

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—

寶 馨

要 旨

従来、水文データは標本サイズが小さく（観測年数が短く）、水文頻度解析においては、確率紙を用いた図式推定法（最小二乗法）や種々の確率分布を用いた解析的方法が用いられてきた。最近、観測年数が100年を超える場所も多くなってきたので、従来の方法を改善する可能性も出てきた。ここでは、小標本（30年未満のデータセット）及び中標本（30年以上）の場合の水文頻度解析の方法を概説するとともに、100年を超えるような大標本に対して、確率分布に頼らない順序統計学的方法とBootstrap法による確率水文学量の推定方法を提案し、その精度を例示した。こうした考え方は、今後ますますデータの蓄積が進む実務現場における一つの客観的な分かり易い新しい方法として一考の価値があると言えよう。

キーワード： 確率水文学量，標本サイズ，プロットイング・ポジション，ブートストラップ法，順序統計学的方法

1. はじめに

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学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（確率水文学量）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、当該の水文学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選び、それによって再現確率統計量を求める

この時、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。特に、標本サイズが小さい（観測年数が短い）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。したがって、小標本の場合、以下のような手順が推奨できる。

2. 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。

- ・ 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。

- ・ 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トング・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、そのほかのプロットング・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、高棹ら，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学量を求めようとする事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. 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（高棹ら，1986；宝・高棹，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、確率水文学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学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学量の変化が小さく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・高棹，1988）。すなわち、適合度に加えて、確率水文学量の変動性（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、確率水文学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に、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、逆に確率水文学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、以下の手順で適合度が良くかつ確率水文学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学を決める。
- 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学の推定精度（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. 大標本時代の水文頻度解析

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年確率水文学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 方法

ここでは、100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100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（リサンプリング法的一种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、以下のようなものである。

Step 1: 大標本 ($N > 100$) の極値データを確率紙にプロットする。分布に依存しないため、単純な普通目盛（横軸は水文学、縦軸は0から1の非超過確率で普通目盛とする）を考える。プロット公式は、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 0.99（すなわち、リターンピリオド 100 年）に相当する水文学の値を線形内挿手法で求める。

Step 4: ブートストラップ標本を多数生成し、Step 1 から Step 3 の手順を繰り返す、確率水文学の推定精度を調べる。

4.2 結果と考察

姉川流域（滋賀県）の年最大2日雨量の水文学時系列 ($N=108$) に適用してみた。極値（グンベル）確率紙にカナン公式でプロットすると **Fig. 1** に示すようである。まず、参考のために従来型の水文学頻度解析の結果を示すとグンベル分布 (**Fig. 1** の回帰直線) のデータへの適合度は悪く、非超過確率 0.990（100年確率）に対して 320 mm という確率水文学が得られた。一般化極値 (GEV) 分布を用いると 357 mm, Log-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年確率水文学を求めると、**Fig. 1** に示すように、440 mm となる。

普通目盛でカナン公式によりプロットすると、**Fig. 2** のようになり、100年確率雨量は 463 mm となる。ワイブル公式を用いると 537 mm となった。

バイアス補正のためにブートストラップ法（宝, 1998; 宝・高棹, 1988）を適用する。試算を繰り返す、ブートストラップ標本の数を100から10000まで変えてみたところ、2000以上とすると結果が安定することがわかった。ブートストラップ法で推定し直した確率雨量はカナン公式で430 mm, ワイブル公式で470 mm となった。

すなわち、ここで提案した手法によれば、100年確率雨量は430 mmである。これは、リターンピリオド 100年を超える極値データの標本に対して確率分布を一切想定しない、客観的な方法による結果であると言える。

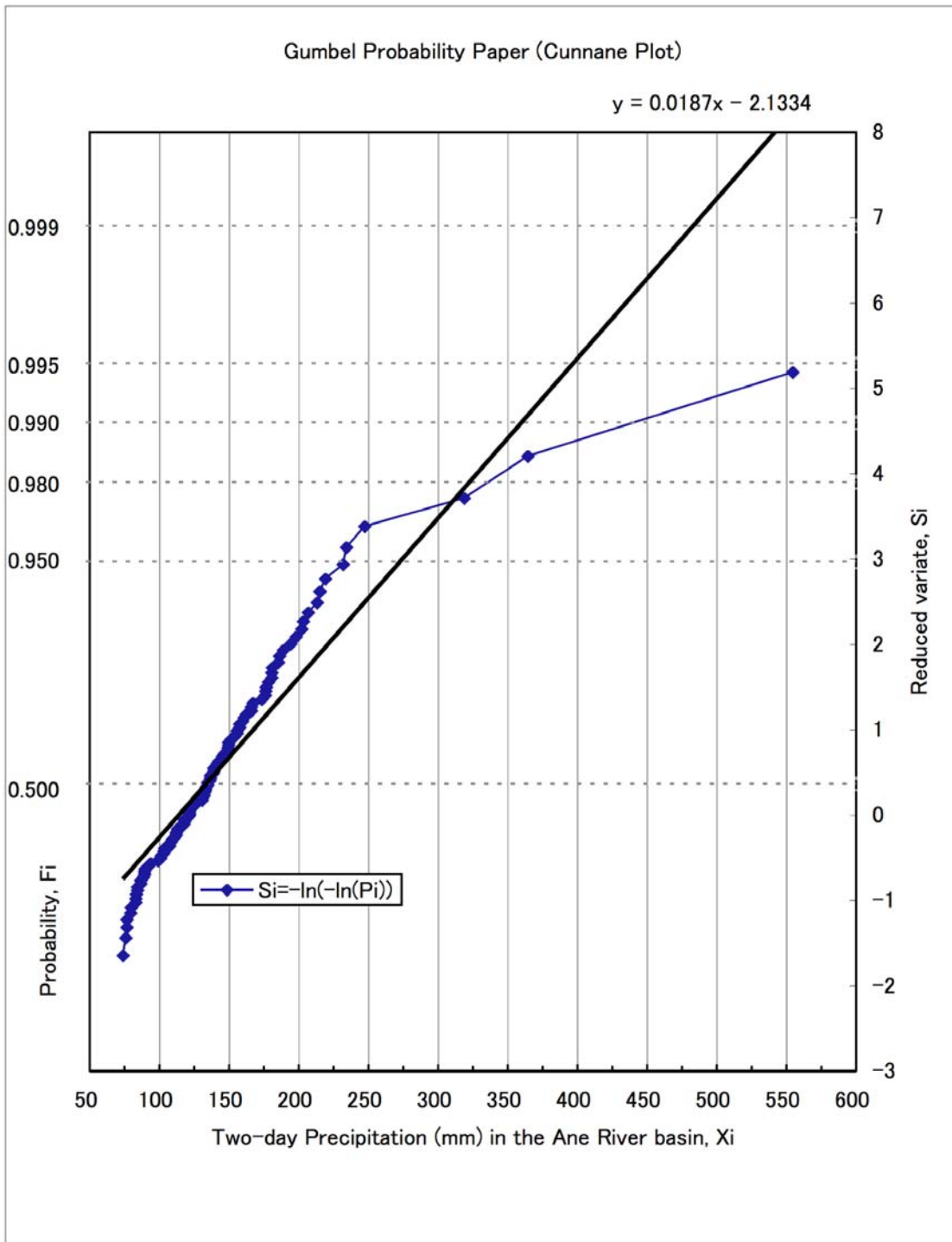


Fig. 1 Probability Plot on the Gumbel probability paper
(annual maximum 2-day precipitation in the Ane River basin, N-108)

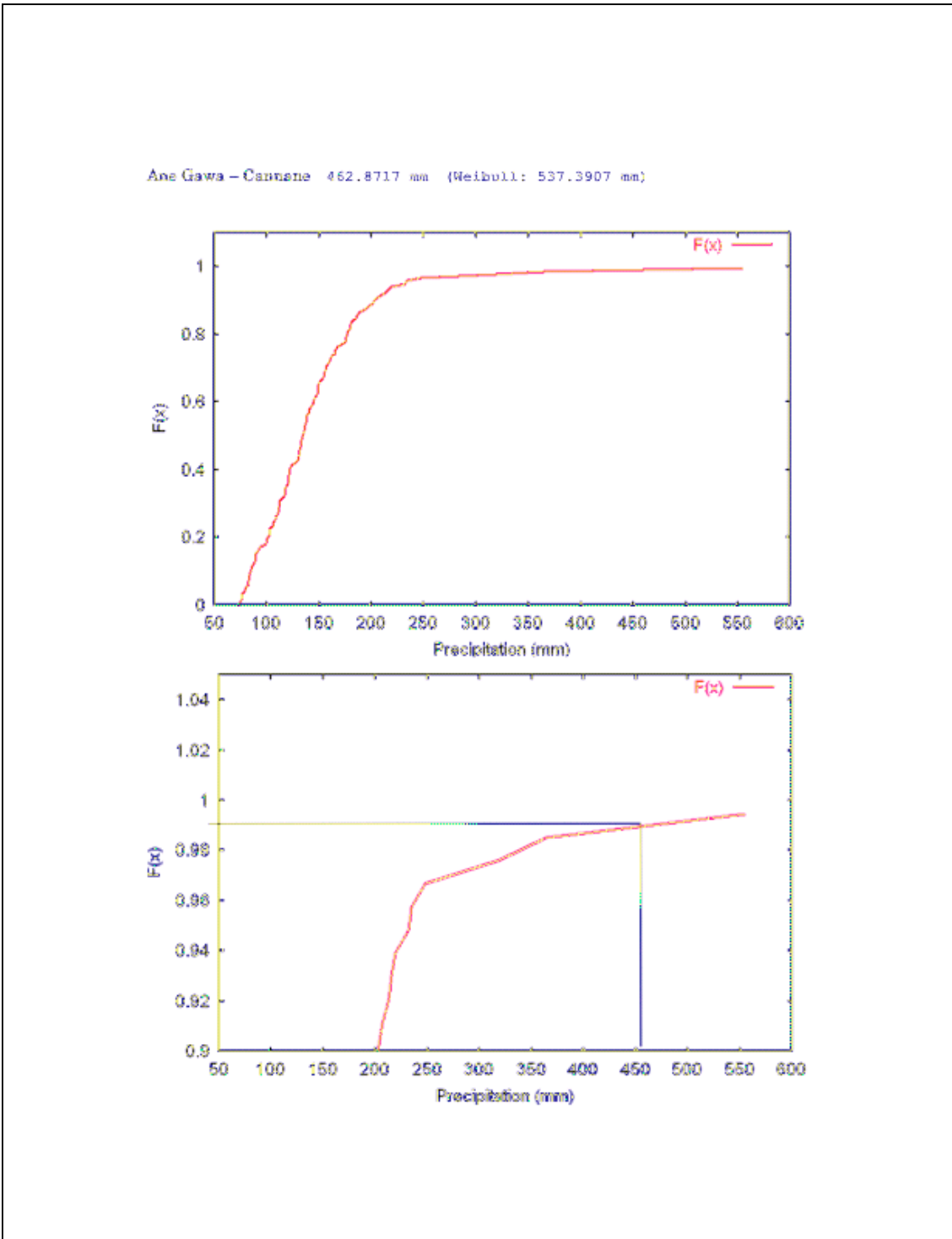


Fig. 2 Plotting on a probability paper with normal-scale Y-Axis
 (Note: The lower panel is an enlarged view for the non-exceedance probability $F_i > 0.9$).

5. 結 論

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）は有用ではないかと考える。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回だけ適用すると過大な100年確率統計量の評価（463 mm）を与える。Bootstrap法によって偏り補正した方がよい（430 mm）。

今後の課題として、これを現場に適用しようとする、以下のような問題が考えられる。すなわち、データ年数がリターンピリオドを超えたとたんに計画・設計値（確率水文量）が変更になることになるがそれでいいのかどうか、上流では50年確率、下流では100年確率という河川がたくさんあるが、上流と下流とで手法や計画の整合性が問題視されないか、といったことである。これらについては、現場の実務者とともに検討を加える必要があるだろう。

謝 辞

本研究で用いた姉川流域年最大2日雨量のデータは、滋賀県河港課より提供を受けたものである。ここに記して謝意を表す。

参考文献

- 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（1998）：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・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.

Frequency Analysis of Larger Samples of Hydrologic Extreme-Value Data --- How to estimate the T-year quantile for samples with a size of more than the return period T

Kaoru TAKARA

Synopsis

This paper describes how to estimate the T-year events (quantiles) for hydraulic systems design and planning in river basins. Traditional frequency analysis methods, which have been providing T-year quantile estimates, include the graphical analysis with probability papers for smaller samples and the parametric approach with a number of probability distribution functions (PDFs) especially for extreme hydrologic variables for larger samples. Recently many hydrological observatories are getting longer historical data and the length of record there is exceeding the return period T. This paper recommends the usage of non-parametric approach for such larger samples, because this approach avoids the difficulty of selecting the best PDF and parameter estimation method from many candidates. The bootstrap method, which is one of the resampling methods, is used for bias correction and quantification of quantile estimation errors. A basic idea and its application results are illustrated here to indicate a direction of hydrologic frequency analysis in the future where more observatories will have longer datasets

Keywords: quantile estimation, sample size, plotting position, the bootstrap, non-parametric approach