

最小2乗法に基づく波高の極値統計解析システムの精度の検討

愛媛大学工学部 正員 山口正隆

東亜建設工業(株)

大木泰憲

愛媛大学大学院 学生員 〇前川隆海

1. はじめに：種々の極値統計解析システムの中で、合田の「最小2乗法に基づく極値統計解析システム」は年最大値資料のみならず極大値資料にも適用できる有用な方法であるが、このシステムの精度についての詳しい検討は行われていない。そこで本研究では、広範囲に入力条件を設定した場合のモンテカルロシミュレーション結果に基づいて、候補確率分布数、データ採択率、確率分布形状および未知母分布が確率波高に及ぼす影響を詳細に検討する。

2. 研究内容：(1)極値統計解析システムの概要；このシステムは、Gumbel分布と形状母数を $k=0.75, 1.0, 1.4, 2.0$ とする4種類の3母数Weibull分布の合計5種類、あるいはGumbel分布と形状母数を $k=2.5, 3.33, 5.0, 10.0$ とする4種類のFT-II型分布の合計5種類（いずれも $k_{ind}=5$ ）あるいは合計9種類（ $k_{ind}=9$ ）の確率分布の中から相関係数基準に従って最適確率分布を選択するとともに、確率波高およびその信頼区間を推定するものである。

(2)シミュレーションの方法；①母集団の確率分布と母数およびデータ採択率 $\nu=N/N_T$ （ N ：採択資料数、 N_T ：資料総数）、年平均発生数 $\lambda=N_T/K$ （ K ：年数）を指定し、 R 年確率波高（ $R=50, 100, 200, 500, 1000$ 年）の真値を求める。②確率分布の逆関数に一樣乱数を与えて N_T 個の標本を発生させ、そのうち上位 N 個を抽出する。③最小2乗法により母数を推定し、相関係数 ρ （適合度指標）、 R 年確率波高 H_R および合田式に基づく確率波高の標準偏差 σ_R を求める。④対象年数 $K=10, 20, 30, 40, 50, 70, 100, 200, 500, 1000$ 年の10ケースの場合に、②および③の手順を M 回（ $M=5000$ ）繰り返し、相関係数 ρ の平均値、確率波高の平均値とbias、標準偏差 $\text{Var}^{1/2}$ 、平方根平均2乗誤差（RMSE）、合田式による標準偏差の平均値などの各種誤差統計量を算定する。一方、未知母分布の影響を考える場合には、100年確率波高を12m、1000年確率波高を15mと設定した場合に、3母数Weibull分布、Gumbel分布、GEV分布、3母数対数正規分布、3母数Gamma分布、3母数対数Gamma分布、3母数超Gamma分布の7種類の確率分布に基づいて、各確率変数を1000回ずつ発生させることにより、未知母分布という条件を近似し、同様の解析を行う。

(3)結果の考察；図-1は、年平均発生数が $\lambda=8$ 、母分布が形状母数 $k=1.4$ の3母数Weibull分布の場合のシミュレーションから得られた極大値資料の解析結果（ $k_{ind}=5$ ）に基づいて、相関係数 ρ 、100年確率波高の真値で無次元化した100年確率波高のbias ΔH 、標準偏差 $\text{Var}^{1/2}$ 、RMSE、および100年確率波高の $\text{Var}^{1/2}$ で無次元化した合田式による100年確率波高の標準偏差の平均値 σ を%表示したものと対象年数 K との関係を6種類（ $\nu=0.125, 0.25, 0.375, 0.5, 0.75, 1.0$ ）の ν に対して示したものである。

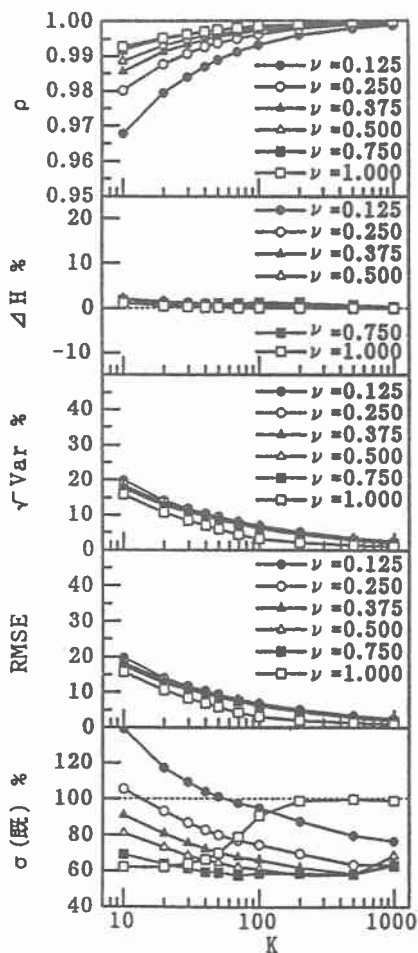


図-1

これからつぎのことがいえる。①資料への確率分布の適合度を表す相関係数 ρ は、データ採択率 ν および対象年数 K が大きくなるほど、1に近づく。②bias、 $\text{Var}^{1/2}$ およびRMSEは ν および K の増加とともに減少する。 ν および K の増加は資料数の増加に相当するので、これは、資料数が増加するにつれて、適合度および確率波高の推定精度が向上することを意味する。③合田式による標準偏差と $\text{Var}^{1/2}$ の差は、 ν が大きくなるほど広がり、最大40%程度に達することから、 $\nu=0.125$ の場合を除いて、合田式は標準偏差を過小評価する。図-2は極大値資料($\lambda=8$ 、 $\nu=0.5$)を対象として、母分布に7種類の確率分布を使用した場合(均等入力)と $k=2.0$ の3母数Weibull分布に固定した場合(固定入力)の結果を示したものであり、候補分布として、Weibull分布4種類とGumbel分布(kind=5)を使用している。相関係数は固定入力(NOR)の方が大きい、bias、 $\text{Var}^{1/2}$ 、RMSEは均等入力(RAN)の方が小さい。とくに、均等入力の場合に、biasはほとんど生じない。また、合田式の精度は対象年数が100年以下の場合によい。

3. まとめ：本研究において得られた知見は、他の条件についての検討結果を含めて、つぎのように要約される。

(1)資料数の増加とともに、適合度は向上し、bias、 $\text{Var}^{1/2}$ およびRMSEは減少するが、とくに、 $\text{Var}^{1/2}$ およびRMSEの減少がより緩やかであることから、確率波高の精度を向上させるためには、資料数の増加が必須である。

(2)候補確率分布の数が多いと、適合度は向上するが、標準偏差は増大する。しかし、biasの挙動は一貫しない。

(3)データ採択率が高いほど、適合度は向上し、標準偏差は減少するが、biasの挙動には、整合性があまり見られない。また、データ採択率の影響は年最大値資料より極大値資料においてより大きい。

(4)Weibull分布に基づく極値統計解析システムの場合、 $k=1.0\sim 1.4$ の範囲におけるbiasが小さいが、 $k < 1$ では負の有意なbiasを、 $k > 1.4$ では正のbiasを生じる。この傾向はFT-II型分布に基づくシステムの場合も同様であり、裾の長い分布である $k=2.5$ の場合には、最大10%以上の負のbiasを示す。つまり、形状母数の変化に伴うFT-II型分布のbiasの変動幅は、分布が裾を長くひくため、Weibull分布の場合より大きい。また、biasや $\text{Var}^{1/2}$ およびRMSEは資料数の多い極大値資料の場合により小さい。

(5)未知母分布の場合のbiasや $\text{Var}^{1/2}$ およびRMSEは既知母分布の場合の値より必ずしも大きいとはいえず、資料分布にあてはめる確率分布の種類、あるいは誤差統計量の種類によって相対的大小関係は変化する。

(6)合田による標準偏差評価式は数十%もの誤差をもち、標準偏差を過小評価する傾向にあるので、実用上問題が多い。合田式は年最大値資料の場合には、20%程度の誤差であるが、極大値資料の場合には、データ採択率の増大とともに、より強い過小評価傾向を示す。

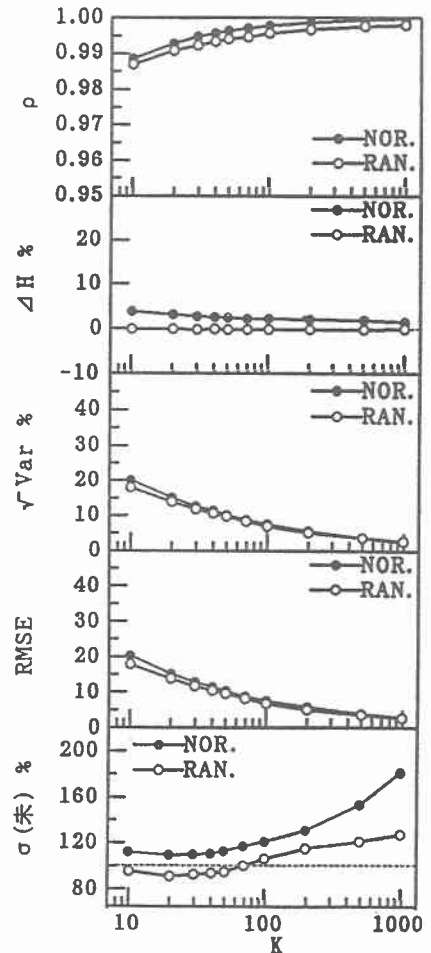


図-2