波浪の極値統計における分布関数のあてはめ基準

合 田 良 実*・小 舟 浩 治**

1. まえがき

海岸や海洋の構造物の計画では，まず設計波浪の選定から作業が始められる．波浪観測資料や推算資料を解析し，高波の出現確率を推定する．これは極値統計解析の一つであり，幾つかの極値分布関数のなかからデータにもっともよく適合すると判断される関数をあてはめるのが通例である．しかし，分布関数のあてはめというのは与えられた一つの極値資料の標本から母集団のタイプを特定しようとすることであるから，標本の大きさが数百以上でなければ，誤った答えを得る危険性が常に存在する．その意味では，最適合の絶対的基準というものはありえない．

ここでは発想を逆転させ，対象とするデータの側からみてあてはめを棄却すべき分布関数の判断基準を検討する．年最大風速や年最大日雨量の全数極値資料については既に小野澤・合田(1989)と合田・李(1989)が検討結果を発表しているので，ここではこれを高波のような部分極値資料に対して拡張した結果を報告する．

2. 極値分布関数およびシミュレーション方法

極値統計の問題では，いろいろな分布関数が用いられる．今回は，著者の一人（合田,1988）が波浪統計に対して解析した 4 種類のワイブル分布（形状母数$k=0.75$, 1.0, 1.4, & 2.0）と極値 I 型分布の組合せに加えて，2 母数対数正規分布を対象として検討を行った．なお，対数正規分布は極値データ x についてその自然対数 $X=\ln x$ をとり，それを正規分布として扱うので，統計的取り扱いは正規分布と同一である．

こうした極値分布関数で表される確率変数の統計的性質については，モンテカルロ法で非常に多くの標本を抽出し，それらの標本について統計解析を行うことによって解明することができる．すなわち，[0, 1] の一様乱数を電子計算機で作成させてこれを x の非超過確率，すなわち分布関数 $F(x)$ の値とし，この逆関数を求めて x とする．一様乱数の作成にあたっては，合田(1985)が 2×10^7 個までの乱数列についてランダム性を確認している乗算合同法の一手法を用いた．

今回の数値的検討では，原則として 1 万組の標本を抽出して各種の統計量を計算した．発生させる標本の大きさ，すなわちデータの総数は $N_T=10\sim800$ の範囲で変え，このうち上位のもののみを取り出す割合，すなわちデータ採択率 $\nu=N/N_T$ は，0.25, 0.5, および 1.0 の 3 種類に固定した．解析は，採択データ個数が $N=10\sim400$ を対象とした．

3. 標本中の最大値の出現特性

極値資料の標本の統計量として一番目立つのはその最大値である．特に，他のデータと懸け離れて異常に大きな値が出現した場合，その取り扱いが問題となる．既に角屋(1962)は，そうした異常値は対象とする母集団以外のデータが混入したためであるとの見方に基づき，異常値の棄却基準を提案している．これは，正規分布における標本の平均値に対する Thompson の棄却検定法（たとえば三上,1959)を応用したものである．この棄却検定法によると，N 個のデータからなる標本において任意のデータ 1 個に注目した場合，その値 x_0 が全体の平均値 x_m から離れている度合いは次式で表される．

$$|x_0-x_m|>s\sqrt{\frac{(N-1)F(1,N-2:\alpha)}{N-2+F(1,N-2:\alpha)}} \quad\cdots\cdots(1)$$

ここに，s は標本全体の標準偏差，$F(1,N-2:\alpha)$ は超過確率 α に対する自由度 $(1,N-2)$ の F 分布である．式 (1) は $x_0>x_m$ および $x_0<x_m$ の両者を同等に扱うので，$x_0>x_m$ のデータに着目するとそれが母集団において占める超過確率は $\alpha/2$ に等しい．

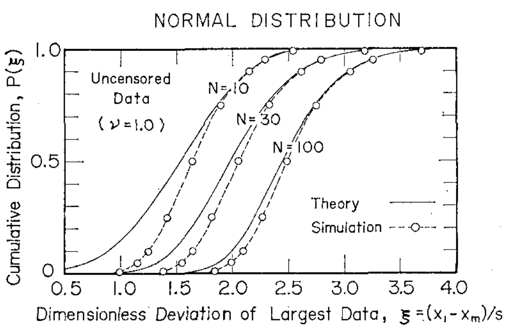
いま x_0 として標本中の最大値 x_1 を考えると，その標本中の全てのデータは x_1 を超えることがないので，N データ中の最大値 x_1 の非超過確率は $P=(1-\alpha/2)^N$ で与えられる．したがって，x_1 の 95% 非超過確率値などは式 (1) に $\alpha=2(1-0.95^{1/N})$ を代入して計算することにより求められる．

この Thompson の棄却検定法では，最大値 x_1 を平均値 x_m と標準偏差 s を使って次のように無次元化する．

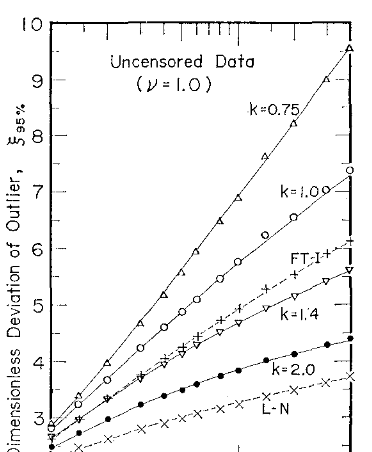
* 正会員　工博　横浜国立大学教授　工学部　建設学科
** 正会員　運輸省港湾技術研究所　海洋水理部　海象調査研究室長

$$\xi=(x_1-x_m)/s \quad \cdots\cdots(2)$$

正規分布について最大値偏差 ξ の累積分布を求めた結果の例が**図―1** である．この分布は，各標本から1個ずつ得られた1万個の ξ の値を単純に大きさの順に並べ替えて求めたものである．数値シミレーションの結果は，累積分布の下の部分で式 (2) の理論値に比べて大きめである．これは，棄却検定理論における標本平均値および標準偏差の算定において，最大値 x_1 の存在が十分に考慮されていないためと思われる．



図―1 標本中の最大値の無次元偏差 ξ の累積分布



図―2 標本中の最大値偏差の 95% 非超過確率値 $\xi_{95\%}$

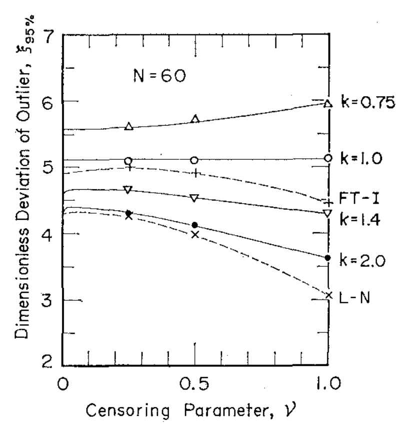
ξ の累積分布が求められると，棄却限界値としてしばしば用いられる 95% 非超過確率値などを算定することができる．**図―2** は全数極値資料に対する算定結果である．部分極値資料の結果も含め，これらに対して次の実験式をあてはめ，係数 a, b, および c を **表―1** のように設定した．

$$\xi_{95\%}=a+b\ln N+c(\ln N)^2 \quad \cdots\cdots(3)$$

あてはめは，まず分布関数およびデータ採択率ごとに式 (3) を適用して最小2乗法で係数値を求め，その結果から係数 c は2種類のデータ採択率の平均値を用いることにして a および b を再計算し，その結果をデータ採択率 ν の比較的単純な関数形として表示した．**図―3** は採択データ数 $N=60$ の場合について ν の影響を示したものである．**図―2, 3** に記入した細線は，**表―1** の値を使って計算した結果である．この実験式のあてはめ誤差は，おおむね±1%以内であり，±2%を超えるものは約200ケース中で3例だけであった．

表―1 標本中の最大値偏差の 95% 非超過確率値 $\xi_{95\%}$ の係数

分布関数	係数 a	係数 b	係数 c
ワイブル (k=0.75)	$-0.256-0.632\nu^2$	$1.269+0.254\nu^2$	0.037
ワイブル (k=1.0)	-0.682	1.600	-0.045
ワイブル (k=1.4)	$-0.548+0.452\nu^{1/2}$	$1.521-0.184\nu$	-0.065
ワイブル (k=2.0)	$-0.322+0.641\nu^{1/2}$	$1.414-0.326\nu$	-0.069
極値I型分布	$-0.579+0.468\nu$	$1.496-0.227\nu^2$	-0.038
対数正規分布	$0.178+0.740\nu$	$1.148-0.480\nu^{3/2}$	-0.035

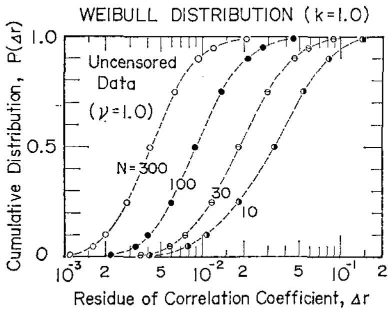


図―3 最大値偏差 $\xi_{95\%}$ に対するデータ採択率の影響

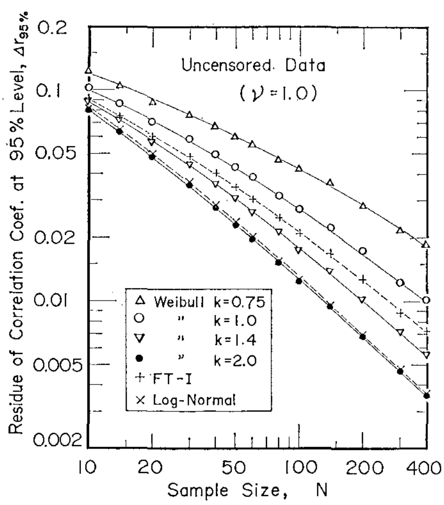
指数分布 ($k=1.0$ のワイブル分布) では，$\xi_{95\%}$ の値がデータ採択率に影響されずに一定である．これは，指数分布が ν としてどのような値を与えてもその分布特性，たとえば s/x_m の比率などが変わらないことによるものである．指数分布以外の関数については，ν の値が小さくなるにつれて分布関数の間の差異が減少する．なお，**表―1** の係数値は ν=0.25, 0.5 および 1.0 の3点のデータで定めているので，$\nu\fallingdotseq 0.15$ 程度以上の範囲で適用するのがよいと思われる．

4. 最小2乗法あてはめ時の相関係数の分布特性

極値資料の標本に対して分布関数をあてはめるには幾つかの方法があるが，中でも最小2乗法が最も簡明であり，かつ偏りのない推定値を与える．ただし，プロッティング・ポジション公式として分布関数毎に最も適切な



図—4　相関係数の残差値 Δr の累積分布



図—5　相関係数の残差の 95% 非超過確率値 $\Delta r_{95\%}$

ものを選択する必要がある．合田(1988)は，ワイブル分布に対しては修正 Petruaskas & Aagaard 公式，極値I型分布にはグリンゴルテン公式，対数正規分布にはブロム公式が最適であることを例証している．今回の検討では，これらの諸公式を使い分けている．

最小2乗法ではまた，データあてはめの適合度が，極値データ x とその基準化変量 y との間の相関係数の値として自動的に表示される．極値分布のあてはめにおける相関係数 r は一般に1に極めて近い値を取るので，適合度の判定には相関係数の1からの残差，すなわち $\Delta r=1-r$ を使うのが適切である．そこでこの相関係数の残差 Δr の分布特性を調べた1例が**図—4** である．これは指数分布から抽出した全数極値資料の場合である．当然のことながら，標本の大きさ N が増すにつれて残差 Δr が減少する．

こうした累積分布から残差 Δr の 95％ 非超過確率値を求め，標本の大きさ N との関係を調べた例を，全数極値資料の場合について**図—5** に示す．$\Delta r_{95\%}$ は $k=0.75$ のワイブル分布が最も大きく，$k=2.0$ のワイブル分布が最小である．**図—5** の $\Delta r_{95\%}$ と N との関係に対しては，$\nu=0.25$ と 0.5 の部分極値資料の結果も含め，次のような2次式をあてはめた．係数値の設定方法は $\xi_{95\%}$ の場合と同じであり，係数値として**表—2** の結果を得ている．

$$\Delta r_{95\%}=\exp\{a+b\ln N+c(\ln N)^2\} \quad \cdots\cdots\cdots\cdots(4)$$

式(4)の実験式による推定値の例は**図—5** に細線で示したが，あてはめ誤差は $\xi_{95\%}$ に対するものよりも大きく，±4％を超えるものが約200ケース中で3例ほどある．ただし，大半は±2％以下である．

表—2　相関係数の残差の 95% 非超過確率値 $\Delta r_{95\%}$ の係数

分布関数	係数 a	係数 b	係数 c
ワイブル ($k=0.75$)	$-1.473-0.049\nu^2$	$-0.2181+0.0505\nu^2$	-0.041
ワイブル ($k=1.0$)	-1.433	-0.2679	-0.044
ワイブル ($k=1.4$)	-1.312	$-0.3356-0.0449\nu$	-0.045
ワイブル ($k=2.0$)	$-1.188+0.073\nu^{1/2}$	$-0.4401-0.0846\nu^{3/2}$	-0.039
極値I型分布	-1.444	$-0.2733-0.0414\nu^{5/2}$	-0.045
対数正規分布	$-1.362+0.360\nu^{1/2}$	$-0.3439-0.2185\nu^{1/2}$	-0.035

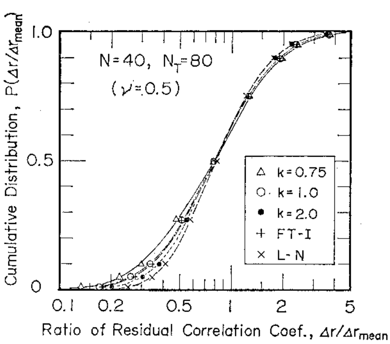
5.　分布関数の棄却基準

いままでに述べた標本中の最大値の偏差および最小2乗法適用の際の相関係数の残差値の分布特性を利用すると，データにあてはめるべき分布関数の棄却検定を行うことが可能である．最大値偏差の場合，角屋(1962)はデータを異常値とみなして棄却すべきか否かの判定に使うことを提案した．しかし，対象とする異常に大きな高波が他とまったく異なる気象原因によって発生したことが明らかでないかぎり，そのデータを極値資料から除外することには無理がある．既に**図—2** に示したように，最大値偏差の95％ 非超過確率値 $\xi_{95\%}$ は分布関数によって異なる値をとるので，対象とする標本の最大値偏差 ξ がある分布関数の $\xi_{95\%}$ を超えていたとしても，他の分布関数の $\xi_{95\%}$ 値は超えていない場合があり得る．こうした場合は，前者の分布関数のあてはめを棄却し，後者の分布関数を標本にあてはめるほうが無理がない．

このように，標本中の最大値の偏差を用いた分布関数の棄却検定の基準をここでは DOL (Deviation of Out-Lier) 基準と呼ぶことにする．同様に，相関係数の残差値を用いるものを REC (REsidue of Correlation coefficient) 基準と呼ぶ．すなわち，想定した分布関数に対して対象とする標本の ξ 値が**表—1** の係数を使って計算される $\xi_{95\%}$ 値を上回る，あるいは最小2乗法適用時の Δr が**表—2** の係数を使って計算される $\Delta r_{95\%}$ 値を上回るときは，その極値資料の標本がその分布関数で表される母集団に所属する確率が5％未満であると判断し，その分布関数のあてはめを棄却する．

6. あてはめ分布関数採択の判断基準について

最初に述べたように，極値資料に対して分布関数をあてはめ，どれか一つを採択するための基準を定めることはむずかしい．上述の二つの棄却基準で棄却されなかった関数はすべて採択される可能性を保持している．50年あるいは100年の再現期間に対する確率波高をどの分布関数を使って推定するかは主任技術者の判断によらざるを得ない．



図—6 相関係数の残差値の平均値に対する比率

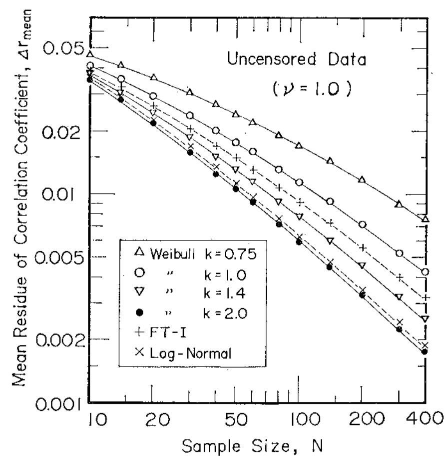
あてはめ分布関数の採択の判断基準として，最小2乗法適用の場合は相関係数の絶対値が用いられることが多く，著者の一人（合田, 1988）もその方式を推奨した．しかし，最小2乗法であてはめを行った場合の相関係数の値は，分布関数毎にその分布特性が異なるので，絶対値での比較は必ずしも適切でない．こうした分布関数毎の違いは，相関係数の残差値をその平均値で無次元化することによってかなり解消することができる．**図—6** はこれを示したもので，80 個のデータ中の上位 40 個を採択した $\nu=0.5$ の部分極値資料の例である．残差値 Δr の平均値に対する比率の累積分布は，非超過確率が 30％程度以下では分布関数による差異がみられるものの，非超過確率の大きなところでは分布関数による違いが目立たない．こうした Δr の平均値に対する比率の累積分布の形状は，標本の大きさやデータ採択率の値によって若干その分布幅が変化する程度であり，分布関数による違いが小さいことには変りがない．

あてはめ分布関数の採択にあたっては，標本に対する相関係数の値を母集団における累積分布と比較し，その非超過確率の低いものを優先させるのが合理的であろう．すなわち，$\Delta r/\Delta r_{\mathrm{mean}}$ の比率が最小のものを優先的に採択する方式が考えられる．この採択基準をここでは MIR (MIninum Ratio of residual correlation coefficient) 基準と呼んでおく．

具体的に MIR 基準を適用するためには，相関係数の残差の平均値 Δr_{mean} を定式化しておく必要がある．図

表—3 相関係数の残差平均値 Δr_{mean} の係数

分布関数	係数 a	係数 b	係数 c
ワイブル ($k=0.75$)	$-2.435-0.168\nu^{1/2}$	$-0.2083+0.1074\nu^{1/2}$	-0.047
ワイブル ($k=1.0$)	-2.355	-0.2612	-0.043
ワイブル ($k=1.4$)	$-2.277+0.056\nu^{1/2}$	$-0.3169-0.0499\nu$	-0.044
ワイブル ($k=2.0$)	$-2.160+0.113\nu$	$-0.3788-0.0979\nu$	-0.041
極値 I 型分布	$-2.364+0.054\nu^{5/2}$	$-0.2665-0.0457\nu^{5/2}$	-0.044
対数正規分布	$-2.153+0.059\nu^{2}$	$-0.2627-0.1716\nu^{1/4}$	-0.045



図—7 相関係数の残差平均値 Δr_{mean} の変化

—7 は全数極値資料の場合の数値実験結果を示したものであるが，Δr_{mean} の実験式としては部分極値資料の結果も含めて，$\Delta r_{95\%}$ と同じく式 (4) を採用し，係数については**表—3** のように定めた．**図—7** 中の細線はこれらの係数値を使って式 (4) で計算した結果である．実験式のあてはめ誤差は $\xi_{95\%}$ の場合と同程度であるが，対数正規分布では ν の影響の取込がむずかしく，±3％を超える場合が 36 ケース中で 6 例ある．

7. 波浪観測資料に対する適用事例

強風や水文統計と違い，波浪統計は観測資料が少なく観測期間も短い．現在の時点で最も大量の観測資料が整理されているのは，運輸省港湾局が中心となって推進している沿岸波浪観測体制による波浪資料である．そこで1984年末までの15か年統計（菅原ほか, 1986）として取りまとめられたデータの原資料に基づき，外海に面している 29 地点について高波の極値統計解析を行った．この結果をまとめて表—4 に示す．分布関数としてはワイブル分布 4 種類と極値 I 型分布を対象とし，対数正規分布は波浪統計ではあまり用いられないので除外している．

こうした解析の際の一つの問題点は，毎日の波浪状況の変化記録をどのように区切って気象擾乱毎の高波として定義するかである．日々の天気図を点検し，洋上のう

表―4 沿岸波浪観測資料に対する極値分布関数のあてはめ結果

地点名	有効年数 K	データ数 N	採択率 ν	棄却検定 ABCDE	ξ 値	MIR 基準での採択関数 分布	相関係数	REC 比
紋別	7.4	75	0.42	○○○○○	3.62	C	0.99518	0.44
留萌	13.1	284	0.44	▲○○○○	3.70	C	0.99500	1.30
瀬棚	4.3	153	0.73	▲▲○○○	3.07	D	0.99758	0.53
深浦	4.2	156	0.78	▲○○○○	3.58	C	0.99720	0.48
酒田	10.1	361	0.72	▲▲▲○▲	3.19	D	0.99629	1.73
弾崎	5.7	173	0.62	▲▲▲○▲	2.69	D	0.99406	1.37
輪島	5.9	196	0.68	▲▲○○○	3.13	D	0.99573	1.13
金沢	12.0	263	0.44	▲▲▲○▲	3.02	D	0.99494	1.56
福井	3.6	94	0.52	○○○○○	3.18	C	0.99246	0.83
鳥取	4.9	143	0.60	○○○○○	4.75	E	0.99578	0.49
浜田	8.0	194	0.49	▲○○○○	4.22	C	0.99738	0.50
玄界灘	3.0	78	0.52	○○○○○	3.00	E	0.98340	1.27
名瀬	7.6	166	0.44	▲▲○○▲	3.16	D	0.99365	1.32
那覇	9.9	283	0.59	▲○○▼○	5.11	E	0.99522	0.92
苫小牧	13.9	213	0.47	○○○▲○	4.78	B	0.99629	0.56
むつ小川原	10.4	250	0.50	▲▲○○▲	3.07	D	0.99589	1.25
八戸	10.6	368	0.61	▲○○▲○	4.14	E	0.99609	0.93
宮古	3.0	138	0.77	○○▲▲▲	4.05	B	0.98858	1.23
釜石	3.3	81	0.43	○○○○○	2.91	C	0.99174	0.80
仙台	5.3	176	0.57	○○○▲○	3.94	B	0.99388	0.78
小名浜	3.7	102	0.49	○○○▲○	3.48	B	0.98827	1.03
常陸那珂	4.8	121	0.43	○○○▲○	3.22	B	0.98814	1.17
鹿島	9.4	407	0.73	▲○○▲○	4.71	B	0.99767	0.55
波浮	9.9	401	0.60	○○▲■○	5.37	B	0.99790	0.49
潮岬	11.8	365	0.56	○○○■○	4.94	E	0.99729	0.63
高知沖	3.6	149	0.75	○▲▲■▲	5.08	A	0.99310	0.53
油津	7.5	71	0.18	○○○■○	4.94	A	0.98878	0.65
志布志	4.5	91	0.41	○▲▲■▲	5.22	A	0.97921	1.29
中城湾	8.9	270	0.68	○○▲▲▲	4.74	A	0.98952	1.17

注 1) 分布関数: A＝ワイブル (k=0.75), B＝ワイブル (k=1.0)
C＝ワイブル (k=1.4), D＝ワイブル (k=2.0)
E＝極値 I 型分布.

注 2) 棄却判定: ○＝棄却基準に該当せず, ▼＝DOL 基準に該当
▲＝REC 基準に該当, ■＝両棄却基準に該当.

注 3) REC 比: $\Delta r/\Delta r_{\mathrm{mean}}$

ねりも考慮して1回毎の高波を識別することが望ましいけれども，地点数が多い場合は作業量が膨大となる．ここでは，地点毎に目分量で有義波高の限界値を設定し，この限界波高を連続して超えている部分を一つの高波の継続期間とみなした．ただし，この期間中に二つ以上のピークが見られる場合には天気図を参照し，複数の高波とみなすべきかどうかを判断した．

観測期間中の高波の総数 N_T もまた，算定がむずかしい問題である．ここでは，高波をかなり綿密に拾い上げた地点の年間平均発生個数を参考にし，日本海沿岸では年間50個程度，太平洋沿岸では50～60個を目安として，これを有効観測年数に乗じた概数を高波の総数とした．

表―4 では，海域毎の高波の極値分布特性がかなり明瞭である．特に，棄却基準に該当しないで残る分布関数に着目すると，日本海沿岸では $k=2.0$ のワイブル分布が最有力であり，東支那海沿岸に入るとワイブル分布でも $k=1.4$ の可能性が強くなる．これらの海域では吹送距離が制限されていることが理由であろう．太平洋沿岸では，鹿島以北では指数分布 ($k=1.0$) あるいは $k=1.4$ のワイブル分布が有力であるが，台風の影響を直接に受ける波浮以西では $k=0.75$ のワイブル分布である可能性が最も高い．もっともこの海域の場合には，台風による高波と低気圧による高波とで極値分布の母集団が異なる可能性があり，両者を混ぜて解析したことによる見掛けの現象かもしれない．今後さらに観測資料が蓄積されるのを待って検討する必要がある．なお，こうした分布関数の棄却判定はデータの個数が多くなるほど感度が向上する．できるだけ観測期間の長い資料に基づき，高波の識別を綿密に行ってデータ個数 N を増すとともに，データ採択率 ν を高めることが望まれる．

8. あとがき

今回の検討により，今まで主観的な判断に偏りがちであった極値分布関数のあてはめがかなり客観的に行い得るようになったと思われる．こうした方式によって高波の出現確率の推定精度が今後さらに向上することを期待する次第である．

参考文献

小野澤昌己・合田良実 (1989): 極値II型分布の特性とその年最大風速資料に対する適用性について，第44回年講 (I).
角屋 睦 (1962): 異常 (確率) 水文量とデータの棄却検定，農業土木研究，別冊第3号，pp. 23-27.
合田良実 (1985): 波浪の統計的性質に関する二,三の数値的検討，港研報告，24巻4号，pp. 65-102.
合田良実 (1988): 極値統計におけるプロッティング公式ならびに推定値の信頼区間に関する数値的検討，港研報告，27巻1号，pp. 31-92.
合田良実・李 孝秀 (1989): 極値分布関数の棄却基準と年最大日雨量資料への適用について，第44回年講 (II).
三上 操 (1959): 応用推計学，内田老鶴圃，pp. 270-273.
菅原一晃ほか (1986): 沿岸波浪観測15か年統計 (昭和45～59年)，港研資料，No. 554，872 p.