

波高の極値統計解析における母数推定法の比較

山口正隆*・畑田佳男**・中村雄二***・大木泰憲****

1. 緒 言

波高の極値統計解析で用いられる各種の理論確率分布の母数推定法として、一般的には、積率法 (MOM)、最尤法 (MLM)、最小 2 乗法 (LSM) のほか、最近では PWM 法の 4 種類が使用されているが、それらの特性についての詳しい検討や比較は行われていないし、分散推定法もあまり明らかでないものが多い。

そこで、本研究では、母集団分布を 3 母数 Weibull 分布、Gumbel 分布、GEV (一般化極値) 分布、3 母数対数正規分布に指定した場合のモンテカルロシミュレーションを実施し、各母数推定法の特性を確率波高の bias (偏倚) および標準偏差に及ぼす標本数の影響という観点から検討する。また、シミュレーション結果に基づいて、分散推定法である jackknife 法および情報行列を用いる方法 (最尤法の場合のみ) の適用性を調べる。

2. モンテカルロシミュレーション

2.1 確率分布と母数・分散推定法

(1) 確率分布

3 母数 Weibull 分布、Gumbel 分布、GEV 分布および 3 母数対数正規分布 (ひずみ係数正) の未超過確率 $F(x)$ はそれぞれ次式で表される。

$$\left. \begin{aligned} F(x) &= 1 - \exp[-\{(x-B)/A\}^k]; B < x < \infty \\ F(x) &= \exp[-\exp\{-(x-B)/A\}]; -\infty < x < \infty \\ F(x) &= \exp[-\exp\{1-(x-B)/kA\}^k] \\ &\quad x < B+kA; k > 0, x > B+kA; k < 0 \\ F(x) &= (1/\sqrt{\pi}) \int_{-\infty}^x \exp(-y^2) dy \\ &\quad y = k \cdot \log\{(x-B)/A\}; B < x < \infty \end{aligned} \right\} \dots\dots\dots (1)$$

ここに、 x : 確率変数、 A : 尺度母数、 B : 位置母数、 k : 形状母数、である。これらは $F(x)$ を独立変数とした表示式に変換されるので、0~1 の一様乱数を与えることにより、任意個数の波高標本をシミュレートできる。

(2) 母数推定法および適合度基準

4 種類の母数推定法のうち、積率法、最尤法および最小 2 乗法はよく知られている。とくに、合田 (1990) は線形最小 2 乗法に基づく系統的で適用範囲の広い極値統計解析システムを提案している。一方、PWM 法は比較的最近 Greenwood et al. (1979) により提案され、Hosking et al. (1985)、竹内ら (1988) により適用範囲が拡張された方法である。PWM 法は積率法で使用する確率変数の原点まわりの積率のかわりに、確率密度関数の未超過確率および超過確率の累乗で重み付けした積率 (PWM) と標本から求められる PWM を等値して母数を推定する。

つぎに、適合度基準として、高棹ら (1986) による SLSC (基準化最小 2 乗平均誤差適合基準) を用いる。SLSC は次式で定義される。

$$SLSC = \left\{ \sum_{i=1}^N (s_i - r_i)^2 / N \right\}^{1/2} / |S_{0.99} - S_{0.01}| \dots\dots (2)$$

ここに、 N : 標本数、 s_i : 順序統計量に対する基準化変量、 r_i : 適当な確率に対応する基準化変量、 $S_{0.99}$ 、 $S_{0.01}$: それぞれ非超過確率 0.99、0.01 に対応する確率分布の基準化変量である。 s_i の評価はプロットングポジション公式を必要とするので、特有の公式が与えられる最小 2 乗法を除いて、Weibull 公式 ($F(x) = 1 - i/(N+1)$) を用いる。

(3) 分散推定法

一般的な分散 (標準偏差) 推定法はあまり明らかでないが、近年、jackknife 法 (Miller, 1974) などの resampling 手法が提案されている。jackknife 法は、 N 個の標本から得られる統計量 $\hat{\theta}$ とこの標本を除いた N 組の標本 (標本数 $N-1$ 個) を用いて得られる統計量 θ_{*i} ($i=1 \sim N$) の演算によって、統計量の bias の補正や分散の推定を行う方法であり、bias を補正した jackknife 推定値 θ_j および分散推定値 σ_j^2 は次式で与えられる。

$$\left. \begin{aligned} \theta_j &= N\hat{\theta} - (N-1)\bar{\theta}_{*}, \quad \bar{\theta}_{*} = \sum_{i=1}^N \theta_{*i} / N, \\ \sigma_j^2 &= (N-1) \sum_{i=1}^N (\theta_{*i} - \bar{\theta}_{*})^2 / N \end{aligned} \right\} \dots\dots (3)$$

ただし、 θ_j の bias 補正量は $O(1/N)$ である。

一方、最尤法による場合、対数尤度関数 L の 2 階微分

* 正 会 員 工 博 愛媛大学教授 工学部土木海洋工学科
 ** 正 会 員 工 修 愛媛大学助手 工学部土木海洋工学科
 *** 正 会 員 工 修 株式会社ハザマ
 **** 学生会員 愛媛大学大学院 工学研究科

値を要素とする行列の逆行列が母数 $(\theta_1, \dots, \theta_r)$ の分散・共分散行列に相当する性質を利用して、確率統計量（確率波高）の分散を推定できる（Phien et al., 1989）。2階微分値そのものを要素とする行列は観測情報行列、2階微分値の期待値を要素とする行列は Fisher 情報行列と呼ばれる。

2.2 シミュレーションの方法

シミュレーションの手順はつぎのようである。

① 理論確率分布と母数を指定し、 n 年 (50 年, 100 年, 200 年, 500 年, 1000 年) 確率波高の真値を求める。

② 理論確率分布に一樣乱数を与えて、任意指定個数 N の標本を抽出し、4 つの母数推定法（最小 2 乗法は Gumbel 分布および Weibull 分布の場合のみ）により母数、適合度指標および n 年確率波高を求める。また、積率法および PWM 法（4 分布）と最尤法（Gumbel 分布）の場合には、jackknife 法によって bias を補正した確率波高と分散 ($\sigma_{JM}^2, \sigma_{JP}^2, \sigma_{Jr}^2$) を、最尤法（4 分布）の場合には、観測および Fisher 情報行列の逆行列から確率波高の分散 ($\sigma_{OM}^2, \sigma_{FM}^2$) を計算する。

③ 同じ手順を M 回 ($M=5000$) 繰り返し、適合度指標の平均値、母数および各確率波高の平均値とそれらの bias、2 乗平均平方根誤差 (RMSE)、各方法に基づく確率波高の分散の平均値を算出する。

④ $N=10, 20, 30, 40, 50, 70, 100, 200, 500, 1000$ の 10 通りの場合に、②、③の手順を繰り返す。

シミュレーションは表-1 に示すように、4 種類の確率分布にそれぞれ母数を 1 種類与えた場合について行った。ここに、 C_s は母ひずみ係数である。また、シミュレーション資料は unsensored data に相当するので、標本数を統計年数とみなす。

表-1 シミュレーションにおける入力母数

prob. distri.	k	A	B	C_s
Lognormal	3.4	8.0	2.9	0.6400
Gumbel		1/0.72	4.5	1.1395
GEV	-5.0	1.0	5.0	3.5351
Weibull	1.8	4.0	1.0	0.7787

3. シミュレーション結果の考察

図-1 は 3 母数 Weibull 分布の場合の、各母数推定法に基づく SLSC、100 年確率波高の真値で無次元化した、bias $\overline{\Delta H}$ 、RMSE および jackknife 法による bias 補正後の 100 年確率波高の bias $\overline{\Delta H}_J$ と、RMSE で無次元化した jackknife 法に基づく標準偏差 (σ_{JM}, σ_{JP}) および 2 種類の情報行列に基づく標準偏差 (σ_{OM}, σ_{FM}) に対する % 表示値（以下の図では、—を省略）と標本数 N の関係であり、これからつぎのことがいえる。

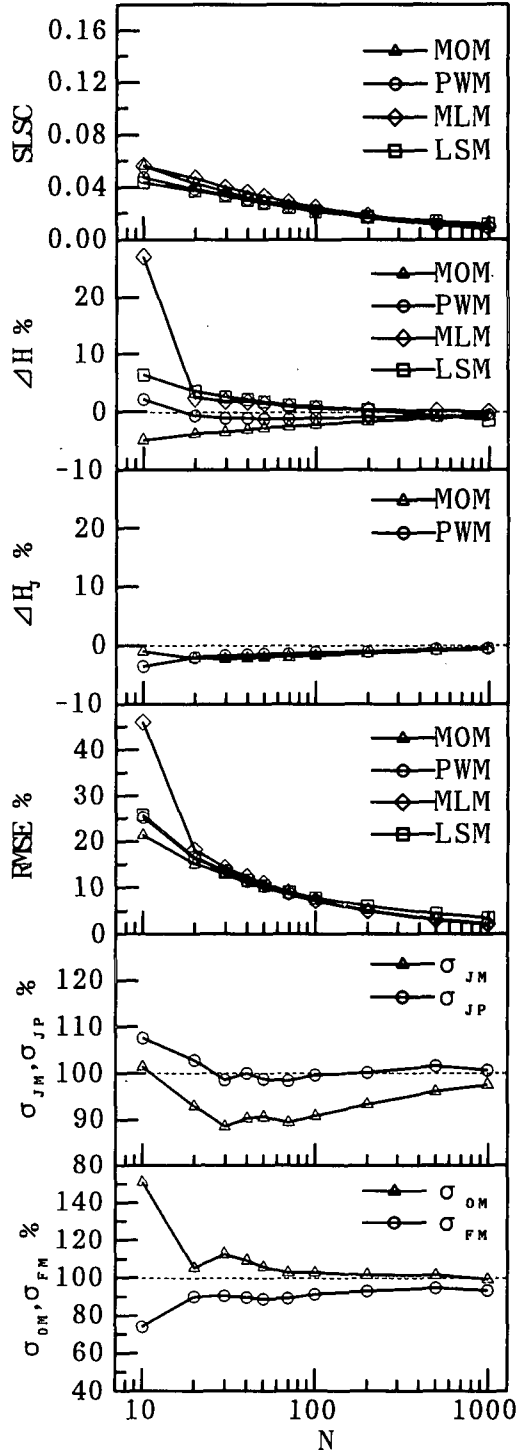


図-1 3 母数 Weibull 分布に対する各種母数推定法の精度

① SLSC によると、 $N \leq 100$ の場合最小 2 乗法、 $N > 100$ では積率法が優れている。絶対基準値である SLSC が 0.02 以下になるのは、いずれの方法でも $N > 200$ の場

合である。

② bias が最も小さいのは $N \leq 50$ の場合 PWM 法, $N \geq 70$ の場合最尤法である。積率法および PWM 法は負の bias を, 最尤法および $N \leq 200$ の場合の最小 2 乗法は正の bias を与える。

③ jackknife 法に基づく bias 補正後の確率波高に対する bias ΔH_j は, 積率法の場合, 標本数によらず非常に小さいのに対して, PWM 法の場合, bias 補正がききすぎて bias が負値をとり, その絶対値も補正前より増加する。

④ $N=10$ の場合の最尤法を除いて, 4 つの方法による RMSE は同程度の値である。

⑤ SLSC, bias (最小 2 乗法を除く) および RMSE のいずれも標本数の増加とともに 0 に漸近する。

⑥ jackknife 法によれば, $N \geq 20$ の場合, 積率法および PWM 法に基づく確率波高の標準偏差を, それぞれ 12 % 以内 (過小評価) および ± 3 % 以内の誤差で推定でき, とくに, PWM 法の場合に精度が高い。また, $N=10$ の場合でも 1.4 % (積率法) および 7.7 % (PWM 法) の過大評価を生じるにすぎない。

⑦ $N \geq 20$ の場合, 観測情報行列に基づく標準偏差推定法は最大 13 % の過大評価値を, Fisher 情報行列に基づく標準偏差推定法は最大 11 % の過小評価値を与えるにすぎず, しかも標本数の増加とともに誤差が減少するので, これらは最尤法に基づく標準偏差推定法として有効である。

図-2 は Gumbel 分布の結果を示す。検討結果はつぎのようである。

① 最小 2 乗法が最も小さい SLSC を与え, $N \leq 200$ では PWM 法が, $N \geq 500$ では積率法がこれにつぐ。また, SLSC が 0.02 以下になるのは $N > 200$ の場合である。

② 4 種類の母数推定法ともほぼ負の bias を生じる。しかし, PWM 法による bias は ± 0.2 % 以下であり, きわめて小さい。また, PWM 法および最小 2 乗法による bias は N に関して極小値をもつ。

③ jackknife 法によって, PWM 法, 積率法および最尤法に伴う bias がほぼ完全に補正される。

④ PWM 法は最尤法よりやや大きい RMSE を与えるけれども, その差はあまり大きくない。また, 最小 2 乗法による RMSE は他の方法に比べてやや大きい。

⑤ jackknife 法によれば, 確率波高の標準偏差を 2 % 以内の誤差 ($N \leq 20$ 以下でも 5 % 以内) で推定できる。

⑥ $N \geq 20$ の場合, 2 種類の情報行列に基づく標準偏差は RMSE とほぼ一致することから, 最尤法の場合の標準偏差推定法として十分有用である。

図-3 は GEV 分布に対する結果であり, つぎのことが

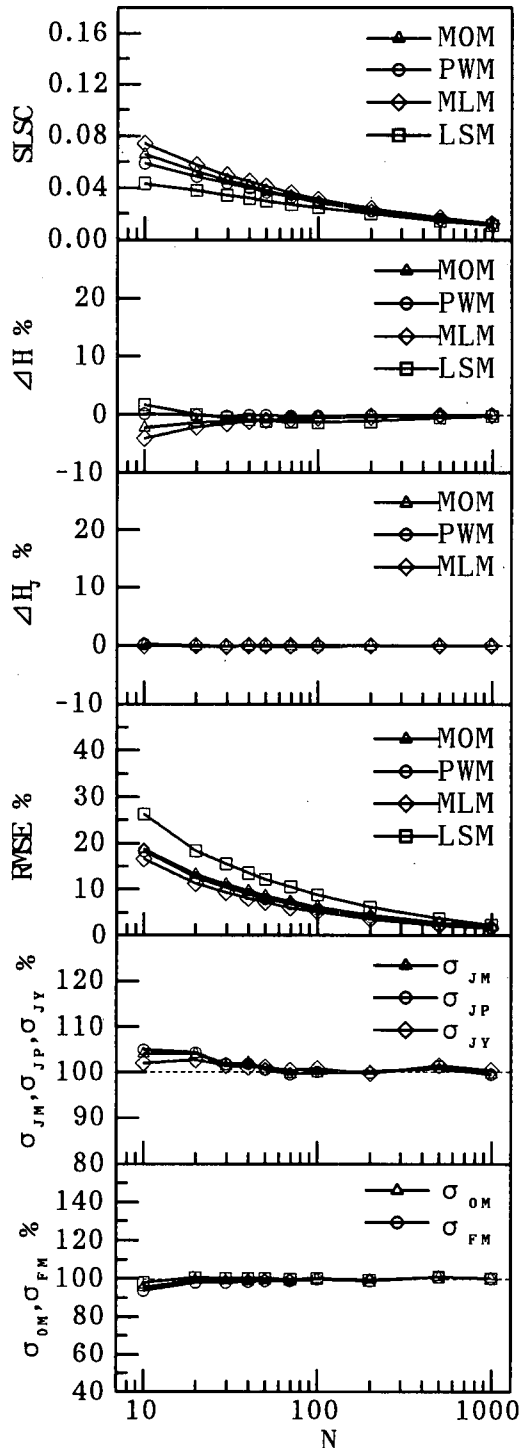


図-2 Gumbel 分布に対する各種母数推定法の精度

わかる。

① PWM 法が最も小さい SLSC を与え, 最尤法がこれにつぐ。しかし, $N \leq 1000$ では, いずれの方法でも

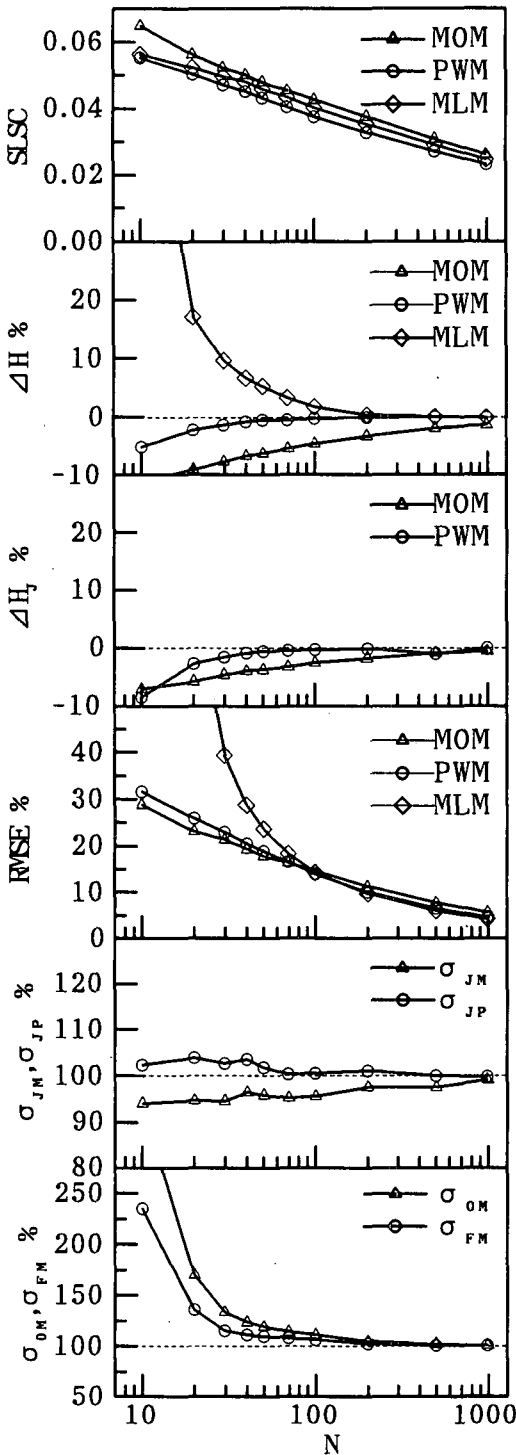


図-3 GEV 分布に対する各種母数推定法の精度

SLSC は 0.02 以下にならない。

② PWM 法はいずれの標本数についても最も小さい負の bias を与える。また、積率法は負の bias (過小評価)。

最尤法は正の bias (過大評価) を与えるが、標本数が少ない場合、bias はかなり大きい。

③ jackknife 法は積率法に基づく確率波高の bias を 1/2 程度にするが、PWM 法に基づく bias に対しては必ずしも適切に働かない。

④ RMSE については、 $N \leq 70$ で積率法が、 $N \geq 200$ で最尤法が最も小さい。PWM 法は 3 つの方法のうち、必ずしも最も小さい RMSE を与えないけれども、これにかなり近い値を生ずる。

⑤ 最尤法は $N \geq 200$ という大標本の場合に、精度の高い推定値を与えるけれども、小標本の場合、bias および RMSE が大きいことから、適切な方法といえない。

⑥ jackknife 法によれば、PWM 法の場合 4 % 以内の誤差 (過大評価)、積率法の場合 6 % 以内の誤差 (過小評価) で確率波高の標準偏差を評価できる。

⑦ 観測および Fisher 情報行列に基づく方法は標準偏差を過大評価するが、標本数の増加とともに、とくに後者の精度が向上し、 $N \geq 50$ では誤差は 10 % 以下になる。

図-4 は 3 母数対数正規分布に対する結果である。これからつぎのことがいえる。

① 同じ標本数に対して最も小さい SLSC を与えるのは PWM 法であるが、 $N \geq 500$ ではこれが積率法に代わる。しかし、 $N \geq 50$ では、これらの値の差は小さい。また、SLSC は $N > 100$ の場合に、0.02 以下をとる。

② $N = 10$ の場合、積率法が最も小さい bias を与えるが、 $N = 20 \sim 100$ では最尤法が、 $N \geq 200$ では PWM 法がこれに代わる。積率法は負の bias を示し、PWM 法および最尤法による bias は、標本数の増加とともに、正值から負値に変化する。しかし、 $N \geq 100$ では、いずれの母数推定法による bias も小さい。

③ jackknife 法は、積率法に伴う bias を 1 % 以内に抑えるが、PWM 法に伴う bias 補正に対して有効に働かない。

④ RMSE については、 $N \leq 200$ では積率法が、 $N > 200$ では最尤法が最も小さい値を与える。

⑤ jackknife 法によって、積率法に基づく確率波高の標準偏差を $N \geq 20$ の場合、5 % 以内の誤差で推定できるが、PWM 法の場合には $N = 10$ で 27 %、 $N = 50$ でも 11 % の誤差を生じる。

⑥ 観測および Fisher 情報行列に基づく方法は標準偏差を過大評価するが、標本数の増加とともに、精度が急速に向上する。とくに、Fisher 情報行列に基づく方法の誤差は、 $N \geq 40$ の場合に 4 % 以下になる。

4. 結 語

本研究の結果はつぎのように要約される。

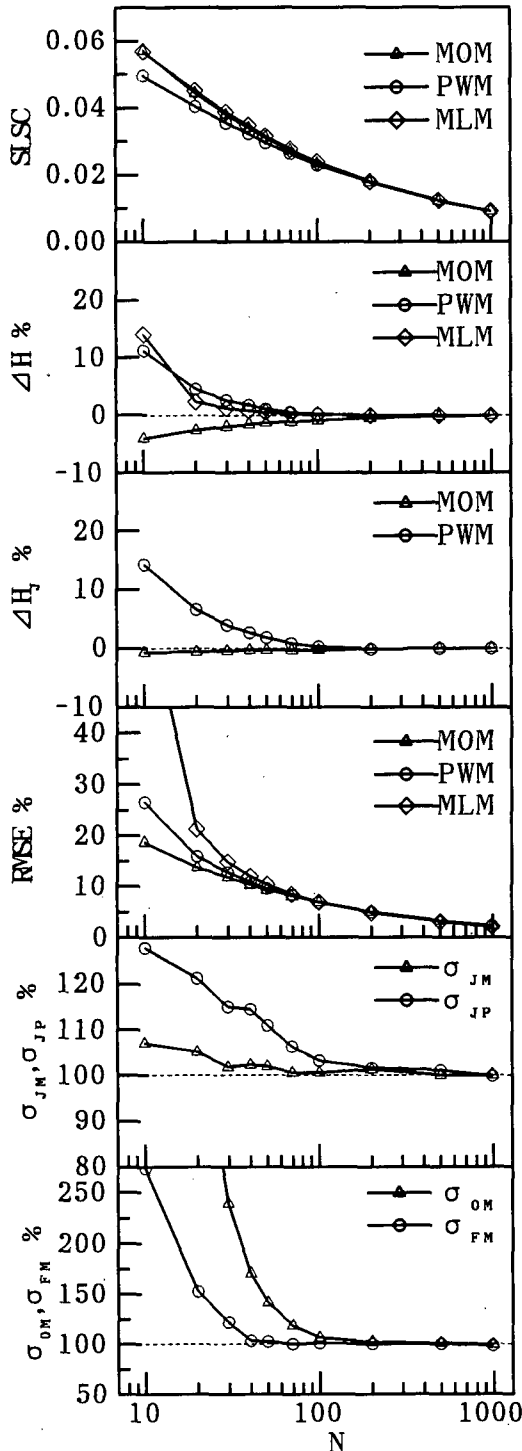


図-4 3母数対数正規分布に対する各種母数推定法の精度

① 3母数 Weibull 分布および GEV 分布の場合、PWM 法により母数および確率波高を、jackknife 法により分散を求めれば、最適な結果が得られる。

② jackknife 法を併用すれば、Gumbel 分布の場合、いずれの母数推定法によっても、また、3母数対数正規分布の場合、積率法によって最適な結果が得られる。

③ 最尤法は、小標本の場合、GEV 分布に対して適切な方法といえないのに対し、3母数対数正規分布に対しては $N \geq 20$ で良好な推定値を与える。

④ 最小2乗法は Weibull 分布および Gumbel 分布に対して平均的に良好な推定値を与えるが、4つの母数推定法のうちで必ずしも最良の方法とはいえない。

⑤ jackknife 法によって、確率波高の標準偏差 (PWM 法を用いた 3母数対数正規分布を除く) を高い精度で推定することが可能である。

⑥ $N \geq 30$ では、Fisher 情報マトリックスに基づく確率波高の標準偏差はいずれの確率分布に対しても 12~20% の誤差内で RMSE と符合するから、最尤法の場合、この方法を用いて標準偏差を推定できる。

なお、一般的な結論を得るためには、より広範な条件のもとでのシミュレーション結果を考察する必要がある。

参考文献

- 合田良賢(1990): 港湾構造物の耐波設計—波浪工学への序説—, 鹿島出版会, 333 p.
- 高棹琢馬・宝 馨・清水 章 (1986): 琵琶湖流域水文データの基礎的分析, 京大防災研年報, 第 29 号 B-2, pp. 157-171.
- 竹内邦良・土屋一仁 (1988): 正規分布, 対数正規分布およびピアソンIII型分布の PWM 解, 土木学会論文集, 第 393 号/II-9, pp. 95-101.
- Greenhood, J. A. et al. (1979): Probability weighted moments; Definition and relation to parameters of several distributions expressible in inverse form, Water Resources Res., Vol. 15, pp. 1049-1054.
- Hosking, J. R. M., Wallis, J. R. and E. F. Wood (1985): Estimation of the generalized extreme value distribution by the method of probability-weighted moments, Technometrics, Vol. 27, No. 3, pp. 251-261.
- Miller, R. G. (1974): The jackknife-A review, Biometrika, Vol. 61, No. 1, pp. 1-15.
- Phien, H. N. and T. S. Emma (1989): Maximum likelihood estimation of the parameters and quantiles of the general extreme-value distribution from censored data, Jour. Hydrol., No. 105, pp. 139-155.