

極値統計における異常値の棄却基準とその応用

REJECTION CRITERIA FOR OUTLIERS AND THEIR APPLICATIONS IN EXTREME STATISTICS

合田良実*

By Yoshimi GODA

An auxiliary statistic is defined for the largest data of a sample using the sample mean and standard deviation. The 5% and 95% nonexceedance levels of this statistic are employed for rejection of distribution fitting to the sample, and these nonexceedance levels are formulated for the Fisher-Tippett I and II, the Weibull, and the log-normal distributions by means of Monte Carlo simulations with a minimum of 10 000 samples for each combination of sample size, censoring rate, and distribution functions. The new rejection criteria are effectively applied for the extreme data of annual maximum wind speed at 128 sites, annual maxima of daily rainfall at 58 sites, and peak storm wave heights at 29 sites. The population distributions for these extreme data emerge through this analysis.

Keywords: extreme statistics, outliers, distribution fitting, maximum wind speed, daily rainfall, storm waves

1. まえがき

土木構造物は自然外力を対象として設計することが多い。また、建設された構造物は数十年にわたって使用することを前提としている。このため、設計荷重の選定にあたっては、そうした長期間に一度だけ現われるような厳しい自然条件を対象とすることになり、そのための統計解析がいろいろ行われる。毎年の最大風速や最大日雨量、台風や低気圧による高波の極大値などが長期間にわたって収集され、解析される。これらは一般に極値統計とよばれる。

極値統計では、得られた1組のデータ（以下では標本とよぶ）の中に他のデータと比べて著しく大きな値のデータが含まれていることが時々ある。そうした場合にはデータの取得状況などをまず点検しなければならないが、特に変わった点が見られない場合には、そうした異常値を標本に含めるか否かを判断しなければならない。このための判断基準が異常値の棄却基準である。水文統計の分野では、1962年に角屋¹⁾が対数正規分布および対数極値分布における異常値の棄却基準を提案しており、実務面でも利用されている。しかし、この2種類以外の

極値分布に対する棄却基準は発表されていない。

標本中の最大値が異常値とみなされるのは、その標本がある1つの極値分布に従うと予測しているにもかかわらず、その極値分布の性質からは考えられないような値を示すからである。ある極値分布では異常値と判断されても、別の極値分布をあてはめると正常なデータとみなされる場合がある。すなわち、異常値の判定は極値分布の選定と深くかかわっている。この意味で、分布特性の異なる各種の極値分布関数ごとに異常値の棄却基準を求めておき、極値統計の標本に対する分布関数のあてはめを適切に行えるようにしておくことが望まれる。

異常値の棄却基準を理論的に導くことは一般的には困難である。そこで本論文では、数値シミュレーションによって分布関数ごとに多数の標本を作成し、そうした標本中の最大値の分布状況を統計的に調べることによって異常値の棄却基準を導く方式を採用した。具体的には、標本中の最大値の5%および95%非超過限界を標本の大きさ（標本中のデータの個数）の関数として実験式の形で表示した。これらの棄却基準を日本各地の年最大風速、年最大日雨量、および高波の極大値資料に対して適用したところ、いろいろ興味ある結果が得られたので併せて報告する次第である。

* 正会員 工博 横浜国立大学教授 工学部建設学科
(〒240 横浜市保土ヶ谷区常盤台156)

2. 対象とする極値分布関数とその検討方法

(1) 極値分布関数

極値統計では、対数正規分布やグンベル分布（本論文では極値Ⅰ型分布の名称を使う）を使用することが多いけれども、こうした特定の分布関数の妥当性が証明されているわけではない。使用例が多いとか、標本にうまく適合するようにみえるという程度の根拠と思われる。したがって、中西²⁾は「用いた確率分布関数は、本当にその母集団を代表しているのだろうか？」との疑問のもとに各種の分布関数の適合度の良否を客観的に評価する方法を検討しており、また宝・高棹³⁾は同じ目的で Jack-knife 法および Boot-strap 法を取り上げている。

極値統計の1つの標本にいろいろな極値分布関数をあてはめて最適な分布関数を選定する場合には、候補とする関数を統一した基準で選ばなければならない。ここでは、確率密度関数の形状が広い範囲で変化する性質をもっているワイブル分布と極値Ⅱ型分布をまず取り上げる。そして、適用事例の多い極値Ⅰ型分布と対数正規分布も対象とする。前2者については形状母数をあらかじめ各4種類に固定し、2母数型分布として取り扱う。これは、形状母数が異なれば分布特性が大幅に変化するため、たとえ名称が同一であっても別の確率分布として取り扱うべきであると考えからである。

なお江藤ほか⁴⁾は、3母数型の対数正規分布を標本にあてはめたときの母数推定値が不安定となることを例示し、他の3母数型の確率分布でも同様であろうとしている。また、宝・高棹³⁾も3母数型関数について over-fitting の危険性を指摘している。

各種の極値分布の適用性について Cunnane⁵⁾、Kumar と Chander⁶⁾は、洪水の極値統計を対象として総括的に論じている。本論文では、強風・降雨・高波などの極値統計で利用されている下記のものについて検討する。

極値Ⅱ型分布：

$$F(x) = \exp\{-[1+(x-B)/kA]^{-k}\}$$

$$: k=2.5, 10/3, 5.0, \& 10.0 \dots\dots\dots(1)$$

極値Ⅰ型分布：

$$F(x) = \exp\{-\exp[-(x-B)/A]\} \dots\dots\dots(2)$$

ワイブル分布：

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

$$: k=0.75, 1.0, 1.4, \& 2.0 \dots\dots\dots(3)$$

対数正規分布：

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}tA} \exp\left[-\frac{(\ln t-B)^2}{2A^2}\right] dt \dots\dots(4)$$

式(1)~(4)において $F(x)$ は確率変数 x の分布関数、 A は尺度母数、 B は位置母数、 k は形状母数である。極値Ⅱ型分布に対する式(1)の表式は小野澤・合田⁷⁾

によるもので、Jenkinson⁸⁾の一般化極値分布(GEV)の形状母数 x を $x = -1/k$ と置き換えたものに等しい。また、角屋⁹⁾の対数極値分布A型とも同等である。形状母数の値は、 $1/k$ が 0.1, 0.2, 0.3, & 0.4 となるように選定した。ワイブル分布における形状母数は、先に合田¹⁰⁾が吟味した結果による。

なお、対数正規分布における A および B は、確率変数 x の自然対数 $X = \ln x$ に対する母数であって次元をもたず、他の分布関数に対するものと意味がやや異なる。また、標本に対して対数正規分布を適用するときは、極値統計量の自然対数について正規分布をあてはめるので、異常値の棄却や分布関数の適合度の検討問題では正規分布としての取り扱いをする。

(2) プロットイング・ポジション公式

標本に対する極値分布関数のあてはめには積率法、最尤法、PWM 法その他が用いられるが、本論文では最小2乗法を使用する。これは以下の理由による。

- ① 再現確率統計量として偏りのない推定値を得ることが可能である。
- ② 標本に対する適合度が相関係数の値で客観的に表示され、分布関数の選別が可能である。
- ③ 部分極値資料に対してもそのまま適用できる。
- ④ 計算が簡単でわかりやすい。

このうち第1の理由が成立するためには、プロットイング・ポジション公式として最適のものを選ぶ必要がある。標本中のデータを大きさの順に並べ替えて順序統計量に組み替えたとき、昇順で i 番目のデータ x_i ($i=1$ は最小値)あるいは降順で m 番目のデータ x_m ($m=1$ は最大値)の非超過確率 F_i あるいは F_m を推定する方法はいろいろあるが、簡略法としては式(5)、(6)が使われる。

$$\hat{F}_i(x_i) = (i-\alpha)/(N_T+\beta) \dots\dots\dots(5)$$

$$\hat{F}_m(x_m) = 1 - (m-\alpha')/(N_T+\beta) \dots\dots\dots(6)$$

ここに、 N_T は標本の大きさ N そのものではなく、対象とする期間中に発生した極値の総数である。毎年最大資料の場合には $N_T = N$ であるが、波浪の極値統計などのように極値の中の大きなものだけを対象とする部分極値資料では $N_T > N$ である。

最適プロットイング・ポジション公式の選択は、式(5)、(6)の定数 α などの選定に依存している。今回取り上げた極値分布に対しては、Table 1 の定数値を

Table 1 Coefficients of plotting position formulas.

分布関数名	α	α'	β	文献
極値Ⅱ型分布	0.44-0.41/k	0.44+0.52/k	0.12-0.11/k	小野澤・合田 ⁷⁾
極値Ⅰ型分布	0.44	0.44	0.12	Gringorten ¹¹⁾
ワイブル分布	0.60-0.50/√k	0.20+0.27/√k	0.20+0.23/√k	合田 ⁷⁾
対数正規分布	0.375	0.375	0.25	Blom ¹²⁾
正規分布	0.375	0.375	0.25	Blom ¹²⁾

使うことにより、再現確率統計量の推定値の偏りをほぼ0とすることができる。

(3) 数値シミュレーションの方法

式(1)~(4)の極値分布関数のおのおのに対して、モンテカルロ法で極値統計の標本を作成した。すなわち、[0, 1]の1様乱数を電子計算機で作成させてこれをxの非超過確率であるF(x)であるとみなし、この逆関数を求めてxの値とする。疑似1様乱数の発生には、合田¹³⁾が2×10⁷個までの乱数列についてのランダム性を確認している乗算合同法の一手法を使用した。

標本としてはN_T=10~800の範囲を対象とし、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ν=N/N_Tを0.25, 0.5, および1.0の3種類に固定した。解析したのは、ν=0.25ではN=10~200, ν=0.5と1.0ではN=10~400の標本である。原則として1万組の標本について統計解析し、Nが100以下の場合にはこれを2~4回繰り返してその平均を用いた。なお著者¹⁰⁾は、データ採択率がν=1の場合の資料を全数極値資料、ν<1の資料を部分極値資料とよんでいる。

3. 標本中の最大値の分布特性の理論的検討

(1) 母集団の統計量が既知の場合

いま、N個のデータからなる標本中の最大値をx₁で表わし、その無次元表示として平均値x̄との差を標準偏差sで除したものをを用いる。これをξで表わすと、

$$\xi = (x_1 - \bar{x}) / s \dots\dots\dots(7)$$

この標本においてはどのデータもx₁を超えることはないので、Nデータ中の最大値x₁の非超過確率は、

$$P(x_1) = [F(x_1)]^N \dots\dots\dots(8)$$

で与えられる。ただしF(x₁)は母集団におけるx₁の非超過確率である。

これから逆に、標本中の最大値としての非超過確率Pを指定したときのx_{1,P}の値は、式(8)を解いて次のように求められる。

$$x_{1,P} = F^{-1}(P^{1/N}) \dots\dots\dots(9)$$

たとえば、極値I型分布については次のようになる。

$$x_{1,P} = B - A \ln(-\ln P^{1/N}) \dots\dots\dots(10)$$

一方、母集団の平均値はx̄=B+Aγ、標準偏差はσ=πA/√6であるので、最大値の無次元偏差ξは、

$$\xi_P = \sqrt{6} [-\ln(-\ln P^{1/N}) - \gamma] / \pi \dots\dots\dots(11)$$

となり、尺度母数Aおよび位置母数Bの値によらないことが明らかである。なお、γはオイラーの定数である。他の極値分布関数についても同様に、ξが母数値に影響されない補助統計量であることが導かれる。

(2) 母集団の統計量が未知の場合

実際の問題では母集団自体が未知であり、その平均値や標準偏差を知ることは不可能である。したがって、式

(7)の定義式では標本平均値と標本偏差を使用する。そうした場合のξ_P値については、正規分布の解のみが知られている。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¹⁴⁾)の応用であり、角屋¹⁾は最大値が異常値として棄却される場合を想定して、最大値x₁を除いたN-1個のデータでx̄およびsを計算する場合を示している。ここではデータを異常値として棄却するのではなく、あてはめ分布関数の方を棄却する場合も考えるので、x̄, sは全データで計算するものとする。このときのξ_PはF分布の理論によって近似的に次式で求められる。

$$\xi_P = \sqrt{\frac{(N-1)F(1, N-2; \alpha)}{N-2+F(1, N-2; \alpha)}} \dots\dots\dots(12)$$

ここに、F(1, N-2; α)は超過確率αに対する自由度(1, N-2)のF分布である。

Thompsonの棄却検定法の考え方では最大値x₁および最小値x_Nを区別せず、平均値からの偏差の絶対値を問題にする。したがって、式(12)のF分布のαは母集団中のx₁の超過確率の2倍に相当し、Pが指定された場合には式(8)によってα=2(1-P^{1/N})として与えられる¹⁵⁾。

4. 数値シミュレーションによる最大値の分布特性と異常値の棄却基準

(1) 最大値偏差の累積分布

式(7)で定義した標本中の最大値の無次元偏差ξをモンテカルロ法で作成した各標本について計算し、その累積分布を求めた。これは、同一条件の標本から1個ずつ得られた各1万個のξ値を単純に大きさの順に並べ替えて作成した。Fig.1は全数極値資料の場合の正規分布に対する結果であり、式(12)によるF分布の理論値も示している¹⁵⁾。ξ値の累積分布の上の部分では数値シミュレーションの結果と理論値がよく一致しているが、累積分布の下の部分では理論値が小さめである。特に、標本が小さいときにこの傾向が強い。

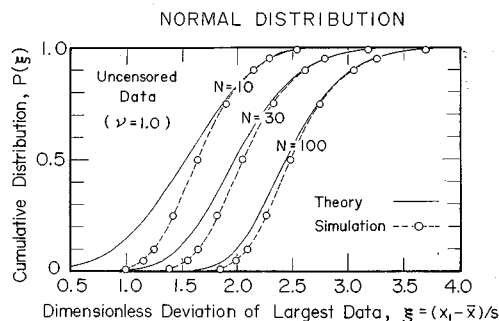


Fig.1 Cumulative distribution of the dimensionless deviation of outlier ξ in the Normal distribution.

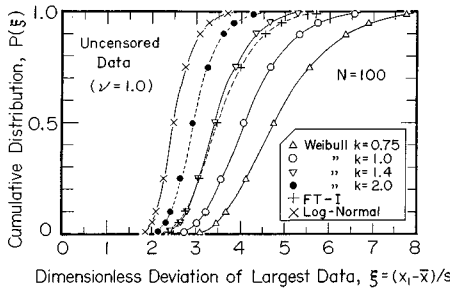


Fig. 2 Cumulative distributions of ξ in various distribution functions ($N=100$).

Thompson の棄却検定理論は、本来は j 組の標本のうちの 1 組が正規分布に所属するか否かが疑問のとき、その標本の平均値を指標として判定するためのものである。標本の大きさについては制限がないところから、 x_1 の所属する組はその大きさが $n=1$ であるとして F 分布の理論を適用している。ただし、この適用に際しては、 x_1 が j 組の全標本中の最大値であるという条件が考慮されていない。このため、最大値 x_1 が相対的に小さい標本では \bar{x} および s がランダムな標本に対する期待値よりも小さくなり、したがって ξ 値が理論値よりも大きくなると考えられる。しかし、標本の大きさによって ξ 値が変化する傾向は一致しているので、数値シミュレーションは最大値偏差 ξ の特性を十分に再現していると考え、以下においては数値シミュレーション結果に基づいて考察を進める。

最大値の無次元偏差 ξ の累積分布は、母集団によって異なる形状を示す。Fig. 2 は、全数極値資料における $N=100$ の標本の場合であり、極値 I 型分布、ワイブル分布、および対数正規分布を比較している。ワイブル分布の $k=0.75$ の場合が最も ξ 値が大きく、かつ累積分布の幅が広い。極値 I 型分布はワイブル分布の $k=1.4$ と非常によく似た累積分布を示す。極値 II 型分布について同じように調べてみると、 $k=10$ の極値 II 型分布は $k=1.0$ のワイブル分布、 $k=5$ の極値 II 型分布は $k=0.75$ のワイブル分布と類似した ξ の累積分布となっていることがわかる。

(2) 最大値偏差の 95% 非超過確率値の特性

Fig. 1, 2 に示すように、標本の大きさが一定であっても、最大値の無次元偏差はある幅で分布する。そうした分布の上限界の指標として、ここでは 95% 非超過限界値を取り上げる。数値シミュレーション結果に基づいて $\xi_{95\%}$ を求め、標本の大きさ N の関数として表示したのが Fig. 3 である。これは極値 II 型分布の部分極値資料の場合であり、極値 I 型分布は $k=\infty$ の極値 II 型分布に相当するので、これも参考として記入してある。なお、他の分布関数について $\xi_{95\%}$ と N の関係を調べた結果は

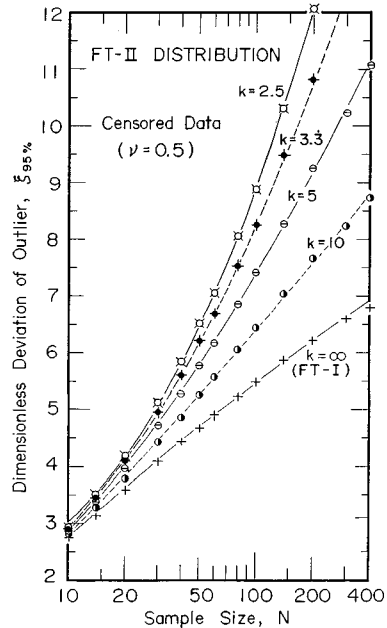


Fig. 3 The 95% nonexceedance level of the dimensionless deviation of outlier $\xi_{95\%}$.

文献¹⁵⁾に発表しているが、 N の増大につれて $\xi_{95\%}$ が増える傾向は同じである。

このような最大値偏差の 95% 非超過確率値 $\xi_{95\%}$ と標本の大きさ N との関係に対して、次のような実験式をあてはめ、係数 a, b , および c を分布関数ごとに設定した。

$$\xi_{95\%} = a + b \ln N + c (\ln N)^2 \dots \dots \dots (13)$$

なお、先に合田・李¹⁶⁾は $N=10\sim 100$ の全数極値資料を対象として、右辺第 3 項を 0 とした直線式を使用したが、標本の大きさを $N=400$ まで拡張し、また部分極値資料についても検討した結果¹¹⁾、一般的には $\xi_{95\%}$ と N との関係を曲線式で表わす必要が認められたので、曲線の中では最も単純な 2 次式を導入したものである。

係数値の設定は次のようにした。すなわち、データ採択率ごとに式 (13) 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 3 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 を再計算した。そして a および b がデータ採択率 ν によって変化する傾向をできるだけ単純な関係形で表示することにして、Table 2 のように取りまとめた。Fig. 3 中の曲線は、式 (13) に Table 2 の係数を使って計算した結果である。また、Fig. 4 は採択データ数が $N=60$ の場合についてデータ採択率 ν の影響を示している。曲線はいずれも実験式による値である。

Fig. 3, 4 にみられるように、実験式は数値シミュレーション結果をよく表示している。あてはめ誤差は、 ± 1

Table 2 Empirical coefficients for the 95 % nonexceedance level of the largest sample data.

分布関数	係数 a	係数 b	係数 c
極値 II 型 (k=10/4)	$4.653 - 1.076 \nu^{1/2}$	$-2.047 + 0.307 \nu^{1/2}$	0.635
同上 (k=10/3)	$3.217 - 1.216 \nu^{1/4}$	$-0.903 + 0.294 \nu^{1/4}$	0.427
同上 (k=10/2)	$0.589 - 0.038 \nu^2$	$0.518 - 0.045 \nu^2$	0.210
同上 (k=10/1)	$-0.371 + 0.171 \nu^2$	$1.283 - 0.133 \nu^2$	0.045
極値 I 型分布	$-0.579 + 0.468 \nu$	$1.496 - 0.227 \nu^2$	-0.038
ワイブル (k=0.75)	$-0.256 - 0.632 \nu^2$	$1.289 + 0.254 \nu^2$	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
対数正規分布	$0.178 + 0.740 \nu$	$1.148 - 0.480 \nu^{3/2}$	-0.035

Table 3 Empirical coefficients for the 5 % nonexceedance level of the largest sample data.

分布関数	係数 a	係数 b	係数 c
極値 II 型 (k=10/4)	$1.481 - 0.126 \nu^{1/4}$	$-0.331 - 0.031 \nu^2$	0.192
同上 (k=10/3)	1.025	$-0.077 - 0.050 \nu^2$	0.143
同上 (k=10/2)	$0.700 + 0.060 \nu^2$	$0.139 - 0.076 \nu^2$	0.100
同上 (k=10/1)	$0.424 + 0.088 \nu^2$	$0.329 - 0.094 \nu^2$	0.061
極値 I 型分布	$0.257 + 0.133 \nu^2$	$0.452 - 0.118 \nu^2$	0.032
ワイブル (k=0.75)	$0.534 - 0.162 \nu$	$0.277 + 0.095 \nu$	0.065
ワイブル (k=1.0)	0.308	0.423	0.037
ワイブル (k=1.4)	$0.192 + 0.126 \nu^{3/2}$	$0.501 - 0.081 \nu^{3/2}$	0.018
ワイブル (k=2.0)	$0.050 + 0.182 \nu^{3/2}$	$0.592 - 0.139 \nu^{3/2}$	0
対数正規分布	$0.042 + 0.270 \nu$	$0.581 - 0.217 \nu^{3/2}$	0

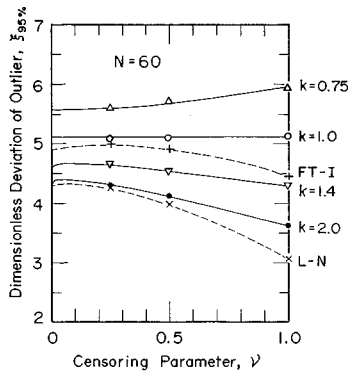


Fig. 4 Effect of censoring parameter ν on $\xi_{95\%}$.

% 以内が 76 %, $\pm 2\%$ 以内が 97 % である。極値分布関数のなかでは、極値 II 型分布が上方にカーブする傾向が強いために誤差がやや大きく、特に $N=10$ のところで誤差が大きくなっている。また、対数正規分布も $\pm 1 \sim 2\%$ の誤差のものが多い。

データ採択率の影響は分布関数によって異なる。極値 I 型分布、対数正規分布、および k が 1 を超えるワイブル分布では、 ν が小さくなるにつれて $\xi_{95\%}$ 値が増大する。 $k=0.75$ のワイブル分布では ν の減少につれて $\xi_{95\%}$ 値も小さくなる。 $k=1.0$ のワイブル分布 (すなわち指数分布) では、データ採択率に影響されることなく一定値をとる。これは、指数分布では分布の裾のどの部分を取り出しても分布形状が変わらないことによるもので、 $\xi_{95\%}$ 値に限らずどの統計量についても成立する。極値 II 型分布の場合は、 k の値が大きくて極値 I 型分布に近いものは同じような傾向を示すが、 $k=10/3$ や 2.5 などでは標本の大きさによって ν の影響が逆に出ることがある。ただし、極値 II 型分布ではデータ採択率の影響が一般的に小さい。

Table 2 の係数値は $\nu=0.25, 0.5$, および 1.0 の 3 点のデータで定めたものなので、Table 2 はデータ採択率が $\nu \geq 0.15$ 程度以上の範囲で使用するのが妥当であろう。次に述べる Table 3 の係数についても同様である。

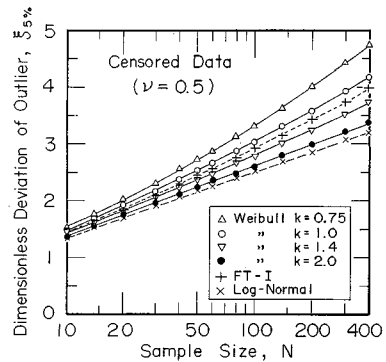


Fig. 5 The 5 % nonexceedance level of the dimensionless deviation of outlier $\xi_{5\%}$ (1).

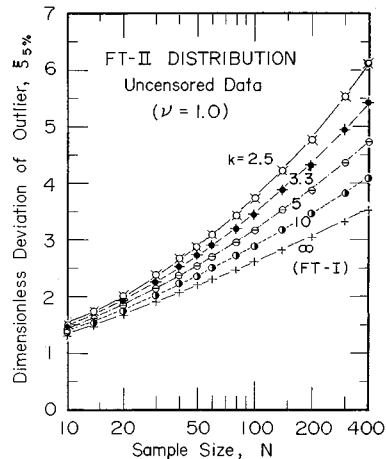


Fig. 6 The 5 % nonexceedance level of the dimensionless deviation of outlier $\xi_{5\%}$ (2).

(3) 最大値偏差の 5 % 非超過確率値の特性

次に、最大値の無次元偏差の累積分布の下限値の指標として 5 % 非超過限界値について調べた例が Fig. 5~7 である。Fig. 5 は極値 I 型分布、ワイブル分布、および対数正規分布のグループについて $\nu=0.5$ の部分極値資料の結果を示しており、Fig. 6 は極値 II 型分布の全数極値資料の結果である。また、Fig. 7 は後者のグループに

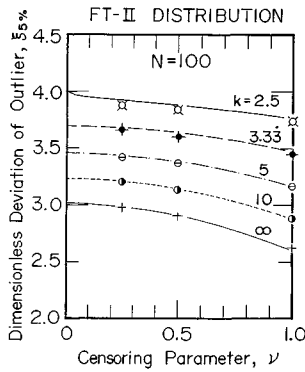


Fig.7 Effect of censoring parameter ν on $\xi_{5\%}$.

ついてデータ採択率の影響を $N=100$ の場合について示したものである。 $\xi_{5\%}$ 値に対して式 (13) をあてはめて係数値を設定した結果は Table 3 に示すとおりであり、またこれによる計算値は Fig. 5~7 の中にそれぞれ曲線で示してある。

実験式のあてはめ精度は $\xi_{5\%}$ の場合よりも高く、全体では $\pm 1\%$ 以内が 92% 、 $\pm 2\%$ 以内が 99% である。特に極値 II 型分布を除くと、 97% が $\pm 1\%$ 以内であって、 $\pm 2\%$ を超えるものはない。

(4) 異常値の棄却基準の提案

前 2 項で提示した最大値の無次元偏差 ξ の 5% および 95% 非超過確率値は、それぞれ標本中の最大値が過小あるいは過大である場合の棄却基準として利用できる。すなわち、標本について求めた ξ 値が $\xi_{5\%} \leq \xi \leq \xi_{95\%}$ の範囲に入っていれば、その標本中の最大値は正常なデータであるとみなし、 $\xi < \xi_{5\%}$ あるいは $\xi > \xi_{95\%}$ のときはそのような事象が起きる確率がそれぞれ 5% 未満であるので、その最大値を異常値とみなす。したがって、母集団から正しく抽出された標本であっても、100 組のうちで 10 組は最大値が異常値であると判定される危険率を伴っている。

このように、標本中の最大値の無次元偏差を利用する異常値の棄却基準をここでは DOL (Deviation of Outlier) 基準とよび、 $\xi_{5\%}$ によるものを下向 DOL 基準、 $\xi_{95\%}$ を使うものを上向 DOL 基準と称する。

与えられた極値資料の標本が特定の極値分布の母集団に属することが一般に認められている場合には、この DOL 基準を使って標本中の最大値の棄却検定を行うことが可能である。しかし、極値資料の母集団が確定していない場合には、標本中の最大値を異常値とみなして棄却するのではなく、逆に標本にあてはめるべき分布関数に対して棄却検定を行う方が合理的である。

DOL 基準の $\xi_{5\%}$ および $\xi_{95\%}$ は、Fig. 3~7 に例示したように分布関数によって異なる値をとる。対象とする

標本の最大値偏差がある分布関数に対する DOL 基準によって異常値と判定されても、他の分布関数の DOL 基準を適用すると正常と判定される場合が起こり得る。その極値データの現象について、今までの多くの検討事例によって前者の分布関数があてはまることが確認されていなければ、その標本に対しては前者のあてはめを棄却し、後者の分布関数をあてはめる方が無理がない。

ここで提案した DOL 棄却基準は、分布関数のあてはめ方法に依存しないことに注意する必要がある。式 (7) で定義したように、 ξ 値は標本中の最大値 x 、平均値 \bar{x} 、および標本標準偏差 s の 3 個の量のみによって定まるものであり、たとえば最尤法を使用する場合でもこの棄却基準は適用できる。

5. 極値分布関数の棄却検定の具体例

(1) 基本的考え方

強風や大雨、あるいは高波などの極値問題では、極値データがどのような母集団に所属するかが確定していない。年最大風速については、気象庁が 1971 年に極値 I 型分布を日本各地 131 地点のデータに適用しており¹⁷⁾、その後のデータを追加した藤野ほか¹⁸⁾の解析でも極値 I 型分布の妥当性がほぼ確かめられている。また、アメリカの風資料を解析した Simiu と Filliben¹⁹⁾ も、ハリケーンの影響が強くて極値 II 型分布が適用される地点もあるけれども、一般的には極値 I 型分布があてはまるとしている。

これに対して、年最大日雨量 (正式には日降水量であるが、年最大値については雨量となる) に関してはいろいろな分布関数のあてはめが試みられているが、年最大日雨量の母集団と認定できるものは見出されていない。雨の場合は、確率分布の裾の方に大きな値のデータが出現することが時々あり、母集団の確定を困難にしている。江藤・室田²⁰⁾は裾を長く引く確率分布として平方根指数型最大値分布を提案し、江藤ほか⁴⁾はこれを全国の年最大日雨量資料に適用して極値 I 型分布よりも適合性が良いことを例証している。ただし、この分布は母数の意味が他の確率分布と異なり、また分布関数の形状が 1 つに固定されているので、一般性にやや欠ける。

高波の極値資料の場合には、資料の統計年数が不足していることもあって、母集団を確定することなしに、与えられた標本に最も適合する分布関数をあてはめるのが一般的である。Petruaskas と Aagaard²¹⁾は極値 I 型分布と形状母数を 7 種類に指定したワイブル分布の中から最適関数を選ぶ方式を提案し、実務面ではこの方式が使われる例が多い。また、対数正規分布を適用する事例もある。

極値資料に対する母集団を探索するためには、多数の

地点の標本について統一的な解析を行い、その結果を統計的に解釈することが必要である。前章で提案した分布関数のDOL棄却基準はこうした目的で使用するのに好適であり、その適用を試みる。ただし、DOL基準は標本中の最大値に着目するものなので、ある分布関数に対する標本のデータ全体の適合度をみるのには適当でない。この適合度は、データのあてはめ方法に依存し、最尤法であれば尤度あるいは対数尤度などが指標となろう。最小2乗法を使う場合には、順序統計量 x_m とその非超過確率の推定値 \hat{F}_m によって定義される基準化変数 y_m との間の相関係数 r の値が指標となる。したがって、この相関係数を使って分布関数の棄却基準を作成することが可能であり、著者は先に REC (REsidue of Correlation coefficient) 基準と称するものを提案^{15),16)}している。そこで、本章における各種分布関数の棄却検定では、DOL基準と併せてREC基準も適用して検定能力を高めることにした。なお、極値II型分布のREC基準については合田・小野澤²²⁾によっている。

(2) 年最大風速資料に対する極値分布

気象庁¹⁷⁾が先に解析した年最大風速資料は1929~1966年間のデータであり、今回はこれに1987年までのデータを追加した。ただし、1967年以降に筑波山、富士山、および雲仙岳が風速統計から除外されていて統計期間が十分に長く取れないので、残りの128地点を対象とした。これらの資料に各分布関数をあてはめた結果のうち、強風統計で使われる極値II型分布、極値I型分布、およびこれらよりも分布幅の狭い $k=2.0$ のワイブル分布に対する棄却検定の結果をTable 4に示す。各地点の観測最大風速は明示していないが、表中の数値を使い $x_1 = \bar{x} + \xi s$ として算定できる。本論文では確率風速の算定が主目的ではないので、長期的な風速の変動傾向については特に検討していない。

最大値の無次元偏差として最も大きいのは甲府の $\xi=5.57$ で、次が大阪の $\xi=5.52$ である。風速の絶対値としては大阪が42.2 m/s、甲府が32.2 m/s であって、大阪の方が大きく、これは1934年の室戸台風のときの値である。 ξ 値がこのように大きいときは極値II型分布の $k=10$ や極値I型分布、ワイブル分布の $k=2.0$ などはあてはめが棄却される。逆に、 ξ 値が最も小さいのは日光の $\xi=1.78$ 、次が仙台の1.89、長野の1.92などで、これらの地点ではここで取り上げたとの極値分布もうまく適合しない。

Table 4 をみると、地点ごとに各分布関数の適合状況が異なり、姫路、神戸などのようにどの分布関数でも棄却されない地点もあり、一方では上記の日光、仙台、長野のようにどの分布関数も棄却される所もある。こうした状況のなかで、棄却されるケースが最も少ない関数を

Table 4 Calibration of various distributions for annual maximum wind speed by rejection criteria.

地点名	年数 N	平均値 \bar{x}	標準偏差 s	ξ 値	棄却検定 I II III IV V IX	地点名	年数 N	平均値 \bar{x}	標準偏差 s	ξ 値	棄却検定 I II III IV V IX
室蘭	59	20.2	4.44	4.31	○○○○○▲	福井	59	15.9	3.61	4.66	○○○×▲▲
苫小牧	45	21.8	3.55	2.92	×○○○○○	敦賀	47	17.3	3.99	3.47	○○○○○▲
苫小牧	59	26.9	4.73	3.09	○○○○○○	舞鶴	40	17.6	4.64	4.28	○○○○○▲
帯広	59	14.5	2.47	3.09	▼××○○○	豊岡	59	15.3	5.43	2.57	▼▼▼○○○
帯広	59	20.5	3.38	2.74	▼▼▼○○○	岡	40	14.0	3.12	3.03	○○○○○○
羽前	59	17.5	2.83	3.31	○○○○○○	姫路	59	18.1	3.58	3.31	○○○○○○
羽前	59	19.8	4.06	2.29	▼▼▼○○○	神戸	59	15.2	4.93	5.52	○○○○▲▲
網走	45	22.1	3.81	2.47	▼▼▼○○○	大阪	59	12.4	3.96	4.41	○○○○○○
網走	45	17.4	3.60	2.09	▼▼▼○○○	京都	59	15.3	3.62	4.57	○○○○○○
雄勝	50	20.3	2.51	2.18	▼▼▼○○○	伊吹山	59	41.8	7.68	2.75	▼▼○○○○
雄勝	43	18.4	3.56	2.23	▼▼▼○○○	上野	50	13.8	3.45	2.97	×○○○○○
雄勝	45	22.5	5.31	2.80	○○○○○○	尾鷲	48	14.0	3.32	3.71	○○○○○○
雄勝	59	11.9	1.92	9.22	××○○○○	潮岬	59	22.1	5.00	2.14	▼▼▼○○○
雄勝	41	16.4	3.29	4.89	○○○○▲▲	和歌山	58	18.5	5.13	3.74	▼▼▼○○○
雄勝	48	13.7	2.27	2.15	▼▼▼○○○	洲本	59	21.0	7.77	3.10	○○○○○○
雄勝	45	16.2	3.67	2.89	○○○○○○						
小豆	44	16.5	5.21	3.51	○○○○○○	徳島	58	20.6	5.99	2.84	▼▼○○○○
小豆	59	29.6	7.13	3.50	○○○○○○	岡山	42	35.7	6.09	3.46	○○○○○○
小豆	47	27.4	4.57	2.04	▼▼▼○○○	室戸	59	31.1	8.40	3.10	○○○○○○
小豆	47	17.1	3.62	2.30	▼▼▼○○○	高知	48	16.1	5.19	2.28	▼▼▼○○○
						高知	47	20.8	4.60	2.92	○○○○○○
むつ	46	19.7	2.83	2.97	×××○○○	宿毛	45	20.4	5.89	3.35	○○○○○○
むつ	51	19.9	4.01	2.83	▼▼○○○○	宇和島	59	20.4	4.59	2.75	▼▼○○○○
深川	58	17.3	3.88	2.86	▼▼○○○○	松山	59	12.9	3.16	3.04	▼○○○○○
深川	48	19.1	4.02	2.89	○○○○○○	多度津	59	19.9	4.82	3.35	×○○○○○
深川	59	21.2	3.30	2.61	▼▼○○○○	高松	48	15.7	3.43	2.07	▼▼▼○○○
深川	51	23.4	5.63	2.25	▼▼○○○○						
山形	59	11.1	2.40	3.39	○○○○○○	岡山	58	13.5	3.10	3.69	○○○○○○
山形	59	15.9	1.88	2.11	▼▼○○○○	福山	42	12.0	2.61	1.97	○○○○○○
宮古	49	13.2	2.85	2.87	××○○○○	呉	40	14.3	3.08	3.04	○○○○○○
宮古	59	17.4	3.42	2.66	▼▼○○○○	広島	53	18.2	5.17	2.67	○○○○○○
仙台	59	15.4	2.58	1.89	▼▼○○○○	津山	45	17.5	4.88	2.79	×○○○○○
仙台	59	17.4	3.38	2.30	▼▼○○○○	鳥取	45	16.9	3.16	3.52	○○○○○○
小名	48	17.4	1.93	2.06	▼▼○○○○	米子	48	15.0	2.95	2.62	○○○○○○
						境	59	14.7	3.32	3.03	○○○○○○
相川	59	23.0	3.30	2.38	▼▼○○○○	西郷	48	18.8	3.19	1.99	▼▼○○○○
新潟	49	19.7	3.31	2.94	×○○○○○	松江	47	18.0	2.47	2.70	▼○○○○○
新潟	59	14.2	1.87	3.63	××○○○○	浜田	59	22.0	2.78	2.76	▼○○○○○
新潟	59	16.8	3.31	1.98	▼▼○○○○	萩	39	19.4	3.54	3.10	○○○○○○
宇都宮	53	14.7	3.78	2.78	▼○○○○○	下野	59	21.0	4.34	2.48	▼○○○○○
日光	44	15.7	1.78	1.78	▼▼○○○○						
日光	59	16.6	4.33	3.02	○○○○○○	姫路	52	15.2	2.80	2.95	×○○○○○
日光	59	14.7	2.36	3.25	○○○○○○	福岡	49	17.1	4.31	3.12	○○○○○○
秩父	59	11.1	2.36	2.88	▼○○○○○	姫路	59	17.3	3.22	3.13	○○○○○○
秩父	58	15.6	3.02	5.57	○○○○▲▲	平戸	48	19.5	5.23	3.14	○○○○○○
河口湖	55	18.0	4.63	3.13	×○○○○○	佐佐木	36	14.4	2.80	2.92	○○○○○○
河口湖	59	19.8	5.40	3.29	○○○○○○	長崎	59	18.3	6.28	3.08	○○○○○○
東武	59	15.5	3.37	4.22	○○○○○○	佐賀	59	15.1	5.07	4.17	○○○○○○
東武	59	25.5	6.91	3.55	○○○○○○	日田	45	12.2	3.14	2.47	▼○○○○○
八丈島	59	22.1	5.59	3.57	○○○○○○	大分	59	14.8	3.19	2.82	▼○○○○○
三宅島	59	27.6	5.59	3.42	○○○○○○	阿蘇山	55	22.6	2.62	3.63	○○○○○○
三宅島	42	26.9	5.56	2.91	○○○○○○	熊本	59	14.3	3.08	3.35	○○○○○○
大島	48	23.7	4.08	2.49	▼▼○○○○	熊本	38	19.8	5.86	2.88	○○○○○○
						人吉	45	16.0	5.57	3.23	○○○○○○
石崎	48	26.3	7.03	3.88	○○○○○○	宮崎	59	17.8	5.51	3.64	○○○○○○
石崎	56	22.1	4.43	3.87	○○○○○○	油津	39	21.3	5.89	2.10	▼▼○○○○
浜松	58	15.4	2.63	3.68	×○○○○○	都立	45	17.9	6.38	2.69	▼○○○○○
静岡	48	14.4	2.94	2.40	▼▼○○○○	阿久根	48	18.5	4.56	3.27	○○○○○○
三島	57	16.9	4.54	3.27	×○○○○○	鹿児島	59	18.5	5.73	2.96	○○○○○○
飯田	59	13.0	2.75	3.57	○○○○○○	枕崎	59	28.1	7.35	2.55	▼▼○○○○
松本	43	15.4	1.65	3.22	○○○○○○	種子島	37	25.8	6.48	2.62	○○○○○○
松本	59	13.7	3.02	2.23	○○○○○○	屋久島	50	27.1	8.39	2.79	○○○○○○
軽井沢	58	12.4	3.71	3.79	○○○○○○	名瀬	58	19.0	5.70	2.27	▼▼○○○○
長野	59	13.7	1.67	1.92	▼▼○○○○						
富山	49	16.6	3.55	2.36	▼▼○○○○						
伏木	59	15.2	3.18	4.12	○○○○○○						
輪島	52	15.8	3.56	2.55	▼○○○○○						
金沢	59	16.8	4.17	3.23	×○○○○○						
高山	59	11.8	2.84	2.90	○○○○○○						
岐阜	59	15.3	4.67	3.54	○○○○○○						
名古屋	59	16.3	4.92	4.03	○○○○○○						

注1) 分布関数:
 I = 極値II型分布 (k = 2.5)
 II = 同上 (k = 3.3)
 III = 同上 (k = 5.0)
 IV = 同上 (k = 10.0)
 V = 極値I型分布
 IX = ワイブル分布 (k = 2.0)

注2) 棄却判定:
 ○ = 棄却基準に該当せず。
 △ = 上向DOL基準に該当するが、REC基準には該当しない。
 ▲ = 上向DOL基準に該当し、かつREC基準にも該当する。
 ▼ = 下向DOL基準に該当するが、REC基準には該当しない。
 ▼ = 下向DOL基準に該当し、かつREC基準にも該当する。
 × = DOL基準には該当しないが、REC基準に該当する。

探すと、全地点を通じてはVの極値I型分布である。ただし、REC基準不適合および下向DOL基準不適合が

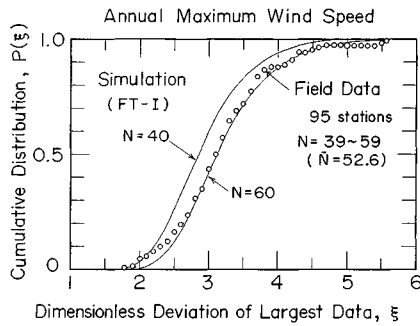


Fig. 8 Cumulative distribution of ξ -value for the annual maximum wind speed in Japan excluding Hokkaido and Tohoku districts.

ともに15地点で12%もあり、全国的に極値Ⅰ型分布があてはまるとは考えにくい。特に、北海道・東北地方は標本中の最大値の伸びが低い傾向があり、ここではむしろ $k=2.0$ のワイブル分布の方が棄却される割合が低い。

北海道・東北地方を除く95地点について ξ 値の累積分布を作成したのが Fig. 8 である。これらの地点の観測年数は39~59年であるので、極値Ⅰ型分布の数値シミュレーション結果のうち、 $N=40$ および60のケースを併せて示してある。実測資料は極値Ⅰ型分布の累積分布にはば一致しており、風速資料に対する極値Ⅰ型分布の適合性を裏付ける。ただし、相関係数に基づく REC 基準では北海道・東北地区を除いても約12%の地点が棄却されるので、極値Ⅰ型分布を完全に推奨できるわけではない。中部から近畿地方にかけては $k=10$ の極値Ⅱ型分布の方が棄却率が低い。年最大風速資料については全国一律に1つの極値分布を適用するのではなく、地方ごとに最適な分布関数を選択するのがよいと思われる。

(3) 年最大日雨量資料に対する極値分布

わが国における降水量の観測は1876年(明治9)から始められ、全国の年最大日雨量のデータが気象庁図書室に収められている。そこで、統計年数ができるだけ長く取れるように、1912年(明治45)以前からの記録がある58地点の資料を解析した。風速資料と同様に、長期変動の補正等は行っていない。各種の極値分布をあてはめた結果のうち、極値Ⅱ型分布のうちの $k=5$ と10、極値Ⅰ型分布、ワイブル分布(4種類)、および対数正規分布に対する棄却検定結果を Table 5 に一覧表として示した。

日雨量の資料では、彦根が飛び抜けて大きな ξ 値 8.23 を示す。これは1896年9月7日の596.9 mm という記録によるもので、ここでの第2位は1959年9月26日の195.8 mm である。非常に局地的な豪雨であったと思われるが、詳しいことはわからない。この彦根の資料

Table 5 Calibration of various distributions for annual maximum daily rainfall by rejection criteria.

地点名	データ数 N	平均値 \bar{x}	標準偏差 s	最大値 偏差 ξ	棄却検定			
					III	IV	V	X
網走	98	53.4	17.7	3.00(2.34)	▽○○	▽▽○○	○	
根室	104	75.0	26.4	2.75(2.14)	▽▽○	▽▽○○	○	
釧路	98	80.6	25.6	3.97(2.87)	○○○	×○○	○	
札幌	110	70.6	29.7	4.59(3.00)	○○○	○○○	○	
寿都	101	65.5	30.0	4.70(3.37)	○○○	○○×	△	
函館	109	78.5	28.9	4.83(3.22)	○○○	○○△	△	
青森	105	66.4	26.5	4.59(3.12)	○○○	○○○	△	×
秋田	104	83.1	24.2	3.87(2.89)	○○○	×○○○	○	
山形	97	72.7	33.3	4.35(3.01)	○○○	○○×	×	
宮古	103	120.1	46.4	3.56(2.57)	○○○	×○○	×	
石巻	100	82.2	27.4	2.82(2.19)	▽○○	▽▽○○	○	
福島	97	81.3	31.8	2.77(2.20)	▽▽○	▽▽○○	○	
小名浜	76	108.4	42.1	2.82(2.27)	▽○○	▽○○×	○	
宇都宮	96	103.9	35.4	3.27(2.35)	○○○	▽○○○	○	
水戸	90	103.0	43.5	3.99(2.84)	○○○	○○○	○	
焼子	99	113.3	46.9	4.23(2.91)	○○○	○○○	○	
前橋	90	96.0	45.8	5.71(3.60)	○○○	○○△	△	
熊谷	90	110.4	56.4	3.40(2.43)	○○○	○○○	○	
東京	111	114.5	49.6	5.60(3.38)	○○○	○○○	△	
横浜	89	124.9	49.3	3.30(2.34)	○○○	▽○○○	○	
八丈島	80	194.8	78.7	3.10(2.26)	○○○	▽○○○	○	
浜松	105	140.1	52.1	3.91(2.86)	○○○	▽○○	×	
飯田	89	104.9	41.2	5.36(3.55)	○○○	○○△	△	
甲府	92	101.2	46.3	3.10(2.26)	○○○	▽○○	○	
本巣	89	70.9	25.9	3.28(2.53)	○○○	▽○○	×	
長野	98	58.3	19.2	2.83(2.30)	▽○○	▽▽○○	○	
新潟	105	74.1	24.1	3.77(2.82)	○○○	○○○	○	
伏木	101	87.1	30.9	3.40(2.66)	○○○	▽○○	×	
金沢	104	95.5	32.2	4.32(3.08)	○○○	○○○	×	
高山	87	97.3	39.2	4.31(3.11)	○○○	○○○	×	
岐阜	104	118.7	41.2	3.44(2.63)	○○○	▽○○	○	
早稲	96	109.4	37.0	3.54(2.55)	○○○	○○○	○	
名古屋	97	133.3	50.5	3.07(2.33)	▽○○	▽○○×	○	
福井	90	93.3	30.6	3.53(2.68)	○○○	○○×	○	
敦賀	85	104.7	36.6	2.91(2.21)	○○○	▽×○○	○	
彦根	93	100.5	60.4	8.23(5.05)	○○○	△△△△	△	
京都	106	105.7	41.2	4.44(3.18)	○○×	○○×	×	
大阪	104	91.2	32.0	4.98(3.30)	○○○	○○△	△	
神戸	90	103.4	50.6	4.27(2.94)	○○○	○○○	×	
和歌山	107	110.2	44.9	4.14(2.82)	○○○	○○○	×	
潮岬	83	175.1	68.3	3.60(2.60)	○○○	○○○	×	
高知	104	188.5	65.9	5.10(3.45)	○○○	○○△	△	
徳島	95	154.8	72.6	4.36(2.93)	○○×	○○×	○	
松山	97	92.8	30.8	3.97(2.84)	○○○	×○○	○	
多度津	94	84.8	34.8	3.26(2.41)	○○○	▽○○	○	
岡山	96	79.0	23.7	4.14(2.91)	○○○	×○○	○	
広島	108	106.9	40.2	5.79(3.69)	○○○	○○△	△	
境	104	108.2	46.8	3.89(2.69)	○○○	○○○	○	
浜田	94	106.3	52.0	4.33(3.07)	○○×	○○×	×	
下関	104	118.0	43.2	5.06(3.37)	○○○	○○△	△	
蔵原	100	171.4	57.9	3.82(2.71)	○○○	×○○○	○	
岡	97	118.8	45.2	3.34(2.44)	○○○	▽○○○	○	
佐賀	96	134.7	48.9	4.74(3.41)	○○×	○○○	△	
長崎	108	146.7	61.6	3.88(2.76)	○○○	○○○	○	
熊本	96	149.1	72.9	4.54(3.03)	○○×	○○×	×	
大分	100	145.8	68.6	4.34(2.83)	○○○	○○○	○	
宮崎	104	181.9	85.0	4.77(3.21)	○○×	○○△	×	
鹿児島	104	144.1	43.3	3.74(2.67)	○○○	×○○○	○	

注1) ξ 値の括弧内の数値は、対数正規分布に対する値である。
 注2) 分布関数： III=極値Ⅱ型 ($k=5.0$), IV=極値Ⅱ型 ($k=10.0$)
 V=極値Ⅰ型分布, VI=ワイブル ($k=0.75$)
 VII=ワイブル ($k=1.0$), VIII=ワイブル ($k=1.4$)
 IX=ワイブル ($k=2.0$), X=対数正規分布
 注3) 棄却基準：表-4に同じ。

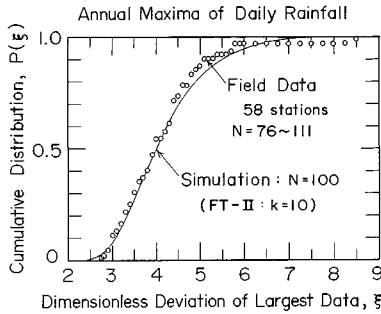


Fig. 9 Cumulative distribution of ξ -value for the annual maxima of daily rainfall in Japan.

に対して適合可能なのは $k=2.5$ の極値 II 型分布だけであり、他の分布関数はすべてあてはめが棄却される。この彦根の 596.9 mm は絶対値としても 58 地点中で最大で、第 2 位は宮崎の 587.2 mm である。しかし、宮崎では普通の年でも降雨量が多いので、最大値偏差としては $\xi=4.77$ にとどまっている。

風速資料の場合は標本中の最大値が比較的小さく、下向 DOL 基準によって棄却されるケースが多かったが、雨資料の場合は最大値の伸びが著しく、上向 DOL 基準で棄却されるケースが多い。全地点に対する各分布関数の棄却検定結果を比べると、IV の $k=10$ の極値 II 型分布が棄却率が最も低い。REC 基準での棄却が 1 地点、上向 DOL 基準による棄却が 1 地点、下向 DOL 基準で棄却されるのが 2 地点のみである。58 地点に対する割合はいずれも 5% 以下である。Fig. 9 は、年最大日雨量の ξ 値の累積分布を $k=10$ の極値 II 型分布に対する数値シミュレーション結果と比較したもので、実測資料の $N=76\sim 111$ に対して数値計算は $N=100$ の場合を示している。非超過確率が 70%~97% の範囲で実測の ξ 値がやや低めであるけれども、全体としてはよく一致している。

今まで年最大日雨量についてはその母集団が不明であったけれども、今回の結果をみるかぎり $k=10$ の極値 II 型分布がその候補としてきわめて有力と考えられる。さらに多数の資料について検討する必要がある。

(4) 高波の部分極値資料に対する極値分布

沿岸の波浪観測資料は、風や雨のデータと異なり観測期間がまだ短く、毎年最大値を使ったのでは標本として小さなものとなり、統計的に信頼できる結果が得られない。このため、高波の極値統計解析ではある大きさ以上の波高の極大値を対象とする部分極値資料が使われる。ここでは、先に合田・小舟¹³⁾が解析した日本沿岸 29 地点の資料に下向 DOL 基準の棄却検定を追加した結果を Table 6 に示す。分布関数としては、波浪の極値統計に対して著者¹⁰⁾が提案している極値 I 型分布とワイブル分

Table 6 Calibration of various distributions for coastal storm wave heights by rejection criteria.

地点名	有効年数 K	データ数 N	採択率 %	ξ 値	棄却判定					地点名	有効年数 K	データ数 N	採択率 %	ξ 値	棄却判定					
					V	VI	VII	III	IV						V	VI	VII	III	IV	V
紋別	7.4	75	0.42	3.62	○	○	○	○	○	苫小牧	13.9	213	0.47	4.78	○	○	○	○	○	×
留萌	13.1	294	0.44	3.70	▽	▽	▽	▽	▽	むつ小川原	10.4	250	0.50	3.07	▽	▽	▽	▽	▽	▽
瀬川	4.3	153	0.73	3.07	▽	▽	▽	▽	▽	八戸	10.6	308	0.61	4.14	○	▽	○	○	○	×
深浦	4.2	156	0.78	3.58	○	○	○	○	○	宮古	3.0	138	0.77	4.05	×	○	○	○	○	×
酒田	10.1	361	0.72	3.19	▽	▽	▽	▽	▽	釜石	5.3	81	0.43	2.91	○	○	○	○	○	○
崎崎	5.7	173	0.62	2.69	▽	▽	▽	▽	▽	仙台	3.3	176	0.57	3.94	○	○	○	○	○	○
崎島	5.9	196	0.68	3.13	▽	▽	▽	▽	▽	小名浜	3.7	102	0.49	3.48	○	○	○	○	○	○
金沢	12.0	263	0.44	3.02	▽	▽	▽	▽	▽	常陸那珂	4.8	121	0.43	3.22	○	○	○	○	○	○
福井	3.6	94	0.52	3.18	○	○	○	○	○	鹿島	9.4	407	0.73	4.71	○	○	○	○	○	×

注1) 分布関数: V=極値 I 型分布, VI=ワイブル (k=0.75), VII=ワイブル (k=1.0), III=ワイブル (k=1.4), IV=ワイブル (k=2.0).
注2) 棄却判定: 表-4 に同じ。

布 4 種類の組合せを対象とした。なお、波高計の欠測期間は統計から除外して有効年数を算出している。

高波の極値資料の場合には、海域ごとに異なる特性が現われている。紋別から福井に至るオホーツク沿岸と日本海北部から中部では ξ 値が低く、分布の裾が広い $k=0.75$ や 1.0 のワイブル分布はほとんどの地点で棄却され、 $k=2.0$ のワイブル分布が有力である。鳥取から那覇にかけての日本海西部から東支那海では $k=1.4$ のワイブル分布がどの地点でも棄却されない。この分布と極値 I 型分布はその性質がかなり似ているのであるが、後者は名瀬の資料に対してうまく適合しない。

一方、苫小牧から鹿島までの太平洋沿岸では、 $k=1.0$ のワイブル分布の棄却される割合が最も低い。しかし、大島の波浮から九州にかけての太平洋沿岸では非常に大きな波高が観測されることが多くて ξ 値が大きいため、 $k=0.75$ のワイブル分布が棄却率ゼロとなっている。

このような海域によって高波の極値分布特性が異なるのは、波浪の発生原因 (低気圧と台風) の違いや吹送距離の差異によると推測される。ただし、今回の解析は統計期間が短い地点が多いため、Table 6 の結果だけで判断するのは早計であり、今後さらに観測資料が蓄積され、統計資料としての信頼性がもう少し高くなった時点で再度検討する必要がある。特に、温帯低気圧による高波と台風による高波が混在している海域については、データの等質性の吟味が必要である。太平洋沿岸の諸地点については、さらに極値 II 型分布の適合性について検討してみる価値があるものと思われる。

5. 結 論

本論文で得られた主な結論は次のとおりである。
(1) 極値統計資料の標本中の最大値についてその無

次元偏差量が定義され、その累積分布特性が数値シミュレーション資料に基づいて明らかにされた。

(2) 上記の最大値偏差の5%および95%非超過確率値が極値Ⅰ型分布、極値Ⅱ型分布、ワイブル分布、および対数正規分布に対して求められ、標本の大きさおよびデータ採択率の関数として実験式の形で取りまとめられた。

(3) 標本中の異常値の棄却基準が、上記の最大値偏差の非超過確率値に基づいてDOL基準の名称で提案された。この基準は、異常値の棄却検定と同時に、標本にあてはめる極値分布関数の棄却検定にも使用可能である。

(4) 日本各地の128地点の年最大風速資料に対して分布関数の棄却検定を行ったところ、極値Ⅰ型分布の棄却率が最小であったが、全地点に対してこの分布関数を適用することには統計学的に疑義があり、地方ごとに最適な分布関数を選んで使用するのが妥当である。

(5) 日本各地の58地点の年最大日雨量に対して分布関数の棄却検定を行った結果、 $k=10$ の極値Ⅱ型分布が母集団分布として最有力であることが示された。

(6) 日本沿岸29地点の高波の部分極値資料については、海域ごとに高波の極値分布特性が異なることが明示された。

なお本研究は、その初期段階において小野澤昌己君および李孝秀君の横浜国立大学工学部における卒業研究として行われたものであり、両君の協力が心から感謝する次第である。

参考文献

- 1) 角屋 睦：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.23~27，1962.
- 2) 中西祐啓：治水計画における外力の確率評価の可能性について，土木計画学研究・論文集，No.4，pp.165~172，1986.
- 3) 宝 馨・高棹琢馬：水文頻度解析における確率分布モデルの評価規準，土木学会論文集，No.393/Ⅱ-9，pp.151~160，1988；b) 同論文への討議・回答，土木学会論文集，No.405/Ⅱ-11，pp.265~272，1989.
- 4) 江藤剛治ほか：大雨の頻度，土木学会論文集，No.369/Ⅱ-5，pp.165~174，1986.
- 5) Cunnanne, C. : Review of statistical models for flood frequency estimation, Hydrol. Freq. Modelling, D. Reidel Pub. Co., pp.49~95, 1987.
- 6) Kumar, A. and Chander, S. : Statistical flood frequency analysis— an overview, Hydrol. Freq. Modelling, D. Reidel Pub. Co., pp.19~35, 1987.
- 7) 小野澤昌己・合田良実：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，土木学会第44回年次学術講演会，第Ⅰ部門，1989.
- 8) Jenkinson, A.F. : The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Quart. J. Roy. Meteor. Soc., Vo.81, pp.158~171, 1955.
- 9) 角屋 睦：極値分布とその一解法，農業土木研究，第23巻，第6号，pp.28~35，1956.
- 10) 合田良実・極値統計におけるプロット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，港湾技術研究所報告，第27巻，第3号，pp.31~92，1988.
- 11) Gringorten, I.I. : A plotting rule for extreme probability paper, J. Geophys. Res., Vol.68, No.3, pp.813~814, 1963.
- 12) Blom, G. : Statistical Estimates and Transformed Beta-Variables, John Wiley & Sons, New York, Chapt. 12, 1958.
- 13) 合田良実・波浪の統計的性質に関する二、三の数値的検討，港湾技術研究所報告，第24巻，第4号，pp.65~102，1985.
- 14) 三上 操：応用推計学，内田老鶴圃，pp.270~273，1959.
- 15) 合田良実・小舟浩治：波浪の極値統計における分布関数のあてはめ基準，海岸工学論文集，第36巻，pp.135~139，1989.
- 16) 合田良実・李 孝秀：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，土木学会第44回年次学術講演会，第Ⅱ部門，1989.
- 17) 気象庁：日本各地の年最大風速(1929~1966)，順位と出現確率，気象庁観測技術資料，第34号，75 p., 1971.
- 18) 藤野陽三・伊藤 学・酒井利夫：年最大風速記録による設計基本風速の算定に関する研究，土木学会論文報告集，第305号，pp.23~34，1981.
- 19) Simiu, E. and Filliben, J.J. : Probability distributions of extreme wind speeds, J. Struct. Div., Proc. ASCE, Vo.102, No. ST 3, pp.1861~1877, 1976.
- 20) 江藤剛治・室田 明：一雨降雨の1確率模型，土木学会論文集，No.345/Ⅱ-1，pp.101~109，1984.
- 21) Petruaskas, C. and Aagaard, P.M. : Extrapolation of historical storm data for estimating design wave heights, Prepr. 2nd OTC, No.1190, 1970.
- 22) 合田良実・小野澤昌己：極値Ⅱ型分布の特性と推定値の信頼区間，土木学会論文集(ノート)，No.417/Ⅱ-13，1990.

(1989.8.21・受付)