

波浪統計における確率波高の信頼区間

合 田 良 実*

1. ま え が き

波浪統計の一つの目的は、台風や低気圧による高波の統計資料を解析して、再現期間 50 年などに対する確率波高を求め、設計波高の選定資料を提供することである。このための手法として種々の極値統計分布を適用する方法が幾つか開発され、実際問題にも使用されている。高波の極値統計資料が与えられれば 50 年確率波高などが一義的に求められるため、確率波高の算定方法は確立されているかのような印象を与えている。

しかしながら、統計資料は常に母集団からの抽出誤差を伴うものであり、確率波高の算定結果も何らかの推定誤差を免れない。こうした誤差の問題については今までほとんど議論されておらず、確率波高の算定結果だけが一人歩きしているきらいがある。

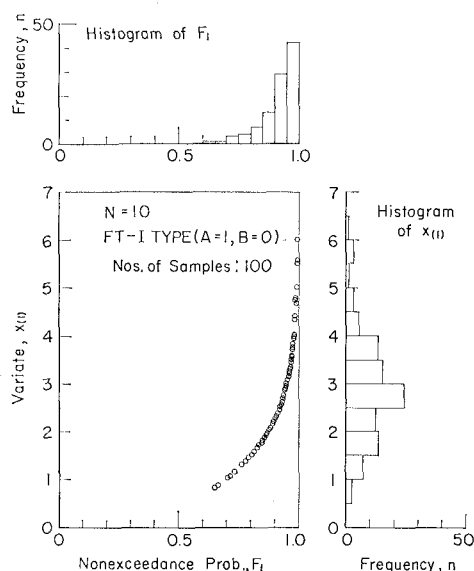
本研究は、理論的には解明困難である確率波高の信頼区間について大規模な数値シミュレーションを実施し、信頼区間の実用的な推定法の提案式を取りまとめたものである。なお、この研究は著者が運輸省港湾技術研究所に在職中に行ない、同所報告として発表したもの¹⁾に基づいていることをお断りする。

2. 極値統計資料の変動性

極値統計は、対象とする極値データがある一つの確率分布に従っており、この確率分布が長年月の間にも変わらないことを前提としている。したがって、気候の長期変動の問題は検討の対象外である。データが確率分布に従うということは、われわれが手にしている 1 組の極値データがある一つの母集団からランダムに抽出された一つのサンプルであることを意味する。データを取得した期間が波浪でいえば比較的穏やかな時期であったかもしれないし、高波が頻発した時期であったかもしれない。そのいずれであるかは、データを取得した時点では知ることができない。統計的なサンプルである以上、その統計量は母集団の値のまわりにある幅で変動する。

図—1 はこうした極値データの変動性の例示である。

これは極値統計分布としてしばしば用いられる Fisher-Tippett I 型分布 (Gumbel²⁾ が詳しく解説したところからグンベル分布ともいわれる、以下 FT-I 型と略称) の一つの母集団を設定し、数値計算でこの母集団から 10 個ずつのデータをランダムに 100 組抽出し、各組の最大値の分布を調べたものである。図は各サンプル中の最大値 $x_{(1)}$ およびそれが母集団において占めている非超過確率 F_1 の頻度分布を表している。この例では、最大値 $x_{(1)}$ が 0.84~6.04 の範囲に広く分布しており、サンプルの数を増すと分布幅はさらに広がる。



図—1 サンプル中の最大値の頻度分布の例

最大値ではなく、各サンプルの標準偏差を調べた例が図—2 である。ここではワイブル分布のうちの $k=1.0$ のものについて、個数 $N=10$ および 100 のサンプルを 10,000 組ランダムに抽出し、各サンプルの標準偏差 σ_x を母集団の値 ($\sigma_x)_0$ と比較してその累積分布を調べた結果である。たとえばデータ個数が $N=10$ と小さいサンプルでは、 σ_x が母集団値の 0.4 倍以下から 1.9 倍以上にまで広く変化する。

この二つの例のように、サンプルの特性は確率のいた

* 正会員 工博 横浜国立大学教授 工学部建設学科

ずらによって大きく変化する。小さなデータの多いサンプルに分布関数をあてはめて確率波高を算定すれば、低い推定値を得るし、大きなデータの多いサンプルであれば、高めの確率波高を推定することになる。

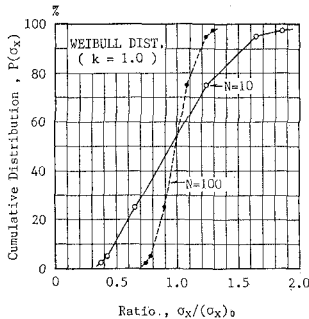


図-2 サンプルの標準偏差の累積度数分布の例

本論文で対象としたのは、式(1)のFT-I型分布と式(2)のワイブル分布である。

$$F(x) = \exp\{-\exp[-(x-B)/A]\} \quad (1)$$

$$F(x) = 1 - \exp\{-(x-B)/A^k\} \quad (2)$$

ここに、 $F(x)$ は極値データ x の非超過確率、すなわち分布関数、 A は尺度母数、 B は位置母数、 k は形状母数である。ワイブル分布は形状母数をあらかじめ固定して取り扱っている。

これらの分布関数に対して、 $0 \leq F \leq 1$ の非超過確率を一様乱数を使ってランダムに選り、これに対する x の値を式(1)、(2)を解いて求め、これをあらかじめ設定した回数だけ繰り返す。この操作で極値データの1組のサンプルが得られるので、種々の統計解析を行なって必要な統計量を算出する。サンプル毎の変動性は以上の操作を繰り返してサンプルの組数を増すことで明らかにできる。本論文では原則として10,000組を用いた。

3. 母分布関数が既知の場合の確率波高の変動

(1) 分布関数の母数および確率波高の推定

1組の極値データが得られれば、まずこれらを大きさの順に並べ替えて順序統計量を作る。ここでは波浪の極値解析のやりかたに従い、最大値を1位とした(すなわち、 $m=1$)。各順序統計量の非超過確率は、FT-I型分布に対しては式(3)のグリンゴルテン公式³⁾、ワイブル公式に対してはペトルアスカア・アーガードの提案⁴⁾を著者¹⁾が若干修正した式(4)で推定する。

$$\hat{F}_m = 1 - (m - 0.44)/(N_T + 0.12) \quad (3)$$

$$\hat{F}_m = 1 - (m - \alpha')/(N_T + \beta) \quad (4)$$

ここに、

$$\alpha' = 0.20 + 0.27/\sqrt{k}, \quad \beta = 0.20 + 0.23/\sqrt{k}$$

また、 m は降順の順位、 N_T は対象とする統計期間中に

発生したと想定される極値データの総個数である。実際のデータ個数 N ではなく N_T を使うのは、波浪統計においては波高の低いデータを棄却した部分極値資料を用いることが多いことを考慮し、その場合でも母分布関数の形状をできるだけ復元できるように考えたものである。この N と N_T との比率をここではデータ採択率 ν と定義する。すなわち、 $\nu = N/N_T$ である。数値シミュレーションでは、あらかじめ設定したデータ採択率の値に応じて上位のデータのみを採択する。

順序統計量 $x_{(m)}$ の非超過確率が推定されれば、これに対して $x_{(m)}$ の基準化変量 $y_{(m)}$ をFT-I型分布は式(5)、ワイブル分布は式(6)で計算する。

$$y_{(m)} = -\ln[-\ln \hat{F}_m] \quad (5)$$

$$y_{(m)} = [-\ln(1 - \ln \hat{F}_m)]^{1/k} \quad (6)$$

そして、 $x_{(m)}$ と $y_{(m)}$ の間に次の直線関係を仮定し、最小2乗法を適用して母数推定値 $\hat{A}$ 、 $\hat{B}$ を得る。

$$x_{(m)} = \hat{A}y_{(m)} + \hat{B} \quad (7)$$

指定された再現期間 R に対する確率波高 x_R は、その基準化変量 y_R を式(8)、(9)で計算し、それを式(7)に代入することによって求められる。

$$y_R = -\ln[-\ln[1 - 1/(\lambda R)]] \quad (8)$$

$$y_R = [\ln(\lambda R)]^{1/k} \quad (9)$$

ただし、 λ は極値データの平均発生率であり、 K 年間に N_T 個発生したのであれば、 $\lambda = N_T/K$ である。

(2) 確率波高の分散例

図-3は、FT-I型分布からの数値サンプルに対して同じくFT-I型をあてはめて $R=100$ に対する確率波高の推定値 $\hat{x}_R$ を求め、真値 x_R との差をサンプルの標準

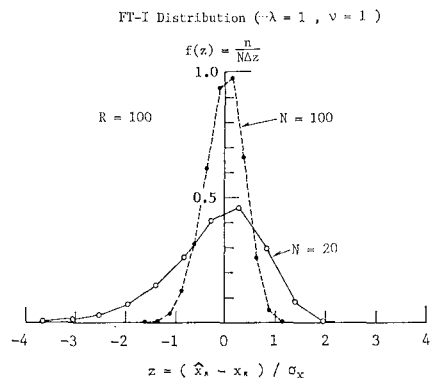


図-3 確率波高の推定値の分散例

偏差 σ_x で除して求めた無次元量、すなわち

$$z = (\hat{x}_R - x_R)/\sigma_x \quad (10)$$

の頻度分布である。図の横軸が推定値の偏差の無次元量 z 、縦軸は頻度を確率密度の形で表したものである。平均発生率は $\lambda=1$ 、データ採択率は $\nu=1$ である。 $N=20$ の場合は、いわば20年間の毎年最大値データに基づい

て100年確率波高を推定しているケースに相当する。

この場合は、確率波高の推定値が真値と $2\sigma_x$ 以上の差を生じることがある。データ個数が100となり統計期間と再現期間が等しい場合は偏差が小さいが、それでもデータの標準偏差の1倍までの変動はある。

(3) 確率波高の推定値の標準偏差

こうした確率波高の推定値の統計的変動性に関してはGumbel²⁾ (6.2.3節) がFT-I型分布 ($\nu=1$ の場合) に対して次の理論式を与えている。

$$\sigma^2(\hat{x}_R) = \sigma^2[1 + 0.8885(y_R - \gamma) + 0.6687(y_R - \gamma)^2]/N \quad \dots\dots\dots(11)$$

ここに、 σ は母集団の標準偏差、 γ はオイラーの定数 (0.5772...) である。母集団の値 σ は未知であるので、式 (11) はこのままでは使えないが、式の形からみて確率波高の推定値の誤差の無次元量 z の標準偏差 σ_z と対応させることは可能である。図-4 は、データ個数 N の影響を小さくするため、FT-I型分布について数値実験で得られた無次元量 z の標準偏差を $\sqrt{N}\sigma_z$ の形で表示したもので、横軸は基準化変量 y_R である。

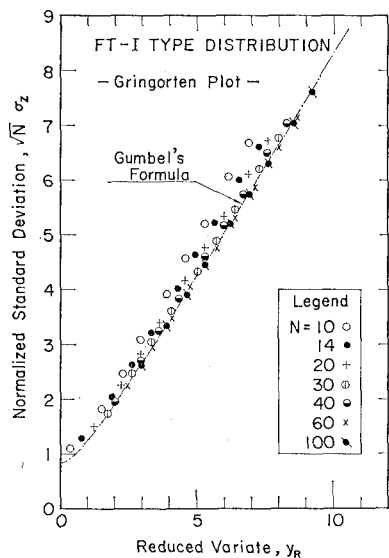


図-4 FT-I型分布の確率波高の標準偏差

図中の1点鎖線は式(11)の σ を σ_x で代替可能であるとして計算したもので、データのはほぼ下限値を与えている。しかし、データ個数の影響がまだ残っており、式 (11) のままでは不十分である。さらに、実際問題で扱う $\nu < 1$ の部分極値資料に対する算定式も必要である。また、ワイブル分布に対しては確率波高の推定値の標準偏差に関する理論式が発表されていない。そこで、次の範囲の条件で数値実験を行ない、その結果を整理した。

分布関数: FT-I型およびワイブル分布
($k=0.75, 1.0, 1.4, \& 2.0$)

データ個数: $N=10, 14, 20, 30, 40, 60, \& 100$

データ採択率: $\nu=1.0, 0.50, \& 0.25$

試行錯誤の結果、次の実験式を採用することとした。

$$\sigma_z = [1.0 + a(y_R - c + \alpha \ln \nu)^2]^{1/2} / \sqrt{N} \quad \dots\dots\dots(12)$$

この式の中の係数 a は次のように与えた。

$$a = a_1 \exp[a_2 N^{-1.3} + \kappa(-\ln \nu)^{1/2}] \quad \dots\dots\dots(13)$$

以上の式 (12), (13) の中の係数 a_1, a_2, c, α および κ は表-1 のように定めた。

表-1 母分布関数既知の場合の確率波高推定値の標準偏差に関する実験式の係数

分 布 関 数	a_1	a_2	κ	c	α
ワイブル ($k=0.75$)	1.65	11.4	-0.63	0	1.15
同上 ($k=1.0$)	1.92	11.4	0	0.3	0.90
同上 ($k=1.4$)	2.05	11.4	0.69	0.4	0.72
同上 ($k=2.0$)	2.24	11.4	1.34	0.5	0.54
FT-I型	0.64	9.0	0.93	0	1.33

数値実験データに対する式 (12), (13) の適合度の1例を図-5に示す。これは $k=0.75$ のワイブル分布から抽出した部分極値データ ($\nu=0.25$) に対する結果であり、図中の記号は数値実験値、破線等は実験式を示している。 $N=100$ のデータに対して実験式がやや過大であるが、全般的には両者の一致は良好である。他の分布関数およびこれと異なるデータ採択率の数値実験値に対しても同程度の一致が得られている。

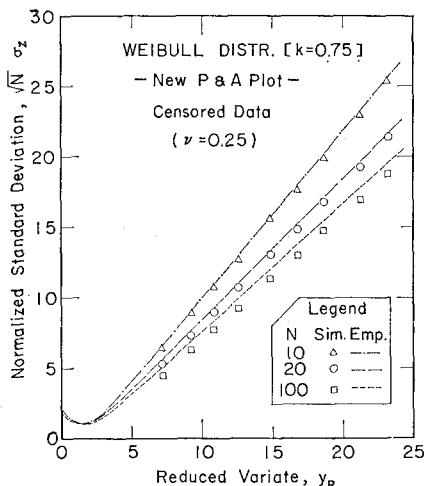


図-5 部分極値資料に対する実験式の適合例

確率波高の推定値の標準偏差の絶対値 $\sigma(\hat{x}_R)$ は、式 (10) の形から近似的に、次のように算定できる。

$$\sigma(\hat{x}_R) = \sigma_z \sigma_x \quad \dots\dots\dots(14)$$

概略的には、データ個数の平方根に逆比例して確率波高の推定誤差が減少することになる。

4. 母分布関数が未知の場合の確率波高の標準誤差

(1) 分布関数あてはめの問題点

前節の検討は、母分布関数が既知の場合である。しかし、実際の波浪などの極値統計の問題では、母分布関数そのものが未知である。経験的あるいは先験的に特定の分布関数、たとえば FT-I 型や対数正規型を採用することが少なくないが、理論的に根拠がない。客観的には複数の候補の中から、対象とするデータに最も適合するものを採用するのがよいと考えられる。

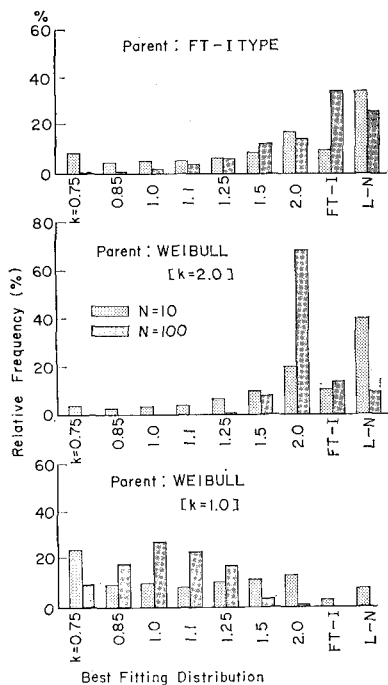


図-6 極値データに対する分布関数の適合度の例

Petruaskas & Aagaard⁴⁾ は FT-I 型分布と 7 種類のワイブル分布を候補として極値データへのあてはめを試みることを提案し、これが我が国の波浪解析でも用いられている。この方式の妥当性を検討するため、ある母集団から抽出された極値データに複数の分布関数をあてはめたとき、母関数にどの程度復帰するかを調べたのが図-6 である。母分布関数は上から順に FT-I 型、ワイブル分布の $k=2.0$ および 1.0 であり、データ個数が $N=10$ と 100 で全データを採択した $\nu=1$ の場合を示している。横軸はあてはめ関数の略記号で、L-N は対数正規分布である。縦軸はサンプルに対して最適と判定された分布関数の出現率である。分布関数のあてはめは式 (7) に対する最小 2 乗法に基づき、その際の $x(m)$ と $y(m)$ との相関係数の最大のものを最適と判定した。

図-6 によると、FT-I 型分布から $N=10$ 個だけ抽出

したサンプルの場合などは、母関数に復帰するものが 10% に過ぎず、34% は対数正規分布が最も適合すると判定されてしまう。 $k=2.0$ のワイブル分布はデータ個数が増えると母関数への復帰率が高いけれども、データ個数が少ないと対数正規分布と見誤られる可能性が高い。このように母関数への復帰率が高くないのは、図-2 で例示したように、サンプル毎のデータの標準偏差が大きく変動するためである。図-7 はこの 1 例である。これは $k=1.0$ のワイブル分布の母集団から $N_T=80$ のデータを抽出し、その上位 40 個を採択して極値データとしたケースである。階段状の折れ線はサンプル 10,000 組の標準偏差の分布、各記号を結ぶ破線はこれらのサンプルのうち特定の分布関数が最適と判定されたものについて調べた標準偏差の分布である。あてはめ関数の候補は FT-I 型および 4 種類のワイブル分布 ($k=0.75, 1.0, 1.4, \& 2.0$) であり、ここではそのうち 3 分布についての結果を示している。明らかに、 $k=2.0$ の分布に適合するサンプルは標準偏差の小さいものが多く、 $k=0.75$ に良く適合するものは標準偏差が大きい。

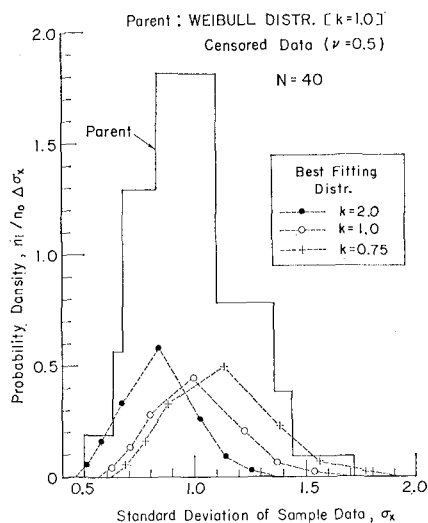


図-7 あてはめ関数別のサンプルの標準偏差の分布

このような極値データのサンプルの統計的変動性の影響を考えるならば、あてはめ関数の候補はその数を制限しておく必要があるといえる。この研究では、FT-I 型と 4 種類のワイブル分布 ($k=0.75, 1.0, 1.4, \& 2.0$) を候補として採用している。

(2) 確率波高の推定値の標準誤差の実験式

母分布関数が未知の場合には、最適と判定された分布以外の関数の可能性を考慮に入れる必要がある。この研究では、5 種類の母分布関数が平等に存在するものとしてこれらから抽出された全サンプル ($N=10, 20, \& 40$ は 50,000 組, $N=100$ は 25,000 組, $N=200$ は 10,000

組) に対してあてはめを行なった。そして、あてはめられた関数毎に確率波高の推定値の分散および真値からの偏りを計算した。推定値の偏り補正については文献¹⁾を参照していただくことにして、この補正を行なった後の推定値の標準誤差を示した1例が図-8である。

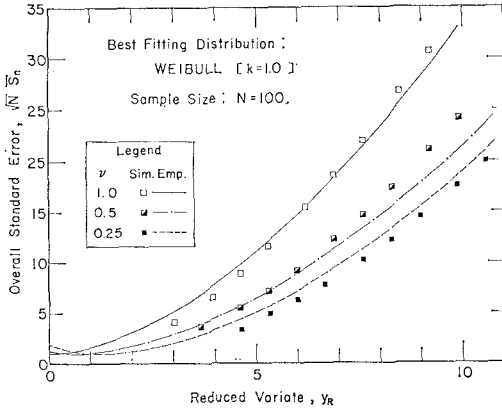


図-8 母分布関数未知の場合の確率波高の標準誤差の例

図の縦軸は式(15)で定義する標準誤差の無次元量であり、各記号は数値実験の結果である。

$$\sqrt{N} S_n = \{E[(\hat{x}_{R,n} - x_R)^2 / \sigma_x^2] - \bar{Z}_n^2\}^{1/2} \dots (15)$$

ここに、 $\hat{x}_{R,n}$ は n 番目の分布関数をあてはめて算定した R 年確率波高の推定値、 x_R は真値、 σ_x は極値データのサンプルの標準偏差、 $\bar{Z}_n$ は偏り補正量の無次元量であり、 $E[\]$ はアンサンプル平均を表す。

表-2 母分布関数未知の場合の確率波高の推定値の標準誤差に関する実験式の係数

分布関数	ν	b_1	b_2	N_c	p	α
ワイブル ($k=0.75$)	1.0	0.57	0.18	20	1.2	2.7
	0.5, 0.25	0.41	0.22	20	1.2	2.7
同上 ($k=1.0$)	1.0	0.55	0.15	15	1.7	1.0
	0.5, 0.25	0.38	0.17	20	1.7	1.0
同上 ($k=1.4$)	1.0	0.37	0.08	1000	2.3	0.5
	0.5, 0.25	0.46	0.09	20	2.3	0.5
同上 ($k=2.0$)	1.0	0.30	0.36	80	3.2	0.35
	0.5, 0.25	0.56	0.20	100	3.2	0.35
FT-I型	1.0	0.24	0.36	80	1.6	0.9
	0.5, 0.25	0.46	0.14	50	1.6	0.9

このような数値実験結果から試行錯誤によって次のような標準誤差推定の実験式を取りまとめた。

$$S_n = \{1.0 + A_s[y_R + \alpha \ln \nu] / \sqrt{N}\} \dots (16)$$

式中の係数 A_s はさらに式(17)のように表し、その係数を α および指数 p とともに表-2のように設定した。

$$A_s = b_1 + b_2 [\log_{10}(N/N_c)]^p \dots (17)$$

以上の実験式は、全データ採用の $\nu=1$ の場合は数値

実験結果にかなり良く一致するが、 $\nu=0.5, 0.25$ の部分極値資料の場合には、実験式の適合性がやや低下する。

5. 確率波高の推定値の信頼区間

確率波高の推定値の分布は図-3にその1例を示したように、必ずしも正規分布に従っていない。しかし、実用上の観点からこれを正規分布で近似させると、推定値の信頼区間は1シグマ、2シグマ限界などで表示することができる。推定値の標準誤差は次式で求めればよい。

$$\sigma(\hat{x}_R) = S_n \sigma_x \dots (18)$$

図-9は、ある港での有効統計年数10.74年にわたる波高計による台風の高波の観測結果に基づき、確率波高を推定した例である¹⁾。推定値の信頼区間の幅は当然のことながら、再現期間が長くなるにつれて広がる。

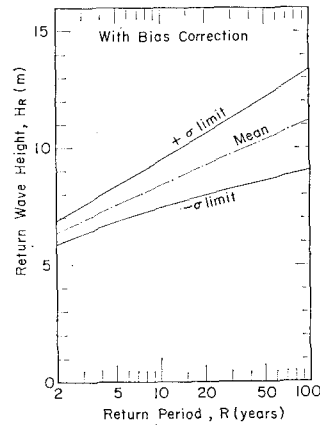


図-9 確率波高の推定値の信頼区間の例

6. あとがき

本論文で例示したように、極値統計はサンプル毎の変動が大きい。与えられたサンプルに最も適合する極値分布をあてはめても、それが真の母集団の分布関数を表しているという保証はない。波浪統計データを使って確率波高を算定し、これに基づいて設計波高を決める場合には、推定値の統計的誤差を考慮し、ある幅の信頼区間の中で主任技術者の判断で選定しなければならない。これについては、今後さらに検討していく必要がある。

参考文献

- 1) 合田良実: 極値統計におけるプロット公式ならびに推定値の信頼区間に関する数値的検討, 港湾技術研究所報告, 27-3, 31-92, 1988.
- 2) Gumbel, E. J.: *Statistics of Extremes*, Columbia Univ. Press., New York, 1958.
- 3) Gringorten, I. I.: A plotting rule for extreme probability paper, *J. Geophys. Res.*, 68-3, 813-814, 1963.
- 4) Petruaskas, C. and P. M. Aagaard: Extrapolation of historical storm data for estimating design wave heights, *Prep. 2nd OTC* 1190, 1970.