

# 年最大値の極値統計量に及ぼす標本サイズの影響 に関する数値実験

Numerical Experiments for Examination of Sample Size Effect on Return Wave Heights  
Estimated from Samples of Annual Maximum Series Data

合田良実

Yoshimi GODA

土木学会名誉会員 工博 横浜国立大学名誉教授 (株) エコー (〒110-0014 東京都台東区北上野 2-6-4)

Distributions of annual maximum wave heights theoretically approach to the generalized extreme-value (GEV) distribution when the sample size is sufficiently large. A series of Monte Carlo simulations have confirmed it for samples of 10,000 years long extracted from the peaks-over-threshold data from the parents of the GEV, generalized Pareto (GPA), and Weibull distributions. However, samples of annual maximum data with duration shorter than several hundred years deviate from the GEV considerably and estimated return wave heights are smaller than the parent values for the return period longer than 50 years. Empirical prediction formulas for the coefficient of variation of return wave heights are derived for the three parent distribution functions.

**Key Words:** *Extreme wave statistics, annual maximum series data, GEV distribution, GPA distribution, Weibull distribution, quantile estimation*

## 1. まえがき

水文統計をはじめとする自然現象の極値事象を扱う分野では、ある閾値を超える極値のすべてを対象とするのか、あるいは各年のうちの最大値に限定して解析すべきであるかについて、早くから議論されてきた。前者のデータは極値時系列データあるいは極大値資料と呼ばれ、後者は年最大値資料と呼ばれる。

統計解析の作業量から言えば毎年最大値資料のほうが楽であるためか、日雨量や高水流量などでは年最大値資料を扱うことが多い。しかし、ある年の年最大値として取り上げられなかった極大値であっても、他の年の年最大値よりも大きいことは珍しくない。そのため、そうした極大値を無視するのは不適切であり、極大値資料に基づいて100年確率統計量などを推定すべきあるとの意見がある。これについては、Langbein (1949) が簡単な確率計算に基づいて極大値と年最大値に対する  $T$  年確率統計量の再現期間の差が小さいことを例示し、実用的には無視できると説明した。また、

著者は極大値の事象がポアソン分布することを前提として年最大値の分布関数を極大値の分布から導き、極大値の分布から推定される確率波高の再現期間が年最大値分布のものよりも 1/2 年短いだけであることを示している (合田 [2008], p.335 参照)。

波浪の分野では計器観測の歴史が短いこともあって、極値解析はもっぱら極大値資料に基づいてきた。しかし、わが国の沿岸波浪観測のデータが蓄積されるにつれて40年に近い期間にわたる年最大波高のデータが解析可能となってきた。こうしたデータについて試算してみると、極大値資料と年最大値資料とでは分布関数に有意な違いが見られ、100年確率波高も後者のほうが低いことが少なくない。このため、一連の数値実験を行って年最大値資料に基づく極値分布ならびに確率波高の特性を解明することにした。

この研究は、先に取りまとめた数値実験の成果 (合田 [2010], 以下では前報と称する) を引き継ぐものであり、データ解析は Hosking (1990) と Hosking・Wallis

(1997)によるL-モーメント法を用いている。したがって、解析法の詳細については前報を参照されたい。

本報の全体構成は次の通りである。まず、2.において極大値資料から年最大値資料を導くときの前提となる高波の発生回数のポアソン分布について、酒田港およびアラス沿岸 Kodiak 島沖の波浪データを用いて吟味する。次に、極大値分布と年最大値分布の理論的關係について3.で説明する。今回の数値実験の条件や計算方法については4.で述べる。波浪条件としては50年確率波高を $H_{50} = 10.0\text{ m}$ に設定し、裾長さパラメータは $\gamma_{50} = 1.2$ に固定する。極大値分布の母集団としては、一般化極値分布(GEV)、一般化パレート分布(GPA)およびワイブル分布の3種類を用い、高波の発生数が年平均3および10回を想定して毎年の最大波高を抽出する。

数値実験では期間が十分に長いデータとして期間長が $K = 10,000$ 年の年最大値を扱った場合を5.で記述する。現実ではあり得ないけれども、このような超長期間にわたる年最大値資料であれば、母集団がいずれであっても一般化極値分布で近似できることを例示する。しかし、期間長が数十年程度の実際の資料であれば、標本サイズの影響によって一般化極値分布への収束が保証できないことを6.で説明し、こうした小標本の年最大値資料の統計的性質の幾つかについて紹介する。確率波高の偏りについては7.で記述し、年最大値資料から推定される確率波高の変動係数について導いた経験的推定式を8.で提示する。さらに、9.で幾つかの現地観測波浪データを用いて、本研究の適用妥当性を検討する。最後に10.において全体のまとめを述べる。

## 2. 高波の発生回数のポアソン分布

統計の分野では、ポアソン分布という確率分布がしばしば用いられる。これは、ある事象が一定期間内に発生する平均回数が分かっている場合、その期間内にその事象が発生する回数 $m$ がどのような確率で分布するかを表すものである。高波の場合であれば、年間の平均発生回数( $\lambda$ で表示)が対象である。

一定期間内に事象が $m$ 回発生する確率 $P(m)$ は次のように表される。

$$P(m) = \frac{1}{m!} e^{-\lambda} \lambda^m \quad : \quad m = 0, 1, 2, \dots \quad (1)$$

ある期間内に1回も発生しない確率は $P(0) = e^{-\lambda}$ であり、 $\lambda = 1, 2, 3$ に応じてそれぞれ $P(0) \approx 0.368, 0.135, 0.050$ となる。

高波の発生回数がポアソン分布で近似できるか否かについて、わが国の港湾関連部局の全国沿岸波浪ネッ

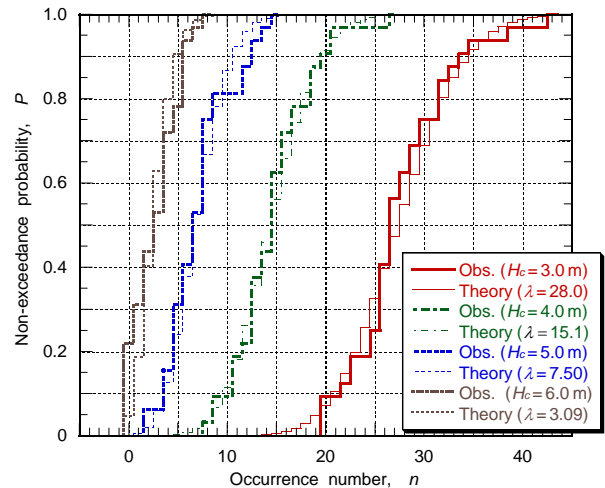


図-1 酒田港の高波の発生回数のポアソン分布との比較

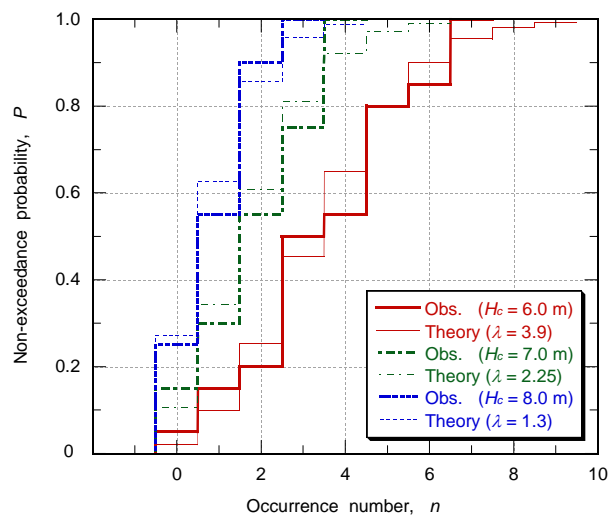


図-2 Kodiak 島沖の高波の発生回数のポアソン分布との比較

トワーク (NOWPHAS, Nationwide Ocean Wave information network for Port and HarbourS) のうちの酒田港の高波資料ならびにアラス沿岸 Kodiak 島沖の高波の追算資料を用いて検討した。酒田港の資料は1970～2008年の39年間の観測データであり、有効統計年数34.4年の間に有義波高3.0 m以上の波高極値を955回記録している。一方、Kodiak 島沖の波浪追算は1956～1975年の20年間を対象として米国陸軍工兵隊海岸工学研究センターが実施したものであり、各年の有義波高極大値の時系列は合田(2008, p.329)に採録されている。

酒田港の資料について高波極大値選定の閾値を $H_c = 3.0, 4.0, 5.0$ , および $6.0\text{ m}$ と高めていくと、極大値の個数が955, 518, 258, および106と減少する。それぞれの波高閾値を超える高波の個数を年ごとに数え、発生回数の累積度数分布を作成して非超過確率を計算した。た

だし、年によって欠測期間の長さが異なるため、年間の発生回数は欠測期間を考慮して補正している。この補正された高波極値の平均発生率は、4通りの閾値に対して $\lambda=28.03, 15.06, 7.50$ および $3.09$ 回/年と算定される。そこで、式(1)のポアソン分布の確率を求めて非超過確率を算定し、実測値と比較したのが図-1である。

図中の凡例でTheoryと記したのはポアソン分布の計算値を表している。ポアソン分布は離散関数であるところからこのような非超過確率の形で比較したものである。実測値と理論値の両者とも累積回数は整数であり、整数値の $\pm 0.5$ に広がる水平線としてデータを表示している。たとえば、閾値が $H_c=4.0$ mのときには年間発生数が8個の年が1回、9個の年が2回あり、年間発生数が10個の年はなかった。したがって、発生回数9と10の累積度数が3、非超過確率が0.094であり、横軸の9~10の個所にこの確率の相当する水平線が描かれている。実測値と理論値の非超過確率に若干のずれはあるものの、全体として両者は合致しているといえる。

図-2はKodiak島沖の高波のデータについて発生回数の非超過確率を比較したものである。波浪追算資料が波高閾値を $H_c=6.0$ mとして提示されたものであり、20年間で78個の高波が報告されていて、そのうち1年は無発生である。平均発生率は $\lambda=3.90$ 回/年である。閾値を7.0および8.0mに高めると、平均発生率が2.25および1.3回/年に低下する。酒田港に比べて高波の総数が少ないために実測回数と理論予測回数との間の差異がやや大きく見えるけれども、全体としてはほぼ一致しているといえる。

この2地点の事例から見ても、高波の発生回数をポアソン分布で近似することには妥当と考えられる。

### 3. 極大値と年最大値の分布関数の相互関係

#### (1) 極大値と年最大値の一般的関係

いま1年間に発生する高波の極大値を発生順に $H_1, H_2, \dots, H_m$ と表示すると、その年の最大波高 $H_{\max}$ は次のように定義される。

$$H_{\max} = \max\{H_1, H_2, \dots, H_m\} \quad (2)$$

ここで高波の極大値の分布関数すなわち非超過確率を $G(x)$ で表す。1年間に $m$ 個発生した高波の極大値のうちの最大波高 $H_{\max}$ がある値 $x$ を超えない事象は、 $H_1, H_2, \dots, H_m$ のいずれもが $x$ を超えない事象であり、その確率は $[G(x)]^m$ で表される。極大値の発生個数 $m$ は $0 \sim \infty$ のうちの任意の値を取り得る。高波の発生個数がポアソン分布に従うとすると、それぞれの生起確率は

式(1)で表されるので、年最大値の非超過確率は次のように求められる。

$$F(x) = \Pr[H_{\max} < x] = \sum_{m=0}^{\infty} P(m) G^m(x) \quad (3)$$

$$= \sum_{m=0}^{\infty} \frac{1}{m!} e^{-\lambda} [\lambda G(m)]^m = \exp\{-\lambda[1-G(x)]\}$$

ここで $\Pr[X]$ は事象 $X$ の起こる確率を表す。なお、式(3)の第2行の演算には指数関数のべき級数展開を利用している。

著者が既に紹介しているように(合田 [2008], p. 335)、 $R$ 年確率統計量は非超過確率 $1-1/R$ に対応する値として定義される。すなわち、

$$F(x_R) = 1 - \frac{1}{R} \quad (4)$$

この非超過確率を式(3)と等値すると、極大値分布の非超過確率が次のように表される。

$$G(x_R^*) = 1 + \frac{1}{\lambda} \ln\left(1 - \frac{1}{R}\right) \quad (5)$$

ここに、 $x_R^*$ は年最大値としての $R$ 年確率統計量 $x_R$ に対応する、極大値としての値である。式(5)はさらに $R$ が1よりも十分に大きい場合について、自然対数のべき級数展開を使って近似計算することによって、次のように書き直される。<sup>1</sup>

$$G(x_R^*) \cong 1 - \frac{1}{\lambda(R-1/2)} \quad (6)$$

すなわち、極大値としての $R$ 年確率統計量 $x_R^*$ は年最大値よりも1/2年だけ短い再現期間に対応する値である。

#### (2) 極大値がワイブル分布に従う場合の年最大値分布

ワイブル分布の分布関数は次のように表される。

$$G(x) = 1 - \exp\left[-\left(\frac{x-B}{A}\right)^k\right] \quad : \quad B < x < \infty \quad (7)$$

ここに、 $k$ は形状母数、 $A$ は尺度母数、 $B$ は位置母数である。式(7)を式(3)に代入すると、極大値がワイブル分布に従う母集団から抽出された年最大値の分布関数

<sup>1</sup>

$$\ln(1-1/R) \cong -\frac{1}{R} - \frac{1}{2R^2} = -\frac{1}{R} \left(1 + \frac{1}{2R}\right)$$

$$= -\frac{1}{R} \cdot \frac{1}{1/(1+1/2R)} \cong -\frac{1}{R} \cdot \frac{1}{1-1/2R} = -\frac{1}{R-1/2}$$

が次のように導かれる．

$$F(x) = \exp\left\{-\lambda \exp\left[-\left(\frac{x-B}{A}\right)^k\right]\right\} \quad (8)$$

$$= \exp\left\{-\exp\left[-\left(\frac{x-B}{A}\right)^k + \ln \lambda\right]\right\}$$

ワイブル分布の中で形状母数が $k=1$ のときは指数分布を表す．極大値が指数分布の場合には，年最大値は次のガンベル分布に従うことになる．

$$F(x) = \exp\left\{-\exp\left[-\frac{x-B^*}{A}\right]\right\} \quad (9)$$

ここに，

$$B^* = B + A \ln \lambda \quad (10)$$

すなわち，尺度母数は極大値分布と同じ値を保ち，位置母数が母集団の値よりも $A \ln \lambda$ だけ大きな値を取ることになる．

式(9),(10)は先にCunnane(1973)が提示したものであり，Cunnaneは指数分布と式(9)のガンベル分布の確率波高推定値の分散の大きさを比較している．ただし，指数関数で表される母分布関数から抽出される個別の標本について吟味しているわけではない．

### (3) 極大値が一般化パレート分布に従う場合の年最大値分布

一般化パレート分布(GPA)の関数形は次のようなものである．

$$G(x) = \begin{cases} 1 - \left[1 - k \frac{x-B}{A}\right]^{1/k} & : k \neq 0 \\ 1 - \exp\left[-\frac{x-B}{A}\right] & : k = 0 \end{cases} \quad (11)$$

このうち $k=0$ は指数分布を表し，このときの年最大値分布は式(9)となる．

$k \neq 0$ の場合の年最大値分布は次のように導かれる．

$$F(x) = \exp\left\{-\left[\lambda^k \left(1 - k \frac{x-B}{A}\right)\right]^{1/k}\right\} \quad (12)$$

Madsen et al. (1997)はこの式をさらに書き換え，年最大値分布が次のような一般化極値分布となることを示している．

$$F(x) = \exp\left\{-\left[1 - k \frac{x-B^*}{A^*}\right]^{1/k}\right\} : k \neq 0 \quad (13)$$

上式中の尺度母数 $A^*$ と位置母数 $B^*$ は一般化パレート分布の母数から次のように算定される．

$$\begin{cases} A^* = A \lambda^{-k} \\ B^* = B + A(1 - \lambda^{-k})/A \end{cases} \quad (14)$$

Madsen et al. は数値シミュレーション手法を用い，式(11)の分布に従う極大値資料と式(13)による年最大値資料のそれぞれから推定される確率統計量の分散の大きさを比較検討している．ただし，あくまでも式(11),(13)の分布関数に基づく検討であって，一般化パレート分布から年最大値の標本を作成してそれを統計解析する作業は行っていない．

## 4. 数値計算の条件と計算方法

### (1) 発生させる高波の特性

数値シミュレーションによって高波の標本を多数作成する際には，高波として現地の波浪特性を反映させたものであることが必要である．高波の極値分布の特性として重要なのは，合田(2002)が提唱した裾長さパラメータである．これは10年確率波高 $H_{10}$ に対する50年確率波高 $H_{50}$ の比として次のように定義される．

$$\gamma_{50} = H_{50}/H_{10} \quad (15)$$

わが国沿岸の高波の極値統計解析の結果では，裾長さパラメータ $\gamma_{50}$ が1.13～1.27の値をとっている(合田2008, 375頁)．前報では裾長さパラメータとして $\gamma_{50} = 1.10, 1.20$ , および1.30の3通りの値を用いたけれども，本研究では $\gamma_{50} = 1.20$ に固定した．その代わりに，後述のように平均発生率として $\lambda = 3$ および10回/年の2通りの値を用いている．

また，50年確率波高としては $H_{50} = 10.0\text{m}$ に設定する．これは数値の換算が容易なように区切りの良い数値としたものである．

裾長さパラメータをある値に設定すると，分布関数の尺度母数と位置母数は任意の値を与えることができず，次式によって一義的に決定される値を使わなければならない．

$$\begin{cases} A = H_{50} \frac{1-1/\gamma_{50}}{y_{50} - y_{10}} \\ B = H_{50} \left[1 - \frac{1-1/\gamma_{50}}{1-y_{10}/y_{50}}\right] \end{cases} \quad (16)$$

ここに， $y_{10}$ と $y_{50}$ は再現期間 $R = 10$ 年と50年に対する基準化変量である．

本研究では、極大値分布として一般化極値分布 (generalized extreme-value: GEV)、一般化パレート (generalized Pareto: GPA) 分布、およびワイブル分布の3種類の極値分布関数を用いる。これらに対する基準化変量は次式で計算される。

1) 一般化極値分布 (GEV) :

$$y_R = \begin{cases} \{1 - [-\ln(1 - 1/\lambda R)]^k\} / k, & k \neq 0 \\ -\ln(-\ln(1 - 1/\lambda R)), & k = 0 \end{cases} \quad (17)$$

2) 一般化パレート分布 (GPA) :

$$y_R = \begin{cases} [1 - 1/(\lambda R)^k] / k, & k \neq 0 \\ \ln(\lambda R), & k = 0 \end{cases} \quad (18)$$

3) ワイブル分布 :

$$y_R = \{\ln(\lambda R)\}^{1/k} \quad (19)$$

(2) 年最大値の母集団である極大値の分布関数

各種の極値分布は、 $L$ -モーメント比  $\tau_3$  ( $L$ -skewness) と  $\tau_4$  ( $L$ -kurtosis) の相関関係によってその特性を表すことができる。図-3は、GEV、GPA、およびワイブル分布の特性曲線を示したもので、図中の、の記号はそれぞれの分布関数において形状母数  $k$  が特定の値をとる場合を示している。これらの形状母数は前報

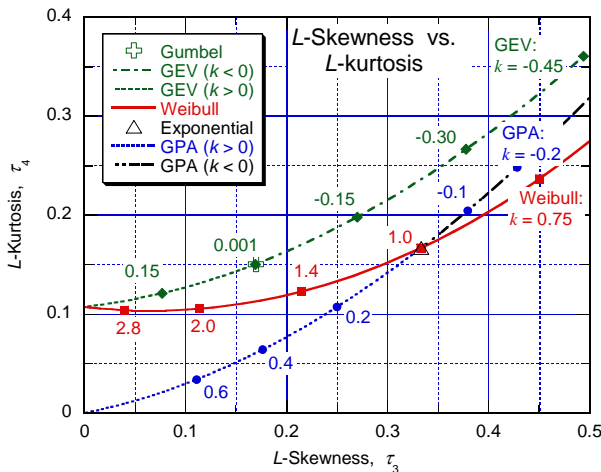


図-3 高波の母分布関数の  $L$ -kurtosis と  $L$ -skewness の関係で用いたものである。

本研究では、年最大波高の標本を抽出する母分布関数として GEV 分布、GPA 分布、およびワイブル分布の3種類を取り上げている。それぞれの分布の形状母数は表-1 に示すような 4 ~ 5 通りのものを標準として用いたが、ときに応じてこれ以外の形状母数の分布についても数値シミュレーションを行っている。

なお、GEV 分布の  $k = 0.001$  はガンベル分布を近似的に表現したものである。また、ワイブル分布の  $k = 1.0$  は指数分布であり、GPA 分布の  $k = 0$  と共通である。ただし、確認のために GPA 分布の  $k = 0.001$  のケースも計算に加えた。

表-1 高波の発生に使用する分布関数

分布関数	形状母数 $k$				
一般化極値 (GEV)	-0.45	-0.3	-0.15	0.001	0.15
一般化パレート (GPA)	-0.2	-0.1	0.001	0.10	0.15
ワイブル	0.75	1.0	1.4	2.0	-

(3) 母分布関数の尺度・位置母数および統計量

既に (1) で述べたように、母分布関数が与えられ、裾長さパラメータが設定されたときには、その尺度母数  $A$  と位置母数  $B$  が式 (16) によって一義的に決定される。今回の数値計算では  $H_{50} = 10.0$  m とし、平均発生率を 2 通りに変えているので、この条件の下で表-1 の母分布関数の尺度・位置母数の値を計算したところ、表-2 ~ 4 の母数値が得られた。また、分布関数の平均値  $\mu[x]$ 、標準偏差  $\sigma[x]$ 、歪み度  $\beta_1$  ならびに尖鋭度  $\beta_2$  を求めた結果も併せて示している。理論値が計算できるものはそれを記入したが、歪み度と尖鋭度については極大値の全データの集計結果から計算した値である。

なお、これらの統計量の定義は次の通りである。また、尖鋭度を除く統計量の理論値は表-5 に提示した。

$$\mu[\xi] = \frac{1}{N} \sum_{i=1}^N \xi_i \quad (20)$$

$$\sigma[\xi] = \left[ \frac{1}{N} \sum_{i=1}^N (\xi_i - \mu[\xi])^2 \right]^{1/2} \quad (21)$$

$$\beta_1 = \frac{1}{\sigma^3} \left[ \frac{1}{N} \sum_{i=1}^N (\xi_i - \mu[\xi])^3 \right] \quad (22)$$

$$\beta_2 = \frac{1}{\sigma^4} \left[ \frac{1}{N} \sum_{i=1}^N (\xi_i - \mu[\xi])^4 \right] \quad (23)$$

表-2 は極大値の母集団が GPA 分布の場合である。このうち、形状母数が -0.45 のケースは分布の裾が非常に長く広がるために、データ個数が 5 千万個と多くなっても尖鋭度を安定して計算できない。表中で \*\*\* の記号を付したのは、信頼できる数値が得られなかったケースである。形状母数が -0.30 のケースでも尖鋭度については信頼度が低い。

さらに、GEV を母分布関数として抽出された年最大値の標本データについてその平均値等も記入してある。各項で POT の列が母分布関数の統計値、AM の列が年最大値の統計値である。表から分かるように、年最大

表-2 GEV分布を母分布関数とする極大値ならびに年最大値の母数値と統計量( $H_{50} = 10.0$  m, 裾長さ  $\gamma_{50} = 1.2$ )

$\lambda$	$k$	L-skew ( $\tau_3$ )	A (m)		B (m)		$\mu(x)$ (m)		$\sigma(x)$ (m)		$\beta_1$		$\beta_2$	
			POT	AM	POT	AM	POT	AM	POT	AM	POT	AM	POT	AM
			3	-0.45	0.4940	0.1520	0.2723	7.122	7.300	7.330	7.653	0.8875	1.324	207.1
-0.30	0.3779	0.2888		0.4517	6.639	6.951	6.926	7.366	0.7029	1.008	10.41	9.393	843.6	587.0
-0.15	0.2700	0.5459		0.7476	5.926	6.473	6.336	6.978	0.8947	1.118	2.530	2.217	19.27	16.38
0.001	0.1693	1.0312		1.2373	4.849	5.810	5.444	6.428	1.321	1.453	1.133	0.824	5.366	4.580
0.15	0.0771	1.922		2.025	3.232	4.904	4.090	5.645	2.102	2.060	0.436	0.074	3.140	3.047
10	-0.45	0.4940	0.0887	0.2586	6.968	7.312	7.089	7.657	0.4588	1.2495	160.6	78.1	***	***
	-0.30	0.3779	0.2020	0.4219	6.330	6.976	6.531	7.381	0.4913	0.9879	9.708	8.570	650.8	363.1
	-0.15	0.2700	0.4579	0.8240	5.330	6.580	5.643	7.010	0.7493	1.0752	2.503	2.368	18.97	18.14
	0.001	0.1693	1.039	1.114	3.567	5.891	4.165	6.496	1.330	1.375	1.133	1.004	5.370	5.068
	0.15	0.0771	2.319	1.789	0.628	5.041	1.663	5.777	2.536	1.896	0.436	0.270	3.127	3.155

表-3 GPA分布を母分布関数とする極大値ならびに年最大値の母数値と統計量( $H_{50} = 10.0$  m, 裾長さ  $\gamma_{50} = 1.2$ )

$\lambda$	$k$	L-skew ( $\tau_3$ )	A (m)		B (m)		$\mu(x)$ (m)		$\sigma(x)$ (m)		$\beta_1$		$\beta_2$	
			POT	AM	POT	AM	POT	AM	POT	AM	POT	AM	POT	AM
			3	-0.20	0.4286	0.4446	0.5122	6.167	6.751	6.723	7.202	0.7175	0.9998	4.595
-0.10	0.3793	0.6793		0.6924	5.582	6.421	6.336	6.937	0.8438	1.104	2.812	2.068	17.77	11.74
0.001	0.3329	1.0399		0.9368	4.802	6.012	5.841	6.609	1.0378	1.281	1.995	1.311	8.956	5.787
0.10	0.2903	1.5753		1.2572	3.792	5.521	5.224	6.210	1.3073	1.5316	1.517	0.845	5.778	3.748
0.15	0.2698	1.9414		1.4574	3.161	5.233	4.850	5.974	1.4806	1.692	1.337	0.662	4.868	3.207
10	-0.2	0.4286	0.3495	0.5536	5.692	6.173	6.128	7.168	0.5639	1.0123	4.648	3.467	63.1	38.1
	-0.1	0.3793	0.6022	0.7578	4.811	6.370	5.480	6.890	0.7481	1.131	2.811	1.909	17.77	10.89
	0.001	0.3329	1.0412	1.0383	3.550	5.943	4.590	6.542	1.039	1.330	1.994	1.135	8.953	5.372
	0.10	0.2903	1.7769	1.4107	1.776	5.429	3.391	6.116	1.475	1.615	1.517	0.639	5.776	3.571
	0.15	0.2698	2.3256	1.6456	0.600	5.126	2.622	5.862	1.774	1.800	1.337	0.438	4.867	3.137

表-4 ワイブル分布を母分布関数とする極大値ならびに年最大値の母数値と統計量( $H_{50} = 10.0$  m, 裾長さ  $\gamma_{50} = 1.2$ )

$\lambda$	$k$	L-skew ( $\tau_3$ )	A (m)		B (m)		$\mu(x)$ (m)		$\sigma(x)$ (m)		$\beta_1$		$\beta_2$	
			POT	AM	POT	AM	POT	AM	POT	AM	POT	AM	POT	AM
			3	0.75	0.4505	0.4818	0.6010	5.869	6.481	6.443	6.982	0.7761	1.0515	3.121
1.0	0.3333	1.0356		0.9341	4.811	6.016	5.847	6.611	1.0356	1.2784	2.000	1.315	9.001	5.819
1.4	0.2150	2.1805		1.3981	3.106	5.497	5.093	6.178	1.4384	1.6002	1.198	0.681	4.838	3.695
2.0	0.1140	4.2279		1.9221	0.536	5.015	4.283	5.747	1.9586	1.9779	0.631	0.194	3.244	3.029
10	0.75	0.4505	0.4428	0.7708	4.941	6.298	5.468	6.832	0.7132	1.1450	3.122	1.712	19.00	8.403
	1.0	0.3333	1.0356	1.139	3.564	6.547	4.600	6.547	1.0358	1.3284	2.000	1.141	9.001	5.406
	1.4	0.2150	2.3453	1.341	1.352	5.595	3.490	6.249	1.5474	1.5468	1.198	0.699	4.839	4.002

表-5 母集団の確率密度関数, 平均値, 標準偏差, および歪み度

分布関数	確率密度関数 $f(x)$	平均値 $\mu(x)$	標準偏差 $\sigma(x)$	歪み度 $\beta_1$
一般化極値 (GEV)	$\frac{[1 - k(x - B) / A]^{1/k-1}}{A} \times \exp\{-[1 - k(x - B) / A]^{1/k}\}$ : $k \neq 0$	$B + A \left[ \frac{1}{-\Gamma(1+k)} \right] / k$	$A \left[ \frac{\Gamma(1+2k)}{-\Gamma^2(1+k)} \right]^{1/2} / k$	$\frac{A^3}{\sigma^3} \left( \frac{E[X^3] - 3E[X^2]}{\times E[X] + 2E^3[X]} \right)$
一般化パ レート (GPA)	$[1 - k(x - B) / A]^{1/k-1} / A$ : $B \leq x, k \neq 0$	$B + A / (1+k)$	$\frac{A}{(1+k)(1+2k)^{1/2}}$	$\frac{2(1-k)(1+2k)^{1/2}}{(1+3k)}$
ワイブル	$k / A \times [(x - B) / A]^{k-1} \times \exp[-\{(x - B) / A\}^k]$ : $B < x < \infty$	$B + A \Gamma(1+1/k)$	$A \left[ \frac{\Gamma(1+2/k)}{-\Gamma^2(1+1/k)} \right]^{1/2}$	$\frac{A^3}{\sigma^3} \left( \frac{\Gamma(1+3/k) - 3\Gamma(1+2/k)}{\times \Gamma(1+1/k) + 2\Gamma^3(1+1/k)} \right)$

$$E[X] = [1 - \Gamma(1+k)] / k: \quad E[X^2] = [1 - 2\Gamma(1+k) + \Gamma(1+2k)] / k^2: \quad E[X^3] = [1 - 3\Gamma(1+k) + 3\Gamma(1+2k) - \Gamma(1+3k)] / k^3$$

値の標本の統計量は母分布の統計量からかなり変化する。これについては、まとめて5.で論述する。

表-3, 4は極大値の母集団がGEVおよびワイブル分布の場合である。記述項目は表-2と同じである。ワイブル分布で平均発生率が $\lambda = 10$ のときには、形状母数を $k = 2.0$ に設定すると位置母数 $B$ が負の値となる、すなわち波高の極大値が負となってしまう現実的ではない。このため、年最大値を抽出するためのワイブル分布の母集団としては、 $k=2.0$ のケースを計算しなかった。

なお、表-2～4中の尺度母数と位置母数のAMの列の値は、統計期間が $K=10,000$ 年の標本500組に対して当てはめた一般化極値(GEV)分布の母数の平均値である。5(2)で述べるように、平均発生率が $\lambda=3$ のときにはGEVへの適合率が0%であるけれども、参考のためにここに記載したものである。

### (3) 高波の平均発生率と年最大値の抽出方法

今回の数値実験では、平均発生率は $\lambda=3$ および10個/年の2通りに設定した。前者は台風等による高波、後者は冬季季節風による高波などを想定したものである。年最大値の抽出は次のようにして行った。

まず、平均発生率を入力として式(1)のポアソン分布に基づいて1年間の発生個数毎の出現確率を計算し、その累積分布確率をあらかじめ求めておく。確率を計算する発生個数の上限値は経験式で設定した。この累積分布確率は0から1まで段階的に増加する関数であり、その計算事例は既に図-1, 2に示したところである。

次に、ある年に対応してコンピュータで疑似一様乱数を発生させ、その乱数値がポアソン分布の累積分布確率を超える一つ前の発生個数を見つけ出す。これがその年に割り当てられる高波の極大値の発生個数である。このようにして年最大値の統計期間の各年について極大値の発生個数を割り当てる。

また、乱数の値によっては発生個数0のポアソン確率 $P(0)$ よりも小さいことが起きる。その年は閾値を超える高波が発生しなかったものであり、その年は極値統計解析の対象から除外する。たとえば、 $\lambda=3$ に対しては $P(0)=0.04976$ であるので、 $K=100$ 年のシミュレーションを行うと、平均して4.98年は高波が発生しないことになる。すなわち、有効統計期間は $K^*=95.024$ 年として統計解析を行う。

全体の極大値の数は、平均的には有効統計期間 $K^*$ 年に平均発生率を乗じた数であるが、乱数の出現の度合いによって $K$ 年の試行ごとに若干の変動がある。

各年の発生個数が設定されたところで、 $K$ 年間の高

波の総個数に相当する極大値をその母分布関数から抽出する。これは前報で用いたモンテカルロ法である。そして、極大値の時系列を各年の発生個数毎に区切り、その中での最大値を判定し、これを年最大値のデータとして採択する。

### (4) 年最大値の統計期間と標本の総数

年最大値の統計期間としては、 $K=10 \sim 10,000$ 年にわたる12種類を対象とした。発生させる極大値の個数は平均発生率を $\lambda$ として、 $n = \lambda K^*$ がその平均値である。ただし、前項に述べたように試行毎に若干の変動がある。 $K=10,000$ 年、 $\lambda=10$ 個/年の場合には平均で $n=99,995$ 個である。

年最大値の極値統計解析では、 $N=50,000$ 組の標本の作製を標準とした。ただし、統計年数が100年を超える場合には計算時間を短縮するため、 $N=5,000,000/K$ に設定した。すなわち、 $K=10,000$ 年に対する年最大値の標本は $N=500$ 個である。

### (5) 年最大値の標本の解析方法

所定の母集団から抽出された年最大値の標本は、まず個々のデータをすべて集積して波高区分0.25 mを級とする度数分布を作成して確率密度分布のデータとした。また、データの1～4乗値の値を集計し、年最大値の平均、標準偏差、歪み度、尖鋭度を計算した。極大値の統計量についても同様である。表-2～4に記載した統計量はこのようにして求めたものである。

年最大値の個別の標本についてはL-モーメント法を適用し、GEV, GPAおよびワイブル分布のそれぞれに当てはめて形状・尺度・位置の3母数を推計した。この際には標本のデータを昇順に並べ替え、式(24)のプロットングポジションの定数としては、前報で推奨された $\alpha=0.45, \beta=0$ の値を用いた。

$$P_i = \frac{i - \alpha}{n + \beta} \quad (24)$$

ここに、 $i$ は昇順に並べ替えられた極値データの順位、 $n$ は年最大値のデータの個数、すなわち平均的に $n = \lambda K^*$ である。

3種類の分布関数の母数値が推定されると、昇順の順位毎の確率統計量が計算される。その計算値と標本のデータとの間の相関係数を算定し、その相関係数をもっとも1に近い分布を最適関数と判定した。すなわち、相関係数の最小残差法を採用した。この方法は、“Probability Plot Correlation Method (PPC)”とも呼ばれて

いる (Chowdhury et al. 1991) .そして、 $N$ 組の標本について最適と判定された回数を分布関数毎に数え、各分布関数の適合率を計算した。

また、年最大値の標本に当てはめられた分布関数の母数に基づいて、再現期間  $R = 1.01 \sim 1000$  年に対する確率波高を推計し、母分布関数の  $R$  年確率波高との差を調べるとともに、確率波高の変動係数を算定した。

### (6) 推計値の偏り率 (Bias) の定義

本研究では、年最大値の標本に基づく各種の推計値が標本の抽出母体である極大値の母集団からどれだけ偏っているかを調べるのが主目的である。このため、偏りおよび偏り率の算定では、極大値分布の母集団の値を偏り判定の基準の真値として扱う。

偏り量の検討に際しては、まず標本ごとの各種統計量の推計値  $\xi_i$  について、その平均  $\mu[\xi]$  ならびに標準偏差  $\sigma[\xi]$  を式 (20), (21) で計算する。その際の標本個数  $N$  は、本研究では  $N = \max\{50,000, 5,000,000/K\}$  としている。

偏りの度合いは絶対値ならびに相対比の二つの方法で定義することができる。すなわち、

$$\text{Absolute Bias} = \mu[\xi] - \xi_0 \quad (25)$$

$$\text{Relative Bias} = (\mu[\xi] - \xi_0) / \xi_0 \quad (26)$$

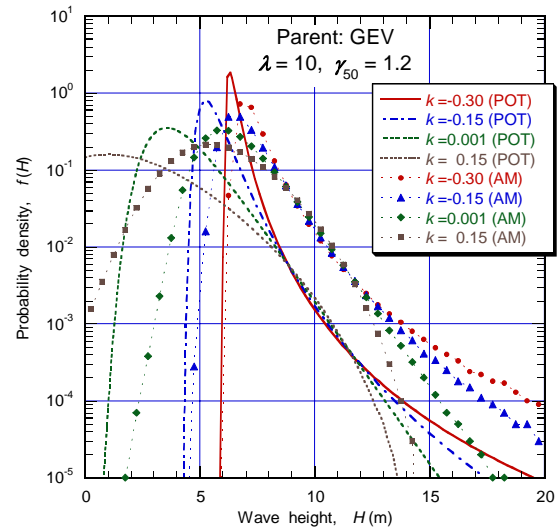
ここに、 $\xi_0$  は極大値の母集団の値である。式 (25) の偏りはここでは偏り量と呼び、これは、分布関数の形状母数などその値が 0 となる可能性のあるものに対して用いる。式 (26) の偏りの相対比はここでは偏り率と呼ぶことにする。なお、表示に際しては実数値ではなく、百分率で表すこともある。

## 5. 年最大値の統計値の特性

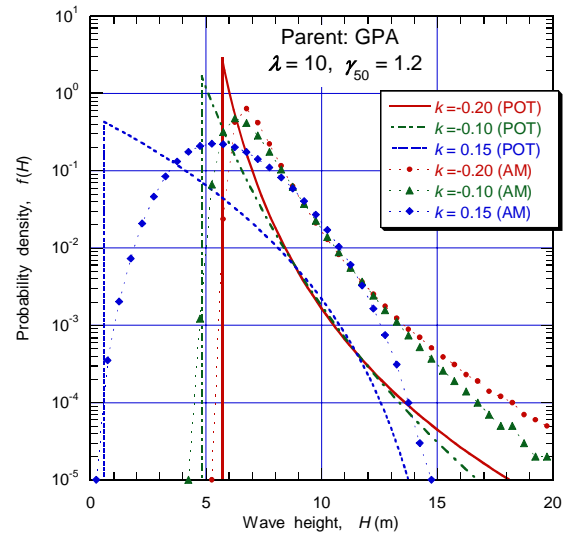
### (1) 極大値および年最大値の出現確率分布

最初に、母集団である極大値とそれから抽出された年最大値のデータの出現確率分布を図-4に示す。極大値については、表-5の確率密度関数で計算される値を示し、年最大値については  $K = 10,000$  年の標本 500 組のデータについて級幅  $\Delta H = 0.25$  m で集計した度数分布から換算したものである。データは形状母数毎に分けて表示しており、極大値のデータは POT、年最大値のデータは AM の凡例で区別している。

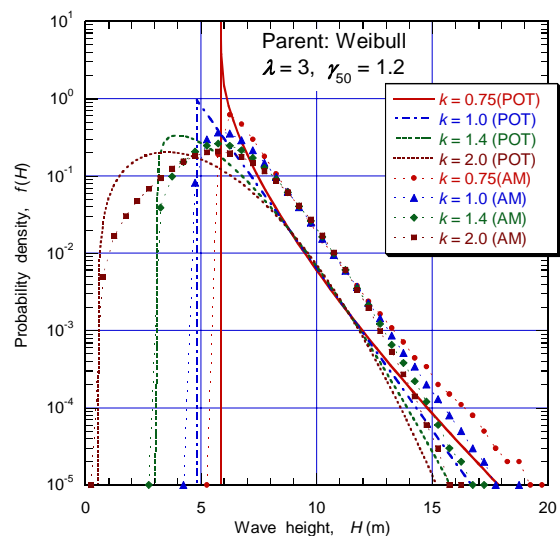
図-4で明らかなように、年最大値のデータは全体として波高の大きなほうへシフトし、ピークの尖り具合も緩和されている。このため、年最大値の平均値は極



(a) 母集団が一般化極値分布の場合



(b) 母集団が一般化パレート分布の場合



(c) 母集団がワイブル分布の場合

図-4 母集団ならびに年最大値のデータの出現確率分布



大値の平均よりも常に大きくなり、歪み度は逆に小さくなる。表-2~4には、こうした統計値が記載されているので、POTとAMの列の数値を比較することで、そうした変化を吟味できる。標準偏差は、形状母数が小さいときは年最大値のほうが極大値よりもかなり大きな値であるけれども、形状母数が大きくなるにつれてその差が縮まる。母集団が一般化極値分布の場合には、 $k > 0$ で年最大値の標準偏差が母集団よりも小さくなっている。

## (2) 年最大値データの一般化極値分布への適合度

年最大値の標本が十分に大きな場合、すなわち非常に長い期間にわたって年最大値のデータが収集された場合には、年最大値のデータが一般化極値分布に従うことが今回の数値実験でも確かめられた。すなわち、 $K=10,000$ 年の標本500組について3種類の分布関数の適合度を調べたところ、母分布関数がどれであっても一般化極値分布への適合率が100%であった。ただし、平均発生率が低いときには一般化パレート分布への適合率が高く、平均発生率が高くなるとワイブル分布へよく適合し、一般化極値分布が最適分布となるのは平均発生率が5程度以上の場合に限られる。

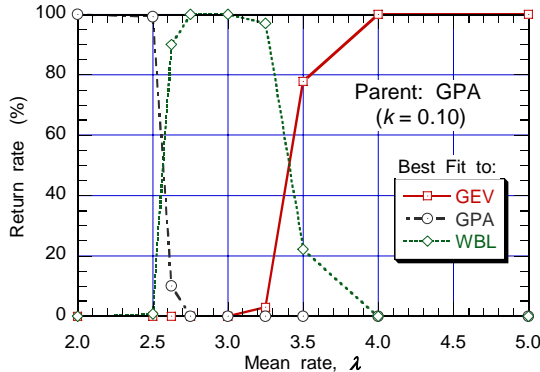


図-5 極値分布関数の適合率に及ぼす平均発生率の影響

図-5は、母集団が一般化パレート分布で形状母数が $k=0.10$ の場合について、平均発生率を $\lambda=2 \sim 5$ に漸変させてみた結果である。 $\lambda \leq 2$ では年最大値が一般化パレート分布によく適合し、 $2.6 \leq \lambda \leq 3.3$ ではワイブル分布の適合率が高く、 $\lambda=4$ になってはじめて一般化極値分布の適合率が100%となる。

母集団が指数分布(ワイブルで $k=1.0$ )のときには、年最大値への最適分布が一般化パレート分布からワイブル分布へ移行するのが $\lambda=2.75$ であり、さらに一般化極値分布へ移行するのが $\lambda \geq 3.25$ である。すなわち、母集団が一般化パレート分布のときよりも平均発生率が

やや低い段階で年最大値の標本が一般化極値分布に適合する。母集団がガンベル分布のときには、一般化極値分布への移行が $\lambda=3.1$ で完了する。

## (3) 年最大値に対する一般化極値分布の形状母数

母集団が一般化パレート分布である極大値から抽出された年最大値のデータは、標本が十分に大きければ式(13)で表されるような一般化極値分布に従い、形状母数は母集団と同一であり、尺度母数と位置母数は式(14)に従って変化する。

今回の数値実験の結果から、平均発生率 $\lambda=10$ 、統計期間 $K=10,000$ 年のデータについて母集団と年最大値に当てはめた一般化極値分布の形状母数を比較したのが図-6である。形状母数は表-1に記載したものに加えてその他のケースについても計算してある。なお、図の横軸にはL-モーメント法で計算されるL-skewnessの値を用いており、分布関数毎のこの値は表-2~4に記載されている。

母集団が一般化パレート分布の場合には、予測通りに一般化極値分布の形状母数は母集団の値と同じである。母集団が一般化極値分布であっても年最大値のデータは一般化極値分布で近似され、そのときの形状母数は母集団よりもやや大きくなる。図-6(a)でのデータに直線回帰式を当てはめると、次のような関係が得られる。

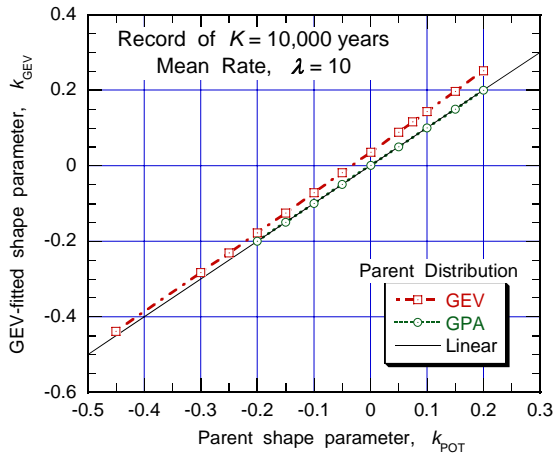
$$k_{AM} = 1.064k_{POT} + 0.036 \quad (27a)$$

一方、母集団がワイブル分布の場合には年最大値分布が式(8)のように導かれており、一般化極値分布に収束することが証明されているわけではない。しかし、数値実験の結果では統計期間が十分に長いときには一般化極値分布が最適分布となっている。その際に推計された形状母数 $k_{AM}$ は母集団の形状母数 $k_{POT}$ と図-6(b)に示すような関係になり、これを2次多項式で近似すると次のように表される。

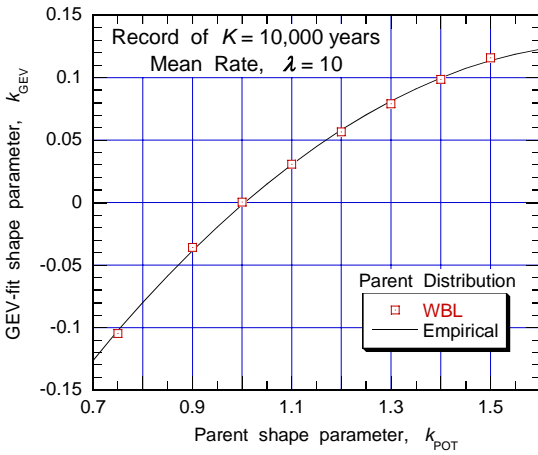
$$k_{AM} = -0.230k_{POT}^2 + 0.806k_{POT} - 0.578 \quad (27b)$$

## (3) 年最大値に対する一般化極値分布の尺度母数と位置母数

母集団が一般化パレート分布である極大値から抽出された年最大値に当てはめられた一般化極値分布の尺度母数は、式(14)によって母集団の尺度母数に関係づけられる。図-7はこの関係を確認するもので、一般化パレート分布のデータは式(14)の理論値に一致している。また、ワイブル分布については理論式がないもの



(a) 母集団が一般化極値分布および一般化パレート分布の場合



(b) 母集団がワイブル分布の場合

図-6 母集団と年最大値に対する一般化極値分布の形状母数の関係

の、結果として式(14)の値でほぼ近似される。

一方、母集団が一般化極値分布の場合には、年最大値に当てはめられた一般化極値分布の尺度母数がかなり大きな値に変化する。この尺度母数の変化率を経験式で近似すると次のように表される。

$$A_{AM}/A_{POT} = 3.668k_{POT}^2 - 2.402k_{POT} + 1.058 \quad (28)$$

また位置母数に関しては、母集団が一般化パレート分布のときは式(14)で変換式が与えられており、年最大値の数値実験の結果でもこの式で変換したとおりになることが確認されている。ただし、平均発生率が $\lambda=3$ のときには図-5に示したように年最大値が一般化極値分布に適合しないため、式(14)が適用できない。また、ワイブル分布に含まれる指数分布を母集団とする年最大値は、式(9)のようにガンベル分布に収束し、そのときの位置母数 $B^*$ は指数分布の位置母数 $B$ と式

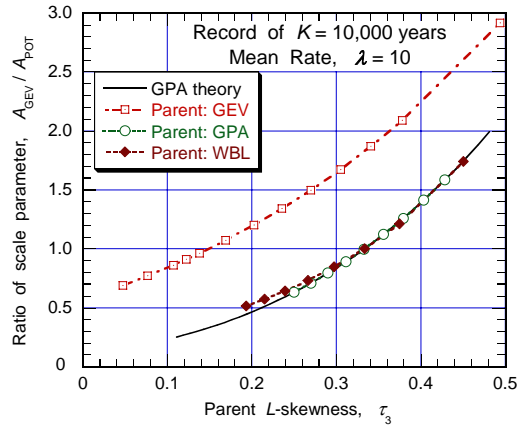


図-7 年最大値に当てはめられた一般化極値分布の尺度母数と母集団の尺度母数の比

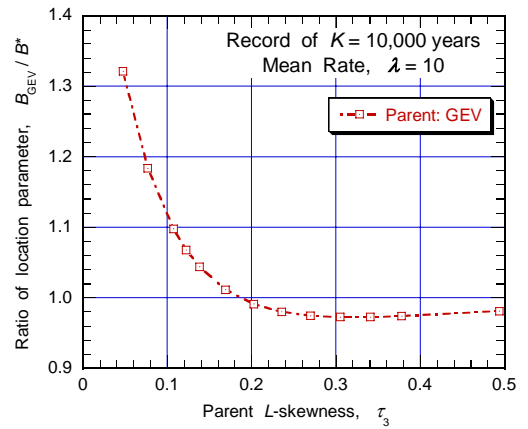


図-8 母集団が一般化極値分布の場合の年最大値分布と極大値分布の位置母数の比

(10)に示したような関係にある。数値実験の結果を調べると、ワイブル分布を母集団とする年最大値に当てはめられた一般化極値分布の位置母数についても、理由は不明であるものの、式(10)の関係が成立することが確認された。

母集団が一般化極値分布の場合には、式(10)、(14)のような変換式が見いだせない。数値実験で当てはめられた位置母数 $B_{AM}$ を式(10)による変換値 $B^*$ に対する比の形で表示したのが図-8である。

#### (4) 大標本の年最大値データによる確率波高推計値の偏り率

前項では、年最大値の標本が十分に大きくて平均発生率が高ければ、母集団の分布特性に関わりなく年最大値のデータが一般化極値分布に適合することを紹介した。また、平均発生率が低いときには図-5のように、大標本の年最大値データが一般化パレート分布あるいはワイブル分布に適合することを示した。しかしなが

ら、これは当てはめ候補として一般化極値分布、一般化パレート分布、およびワイブル分布の3種類に限定したためであり、適合率が100%であっても当てはめられた分布関数から推計される確率波高が母集団の値に一致するわけではない。

例外は母集団が一般化パレート分布の場合であって、このときには式(13)のように年最大値データが一般化極値分布となる。今回の数値シミュレーション結果を調べると、当てはめられた一般化極値分布から推計される再現期間 $R=20\sim 1000$ 年に対する確率波高の値は、母集団である一般化パレート分布の値からの偏り率が $-0.3\%\sim +0.04\%$ 異なるだけである。

母集団がワイブル分布の場合には、形状母数が $k=1.0$ 、すなわち指数分布の場合だけ年最大値データが一般化極値分布となる。これは式(9)、(10)で示したところであり、確率波高の偏り率もほぼ0である。しかし、形状母数が $k<1$ のときには当てはめられた一般化極値分布から推計した確率波高がやや過大評価となる。母集団の形状母数が $k=0.75$ であれば、1000年確率波高の推計値が $+4.6\%$ の偏り率を持つ。逆に形状母数が $k>1$ であれば確率波高が過小評価となる。母集団の形状母数が $k=1.5$ であれば、1000年確率波高の推計値が $-4.4\%$ の偏り率を持つ。

一般化極値分布を母集団として持つ極大値データから年最大値を抽出した場合には、そのデータに当てはめられた一般化極値分布から推計した確率波高が負の偏り率を示す。母集団の形状母数が $k=-0.45$ および $0.20$ のときの1000年確率波高の偏り率がそれぞれ $-1.2\%$ と $-3.5\%$ である。100年確率波高であればそれぞれ $-0.2\%$ と $-1.6\%$ である。

図-5に示したのは母集団が一般化パレート分布( $k=1.0$ )の場合である。統計期間が1万年という大標本であっても平均発生率が $\lambda=2.5$ のときには一般化パレート分布が最適となる。このときに当てはめられて一般化パレート分布から推計された100年および1000年確率波高はそれぞれ $-5.8\%$ と $-14.8\%$ の偏り率をもつ。平均発生率が $\lambda=3.0$ のときにはワイブル分布が最適となり、この場合の100年と1000年確率波高の偏り率 $-1.2\%$ と小さい。

図には示していないが、母集団が一般化極値分布やワイブル分布の場合も状況は同様である。いずれも平均発生率が低い場合には一般化パレート分布またはワイブル分布が最適となる。一般化パレート分布が当てはめられたときには、推計された確率波高の偏り率が最大で $-22\%$ と大きなものとなる。ワイブル分布が当てはめられたときも負の偏り率を生じるけれども、

100年と1000年確率波高の偏り率はそれぞれ $-5\%$ と $-9\%$ とやや小さくなる。

実際の極値統計解析では統計期間が1万年のような大標本の年最大値データが取得できないので、確率波高の偏り率については7.を参照して頂きたい。

## 6. 年最大値の統計分布に及ぼす標本サイズの影響

### (1) 年最大値の統計期間による極値分布適合率の変化

前節においては年最大値の標本が十分に大きい場合として、年最大値が $K=10,000$ 年にわたって収集できたという仮想的なケースを対象として議論をしてきた。しかし、自然現象を対象とする実際の極値統計解析ではそのような長期間のデータを取得することは不可能であり、数十年程度のデータを用いて極値統計解析を行い、100年確率事象などの予測を行うのが現実である。

一方、数値実験では統計期間の長さを変える、すなわち標本の大きさを変えることには何も差し支えない。そこで、統計期間を $K=10\sim 10,000$ 年の範囲で12段階に変え、3種類の極値分布関数への適合率を調べてみた。図-9は形状母数が $k=1.0$ のワイブル分布(指数分布)で平均発生率が $\lambda=10$ の母集団から年最大値を抽出した場合の計算結果である。一般化極値分布(GEV)、一般化パレート分布(GPA)、およびワイブル分布(WBL)への最適判定は、4(5)に述べた相関係数の最小残差法に基づいている。

この図-9の結果で明らかのように、統計期間が100年程度以下では3種類の極値分布に対する当てはめ適合率が拮抗しており、最適となる分布は個別の標本によって変動する。統計期間が100年程度以下では一般化極値分布が適合する割合が高いけれども、統計期間が15年以下では一般化パレート分布も同程度の割合で適合する。ワイブル分布への適合率は標本が大きくなるにつれて増大し、 $30\%\sim 35\%$ の割合にまで達する。他の分布に紛れることなしに一般化極値分布が当てはめられるのは、統計期間が5000年を超える場合である。

母集団が指数分布の場合に加えて、 $k=0.001$ の一般化極値分布と $k=0.10$ の一般化パレート分布の母集団から抽出された年最大値のデータが一般化極値分布に最適となる割合を調べたものが図-10である。母集団が一般化極値分布の場合には、年最大値のデータが一般化極値分布に適合する割合が比較的に高い。とは言っても、確実に一般化極値分布に適合するようになるのは、統計期間が1000年を超えるような場合である。

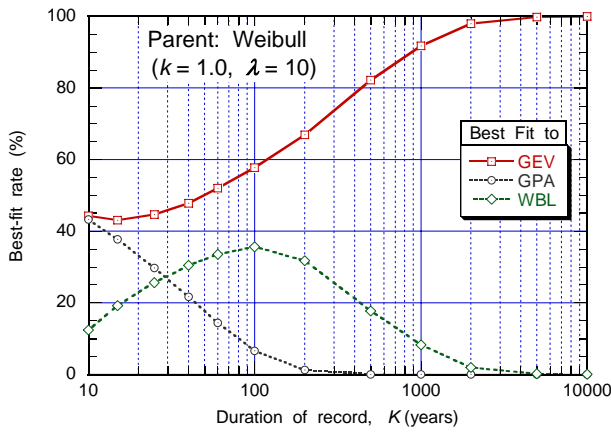


図-9 指数分布から抽出された年最大値のデータが3種類の極値分布関数に適合する割合

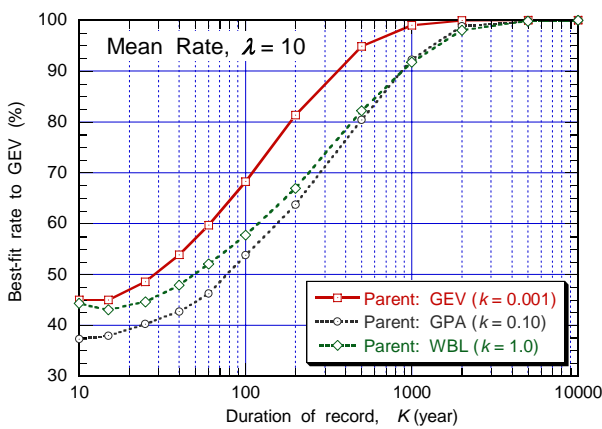


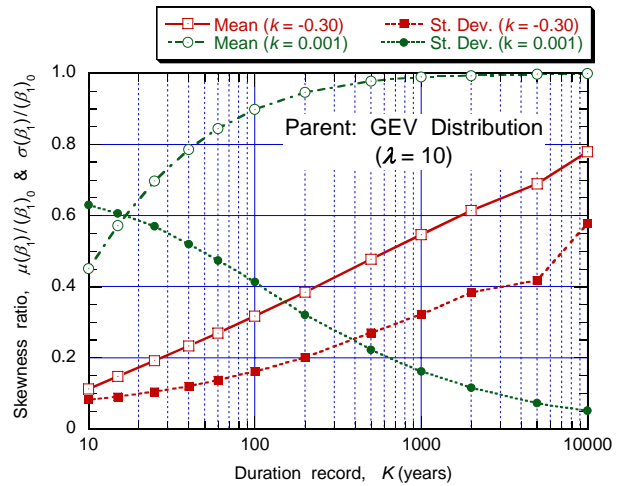
図-10 3種類の母分布関数から抽出された年最大値のデータが一般化極値分布関数に適合する割合

$k=0.10$ の一般化パレート分布の場合には指数分布と同じような挙動である。

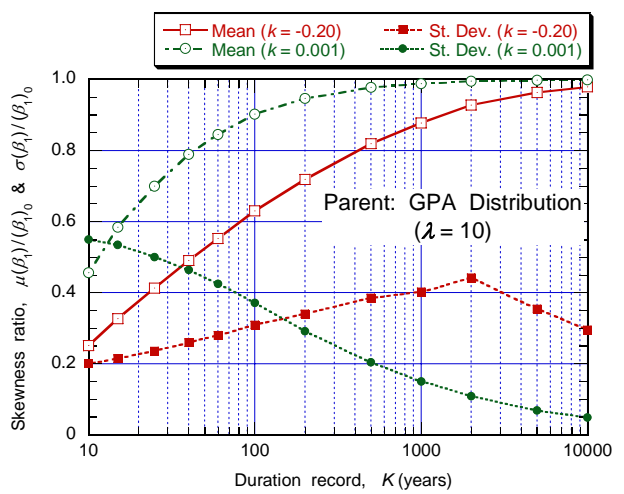
(2) 年最大値の統計期間による標本の歪み度の変化

このように年最大値の標本が小さいときに一般化極値分布に収束しない基本的な原因は、母集団から年最大値の分布関数を導く式(3)が確率の期待値としての計算であるためである。計算の前提としているポアソン分布の式(1)はデータの数十分に大きいときのものであり、データの数少なければ既に図-1, 2に示したように、標本毎にポアソン確率とは異なる値を取る。標本毎の統計的変動は、母集団の分布関数の計算に直接に影響する。

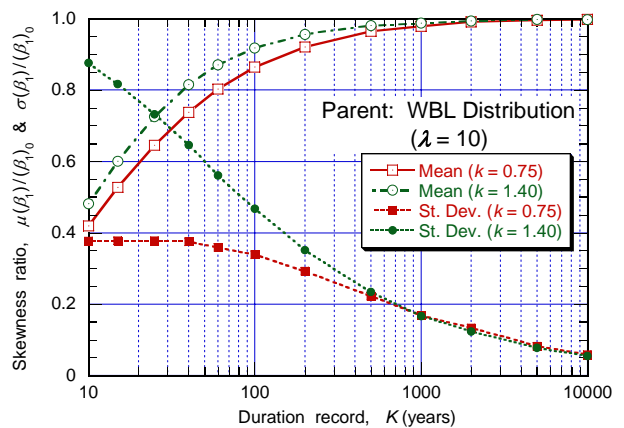
標本毎の統計的変動の影響を直接に示すのは、標本が小さいときにデータの歪み度が小さめに偏って現れる現象である(付録参照)。図-11は、表-2~4に記載した年最大値の標本全体の歪み度 $(\beta_1)_0$ を基準として、標本の歪み度の平均値と標準偏差が年最大値の統計期



(a) 母集団が一般化極値分布の場合



(b) 母集団が一般化パレート分布の場合



(c) 母集団がワイブル分布の場合

図-11 年最大値の標本の歪み度の平均と標準偏差

間によって変化する状況を示したものである。3種類の母集団について2通りの形状母数の値の場合を例示している。一般に、形状母数が小さいほど歪み度の低下が著しい。また、統計期間が短い場合には標準偏差

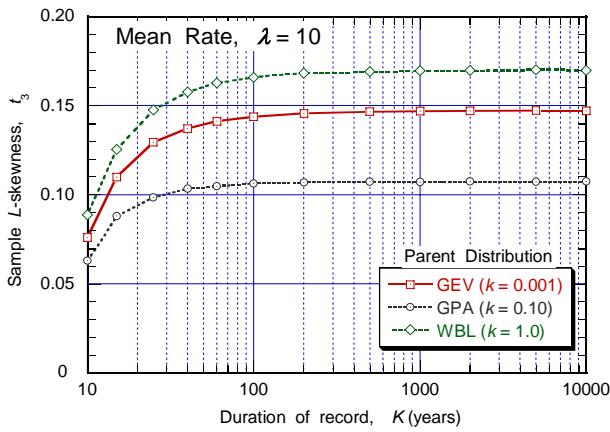


図-12 統計期間による年最大値の標本L-skewnessの変化

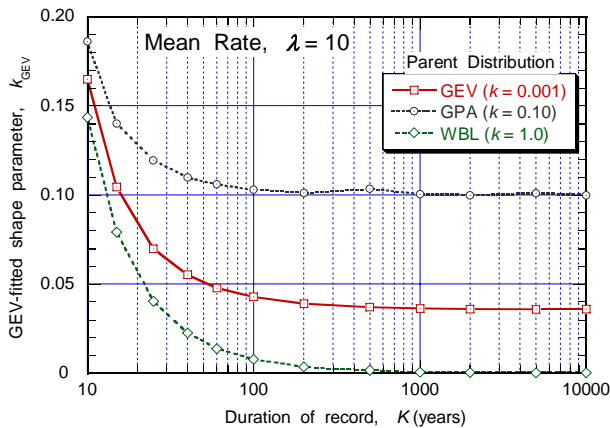
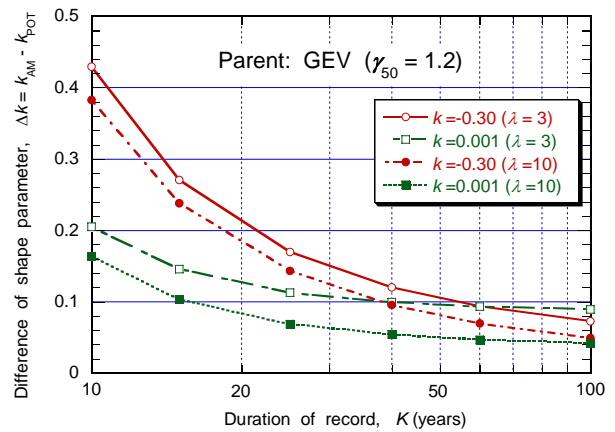


図-13 統計期間による年最大値の標本に当てはめた一般化極値分布の形状母数の変化

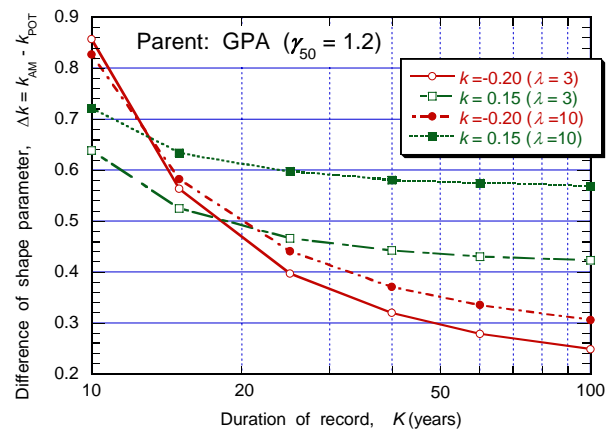
が平均に比べてかなり大きく、ここに例示したものでは統計期間が10年の場合など、歪み度の標準偏差が平均値よりも大きくなっている。すなわち、標本毎の変動が極めて大きい。

L-モーメント法では、標本に対して算定されたL-skewnessの値が当てはめられる分布関数の特性を規定し、このパラメータは標本の歪み度と密接に関連している。したがって、標本の歪み度が年最大値のデータ全体の統計量から低落すると、それに応じてL-skewnessの値も低減し、本来の極値分布に当てはめられる割合が低下するのである。

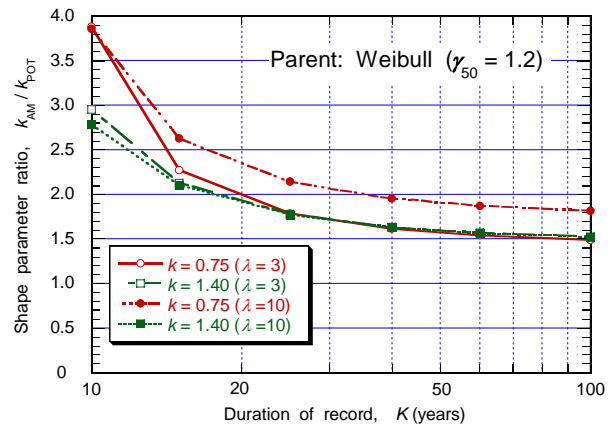
図-12は標本について計算したL-skewnessの値が標本の統計期間に変化する状況を例示したものである。こうしたL-skewnessの値の変化に応じ、年最大値の標本に当てはめた一般化極値分布の形状母数の値は、統計年数が短いほど大きめの値となる。図-13は、図-12に例示したケースについて形状母数 $k_{GEV}$ の値の変化を示したものである。統計期間が十分に長い $K = 10,000$ 年における形状母数の値は、既に図-6に示したところであり、統計期間が短くなるにつれて一般化極値分布に



(a) 母集団が一般化極値分布の場合



(b) 母集団が一般化パレート分布の場合



(c) 母集団がワイブル分布の場合

図-14 年最大値の標本を母集団の極値分布に当てはめたときの形状母数の値

当てはめて推計した形状母数の値が増加する状況が明らかである。

### (3) 統計期間が短い標本の年最大値分布の形状母数

先に図-9に例示したように、統計期間が数十年程度

である極値統計解析の実務においては、年最大値の標本に対して一般化極値分布を一律に当てはめるのは標本毎の特性を考慮しない方法であり、推奨すべきでないであろう。実際問題では、標本に最も適合する極値分布を選び、推計された形状・尺度・位置母数の値を用いて $R$ 年確率統計量を推計することになる。本項では、母集団と同じ極値分布に当てはめられたときに形状母数の値がどの程度変化するかを検討する。

図-14は統計期間が $K = 10 \sim 100$ 年の年最大値標本について、母集団と同じ分布関数を当てはめたときの形状母数を母集団の形状母数と比べたものである。図-14(a), (b)は、一般化極値分布と一般化パレート分布の場合であり、これらの分布は形状母数が0を挟んで正負の範囲に広がるので、形状母数の値の差を表示している。ワイブル分布の場合には形状母数が常に正の値なので、図-14(c)のように母集団の形状母数に対する比として表示している。

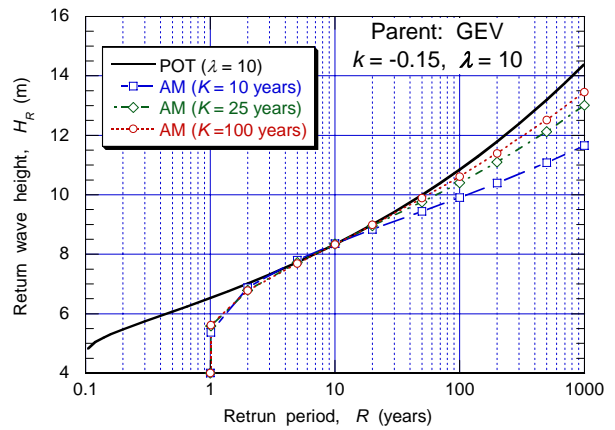
図で明らかのように、統計期間が短くなるにつれて形状母数の推計値が大きくなっている。また、母集団の形状母数の影響は極値分布によって異なり、一般化極値分布とワイブル分布では母集団の形状母数が小さいほうが推計値の変化が大きくなっている。ただし、一般化パレート分布では統計期間によってその関係が逆転している。平均発生率の影響は比較的小さい。

図-14では統計期間が100年までしか図示していないが、形状母数の値の差異は統計期間が長くなっても解消するわけではない。統計期間が $K = 10,000$ 年の標本の計算結果を調べると、一般化極値分布では $\Delta k = 0.02 \sim 0.05$ 程度の差が残る。一般化パレート分布では、 $\Delta k = 0.27 \sim 0.56$ とかなり大きな差である。ワイブル分布では $k_{AM}/k_{POT} = 1.4 \sim 1.7$ である。したがって、たとえば年最大値の標本がワイブル分布に適合したとしても、そのときに推計される形状母数は母集団である極大値の分布関数に対するものよりも1.5～2.0倍程度大きくなっている可能性が高い。これについては、実際の波浪データとの比較を9.で述べる。

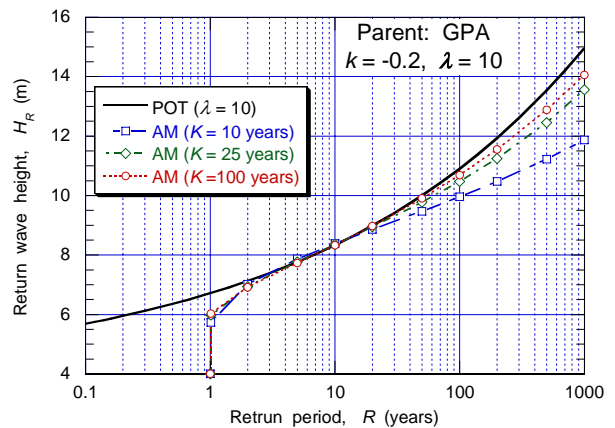
## 7. 年最大値資料に基づく確率波高の偏り率

### (1) 確率波高の推計手順

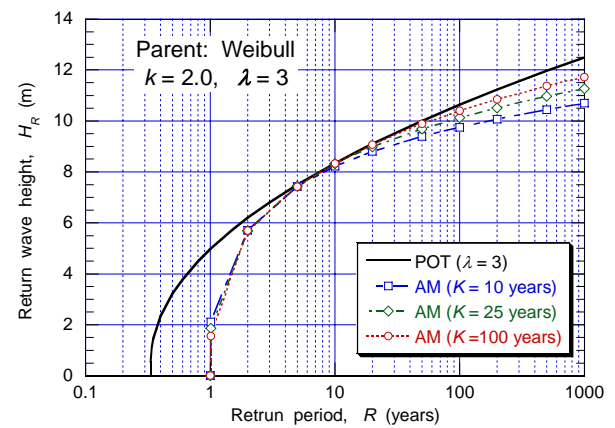
前節で述べたように年最大値の分布が数十年程度の場合には、その分布関数が一般化極値分布に収束するという保証がない。図-9, 10に示したように、一般化パレート分布もワイブル分布もかなりの割合で最適分布として判定される。極値統計解析の実務では、個別



(a) 母集団が一般化極値分布の例



(b) 母集団が一般化パレート分布の例

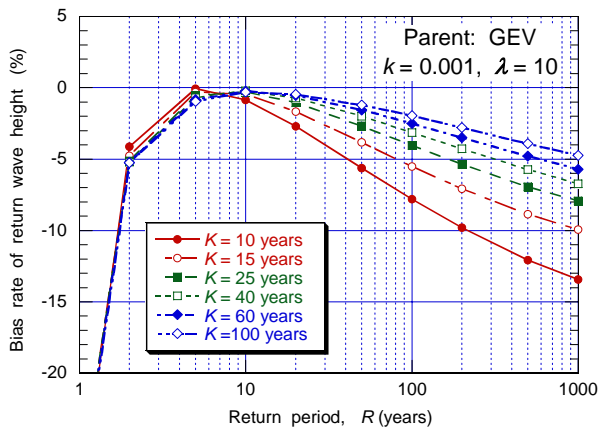


(c) 母集団がワイブル分布の例

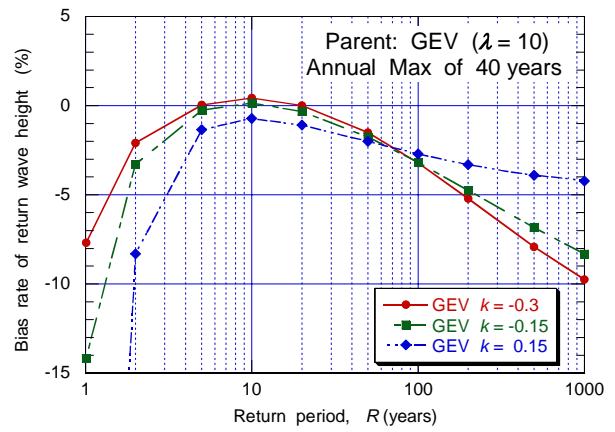
図-15 母集団と年最大値の標本から推計される確率波高の比較

の標本について最適と判定された分布関数を用いて $R$ 年確率波高を推計することになる。

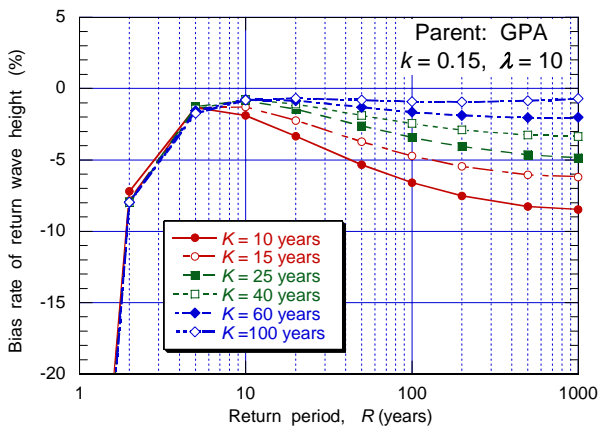
本研究では、個々の標本ごとに最適分布を判定し、その分布に当てはめられた形状・尺度・位置母数の数値を用いて、 $R = 1, 01, 2, \dots, 500, 1000$ 年に対する確率



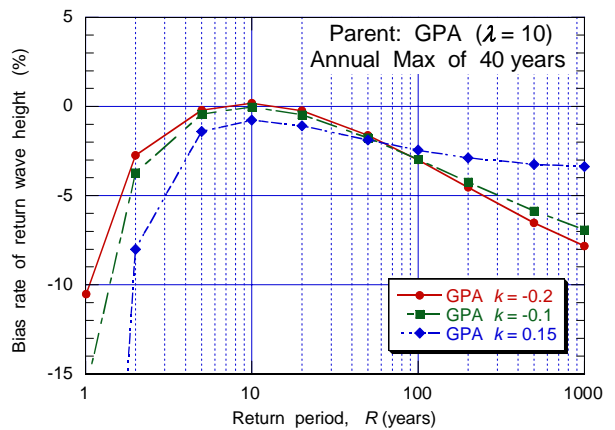
(a) 母集団が一般化極値分布の例



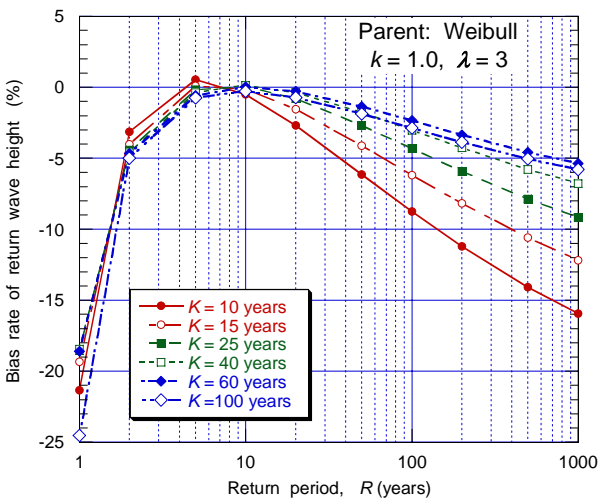
(a) 母集団が一般化極値分布の場合



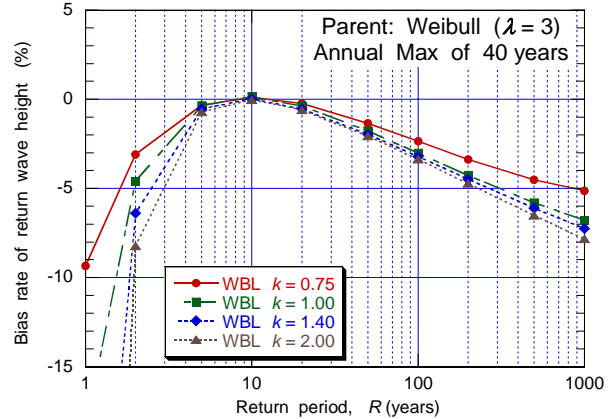
(b) 母集団が一般化パレート分布の例



(b) 母集団が一般化パレート分布の例



(c) 母集団がワイブル分布の例



(c) 母集団がワイブル分布の例

図-17 統計期間40年の年最大値資料から推計される確率波高の偏り率

図-16 年最大値の標本から推計される確率波高の偏り率

波高を推計した。そして、 $N$ 組の標本から得られた再現期間毎の確率波高の平均値をもって、 $N$ 組全体に対する確率波高と定義した。また、 $N$ 組の標本に対する確率波高の標準偏差を計算した。標準偏差と平均値と

の比は変動係数であり、これについては8. で述べる。3種類の母集団から抽出された年最大値の標本について確率波高を推計し、母集団における確率波高と比較した事例を、母集団毎に図-15に示す。年最大値というのは、1年間に出現する極値の中の最大値であるか

ら、再現期間としては $R < 1$ 年はあり得ない。したがって、年最大値の確率波高は $R = 1$ 年から立ち上がる形となる。図-15には母集団毎に統計年数が $K = 10, 25$ , および100年の場合に推計される確率波高を示しており、母集団における確率波高と比較できるようにしてある。図から明らかなように、再現期間10年付近では年最大値と母集団との確率波高に差異はないけれども、再現期間が短い場合ならびに長い場合には年最大値が確率波高を過小評価している。再現期間が5年未満については年最大値という定義に起因するものであり、是正は不可能である。

しかし、再現期間50年以上に対する過小評価は、年最大値の標本がそのサイズの影響を受けた結果である。すなわち、図-14に示したように、年最大値の標本の統計期間が数十年程度の場合には形状母数の推計値が大きめに見積もられる。極値分布関数は形状母数が大きくなるにつれて確率密度関数の裾が短くなり、確率波高が低めに推計されるようになる。母集団の確率波高との差は統計年数が長くなるにつれて減少するけれども、 $K = 100$ 年程度ではまだ明らかな差が認められる。

## (2) 年最大値から推計される確率波高の偏り率

確率波高の推計値の差異を偏り率として表示した例が図-16である。図-14とは異なる形状母数の場合について、統計年数が10～100年の6通りの場合を示している。母集団が一般化パレート分布の場合には再現期間が200年を超えるあたりから偏り率の変化が小さくなるけれども、母集団が一般化極値分布とワイブル分布の場合には再現期間が伸びるに従って偏り率が単調に増加する。

高波の極値統計の場合には実際の波浪データの統計期間の上限が40年程度であるので、 $K = 40$ 年のケースについて確率波高推計値の偏り率を提示したのが図-17である。3種類の母集団毎に形状母数を3～4通りに変

えた事例を示している。母集団の分布関数によって若干の差はあるけれども、50年確率波高で2%程度、100年確率波高で3%程度の過小評価が見込まれる。

高波の極値統計では再現期間として100年を超える事象を推計する例はあまりないけれども、高潮などの場合にはオランダで行われているように1万年確率偏差を求めることもあり得る。高潮などについては裾長さの設定など高波と異なる点もあるので、図-16で示されているトレンドをそのまま引き延ばすわけにはいかいであろう。それでも、年最大値資料が確率統計量を過小評価する可能性が大きいことを考えるならば、高潮についても極大値資料を使用することが推奨される。

## 8. 年最大値資料に基づく確率波高の変動係数

極大値資料から推計される確率波高に関しては、前報においてその変動係数の経験的推定式を導いた。本研究においてもその推定式を参考として、年最大値資料から推計される確率波高の変動係数について作業を行った。ただし、前報では裾長さとして3通りの値もを用いて裾長さの影響を定式化したけれども、今回は $\gamma_{50} = 1.20$ に固定している。そこで、裾長さに関しては前報の定式化をそのまま使用した。

さらに、年最大値資料の場合には変動係数が標本の大きさ $n$ の $-1/2$ 乗にほぼ比例したけれども、年最大値資料の場合には母集団によって $K^{0.4} \sim K^{0.45}$ に比例する形となった。そこで、年最大値資料の確率波高推計値の変動係数に関する推定式を次のように定式化して、そこに用いられる定数 $a, b, c$ を数値計算の結果から定めることにした。

$$CV[H_R] = (ay_R^2 + by_R + c)(\gamma_{50} - 1)^{0.7} / K^\kappa \quad (29)$$

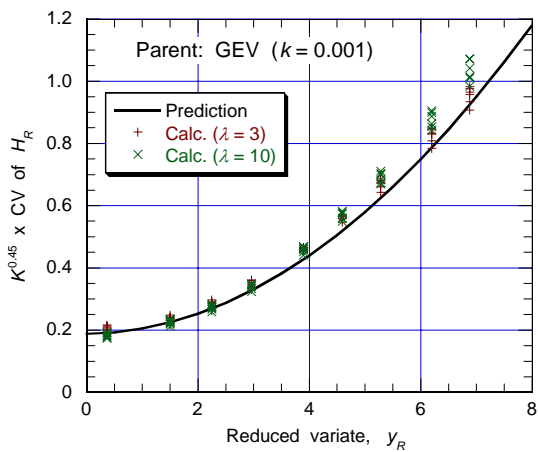
ここに、 $\kappa = 0.45$ （一般化極値分布）および $\kappa = 0.4$ （一般化パレート分布およびワイブル分布）と設定する。

定数 $a, b, c$ については表-6のように設定した。

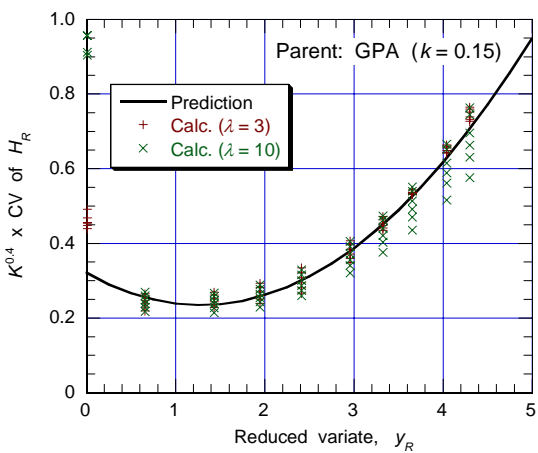
表-6 年最大値資料から推計された確率波高の変動係数の近似推定式の係数表

分布関数	係数の推定式
一般化極値分布 (GEV)	$a = \exp(-0.00088k^2 + 0.2458k + 0.01295)$ $b = -4.0490k^2 - 1.8910k + 0.0062$ $c = 4.2782k^2 + 2.9640k + 0.5782$
一般化パレート分布 (GPA)	$a = 24.594k^4 + 6.587k^3 + 1.0423k^2 + 0.3495k + 0.0493$ $b = -68.204k^4 - 18.949k^3 - 4.1795k^2 - 1.5050k + 0.0077$ $c = 20.605k^3 + 6.878k^2 + 2.3229k + 0.4161$
ワイブル分布	$a = 0.3385k^2 - 0.4510k + 0.1280$ $b = -0.9265k^2 + 1.3062k + 0.3658$ $c = 0.5342k^2 - 0.3864k + 0.2756$

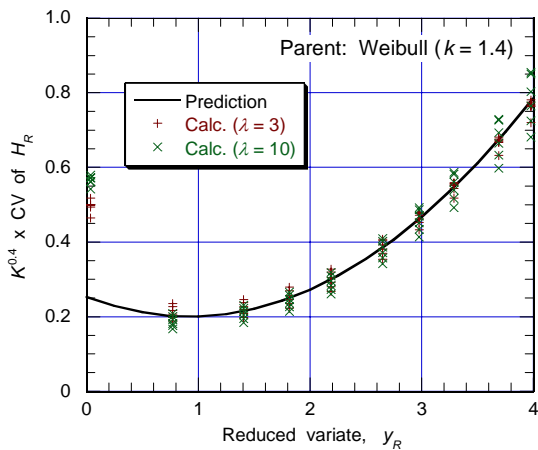




(a) 母集団が一般化極値分布の例



(b) 母集団が一般化パレート分布の例



(c) 母集団がワイブル分布の例

図-18 年最大値資料に基づく確率波高の変動係数の推定式と数値計算結果の比較

これらの経験式による確率波高推計値の変動係数の予測値を数値実験で得られたデータと比較した結果の例を図-18に示す。予測値は曲線で示し、数値実験結

果のデータは+およびxで示してある。横軸の同じ値の位置に複数のデータがあるのは、年最大値の異なる統計期間の計算結果である。図-18(b), (c)で横軸の値数が0の付近で数値実験の値が異常に大きいのは再現期間1.01年に対するものであり、無視してよい。

なお、式(29)の基準化変数 $y_R$ および表-6の定数 $a, b, c$ の算定式は、抽出された極大値の母集団の形状母数を引数としている。年最大値の標本を解析するときにはその極大値の母集団は確定しておらず、したがってその形状母数も分からない。実際問題への適用にあたっては、図-14に示す年最大値の小標本における形状母数のデータを参考とし、年最大値が抽出されたであろう極大値の母集団の形状母数を推定する作業が必要となる。これについては9(2)で計算例を示すので参照されたい。

## 9. 現地の高波データへの適用性について

### (1) 現地波高データの極大値と年最大値解析の比較

先に合田・久高(2010)は日本海沿岸11地点の高波データについてL-モーメント法による地域頻度解析を行った。そのデータは極大値資料であったが、各地点について年最大値のデータを抽出し、極値解析を行ってみた。欠測を除いた有効統計年数は $K^* = 18.0 \sim 37.2$ 年であった。また、ポアソン分布について検討したアラスカ沿岸Kodiak島沖の波浪データについても年最大値を抽出して極値解析を行った。

これら12地点の極大値の標本はいずれもワイブル分布によく適合し、また年最大値の標本もワイブル分布に適合した。そこで、ワイブル分布として推計された形状母数の値の比率を求めた結果を図-19に示す。

日本海沿岸の高波の極大値資料では、ワイブル分布の形状母数が $k_{\text{POT}} = 1.060 \sim 1.344$ 、平均で1.225であり、Kodiak沖では $k_{\text{POT}} = 1.518$ であった。数値シミュレーションでは概略値として $k_{\text{POT}} = 1.25$ を用いた。シミュレーションによる形状母数の比は、平均値を実線で表し、その上下に標準偏差の1倍をとった値を一点鎖線および二点鎖線で示している。形状母数の比が正規分布すると仮定すると、この二つの鎖線は形状母数比の68%信頼区間を表すことになる。現地の高波のデータはばらつきが大きいものの、十分に信頼区間内には入っており、今回の数値シミュレーションの結果が現地にも適用できることを示している。

また100年確率波高について、極大値資料から推計された値を真値 $\xi_0$ とみなし、式(26)を用いて年最大値資料による確率波高推計値の偏り率を計算した結果を

図-20 に示す．各地点の偏り率は-9.9% ~ 0.8% とばらついており，算術平均が-4.6% である．12 地点の平均統計年数 26.8 年に対応する数値シミュレーションによる 100 年確率波高の平均偏り率は-3.4% であり，現地データよりはわずかに小さい．しかし，偏り率の 68% 信頼区間の幅が非常に広いため，この程度の平均値の差は統計的に有意なものではない．したがって，確率波高の推計値についても数値シミュレーションの結果

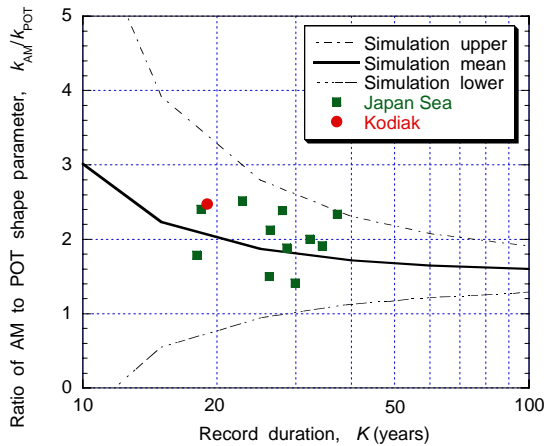


図-19 日本海沿岸および Kodiak 沖の波高に当てはめられたワイブル分布の形状母数の比率

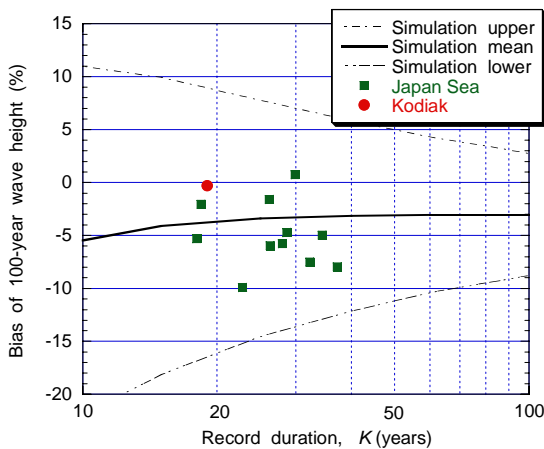


図-20 極大値資料から推計された 100 年確率波高を基準とする年最大値資料の 100 年確率波高の偏り率

が現地データで検証されたといえる．

(2) 年最大値から推計された確率波高の信頼区間の試算例

ここで，酒田港の年最大値資料から推計された確率波高についてその信頼区間の推定を試みる．酒田港の年最大値資料に対して L-モーメント法で推計した母数値と確率波高は次の通りである．

$$K^* = 34.4 \text{ 年}, \lambda = 15.06 \text{ 回/年} (H_c = 4.0 \text{ m})$$

$$k_{AM} = 2.313, A_{AM} = 3.230 \text{ m}, B_{AM} = 4.331 \text{ m}$$

$$H_{10} = 8.96 \text{ m}, H_{50} = 10.16 \text{ m}, H_{100} = 10.58 \text{ m}, \gamma_{50} = 1.13$$

8. で紹介した確率波高の変動係数の推定式を使うためには，まず極大値の母集団に対する形状母数を推定する必要がある．そこで図-14 を参照し， $k_{AM}/k_{POT} = 1.8$  と想定する．それによって， $k_{POT} = 2.313/1.8 = 1.29$  と推定される（実際の酒田港の極大値の形状母数は波高閾値によって  $k = 1.17 \sim 1.27$  の間で変化した）．

この形状母数の推定値  $k_{POT} = 1.29$  を表-6 の定数算定式に代入すると，ワイブル分布の確率波高の変動係数の定数が次のように計算される．

$$a = 0.3385 \times 1.29^2 - 0.4510 \times 1.29 + 0.1280 = 0.1095$$

$$b = 0.9265 \times 1.29^2 + 1.3062 \times 1.29 + 0.3658 = 0.5090$$

$$c = 0.5342 \times 1.29^2 - 0.3864 \times 1.29 + 0.2756 = 0.6661$$

この定数を  $K^* = 34.4$  年， $\gamma_{50} = 1.13$  とともに式 (29) に代入する．年最大値の場合の再現期間  $R = 100$  年に対する基準化変数は  $y_{100} = (\ln 100)^{1/2.313} = 1.935$  である．したがって，100 年確率波高に対する変動係数が次のように計算される．

$$CV = (0.1095 \times 1.935^2 + 0.5090 \times 1.935 + 0.661) \times (1.13 - 1.00)^{0.7} / 34.4^{0.4} = 0.0761$$

これによって  $H_{100} = 10.58$  m に対する標準偏差が  $\sigma = 0.0761 \times 10.58 = 0.805$  m と算定される．さらに，図-14(c) を参照して 100 年確率波高に対する偏り率を -3% と見積もると， $H_{100} = 10.58$  m に対する 68% 信頼区間が  $H_{100} = 10.09 \sim 11.70$  m と推定される．

なお，極大値資料から推計される確率波高の変動係数については前報においてその推定式を次のように提示している．

$$CV[H_R] = (ay_R^2 - by_R + c)(\gamma_{50} - 1)^{0.7} / \sqrt{n} \quad (30)$$

酒田港の極大値資料では  $n = 34.4 \times 15.06 = 518$  であった．極大値資料に対する上式の定数は，前報の表-8 の算定式を用いて次のように計算される．

$$a = 0.9742 \times (\ln 1.29)^2 + 0.6949 \times (\ln 1.29) + 0.1517 = 0.3918$$

$$b = 1.9049 \times (\ln 1.29)^2 + 1.8142 \times (\ln 1.29) + 0.3677 = 0.9532$$

$$c = 0.6349 \times (\ln 1.29)^2 + 1.8561 \times (\ln 1.29) + 0.9095 = 1.423$$

一方， $\lambda = 15.06$  回/年のときの再現期間 100 年に対する基準化変数は  $y_R = (\ln(15.06 \times 100))^{1/1.29} = 4.678$  であるので，これらの数値を式 (30) に代入することによって，変動係数が次のように計算される．

$$CV = (0.3918 \times 4.678^2 - 0.9532 \times 4.678 + 1.423) \\ \times (1.13 - 1.00)^{0.7} / 518^{0.5} = 0.0583$$

すなわち、この事例の場合には、極大値資料から推計される100年確率波高の変動係数は5.8%であり、年最大値資料の場合の変動係数7.6%よりも23%小さい。確率波高の変動係数の値を極大値資料と年最大値資料について試算してみると、平均発生率が低くて再現期間が長い場合には年最大資料のほうが小さいことがあるけれども、平均発生率が高ければ極大値資料のほうが変動係数が小さくなる。この大小関係は条件によって異なるので、本論文の表-6および前報の表-8に記載した推定式を使って試算してみる必要がある。

## 10. まとめ

今回のモンテカルロ法による一連の数値シミュレーションによって明らかになった諸点をまとめると次のようになる。

- 1) 高波の年間発生回数がポアソン分布に従うことが酒田港およびアラスカ沿岸 Kodiak 島沖の波高データで例証された。
- 2) モンテカルロ法を用い、一般化極値分布、一般化パレート分布、ワイブル分布の3種類の母集団から年最大値を抽出し、その極値統計特性を検討した。年最大値の標本は、統計期間が最長1万年に相当するものまでを作成した。
- 3) 年最大値の極値データは、統計期間が数千年以上と十分に長く、平均発生率が年5回程度以上であれば、母集団が一般化極値分布、一般化パレート分布、ワイブル分布のいずれであっても、一般化極値分布に従うことが確認された。
- 4) 上記の3種類の母集団から抽出された大標本の年最大値データに当てはめられた一般化極値分布の形状・尺度・位置母数に関して、母集団の母数からの変換式が明示された。
- 5) 実際の統計解析で入手可能な統計期間が数十年程度の年最大値データ（小標本）の場合には、一般化極値分布への収束は期待できず、3種類の極値分布関数のいずれもが最適となる可能性がある。
- 6) 上記の減少は、標本の大きさが縮小するにつれて標本の歪み度が低減することに関連している。標本が小さくなるにつれて  $L$ -skewness が減少し、標本に当てはめられる形状母数の値が増大する。
- 7) 小標本の年最大値資料から推計される確率波高は、再現期間が50年程度以上の場合に母集団の確率波

高に比べて過小評価（負の偏り）となり、再現期間が長くなるほど過小評価の度合いが強くなる。

- 8) 上記のことから、高潮偏差など再現期間の長い確率事象を推計する場合には、小標本の年最大値資料ではなく、極大値資料を使用することが推奨される。
- 9) 統計期間40年の年最大値の標本から推計した100年確率波高は、母集団の値に対して-3%程度の過小評価となることを例示した。
- 10) 小標本の年最大値資料の形状母数値が増大し、確率波高が過小評価される傾向は、日本海沿岸11地点および Kodiak 島沖の高波データで確認された。
- 11) 年最大値から推計される確率波高の変動係数の経験的推定式が導かれた。
- 12) 酒田港の波高データを用いて、年最大値資料による100年確率波高の変動係数の試算ならびに極大値資料による変動係数との比較の事例を示した。

最初にも述べたように、これまでの高波の極値統計解析では年最大値資料がほとんど使われていなかったけれども、今後は高波データの蓄積とともに年最大値資料の利用が試みられる可能性がある。しかし本論文で述べたように、統計期間が数十年程度の年最大値資料は確率波高の推計に負の偏りをもたらし危険性が高い。既に年最大値資料の利用が一般的である高潮その他の極値統計についても、極大値資料との比較検討を進めることが望まれる。

本論文の成果が種々の分野の極値統計解析に活用されることを願う次第である。

## 付録：非対称分布における標本 Skewness の特性

本文の6(2)で述べたように、標本が小さいときには標本の歪み度 (skewness) が母集団よりも平均的に小さくなる。この問題について別途の数値実験を行ったので、以下にその概要を記載する。

### A.1 対象とした非対称分布関数

数値実験では式 (A.1) ~ (A.3) に示すべき乗関数、一般化パレート (GPA) 分布関数、およびワイブル分布関数の3種類の関数を対象とした。ただし、取り扱いを簡単にするため、尺度母数を  $A=1$ 、位置母数を  $B=0$  に設定した。このうちのべき乗関数は、 $[0, 1]$  の範囲でのみ定義され、確率密度関数  $f(x)$  は  $k > 1$  であれば  $x=0$  で最大値を取り、その後次第に減少して  $x=1$  で0となる。形状母数が  $k=2$  のときの  $f(x)$  の減少は放物線型である。

A. べき乗関数

$$\left. \begin{aligned} F(x) &= 1 - (1-x)^k \quad : 0 \leq x \leq 1, \\ f(x) &= k(1-x)^{k-1}, \\ x &= F^{-1}(P) = 1 - (1-P)^{1/k}. \end{aligned} \right\} \quad (A.1)$$

B. 一般化パレート (GPA) 分布関数

$$\left. \begin{aligned} F(x) &= 1 - (1-kx)^{1/k} \quad : 0 \leq x \leq 1/k \text{ for } 0 < k \leq 1, \\ & \quad : 0 \leq x < \infty \text{ for } k < 0, \\ f(x) &= (1-kx)^{1/k-1} \quad : k \neq 0, \\ x &= F^{-1}(P) = [1 - (1-P)^k] / k. \end{aligned} \right\} \quad (A.2)$$

C. ワイブル分布関数

$$\left. \begin{aligned} F(x) &= 1 - \exp[-x^k] \quad : 0 \leq x < \infty, \\ f(x) &= kx^{k-1} \exp[-x^k], \\ x &= F^{-1}(P) = [-\ln(1-P)]^{1/k}. \end{aligned} \right\} \quad (A.3)$$

上記の3分布関数の平均，標準偏差，歪み度 (Skewness) などはすべて形状母数  $k$  の関数として表-A.1のように表される（一部は表-5と重複）．これに基づいて3分布関数の特性値を幾つかの形状母数の値について計算した結果を表-A.2に示す．

表-A.1 非対称分布関数の特性値の計算式

分布関数	平均 $E[x] = \mu [x]$	$E[x^2]$	$E[x^3]$	標準偏差 $\sigma[x]$	歪み度 (Skewness) $\beta_1$
べき乗関数	$\frac{1}{k+1}$	$\frac{2}{(k+1)(k+2)}$	$\frac{6}{(k+1)(k+2)(k+3)}$	$\frac{1}{k+1} \left[ \frac{k}{(k+2)} \right]^{1/2}$	$\frac{2(k-1)(k+2)^{1/2}}{k^{1/2}(k+3)}$
一般化パ レ - ト (GPA)	$\frac{1}{1+k}$	$\frac{2}{(1+k)(1+2k)}$	$\frac{6}{(1+k)(1+2k)(1+3k)}$	$\frac{1}{(1+k)(1+2k)^{1/2}}$	$\frac{2(1-k)(1+2k)^{1/2}}{(1+3k)}$
ワイブル	$\Gamma\left(1 + \frac{1}{k}\right)$	$\Gamma\left(1 + \frac{2}{k}\right)$	$\Gamma\left(1 + \frac{3}{k}\right)$	$\left\{ \Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right\}^{1/2}$	$\frac{E[x^3] - 3E[x^2] \times E[x] + 2E^3[x]}{\sigma^3[x]}$

表-A.2 非対称分布関数の特性値の算定例

特性値	べき乗関数		一般化パレート (GPA)		ワイブル	
	$k=2$	$k=3$	$k=0.2$	$k=-0.2$	$k=1.0$	$k=2.0$
平均	0.3333	0.2500	0.8333	1.2500	1.0	0.8862
標準偏差	0.2357	0.1937	0.7043	1.6137	1.0	0.4633
歪み度	0.5657	0.8607	1.1832	4.6476	2.0	0.6311

表-A.3 標本の大きさによる非対称分布関数の歪み度の簡易計算

標本サ イズ $n$	べき乗関数		一般化パレート (GPA)		ワイブル	
	$k=2$	$k=3$	$k=0.2$	$k=-0.2$	$k=1.0$	$k=2.0$
5	0.378	0.512	0.619	0.919	0.775	0.262
10	0.485	0.685	0.858	1.412	1.132	0.392
20	0.533	0.776	1.001	1.841	1.394	0.482
50	0.556	0.830	1.103	2.346	1.636	0.555
100	0.562	0.847	1.142	2.687	1.759	0.587
1000	0.566	0.860	1.179	3.579	1.949	0.625

A.2 小標本の歪み度の簡易計算

ここで標本が小さいときの歪み度を求めることを考える．大きさが  $n$  の標本中のデータ  $x$  が昇順，すなわち  $x_1 \leq x_2 \leq \dots \leq x_j \dots \leq x_n$  と並べ替えられている

とする．これらのデータの非超過確率  $P_j$  は非減少の数値列を形成する．すなわち次の関係を満たす．

$$0 < P_1 \leq P_2 \leq \dots \leq P_j \leq \dots \leq P_n < 1 \quad (A.4)$$

非超過確率  $P_j$  の値は標本によって異なるけれども，多数の標本についての平均，すなわち期待値

$E[P_j]$  は  $(j-1)/n$  と  $j/n$  の間の値を取ると考えられる。そこで、 $E[P_j]$  がその中間値である  $(j-1/2)/n$  の値をとると仮定すると、各データの期待値  $E[x_j]$  は  $F^{-1}[(j-1/2)/n]$  として計算できる。そうすると、標本の平均  $\mu[x]$ 、分散  $\sigma^2[x]$ 、および歪み度  $\beta_1$  はデータの1~3次の積率を用いて次のように求められる。

$$\mu[x] = \frac{1}{n} \sum_{j=1}^n x_j = E[x] \quad (\text{A.5})$$

$$\sigma^2[x] = \frac{1}{n} \sum_{j=1}^n (x_j - \mu[x])^2 = E[x^2] - E^2[x] \quad (\text{A.6})$$

$$\begin{aligned} \beta_1 &= \frac{1}{\sigma^3[x]} \frac{1}{n} \sum_{j=1}^n (x_j - \mu[x])^3 \\ &= \frac{1}{\sigma^3[x]} (E[x^3] - 3E[x^2] \times E[x] + 2E^3[x]) \end{aligned} \quad (\text{A.7})$$

ここに、

$$E[x^2] = \frac{1}{n} \sum_{j=1}^n E^2[x_j], \quad E[x^3] = \frac{1}{n} \sum_{j=1}^n E^3[x_j] \quad (\text{A.8})$$

標本の大きさを  $n = 5 \sim 1000$  で変化させて式 (A.1) ~ (A.3) の3種類の分布関数について歪み度を計算した結果を表-A.3に示す。標本が小さくなるにつれて歪み度が減少することが明らかに見とれる。 $n = 1000$  の場合、べき乗関数では母集団の歪み度とほぼ同じ値が得られるが、非対称の度合いが強い一般化パレート分布の  $k = -0.2$  の場合には、 $n = 1000$  であっても母集団よりも小さな値にとどまる。なお平均  $\mu[x]$  および  $\sigma[x]$  についても  $E[x_j] = F^{-1}[(j-1/2)/n]$  と仮定して計算を行ったが、標本サイズの影響は僅かであった。

標本が小さいときに歪み度が母集団よりも小さくなるのは、データの最大値  $x_n$  の非超過確率の期待値が  $P_n = 1 - 1/2n$  であるために最大値の期待値も低く抑えられ、その結果として式 (A.8) の3次の積率  $E[x^3]$  が小さな値になるためである。

### A.3 数値実験による非対称分布の歪み度の算定

次に、モンテカルロ法を用いて3種類の分布関数から多数の標本を抽出し、標本毎の歪み度を計算した結果を述べる。標本の大きさは  $n = 10 \sim 10,000$  で段階的に変化させ、標本数は  $n = 1000$  以下では  $N = 100,000$  とし、 $n = 2,000 \sim 10,000$  では計算時間を節約するために  $N = 10,000,000/n$  とした。

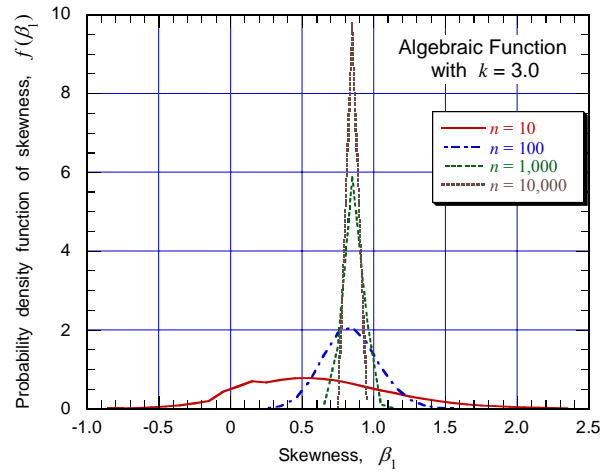


図-A.1 べき乗関数 ( $k=3$ ) の歪み度の頻度分布

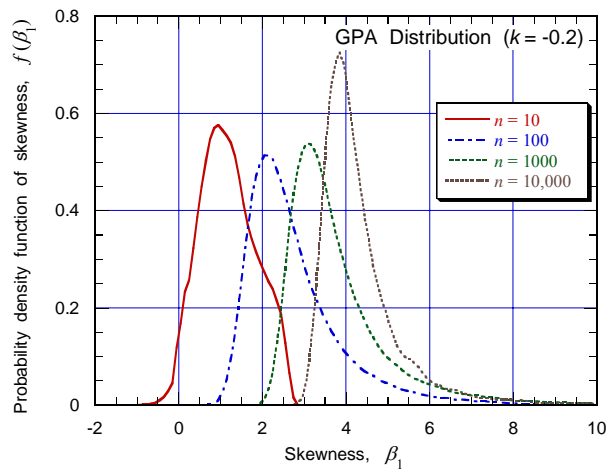


図-A.2 GPA分布 ( $k=-0.2$ ) の歪み度の頻度分布

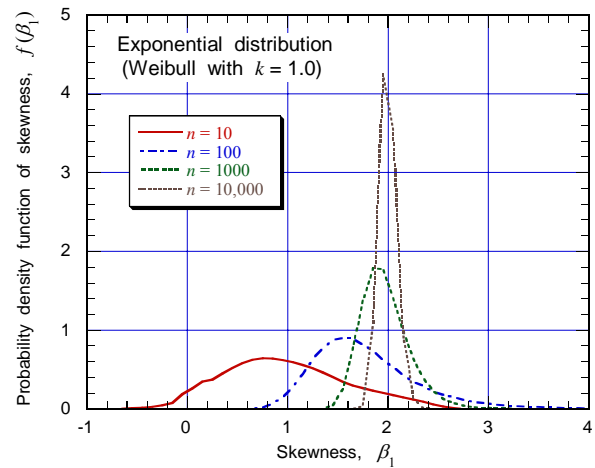


図-A.3 ワイブル分布 ( $k=1.0$ ) の歪み度の頻度分布

数値実験で得られた歪み度の頻度分布を確率密度関数の形で示したのが図-A.1 ~ A.3である。それぞれべき乗関数 ( $k=3$ )、一般化パレート分布 ( $k=-0.2$ )、およびワイブル分布 ( $k=1.0$ : 指数分布) の場合であ

る。

頻度分布は級幅  $\Delta\beta_1 = 0.1$  で作成している。なお、標本が  $n = 10$  と小さい場合には、級毎の頻度がやや不規則に変動したため、連続する三つの級にわたる移動平均を行って平滑化している。3分布関数いずれのケースも、標本が小さくなるにつれて頻度分布が左にシフト（歪み度が減少）し、歪み

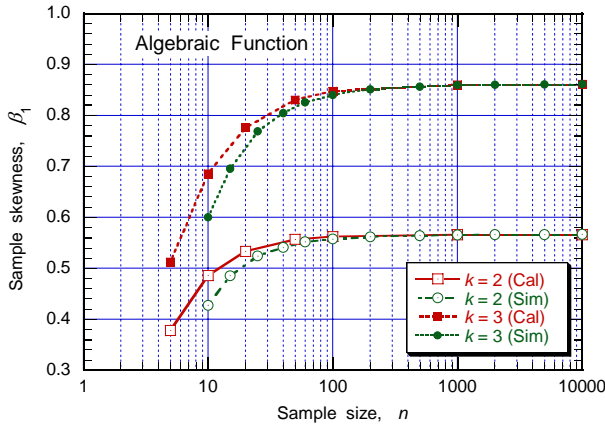


図-A.4 標本サイズと乗関数の歪み度の関係

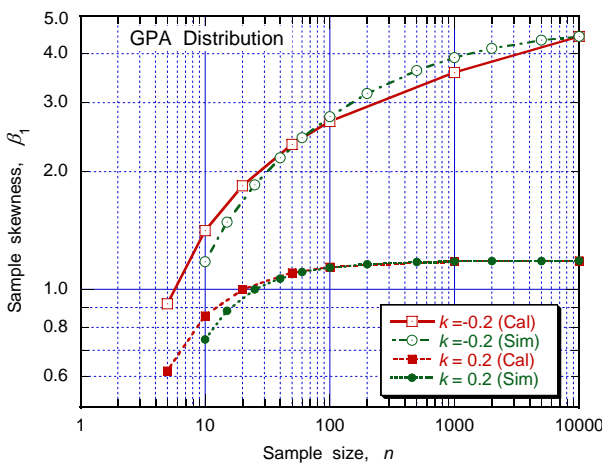


図-A.5 標本サイズとGPA分布の歪み度の関係

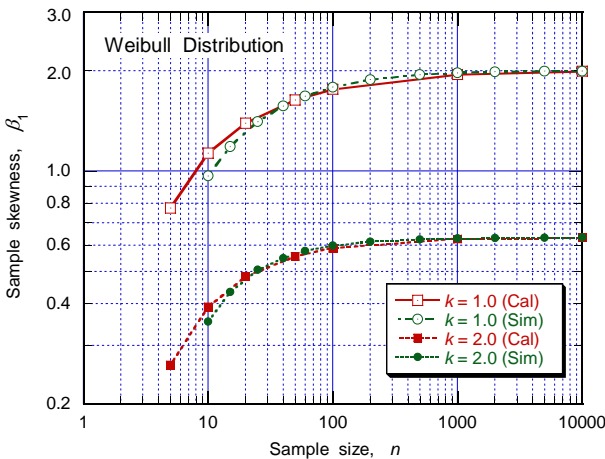


図-A.6 標本サイズとワイブル分布の歪み度の関係

度の分布幅が広がる。

図-A.4 ~ A.6は数値実験で得られた歪み度の平均値と標本サイズとの関係を、表-A.3に示した簡易計算の結果と比較して示したものである。簡易計算は  $n = 1000$  までしか行っていないので、 $n = 10,000$  については表-A.2の母集団の値を記載している。数値実験の結果はSim、簡易計算はCalの凡例で示してある。これらの図に見られるように、両者の値には若干の差があるものの、標本の大きさに応じて歪み度が変化する様相は同じである。

この結果は、A.2節で紹介した考え方、すなわち小標本における歪み度の低下は標本中の最大値が低めに抑えられることに起因するというメカニズムを確認するものである。本文の図-11, 12に示した年最大値資料の歪み度が統計期間の長さによって変化する現象は、こうした非対称分布の特性によるものである。

#### 参考文献

- 合田良実 (2008) : 「耐波工学 - 港湾・海岸構造物の耐波設計-」, 鹿島出版会, 430p.
- 合田良実 (2010) : *L*-モーメント法を用いた高波の極値統計解析におけるプロットングポジション公式の選択と各種統計量の信頼区間, ECOH/YG 技術論文 No. 13, 2010年9月30日, 29p.
- 合田良実・久高将信 (2010) : *L*-moment法による日本海沿岸の高波の地域頻度解析について, ECOH/YG 技術論文 No. 12, 2009年4月12日, 13p.
- Chowdhury, J. U., J. R. Stedinger, and L.-H. Lu (1991): Goodness-of-fit tests for regional generalized extreme value flood distributions, *Water Resources Res.*, Vol. 27, No. 7, pp. 1765-1776.
- Cunnane, C. (1973): A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, *J. Hydrology*, Vol. 18, pp. 257-271.
- Hosking, J. R. M. (1990): *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. Roy. Statistical Soc., Series B*, 52, pp. 105-24.
- Hosking, J. R. M. and J. R. Wallis (1997): “*Regional Frequency Analysis*,” Cambridge Univ. Press, 224p.
- Madsen, H., C. P. Pearson, and D. Rosbjerg (1997): Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling, *Water Resources Res.*, Vol. 33, No. 4, pp. 759-769.