水工学論文集，第43巻，1999年2月

毎年資料と非毎年資料による確率水文量の評価

HYDROLOGIC FREQUENCY ANALYSIS WITH ANNUAL MAXIMUM SERIES AND PARTIAL DURATION SERIES

田中茂信[1]・宝　馨[2]

Shigenobu TANAKA and Kaoru TAKARA

[1]正会員　工修　(財) 国土開発技術研究センター　調査第一部 (〒105-0001 港区虎ノ門2-8-10)

[2]正会員　工博　京都大学防災研究所教授　水災害研究部門 (〒611-0011 宇治市五ヶ庄)

Hydrologic frequency analysis often uses the annual maximum series (AMS) of rainfall and river discharge. However, sometimes the use of AMS is insufficient because AMS may include non-flood annual maximum discharges. This paper describes the use of the partial duration series (PDS) as well as the AMS, giving the theoretical background of the relationship between the generalized Pareto (GP) distribution for the PDS and the generalized extreme-value (GEV) distribution for the AMS. These distributions are evaluated for seventy-nine river discharge samples in Japan from the view point of both goodness of fit and stability of quantile estimates obtained by the jackknife resampling method.

Key Words : *Annual maximum series, Partial duration series, Generalized extreme-value distribution, Generalized Pareto distribution, Resampling method*

1．はじめに

従来，河川計画の基本量である外力を定める際には，毎年最大値資料(AMS, Annual Maximum Series)を対象とした水文頻度解析が主に用いられてきた．しかしながら，毎年資料の中には寡雨年の年最大流量等のように洪水とは呼べないようなものも少なからず含まれている場合がある．このような場合，毎年資料を用いて確率年100年から200年の確率水文量を求めるとその精度や信頼性について疑問が残る．

また，河川によっては第二次世界大戦前後において未曾有の洪水を経験しているものがあり，その被災経験が流域住民に強く残っている場合がある．このような場合，その大洪水の流量も含めて，その流域の洪水の発生頻度を推定すべきであるが，一般に戦中から戦後にかけての数年から十数年間は雨量や流量の記録が乏しく，従来の毎年最大値を用いる方法が適用できない．

このような問題を回避できる方法として，時系列資料の中からある値より大きな水文量のみを抽出して標本とする非毎年資料(PDS, Partial Duration Series)による頻度解析がある．

本研究では，非毎年資料を用いて確率水文量を評価する方法について紹介する．また，毎年資料を用いた確率水文量との比較を行うため，非毎年資料のうち毎年値と同じ個数だけデータを抽出する非毎年年数最大値を用いた頻度解析を行なうとともにその特性を検討した結果を報告する．

2．毎年資料と非毎年資料の理論的関係

毎年資料と非毎年資料の理論的関係およびその解析方法について，Handbook of Hydrology に詳しく記述されている[1)]のでそれを紹介する．

洪水、降雨、およびそのほかの多数の水文資料のモデル化には、一般的に2通りの方法が使用されている．毎年最大値資料(AMS)では、各年の最大の事象について検討することができる．部分的水文資料(PDS) または閾値超過ピーク (POT, Peaks Over Threshold) とも呼ばれる資料を使用する方法は、打切レベルまたは閾値レベルを超える全てのピークについて解析するものである．AMSを使用する方法では，各年の最大事象のみが用いられるため，ある年の2番目に大きな事象が他の年の最大事象を上回るか否かが考慮されないという問題がある．さら

に，乾燥地域または準乾燥地域においては，渇水年における年最大洪水流量が0となる，またはあまりにも小さいため洪水と称するには無理がある場合も考えられる．

PDSでは，所定の閾値を上回るすべての独立なピークが対象とされるため，上記の問題は回避される．幸いにも，年間超過確率は後述する式（4）を用いたPDS解析や経験的関係から推定することができる．PDS解析の擁護者の主張によるとPDS記録は比較的長期的で信頼性の高い記録が使用できることが多く，閾値を超えるピークの発生率が十分に大きいとき（即ち，閾値を超過する確率が指数分布で表され，閾値を超える事象の発生間隔がポアソン分布にしたがう場合，1.65事象／年より大きいとき）PDS解析を用いることによって，対応する毎年最大値資料によるよりも高精度の確率水文量を推定できる．このような長所を持つ一方で，PDS解析には，（同一の事象に関する複数のピークではない）独立ピークを特定するための判定基準が必要になるという短所がある．したがって，PDS解析は，毎年最大値資料を使用する解析法に比して複雑な方法となる．

PDSのモデル化においては二つの問題が生じる．一つは，閾値レベルを超える大規模事象の発生率をモデル化しなければならないという問題であり，もう一つはそれらの事象の大きさをモデル化することである．たとえば，事象の発生率のモデル化にはポアソン分布が使用されることが多く，閾値を上回るピークの規模を記述するためには，指数分布が使用される．非超過確率の大きい事象の場合，複数のモデルにおいて1年当たりの平均発生回数が一致するときは，発生率に関する確率モデルはそれほど重要ではない．

毎年最大値資料と部分的水文資料における事象の頻度の間には，いくつかの一般的な関係が存在する．PDSについて，λを閾値 x_0 を上回る事象の1年当たりの平均発生回数とする．$G(x)$ は，事象がxを下回る確率，したがって (x_0, x) の範囲をとる確率とする．$x \geq x_0$ となる任意のレベル xについて，発生率は次の式で表される．

$$\lambda^* = \lambda[1 - G(x)] \tag{1}$$

対応する毎年最大値資料列に関する累積分布関数 $F_a(x)$ は，ある年に関する年間最大値がxを超えない確率である．独立事象については，1年間を通じてxを超過しない確率はポアソン分布によって与えられ，次の式で表される．

$$F_a(x) = \exp(-\lambda^*) = \exp\{-\lambda[1 - G(x)]\} \tag{2}$$

この関係は，1年を，発生率が λ^*/m の m個の区間に分割することによって導出される．λ^*/mが小さいときは，1年間の非出現確率は基本的に $(1-\lambda^*/m)^m$ となる．式（2）は，$m \Rightarrow \infty$ のときの極限において得られる．

式（2）から，年最大事象に関する累積分布関数と，部分的水文資料ピークに関する発生率および分布との関係性が明らかになる．リターンピリオド T_a とすると年超過確率 $1 - F_a(x)$ は $1/T_a$ で表される．部分的水文資料においてこれに対応する超過確率 $[1 - G(x)]$ を q_e とすると，式（2）は，次のように表すことができる．

$$\frac{1}{T_a} = 1 - \exp(-\lambda q_e) = 1 - \exp(-\frac{1}{T_p}) \tag{3a}$$

ここに，$T_p = 1/\lambda q_e$ はPDSにおけるレベル xに関するリターンピリオドである．T_p について式（3a）を解くと、次の関係式が得られる．

$$T_p = -\frac{1}{\ln(1 - 1/T_a)} \tag{3b}$$

PDSにおいては1年間に起こりうる事象は1を上回るため，T_p は T_a よりも小さい．

式（3a）によって，xを上回る事象の平均発生率 λq_e は，毎年最大値資料における年超過確率 $1/T_a$ に変換される．$T_a > 10$，すなわち発生頻度の低い事象については，年超過確率 $1/T_a$ は，PDSに関する平均発生率 $\lambda q_e = \lambda[1 - G(x)]$ にほぼ等しくなる．したがって，$T_a = T_p$ となる．

式（2）において，閾値 x_0 を超える事象の規模を一般パレート（GP, Generalized Pareto）分布

$$G(x) = 1 - \left[1 - \kappa\left(\frac{x - x_0}{\alpha}\right)\right]^{1/\kappa} \qquad \kappa \neq 0 \tag{4a}$$

で表すことができるとすると，κが正の値のとき，累積分布関数は上界 $x_{\max} = x_0 + \alpha/\kappa$ を持つ．$\kappa < 0$ の

ときは非有界であり，長く尾を引く分布となる．$\kappa=0$ のときは，式（4b）のような2母数指数分布となる．

$$G(x)=1-\exp\left[-\frac{x-x_0}{\alpha}\right] \qquad \kappa=0 \qquad (4b)$$

式（4a）を式（2）に代入すると，$\kappa\neq 0$ のときに x_0 を上回る毎年最大値資料に関する一般極値（GEV, Generalized Extreme Value）分布が得られる．

$$F_a(x)=\exp\left[-\left(1-\kappa\frac{x-\xi}{\alpha^*}\right)^{1/\kappa}\right] \qquad \kappa\neq 0 \qquad (5a)$$

$\kappa=0$ のときはグンベル分布となり，次の式で表される．

$$F_a(x)=\exp\left[-\exp\left(-\frac{x-\xi}{\alpha}\right)\right] \qquad \kappa=0 \qquad (5b)$$

$x\geq x_0$ の時、変換パラメータ ξ および α^* は次のように定義される．

$$\left.\begin{aligned}&\xi=x_0+\frac{\alpha(1-\lambda^{-\kappa})}{\kappa},\ \alpha^*=\alpha\lambda^{-\kappa}\ \text{ for }\ \kappa\neq 0\\&\xi=x_0+\alpha\ln(\lambda)\qquad \text{for}\quad \kappa=0\end{aligned}\right\} \quad (6)$$

この一般的なポアソン－パレートモデルは，柔軟性が大きく，多くの現象について妥当なモデルとなっている．毎年最大値資料および部分的水文資料から得られたGEV分布の形状母数 κ の地域推定値は互換的に使用できる．

実際には発生率 λ は，1年当たりの x_0 を超過する平均回数という単純な方法で推定できる．そのほかのパラメーターは，次式により求められる．

$$\left.\begin{aligned}&\kappa=\frac{\mu-x_0}{\lambda_2},\ \alpha=(\mu-x_0)(1+\kappa)\ \text{ for }\ \kappa\neq 0\\&\frac{1}{\beta}=\alpha=\mu-x_0\quad \text{for}\quad \kappa=0\end{aligned}\right\} \quad (7)$$

ここに，$\mu=\lambda_1$ は標本の平均，λ_2 は2次L積率，β は指数分布の尺度母数である．

3．研究方法

本研究では，対象地点の確率流量を毎年最大値と非毎年年数最大値（PDSのうち統計年数と同じ数だけ上位から取り出したもの）の両面から適合度と安定性の指標を用いて評価し[2)]，両者を比較する．理論的には，年最大値と非毎年年数最大値は再現期間が10年より大きい範囲で一致する．前述したように，われわれが利用できる資料の観測期間は40年前後が多く，資料は多少にかかわらずバラツキを持っている．また，梅雨や台風による洪水がない年の最大流量は洪水と見なせないような値である場合がある．AMSを用いて求めた確率水文量とPDSを用いて求めた確率水文量は必ずしも一致しない．このような場合，AMSのみを用いて確率水文量を推定するより，PDSからも確率水文量を推定し，両者を比較するほうがよりよい確率水文量の推定が行えると考えるからである．

確率水文量の値やその変動特性は，確率分布モデルだけでなく用いる母数推定法によっても異なる．本研究で検討対象とした確率分布モデルと母数推定法は，AMSについては，水文統計の分野で用いられてきた8種類の確率分布モデル（指数分布，グンベル分布，平方根指数型最大値分布[3)]（SQRT-ET），GEV分布，対数ピアソンIII型分布（原標本の積率解(LP3Rs)，対数標本の積率解(LP3Ls)），対数正規分布（岩井法，2母数(LN2)の場合最尤法））を用いた．PDSについては，GP分布と指数分布（角屋の方法[4)]，GP分布で形状母数 $\kappa=0$ としたもの(GPExp)）を用いた．GP分布の母数推定はLモーメント法を用いた[1)]．

4．解析結果

(1) 解析に用いた資料

a)標本の観測年数

今回，解析に用いた資料はわが国の一級河川における79箇所の流量である．流量の観測が始まった時期が異なるため，標本の統計年数は，最大が62年，最小が27年、平均値が42.9年であり、40年～44年の階級が35標本あり，全体の約半数をしめる．

b)AMSとPDSの比較

PDSには年間第2位や第3位等の洪水流量も含まれる．今回の検討では非毎年年数最大値を対象としているので，標本の大きさはAMSと同じである．もし，1年に1回しか洪水が発生せず，これ以外の期間は極めて小さい流量（AMSの最小値より小さい流量）だとすると，AMSとPDSは一致する．実際には大洪水のあった年の第2位の洪水が毎年値の上位m＋1番目洪水より大きいなどの状況になり，AMSとPDSを併せてプロットすると上位からm個

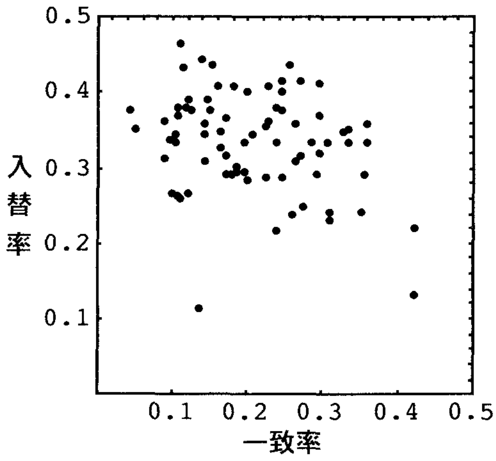


図-1　一致率と入替率

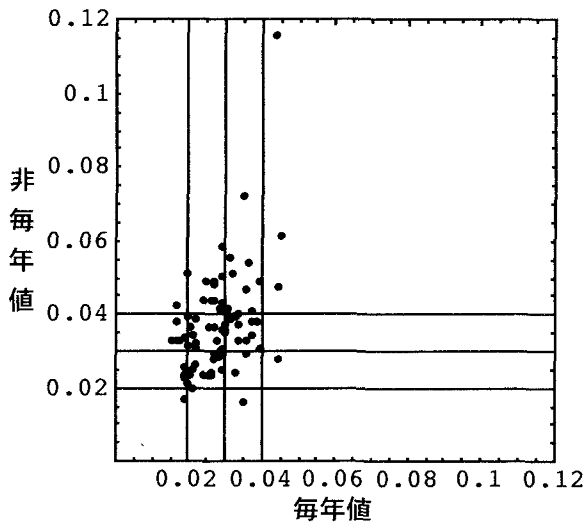


図-4　毎年値と非毎年値の最小SLSCの比較

までは一致するがそれ以降は離れてプロットされる．そして，PDSの最小値は毎年値の小さい方からk番目よりも大きくなる(k>0)．mと標本の大きさNの比を一致率，kとNの比を入替率と呼ぶことにすると，それらは図-1のようになる．洪水が毎年のように発生するが年間の発生回数が少ない河川では，一致率は高くなり結果として入替率が低くなる傾向がみてとれる．一方，一致率が小さい河川では入替率が大きくなる傾向にある．図-1の右下にある2点のプロットは九州の標本である．

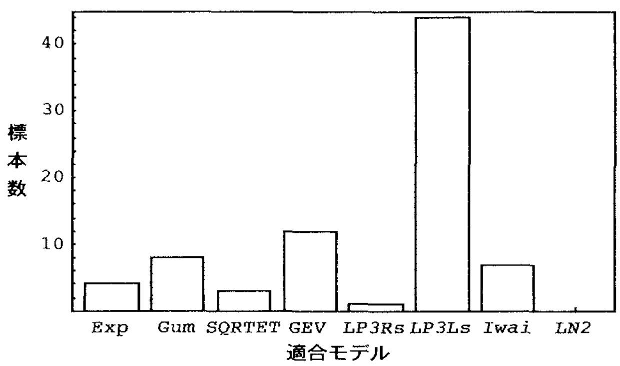


図-2　毎年値のSLSC最小モデルの頻度分布

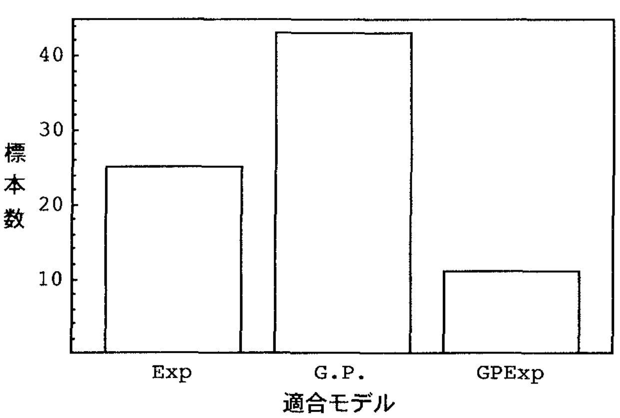


図-3　非毎年値のSLSC最小モデルの頻度分布

(2)確率分布モデルの適合度

a)SLSCの算定

確率分布モデルによっては標準変量の考え方がいくつもあり得るものがある．本研究では，平方根指数型最大値分布と一般極値分布のs_i^*はグンベル分布のs_i^*と同じものを用いた．

また，SLSCの算定にあたっては，各種の分布に対して近似的に不偏なクォンタイルを与えるCunnaneの公式をプロッティングポジションに用いた[1]．

b)SLSCの算定結果と評価

図-2はAMSについて最小のSLSCとなった確率分布モデルのヒストグラムを示したものである．LP3Lsが44標本で過半数を占める．次に適合度がよいのはGEV分布であり，これにグンベル分布，岩井法が続く．

PDSについて同様に整理したものが図-3であり，GP分布が最も多く43標本であった．

この結果を見るとどちらの場合も3母数の確率分布モデルがデータへの適合という面では有利であることを示している．

図-4は横軸および縦軸に，それぞれ毎年値の確率分布モデルの中での最小のSLSCおよび非毎年値の確率分布モデルの中での最小のSLSCをとって比較したものである．毎年値のSLSCはほぼ河川流量に対する良好な適合度の判定基準である0.04以内[5]に納まっているが非毎年値の方は0.04で区切ると26標本において適合度が満足できる確率分布モデルを失ってしまう．全体的には，少し正の相関があるが，強くはない．この図にプロットされていないが，SLSCが最小ではなく0.04より小さいものはたく

さんあるので，毎年値と非毎年値のどちらかで見るのではなく，両者を見た上で判断することが望ましいといえる．

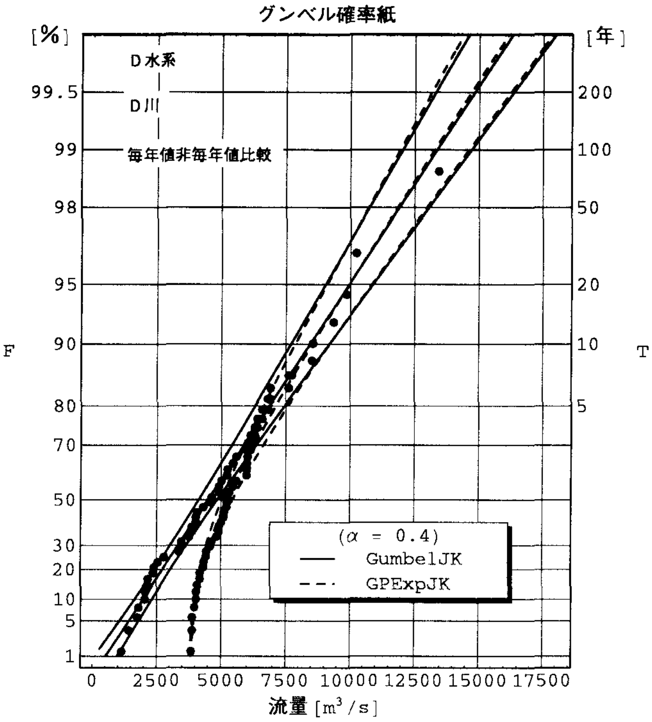
(3)jackknife 法による確率水文量の不偏推定と毎年値と非毎年値の比較評価

確率分布モデルの安定性の評価はリサンプリング手法[6)]で行う．リサンプリング手法にはjackknife法やbootstrap法がある．jackknife法は大きさｎ個の標本のうち任意の１データを欠いた大きさｎ－１個の標本をｎセット作成し，これらの標本から求めた確率水文量をもとに不偏推定値およびそのまわりの推定誤差を算定する手法である．一方，bootstrap法は大きさｎ個の標本から重複を許して任意にｎ個取り出した標本を複数作成し，これらの標本から求めた確率水文量をもとに不偏推定値およびそのまわりの推定誤差を算定する手法である．jackknife 法は計算回数が少なく，作成する標本数および不偏推定値，推定誤差が一意的に定まるのに対し，bootstrap 法は作成する標本数が任意に設定でき，作成する標本数によって不偏推定値や推定誤差が異なる．このため，今回の検討における確率分布モデル相互の安定性評価の指標としてはjackknife法を用いた．安定性の評価は確率水文量の推定誤差の大きさで判断した．

図－5 はＤ川について，適合度のよいグンベル分布（ＳＬＳＣ＝０．０１８，実線）とＧＰＥｘｐ分布(SLSC=0.021，破線)のjackknife法による推定曲線であり，中央の線がjackknife 推定値，両外側の線はjackknife推定値の両側に推定誤差の幅を示している．この両外側の曲線は観測値をほとんど包絡するように引かれている．また，毎年値から推定される確率水文量と非毎年値から推定される確率水文量が非常によく一致していることがわかる．

図－6 はＣ川について適合度のよいLP3Rs 分布(SLSC=0.018)とGPExp 分布(SLSC=0.031)をあてはめたものである．全資料を用いた推定曲線（図中LP3rs）は再現期間20年より大きい範囲で非毎年値による推定値と乖離している．しかし，jackknife推定値（LP3RSJK）は非毎年値による推定値とほぼ一致している．このようにjackknife 法により確率水文量の偏りが補正されることがわかる．

また，Ｃ川とＤ川の低水路満杯流量はそれぞれ $2{,}500\mathrm{m^3/s}$ および $3{,}500\mathrm{m^3/s}$ であり，非毎年値のデータは，これより大きくなっている．したがって，非毎年値のデータはすべて高水敷を流れる規模の洪水となっている．一方，毎年値の方は低水路内だけで流れる出水も含まれていることになる．また，Ｄ



図－5　Ｄ川の毎年値および非毎年値にそれぞれグンベル分布およびGPExp 分布をあてはめたもののjackknife 推定値，推定誤差

川の毎年値には $3{,}000\mathrm{m^3/s} \sim 3{,}500\mathrm{m^3/s}$ 付近のデータが少なく，この周辺でデータが階段状にプロットされ，これが適合度を下げる一因にもなるが，Ｄ川の場合，非毎年値ではこういったことも回避できる．

このように，リサンプリング手法およびAMS とPDS を併せて評価することにより，偏りを補正した確率水文量の推定が可能となるばかりでなく，AMS 固有の問題も回避でき，確率水文量の信頼性を向上させることができる．また，はじめに述べたように，AMS が適用できない標本に対して，残された手法はPDSである．ここで用いた手法は，PDS のみしか使えない標本に対しても確率水文量を精度よく推定する有力な手段となろう．

５．結論

全国の１級水系の79地点の流量資料について，適合度と安定性から評価するとともに毎年値と非毎年値の両面から評価することにより以下のことがわかった．

（1）非毎年値は満足すべき適合度の基準としてSLSC ≒ 0.04 をとると適合度を満足しなくなる標本が少なからずある．

(2)非毎年値で評価することにより，洪水規模の均

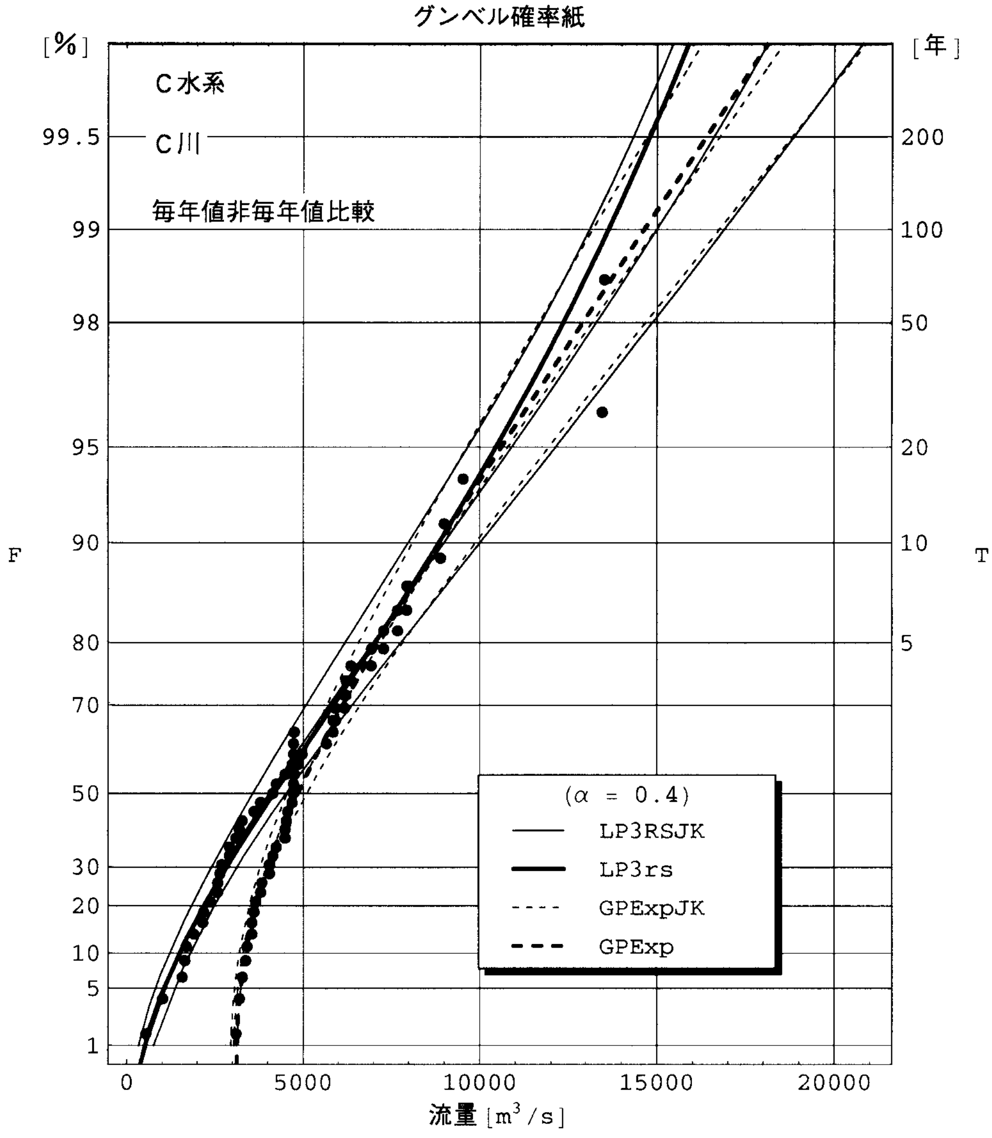


図-6 C川の毎年値および非毎年値にそれぞれLP3RsおよびGPExp分布をあてはめたもののjackknife推定値，推定誤差

質性が確保されるとともに，水位流量曲線の接続点周辺のデータのバラツキ等の問題を回避できる場合がある．

(3)適合度評価，リサンプリング手法と併せて毎年値と非毎年値を併せて評価することにより，確率水文量の信頼性を向上させることができる．

(4)観測期間が連続しない資料について，非毎年値の頻度解析が有力な手段となり得る．

謝辞：今回の検討にあたって，建設省には貴重な資料を提供いただいた．ここに記して感謝の意を表します．

参考文献

1) Stedinger, J. R., Vogel, R. M. and Foufoula-Georgiou, E. : Frequency Analysis of Extreme Events, Chap. 18, Handbook of Hydrology, (Ed.) D. R. Maidment, McGraw-Hill, New York, pp.18.1-18.66, 1993.

2) 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，第393号/II-9，pp.151-160，1988.

3) 江藤剛治，室田　明，米谷恒春，木下武雄：大雨の頻度，土木学会論文集，第369号/II-5，pp.165-174，1986.

4) 角屋　睦：水文統計論，水工学シリーズ，土木学会水理委員会，64-02，59pp.，1964.

5) 田中茂信，宝　馨：河川流量の頻度解析における適合度と安定性の評価，水工学論文集，第43巻，土木学会水理委員会，1999.

6) Efron, B. : The Jackknife, the bootstrap and Other Resampling Plans, SIAM Monograph, No.38, p.92, 1982.

(1998.9.30受付)