

# censoring を考慮した極値統計解析モデルおよび 現地観測資料への適用

山口正隆\*・畑田佳男\*\*・大福 学\*\*\*・前川隆海\*\*\*\*

## 1. 緒 言

高潮などの極値統計解析において最も重要な点はできるだけ長期の連続した（年最大）観測資料を得ることである。しかし、連続記録が得られるのは最近の数十年であり、それ以前では、既往最大値など巨大台風時の記録に限られることが多い。こうした欠落を伴う資料は censored data とみなせることから、censoring の影響を含めた尤度関数の最大化に基づく方法（最尤法）によって解析可能となるが、高潮などの極値統計分野では、この方法の適用性は全く検討されていない。

そこで、本研究では母集団分布を Gumbel 分布、Weibull 分布、GEV 分布に指定した場合に、広範な条件下でのモンテカルロシミュレーション結果の解析に基づいて、censoring の影響を考慮した場合の、最尤法、bias 補正法および分散推定法である jackknife 法、および分散推定法である観測情報行列法の適用性を明らかにする。また、欠落を伴う高潮偏差などの観測資料に適用して、それらの再現確率統計量と標準偏差を推定するとともに、censoring を考慮しない解析結果との比較から、極値統計解析結果に及ぼす censoring の影響を考察する。

## 2. 確率分布と母数・分散推定法

### (1) 確 率 分 布

Gumbel 分布, 3 母数 Weibull 分布および GEV 分布の非超過確率  $F(x)$  はそれぞれ次式で表される。

$$\left. \begin{aligned} F(x) &= \exp[-\exp\{-(x-B)/A\}] & ; -\infty < x < \infty \\ F(x) &= 1 - \exp[-\{(x-B)/A\}^k] & ; B \leq x < \infty \\ F(x) &= \exp[-\{1 + (x-B)/kA\}^{-k}] & \\ & \quad -\infty < x \leq B - kA & ; k < 0 \\ & \quad B - kA \leq x < \infty & ; k > 0 \end{aligned} \right\} \dots\dots\dots (1)$$

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。

### (2) 母数推定法

censoring は censoring 基準値が固定され、censoring を受ける個数がランダムである第 1 種 censoring と、censoring 基準値がランダムで資料数が固定される第 2 種 censoring に、また censoring 基準値によって上限、下限、上下限 censoring に分類される。つまり、censoring は第 1 種上限・下限・上下限 censoring と第 2 種上限・下限・上下限 censoring の 6 種類に分けられる。

$N$  個の標本のうち下限値  $x_L$  以下の  $r$  個および上限値  $x_U$  以上の  $s$  個が censoring を受けている場合に、第 1 種 censoring 問題（下限値  $x_L$ 、上限値  $x_U$  は固定、 $r, s$  はランダム）に対する尤度関数（Phien・Emma, 1989）は次式で表される。

$$L_I = \{N!/(r! \cdot s!)\} \{F(x_L)\}^r \prod_{i=1}^n f(x_i) \{1 - F(x_U)\}^s \dots\dots\dots (2)$$

ここに、 $f(x_i)$  は確率密度関数、 $n = N - r - s$ 、である。一方、第 2 種 censoring 問題（ $r, s$  は固定、 $x_{r+1}, x_{N-s}$  はランダム）に対する尤度関数は、

$$L_{II} = \{N!/(r! \cdot s!)\} \{F(x_{r+1})\}^r \prod_{i=r+1}^{N-s} f(x_i) \times \{1 - F(x_{N-s})\}^s \dots\dots\dots (3)$$

である。したがって、それぞれの対数尤度関数を最大化する母数を求めれば最尤解が得られる。このときの最大対数尤度 MLL は標本に対する候補確率分布の適合度を表す指標となる。数値計算には、負符合を付けた対数尤度関数に対して極小値を求める最適化法を適用する。

### (3) 分散推定法

再現確率統計量の bias 補正法や分散（標準偏差）推定法として、censored data に対しても jackknife 法を用いる。ただし、jackknife 法は、多くの計算時間を要するので、その適用を  $N \leq 200$  の場合に限定する。また、最尤法による漸近的分散推定法として、観測情報行列法および Fisher 情報行列法がある。本研究で取り扱う理論確率分布に対する観測・Fisher 情報行列は上・下限の censoring を受けた場合を含めて導かれている（たとえば、Phien・Emma, 1989）が、Fisher 情報行列は多くの印刷上の誤りを含むので、検討対象としない。

\* 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 愛媛大学助手 工学部環境建設工学科  
\*\*\* 愛媛大学技官 工学部環境建設工学科  
\*\*\*\* 工修 日本建設コンサルタント株式会社

### 3. モンテカルロシミュレーションに基づく精度の検討

#### (1) 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 母数を指定した確率分布に一樣乱数を与えて、資料年数に相当する $N$ 個の標本を作成し、第1種 censoring の場合には下限 censoring 値  $x_L$  以下あるいは上限 censoring 値  $x_U$  以上の資料を別個に、または同時に削除した第1種下限・上限・上下限 censoring 資料を、第2種 censoring の場合には censoring 比  $\mu$  に応じてその相当する下限  $r$  個、上限  $s$  個、上下限  $r+s$  個を削除した第2種下限・上限・上下限 censoring 資料を抽出する。

② 最尤法による母数の推定および再現確率統計量と bias  $\Delta H$  の計算、jackknife 法による bias 補正後の bias  $\Delta \tilde{H}_j$  と分散  $\sigma_j^2$  の計算ならびに観測情報行列法による分散  $\sigma_{OM}^2$  の演算を行う。

③ 同じ手順を  $M$  ( $=5000$ ) 回繰り返し、②の諸量の平均値および再現確率統計量の分散  $Var$  を求める。

④ 資料年数を  $N=10\sim 1000$  の間の10通りとして、

以上の計算を繰り返す。

シミュレーションは censoring 値条件 ( $x_L, x_U$ ) や censoring 比  $\mu$  を適宜変えて行い、Weibull 分布および GEV 分布の場合には形状母数  $k$  を変化させる。

#### (2) シミュレーション結果の考察

図-1は、Gumbel 分布 ( $A=1.39, B=4.5$ )、3 母数 Weibull 分布 ( $A=4.0, B=1.0, k=2.0$ ) および GEV 分布 ( $A=1.0, B=5.0, k=5.0$ ) において、censoring の下限値を非超過確率  $F(x_L)=0.1$  に対応する値  $x_L$ 、上限値を  $F(x_U)=0.9$  に対応する  $x_U$  とした場合の、下限 (lower), 上限 (upper), 上下限 (double) の第1種 censoring 資料に対する誤差統計量と資料年数  $N$  との関係を示したものである。下限 censoring 値  $x_L$  (固定) に対する資料個数  $r$  あるいは上限 censoring 値  $x_U$  (固定) に対する資料個数  $s$  は1回のシミュレーションごとに異なる。誤差統計量は、再現期間100年の確率統計量の真値で無次元化した bias ( $\Delta \tilde{H}, \Delta \tilde{H}_j$ )、 $\bar{Var}^{1/2}$ 、および  $Var^{1/2}$  で無次元化した標準偏差推定値 ( $\tilde{\sigma}_j, \tilde{\sigma}_{OM}$ ) に対する%値で表されており、censoring を考慮しない資料 (complete data) に対する結果も与えられる。この図および他の条

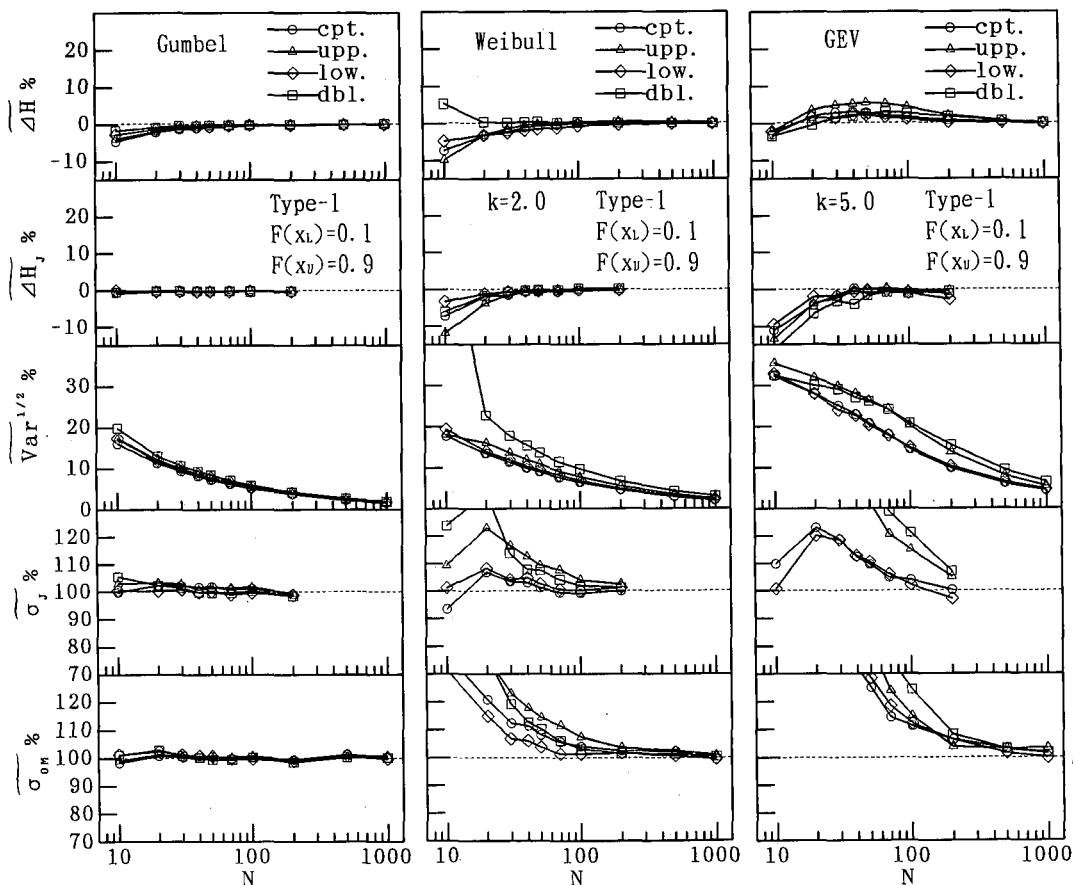


図-1 censoring 資料に対する誤差統計量 (1)

件に対する図から、つぎの特徴が見出される。

① Gumbel 分布の場合,  $\text{bias } \Delta H$  は上下限 censoring, 上限 censoring, 下限 censoring, censoring なし, の順に小さいし, 資料年数の増加とともにほぼ消失する. jackknife 法による  $\text{bias}$  補正後の  $\text{bias } \Delta H_j$  は資料年数が短い場合でもほぼ 0 になることから, jackknife 法は censoring の有無によらず, 再現確率統計量の  $\text{bias}$  補正に対してきわめて有効である. また, 例示した条件に対する Gumbel 分布の場合には明瞭でないが,  $\text{Var}^{1/2}$  は資料年数が小さく, かつ censoring の割合が高いほど, とくに上下限 censoring および上限 censoring の場合に大きい. この傾向は 2 母数分布である Gumbel 分布より②, ③に述べる 3 母数分布である Weibull 分布や GEV 分布の場合に著しい. jackknife 法, 観測情報行列法はいずれの資料についても標準偏差を適切に評価するが, 計算時間および精度を考慮すると, 標準偏差推定法として観測情報行列法の方が適切である.

② 3 母数 Weibull 分布の場合,  $\text{bias } \Delta H$  の絶対値は下限 censoring, 上限 censoring の順に大きくなるが, 形状母数が大きいときほど小さい. しかし, 上下限 censor-

ing に対しては,  $\text{bias}$  は  $N \geq 20$  では形状母数によらずほとんど生じない. jackknife 法による  $\text{bias}$  補正は  $N \geq 20 \sim 30$  である程度有効に作用し,  $\text{bias } \Delta H_j$  が減少する. なお, 最尤法による形状母数の解の存在範囲は  $k > 1$  であり,  $k < 1.4$  では安定した解が得られないこともある. jackknife 法による標準偏差推定法はとくに上限 censoring 資料に対して過大評価傾向をもつが, 他の censoring 資料では,  $N \geq 30$  で 10% 以内の推定誤差となる. ただし, censoring の割合が高くなれば, 上下限 censoring 資料に対しても過大評価傾向が現れる. 観測情報行列法も同様の特性を示すが, censoring 条件の相違に伴う誤差特性が明確に分れる.

③ GEV 分布の場合, 形状母数  $k \geq 5$  では  $\text{bias}$  は全般的に小さく, jackknife 法に基づく  $\text{bias}$  補正機能も  $N \geq 30$  では比較的有效に作用する. しかし, 形状母数が小さくなれば, 資料年数が短いほど負の  $\text{bias}$  が増加し, jackknife 法によっても  $\text{bias}$  は縮小しない. また, Weibull 分布の場合と同様, 資料年数が短く, 形状母数が小さいときほど, jackknife 法はとくに上限 censoring 資料および上下限 censoring 資料に対して標準偏差をかなり過大評

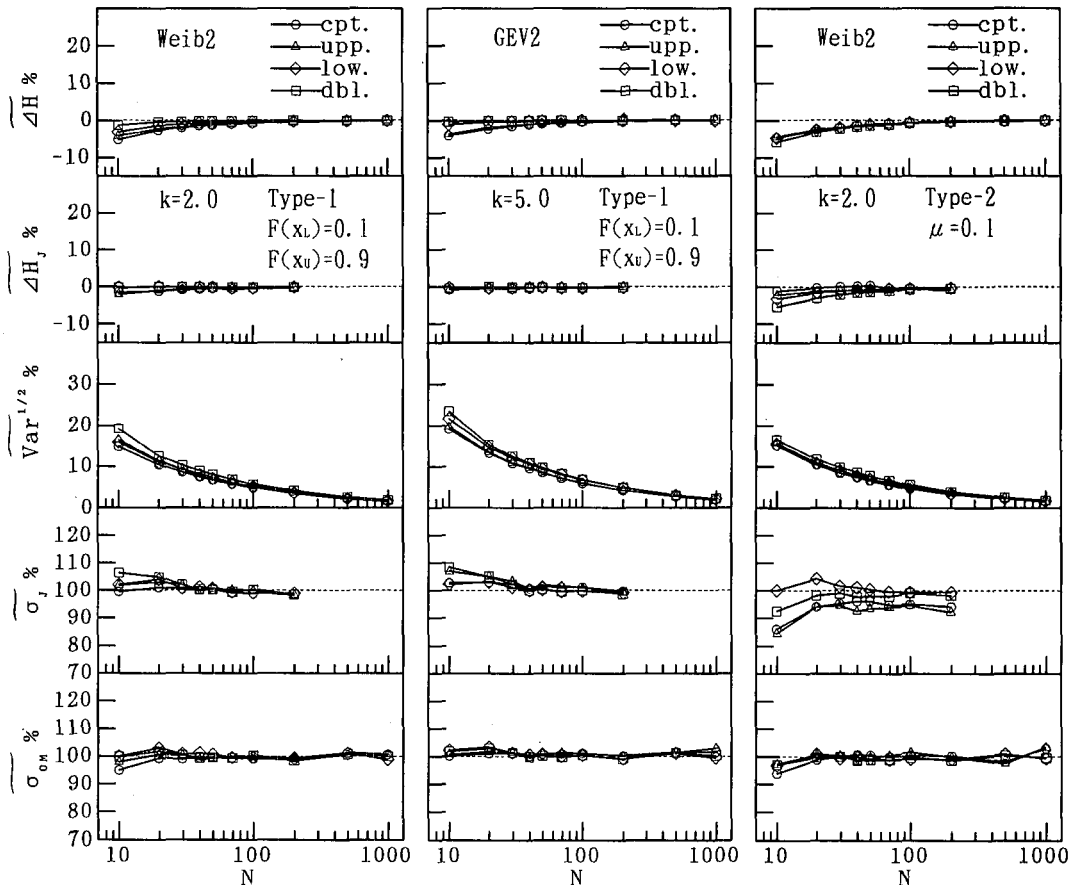


図-2 censoring 資料に対する誤差統計量 (2)

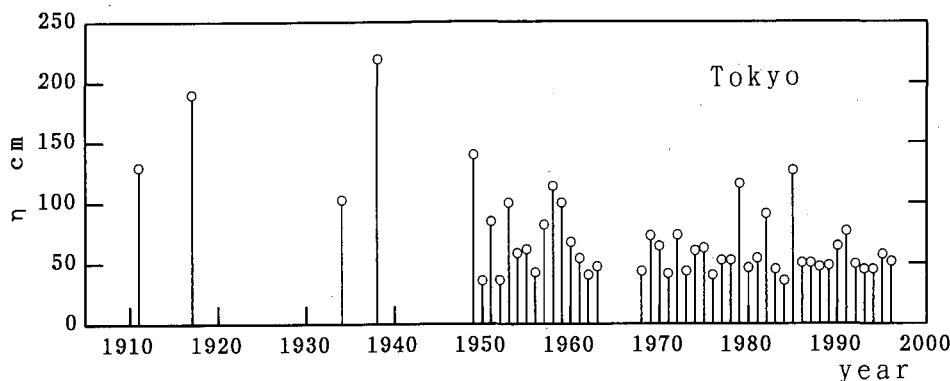


図-3 東京における年最大高潮偏差の時系列

価するし、観測情報行列法も全体的に大き目の標準偏差推定値を与える。

図-2は3母数 Weibull 分布 ( $A=4.0$ ,  $B=1.0$ ,  $k=2.0$ ) および GEV 分布 ( $A=1.0$ ,  $B=5.0$ ,  $k=5.0$ ) による第1種 censoring 資料 ( $F(x_L)=0.1$ ,  $F(x_U)=0.9$ ) の解析にそれぞれ形状母数固定型 Weibull 分布 (Weib2;  $k=2.0$ ) および形状母数固定型 GEV 分布 (GEV2;  $k=5.0$ ) を適用した場合と、3母数 Weibull 分布 ( $A=4.0$ ,  $B=1.0$ ,  $k=2.0$ ) による第2種 censoring 資料 (censoring 比  $\mu=0.1$ ) の解析に形状母数固定型 Weibull 分布 (Weib2;  $k=2.0$ ) を適用した場合の誤差統計量図である。第2種の下限 censoring あるいは上限 censoring では全資料年数に対して、それぞれ下位あるいは上位の10%の資料が、上下限 censoring では、上位および下位の10%の資料が削除されており、上・下限 censoring 値が資料ごとに異なる。

第1種 censoring 資料に対する誤差統計量の特性は Gumbel 分布に対するものと類似しており、censoring 資料の種類によらず、jackknife 法は bias 補正に有効に機能し、標準偏差を比較的精度よく評価すること、および観測情報行列法に基づく標準偏差推定法の精度も良好であることがわかる。また、いずれの方法の推定精度も censoring 値にあまり依存せず、形状母数の増加に伴い向上する。

一方、第2種 censoring 資料の場合、jackknife 法による bias 補正後の  $\text{bias} \Delta H_f$  は依然として負値をとっており、jackknife 法による bias 補正は第1種 censoring の場合ほど有効に機能しない。また、jackknife 法は上限 censoring 資料や censoring なしの資料に対して標準偏差をやや過小評価する点を除いて比較的良好な精度をもつことや、観測情報行列法の標準偏差推定精度が高いことなどが指摘される。これらの結果は Gumbel 分布や GEV 分布についても云えるが、その場合 jackknife 法の標準偏差推定精度は censoring 資料の種類によらずかな

り高い。なお、 $\text{Var}^{1/2}$  は図示した例では censoring 資料の種類にあまり依存しないが、censoring の割合が増加すれば、とくに上下限 censoring の場合に大きくなる。

要するに、censoring 資料に対して極値統計解析を行う場合、3母数分布 (Weibull 分布、GEV 分布) の適用は数値計算結果の安定性および再現確率統計量とその標準偏差の推定精度の点で得策でないで、最小2乗法に基づく極値統計解析システム (合田, 1990) にならって、Gumbel 分布や形状母数を固定した複数の Weibull 分布あるいは GEV 分布の中から MLL 基準や AIC 基準および分散最小基準に従って最適分布を選択するシステムを構築することが望ましいと考えられる。ただし、本研究では形状母数固定型3母数分布の適用にあたって、3母数分布に対して得られた形状母数をそのまま使用する。

#### 4. 現地観測資料への適用

##### (1) 高潮偏差資料

今回入手した東京における年最大高潮偏差資料は図-3に示すように、1911~1996年の86年間相当分であるが、このうち1948年以前の38年間では巨大台風に伴う4ケースの異常高潮偏差しか明らかでない。しかもこの中に、観測期間中の既往最大値 (1938年の220 cm) および第2位最大値 (1917年の190 cm) が含まれており、最も低い値でも103 cmである。そこで、東京における年最

表-1 年最大高潮偏差資料に対する極値統計解析結果

|                  | Gumbel | GEV   | GEV2  | GEV2* |
|------------------|--------|-------|-------|-------|
| MLL              | 185.2  | 193.0 | 193.0 | 0.019 |
| $k$              |        | 2.516 | 2.516 | 1.942 |
| $\eta_{100}$ cm  | 144    | 225   | 225   | 341   |
| $\eta_{500}$ cm  | 10     | 62    | 25    | 49    |
| $\eta_{1000}$ cm | 154    | 280   | 249   | 343   |
| $\eta_{99}$ cm   | 15     | 49    | 24    | 48    |

\* complete data

表-2 年最大風速資料に対する極値統計解析結果

|                  | Gumbel | Weibull | Weib2  | Weib2* |
|------------------|--------|---------|--------|--------|
| MLL              | -9.660 | -8.625  | -8.625 | -93.68 |
| $k$              |        | 1.512   | 1.512  | 1.646  |
| $U_{100}$ m/s    | 41.3   | 41.4    | 41.4   | 38.7   |
| $U_{50M}$ m/s    | 2.8    | 3.9     | 2.7    | 2.3    |
| $U_{1100}$ m/s   | 41.5   | 38.7    | 42.7   | 39.7   |
| $U_{\sigma}$ m/s | 2.9    | 4.3     | 2.9    | 2.3    |

\* complete data

表-3 年最大波高資料に対する極値統計解析結果

|                | Gumbel | Weibull | Weib2 | Gumbel* |
|----------------|--------|---------|-------|---------|
| MLL            | 3.350  | 3.305   | 3.305 | -24.39  |
| $k$            |        | 1.515   | 1.515 |         |
| $H_{100}$ m    | 10.84  | 10.70   | 10.70 | 10.57   |
| $H_{50M}$ m    | 1.19   | 1.59    | 1.15  | 1.09    |
| $H_{1100}$ m   | 10.71  | 11.10   | 10.71 | 10.74   |
| $H_{\sigma}$ m | 1.16   | 2.14    | 1.26  | 1.13    |

\* complete data

大高潮偏差資料を、資料年数  $N=86$  年、資料数  $n=48$ 、下限 censoring 値  $\eta_L=103$  cm、下限 censoring 資料数  $r=38$  の第 1 種下限 censoring 資料とみなして、censoring の影響を考慮した場合および無視した場合（資料数が資料年数に等しいとする）の極値統計解析を行った。表-1 はその結果をまとめたものである。候補分布は Gumbel 分布、GEV 分布、形状母数固定型 GEV 分布である。資料分布が正側に長く裾を引くため、Weibull 分布の最尤解は得られない。この表によると、最大対数尤度 MLL（適合度）が大きい分布のうち標準偏差が小さい（有効性）分布は形状母数固定型 GEV 分布であり、jackknife 法に基づく 100 年確率高潮偏差  $\eta_{100}$  とその標準偏差  $\eta_{\sigma}$  は、 $249 \pm 24$  cm と推定される。これらの推定値は下限 censoring 値を  $\eta_L=80$  cm まで下げて 3~4 cm 程度しか変化しない。また censoring の影響を考慮しない場合には、資料年数を短く評価しているので、確率高潮偏差は 1 m 近く過大な値をとる。

(2) 風速資料

風速資料は志摩半島先端に位置する大王崎灯台（風速計の海拔高度 37.2 m）において得られた 1953~1986 年の 34 年間の年最大風速資料である。このうち、1959 年の年最大風速は台風 5915 号（伊勢湾台風）によって生じた既往最大風速に対応するはずであるが、欠測となっており、1972 年の 36.6 m/s が第 1 位風速となっている。そこで、資料年数  $N=34$  年、上限 censoring 値  $U_U=36.6$  m/s、上限 censoring 資料数  $s=1$  の第 2 種上限 censoring 問題としての解析を行った（表-2）。候補分布は Gumbel 分布、3 母数 Weibull 分布、形状母数固定型 Weibull 分布、GEV 分布および形状母数固定型 GEV 分布であるが、GEV 分布は最適分布とならないので、表から削除している。この場合の最適分布は形状母数固定型 Weibull 分布であり、jackknife 法に基づく 100 年確率風速  $U_{100}$  およびその標準偏差  $U_{\sigma}$  は  $42.7 \pm 2.9$  m/s と評価される。すなわち、100 年確率風速は censoring の影響を考慮しない結果より 3 m/s 大きい。

(3) 波高資料

波高資料は宮崎県油津港の水深約 50 m 地点において

1975~1990 年の 16 年間にわたり得られた観測資料に基づく年最大波高資料であるが、1984 年における波高記録は取得されていない。観測期間中の第 1 位波高は台風 8013 号時に生じた 10.34 m であり、一方、1984 年における年最大波高は周辺海域における観測記録から、年最大波高の最小値（1985 年の 4.57 m）より低いと推定される。そこで、油津における年最大波高資料を、資料年数  $N=16$  年、資料数  $n=15$ 、下限 censoring 値  $H_L=4.57$  m、下限 censoring 資料数  $r=1$  の第 1 種下限 censoring 資料とみなして、解析を行った（表-3）。候補分布は上記の 5 種類である。このうち、GEV 分布の形状母数は 200 以上と非常に大きい値をとり、Gumbel 分布とほぼ一致するので、表から削除している。本資料に対する最適分布は Gumbel 分布、jackknife 法に基づく 100 年確率波高  $H_{100}$  とその標準偏差  $H_{\sigma}$  は  $10.71 \pm 1.16$  m であるから、censoring の影響を考慮しない結果とほぼ同じ値を示す。

5. 結 語

本研究の検討から、censoring 資料を扱う場合、2 母数分布（Gumbel 分布、形状母数固定型 Weibull 分布、形状母数固定型 GEV 分布）では、再現確率統計量は jackknife 法に基づく bias 補正值、その標準偏差は観測情報行列法および jackknife 法に基づく推定値によって精度よく与えられるが、3 母数分布（Weibull 分布、GEV 分布）では、資料年数が小さいほど再現確率統計量および標準偏差は必ずしも適切に評価されない傾向にあること、および欠落を伴う年最大値資料の極値統計解析にあたっては、censoring の影響を考慮する必要があることが見出された。

参 考 文 献

合田良實(1990): 港湾構造物の耐波設計一波浪工学への序説一、鹿島出版会、333 p.  
Phien, H. N. and F. T. S. Emma (1989): Maximum likelihood estimation on the parameters and quantiles of the general extreme-value distribution from censored sample, Jour. Hydrol., Vol. 105, pp. 139-155.