

【基調講演】

地域頻度解析に基づく極値統計解析の
新手法について

(株) エコー 顧問

横浜国立大学 名誉教授

合田 良実

土木学会
水工学委員会・海岸工学委員会

2011 年 8 月

地域頻度解析に基づく極値統計解析の新手法について

A New Methodology of Extreme Data Analysis with Regional Frequency Analysis

合 田 良 実

Yoshimi Goda

1. はじめに

土木防災施設の設計においては、所定の再現期間に対する外力条件を最初に設定する必要がある。そのため、最大風速、高水流量、高潮偏差、高波時の有義波高などの極値データを長期間にわたって収集し、それらの出現確率を推計する。いわゆる極値統計解析である。外力の極値データとしては、年最大値 (Annual Maximum, AM)、またはある閾値を超える外力の時間変化のなかでの極大値 (Peaks-over-Threshold, POT) のいずれかが用いられる。前者は年最大値資料、後者は極大値資料であり、ときに部分的水文資料 (Partial duration series) と呼ばれることがある。水文統計では前者を用いることが多いようであるが、極値波浪統計では後者が標準的である。同じ統計データから抽出した年最大値資料と極大値資料とでは、極値の累積分布関数 (Cumulative Distribution Function, CDF) がやや異なり、所定の再現期間に対する確率統計量 (Quantile) に差異がある。これについては7節で論述する。

極値データの資料 (標本) に適合する CDF を推計する方法としては、積率法、最小2乗法、最尤法、確率重み付き積率法 (PWM) などが提案され、それぞれの優劣が議論されてきた。近年は、Hosking(1990) が PWM 法の改良として提案した L -モーメント法が米国を中心に広く用いられており、さらにある地域内の複数地点のデータを解析して地域共通の CDF を推計する地域頻度解析にも活用されている。わが国では、外山・水野 (2002) が全国のアメダス地点の確率降水量をこの手法で解析している。しかし、これ以外の極値事象については L -モーメント法に基づく地域頻度解析の事例が少ないと思われる。本稿では、著者が最近に検討してきた有義波高の極値統計解析への適用を中心に、 L -モーメント法を用いた地域頻度解析手法について紹介する。

2. 極値有義波高に対する累積分布関数

極値データの累積分布関数 (CDF) については多様な関数が提案され、極値データの標本への当てはめが試みられてきた。しかしながら、極値波高その他の極値データがいかなる母分布に従うかについては理論的根拠が皆無であり、極値の母分布関数は未知と言わざるを得ない。Gumbel(1953) は二重指数分布が数多くの極値事象に適用できることを例示したけれども、理論的裏付けは特定の初期分布に限定される。

極値事象に関しては、特定の CDF を母分布関数として設定する方式と、複数の CDF を候補として取り上げ、標本に最も適合するものを母分布関数とみなす方式とがある。著者は後者の方式を採択しており、本稿でもこれに従う。母分布関数の候補として取り上げるのは、一般化極値分布 (Generalized extreme-value: GEV)、一般化パレート分布 (Generalized Pareto: GPA)、およびワイブル分布の3分布である。これらはいずれも形状母数 (k)、尺度母数 (A)、および位置母数 (B) の3個の母数でその特性が定められる。3母数型の CDF としてはピアソンⅢ型分布もある。しかし、後述する L -モーメント法における L -skewness と L -kurtosis の相関関係をみると、ピアソンⅢ型分布の特性はワイブル分布と類似しており、後者で十分に代替可能である。本稿で取り上げる3分布の関数形等を表-1に示す。

表-1 極値波高の母分布関数の候補としての累積分布関数

関数名	累積分布関数: $F(x)$	基準化変量: y_i	平均: $\mu[x]$	分散: $\sigma^2[x]$
一般化極値 (GEV)	$\exp\left[-\left\{1-k\frac{x-B}{A}\right\}^{1/k}\right] \quad k \neq 0$	$[1 - (-\ln P_i)^k]/k$	$B + \frac{1}{k}A\left[\frac{1 - \Gamma(1+k)}{\Gamma(1+k)}\right]$	$\frac{A^2}{k^2}\left[\frac{\Gamma(1+2k)}{\Gamma^2(1+k)}\right]$
一般化パレ ート (GPA)	$1 - \left[1 - k\frac{x-B}{A}\right]^{1/k} \quad k \neq 0$	$[1 - (1 - P_i)^k]/k$	$B + \frac{A}{1+k}$	$\frac{A^2}{(1+k)^2(1+2k)}$
ワイブル (WBL)	$1 - \exp\left[-\left(\frac{x-B}{A}\right)^k\right]$	$[-\ln(1 - P_i)]^{1/k}$	$B + A\Gamma\left(1 + \frac{1}{k}\right)$	$A^2\left[\frac{\Gamma\left(1 + \frac{2}{k}\right)}{\Gamma^2\left(1 + \frac{1}{k}\right)}\right]$

なお、ここでは GEV と GPA について $k=0$ のケースを外している。これは、標本に当てはめて母数を推計したときには、形状母数が如何に微小であっても $k=0$ の結果を得ることがないためである。

CDF は累積確率あるいは非超過確率を表し、 $[0, 1]$ の値を取る。それを確率変数 x で微分したものが確率密度関数である。その右側が長く裾を引いているときには、異常に大きな極値が出現する可能性があり、土木施設の設計で慎重に対処する必要がある。確率密度関数の右裾の広がり、合田 (2002) が提案した裾長さ、すなわち $\gamma_{50} = H_{50}/H_{10}$ を用いて表示可能である。ここに、 H_{50} と H_{10} はそれぞれ 50 年および 10 年確率波高である。極値有義波高に関しては $\gamma_{50} = 1.13 \sim 1.27$ の値を取ることが知られている (合田 2008)。

CDF の 3 母数のうち、尺度母数 A と位置母数 B は恣意的に設定できるものではない。合田 (2002, 2008) が論じたように、50 年確率波高 H_{50} と裾長さ γ_{50} が与えられると、 A と B は式 (1) によって自動的に定まる。

$$\frac{A}{H_{50}} = \frac{1 - 1/\gamma_{50}}{\gamma_{50} - \gamma_{10}}, \quad \frac{B}{H_{50}} = 1 - \frac{1 - 1/\gamma_{50}}{1 - \gamma_{10}/\gamma_{50}} \quad (1)$$

ここに、 γ_{50} と γ_{10} はそれぞれ再現期間 $R = 50$ と 10 年に対する基準化変量であり、表-1 において非超過確率を $P_i = P_R = 1 - 1/(\lambda R)$ と設定して算定される。ここに λ は平均発生率すなわち年間の極値の平均出現回数である。したがって、 H_{50} と γ_{50} が同じであっても、平均発生率が異なると尺度母数と位置母数の値が異なることに注意する必要がある。

3. L -モーメントと CDF の母数推定

L -モーメントは、Greenwood ほか (1978) が提唱した確率重み付き積率 (PWM) の線形和として Hosking (1990) が提案したもので、その理論と応用については Hosking・Wallis (1997) が詳しく解説している。本稿では紙数の制限のため説明を割愛するので、合田 (2009) などを参照されたい。1 次の L -モーメント (λ_1) は平均値であり、積率の次数に応じて 2 次 (λ_2)、3 次 (λ_3)、4 次 (λ_4) 等のモーメントが定義される。さらに、 L -モーメントの比として L -CV ($\tau = \lambda_2/\lambda_1$)、 L -skewness ($\tau_3 = \lambda_3/\lambda_2$)、および L -kurtosis ($\tau_4 = \lambda_4/\lambda_2$) も定義される。

CDF が与えられると、 L -モーメントおよびその相互比が理論的に計算できる。表-2 は、本稿で扱う 3 種類の分布関数の 2 次 L -モーメントならびに L -skewness と L -kurtosis の理論値を示す。このうち GEV と GPA については Hosking・Wallis (1997) によるもので、ワイブル分布に関しては合田 (2009, 2010, 2011) が導いたものである。

表-2 L-モーメント諸量の理論値

分布関数	λ_2	L-skewness (τ_3)	L-kurtosis (τ_4)
GEV	$\lambda_2 = A(1 - 2^{-k}) \times \Gamma(1 + k)/k$	$2(1 - 3^{-k})/(1 - 2^{-k}) - 3$	$[5(1 - 4^{-k}) - 10(1 - 3^{-k}) + 6(1 - 2^{-k})]/(1 - 2^{-k})$
GPA	$\lambda_2 = A/[(1 + k)(1 + 2k)]$	$(1 - k)/(3 + k)$	$(1 - k)(2 - k)/[(3 + k)(4 + k)]$
ワイブル	$\lambda_2 = A(1 - 2^{-1/k}) \times \Gamma(1 + 1/k)$	$3 - 2(1 - 3^{-1/k})/(1 - 2^{-1/k})$	$[5(1 - 4^{-1/k}) - 10(1 - 3^{-1/k}) + 6(1 - 2^{-1/k})]/(1 - 2^{-1/k})$

表-3 分布関数の母数推計式

分布関数	形状母数: $\hat{k}$	尺度母数: $\hat{A}$	位置母数: $\hat{B}$
GEV	$\approx 7.8590c + 2.9554c^2$ $c = 2/(3 + t_3) - \log 2 / \log 3$	$l_2 \hat{k} / [(1 - 2^{-\hat{k}}) \times \Gamma(1 + \hat{k})]$	$l_1 - \hat{A}[1 - \Gamma(1 + \hat{k})] / \hat{k}$
GPA	$(1 - 3 t_3)/(1 + t_3)$	$l_2(1 + \hat{k})(2 + \hat{k})$	$l_1 - l_2(2 + \hat{k})$
ワイブル	$285.3 t_3^6 - 658.6 t_3^5 + 622.8 t_3^4 - 317.2 t_3^3$ $+ 98.52 t_3^2 - 21.256 t_3 + 3.5160$	$l_2 / [(1 - 2^{-1/\hat{k}}) \times \Gamma(1 + 1/\hat{k})]$	$l_1 - \hat{A}[1 - \Gamma(1 + 1/\hat{k})]$

一方、極値の標本についてはデータを昇順に並べ替えた上でL-モーメント諸量を計算する。ただし、CDFの理論値と区別するために、標本諸量に対しては l_2, t, t_3, t_4 などの記号を割り振る。なお、L-モーメント諸量の計算に当たっては、Hosking(2005)が開発したFORTRANルーティン(<http://lib.stat.cmu.edu/>からダウンロード)を利用するとよい。標本L-モーメント諸量が算定されたならば、分布関数の母数が表-3に記載した諸式によって直裁に推計される。表-3のうちのGEVとGPAの母数推定式はHosking・Wallis(1997)によるもので、ワイブル分布に関しては合田(2009, 2010, 2011)による。なお、ワイブル分布の形状母数推定式は、表-2による τ_3 と k との関係の数値計算結果に対して6次回帰式を当てはめた結果である。

極値の標本についてL-モーメント諸量を計算するときには、あらかじめ極値データのそれぞれについて非超過確率 P_i 、すなわちプロテティングポジションを式(2)で算定しておく必要がある。

$$P_i = (i - \alpha)/(n + \beta) \quad (2)$$

ここに、 i は昇順の順序数、 n は標本中のデータ数すなわち標本の大きさ、 α と β は定数である。

標本に対する母数推計を最小2乗法で行うときには、推計値の偏りを最小にするために当てはめる分布関数ごとに最も適切な定数値を選ばなければならない。正規分布に対するBlom公式(1965)や二重指数分布に対するGringorten公式(1963)は良く知られている。L-モーメント法においては、Hosking・Wallis(1997)が $\alpha = 0.35, \beta = 0$ の組合せと $\alpha = 0, \beta = 0$ の組合せの二つを提示した上で後者を“Unbiased Estimator”と称して推奨している。しかしながら、推計値の偏り量については具体的なデータを提示していない。また、 $\alpha = 0, \beta = 0$ の組合せによるプロテティングポジションでは、標本中の最大値 x_n に対する非超過確率が $P_n = 1$ となってしまう、 x_n に対する期待値が計算できないという問題が生じる。このため合田(2010)は大規模な数値実験を行って各種のプロテティングポジションを比較検討した。

図-1はL-CVの偏り率が標本の大きさによって変化する状況を示したもので、 $\alpha = 0, \beta = 0$ のプロテティングポジションの偏り率がほぼ0であり、 $\alpha = 0.45, \beta = 0$ の組合せがこれに次いで小さな偏り率を示す。ここで偏り率とは、各条件において抽出した5万組の標本から得た平均値を求め、その値と母集団の値との差を母集団値で除したものである。一方、図-2は確率波高の推計値の偏り率であるが、偏り率に標本の大きさを乗じて正規化した結果として示している。標本はGEVの母集団から抽出し、それにGEV分布を当てはめている。プロテティングポジションとしては $\alpha = 0, \beta = 0$ および $\alpha = 0.45, \beta = 0$ の組合せを取り上げて比較している。

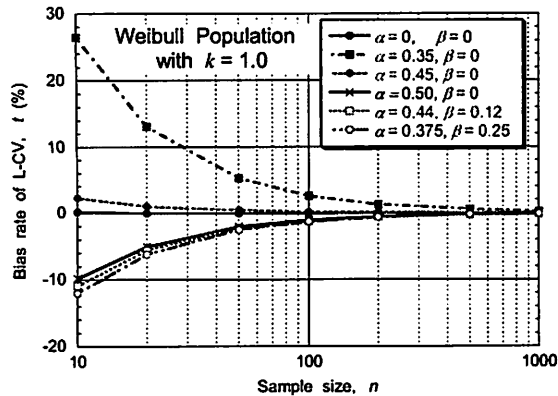


図-1 標本の大きさによる L-CV の偏り率

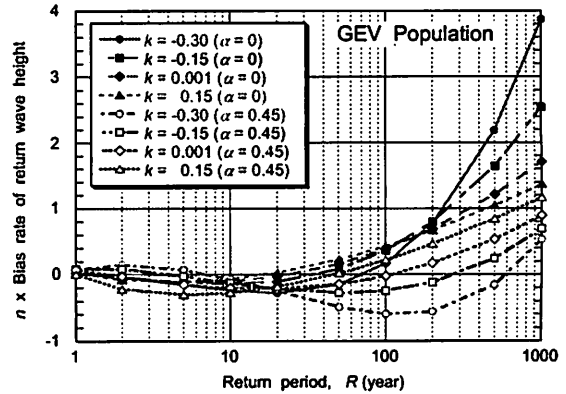


図-2 GEV 母集団から抽出された標本から推計された確率波高の偏り率 (n を乗じて正規化)

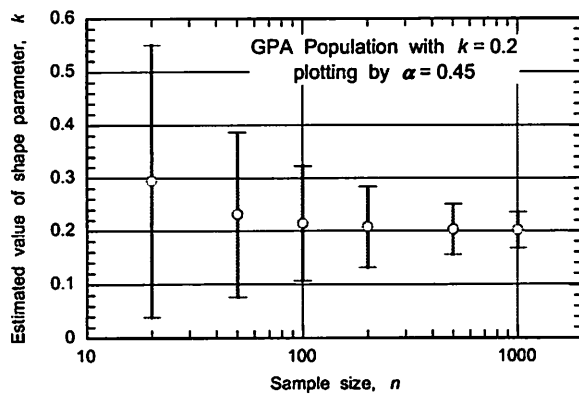


図-3 GPA 母集団 ($k = 0.2$) から抽出された標本から推計された形状母数の平均値と標準偏差

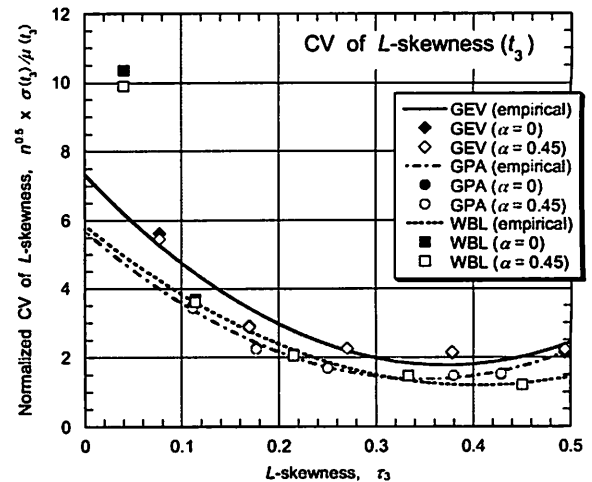


図-4 $n^{1/2}$ を乗じて正規化された L -skewness の変動係数

図-1, 2 で例示されるように、 $\alpha = 0, \beta = 0$ の組合せは L -モーメント諸量の偏りは小さいものの確率波高の推計値にかなりの偏りを生じる。このため、 L -モーメント法では $\alpha = 0.45, \beta = 0$ の組合せのプロットングポジションを採択することが適切であると判断される。本稿では、プロットングポジション式として $\alpha = 0.45, \beta = 0$ の組合せを用いて計算した結果のみを記述する。

極値データの標本から推計した CDF の母数の値は、かなりの統計的変動性を示し、平均値としても真値からずれた値を取る。図-3 は形状母数が $k = 0.2$ の GPA 母集団から抽出された 5 万組の標本に GPA 分布を当てはめて推計した形状母数の平均値と標準偏差（上下の線分の長さ）が標本の大きさ n に応じて変化する状況を示している。また、図-4 は標本から推計された L -skewness の変動係数 ($CV[t_3] = \sigma[t_3]/\mu[t_3]$) が母集団の L -skewness に応じて変化する状況を示したものである。図中の曲線は数値実験結果に対して 2 次回帰曲線を当てはめたものであり、その関数形は L -CV および L -kurtosis とともに表-4 のように取りまとめられている。図には $\alpha = 0, \beta = 0$ および $\alpha = 0.45, \beta = 0$ の組合せの結果をプロットしているが、両者の間の差異はほとんど見られない。なお、図-4 のうちでワイブル分布の $\tau_3 = 0.040$ ($k = 2.8$) のデータは他の形状母数のデータから大きくずれているため、回帰曲線はこのデータを除外して求めている。なお、 L -CV と L -kurtosis に対する回帰曲線の適合状況については合田 (2010) を参照されたい。

表-4 L-モーメント比の変動係数の推定式

L-モーメント比	分布関数	変動係数 CV の推定式
L-CV (t)	一般化極値分布 (GEV)	$CV(t) = (21.6745t_3^2 - 6.7226t_3 + 1.256)/n^{0.5}$
	一般化パレート分布 (GPA)	$CV(t) = (7.2791t_3^2 - 0.4957t_3 + 0.4973)/n^{0.5}$
	ワイブル分布 (WBL)	$CV(t) = (6.4538t_3^2 - 1.4805t_3 + 0.7362)/n^{0.5}$
L-skewness (t_3)	一般化極値分布 (GEV)	$CV(t_3) = (40.1193t_3^2 - 29.8590t_3 + 7.3288)/n^{0.5}$
	一般化パレート分布 (GPA)	$CV(t_3) = (35.5486t_3^2 - 24.7527t_3 + 5.6864)/n^{0.5}$
	ワイブル分布 (WBL)	$CV(t_3) = (28.5170t_3^2 - 23.0420t_3 + 5.8494)/n^{0.5}$
L-kurtosis (t_4)	一般化極値分布 (GEV)	$CV(t_4) = (5.8445t_3^2 - 3.5959t_3 + 3.4001)/n^{0.5}$
	一般化パレート分布 (GPA)	$CV(t_4) = (93.1882t_3^2 - 68.4313t_3 + 15.0430)/n^{0.5}$
	ワイブル分布 (WBL)	$CV(t_4) = (-6.0637t_3^2 + 2.2134t_3 + 2.6893)/n^{0.5}$

4. 単一地点の資料に基づく再現確率波高の推計

4.1. 極値データの確認と選定

極値統計解析では、対象とする統計期間内の極値事象が確実に捕捉されていることを確認しなければならない。既往最大の事象が発生したにもかかわらず適正な観測値が得られていない場合には、適切な数値モデルで推算値を得るように努力する。波高の場合には高波のピーク値が欠測することがあり、スプライン補間あるいは波浪推算によってピーク値を補足するように努める。極値の大きさが推定できない場合にはその年および発生期間を統計期間から除外し、有効統計期間を補正する措置をとる。

極値事象は、発生原因ごとに異なる累積分布に従うことが多い。たとえば、台風系の高波と温帯低気圧系の高波とでは、前者のほうが確率密度関数の右裾が長く広がるのが普通である。このため、極値事象は発生原因別に分離して標本作製しなければならない。

極値時系列 (POT) データの場合には、標本の特性が一つの極値事象を定義する閾値によって変わることがある。閾値のレベルを少しずつ変えて標本作成し、たとえば確率波高の値が閾値によってあまり変化しないことを確認する必要がある。閾値の影響が顕著な事例では、複数の母集団の事象を分離せずに取り扱った場合が多い。閾値レベルは確率波高などへの影響が小さい範囲でできるだけ低く抑え、平均発生率を高めて標本サイズを大きくするのがよい。以下においては、こうした検討を済ませた標本が得られたものとして記述する。

4.2. 標本に対する CDF の適合度の判定

極値データの母分布関数を一つに設定する場合には、標本の L-モーメント諸量を計算し、表-3 の母数推計式を用いて形状・尺度・位置の 3 母数の値を推計する。それによって CDF の特性が定められるので、再現期間 R に対する確率波高は次の式 (3) で推計される。

$$\hat{x}_R = F^{-1}(P_R) = \hat{B} + \hat{A} y_R \quad (3)$$

ここに、 y_R は再現期間 R に対する基準化変量であり、表-1 の諸式で非超過確率を $P_i = P_R = 1 - 1/(\lambda R)$ と設定して算定される。記号 x_R, A , および B の上の山形記号は、これらが標本からの推計値であることを表す。

本稿で著者が採択している極値データの母集団が未知であるとの方式の場合には、一つの標本に対して複数の CDF をに当てはめ、そのうちで適合度がもっとも高いものを母集団分布であると想定する。標本に対する CDF の適合度を測る方式はいろいろ考えられるけれども、ここでは最も分かりやすい“Probability Plot Correlation (PPC)”法を採用する。すなわち、標本中の昇順の各データ x_i に対して、標本に当てはめられた

CDF による期待値 $\hat{x}_i = F^{-1}(P_i) = \hat{B} + \hat{A} y_i$ を計算し、 x_i と $\hat{x}_i$ の間の相関係数 $r(x_i, \hat{x}_i)$ を計算する。そして、相関係数が極限值 1 に近いほど適合度が高いと判断する方法である。PPC 法は Filleben (1975) が使用し、Chowdhury ほか (1991) が他の適合度判定法と併用している。著者 (2008) が最小 2 乗法で用いている相関係数の最小残差法と同じ方式である。

4.3. 母集団が未知のときの確率波高の偏り率

先に図-2 で示した確率波高の偏り率は、GEV の母集団から抽出した標本を GEV 分布に当てはめて推計した結果である。実際の極値統計の実務では母集団を確定することができないので、GEV, GPA, およびワイブルの 3 分布関数の中で最も適合度の高い分布を用いることになる。数値実験では、分布関数とその形状母数、標本の大きさ、裾長さの条件ごとに 5 万組の標本を抽出しているので、標本ごとに最適な分布関数を用いて $R=1 \sim 1000$ 年の 10 レベルの確率波高を推計し、5 万組の標本の確率波高の平均値と標準偏差を算定した。平均値については抽出された母集団の値を基準として偏り率を計算した。図-5~7 はこうして求めた確率波高の偏り率を、標本の大きさ別に再現期間に対して表示したものである。いずれも裾長さが $\gamma_{50} = 1.2$ の場合である。図-5 は母集団が GEV 分布、図-6 は母集団が GPA 分布、図-7 は母集団がワイブル分布の場合であり、形状母数をそれぞれ 4 通りに変えたケースについて図示している。

なお、偏り率は裾長度が大きいほど増大する。計算事例は合田 (2010) を参照されたい。

これらの図の使い方としては、極値データの一つの標本に対して 3 通りの分布関数のいずれかが最適と判定されたとして、推計された形状母数と標本の大きさを勘案し、これらの図の曲線群を内挿して偏り率を推定すればよい。図の凡例である標本の大きさ n はすべて平均発生率が $\lambda=2$ の場合である。たとえば、 $n=100$ は $K=50$ 年の極値データに相当する。平均発生率が $\lambda=10$ であれば、 $K=50$ 年の極値データは $n=500$ となる。ただし、これらの図の確率波高は $\lambda=2$ に対する非超過確率としての再現期間に対して表示しているので、平均発生率が異なるときには再現期間を補正して読み取る必要がある。すなわち、 $\lambda=10$ であれば再現期間 100 年に対しては $R=500$ 年の値を用いる必要がある。

4.4. 母集団が未知のときの確率波高の変動係数

上述の各 5 万組の標本から算定された標準偏差と平均値に基づき、確率波高の変動係数が求められる。これについては試行錯誤によって式 (4) のような推定式を導いた。

$$CV[H_R] = \sigma[H_R] / \mu[H_R] = (ay_R^2 - by_R + c)(\gamma_{50} - 1)^{0.7} / \sqrt{n} \quad (4)$$

そして、母集団の分布関数の形状母数毎に定数 a, b, c を算定し、さらにその結果を形状母数の関数として表示することを試みた。これによって、分布関数毎に表-5 のような経験式を得ることができた。

表-5 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)	$a = 2.7037k^3 + 2.4351k^2 + 0.7325k + 0.0830$ $b = 10.953k^3 + 10.509k^2 + 3.1359k + 0.0372$ $c = 11.311k^3 + 12.526k^2 + 5.6088k + 0.7719$
一般化パレート分布 (GPA)	$a = \exp[-5.3844k^3 - 0.6089k^2 + 7.7316k - 1.9805]$ $b = -22.438k^3 + 24.678k^2 + 6.1046k + 0.1062$ $c = -28.287k^3 + 18.860k^2 + 6.9797k + 0.6776$
ワイブル分布	$a = 0.9742(\ln k)^2 + 0.6949(\ln k) + 0.1517$ $b = 1.9049(\ln k)^2 + 1.8142(\ln k) + 0.3677$ $c = 0.6349(\ln k)^2 + 1.8561(\ln k) + 0.9095$

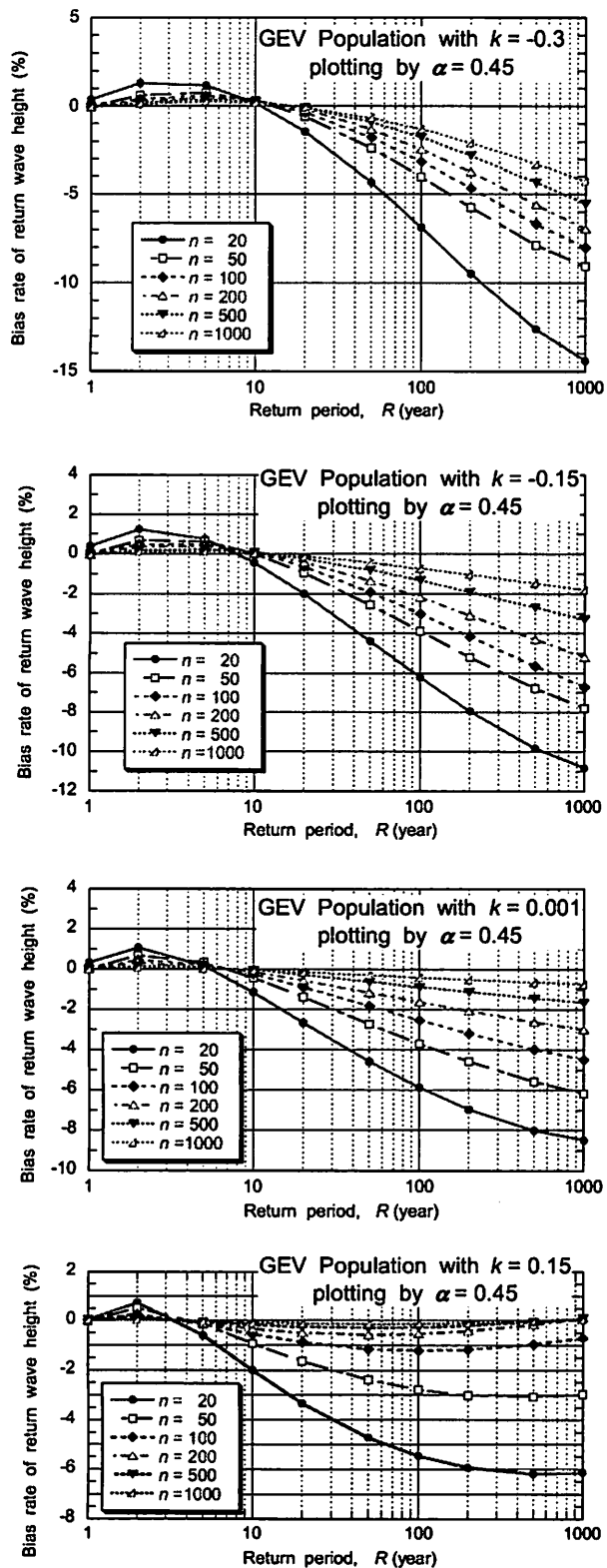


図-5 母集団が未知の場合に一般化極値分布の標本への最適合分布から推計された確率波高の偏り率 ($\gamma_{50}=1.2$, $\alpha=0.45$)

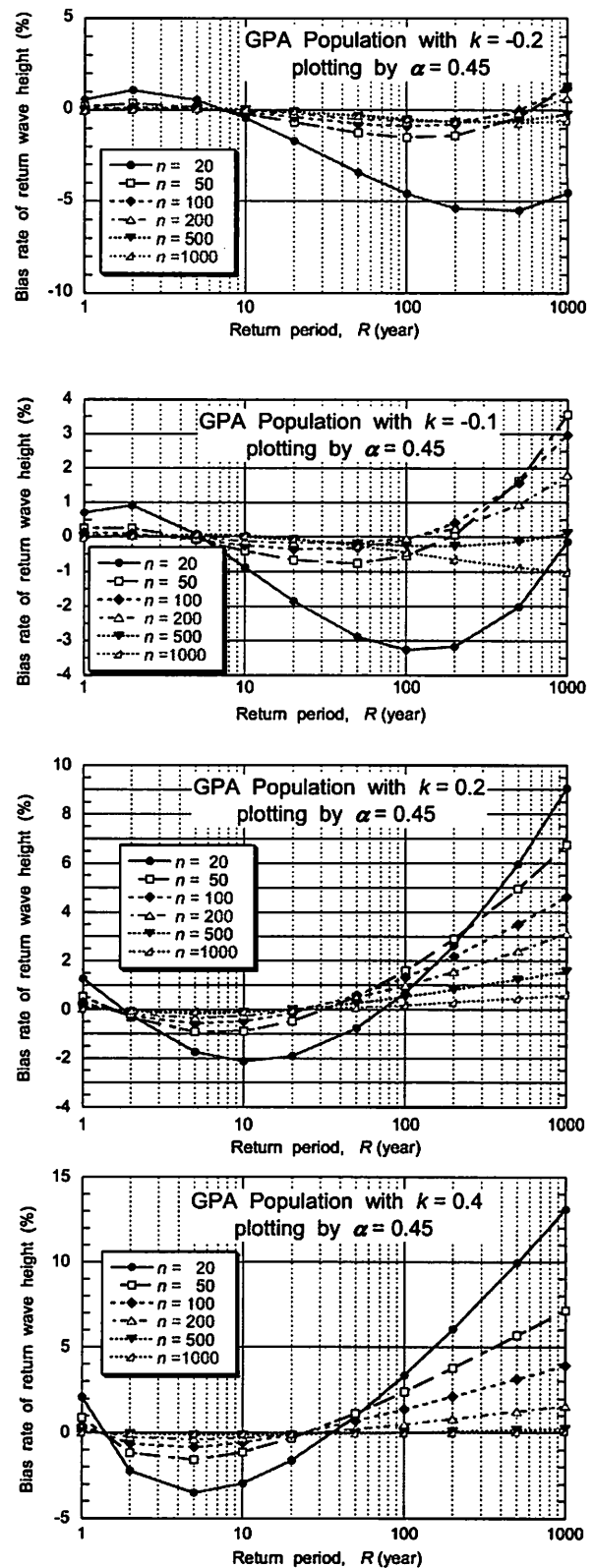


図-6 母集団が未知の場合に一般化パレート分布の標本への最適合分布から推計された確率波高の偏り率 ($\gamma_{50}=1.2$, $\alpha=0.45$)

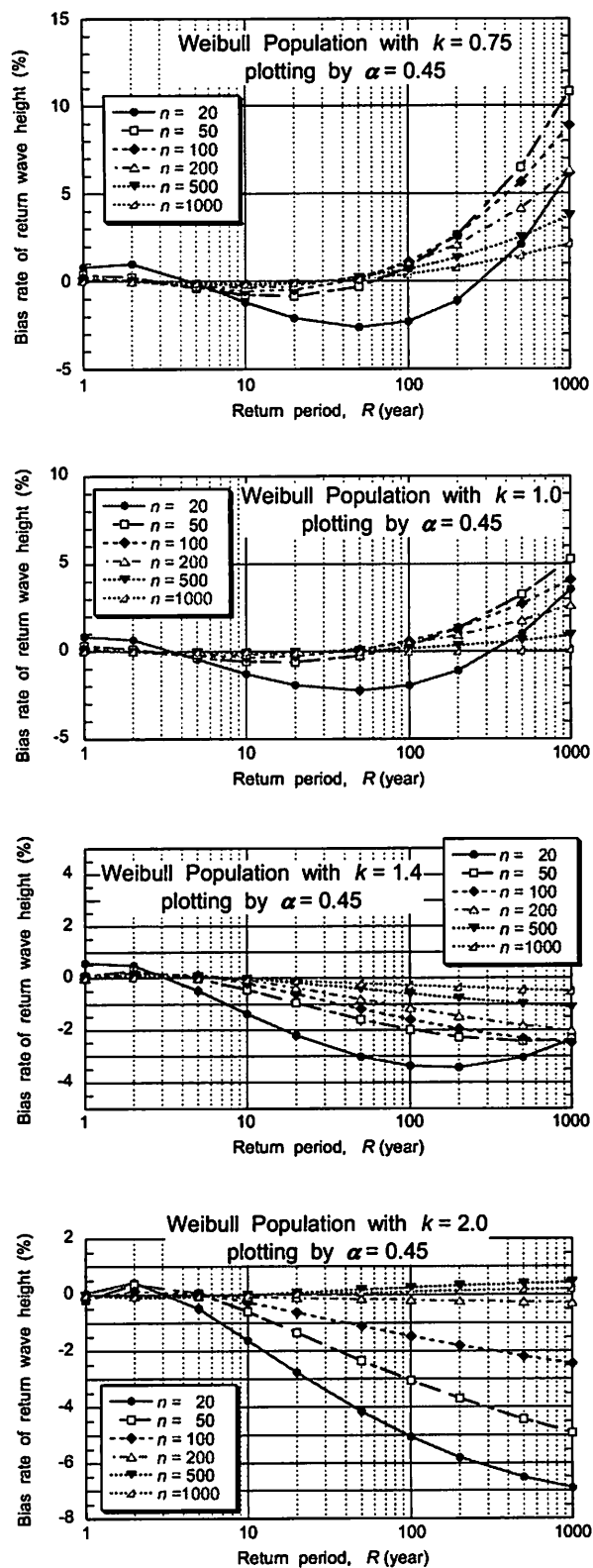


図-7 母集団が未知の場合にワイブル分布の標本への最適適合分布から推計された確率波高の偏り率 ($\gamma_{50} = 1.2, \alpha = 0.45$)

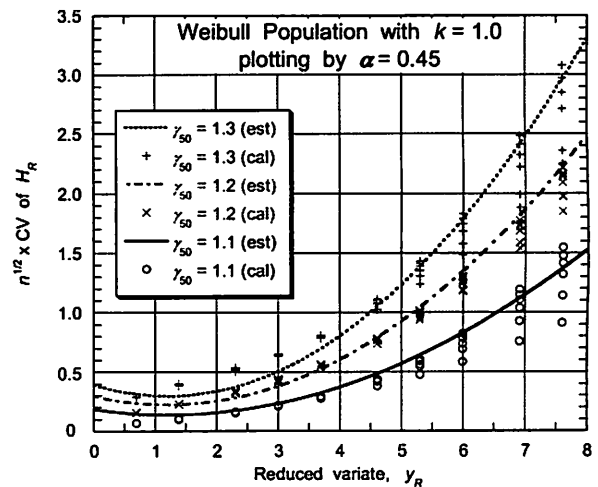


図-8 指数分布から抽出された標本に対し、母集団が未知であるとして推計された確率波高の変動係数

図-8は、表-5の係数を用いて式(4)で確率波高の変動係数を算定し、数値実験の結果と比較した例である。母集団が $k = 1.0$ のワイブル分布、すなわち指数分布から抽出された標本に対してGEV、GPA、およびワイブルの3通りのCDFの中から最適適合の分布を選び、確率波高を推計したときの変動係数に $n^{0.5}$ を乗じて正規化した結果を示している。裾長度は $\gamma_{50} = 1.1, 1.2, 1.3$ の3通りに変えて標本を抽出している。

式(4)の推定値(est)の上下に数値実験の値(cal)が広がっているのは、確率波高の変動係数が $n^{0.5}$ に逆比例する関係からやや逸脱していることを表している。すなわち、式(4)は確率波高の変動係数の概略値を与えるものにとどまる。より信頼できる変動係数を求めるためには、標本に対して当てはめられたCDFの母数値を使い、モンテカルロ法で数値シミュレーションを行うことが推奨される。

5. 地域頻度解析による現確率波高の推計法

5.1. 地域頻度解析の手順

観測資料などから得られる極値データは、たまたま得られた一つの標本であり、母集団の特性を的確に表示したものとはいえない。観測の年次が異なれば、標本もまた異なったものが得られるであろう。したがって、一つの地点の標本から推計される極値の累積分布関数は本来の母分布関数とは異なる可能

性があり、それに基づいて推計される確率波高もかなりの誤差を伴うと考えられる。

こうした標本の統計的変動性の影響を減殺する方法として、極値事象の特性がほぼ同じ地域内の複数の地点のデータを同時に解析し、推計値の信頼性を高めることが考えられる。これが地域頻度解析であり、Hosking・Wallis(1997)が詳しく解説している。この地域頻度解析は次のような手順で進めればよい。

- 1) 地理的な位置その他を勘案し、対象とする地域を選定してその地域内の極値データを収集する。
- 2) 各地点のデータをその平均値で除し、無次元化する。
- 3) 無次元化された標本について、単一地点と同じ手順でL-モーメント法による解析を行う。
- 4) 各地点のL-CV, L-skewness, およびL-kurtosisの値について各地点のデータ数を重みとする加重平均を行い、地域としての3種類のL-モーメント比を算出する。
- 5) 各地点の3種類のL-モーメント比の地域全体の値(重心)からの距離 D を計算し、その距離が著しく大きければその地点は不整合であるとして当該地域から除外する。これはDiscordance Test (不整合性テスト)と名付けられており、その作業指針はHosking・Wallis(1997)が示している。具体的にはHosking(2005)が作成したFORTRANルーティンを利用して不整合性を検討する。
- 6) さらに、各地点の3種類のL-モーメント比を用いて地域内の標本の均質性をチェックし、地域の均質性が確保できないときには幾つかの地点を除外あるいは入れ替えて組み直す。この均質性は、Heterogeneity Measure (不均質性指標)と名付けられた値を計算してチェックする。
- 7) 均質性が確認された地域について、地域全体の3種類のL-モーメント比を使って地域共通の分布関数を選定する。
- 8) 地点毎の極値分布関数は、形状母数は地域共通の値を用い、尺度母数と位置母数は地域共通の母数値に各地点の平均値を乗じて換算する。
- 9) 地点毎に換算された分布関数の母数を使い、各種の再現期間に対する確率波高を算定する。
- 10) 確率波高の信頼区間等については、地域共通の分布関数を使って数値シミュレーションを行い、変動範囲を調べて推定する。

5.2. Hosking・Wallisによる不均質性指標

Hosking・Wallis(1997)は不均質性指標として次のような指標を提示している。

$$H = \frac{(V - \mu_V)}{\sigma_V} \quad (5)$$

上式中の V は次式で定義される量であり、 μ_V と σ_V はカッパ分布を当てはめて行う数値シミュレーションによって求められる V の平均値と標準偏差である。

$$V = \left\{ \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{1/2} \quad (6)$$

ここに、 N は地点数、 n_i は i 番目の地点のデータ数、 $t^{(i)}$ はその地点のL-モーメント比に係わる統計量、 t^R はその統計量に関してカッパ分布の当てはめシミュレーションで得られた地域平均の値である。

不均質性指標は、3種類の統計量について検討する。第1は、2次のL-モーメントと1次のL-モーメントの比であるL-CV[t]である。すなわち、ある地点の t の値が他よりも著しくかけ離れていると、その地域は不均質と見なされる。第2の統計量としては、L-CVとL-skewnessの相関プロット図において、ある地点が全体の重心からどれだけ離れているかを検討する。第3の統計量としては、L-kurtosisとL-skewnessの相関プロット

図における重心からの距離を用いる。なお、第2と第3の不均質性指標については、 r として2次元座標上の距離を用いて算定する。

地域が均質であるか否かを判断する不均質性指標の閾値として Hosking・Wallis (1997) は H が1以下であれば均質性に問題はなく、 H が2を超えると不均質である公算が大きいとしている。

5.3. 新しい不均質性指標の提案

極値事象に関する地域頻度解析としては、河合ら (2010) による日本沿岸の高潮偏差ならびに最高潮位の解析事例が挙げられる。その事例では、 L -CV に係わる $H(1)$ の不均質性の条件が厳しく、観測地点の地理的位置をかなり無視する形で地域を細分しなければならなかった。また、次節に述べる日本海沿岸の高波の解析でも、 $H(1)$ の不均質性指標が地域設定の制約条件となった。何よりも、Hosking・Wallis (1997) による不均質性指標は設定した地域に対して単一の値が示されるのみであり、その地域が不均質であると判定されてもどの地点が不均質性を招いているかを知ることができない。そのため、地点入れ替えの試行錯誤によって均質な地域を設定しなければならない。また、Hosking・Wallis (1997) の不均質性指標は4母数型の極値分布であるカッパ分布の数値シミュレーションに基づいているけれども、カッパ分布を導入する意義が明確でない。

こうしたところから、合田 (2010) は表-5 に導いた L -モーメント比の変動係数を用いた新しい不均質性指標を提案している。すなわち、地域内の各地点について L -モーメント比を計算し、地域平均の L -モーメント比とその標準偏差を算定する。そして、地域平均の L -モーメント比と各地点の値との差を標準偏差で除したものを偏心率と名付け、地点毎の偏心率の分散の度合いが正規分布に近ければその地域を均質とみなし、偏心率の分散が正規分布から外れていれば不均質とみなす方式である。

偏心率は、 L -CV, L -skewness, および L -kurtosis については次式で計算する。

$$\psi_{k,i} = [\zeta_{k,i} - \mu(\zeta_k)] / \sigma(\zeta_k) : k = 1, 2, 3; i = 1, 2, \dots, N \quad (7)$$

ここに、 $\zeta_1, \zeta_2, \zeta_3$ はそれぞれ L -CV [t_1], L -skewness [t_3], および L -kurtosis [t_4] を代表する。 $\zeta_{k,i}$ は i 番目の地点の L -モーメント比、 $\mu(\zeta_k)$ と $\sigma(\zeta_k)$ は地域全体の L -モーメント比の平均と標準偏差である。さらに、 L -CV と L -skewness の相関プロット図上の偏心率として次のものを定義する。

$$\psi_{4,i} = \left[\frac{[\zeta_{1,i} - \mu(\zeta_1)]^2 + [\zeta_{2,i} - \mu(\zeta_2)]^2}{\sigma^2(\zeta_1) + \sigma^2(\zeta_2)} \right]^{1/2} \quad (8)$$

なお、合田 (2010) は L -kurtosis と L -skewness の相関プロット図上の偏心率も定義したがここでは割愛する。

こうした各地点の偏心率の分散の度合いを正規分布と比べるには、偏心率の最大値を用いるのが簡便である。すなわち、

$$\Psi_k = \max \{ |\psi_{k,i}| : i = 1, 2, \dots, n \} : k = 1, 2, \dots, 4 \quad (9)$$

正規分布であれば、偏心率の最大値が1.6を超える確率は11%、2.0を超える確率は4.6%であるので、4種類の偏心率について正規分布と比較しながら均質性を調べることができる。この偏心率の適用事例については次節で紹介する。

6. 日本海沿岸の波浪データに対する地域頻度解析法の適用事例

6.1. 解析した波浪データの概要

わが国の国土交通省の港湾関係部局では1970年から全国波浪観測体制を構築しており、24時間連続の波浪観測データ（一部は2時間毎20分観測）が独立行政法人港湾空港技術研究で集中管理されている。この観

測体制はNOWPHAS(Nationwide Ocean Wave information network for Port and HArbourS)と呼ばれており、2009年現在は61地点で海底設置型波浪計、11地点でGPSブイによる観測を継続している。

本稿で紹介するのは、表-6に示す日本海沿岸11地点(北から南へ順に記載)のデータである。すなわち、有義波高の時系列において角括弧内の波高閾値を超え、波高値が第2閾値2.0mを下回るまでの期間を一つの高波とみなし、その間の最大有義波高を極大値波高として定義した高波データである。また有効統計年数というのは、何らかの原因で欠測が生じた期間を日単位で積算し、それを観測年数から年単位で差し引いたものである。したがって、有効統計年数と観測年数の比が測得率を与える。高波極値としては、閾値を0.5mずつ段階的に変化させ、確率波高がほぼ安定する範囲で低めの閾値に基づいて採択した。平均発生率は $\lambda=10\sim15$ である。なお、この高波極値の解析の詳細は合田・久高(2011)を参照されたい。

表-6には地点名と高波資料の概要を示しているが、ほとんどは冬季季節風によって発達した高波である。なお、超音波式波高計で高波時に欠測が生じたケースについては、併設の水圧式センサーの記録を用いて可能な限り補足した。補足が難しかった場合には、その高波の期間を欠測扱いとした。この表で明らかなように、極値統計解析に用いた極値有義波高の平均値は4.3~5.2m、標準偏差が0.75~1.08m、極値有義波高の変動係数が $CV=0.17\sim0.21$ と比較的に均一な資料である。

表-6 検討に用いた高波資料の概要

地点名	観測期間	観測年数	有効統計年数	データ総数 [H_c (m)]	閾値 H_c (m)	極値 個数	平均発生率 λ	平均波高 (m)	標準偏差 (m)	最大波高 ($H_{1/3}$) _{max} (m)
留 萌	1970-2008	39	37.2	1214 [2.5]	3.5	554	14.89	4.440	0.790	7.83
瀬 棚	1980-2008	29	26.3	660 [3.0]	4.0	330	12.55	4.989	0.870	9.43
深 浦	1979-2008	30	28.0	779 [3.0]	4.0	408	14.57	5.097	0.971	10.36
秋 田	1981-2008	28	22.8	594 [3.0]	4.0	314	13.77	5.202	1.030	8.74
酒 田	1970-2008	39	34.4	955 [3.0]	4.0	518	15.06	5.249	1.081	10.65
新潟沖	1989-2008	20	18.4	393 [3.0]	3.5	252	13.70	4.437	0.895	8.48
輪 島	1979-2008	30	28.7	658 [3.0]	4.0	298	10.38	4.934	0.749	7.73
金 沢	1970-2008	39	32.3	743 [3.0]	4.0	387	11.98	5.104	0.889	8.14
福 井	1980-2008	29	18.0	347 [3.0]	3.5	248	13.78	4.594	0.946	7.79
鳥 取	1979-2008	30	26.2	542 [3.0]	3.5	320	12.21	4.346	0.745	7.54
浜 田	1974-2008	35	29.9	632 [3.0]	3.5	400	13.38	4.354	0.755	7.93

6.2. 各地点のL-モーメント比と地域区分

表-6の高波資料についてL-モーメント比であるL-CV, L-skewness, およびL-kurtosisを計算した結果を表-7に示す。これらのL-モーメント比は、L-CVが0.082~0.111, L-skewnessが0.224~0.297, L-kurtosisが0.095~0.162と比較的に狭い範囲に分布している。この表には、11地点を北部6地点、南部5地点の2地域に分割したときの地域加重平均値も記載してある。また、不整合距離 D_i 値は南北2地域に分割したときの値である。 D_i 値はどの地点も2以下であり、地域内で整合している。

偏心率を用いた不均質性指標は、北部地域ではL-CVに係わる ψ_1 が瀬棚で-1.96と大きな値を示すものの、 $\psi_2\sim\psi_4$ は絶対値が1.1以下であり、この地域は均質であると判断される。南部地域では、輪島のL-CVが0.082と小さい一方で福井がL-CV=0.111と大きな値を示すため、L-CVに係わる ψ_1 の最大値が4.12となり、均質性に問題があることを示す。しかし、CDFの形状母数を支配するL-skewnessに係わる ψ_2 は最大でも1.55である。L-CVというのは、L-モーメントの2次と1次の比であり、coefficient of L-variationと名付けられている。CDFの母数推定式を記載した表-3に見られるように、標本から推計する母数値を支配するのはL-skewness(l_3)、2次のL-モーメント(l_2)、および1次のL-モーメント(l_1)すなわち標本平均値である。その意味で、均質性に係わるL-CVの意義についてはさらに検討する必要がある。しかし、日本海沿岸の11地点に

表-7 各地点の L -モーメント比および不均質性指標

地点名	データ 個数, n	L -CV	L -skewness	L -kurtosis	D_i	ψ_1 (L -CV)	ψ_2 (L -skew)	ψ_3 (L -krt)	ψ_4 (CV-skew)
留 萌	554	0.0961	0.2475	0.1145	1.42	-1.15	-0.65	-0.58	0.67
瀬 棚	330	0.0921	0.2749	0.1328	0.97	-1.96	0.28	0.27	0.42
深 浦	408	0.1002	0.2779	0.1385	0.82	-0.23	0.39	0.49	0.39
秋 田	314	0.1060	0.2660	0.1262	0.24	0.87	-0.02	-0.01	0.16
酒 田	518	0.1092	0.2586	0.1263	0.92	1.45	-0.28	-0.01	0.38
新潟沖	252	0.1053	0.2971	0.1256	1.62	0.68	1.07	-0.04	1.06
北部平均・最大	2,376	0.1014	0.2667	0.1265	1.62	1.96	1.07	0.58	1.06
輪 島	298	0.0819	0.2515	0.1226	0.66	-3.54	-0.39	-0.07	0.61
金 沢	387	0.0956	0.2238	0.0948	0.80	0.70	-1.55	-2.13	1.54
福 井	248	0.1110	0.2568	0.1162	1.08	4.12	-0.16	-0.46	0.77
鳥 取	320	0.0892	0.2954	0.1620	1.16	-1.02	1.54	1.67	1.53
浜 田	400	0.0912	0.2772	0.1271	1.30	-0.52	0.73	0.18	0.72
南部平均・最大	1,653	0.0931	0.2605	0.1239	1.30	4.12	1.55	2.13	1.54

ついで言えば、これらの地理的位置から見て、高波の発生特性に大きな差異があるとは思われない。したがって、均質性にはやや問題が残るけれども、本稿では日本海沿岸を北部地域6地点、南部地域5地点に区分して地域頻度解析を行った結果について記述する。

6.3. 観測データに適合する極値分布関数

前述のように、本稿では高波の母分布関数の候補として CDF, GPA, およびワイブルの3分布を取り上げ、標本に対する最適分布を PPC 法あるいは最小残差法で選択している。後者は、標本の各データとそれに対する確率期待値の間の相関係数 $r(x_i, \hat{x}_i)$ を計算し、その1からの残差 $\Delta r = 1 - r(x_i, \hat{x}_i)$ が最小のものを最適と判定する方式である。日本海沿岸11地点について最適分布を判定した結果が表-8である。

表-8 波高データと確率期待値との相関係数の残差値と最適分布

地点名	相関係数の残差値			最適分布		地点名	相関係数の残差値			最適分布
	GEV	GPA	WBL				GEV	GPA	WBL	
留 萌	0.0245	0.0019	0.0050	GPA		輪 島	0.0208	0.0013	0.0042	GPA
瀬 棚	0.0146	0.0030	0.0032	GPA		金 沢	0.0301	0.0032	0.0082	GPA
深 浦	0.0116	0.0021	0.0016	WBL		福 井	0.0227	0.0018	0.0052	GPA
秋 田	0.0262	0.0027	0.0064	GPA		鳥 取	0.0118	0.0036	0.0029	WBL
酒 田	0.0127	0.0015	0.0012	WBL		浜 田	0.0103	0.0029	0.0017	WBL
新潟沖	0.0122	0.0040	0.0033	WBL						
北部地域 最適地点数	0	3	3	-		南部地域 最適地点数	0	3	2	-

この表-8から分かるように、日本海沿岸の高波に関しては GEV の適合度が低く、母分布関数の可能性が低い。これに対して GPA 分布とワイブル分布は最適の地点数が拮抗しているといえる。この二つの分布に対する母数推計値は表-9, 10の通りである。各母数で「地域」の列は、表-7の地域平均 L -モーメント比の値を用いて無次元母数値を求め、形状母数についてはそのまま、尺度・位置母数については無次元母数値に各地点の平均波高を乗じて次元量に換算したものである。一方「単独」の列は、各地点の L -モーメント諸量を直接に使って母数値を推計した結果である。当然ながら、地点単独に推計した母数値のほうが地点ごとの変動が大きく現れている。

表-9 GPA 分布の母数推計値

地点名	形状母数 k		尺度母数 A		位置母数 B		地点名	形状母数 k		尺度母数 A		位置母数	
	地域	単独	地域	単独	地域	単独		地域	単独	地域	単独	地域	単独
留 萌	0.1579	0.2063	1.125	1.094	3.468	3.425	輪 島	0.1733	0.1963	1.172	1.026	3.936	4.046
瀬 棚	0.1579	0.1376	1.264	1.118	3.897	4.007	金 沢	0.1733	0.2686	1.212	1.404	4.071	3.997
深 浦	0.1579	0.1302	1.292	1.230	3.982	4.009	福 井	0.1733	0.1827	1.091	1.317	3.664	3.481
秋 田	0.1579	0.1595	1.318	1.381	4.064	4.011	鳥 取	0.1733	0.0878	1.032	0.881	3.467	3.537
酒 田	0.1579	0.1781	1.330	1.472	4.100	4.000	浜 田	0.1733	0.1319	1.034	0.958	3.473	3.508
新潟沖	0.1579	0.0837	1.124	1.055	3.466	3.463							

表-10 ワイブル分布の母数推計値

地点名	形状母数 k		尺度母数 A		位置母数 B		地点名	形状母数 k		尺度母数 A		位置母数	
	地域	単独	地域	単独	地域	単独		地域	単独	地域	単独	地域	単独
留 萌	1.2024	1.272	1.093	1.094	3.412	3.425	輪 島	1.2241	1.257	1.136	1.026	3.871	3.980
瀬 棚	1.2024	1.174	1.228	1.090	3.834	3.958	金 沢	1.2241	1.367	1.175	1.340	4.004	3.878
深 浦	1.2024	1.164	1.255	1.201	3.917	3.959	福 井	1.2241	1.238	1.057	1.274	3.605	3.405
秋 田	1.2024	1.205	1.281	1.342	3.998	3.942	鳥 取	1.2241	1.108	1.000	0.866	3.410	3.513
酒 田	1.2024	1.231	1.292	1.425	4.034	3.917	浜 田	1.2241	1.167	1.002	0.935	3.416	3.468
新潟沖	1.2024	1.103	1.092	1.039	3.410	3.435							

6.4. 日本海沿岸の確率波高

日本海北部および南部地域について、当てはめた地域共通分布を無次元化した高波データと比較したのが図-9, 10 である。昇順の高波データは非超過確率 P_i を式 (2) で与え、それを $P_i = 1 - 1/(\lambda R_i)$ の関係式で再現期間に換算している。両図ともグラフ上の点の輻輳を避けるため、上位から第13位以下の波高データを適宜間引いている。北部地域に対する図-9 をみると、留萌の観測値や秋田の上位のデータは地域共通の GPA 分布よりも下に位置しており、一方、酒田や新潟沖の上位のデータは地域共通のワイブル分布の上にある。瀬棚の波高値は全体として地域共通分布よりも低めであるが、第1位のデータは地域共通のワイブル分布に近い。深浦のデータも瀬棚と類似の傾向を示す。

南部地域に対する図-10 では、鳥取と浜田の波高値が地域共通のワイブル分布にほぼ沿っているのに対し、輪島と金沢のデータは地域共通の GPA 分布よりも下に位置している。福井港のデータは地域共通分布よりも上にずれた位置にあり、これは表-8 の L -CV の偏心率が $\psi_1 = 4.12$ と大きいことに関係していると思われる。

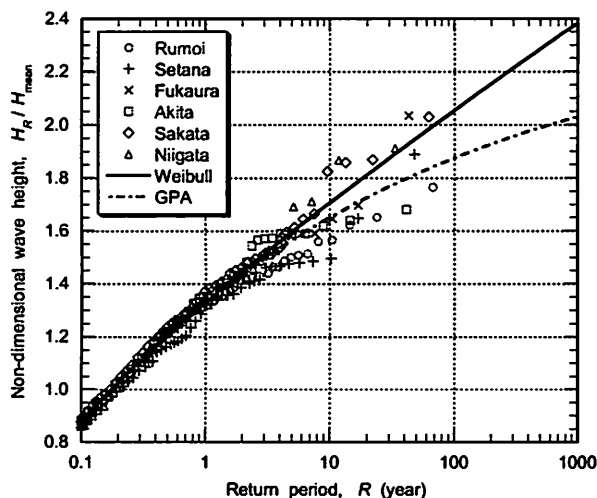


図-9 北部地域の極値有義波高に対する極値分布の当てはめ結果

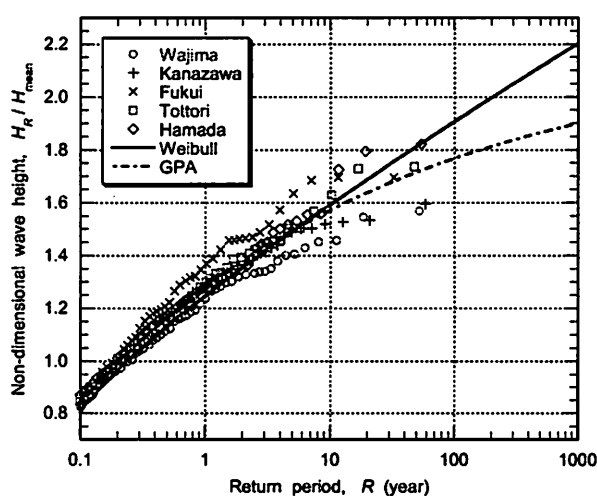


図-10 南部地域の極値有義波高に対する極値分布の当てはめ結果

地域頻度解析というのは、図-9、10に見られるように所定の再現期間に対する地点毎の無次元確率波高の値が分散するものの、分散は標本の統計的変動性によるものとみなし、地域共通の極値分布を用いて確率波高を推計するものである。したがって、地点単独に当てはめた極値分布から推計した確率波高と地域共通分布から推計した確率波高とでは、そこにある程度の差が生じる。しかし、地域共通分布を用いることによって確率波高の空間分布が滑らかに変化し、またデータ数の総計が大きくなることによって推計された確率波高の信頼区間が狭まると考えられる。

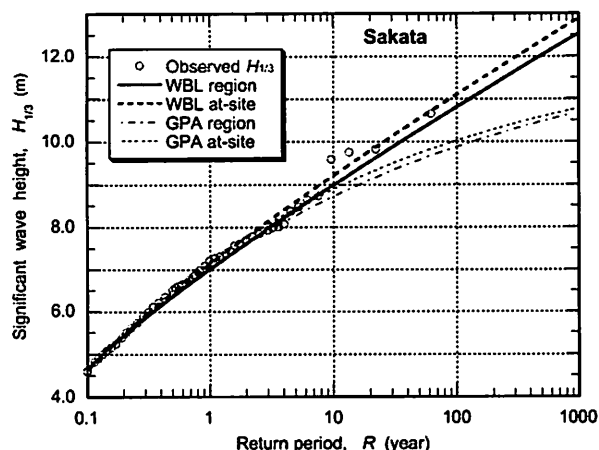


図-11 酒田港の極値有義波高に対する GPA およびワイブル分布の当てはめ結果

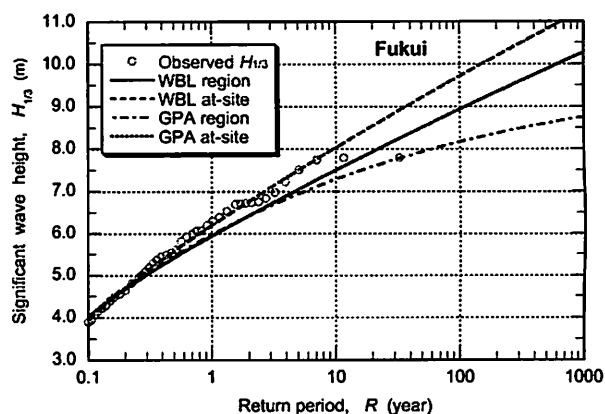


図-12 福井港の極値有義波高に対する GPA およびワイブル分布の当てはめ結果

個別地点に対する CDF の当てはめの事例を図-11、12に示す。前者は酒田、後者は福井の場合であり、地域共通分布ならびに地点単独の当てはめ分布を示している。CDF としては GPA とワイブル分布を用いている。図-11 の酒田のケースでは、GPA・ワイブルの両分布とも地域共通分布と単一地点当てはめ分布の差異がほとんどない。しかし、図-12 の福井のケースでは単一地点当てはめ分布（GPA・ワイブルの両分布がほぼ同一）が地域共通分布と大きくずれている。この逸脱傾向は既に図-10 で認められているところであるが、隣接する地点である金沢や鳥取と比べて地理的条件で高波の出現特性が異なるとは考えにくい。福井の観測データがたまたまそうした統計的変動性を伴っていたと解釈せざるを得ない。

日本海沿岸の 11 地点に対して 100 年確率波高を推計した結果を表-11 に示す。単一地点当てはめによる推計値に比べて、地域共通分布に基づく推計値のほうが北から南へ向かって滑らかに変化していることが分かる。これが地域頻度解析の特長である。

表-11 の 100 年確率波高の推計値のうちの斜体文字は、有効統計期間 18～37 年の間に観測された最大有義波高よりも小さいケースを示している。高波データへの適合度は表-9 に示したように GPA 分布とワイブル

表-11 100 年確率有義波高の推定値の比較

地点名	最大有義波高 (m)	100 年確率有義波高 (m)				地点名	最大有義波高 (m)	100 年確率有義波高 (m)			
		GPA 地域	GPA 単独	WBL 地域	WBL 単独			GPA 地域	GPA 単独	WBL 地域	WBL 単独
留 萌	7.83	8.32	7.78	9.08	8.65	輪 島	7.73	8.72	8.07	9.50	8.77
瀬 棚	9.43	9.35	9.09	10.21	9.77	金 沢	8.14	9.02	8.44	9.83	9.49
深 浦	10.36	9.55	9.80	10.43	10.57	福 井	7.79	8.12	8.76	8.85	9.70
秋 田	8.74	9.75	9.94	10.64	10.87	鳥 取	7.54	7.68	8.19	8.37	8.60
酒 田	10.65	9.84	10.02	10.74	11.09	浜 田	7.93	7.70	7.96	8.38	8.55
新潟沖	8.48	8.32	9.18	9.08	9.68						

分布が拮抗しているけれども、GPA 分布は再現期間が長くなるにつれて確率波高が頭打ちになる傾向がある（形状母数が正の場合）。GPA 分布を当てはめたときに、観測最大波高が100年確率波高を上回る地点が 11 地点中で5箇所も出現するようでは、そうした分布の選択をためらわざるを得ない。構造物設計の立場からは、構造物の安全性を重視し、ワイブル分布を採択するのが適切と考えられる。

7. AM 資料と POT 資料に基づく極値統計解析の比較

7.1. 高波データに対する AM 資料と POT 資料解析の比較

極値統計の教科書では、年最大値 (AM) 資料から推計した確率波高は POT 資料から推計した波高よりも再現期間にして 1/2 年長いだけの差であると説明されている。しかし、表-6 の日本海 11 地点の観測データならびにアラスカ沖 Kodiak 島の推算データ (Andrew et al., 1985) について年最大波高の標本を作成して解析したところ、前節で紹介した POT 資料による推計結果とかなり異なる結果が得られた。図-13 は、AM 資料にワイブル分布を当てはめたときの形状母数を、同じ地点の POT 資料に当てはめたワイブル分布の形状母数と比較した結果である。縦軸は形状母数の比、横軸は資料の統計年数である。図中の実線は数値実験による平均値、上下の鎖線は平均値の上下に標準偏差の 1 倍をとった値を示している。明らかに、AM 資料に対して当てはめたワイブル分布は形状母数が 2 倍程度大きくなっている。

一方、図-14 は 100 年確率波高の推計値を比べたもので、POT 資料による推計値を真値とみなしたときの AM 資料の推計値の偏り率を表している。地点ごとの変動が大きいものの、算術平均で-4.6%の偏り率を示している。

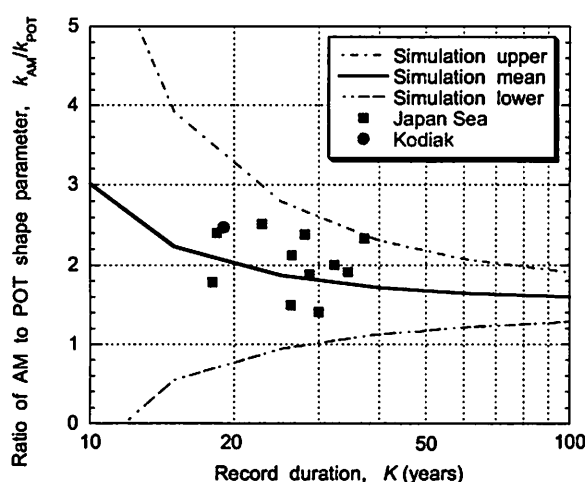


図-13 日本海沿岸および Kodiak 沖の波高に当てはめられたワイブル分布の形状母数の比率

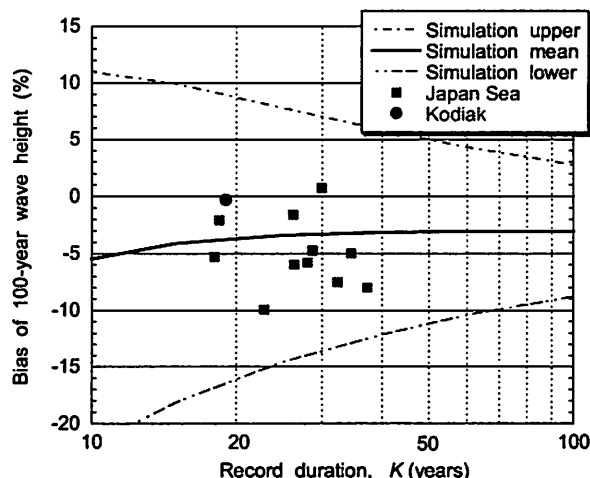


図-14 極大値資料から推計された 100 年確率波高を基準とする年最大値資料の 100 年確率波高の偏り率

7.2. AM 資料と POT 資料の累積分布関数の理論的關係

年最大値 (AM) というのは 1 年間に発生する極値 (POT) の中の最大値である。いま POT の累積分布関数 (CDF) すなわち非超過確率を $G(x)$ で表すと、年間に極値が m 個発生したときの AM の非超過確率は $[G(x)]^m$ で表される。高波の場合、極値の発生回数は次のポアソン分布に従うことが例証されている (合田 2011a)。

$$P(m) = e^{-\lambda} / m! \quad (10)$$

ここに、 λ は平均発生率すなわち 1 年間の平均発生回数である。

このポアソン分布を導入することで、AM の累積分布関数が次のように POT の CDF に関係づけられる。

$$F(x) = \Pr[H_{\max} < x] = \sum_{m=0}^{\infty} P(m) G^m(x) = \sum_{m=0}^{\infty} \frac{1}{m!} e^{-\lambda} [\lambda G(m)]^m = \exp\{-\lambda[1-G(x)]\} \quad (11)$$

POT の CDF が形状母数 $k=1$ のワイブル分布、すなわち指数分布のときには、それに対する AM 資料の CDF が次の二重指数分布に漸近することを Cunnane (1973) が示している。

$$F(x) = \exp\{-\exp[-(x - B^*)/A]\} \quad : \quad B^* = B + A \ln \lambda \quad (12)$$

また、POT の CDF が GPA 分布のときには、AM 資料の CDF が次の GEV 分布に漸近することを Madsen et al. (1997) が示している。

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

この尺度・位置母数は、GPA 分布の母数と次のように関係づけられる。形状母数は共通である。

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

7.3. 数値実験による AM 資料の統計的性質

式 (12), (13) のような AM 資料の漸近分布の妥当性を確認するため、合田 (2011a, 2011b) は一連の数値実験を行った。GEV, GPA, およびワイブルの 3 分布の母集団から多数の POT データを抽出し、平均発生率を設定した上で各年の POT データがポアソン分布に従うとして年最大値を検出し、所定の統計年数に対する AM データの標本を作成した。AM 資料の統計年数は $K=10 \sim 10,000$ 年にわたる 12 種類を対象とし、標本数は各条件で 5 万組である。ただし、 $K > 100$ 年については計算時間を短縮するため $N=5$ 万/ K に縮減した。平均発生率は $\lambda=3$ と 10 の 2 通りを用いた。POT データの母集団としては、3 種類の CDF の形状母数を 3~5 通りに変え、2 種類の平均発生率をあわせて全体で 27 種類を用いた。尺度・位置母数は 50 年確率波高が $H_{50} = 10.0\text{m}$ 、裾長度が $\gamma_{50} = 1.2$ の条件で式 (1) に従って与えた。作成された AM データの標本はこれまでに述べた L-モーメント法の手法で解析し、POT の母集団に係わりなく最適分布を当てはめた。

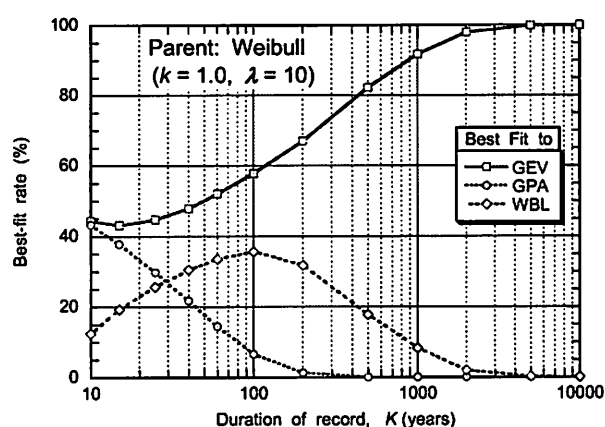


図-15 指数分布から抽出された AM データが 3 種類の極値分布関数に適合する割合

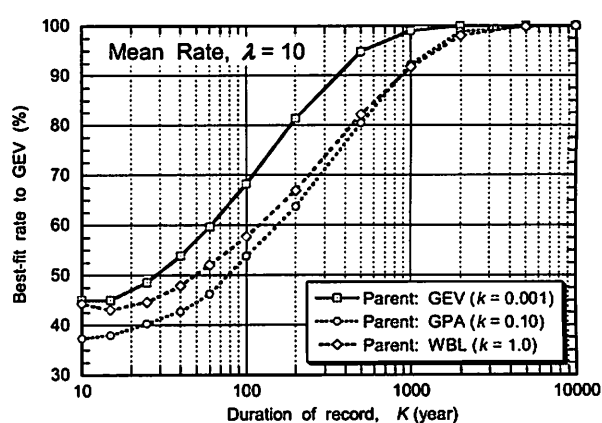


図-16 3 種類の POT 分布から抽出された AM データが GEV 分布関数に適合する割合

AM データの統計期間が $K=1$ 万年と十分に長いときには、式 (12)~(14) の漸近分布が成立し、尺度母数と位置母数が漸近理論の値に一致することが確認されている (合田 2011a)。しかし、統計期間が 1,000 年程度未満では漸近分布が成立しない。図-15 は指数分布に従う POT データから抽出した AM データが 3 種類の CDF に適合する割合を示したものである。式 (12) によれば、こうした AM データはすべて GEV 分布に適合するはずであるが、数値実験の結果では GPA やワイブル分布に適合する割合が無視できない。一方、図-16

は3種類のPOT分布から抽出されたAMデータがGEV分布に適合する割合を示している。式(13)によれば、GPA分布を母集団とするAMデータはGEV分布に漸近するはずである。しかし、数値実験では標本の大きさが $K=1000$ 年であっても、GEV分布への適合率が90%強に留まる。

統計期間が数十年程度のAMデータに対して漸近理論が成立しない原因は、確率密度関数が非対称な母分布において、それから抽出した標本のSkewness (β_1) が母分布のSkewnessよりも平均的に小さな値をとることにある。図-17は形状母数が $k=-0.2$ および $+0.2$ のGPA分布の母集団から抽出した標本について、標本のSkewnessがその大きさによって変わる状況を数値シミュレーション(Sim)および理論計算(Cal)で検証した事例である(Goda, 2011a)。分布のSkewnessはデータの3次の積率で定義され、上位のデータの値の影響を強く受ける。小さな標本の中の最大値は、大きな標本中の最大値に比べて平均的に小さくならざるを得ない。標本の大きさが n のとき、標本中の昇順第 j 番目のデータは期待出現確率が $E[P_j]=(j-1)/n \sim j/n$ の間の値を取ると考えられる。図-17に示した理論計算値(Cal)は、データの期待出現確率を $E[P_j]=(j-1/2)/n$ と設定して計算のものであり、数値シミュレーションの値と整合している。ワイブル分布の母集団から抽出した標本についても、標本が小さくなるにつれてSkewnessが減少することが確認されている。

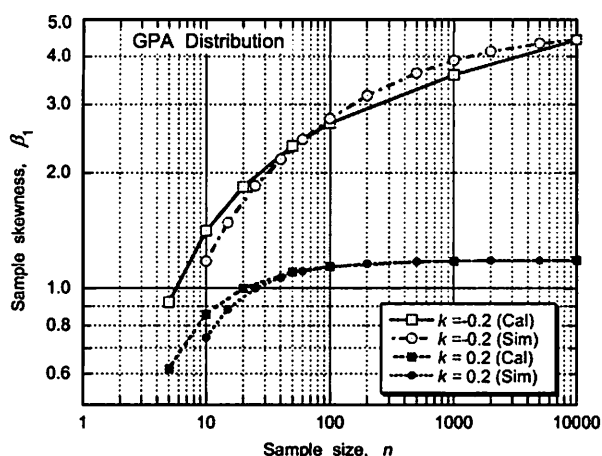


図-17 GPA分布から抽出した標本のSkewnessと標本の大きさの関係

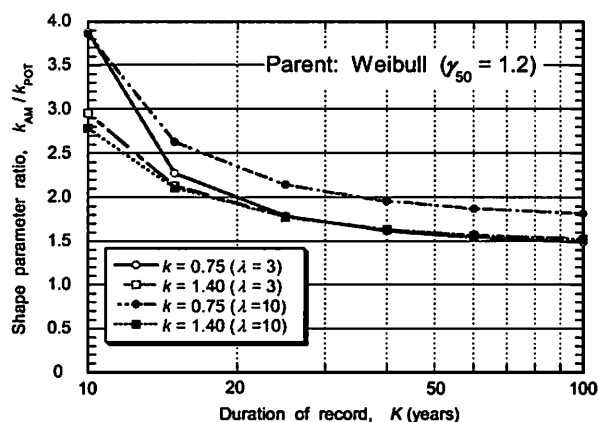


図-18 AMデータを母集団であるワイブル分布に当てはめたときの形状母数の変化

標本のSkewnessが減少すると、それにつれて L -skewness (t_3) も減少し、その結果として標本に当てはめたCDFの形状母数 k の値が増大する。図-18は母集団が $k=0.75$ および 1.4 のワイブル分布から抽出したAMデータにワイブル分布を当てはめたときの形状母数の変化を、 k_{AM}/k_{POT} の比として表示したものである。標本の大きさが $K=20 \sim 100$ 年の範囲では、AMデータに当てはめた形状母数が $1.5 \sim 2.5$ 倍に増大しており、図-13に示した現地波浪のデータを検証している。

7.4. AM資料に基づく確率波高推計値の負の偏り

統計期間が短いAM資料では図-15, 16に例示したように、あらかじめ適合するCFDを選定することができない。そのため、数値実験では確率波高の推計に当たって標本ごとに最適のCFDを用いて10通りの再現期間に対する確率波高を求め、5万組の標本の平均値を算出した。そのようにして求めた確率波高をPOT母集団の値と比べた事例が図-19, 20である。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 $R < 1$ 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-19, 20にはPOTの母集団における確率波高と比較できるように、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ある。再現期間10年付近では年最大値と母集団との確率

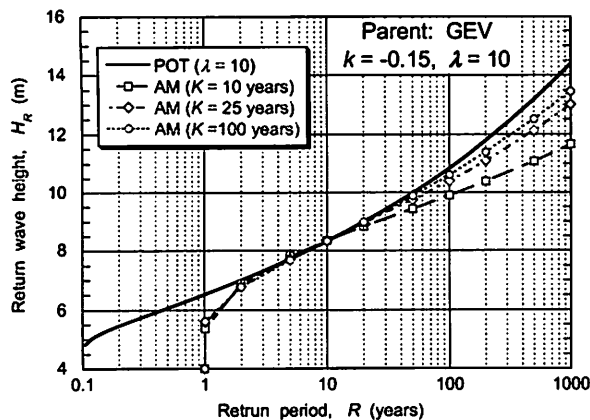


図-19 GEV 分布 ($k = -0.15$) の母集団から抽出された AM 資料の確率波高推計値

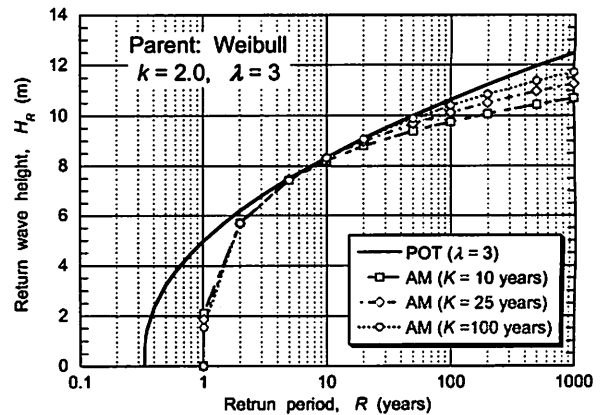


図-20 ワイブル分布 ($k = 2.0$) の母集団から抽出された AM 資料の確率波高推計値

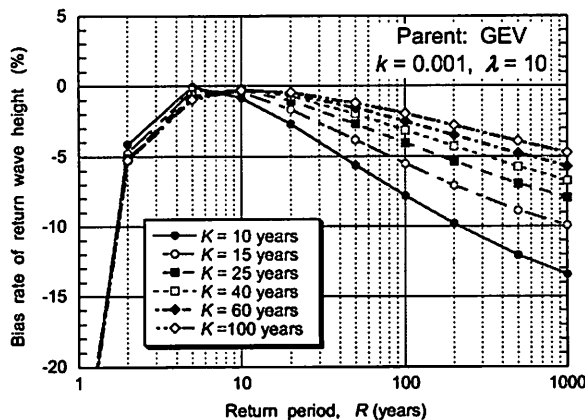


図-21 GEV 分布 ($k = 0.001$) の母集団から抽出された AM 資料の確率波高推計値の偏り率

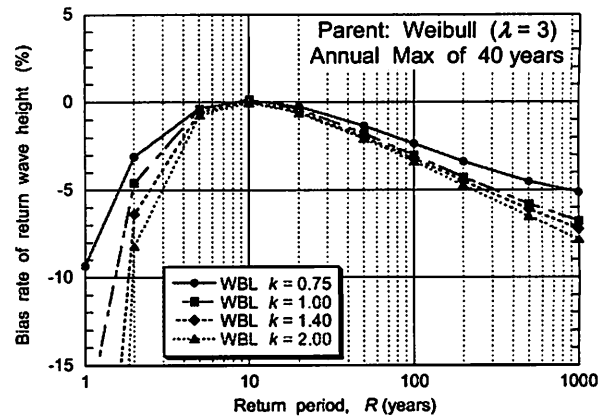


図-22 ワイブル分布の母集団から抽出された統計期間 40 年の AM 資料の確率波高推計値の偏り率

波高に差異はないけれども、再現期間が長い場合には AM 資料が確率波高を過小評価することが明らかである。

こうした AM 資料による確率波高の推計値を母集団である POT 分布の確率波高を基準として偏り率を計算した一例が図-21, 22 である。前者は POT 母集団が $k = 0.001$ の GEV 分布（ほぼ二重指数分布）の場合であって、AM 資料の統計期間が短くなるにつれて負の偏り率が増大することを示している。統計期間が $K = 100$ 年であっても、100 年確率波高が約-2%、1000 年確率波高では約-5%の偏り率である。一方、図-22 は高波の統計資料が最大でも 40 年前後であることを考慮し、統計期間が $K = 40$ 年の AM 資料から推計した確率波高の偏り率を示している。POT 母集団はワイブル分布であるが、母集団の形状母数の影響は小さい。

以上に示したように、数十年程度の観測期間から求めた年最大値資料は、それに基づいて推計した確率波高が数パーセントの負の偏り率、すなわち過小評価を伴うと考えられる。この知見は大規模数値実験によって初めて見いだされたものであって理論的根拠があるわけではない。しかし、極値統計の多くの分野において年最大値資料の統計解析に基づいて土木施設的设计条件を選定しているところから、これまでの設計条件の見直しを行う必要があるのではないかとと思われる。

8. むすび

極値統計解析というのは、たまたま入手できた極値の一つの標本に基づいて対象とする極値事象の長期確

率を推計する作業である。大半の場合に極値事象の母分布関数は未知であり、標本もデータの偶発性あるいは統計的変動の影響を強く受けている。そのため、再現確率統計量が推計されたとしてもその値は相当量の推計誤差を伴っている。これまでに種々の統計理論が展開されてきたものの、いずれもある種の前提条件の下で導かれたものであり、現実の極値事象に普遍的に適用可能なものではない。

本稿では、わが国にあまりなじみのない L -モーメント法に基づく頻度解析法を紹介し、日本海沿岸の高波データへの適用結果を記述した。 L -モーメント法については大規模な数値実験に基づいて、これまで不明瞭であった諸点を明らかにするとともに、理論的根拠には乏しいものの、実務への適用のための幾つかの経験公式を導いた。また、年最大値資料についてこれまで提唱されてきた漸近理論は、標本の大きさが数千以上でなければ成立しないことを数値実験で明らかにした。現実の年最大値資料は統計期間が最長でも 100 年程度であるので、統計理論で言うような GEV 分布の普遍的適用性を受け入れることができない。さらに、統計期間が数十年程度の年最大値資料では、再現期間が 100 年程度以上の再現確率期待値が負の偏り（過小評価）を持つ傾向がある。したがって、極値統計解析では年最大値資料ではなく、できるだけ POT 資料を用いることが望ましい。

本稿で示した事例は高波の極値データであるが、最大風速、最大降水量、高水流量その他の極値事象へも全く同じようにして適用可能である。今後、 L -モーメント法に基づく地域頻度解析法が数多くの分野で活用されることを願って本稿を終える次第である。なお、参考文献中の ECOH/YG 技術論文の諸編は、(株)エコーのホームページ (<http://www.ecoh.co.jp/comp/paper/paper.htm>) からダウンロードし、参照していただければ幸いである。

参考文献

- 河合弘泰・合田良実・久高将信・仲井圭二 (2010) : 地域頻度解析手法によるわが国沿岸の高潮偏差の極値統計解析, 土木学会論文集 B2 (海岸工学), Vol. 66, No. 1, pp. 121-125.
- 合田良実 (2002) : 設計波高に係わる極値統計分布の裾長さパラメータとその意義, 海岸工学論文集, 第 49 巻, pp. 171-175.
- 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, 第 13 章.
- 合田良実 (2009) : 高波の極値統計解析に対する L -moments 法の適用と分布関数の選択について, ECOH/YG 技術論文 No. 10, 2009 年 5 月 7 日, 12p.
- 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 10 月 11 日, 29p.
- 合田良実 (2011a) : 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, ECOH/YG 技術論文 No. 14, 2011 年 3 月 29 日, 22p.
- 合田良実 (2011b) : 年最大値資料に基づいて推計した再現確率統計量の負の偏りについて, 土木学会論文集 B3 (海洋開発), Vol. 67, No. 4 (印刷中).
- 合田良実・久高将信 (2009) : 高波の極値統計解析に対する L -moments 法の適用と分布関数の選択について, ECOH/YG 技術論文 No. 10, 2009 年 5 月 7 日, 12p.
- 合田良実・久高将信 (2011) : 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 (その 2), ECOH/YG 技術論文 No. 15, 2011 年 3 月 22 日, 17p.
- 合田良実・久高将信・河合弘泰 (2009) : L -moments 法を用いた波浪の極値統計解析について, 海岸工学論文集, 第 56 巻, pp. 161-165.
- 合田良実・久高将信・河合弘泰 (2011) : 地域頻度解析手法による日本海沿岸の確率波高と推計と L -モーメント法の改良, 土木学会論文集 B2 (海岸工学), Vol. 67 (印刷中).
- 外山奈央子・水野 量 (2002) : L -moments を用いた地域頻度解析による全国アメダス地点における確率降水量の推定, 気象庁研究時報, 54 巻 5-6 号合併号, pp. 55-100.
- Andrew, M., Smith, O. P. and McKee, J. M. (1985): "Extremal analysis of hindcast wind and wave data at Kodiak, Alaska," Tech. Rept., CERC-85-4, U.S. Army Corps of Engineers, Waterways Experiment Station (Vicksburg, Miss.).

- Blom, G. (1965): "Statistical Estimates and Transformed Beta-Variables," John Wiley & Sons, New York, Chapter 12.
- Chowdhury, J. U., Stedinger, J. R., and Lu, L.-H. (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.
- Filliben, J. J. (1975): "The probability plot correlation coefficient test for normality," *Technometrics*, Vol. 17, No. 1, pp. 111-117.
- Goda, Y. (2010a): "Plotting-position estimator for the *L*-moment method and quantile confidence interval for the GEV, GPA, and Weibull distributions applied for extreme wave analysis," *Coastal Engineering Journal* (accepted for publication).
- Goda, Y. (2010b): "Inherent negative bias of quantile estimates of annual maximum data due to sample size effect: A numerical simulation study," submitted to *Coastal Engineering Journal* for possible publication.
- Goda, Y. (2010c): "A note on sample skewness of asymmetric distribution," submitted to *Coastal Engineering Journal* for possible publication.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1978): "Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form," *Water Resources Res.*, Vol. 15, No. 5, pp. 1049-1054.
- Gringorten, I. I. (1963): "A plotting rule for extreme probability paper," *J. Geophys. Res.*, Vol. 68, No. 3, pp. 813-814.
- Gumbel, E. J. (1953): "Statistics of Extremes," Columbia Univ. Press, New York, 375p.
- Hosking, J. R. M. (1990): "*L*-moments: Analysis and estimation of distributions using linear combinations of order statistics," *J. Roy. Statistical Soc., Series B*, Vol. 52, pp. 105-24.
- Hosking, J. R. M. (2005): LMOMENTS: Fortran routines for use with the method of *L*-moments, Version 3.04, July 2005. (Documentation for the lmoments package is also available at <http://lib.stat.cmu.edu/>)
- Hosking, J. R. M. and Wallis, J. R. (1997): "Regional Frequency Analysis," Cambridge Univ. Press, 224p.
- Jenkinson, A. F. (1955): "The frequency distribution of the annual maximum (or minimum) of meteorological elements," *Quarterly J. Royal Meteorol. Soc.* Vol. 81, pp. 158-171.
- Madsen, H., Pearson, C. P., and Rosbjerg, D. (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.