水工学シリーズ　11-A-4

水文統計解析の最前線

京都大学防災研究所　教授

寶　馨

土木学会
水工学委員会・海岸工学委員会

2011 年 8 月

水文統計解析の最前線

Recent Advancement in Statistical Analysis in Hydrology

寶　馨
Kaoru TAKARA

1．はじめに

統計学は,データに基づく解析の学問分野である．扱いたい対象や現象について，調査したり観測したりしてデータを集める．こうしてその対象や現象の特徴・性質をデータに基づいて明らかにしようとするのである．したがって，データの収集については，解析の目的に応じて適切に行わなければならない．また，データの品質や個数も問題となる．ごく限られたデータからは，対象や現象の特徴・性質を明らかにするまでには至らないであろう．得られたデータから，解析の目的に沿った有用な情報を導き出す，そしてその情報を整理して知識の体系にする．そのためには，専門分野，すなわち，気象学，水文学，水工学などの知識が必要となる．

本講義では，豪雨や洪水といった水文極値の統計解析，特に，頻度解析の手法について述べる．水文極値データに確率分布をあてはめ，その適合度や確率水文量の変動性を求め最適モデルを求める手法は，20年以上前に確立されており，1998年の夏期講習会においても解説されている（宝，1998a）．

ここでは，特に，一般的な収集されたデータの個数と頻度解析の方法の関係について説明する．さらに，データ収集の期間が長期となってきた実情を踏まえ，従来型の頻度解析の手法（確率分布を用いるパラメトリックな方法）ではなく，経験分布を用いるノンパラメトリックな方法について概説する．ここで述べる手法は，まだ実務に適用された例は少ないが，今後，観測年数は確実に長くなっていき，早晩，データの蓄積が対象とする再現期間を超えることになる．すでにそうなっているところも少なくない．そうした状況下で有用な手法であると考えている．

2．水文頻度解析の手順

水害の防止・軽減のために種々の水工構造物を計画・設計する．その際，降雨量や洪水流量といった水文量の極値データを用いて年超過確率あるいは再現確率統計量（確率水文量）を求める．これを水文頻度解析という．洪水防御計画や河川計画を立案する際の基本的作業である．この際，当該の水文量の観測年数が必ずしも長くない状況で，その年数を上回る100年に1度，200年に1度発生するような規模の豪雨や洪水（それは100年後，200年後に起こるとは限らない．来年起こるかも知れないのである．）を推定する．

この水文頻度解析の実際的な手順は，以下のようである．

① 水文極値データの収集・吟味　（独立性・均質性などをチェック）

② 候補となる確率分布を複数選ぶ　（対数正規分布，極値分布，ガンマ分布などが候補となる）

③ データに確率分布を当てはめる（確率紙や解析的方法を用いる）

④ 当てはまり具合（適合度）を評価する．　（目視，客観的適合度評価規準による）

⑤ 適合度の良い確率分布を複数個選ぶ．1つしかない場合はその確率分布を用いて確率水文量を決める．

⑥ 同程度の適合度を示す複数の確率分布に対して，ジャックナイフ法を適用し確率水文量の推定精度（ジャックナイフ推定誤差）を求め，これの小さいモデルを選ぶ．

この手順については，1980年代後半に宝・高棹（1988, 1989）によって提案され，その後，我が国の河川計画において標準的な手法として用いられている（宝，1998ab）．特に，上記手順④の客観的適合度評価規準として，高棹ら（1986）によって提案された標準最小二乗規準（SLSC）は，異なる確率分布の適合度を簡便に評価するのに有用であり，従来の水文頻度解析手法で曖昧であった適合度評価の方法を客観化するのに貢献した．また，上記手順⑥において，水文極値データセットに確率分布を当てはめて得られる再現確率統計量（確率水文量）の推定精度をジャックナイフ法やブートストラップ法によって統計的・定量的に明らかにする手法は斬新な考え方であり，また，直観的にも分かりやすい．

この方法は，上述のようにすでに我が国のほとんどの一級河川，二級河川等の実務において採用されており，その基本的な解析ソフト「水文統計ユーティリティ」が国土技術研究センターのウェブ（http://www.kasen-keikaku.jp/suimon/）から参照して利用できる．

3．標本サイズによって異なる水文頻度解析の考え方

上記の水文頻度解析の手順は，水文極値データセットに含まれるデータの数，すなわち，標本サイズ（観測年数）に応じて取り扱いを変えねばならない．標本サイズが30以上の場合，標本（データセット）への確率分布の当てはめの精度が比較的良くなる，言い換えると，確率水文量の推定値が安定してくると言える．しかしながら，標本サイズが小さい（観測年数が短い）いわゆる小標本のときには，確率分布モデルを当てはめるのに注意が必要である．小標本の場合，確率分布を解析的方法で当てはめても，再現確率統計量の推定精度が悪い．新たに付け加えられるデータによって再現確率統計量が大きく変わりうるのである．以下，本章では，小標本の場合（標本サイズが30未満）と標本サイズが30以上の場合について，異なる取り扱いをする必要があることを述べる．

3.1 小標本の場合の水文頻度解析

ここで「小標本」とは，観測年数が30年に満たない場合とし，特別な取り扱いをすることとする．なぜならば，次の二つの理由による．

（1）気象諸量の平年値は西暦で10年ごとにそれまでの30年の平均値をとっている．たとえば，西暦2005年にいるとすれば，1971年から2000年までの30年間の平均値を当該気象量の平年値としており，それを日常生活における一つの目安としている．

（2）確率分布を当てはめる場合，30～40程度の標本サイズから再現確率統計量の安定性が増してくる．

したがって，30年未満の小標本の場合は特別な取り扱いをする必要があると考えて良いという立場を取る．

これまでの国内外の長年にわたる研究によって，水文極値データ（年最大雨量，年最大河川流量）は，一般に右にひずむ分布をすることが知られている．したがって，観測年数 N が短い場合においても，極値確率紙（グンベル確率紙）あるいは対数正規確率紙にデータをプロットし，それに平分線（fair curve）を引いて所定の非超過確率（あるいはリターンピリオド）に対応する再現確率統計量（確率点，クオンタイル，quantile）を求める．このとき，N個のデータのうち小さいほうから

i番目のデータ（順序統計量）に非超過確率を与えるプロッティング・ポジション公式として，ワイブル（Weibull）公式 $i/(N+1)$ を用いると良い．なぜなら，この公式は，もともと一様分布を仮定したときに誘導されるものであり，N<30 というような小さい標本のそれぞれのデータは，等確率で発生したと見るのが，公平・客観的な立場であるからである．

もう一つの理由は，そのほかのプロッティング・ポジション公式に比べて，大きい目の再現確率統計量を与えるからである（たとえば，髙棹ら, 1986）．短い観測年数のうちには，大きな（危険な）事象が含まれないことが容易に想定されるので，こうした小標本にはワイブル公式を適用して大きい目の（すなわち，安全側の）再現確率統計量を求めておいたほうが良いといえるのである．

用いる確率紙としては，極値（グンベル）確率紙をまず推奨する．なぜならば，最大極値データを扱うのであり，それに対して理論的背景を持つ確率分布ならびに確率紙を用いるのが正当な態度である．ただし，小標本であるから，極値確率紙上でプロットする点が直線に並ぶことを期待しすぎてはいけない．この場合，平分線を引くのが難しいが，目分量で引くと客観性に欠けるので，最小二乗法で求めるのが良い．プロットされたデータ点と平分線との乖離がある程度あってもやむをえない．

小標本を使って 50 年，100 年という再現期間の確率水文量を求めようとすること自体そもそも無理があるのであるから，この場合は，あくまでもラフな推定であると割り切って，一時近似的な推定に過ぎないことを解析者も解析結果を利用しようとする者も十分理解しておく必要がある．

極値確率紙が利用できないときは，データの対数をとって，対数正規確率紙にワイブル公式でプロットすると良い．対数正規分布はグンベル分布に近い形状を表現できるからである．あくまでもラフな推定であるから，極値確率紙（グンベル分布）にこだわりすぎる必要はない．

3.2　中標本の場合の水文頻度解析

標本サイズが 30～40 以上になってくると，極値データに色々な確率分布を当てはめて頻度解析を行うことができるようになる．すなわち，上で述べた①～⑤の手順を行うのが一般的である．

この場合，候補となる確率分布は多数あり，そのどれもが似たような良い適合度を示す．適合度の評価指標として，筆者は標準最小二乗基準（SLSC）を提案した（髙棹ら, 1986；宝・髙棹, 1988）．これは日本全国の国レベル，地方レベルの水文頻度解析において広く使われている．

似たような適合度は示しても，確率水文量は少なからず異なる．また，データがさらに蓄積したときに確率水文量の値が大きく変わるようでは現場の設計・計画において大きな変更や見直しが要請されることにもなりかねない．したがって，適合度が良く，かつ，データが増えても確率水文量の変化が小さくてすむような確率分布を採用するのが望ましい．

こうした観点から，上記の⑤のステップでとどまっていた従来の水文頻度解析に，もう 1 ステップ追加することを提案した（宝・髙棹, 1988）．すなわち，適合度に加えて，確率水文量の変動性（あるいは推定精度）を評価の基準に加えることを提案したのである．計算機集約型統計学（computer intensive statistics）として近年よく使われるようになったジャックナイフ法やブートストラップ法に代表されるリサンプリング（resampling）の手法を導入することを試みた．これによって，確率水文量の推定値のバイアス補正ならびに推定精度の定量化ができる．同じような適合度を示す確率分布が複数ある場合にはジャックナイフ推定誤差が小さい分布のほうを選ぶことを推奨する．

当然のことながら，母数の個数が少ない確率分布のほうが，リサンプリングによって生成されるデータ（標本）群に対して変動が小さくなる．母数の個数が多いとそれぞれの標本に対して適合度

が良くなる分，逆に確率水文量の推定値の変動は大きくなるのである．

３母数の分布よりも２母数の分布のほうが，安定性が良い，あるいはロバスト(robust)であると言える．したがって，グンベル分布の適合度が優れていればそれを用いるのが良い．グンベル分布が適合しない場合には，３母数の分布を試み，以下の手順（上記**２**．で述べたものを再掲）で適合度が良くかつ確率水文量の変動が少ないモデルを選ぶ．

① 水文極値データの収集・吟味（独立性・均質性などをチェック）
② 候補となる確率分布を複数選ぶ．(対数正規分布，極値分布，ガンマ分布などが候補となる)
③ データに確率分布を当てはめる．(確率紙や解析的方法を用いる)
④ 当てはまり具合（適合度）を評価する．(目視，客観的適合度評価規準による)
⑤ 適合度の良い確率分布を複数個選ぶ．１つしかない場合はその確率分布を用いて確率水文量を決める．
⑥同程度の適合度を示す複数の確率分布に対して，ジャックナイフ法を適用し確率水文量の推定精度（ジャックナイフ推定誤差）を求め，これの小さいモデルを選ぶ．

４．大標本に対する水文頻度解析

近代的な水文観測が始められたのは明治時代であり，当時は19世紀の終盤であった．いまや21世紀であり，１世紀（100年）以上のデータが蓄積されてきたことになる．これからもデータの蓄積は進んでいく．

わが国の主要河川では水文頻度解析によって100年確率水文量を求めることが多い．そして，100年以上のデータを備える流域が出てき始めたのである．このような時代になっても，上述した中標本に対する解析手法を使い続けなければならないのであろうか．このような場合には，解析的方法（確率分布に頼るパラメトリックな方法）は必要なく，ノンパラメトリックな方法を用いたら良いのではないか．本章では，これについて考察する．

4.1 パラメトリックな従来型の手法による結果

まず，参考のために従来型の水文頻度解析の結果を示す．滋賀県北東部に位置する姉川流域（標本サイズN=108），天野川流域（N=107），芹川流域（N=107），彦根（N=100），余呉川流域（N=106），愛知県・豊川流域（石田地点 N=104）の6カ所の年最大雨量データを極値（グンベル）確率紙にカナン公式でプロットする．図-１は姉川流域年最大２日雨量の例である３)．データ点が直線状に並んでいないので，グンベル分布の適合が極めて悪いことは，一見して明らかである．このプロットに最小二乗法を適用すると，推定されたグンベル分布は，図１の回帰直線のようになり，非超過確率 0.990（再現期間100年に相当）とこの回帰直線の交点から x軸の雨量を読み取ると 360 mm となる．これが，図解法（ただし，カナン公式を用いた）による推定値（小標本の場合と同様のやり方）である．

同じ手法を，天野川流域，芹川流域，彦根，余呉川流域，豊川流域に対して適用すると，順に 330, 430, 335, 260, 302 mm という100年確率水文量が得られた．３．で述べた，中標本の場合の方法を姉川流域の年最大２日雨量に適用してみたところ，グンベル分布で 313mm という確率水文量が得られた． 一般化極値（GEV）分布を用いると357 mm，対数Pearson III型分布を用いると 373 mm となった．

グンベル確率紙上でプロット点を直線でつなぎ，100年確率水文量を求めると，**図-１**に示すように，440 mmとなる．同様に，天野川流域，芹川流域，彦根，余呉川流域，豊川流域に対して適用す

ると，順にそれぞれ 380,480, 380, 285, 285 mm という100年確率水文量になった．これらの結果は，グンベル分布とGEV分布によるパラメトリックな解析の場合と併せて，表-１に整理されている．

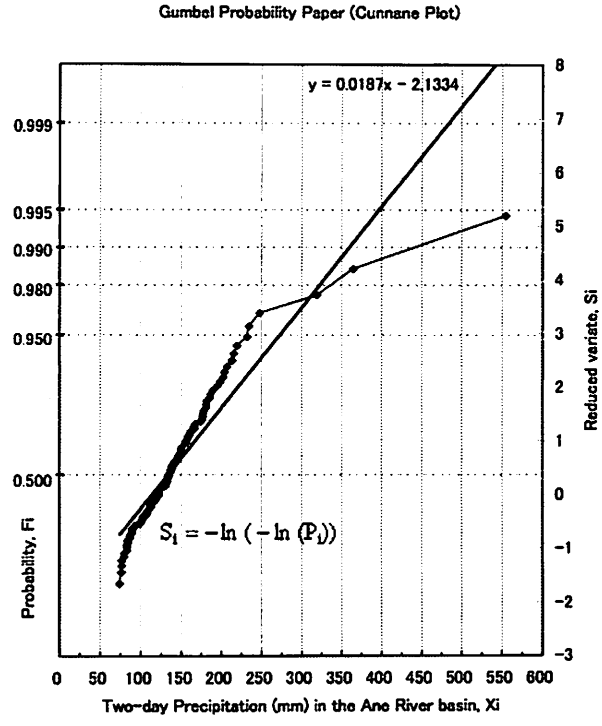

図-1 姉川流域年最大２日雨量の極値確率紙へのプロット

表-1 従来型のパラメトリックな手法による100年確率日雨量(mm) [姉川のみ２日雨量]

確率分布	グンベル分布（極値確率紙）		グンベル分布		GEV分布	
推定法	最小二乗法	経験分布法	L積率法	SLSC	L積率法	SLSC
姉川流域	360	440	290	0.033	298	0.026
天野川流域	330	380	250	0.032	281	0.021
芹川流域	430	480	313	0.149	401	0.039
彦根	335	380	231	0.181	292	0.044
余呉川流域	260	285	195	0.035	177	0.020
豊川流域	302	285	302	0.021	300	0.021

表-2 ノンパラメトリックな手法による100年確率日雨量(mm) [姉川のみ２日雨量]

	統計年数 N (年)	(I) ワイブル公式	(II) カナン公式	(II') ブートストラップ法
姉川流域	108	537	463	426
天野川流域	107	517	413	383
芹川流域	107	735	541	494
彦根	100	560	436	381
余呉川流域	106	413	315	288
豊川流域	104	286	284	280

なお，この計算については，前述の国土技術研究センターの水文統計ユーティリティを用いた．**表-１**からわかるように，適合度指標SLSC＞0.03 となる場合が多く，グンベル分布はデータへの適合度が全体的に悪いと言える（宝，2006）．またグンベル分布，GEV分布の両者とも芹川流域，彦根に対しての適合度が特に悪い（芹川に対してSLSC=0.149, 0.039, 彦根に対してSLSC=0.181, 0.044）．

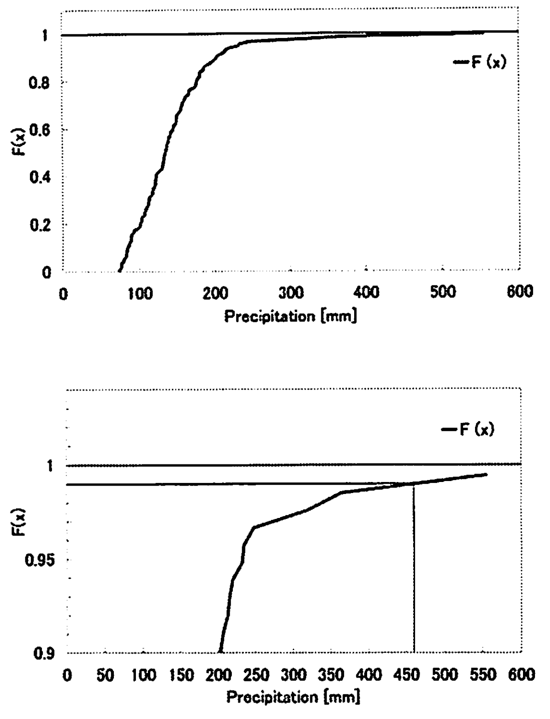

図-2 普通目盛にプロットした経験分布(上:全データ,
下:非超過確率0.90以上の部分を拡大表示)

4.2 新しい方法（ノンパラメトリック手法）

ここでは，対象とする確率年(T 年)以上の観測年数をもつ観測所の極値データを用いて，ノンパラメトリック法でT 年確率統計量を求める方法を提案する．計算機集約型統計手法（リサンプリング法の一種であるブートストラップ法）でその精度を検証する．具体的な手順は，以下のようである．

Step 1: 大標本極値データ（$N>T$）を確率紙にプロットする．確率分布に依存しないため，非超過確率を表す縦軸には単純な普通目盛（横軸は水文量，縦軸は0から1の非超過確率で普通目盛とする）を用いる．こうしてプロット（カナン公式）すると，図-2のようになる．図-2の上のグラフが，標本に含まれる全データに対する確率分布関数に相当する．このままでは見にくいので，非超過確率の大きい部分だけ拡大したのが，図-2の下のグラフである．プロッティング公式は，近年よく使われるカナン公式（$(i-0.4)/(N+0.2)$）を採用する．(比較のためワイブル公式 $i/(N+1)$ も併用する.)

Step 2: 各プロット点を順順に直線でつないでいく．すなわち，これがいわゆる経験分布である．

Step 3: 非超過確率 $1-1/T$（すなわち，リターンピリオドT年）に相当する水文量の値を線形内挿手法で求める．

Step 4: ブートストラップ法により確率水文量の偏りを補正し，その推定精度を調べる．すなわち，N個あるデータの内から繰り返しを許してN個のデータを抽出することによりブートストラップ標本を生成する．これに Step 1 から Step 3 の手順を多数回繰り返して確率水文量のブートストラップ推定量を得る．

ブートストラップ法は，元の標本から繰り返しを許して同じ標本サイズNの標本をB個作り出すもので，そのアルゴリズムは以下のようである(Efron, 1982, 1982)．

i) 一様乱数によるランダム抽出法を用いて，N 個のデータ $\{x_1, x_2, \cdots, x_N\}$ をもつ標本から，繰り返しを許して同じ N 個のデータをもつ標本を作り出す．作り出した標本を$\{x_1^*, x_2^*, \cdots, x_N^*\}$．と表す．これをブートストラップ標本（bootstrap sample）という．このブートストラップ標本を用いて，次のブートストラップ統計量が得られる．

$$\psi^* = \psi(x_1^*, x_2^*, \cdots, x_N^*) \tag{1}$$

ここに，ψ は，N個のデータから統計量を求める関数である．

ii) 前のステップ i) における操作を独立に多数回（B回）繰り返すことにより，B 個のブートストラップ標本が得られ，それぞれのブートストラップ標本に対して ψ^* を得ることができる．

第 b 番目のブートストラップ標本に対して得られる統計量を ψ^{*b} $(b = 1,2,\cdots,B)$ と表すこととする．

iii) 前のステップで得られた B 個の ψ^{*b} の平均値 $\psi^{*\cdot}$ は

$$\psi^{*\cdot} = \frac{1}{B}\sum_{b=1}^{B}\psi^{*b} \tag{2}$$

で計算できる．これが統計量 ψ のブートストラップ推定値である．

iv) 統計量 ψ の分散のブートストラップ推定値は，次式で求められる．

$$\hat{s}_B^2 = \frac{1}{B-1}\sum_{b=1}^{B}(\psi^{*b} - \psi^{*\cdot})^2 \tag{3}$$

ブートストラップ（bootstrap）という名称は，そのアルゴリズムが，自分自身で何とかする性質（self-help nature）を持つことから名づけられたと言う（Efron, 1979）．例えば，The American Heritage Dictionary では，"by one's own bootstraps" = by one's own efforts とある．編み上げ靴のかかとに付けられたストラップ（つまみ革）を自分で引っ張って，履きにくい編み上げ靴を人の力を借りずに履くというようなことから，こうした表現が使われるようになった．

4.3 結果と考察（ノンパラメトリック手法）

極値確率紙ではなく，4.2で示した普通目盛り確率紙を用いる方法を適用する．確率分布を用いないので確率分布関数に含まれる母数（parameters）の推定が必要でないため，ノンパラメトリックな方法である．ただし，各順序統計量（x_1, x_2, ・・・, x_N）に非超過確率を与える必要がある．ここでは，ワイブル公式，カナン公式によりプロットして得られるT年確率雨量を求める．T=100 年を超えるような大きさの標本（N>100）は，水文統計では「大標本」と称して良いであろう．

上述の図-２のグラフは，姉川流域の年最大２日雨量に適用したものであって，この方法によれば，100年確率雨量は463 mm となった．同様にしてワイブル公式を用いると，100年確率雨量は537 mm となった．これは，ここで扱った年最大雨量系列に対して，従来のパラメトリックな方法では，確率分布の適合度が悪く，図-１に示すような，最大値が飛び抜けて大きいような標本（データセット）に対して再現確率統計量（確率水文量）を過小推定することによる．

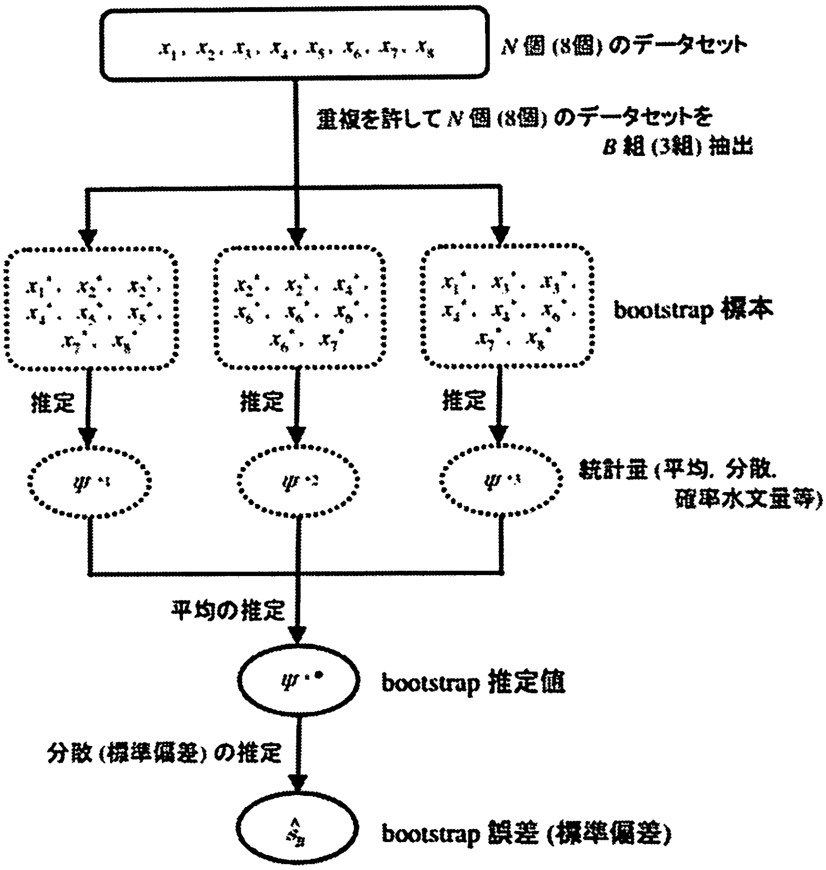

図-3　ブートストラップ法の概略説明図

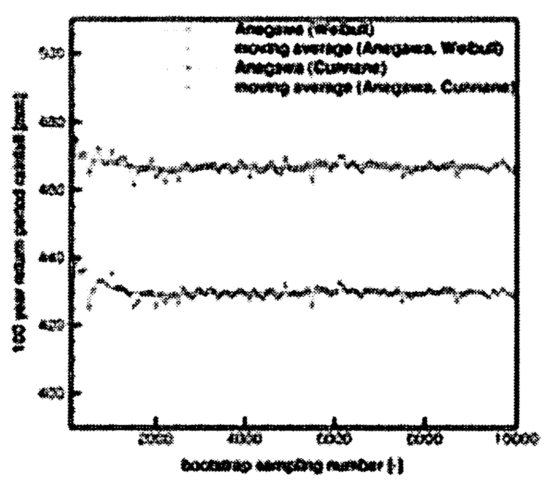

図-4　ブートストラップ標本の数Bとそれに応じて得られる100年確率雨量の変動
(姉川流域平均２日雨量；上：ワイブル公式を使った場合，下：華南公式を使った場合；
両者とも移動平均をとって平滑化した値も併記してある)

他の４カ所についても同様の方法を適用して得られた結果を整理すると，表-２の（I），(II)のようになる．これと表-１の結果とを比べると分かるように，提案する方法の方が確率雨量を多く見積もることになる．バイアス補正のためにブートストラップ法を適用する．　試算を繰り返し，ブートストラップ標本の数BをB=100 から10000まで100個きざみに変えてみたところ，2000以上とすると結果が安定することがわかった．図-４は，姉川流域の年最大２日雨量に対して，試算した

結果を示している．横軸がブートストラップ標本の数で，縦軸がブートストラップ法によって得られる100年確率雨量である．図中，赤線（上）がワイブル公式を使った場合，青線（下）がカナン公式を使った場合で，両者とも移動平均をとって平滑化した値も併記してある．

結局 B=2000 として，ブートストラップ法で推定し直した姉川流域の100年確率２日雨量はカナン公式で426 mm となった．すなわち，ここで提案した手法によれば，100年確率２日雨量は426 mm である．これは，図-１，表-１の極値確率紙における経験分布法による440mm を少し下回る結果である．なお，参考のため試みたワイブル公式の場合は 470 mm となった．これは，このプロッティング・ポジション公式がもつ特徴からして，過大評価であるとみなすことができる．

カナン公式を使って同様に天野川流域，芹川流域，彦根，余呉川流域，豊川流域における100年確率日雨量をブートストラップ法で求めた結果は，それぞれ，383，494，381，288，280 mm となった（表-２の（II’））．豊川流域の年最大日雨量データを極値確率紙にプロットすると，図-５のようになる．このデータセットの特徴は，第１位の雨量と第２位の雨量の値が近いことであり，図-１で見たような姉川ほか滋賀県北東部の雨量のように第１位が大きく飛び離れているものとは異なる．すなわち，表-２の豊川流域の行を見てわかるように，(I)，(II)，(II’) の推定値の差が小さくなる．第１位と第２位との狭い区間において100年確率の値が決まるからである．

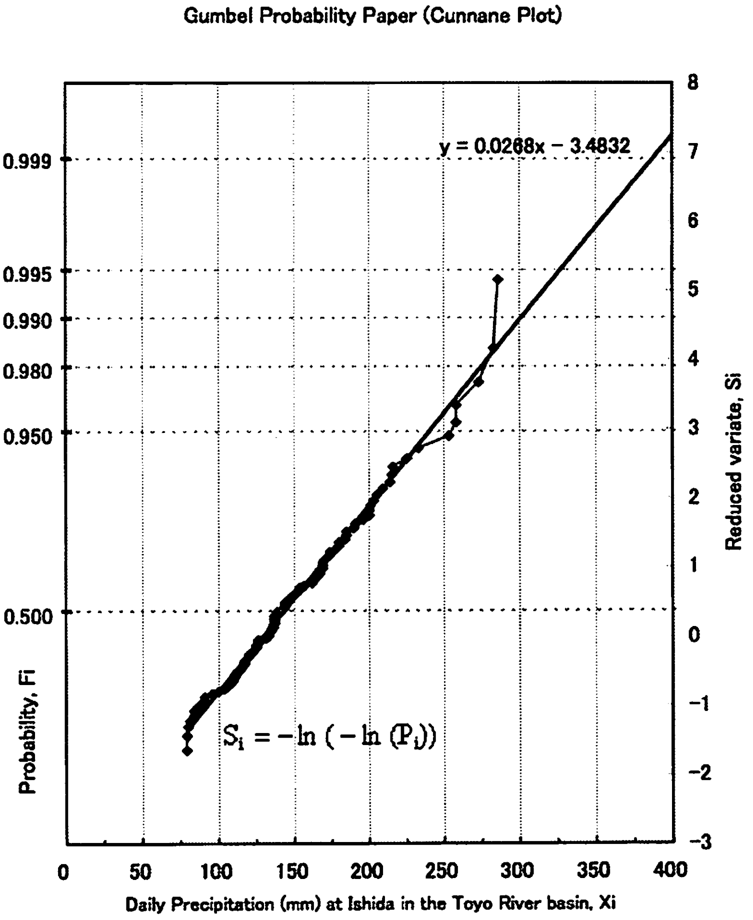

図-５　豊川流域における年最大日雨量の確率プロット

図-５からわかるように，パラメトリックな従来法であれば，極値（グンベル）分布に当てはめた解は，図中の回帰直線であり，これによると100年確率（すなわち非超過確率0.990の線との交点）の日雨量値は 302 mmということになる（表-１のグンベル分布の最小二乗法の欄）．なお，極値確

率紙上でのプロットがこの程度の高い直線性を示すと，グンベル分布の他の推定法（L積率法）でも302 mm，GEV分布でも 300 mm とほぼ近い推定値を与えることになる．

ただし，ノンパラメトリック法では，この豊川のデータセットでは第１位と第２位とが近い値であるので，プロット点が縦に立った形でプロットされるから，これらの推定値よりも15 mm 以上下方修正されることになる．一方，図-１で見た姉川のように第１位と第２位とが離れていると，パラメトリック法による解析解は，ノンパラメトリック法によって上方修正されることになる．

この方法は，筆者らが提案してきた考え方（宝，2006；宝・小林，2008）を整理して再編集したものである．最近になって，石原・仲江川（2008）が，この方法を気象庁の全国51地点の日降水量データ（1901年～2006年の106年の系列，那覇のみ戦中データの欠損により101年）に適用した結果を報告している．その報告によると，ノンパラメトリック法で得られた100年確率日雨量推定値（A）と従来型のパラメトリックな方法で得られた推定値（B）とを比較したところ，両者の比（A/B）は51地点平均で1.03であった．

すなわち，ノンパラメトリック法が，従来法よりも3%程度大きい確率水文量を与える結果を示している．106年のうちの最大値が，ノンパラメトリック法で得られる推定値よりもかなり大きい地点がいくつもあり（名古屋，高知，宮崎，津など），これらがその結果に影響しているとも言える．ノンパラメトリック法の推定値の方が小さくなる地点もあることに留意しなければならない．いずれにせよ，さらに検討を重ねる必要があると考えている．

５．おわりに

従来，求めようとする再現確率統計量（確率水文量）のリターンピリオドに対して，標本サイズが相対的に小さいことが多かった．よって，確率分布を用いるパラメトリックな手法により，極値データの母集団を推定する方法がとられてきた．しかしながら，データ数が極端に少ない小標本の場合には，推定精度が悪いことが水文統計学的に知られている．中標本の場合においては，いくつもの確率分布を検討しなければならないし，その作業における実務者の悩みは今なお大きいものがある．

標本サイズの大きさによって，頻度解析の手法を変えるべきである．本稿では，小標本には確率紙を用いた安全側（ワイブル公式を用いる）の簡易推定，中標本に対しては従来のパラメトリックな方法，そして，リターンピリオドを越えるような観測年数を持つ標本に対しては，ノンパラメトリックな方法を推奨した．特に，リターンピリオドを超えるようなサンプルサイズを持つ標本に対してノンパラメトリックな方法（経験分布を用いる方法）は有用であろうと考える．ただし，上で見たように，ノンパラメトリックな方法（経験分布を用いる方法）を標本に対して１回だけ適用すると，偏りのある100年確率統計量の評価を与えるので，ブートストラップ 法によって偏り補正した方が良い．

50年確率の水文量を求める場合，50年程度の水文極値データは，多くの場所で得られているものと考えてよい．ここでは，100年確率水文量を求める例を示したが，30～100年の中標本に対しても，再現確率年が30～80年程度の場合は，このノンパラメトリック法は利用可能である．

100年以上データが蓄積されていても，200年確率の水文量を求めるときは，やはり外挿問題になる．中標本に対して，100～300年確率の水文量を求める場合と同じである．これについて，筆者は，可能最大降水量（PMP）や可能最大洪水流量（PMF）の概念を応用することを考えている．このアイデアについては，講義中に少し述べたい．

引用文献

石原幸司・仲江川敏之（2008）：全国51地点におけるノンパラメトリック手法を用いた確率降水量の算出，水文・水資源学会誌，第21巻，第6号，pp. 459-463，2008.

Efron, B. (1979): Computers and the Theory of Statistics: Thinking the Unthinkable, *SIAM REVIEW*, Vol. 21, No. 4, pp. 460 -480.

Efron, B. (1982): *Jackknife, the Bootstrap and Other Resampling Plans*, SIAM Monograph, No. 38, 92 pp.

高棹琢馬・宝　馨・清水　章（1986）：琵琶湖流域水文データの基礎的分析，京都大学防災研究所年報，第29号B-2，pp. 157-171.

宝　馨（1998a）：現代水文統計論―水文頻度解析のモデル，手法とその評価―，1998年度（第34回）水工学に関する夏期研修会講義集，Aコース，土木学会水理委員会・海岸工学委員会，1998年7月，pp.　A-8-1～A-8-20.

宝　馨（1998b）：水文頻度解析の進歩と将来展望，水文・水資源学会誌，第11巻，第3号，1998，pp. 740-756.

宝　馨（2006）：大標本時代の水文頻度解析手法―リターンピリオドを超えるようなサイズの標本に対する極値データ解析―，京都大学防災研究所年報，第49号B，pp. 7-12.

宝 馨・小林健一郎（2008）：水文極値の大標本に対するノンパラメトリックな頻度解析，日本自然災害学会年次学術講演会，九州大学，II-4-3.

宝　馨・小林健一郎（2009）：標本サイズと水文頻度解析，水工学論文集，土木学会，第53巻，pp. 205-210.

宝　馨・高棹琢馬（1988）：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，第393号/II-9，pp. 151-160.

宝　馨・高棹琢馬（1989）：「水文頻度解析における確率分布モデルの評価規準」への討議に対する回答，土木学会論文集，第405号/II-11，1989，pp. 267-272.