

censoring を伴う年最大値資料に対する極値統計解析における PWM 法の適用性

山口 正 隆* ・ 宇都宮 好 博** ・ 野 中 浩 一***
真 鍋 晶**** ・ 畑 田 佳 男*****

Gumbel 分布, 形状母数固定型 Weibull・GEV 分布の 2 母数分布, および Weibull・GEV 分布の 3 母数分布を対象とした場合に, 第 2 種 censored sample に対する PPWM (部分確率加重積率) 法の精度を, censoring 比など種々の条件を変えたモンテカルロシミュレーション結果に基づいて検討したのち, 候補分布を上記の 2 母数分布および Gumbel 分布と 3 母数分布とする 2 種類の極値統計解析モデルを, 欠落を伴う年最大高潮偏差観測資料の解析に適用した. そして, 各確率分布に対する PPWM 法の適用条件ならびに標準偏差推定に対する jackknife 法の精度を明らかにするとともに, 当該モデルは確率高潮偏差と標準偏差を適切に評価することを示した.

1. 緒 言

欠落を伴う年最大値資料 (censored sample) に対する極値統計解析において, 種々の制約 (たとえば, 第 2 種 censored sample のみ対象) を伴うものの, 候補分布の母数推定を部分確率加重積率 (PPWM) 法によるモデルの適用も可能になっている. しかし, その適用性に関しては, Wang (1990, 1996) の GEV 分布および Gumbel 分布に対する少数の検討事例があるのみで, 広範な条件下での考察は行われていないし, 海岸工学の分野で多用される Weibull 分布に対する検討事例も見当たらない.

そこで本研究では, Gumbel 分布, 形状母数固定型 Weibull 分布, 形状母数固定型 GEV 分布の 2 母数分布, Weibull 分布, GEV 分布の 3 母数分布を対象とする大規模なモンテカルロシミュレーション結果の検討に基づいて, 各分布に対する PPWM 法の精度を調べたのち, これらの分布のいくつかで構成される極値統計解析モデルを大阪における欠落を伴う年最大高潮偏差資料に適用して, 確率高潮偏差とその標準偏差を推定する.

2. PPWM 法と精度の検討方法

(1) 候 補 分 布

Gumbel 分布, GEV 分布および Weibull 分布の確率分布 $F(H)$ は H を確率変数, k を形状母数, A を尺度母数, B を位置母数として, それぞれ次式で表される.

$$F(H) = \exp[-\exp\{-(H-B)/A\}], \quad -\infty < H < \infty \quad \dots\dots\dots (1)$$

$$F(H) = \exp[-\{1 + (H-B)/kA\}^{-k}], \quad \begin{aligned} k > 0: B - kA < H < \infty, \\ k < 0: -\infty < H < B - kA \end{aligned} \quad \dots\dots\dots (2)$$

$$F(H) = 1 - \exp[-\{(H-B)/A\}^k], \quad B < H < \infty \quad \dots\dots\dots (3)$$

(2) PPWM 法

PPWM は一般に次式で定義される.

$$M_{p,r,s} = \int_{F_1}^{F_2} H(F)^p F^r (1-F)^s dF \quad \dots\dots\dots (4)$$

ここに, F_1, F_2 はそれぞれ censoring 条件の下限値と上限値を与える非超過確率, $H(F)$ は式 (1)~式 (3) の inverse form のいずれかから計算される確率変数, である. 式 (4) において, Gumbel 分布と GEV 分布の場合には, $p=1, s=0$ とした $M_{1,r,0}(=\beta_r)$, Weibull 分布の場合には, $p=1, r=0$ とした $M_{1,0,s}(=\beta_s)$ を用いる. 以下では, Weibull 分布に限定して説明を行う. まず β_s は $k'=1/k$ として, 次式で表される.

$$(1+s)\beta_s = [-B(1-F)^{1+s} - A(1+s)^{-k'}\gamma(k'+1; t_s)]_{F_1}^{F_2} \quad \dots\dots\dots (5)$$

$$\gamma(k'+1; t_s) = \int_{t_s}^{\infty} t^{(k'+1)-1} e^{-t} dt \quad \left. \begin{aligned} t_s = -(1+s)\log(1-F) \end{aligned} \right\} \quad \dots\dots\dots (6)$$

ここに, $\gamma(k'+1; t_s)$ は不完全 γ 関数, および $[D]_{F_1}^{F_2} = D(F_2) - D(F_1) \quad \dots\dots\dots (7)$ である. ついで, PPWM を,

$$\beta_s = \beta_s / \{(1+s) \int_{F_1}^{F_2} (1-F)^s dF\} \quad \dots\dots\dots (8)$$

で再定義すると, 式 (5) は次式になる.

$$(1+s)\beta_s = -B - A(1+s)^{-k'} [\gamma(k'+1; t_s)]_{F_1}^{F_2} / [-(1-F)^{s+1}]_{F_1}^{F_2} \quad \dots\dots\dots (9)$$

下限, 上限 censoring 条件ではそれぞれ $F_2=1, F_1=0$ とする.

一方, 標本 PPWM の再定義値 b'_s は $H_1 < H_2 < \dots < H_N$ として昇順に順序づけられた標本資料 H_i に対して次式で定義される.

$$b_s = (1/N) \sum_{i=1}^N \{(N-1)(N-2)\dots(N-i-s+1)\} H_i / \{(N-1)(N-2)\dots(N-s)\} \quad \dots\dots\dots (10)$$

$$b'_s = b_s / [-(1-F)^{s+1}]_{F_1}^{F_2} \quad \dots\dots\dots (11)$$

ここに,

$$H_i = 0; H_i < H(F_1), \quad H_i = 0; H_i > H(F_2) \quad \dots\dots\dots (12)$$

* 正 会 員 工 博 愛媛大学教授 工学部環境建設工学科
** 正 会 員 工 修 (財)日本気象協会首都圏支社
*** 正 会 員 博(工) 愛媛大学契約職員 工学部環境建設工学科
**** 学生会員 愛媛大学大学院理工学研究科博士前期課程
***** 正 会 員 博(工) 愛媛大学助手 工学部環境建設工学科

である。ここで、理論 PPWM β_s と標本 PPWM b_s を等値し、 $s=(0, 1, 2)$ に対応する式 (9) と式 (11) を用いれば、Weibull 分布の母数が算出される。また、 $s=(1, 2, 3)$ および $s=(2, 3, 4)$ の高次 PPWM に対しても同様の方法で母数推定を行う。形状母数固定型 2 母数 Weibull 分布に対しては、 $s=(0, 1)$ 、 $s=(1, 2)$ 、 $s=(2, 3)$ と置いた式 (9) の 2 つの式に対する代数計算から母数 A 、 B を算出できる。Gumbel 分布 (下限 censoring 条件のみ) および GEV 分布 (上下限 censoring 条件) に対する以上の諸式は Wang (1990, 1996) に与えられる。

(3) シミュレーションの方法

モンテカルロシミュレーションは、尺度母数 A と位置母数 B を固定し、種々の形状母数 k をとる上記の 5 種類の分布を母分布として、censoring 条件 (下限 censoring 比 $\mu_L=F_1$ 、上限 censoring 比 $\mu_T=1-F_2$ 、上位削除数 M_T) や PPWM の次数を変えた場合に、資料年数 $K=10\sim 100$ の 7 通りについて 10,000 回実施し、bias ΔH 、分散 Var 、分散の jackknife 推定値 σ_j^2 の平均値を求める。用いる標本資料は、母数を指定した確率分布に一樣乱数を与えて作成した資料年数 K に相当する大きさ N の標本から、censoring 比 μ_L 、 μ_T に応じて下限 p 個、上限 q 個、上下限 $p+q$ 個を削除した第 2 種下限・上限・上下限 censored sample である。また解析においては、標本の作成に使用した確率分布のみを候補分布とする。すなわち、Gumbel 分布を除く 2 母数分布では、形状母数を固定して解析を行う。

図は再現期間 100 年について、入力母分布から得られる確率統計量の真値で無次元化 bias $\Delta \hat{H}$ と $\bar{V}ar^{1/2}$ 、および $Var^{1/2}$ で無次元化した分散の jackknife 推定値の平均値に対する平方根 σ_j の % 値の 3 つ、あるいは $\Delta \hat{H}$ と σ_j の % 値の 2 つの誤差統計量と資料年数 K との関係として示す。

(4) 極値統計解析モデルの構成

母分布が未知である現地観測資料に適用しうる極値統

計解析モデルとして、候補分布を Gumbel 分布、形状母数固定型 Weibull 分布 (形状母数 k は 0.5~10 の間の 27 種類)、形状母数固定型 GEV 分布 (形状母数 k は 2.5~40 の間の 20 種類) の 2 母数分布、および Gumbel 分布と、Weibull 分布・GEV 分布の 3 母数分布とし、最適分布の選択を相関係数最大基準による 2 種類のモデルを作成する。相関係数の算出においては、個々の標本資料への非超過確率値の割当てに、汎用性の高い Cunnane (1978) のプロットング公式、

$$F=(i-0.4)/(N+0.2); \quad i=1\sim N \cdots \cdots (13)$$

を適用する。2 母数分布の使用は線形最小 2 乗法を用いたモデルに準拠しており、モデルの適用範囲の拡大と安定した解を得ることを目的とする。これらのモデルは候補分布を単独の分布あるいは 2 種類の分布とすることもできる。

3. PPWM 法の精度の検討

各母分布に対する誤差統計量に影響を及ぼす要因は、資料年数 K のほか、PPWM の次数 (I_{rs})、下限 censoring 比 μ_L 、形状母数 k 、上位削除数 M_T 、上限 censoring 比 μ_T であるので、以下ではこれらのうち 3 つのパラメータについて図を与える。

図-1 は Gumbel 分布 ($A=1.39$, $B=4.5$) に対する誤差統計量と資料年数の関係を示す。図は左からそれぞれ PPWM の次数 I_{rs} 、下限 censoring 比 μ_L 、上限 censoring 比 μ_T の影響を表す。これから、つぎのことが言える。

① 下限 censoring 条件 ($\mu_L=0.2$) では、bias は (0, 1) 次積率 ($I_{rs}=0$) より (1, 2) 次積率 ($I_{rs}=1$) または (2, 3) 次積率 ($I_{rs}=2$) を用いる場合に小さい。これは広範な μ_L 条件について成り立つ。一方、上限 censoring 条件では、bias は (0, 1) 次積率を用いる場合に小さい。したがって、Gumbel 分布の場合、下限 censoring 条件では (1, 2) 次積率または (2, 3) 次積率、上限 censoring 条件では (0, 1) 次積率を使用する必要がある。

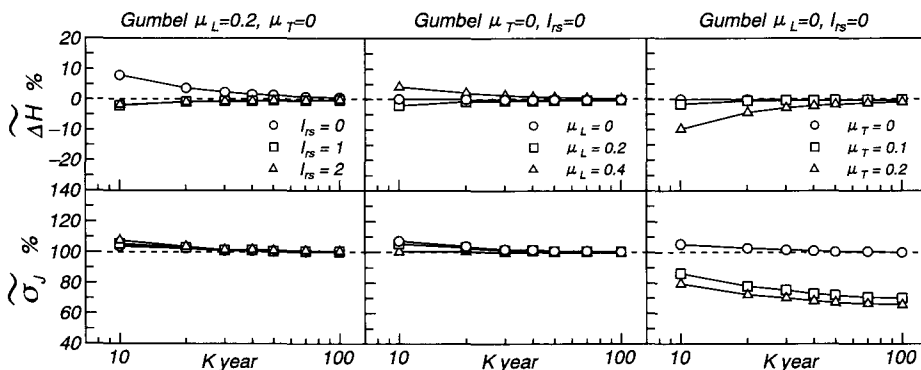


図-1 誤差統計量と資料年数の関係 (Gumbel 分布)

② 下限 censoring 比 $\mu_L < 0.4 \sim 0.5$ に対して, bias はほとんど生じないが, μ_L が大きくなれば, bias が急増する. 上限 censoring 条件では, $\mu_T = 0.1$ の場合 bias はほとんどない. $\mu_T = 0.2$ では, 有意な負の bias を生じる. Gumbel 分布に対する PPWM 法の適用範囲は $\mu_L < 0.5$, $\mu_T < 0.2$ である.

③ jackknife 法は積率の次数および下限 censoring 比によらず, 適切な標準偏差の推定値を与える. しかし上限 censoring 条件では, 10~30% 低く評価し, しかも資料年数が大きいほどその傾向が強まる.

図-2 は形状母数固定型 GEV 分布 ($A=1.0$, $B=5.0$) に対する誤差統計量を示す. その挙動は Gumbel 分布の場合と類似するので, 上記以外の特徴を述べればつぎのようである.

① $Var^{1/2}$ は高次積率を用いる場合ほど大きい傾向にあるので, (1, 2) 次積率 ($I_{rs}=1$) を用いる方が適切である. これは Gumbel 分布についても言える.

② 下限 censoring 条件では, bias は下限 censoring 比 μ_L や形状母数 k によらず小さい. しかし $\mu_L > 0.5$ では, bias が増大することから, $\mu_L < 0.5$ が PPWM 法の適用範囲である.

③ $Var^{1/2}$ は下限 censoring 比 μ_L が大きくなると若干増加し, また当然, 形状母数 k に強く依存する.

図-3 は GEV 分布 ($A=1.0$, $B=5.0$) に対する結果であり, つぎの特徴を指摘できる.

① 下限 censoring 条件では, bias は (0, 1, 2) 次積率 ($I_{rs}=0$) より (1, 2, 3) 次積率 ($I_{rs}=1$) および (2, 3, 4) 次積率 ($I_{rs}=2$) を用いる場合に小さい. 広範な下限 censoring 条件のもとでは, (2, 3, 4) 次積率 ($I_{rs}=2$)

を用いる場合に小さい. また, 上限 censoring 条件では (0, 1, 2) 次積率を用いる必要がある.

② 下限 censoring 比 μ_L が 0.4 以下では, bias は小さい. また, 正側にきわめて長く分布の裾を引く形状母数 $k=2.5$ のとき, 負の bias が有意な値をとるが, k が 5 より大きくなると, bias は実際上無視しうる. 上限 censoring 条件では, μ_T が 0.1 を越えると bias が増大する.

③ 下限 censoring 条件では, jackknife 法は, 形状母数が負値をとり絶対値が小さいときほど, やや過大な値を与える傾向にあるが, 資料年数 K が 20~30 以上であれば, いずれのケースについても比較的高い精度を与える. しかし上限 censoring 条件では, 標準偏差を 10~30% 過小評価する.

④ GEV 分布は Gumbel 分布や形状母数固定型 GEV 分布に比べて母数が 1 つ多いことから, bias がより大きく, jackknife 法の精度もやや低い傾向にある.

図-4 は形状母数固定型 Weibull 分布 ($A=4.0$, $B=1.0$) に対する結果を与える. その特徴はつぎのようである.

① 下限 censoring 条件では, (0, 1) 次積率 ($I_{rs}=0$) を用いる場合の bias が (1, 2) 次積率 ($I_{rs}=1$) や (2, 3) 次積率 ($I_{rs}=2$) という高次積率を用いる場合よりはるかに小さい. 一方, 上限 censoring 条件では (2, 3) 次積率 ($I_{rs}=2$) を用いる方が, 相対的に適切な結果を与える. これは Weibull 分布の場合にも同じである.

② 同一の形状母数の場合, 下限 censoring 比 μ_L の増加とともに, あるいは同一の下限 censoring 比の場合, 形状母数 k の増加, すなわち分布形状の尖鋭化とともに, bias が大きくなる. 下限 censoring 比に対する限界値は $\mu_L =$

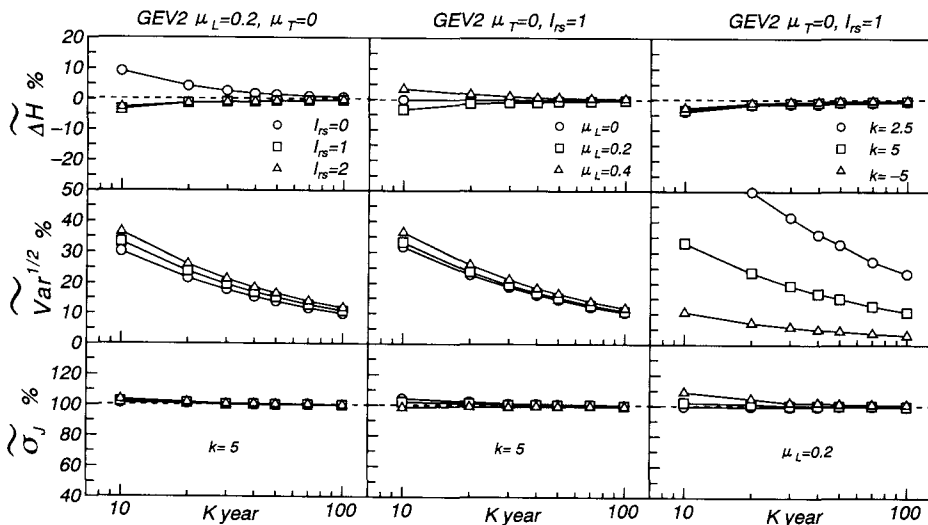


図-2 誤差統計量と資料年数の関係 (形状母数固定型 GEV 分布)

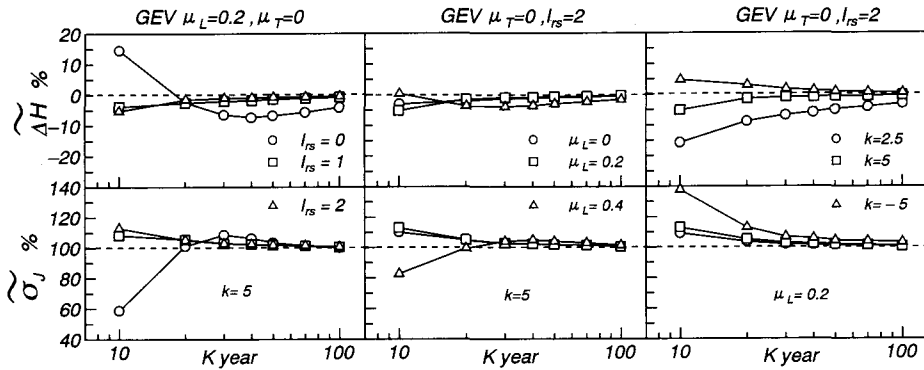


図-3 誤差統計量と資料年数の関係 (GEV 分布)

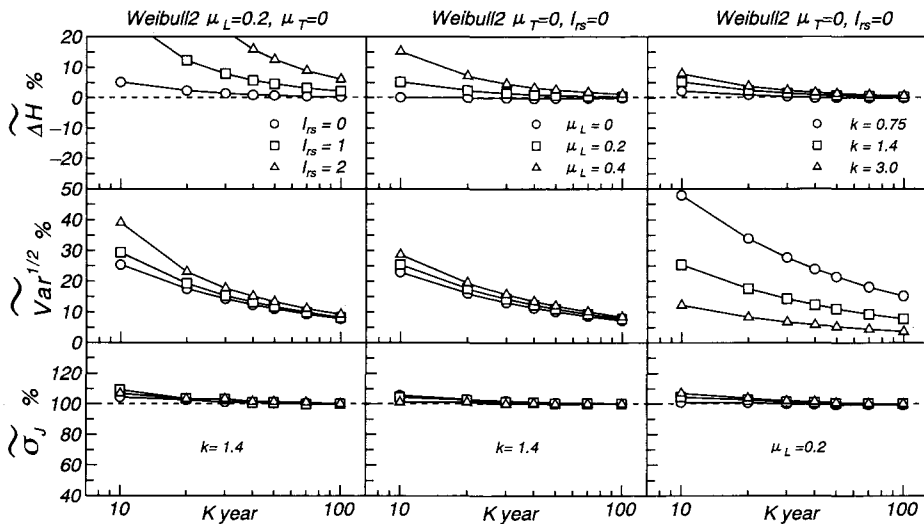


図-4 誤差統計量と資料年数の関係 (形状母数固定型 Weibull 分布)

0.3~0.4 であり、Gumbel 分布や GEV 分布の場合よりやや小さい。しかし、bias の形状母数への依存性は弱い。

③ $Var^{1/2}$ は、高次積率を用いる場合ほど、下限 censoring 比が大きいほど、また形状母数が小さいほど、増加する。 $Var^{1/2}$ の前 2 者への依存性は弱い、形状母数への依存性は当然のことながら非常に強い。

④ jackknife 法は下限 censoring 条件の場合のみならず上位削除数 M_T が 2 以下の場合に、標準偏差推定に対して高い精度を有する。しかし、上限 censoring 比 $\mu_T=0.1$ の場合で 10%、 $\mu_T=0.2$ の場合で 20% の過小評価傾向をもつ。

図-5 は Weibull 分布 ($A=4.0$, $B=1.0$) に対する結果を与える。これから以下の事項が指摘できる。

① 下限 censoring 条件では、bias は下限 censoring 比 μ_L にかなりの程度依存し、 μ_L が増加すると大きくなる。PPWM 法の適用範囲は $\mu_L < 0.3 \sim 0.4$ とみられる。また

形状母数が大きいほど、bias は増加するが、資料年数 K が 20 以上であれば、bias は有意な値をとらない。一方、上限 censoring 条件では、上位削除数 M_T が 2 以下で資料年数 K が 30~40 を越えると、bias は小さくなる。

② jackknife 法の精度は、下限 censoring 比 μ_L が 0.4 以下あるいは形状母数 k が 3 以下で資料年数 K が 30 以上の場合比較的良好である。上限 censoring 条件では、上位削除数 M_T が 2 以下で資料年数 K が 40 以上の場合に高い。

③ Weibull 分布も GEV 分布と同様に 3 母数をもつので、bias と jackknife 法の誤差が 2 母数分布の場合より大きい。

4. 現地観測資料への適用

大阪では 1902~2003 年の 102 年間に於いて 0.26 m より大きい年最大高潮偏差観測資料 81 個が得られてい

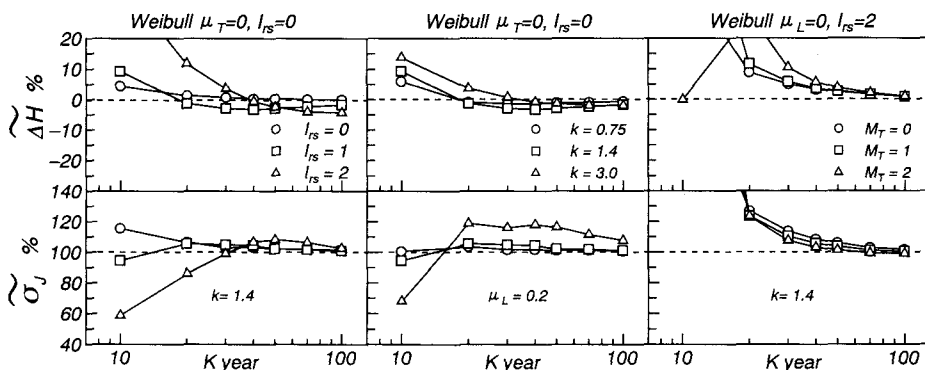


図-5 誤差統計量と資料年数の関係 (Weibull 分布)

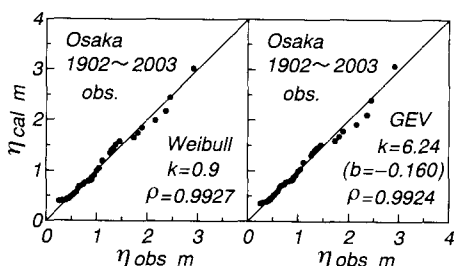


図-6 年最大高潮偏差資料に対する Q-Q プロット

る。このうち、第1位～第4位値はそれぞれ台風3412号(室戸台風)時の $\eta=2.92$ m, 台風6118号(第2室戸台風)時の2.45 m, 台風5028号(ジェーン台風)時の2.37 m, および台風6523号時の2.16 mであり、102年間で2 mを越える高潮偏差はこれら4ケースとなっている。

ここでは、候補分布を上述した2母数分布および3母数分布とする2種類のモデルによる極値統計解析を、下限 censoring 比 $\mu_L=0.206$ の条件のもとに行った。図-6は2母数分布および3母数分布を適用した場合のQ-Qプロットを示し、いずれも計算値と観測値の一致度は高い。この結果によると、2母数分布では最適分布は形状母数 $k=0.90$ のWeibull分布、相関係数は0.9927、200年確率高潮偏差とその標準偏差は 3.11 ± 0.35 m、3母数分布では形状母数 $k=6.24$ のGEV分布、0.9924、 3.20 ± 0.55 mと評価される。第1位高潮偏差2.92 mの再現期間はいずれのモデルでも140年前後であるから、合理的な推定値が得られている。因みに、母数推定を最小2乗法による場合、最適分布は形状母数 $k=0.9$ のWeibull分布、相関係数は0.9924、200年確率高潮偏差とその標準

偏差は 3.07 ± 0.36 mであるから、2母数分布による結果とはほぼ一致する。

5. 結 語

本研究で得られた結果の要点はつぎのようである。

- ① Gumbel分布とGEV分布の場合、下限 censored sample に対して高次積率を、上限 censored sample に対して低次積率をPPWMに用いる方が高い精度が得られる。Weibull分布では逆の関係になる。
- ② PPWM法の適用限界は $\mu_L < 0.4 \sim 0.5$, $\mu_T < 0.1 \sim 0.2$ である。jackknife法は下限 censored sample に対して適切な標準偏差の推定値を与えるが、上限 censored sample に対してはGumbel分布やGEV分布の場合10～30%低く評価する。いずれにせよ、標本の大きさの増加が精度の高い推定値を得るために最も重要である。
- ③ 2母数分布あるいは3母数分布を候補分布として母数推定をPPWM法による極値統計解析モデルは欠落を含む年最大値資料に対しても適用可能であり、その推定値も合理的と言える。

参 考 文 献

- Cunnane, C. (1978): Unbiased plotting position, -a review, Jour. Hydrol., Vol. 37, pp. 205-222.
- Wang, Q. J. (1990): Estimation of the GEV distribution from censored samples by method of partial probability weighted moments, Jour. Hydrol., Vol. 120, pp. 103-114.
- Wang, Q. J. (1996): Using partial probability weighted moments to fit the extreme value distributions to censored samples, Water Resour. Res., Vol. 32, No. 6, pp. 1767-1771.