

極値統計解析における層別化手法の適用性の検討

山口正隆*・畑田佳男**・野中浩一***・大福学****

1. 緒 言

極値統計解析では、標本の等質性を確保する観点から、層別化された極値資料を用いることが望ましいが、実際の解析事例はあまり多くない。この方法は層別資料に対して別々に極値統計解析を行ったのち、それぞれの結果の結合に基づいて合成確率統計量と合成分散を推定するというものであり、その一般的手法（層別化解析法）が合田（1990）により提案されている。このうち、合成分散推定法には合田（1990）の経験式のほか、最近提案された泉宮（2000）の理論式があるが、とくに後者の適用性は広範囲には調べられていない。

そこで本研究では、候補分布の母数推定に線形最小 2 乗法を用いる極値統計解析モデル（合田，1990；Yamaguchi ら，1997）を適用した場合に、合田（1990）および泉宮（2000）による合成分散推定式の精度を広範な入力条件のもとでのモンテカルロシミュレーション結果に基づいて検討するとともに、わが国沿岸における観測資料より得た風速や波高の年最大値資料に対する層別化解析法の適用性を調査する。

2. 合成確率統計量および合成分散の推定法

合成確率分布 $F(H)$ は、層別年最大値資料あるいは層別極大値資料に対する確率分布 $F_j(H)$ の積として、それぞれ次式で与えられる（合田，1990）。

$$F(H) = \prod_{j=1}^n F_j(H) \quad \cdots \cdots \cdots (1)$$

$$F(H) = \exp \left[- \sum_{j=1}^n \lambda_j \{ 1 - F_j(H) \} \right] \quad \cdots \cdots \cdots (2)$$

ここに、 n は極値資料の層別数、 λ_j は層別極大値資料の年平均発生数、である。

一方、層別確率統計量の分散に対する合成式は、合田（1990）の経験式では、

$$\sigma_c^2(H_R) = \sum_{j=1}^n N_j \{ 1 - F_j(H_R) \} \sigma_j^2(H_R) / \sum_{j=1}^n N_j \{ 1 - F_j(H_R) \} \quad \cdots \cdots \cdots (3)$$

泉宮（2000）の理論式では、

$$\sigma_j^2(H_R) = \sum_{j=1}^n \lambda_j^2 f_j^2(H_R) \sigma_j^2(H_R) / \left\{ \sum_{j=1}^n \lambda_j f_j(H_R) \right\}^2 \quad \cdots \cdots \cdots (4)$$

により表される。ここに、 N_j ：層別データ数、 H_R ：式（1）あるいは式（2）の合成確率分布より得られる合成確率統計量、 $\sigma_j^2(H_R)$ ：合成確率統計量 H_R に対する層別分散、 $f_j(H_R)$ ：層別極値資料の確率密度関数、である。なお、以上の諸式は確率分布や母数推定法の種類に依存しない。

3. シミュレーションに基づく精度の検討

（1）シミュレーションの方法

モンテカルロシミュレーションは Gumbel 分布、Weibull 分布、FT-II 型分布を対象分布として、確率分布の形状母数 k 、尺度母数 A 、位置母数 B 、層別数 n の組合せを適宜変えて行う。入力条件は泉宮（2000）の解説に与えられたケース、各分布の母数を任意に設定するケース、およびわが国沿岸の灯台における観測風速資料に対する極値解析から推定した母数を与えるケースとする。年最大値（AM）資料および極大値（POT）資料は各確率分布の inverse form（合田，1990）の中の非超過確率 $F(H)$ に一様乱数を入力することによって作成されるが、極大値資料の場合には、 $F(H)$ のかわりに $1 + (1/\lambda) \log F(H)$ を用いる。

シミュレーションの手順は山口ら（1999）の論文に記載されている。極値統計解析は、候補分布の母数推定に線形最小 2 乗法を用いる合田（1990）のモデルを変更した Yamaguchi ら（1997）のモデル（拡張型モデル）による。シミュレーションは既知母分布を前提として行うので、候補分布の形状母数を入力分布のそれに固定することとし、最大相関係数基準に基づく最適分布選択法を適用しない。また、jackknife 法による確率統計量の分散推定も行わない。

山口ら（1999）のシミュレーションでは、同じ層別確

* 正 会 員 工 博 愛媛大学教授 工学部環境建設工学科
** 正 会 員 工 修 愛媛大学助手 工学部環境建設工学科
*** 学生会員 工 修 愛媛大学大学院理工学研究科
**** 愛媛大学教官 工学部環境建設工学科

率分布のとき層別分散と相等しい合成分散を得るために、同一の一樣乱数の組を入力条件としているが、この場合層別資料間の独立性が確保されないことから、今回の検討では異なる一樣乱数の組を用いる。この結果、同じ層別確率分布で層別数 $n=2$ および 4 の場合、合田式による合成分散推定値は層別分散にほぼ等しく、泉宮式による合成分散推定値はそれぞれ層別分散のほぼ 1/2 および 1/4 になることがシミュレーション結果より確認された。

(2) 結果の考察

再現期間 100 年に対して合成確率統計量の真値 H_{UR} で無次元化した $\text{bias}\Delta\tilde{H}$ および分散平方根 $\widetilde{Var}^{1/2}$ と、 $\widetilde{Var}^{1/2}$ で無次元化した合成分散平方根推定値 (合成標準偏差) $\tilde{\sigma}_c$, $\tilde{\sigma}_i$ の%値とデータ年数 K の関係で結果を図示する。まず泉宮 (2000) および本研究におけるシミュレーション結果の整合性を検討するため、泉宮 (2000) が与えた条件、すなわち層別数 $n=2$ とし、 $A=0.7$, $B=3.0$ の Gumbel 分布 ($\lambda_1=4$) および $k=1.0$, $A=0.6$, $B=2.3$ の Weibull 分布 ($\lambda_2=14$) のもとでの事例を示したのが図-1 である。この場合、泉宮 (2000) の解説においてデータ採択率 ν が不明であるので、 ν を 1/8~1 の間で変化させた。 $\nu=1$ のケースでは、データ年数 $K=10$ 年における $\text{bias}\Delta H$ は 0.082 と泉宮 (2000) の図-12 における値に近い値をとっており、データ年数 (データ数) の増加とともに 0 に漸近する。無次元合成分散平方根推定値 $\tilde{\sigma}$ は合田式で 127~128% (図省略)、泉宮式で 99~100% をとり、いずれも泉宮 (2000) の図-13 における値とほぼ一致する。また、データ採択率 (データ数) の増加とともに、

データ年数 $K=10$ 年の場合 $\text{bias}\Delta\tilde{H}$ は 3.8% から 1.1% に、 $\widetilde{Var}^{1/2}$ は 17.0% から 7.3% に減少し、他のデータ年数に対しても同様の挙動を示すことから、 bias や $\widetilde{Var}^{1/2}$ に及ぼすデータ数 N の影響がかなり大きい。合田式による $\tilde{\sigma}_c$ は 127~130% の範囲 (図省略) にある。一方、泉宮式による $\tilde{\sigma}_i$ は全ケースについて 99~103% の範囲にあり、泉宮式の精度はこのケースに対して非常に高い。

ついで、図-2 は層別数 $n=2$ とした場合に $A=1.3$, $B=4.5$ の Gumbel 分布と $k=1.4$, $A=5.0$, $B=1.0$ の Weibull 分布、および $A=1.3$, $B=4.5$ の Gumbel 分布と $k=5.0$, $A=1.0$, $B=5.0$ の FT-II 型分布に従う極大値資料 ($\lambda=8$, $\nu=1/8$) に対する解析結果を表す。 $\text{bias}\Delta\tilde{H}$ は全般的に小さいが、層別確率分布として正側に長く裾を引く FT-II 型分布を含むケースでは、データ年数 $K=10$ 年の場合 $\text{bias}\Delta\tilde{H}$ は 2.5% となり相対的にやや大きい値をとる。 $\widetilde{Var}^{1/2}$ は Weibull 分布を含むケースでは $K=10\sim100$ 年に対して 17.5~5.7% であるが、FT-II 型分布を含むケースでは 32.1~12.7% とかなり大きい。無次元合成分散平方根推定値 $\tilde{\sigma}$ は Weibull 分布を含むケースでは $K=10\sim100$ 年に対して合田式で 105~101%、泉宮式で 104~100% をとり、いずれの式による値も $\widetilde{Var}^{1/2}$ とほぼ符合するが、FT-II 型分布を含むケースでは合田式で 112~105%、泉宮式で 110~103% となって、いずれの式も $\widetilde{Var}^{1/2}$ をやや過大評価する。

図-3 は層別数 $n=4$ とした場合に、($k=0.75$, $A=2.0$, $B=0.4$), (1.0 , 3.0 , 0.6), (1.4 , 4.0 , 0.8), (2.0 , 5.0 , 1.0) の Weibull 分布、および ($k=5.0$, $A=1.0$, $B=5.0$), (5.0 , 0.8 , 4.0), (5.0 , 0.6 , 3.0), (5.0 , 0.4 ,

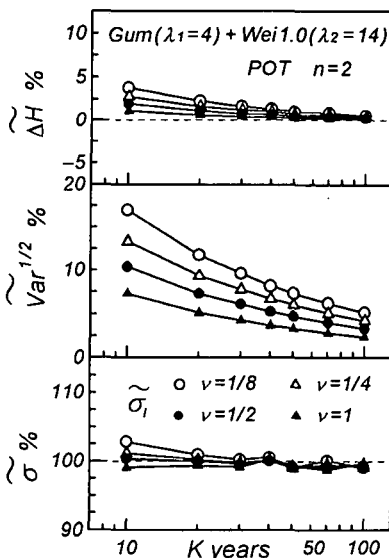


図-1 $\Delta\tilde{H}$, $\widetilde{Var}^{1/2}$, $\tilde{\sigma}_i$ に及ぼす ν の影響 (泉宮ケース)

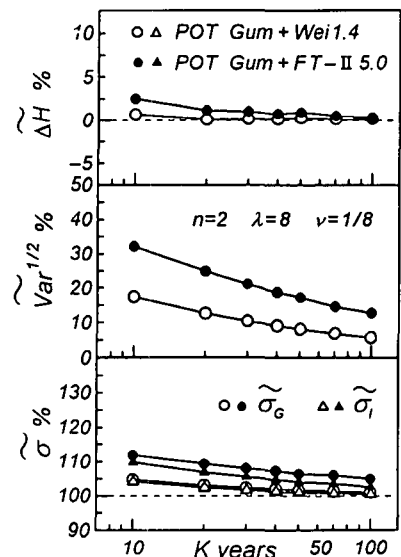


図-2 $\Delta\tilde{H}$, $\widetilde{Var}^{1/2}$, $\tilde{\sigma}$ と K の関係 ($n=2$, POT 資料)

2.0) の FT-II 型分布に従う極大値資料 ($\lambda=8, \nu=1/8$) に対する解析結果を示す。これらのケースでは正側に長く裾を引く $k=0.75$ の Weibull 分布や、FT-II 型分布を含み、しかも層別数が多いことから、 $\text{bias } \Delta \tilde{H}$ は $K=10 \sim 100$ 年に対して前者で $13.6 \sim 1.9\%$ 、後方で $8.1 \sim 1.5\%$ 、 $\tilde{Var}^{1/2}$ も $33.9 \sim 10.7\%$ 、 $33.4 \sim 11.4\%$ とかなり大きい。 $\tilde{Var}^{1/2}$ の推定値 $\tilde{\sigma}$ は合田式では Weibull 分布ケースで $162 \sim 153\%$ (図省略)、FT-II 型分布ケースで $128 \sim 130\%$ と過大な値をとるが、泉宮式では前者で $113 \sim 98\%$ 、後方で $97 \sim 98\%$ となり、Weibull 分布ケースにおける $K=10$ 年の場合 (113%) を除いて、泉宮式の誤差は $-3 \sim 5\%$ とその精度はかなり高い。

さらに以下では、観測資料の解析より得た層別確率分布を用いた検討結果を示す。図-4 の上図は層別数 $n=2$ とした場合に、($k=1.3, A=15.3, B=19.6$)、($3.0, 15.3, 18.1$) の Weibull 分布 (第 1 ケース)、および $k=1.7, A=10.6, B=13.5$ の Weibull 分布と $A=2.29, B=20.2$ の Gumbel 分布 (第 2 ケース) に従う年最大値資料 ($\nu=1$) に対する $\tilde{\sigma}$ を、下図は ($k=1.1, A=5.23, B=21.3$)、($2.0, 6.82, 22.9$) の Weibull 分布 (第 1 ケース)、および ($k=1.7, A=8.83, B=13.7$)、($1.9, 10.1, 15.1$) の Weibull 分布 (第 2 ケース) に従う年最大値資料 ($\nu=1$) に対する $\tilde{\sigma}$ を表す。上図の第 1 ケースは 100 年確率統計量がそれぞれ $69.1, 43.6$ と両者に大きな差を生じる場合である。合田式および泉宮式はデータ年数 (データ数) によらず $\tilde{Var}^{1/2}$ とほぼ一致する値を与え、両式の精度は高い。上図の第 2 ケースでは 100 年確率統計量が $39.5, 30.7$ と相対的に差がやや小さい場合であり、いずれの式

も $\tilde{Var}^{1/2}$ をやや過大評価するが、泉宮式の誤差はデータ年数の増加とともに急速に減少する。

図-4 の下図の第 1 ケースは 100 年確率統計量が $42.3, 37.5$ とかなり接近している場合である。合田式はデータ年数 $K=10 \sim 100$ 年に対し $28 \sim 13\%$ の過大評価、泉宮式は $22 \sim 5\%$ の過大評価となっている。泉宮式の精度が相対的に高いが、データ年数が少ない場合の絶対誤差は小さいとは云えない。下図の第 2 ケースは 100 年確率統計量が $34.4, 37.7$ とほぼ拮抗する場合である。合田式はデータ年数によらず約 33% の過大評価となるが、泉宮式はほぼ $\tilde{Var}^{1/2}$ に一致する値を与える。前述のように、層別確率統計量が同程度の場合、合田式は $\tilde{Var}^{1/2}$ をほぼ n 倍過大評価するのに対し、泉宮式は $\tilde{Var}^{1/2}$ をほぼ適切に評価する。

図-5 は層別数 $n=4$ とした場合に、 $A=3.14, B=18.3$ の Gumbel 分布と ($k=2.4, A=10.5, B=11.0$)、($2.2, 20.0, 8.36$)、($2.8, 10.1, 8.79$) の Weibull 分布 (第 1 ケース)、および ($k=0.75, A=2.38, B=21.2$)、($0.90, 2.16, 16.3$)、($2.2, 13.3, 9.90$) の Weibull 分布と $A=2.68, B=21.8$ の Gumbel 分布 (第 2 ケース)

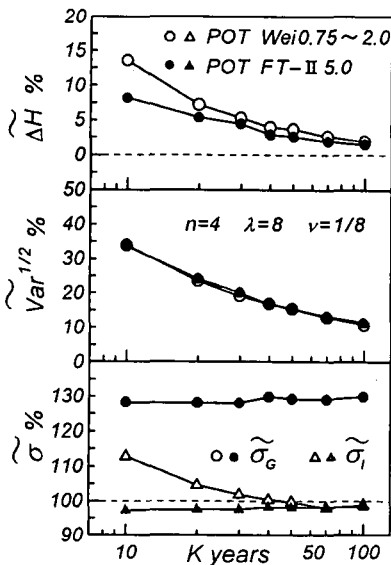


図-3 $\Delta \tilde{H}$, $\tilde{Var}^{1/2}$, $\tilde{\sigma}$ と K の関係 ($n=4$, POT 資料)

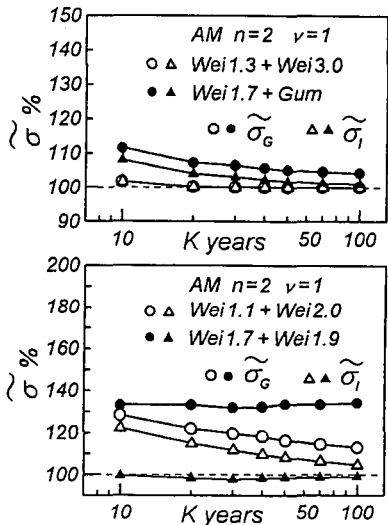


図-4 $\tilde{\sigma}$ と K の関係 ($n=2$, AM 資料)

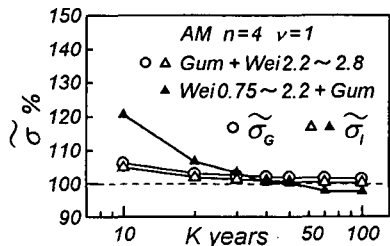


図-5 $\tilde{\sigma}$ と K の関係 ($n=4$, AM 資料)

に従う年最大値資料 ($\nu=1$) に対する解析結果である。第1ケースは100年確率統計量が32.7, 30.9, 48.5, 26.3であり、第1位値と第2位値以下の差が大きい場合に相当し、合田式、泉宮式の精度は高い。第2ケースは100年確率統計量が39.4, 28.0, 36.5, 34.2であり、第1位値と第2位・第3位値が接近した値をとることから、合田式は $K=10\sim 100$ 年にに対し59~47%の過大評価(図省略)となっている。層別分布に小さい形状母数のWeibull分布が含まれるため、泉宮式も $K=10$ 年にに対し21%の過大評価となるが、 $K\geq 20$ 年では誤差は7~2%であり、高い精度を与える。

全検討ケースを通じてみても、合田式や泉宮式の一般の適用条件を明確にすることは困難であるが、層別確率統計量の差が大きい場合、泉宮式や合田式は高い精度を与える。しかし、その差が小さい場合、泉宮式はかなり高い精度をもつものに対して、合田式は合成分散を過大評価する。泉宮式は合成分散推定に対して全般的に高い精度を与えるが、その精度は合成確率統計量推定値のbiasと連動しており、biasが大きいほど、すなわち層別分布に小さい形状母数の確率分布と異なる種類の確率分布が混在するときほど、過大あるいは過小な値を与える傾向にあり、層別数が増加すれば、この傾向が助長される。

4. 層別化解析法の現地観測資料への適用

ここでは、2種類の現地資料(山口ら, 1998)、すなわちわが国沿岸7箇所の海上保安庁灯台における風速観測資料およびわが国太平洋岸から東シナ海沿岸の9地点における波浪観測資料を使用する。そして風速や波高の年最大値資料および層別年最大値資料に拡張型モデルと層別化解析法を適用し、層別100年確率統計量とその標準偏差およびそれらの合成値を推定する。極値統計解析では、形状母数が大きいWeibull分布が最適分布として抽出されるケースが大部分であるので、Gumbel分布と27種類の形状母数に対するWeibull分布の計28種類を候補分布とし、FT-II型分布を加えていない。

表一は年最大風速資料に対する一括解析結果および

気象擾乱別(台風とそれ以外の気象擾乱)年最大風速資料(層別数 $n=2$)に対する層別解析結果の一覧である。表中の記号は K が観測年数、 k が最適分布として選択されたWeibull分布の形状母数、 G がGumbel分布、 ρ が相関係数、 U_{100} が100年確率風速、 U_G が100年確率風速の標準偏差、 U_G 、 U_I が合田式、泉宮式による合成標準偏差、太字は合成標準偏差と一括解析に基づく標準偏差のうちで小さい方のケースを表す。この表によると、一方の層別資料に対する最適分布の適合度は向上するが、それ以外の層別資料に対する適合度は必ずしも高くないケースが多い。一括解析に基づく確率風速と合成確率風速の差は一括解析に基づく確率風速に対して-2.7~3.7%であり、一括解析に基づく確率風速がより大きい値をとる地点とより小さい値をとる地点が混在する。また、合田式および泉宮式による合成標準偏差は、気象擾乱別確率風速に差がある太平洋岸南西部の地点ではほぼ符合するが、気象擾乱別確率風速が拮抗する太平洋岸北東部や日本海沿岸では、泉宮式による値が合田式による値より小さい。しかし、観測年数が33~34年と短い年最大風速資料の層別化は資料の統計的変動を増加させるので、合成標準偏差は必ずしも一括解析に基づく標準偏差より小さい値をとらない。

表二は年最大風速資料および各種層別年最大風速資料に対する解析結果をまとめたものである。これによると、層別数とともに、合成確率風速および合成標準偏差は増加する傾向にあることから、短い期間(データ数)の資料を解析する場合、層別数を多くしても層別化解析法の有用性は見出されない。

表三は年最大波高資料および気象擾乱別年最大波高資料に対する解析結果の一覧である。風速資料の場合より観測期間が一層短い波高資料の場合にも、層別化によって適合度が向上する地点としない地点および標準偏差が減少する地点としない地点が混在することや、台風の影響が卓越する太平洋岸南西部では泉宮式と合田式による合成標準偏差はほぼ符合するが、台風とそれ以外の気象擾乱の影響が拮抗する太平洋岸北東部では泉宮式が

表一 年最大風速資料に対する極値解析結果の一覧

地点名	年最大風速					合成風速			台風年最大風速				非台風年最大風速			
	K	k	ρ	U ₁₀₀	U _σ	U ₁₀₀	U _G	U _I	k	ρ	U ₁₀₀	U _σ	k	ρ	U ₁₀₀	U _σ
襟裳岬	34	1.9	.994	50.0	1.8	50.3	2.0	1.8	2.4	.986	44.0	2.2	1.8	.995	50.2	1.9
恵山岬	34	1.3	.995	33.2	2.4	32.3	3.3	2.0	G	.983	30.1	3.6	1.6	.991	30.7	1.8
尻屋崎	34	1.7	.997	41.3	1.8	42.0	3.5	2.6	1.3	.994	41.3	3.6	1.8	.995	38.9	1.6
大王崎	34	1.7	.996	38.2	2.1	39.6	2.5	2.4	1.7	.993	39.5	2.4	G	.995	30.7	1.8
土佐沖ノ島	33	1.5	.994	61.8	4.5	60.8	4.1	4.1	2.0	.992	60.8	4.1	1.9	.990	35.3	1.2
女島	34	1.1	.986	50.0	5.3	50.2	4.6	4.6	G	.989	50.2	4.6	2.2	.985	30.2	0.7
融倉島	34	G	.989	39.5	2.7	40.5	4.2	3.4	3.0	.986	36.2	1.8	0.9	.992	39.9	4.3

K : 年数, G : Gumbel 分布, U_{100} , U_G , U_G , U_I : m/s, U_G : 合田式, U_I : 泉宮式

表-2 年最大風速資料に対する極値解析結果に及ぼす層別数の影響

地点名	年最大風速					擾乱別			2 期間別			季節別			月別		
	K	k	ρ	U_{100}	U_{σ}	U_{100}	U_C	U_I	U_{100}	U_C	U_I	U_{100}	U_C	U_I	U_{100}	U_C	U_I
襟裳岬	34	1.9	.994	50.0	1.8	50.3	2.0	1.8	50.7	2.2	2.1	50.5	2.1	2.0	53.7	3.7	2.4
恵山岬	34	1.3	.995	33.2	2.4	32.3	3.3	2.0	31.5	2.8	1.8	32.3	3.8	2.2	33.4	4.5	2.1
尻屋崎	34	1.7	.997	41.3	1.8	42.0	3.5	2.6	42.4	3.4	3.1	43.1	4.0	2.6	44.0	4.7	3.2
大玉崎	34	1.7	.996	38.2	2.1	39.6	2.5	2.4	38.2	2.0	2.0	40.4	2.9	2.8	42.2	4.3	3.1
土佐沖ノ島	33	1.5	.994	61.8	4.5	60.8	4.1	4.1	60.5	4.1	4.1	61.3	10.4	4.7	65.2	13.3	5.0
女島	34	1.1	.986	50.0	5.3	50.2	4.6	4.6	49.3	4.6	4.6	50.1	4.6	4.6	55.9	8.2	5.3
舩倉島	34	G	.989	39.5	2.7	40.5	4.2	3.4	40.5	4.8	3.8	40.9	5.6	3.7	42.6	6.8	3.7

K : 年数, G : Gumbel 分布, U_{100} , U_{σ} , U_C , U_I : m/s, U_C : 合田式, U_I : 泉宮式

表-3 年最大波高資料に対する極値解析結果の一覧

地点名	年最大波高					合成波高			台風年最大波高				非台風年最大波高			
	K	k	ρ	H_{100}	H_{σ}	H_{100}	H_C	H_I	k	ρ	H_{100}	H_{σ}	k	ρ	H_{100}	H_{σ}
苫小牧	25	1.8	.996	6.48	0.36	6.27	0.38	0.38	4.0	.976	5.29	0.44	3.5	.988	6.27	0.38
むつ小川原	21	G	.970	9.84	1.32	9.80	1.53	1.14	1.4	.985	8.81	1.20	G	.952	9.41	1.43
江ノ島	17	2.4	.993	9.68	0.73	9.82	0.81	0.70	3.0	.993	8.76	0.69	2.6	.992	9.74	0.79
鹿島	22	2.6	.990	7.95	0.36	7.67	0.50	0.35	10	.990	7.54	0.43	2.8	.969	7.34	0.51
波浮	22	2.2	.994	9.01	0.52	9.29	0.71	0.70	2.0	.980	9.29	0.71	2.6	.982	7.71	0.38
石廊崎	19	G	.980	12.04	1.38	12.01	1.29	1.29	1.8	.989	12.01	1.29	10	.985	6.02	0.18
佐喜浜	18	2.2	.989	9.60	0.76	10.14	0.83	0.83	2.0	.990	10.14	0.83	5.0	.983	6.36	0.28
福江島	15	2.0	.996	7.23	0.44	7.50	0.62	0.60	2.4	.985	7.49	0.62	2.2	.991	6.58	0.29
名瀬	18	G	.978	7.94	0.79	7.84	0.84	0.84	5.0	.992	6.45	0.30	G	.985	7.84	0.84

K : 年数, G : Gumbel 分布, H_{100} , H_{σ} , H_C , H_I : m, H_C : 合田式, H_I : 泉宮式

合田式より小さい合成標準偏差を与えることがわかる。

に謝意を表します。

5. 結 語

本研究で得られた結果の概要はつぎのようである。

1) 泉宮による合成分散推定式は全般的に高い精度をもつが、その際再現確率統計量推定値の bias を小さくするため、データ数を多くとる必要である。

2) 限られた期間で取得された風速や波高の極値資料に対する層別解析は、標本の統計的変動を増加させるため、一括解析より高い精度の再現確率統計量推定値を与えるとは限らないことから、層別化解析法は現時点ではすべてのケースに対して一様に有用な方法とは云えない。

最後に、シミュレーションの方法や合成分散推定法に関して種々の教示を戴いた新潟大学工学部泉宮尊司教授

参 考 文 献

- 泉宮尊司(2000): 設計波の合理的な設定法, 2000 年度(第 36 回) 水工学に関する夏期研修会講義集, B コース, 土木学会海岸工学委員会・水理委員会, pp. B3-1-20.
- 合田良實(1990): 港湾構造物の耐波設計—波浪工学への序説—, 鹿島出版会, 333 p.
- 山口正隆・大福 学・畑田佳男(1998): 類別化手法に基づくわが国沿岸での風速および波高の極値の推定, 自然災害科学 J. JSNDS, 17-1, pp. 45-55.
- 山口正隆・畑田佳男・大福 学(1999): 波高の極値統計解析における層別化手法の精度の検討, 海岸工学論文集, 第 46 巻, pp. 241-245.
- Yamaguchi, M. and Y. Hatada (1997): An extremal analysis system and its application to the estimation of extremes of meteorological and oceanographic elements around the coasts of Japan, Proc. Waves '97 Conf., Vol. 2, pp. 932-946.