様の結果を示したものである。RMSE の順位は、最尤法、最大エントロピー法、PWM 法の順に良く、降水量の場合と同じ結果である。BIAS の順位は、 $N \geq 30$ のとき PWM 法、積率法、最大エントロピー法、最尤法で、 $N \leq 20$ のとき積率法、PWM 法、最大エントロピー法である。最小二乗法はいずれについても良くない。極値降水量の場合とほぼ同じ結果である。

4. 一般化極値分布の母数推定法の比較評価

4.1 一般化極値分布

Jenkinson²²⁾ は、最大値分布に従う変量が、その分布の形式にかかわらず、次のような累積分布関数で統合的に表わされることを示し、その後これは一般化極値 (generalized extreme value, GEV) 分布と呼ばれるようになった。

$$F(x) \begin{cases} \exp \left[- \left\{ 1 - \frac{k(x-x_0)}{a} \right\}^{1/k} \right], & k \neq 0 \\ \exp \left[- \exp \left(- \frac{x-x_0}{a} \right) \right], & k = 0 \end{cases}$$

母数は、 k, a, x_0 の三つである。この分布は、1975年に Natural Environment Research Council によって、イギリスの河川の日流量の年最大値の分布に推奨されて以来、イギリスでは水文頻度解析における重要な役割を担うようになった。海水位の極値にも用いられているようである²³⁾。母数 $k=0$ のとき、GEV 分布は Gumbel 分布 (最大値に関する第 I 種極値分布) に一致する。また、 $k \neq 0$ のとき $F(x)$ は次のように変形できる。

$$F(x) = \exp \{ - \exp(-s) \}$$

$$s = - \frac{1}{k} \ln \left[- \left\{ x - \left(x_0 + \frac{a}{k} \right) \right\} / \left(\frac{a}{k} \right) \right]$$

この式から、 $k < 0, k > 0$ のとき、それぞれ対数正規分布 A 型、B 型^{24, 25)} となる。

Table 2. Comparison of the results for the Gumbel distributed rainfall: \hat{x}_{100} , $Bias\{\hat{x}_{100}\}$, $Sd\{\hat{x}_{100}\}$ and RMSE

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	184.31	187.59	189.13	189.91	190.31	190.54	190.85	191.01	191.11	191.08	191.50	191.76	191.90
	-7.70	-4.41	-2.87	-2.09	-1.69	-1.46	-1.16	-0.99	-0.89	-0.92	-0.50	-0.24	-0.11
	32.37	23.08	18.50	16.08	14.31	13.09	12.08	11.29	10.69	9.89	7.10	4.48	3.18
	33.27	23.49	18.72	16.14	14.41	13.17	12.14	11.33	10.73	9.94	7.11	4.49	3.18
MoM	192.53	191.00	190.97	191.08	191.20	191.23	191.30	191.38	191.39	191.37	191.68	191.82	191.92
	0.52	-1.01	-1.04	-0.92	-0.80	-0.77	-0.71	-0.62	-0.61	-0.63	-0.32	-0.18	-0.09
	38.38	27.46	22.33	19.45	17.49	15.98	14.87	13.79	13.18	12.15	8.79	5.60	3.99
	38.39	27.48	22.36	19.47	17.51	16.00	14.88	13.81	13.19	12.17	8.80	5.60	3.99
LS (H)	202.31	197.20	195.66	194.95	194.52	194.11	193.87	193.70	193.51	193.33	192.83	192.37	192.23
	10.31	5.20	3.66	2.95	2.52	2.11	1.86	1.69	1.51	1.33	0.82	0.37	0.22
	42.58	29.57	23.66	20.47	18.32	16.64	15.40	14.26	13.61	12.54	8.97	5.67	4.01
	43.81	30.02	23.94	20.68	18.49	16.77	15.51	14.36	13.69	12.61	9.01	5.68	4.02
LS (W)	228.93	213.57	288.04	205.10	203.23	201.79	200.76	199.98	199.30	198.71	196.11	194.05	193.22
	36.93	21.57	16.03	13.10	11.23	9.78	8.76	7.98	7.29	6.70	4.11	2.05	1.22
	51.93	33.96	26.47	22.55	19.95	17.97	16.51	15.22	14.46	13.29	9.32	5.79	4.07
	63.72	40.23	30.95	26.08	22.90	20.46	18.69	17.18	16.19	14.88	10.19	6.14	4.25
PWM	196.64	193.15	192.49	192.28	192.15	192.04	192.03	192.01	191.98	191.87	191.92	191.93	191.97
	4.64	1.15	0.49	8.27	8.15	0.04	0.03	0.01	-0.03	-0.13	-0.09	-0.87	-0.04
	37.48	26.03	20.82	17.95	16.03	14.63	13.60	12.61	11.98	11.00	7.88	4.99	3.57
	37.77	26.06	20.83	17.96	16.03	14.63	13.60	12.61	11.98	11.00	7.88	4.99	3.57
PME	187.11	188.87	189.87	190.41	190.70	190.84	191.06	191.19	191.25	191.22	191.58	191.80	191.91
	-4.89	-3.13	-2.14	-1.59	-1.31	-1.16	-0.94	-0.82	-0.75	-0.78	-0.42	-0.21	-0.10
	33.74	23.89	19.11	16.50	14.74	13.47	12.49	11.62	11.02	10.15	7.28	4.61	3.28
	34.09	24.09	19.23	16.57	14.80	13.52	12.52	11.64	11.04	10.18	7.29	4.61	3.28

Table 3. Ranking of the methods in terms of the RMSE for the Gumbel distributed rainfall

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	1	1	1	1	1	1	1	1	1	1	1	1	1
MoM	4	4	4	4	4	4	4	4	4	4	4	4	4
LS (H)	6	6	6	6	6	6	6	6	6	6	6	6	6
LS (W)	5	5	5	5	5	5	5	5	5	5	5	5	5
PWM	3	3	3	3	3	3	3	3	3	3	3	3	3
PME	2	2	2	2	2	2	2	2	2	2	2	2	2

Table 4. Ranking of the methods in terms of the bias for the Gumbel distributed rainfall

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	4	4	4	4	4	4	4	4	4	4	4	4	4
MoM	1	1	2	2	2	2	2	2	2	2	2	2	2
LS (H)	6	6	6	6	6	6	6	6	6	6	6	6	6
LS (W)	5	5	5	5	5	5	5	5	5	5	5	5	5
PWM	2	2	1	1	1	1	1	1	1	1	1	1	1
PME	3	3	3	3	3	3	3	3	3	3	3	3	3

4.2 一般化極値分布の母数推定法

以下の三つの母数推定法を比較する。

(1) 最尤法

GEV 分布の最尤法による解法は、1969年の Jenkinson や Prescott and Welden^{23, 26)} の反復解法がある。ここでは、Gumbel 分布と同様、最適化プログラム DMINF 1 を用いて、尤度関数を最大化するような母数の組み合わせを求める。

(2) 積率法

GEV 分布は3個の母数をもつので3次までの積率を用いる。すなわち、次式で与えられる平均 μ 、分散 σ^2 、歪係数 β と、標本から得られるそれらを等しく置いて a 、 k 、 x_0 を求める。

$$\mu = x_0 + a \{1 - \Gamma(1+k)\} / k$$

$$\sigma^2 = a^2 \{ \Gamma(1+2k) - \Gamma^2(1+k) \} / k^2$$

$$\beta = \frac{\Gamma(1+3k) - 3\Gamma(1+2k)\Gamma(1+k) + 2\Gamma^3(1+k)}{\{ \Gamma(1+2k) - \Gamma^2(1+k) \}^{3/2}}$$

ここに、 $\Gamma(\cdot)$ はガンマ関数である。

ここでは、標本歪係数を用いる場合と不偏歪係数を用いる場合の二通りを考えた。

(3) PWM 法

Hosking ら²⁷⁾ は、PWM 法を GEV 分布に適用し、この分布に対する PWM 法の有用性を示した。GEV 分布の PWM と母数との関係は次のようである。

Table 5. Comparison of the results for the Gumbel distributed flood : \hat{x}_{100} , $Bias\{\hat{x}_{100}\}$, $Sd\{\hat{x}_{100}\}$ and RMSE

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	29214.73	29708.06	29969.35	30088.23	30148.30	30180.25	30226.07	30249.83	30264.17	30258.95	30322.39	30362.53	30383.89
	-1185.87	-692.54	-431.25	-312.37	-252.30	-220.35	-174.53	-150.77	-136.43	-141.65	-78.21	-38.07	-17.51
	5218.67	3702.75	2969.73	2569.32	2294.50	2098.44	1937.56	1808.86	1712.78	1584.92	1135.78	717.76	509.12
MoM	5351.72	3766.95	3000.88	2588.24	2308.41	2109.98	1945.41	1815.14	1718.21	1591.24	1138.47	718.77	509.42
	30484.11	30239.51	30234.71	30253.27	30272.79	30276.97	30287.76	30308.77	30303.05	30299.29	30348.79	30371.27	30386.64
	83.51	-161.09	-165.89	-147.33	-127.81	-123.63	-112.84	-99.83	-97.55	-101.31	-51.81	-29.33	-13.96
LS (H)	6141.09	4394.16	3573.39	3112.04	2798.21	2557.19	2378.68	2207.23	2108.09	1943.69	1406.60	896.11	638.28
	6141.65	4397.12	3577.24	3115.53	2801.13	2560.18	2381.35	2209.49	2110.34	1946.33	1407.56	896.59	638.43
	32049.65	31232.52	30986.40	30872.22	30803.25	30737.52	30698.70	30671.72	30642.27	30613.36	30532.23	30459.40	30436.29
LS (W)	1649.05	831.92	585.80	471.62	402.65	336.92	298.10	271.12	241.67	212.76	131.63	58.80	35.69
	6813.05	4730.68	3785.54	3275.73	2930.66	2662.53	2463.83	2282.32	2177.12	2006.95	1435.53	986.38	642.44
	7009.78	4803.28	3830.60	3309.51	2958.19	2683.76	2481.80	2298.36	2190.49	2018.20	1441.55	908.29	643.43
PME	36309.22	33851.65	32966.12	32496.77	32197.43	31966.10	31801.74	31677.48	31567.80	31473.11	31057.92	30728.18	30596.08
	5908.62	3451.05	2565.52	2096.17	1796.83	1565.49	1401.14	1276.80	1167.20	1072.51	657.32	327.58	195.48
	8308.62	5433.18	4235.50	3608.07	3192.48	2875.42	2641.91	2434.60	2312.80	2125.74	1491.61	926.14	651.25
PWM	10195.34	6436.55	4951.90	4172.78	3663.41	3273.96	2998.47	2749.08	2590.64	2380.97	1630.02	982.37	679.95
	31142.92	30584.40	30479.12	30444.56	30424.53	30406.45	30405.43	30402.46	30396.64	38379.90	30386.51	30388.88	30394.74
	742.32	183.80	78.52	43.96	23.93	5.85	4.83	1.86	-3.96	-20.70	-14.09	-11.72	-5.86
PME	5997.05	4165.08	3331.38	2872.86	2564.60	2340.45	2175.70	2017.31	1916.17	1760.62	1261.01	798.80	571.08
	6042.81	4169.14	3332.30	2873.20	2564.71	2340.46	2175.71	2017.31	1916.18	1760.74	1261.08	798.88	571.11
	29617.40	29899.83	30058.63	30145.86	30191.73	30215.27	30250.47	30269.90	30280.80	30275.18	30332.99	30367.40	30385.07
PME	-783.20	-500.77	-341.97	-254.74	-208.87	-185.33	-150.13	-130.70	-119.80	-125.42	-67.61	-33.20	-15.53
	5398.25	3821.52	3058.40	2639.80	2358.95	2155.40	1997.85	1858.50	1763.08	1623.91	1164.96	737.68	525.24
	5454.77	3854.19	3077.45	2652.06	2368.18	2163.35	2003.49	1863.09	1767.15	1628.75	1166.93	738.43	525.47

Table 6. Ranking of the methods in terms of the RMSE for the Gumbel distributed flood

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	1	1	1	1	1	1	1	1	1	1	1	1	1
MoM	4	4	4	4	4	4	4	4	4	4	4	4	4
LS (H)	6	6	6	6	6	6	6	6	6	6	6	6	6
LS (W)	5	5	5	5	5	5	5	5	5	5	5	5	5
PWM	3	3	3	3	3	3	3	3	3	3	3	3	3
PME	2	2	2	2	2	2	2	2	2	2	2	2	2

Table 7. Ranking of the methods in terms of the RMSE for the Gumbel distributed rainfall

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	4	4	4	4	4	4	4	4	4	4	4	4	4
MoM	1	1	2	2	2	2	2	2	2	2	2	2	2
LS (H)	6	6	6	6	6	6	6	6	6	6	6	6	6
LS (W)	5	5	5	5	5	5	5	5	5	5	5	5	5
PWM	2	2	1	1	1	1	1	1	1	1	1	1	1
PME	3	3	3	3	3	3	3	3	3	3	3	3	3

$$M_j = \frac{x_0 + a \{1 - (j+1)^{-k} \Gamma(1+k)\} / k}{j+1}$$

$$M_0 = x_0 + \frac{a \{1 - \Gamma(1+k)\}}{k}$$

$$2M_1 - M_0 = a \Gamma(1+k) \frac{1 - 2^{-k}}{k}$$

$$\frac{3M_2 - M_0}{2M_1 - M_0} = \frac{1 - 3^{-k}}{1 - 2^{-k}}$$

ここに、 j は非負の整数であり、 $k > -1$ である。

4.3 結果と考察

Tables 8, 9, 10 は、100年確率降水量（真値は 191.82 mm）についての結果を示したものである。 $N=1000$ のとき RMSE, BIAS のどちらについても最尤法が良いが、 $N \leq 500$ のとき PWM 法が常に第 1 位である。この理由の一つに、最尤法の overfitting があると思われる。一般化極値分布は 3 個の母数を持ち多様な形状の標本に適合し易い柔軟な構造を有しており、これに最尤法を用いると、標本ごとによく適合することになるので、かえって確率水文量の推定誤差が大きくなってしまふのである。標本サイズ N がかなり大きいとこの問題はなくなるが、 N があまり大きくない場合、こうした overfitting には注意する必要がある。

PWM 法の RMSE は、 $N \leq 100$ のとき最尤法の 1/2 以下である。また、BIAS は $N=10$ のときでも約

Table 8. Comparison of the results for the GEV distributed rainfall: \hat{x}_{100} , Bias (\hat{x}_{100}), $Sd(\hat{x}_{100})$ and RMSE

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	238.88 47.06 151.92 159.04	230.54 38.72 121.55 127.57	213.43 21.61 98.39 100.73	212.90 21.08 85.16 87.73	209.04 17.22 84.18 85.93	199.96 8.14 57.46 58.03	201.17 9.35 59.84 59.77	199.14 7.32 51.91 52.43	199.35 7.53 52.25 52.79	199.26 7.44 52.01 52.54	192.45 0.63 19.68 19.70	192.03 0.21 12.22 12.22	191.81 -0.01 7.10 7.10
PWM	208.84 17.02 74.63 76.55	197.32 5.50 53.07 53.36	194.71 2.89 43.08 43.18	193.73 1.91 37.59 37.64	193.36 1.54 33.81 33.85	193.20 1.38 30.29 30.32	192.68 0.86 28.22 28.24	192.60 0.78 25.97 25.98	192.32 0.50 24.90 24.90	192.21 0.39 23.03 23.03	192.04 0.23 16.39 16.39	191.82 -0.00 10.26 10.26	191.84 0.02 7.30 7.30
MoM (sample skew)	93.17 -98.65 84.80 130.08	107.35 -84.47 82.97 118.48	118.10 -73.71 81.04 109.55	125.19 -66.63 79.52 103.74	132.18 -59.64 77.45 97.75	136.91 -54.91 75.43 93.30	141.88 -49.94 73.26 88.66	146.38 -45.44 70.81 84.14	149.17 -42.65 69.23 81.31	151.87 -39.95 67.39 78.34	170.68 -21.22 51.86 56.04	186.07 -5.75 25.87 26.50	191.11 -1.71 11.33 11.46
MoM (unbiased skew)	108.75 -83.07 88.84 121.63	114.93 -76.89 83.70 113.66	123.21 -68.61 80.88 106.07	129.92 -61.98 78.89 100.28	135.73 -56.09 76.76 95.07	140.16 -51.66 74.55 90.70	144.54 -47.28 72.36 86.43	148.67 -43.14 69.87 82.12	151.49 -40.33 68.16 79.19	154.08 -37.74 66.24 76.23	171.40 -20.42 51.09 55.02	186.25 -5.57 25.52 26.12	190.14 -1.68 11.33 11.45

差が、標本サイズ N に応じてどのように変化するか（それは、Table 2, 5, 8 から読みとることができる）、という点に関して検討したものである。実際問題としては、データが与えられ、その母分布を推定するという場面がほとんどである。この場合の分布形の選択には、jackknife や bootstrap といったリサンプリング手法を用いて、適合度と確率水文学の変動性を評価規準とする筆者らの方法²⁸⁻³⁰⁾が有用である。

17 mm で他の推定法よりかなり小さく、 $N=70$ では 1 mm 未満となる。PWM 法は GEV 分布に対して非常に優れた方法であると言える。極値流量についても同様の結果を得た。

5. おわりに

極値データの頻度解析において重要な極値分布の種々の母数推定法を比較評価した。

Gumbel 分布については最尤法を用いるのがよい。確率水文学の推定値の偏倚が最も小さいのは PWM 法であり、続いて積率法、最大エントロピー法、最尤法であるが、それらの差は小さい。ここでは、議論を簡明にするため 100 年確率水文学の結果のみを示したが、50 ~ 200 年確率水文学についても全く同様の結果を得ている。また、母数推定値に関する検討も同時に行なったが、複数の母数の推定精度の良否は、結局、確率水文学の推定精度に集約されるので、ここでは示さなかった。個々の母数の推定値を問題とするときは、3.2 (4) で述べた Landqehr¹³⁾の結果（最尤法、積率法、PWM 法を比べている）が参考となろう。本論文の結果は、Gumbel 分布に関するこれまでの多数の評価結果⁴⁻⁸⁾を追認したことになる。

一般化極値分布については PWM 法を用いるのがよい。わが国では、角屋が対数極値分布としてこの分布の解法を早い時期から与えていた^{24, 25)}が、最尤法や PWM 法が用いられるようになったのはごく最近であり、今後さらに検討する余地がある。筆者らの経験では、標本によく適合する分布であり、適合しすぎて確率水文学の変動が大きくなるという側面も有している。

本論文では、母分布が判っている場合にどの推定法を用いればよいか、また、各推定法によって得られる 100 年確率水文学の推定誤

Table 9. Ranking of the methods in terms of the RMSE for the GEV distributed rainfall

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	4	4	2	2	2	2	2	2	2	2	2	2	1
PWM	1	1	1	1	1	1	1	1	1	1	1	1	2
MoM (sample skew)	3	3	4	4	4	4	4	4	4	4	4	4	4
MoM (unbiased skew)	2	2	3	3	3	3	3	3	3	3	3	3	3

Table 10. Ranking of the methods in terms of the bias for the GEV distributed rainfall

N	10	20	30	40	50	60	70	80	90	100	200	500	1000
MLE	2	2	2	2	2	2	2	2	2	2	2	2	1
PWM	1	1	1	1	1	1	1	1	1	1	1	1	2
MoM (sample skew)	4	4	4	4	4	4	4	4	4	4	4	4	4
MoM (unbiased skew)	3	3	3	3	3	3	3	3	3	3	3	3	3

参考文献

- 1) 神田 徹・藤田睦博：水文学——確率論的手法とその応用——，技報堂出版，1982，p.49.
- 2) Kite, G.W.: Frequency and Risk Analysis in Hydrology, Water Resources Publications, Fort Collins, Colorado, U.S.A., 1977, p.9.
- 3) Gumbel, E.J.: Statistics of Extremes, Columbia University Press, 1958, pp.113-155.
- 4) Lowery, M.D. and J.E. Nash: A Comparison of Methods of Fitting the Double Exponential Distribution, Journal of Hydrology, Vol.10, 1970, pp.259-275.
- 5) Lettenmaier, D.P. and S.J. Burges: Gumbel's Extreme Value I Distribution——A New Look, Proc. ASCE, Vol.108, No. HY 4, 1982, pp.502-514.
- 6) Raynal, J.A. and J.D. Salas: Estimation Procedures for the Type-1 Extreme Value Distribution, Journal of Hydrology, Vol.87, 1986, pp.315-336.
- 7) Phien, H.N.: A Review of Methods of Parameter Estimation for the Extreme Value Type-1 Distribution, Journal of Hydrology, Vol.86, 1987, pp.251-268.
- 8) Jain, D. and V.P. Singh: Estimating Parameters of EV I Distribution for Flood Frequency Analysis, Water Resources Bulletin, Vol.23, No.1, 1987, pp.59-71.
- 9) 前出3) の pp.231-232.
- 10) 富士通(株): FACOM FORTRAN SSL II 使用手引書, 1980, pp.403-406.
- 11) 高棹琢馬・宝 馨・清水 章: 琵琶湖流域水文データの基礎的分析, 京都大学防災研究所年報,

- 第 29 号 B-2, 1986, pp. 157-171.
- 12) Greenwood, J. A., J. M. Landwehr, N. C. Matalas and J. R. Wallis: Probability Weighted Moments—Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form, *Water Resources Research*, Vol. 15, No. 5, 1979, pp. 1049-1054.
 - 13) Landwehr, J. M., N. C. Matalas and J. R. Wallis: Probability Weighted Moments Compared With Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles, *Water Resources Research*, Vol. 15, No. 5, 1979, pp. 1055-1064.
 - 14) 高棹琢馬・池淵周一：長期間流出機構の情報理論的研究，京都大学防災研究所年報，第 12 号 B, 1969, pp. 273-293.
 - 15) 高棹琢馬・池淵周一：降水観測がもたらす情報量とその観測網配置計画への応用，土木学会論文報告集，第 234 号，1975, pp. 83-95.
 - 16) 高棹琢馬・池淵周一：エントロピー的に見た降雨・流出変換特性とそのモデル化，京都大学防災研究所年報，第 23 号 B-2, 1980, pp. 193-209.
 - 17) Sonuga, J. O.: Principle of Maximum Entropy in Hydrologic Frequency Analysis, *Journal of Hydrology*, Vol. 17, 1972, pp. 177-191.
 - 18) 寒川典昭・荒木正夫・佐藤健次：多変数最大エントロピー分布とその基礎特性に関する研究，土木学会論文集，第 375 号/II-6, 1986, pp. 89-98.
 - 19) 寒川典昭：最大エントロピー法を用いた水文量の頻度分析手法に関する基礎的研究，京都大学学位論文，1988, 176 pp.
 - 20) Jowitt, P. W.: The Extreme-Value Type-1 Distribution and the Principle of Maximum Entropy, *Journal of Hydrology*, Vol. 42, 1979, pp. 23-38.
 - 21) Phien, H. N.: Sampling Properties of the Maximum Entropy Estimators for the Extreme-Value Type-1 Distribution, *Journal of Hydrology*, Vol. 86, 1986, pp. 391-398.
 - 22) Jenkinson, A. F.: The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements, *Quarterly Journal of the Royal Meteorological Society*, 81, 1955, pp. 158-171.
 - 23) Prescott, P. and A. T. Walden: Maximum Likelihood Estimation of the Parameters of the Three-Parameter Generalized Extreme-Value Distribution from Censored Samples, *J. Statist. Comput. Simul.*, Vol. 16, 1983, pp. 241-250.
 - 24) 角屋 睦：極値分布とその一解法，*農業土木研究*，第 23 卷，第 6 号，1956, pp. 350-357.
 - 25) 角屋 睦：水文統計論，*水工学シリーズ*，64-2，土木学会水理委員会，1964, 59 pp.
 - 26) Prescott, P. and A. T. Walden: Maximum Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution, *Biometrika*, Vol. 63, No. 3, 1980, pp. 723-724.
 - 27) Hosking, J. R. M., J. R. Wallis and E. F. Wood: Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments, *Technometrics*, Vol. 27, No. 3, 1985, pp. 251-261.
 - 28) 宝 馨・高棹琢馬：水文頻度解析における確率分布モデルの評価規準，土木学会論文集，第 393 号/II-9, 1988, pp. 151-160.
 - 29) 合田良実：“水文頻度解析における確率分布モデルの評価規準”への討議，土木学会論文集，第 405 号/II-11, 1989, pp. 265-267.
 - 30) 宝 馨・高棹琢馬：“水文頻度解析における確率分布モデルの評価規準”への討議に対する回答，土木学会論文集，第 405 号/II-11, 1989, pp. 267-272.