

両側有界分布の水文頻度解析への応用

APPLICATION OF PROBABILITY DISTRIBUTIONS WITH LOWER AND UPPER BOUNDS TO HYDROLOGIC FREQUENCY ANALYSIS

宝 馨¹・土佐 香織²
Kaoru TAKARA and Kaori TOSA

¹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 京都府宇治市五ヶ庄)

²学生会員 京都大学大学院工学研究科修士課程 (〒606-8501 京都府京都市左京区吉田本町)

This paper describes the applicability of probability distributions with both upper- and lower-bounds to hydrologic frequency analysis. Two distributions are applied to some data sets of extreme-value precipitation and river discharge: the Slade-type four-parameter log-normal distribution and the extreme-value distribution with upper- and lower-bounds (EVLUB) proposed by Kanda (1981). Their goodness of fit to the data sets is assessed in terms of the standard least-square criterion (SLSC). The analysis using the bootstrap method indicates that these distributions with the upper-bound give less variability in quantile estimates than the three-parameter log-normal distribution with the infinite upper-bound. Finally, this paper discusses incorporation of the probable maximum precipitation (PMP) or probable maximum floods (PMF) into the frequency analysis models as the upper bounds.

Key Words : *Upper- and lower-bounds, Probable maximum hydrologic quantities, Goodness-of-fit criterion, The bootstrap*

1. はじめに

種々の水工構造物の設計や計画においては、対象とする事象に関する変量の極端な実現値(極値データ)を用いて、当該事象の頻度解析を行う。その結果得られる確率水文量(T年確率降水量あるいは流量など)が設計・計画における基本量として用いられることになる。すなわち、当該の水文データを収集し、そのデータに水文学的および確率統計的観点から一定の評価を加えた後、その水文量の所定の再現期間に対応する規模を求めるという手順を踏む。この手順を水文頻度解析という¹⁾。

従来、水文頻度解析には対数正規分布、Gumbel分布、ガンマ分布など上限無限大の確率分布が多く用いられてきたが、これらは慣用的に用いられてきた側面もあり、必ずしも物理的根拠が明確なわけではない。著者らは、負の降水・流出量が存在するはずがなく、さらに物理的に可能な上限値が存在するはずであるという考えに基づき、上下限ともに有界の確率分布を頻度解析に適用してきた²⁾。そこで本研究では、水文学の分野においてまだ十

分に吟味されていない、上下限界を持つ確率分布を取り扱う。すなわち、Slade型の4母数対数正規分布と極値分布(EVLUB分布)である。これらを、従来の上限無限大の分布(3母数対数正規分布)と比較することにより、その特性を明らかにする。さらに、これらの両側有界分布を採用することにより、確率水文量の推定精度がどの程度向上するのかを例示する。

2. 水文頻度解析

(1) 水文頻度解析モデル

水文頻度解析に用いる確率分布を水文頻度解析モデルと呼ぶ。水文頻度解析モデルとしてよく用いられる極値分布は、上下限無限大のGumbel分布、下限値のみ有するFrechet分布、そして上限値のみ有するWeibull分布である。これら3つを統合した形式が、一般化極値分布(GEV)である。自然現象を考えると、両側有界の分布を考えるのが現象合理的である。ここでは、次の2つの分布を取り扱う。

a) 両側有界極値分布³⁾ (EVLUB分布—Extreme Value Distribution with Lower and Upper Bounds—)

$$F(x) = \exp \left[- \left\{ \frac{g-x}{v(x-a)} \right\}^{\kappa} \right] \quad (1)$$

$$f(x) = \frac{\kappa(g-x)^{\kappa-1}(g-a)}{v^{\kappa}(x-a)^{\kappa+1}} \times \exp \left[- \left\{ \frac{g-x}{v(x-a)} \right\}^{\kappa} \right] \quad (2)$$

を用いる。ここに、 $F(x)$ は変量の分布関数、 $f(x)$ は確率密度関数、 a 、 g はそれぞれ変量の下限および上限母数、 v は尺度母数、 κ は形状母数である。この分布は上記3つの極値分布から経験的に求められ、神田(1981)によって地震動や風速の最大荷重強度を表すための確率分布モデルとして紹介された。この分布は水文学の分野において、まだその特性が吟味されていない。

b) Slade型4母数対数正規分布

$$f(x) = \frac{g-a}{(x-a)(g-x)\sigma_Y\sqrt{2\pi}} \times \exp \left[- \frac{1}{2} \left\{ \frac{\ln\{(x-a)/(g-x)\} - \mu_Y}{\sigma_Y} \right\}^2 \right] \quad (3)$$

この分布は、岩井(1949)⁴⁾によって紹介された。

さらに、慣用されている3母数対数正規分布(上限無限大)：

$$f(x) = \frac{1}{(x-a)\sigma_Y\sqrt{2\pi}} \times \exp \left[- \frac{1}{2} \left\{ \frac{\ln(x-a) - \mu_Y}{\sigma_Y} \right\}^2 \right] \quad (4)$$

を取り上げ、比較に用いる。

(2) 上下限値の推定

下限値は、クオンタイル法を用いて推定する。推定された値が負であった場合、ゼロとする。本研究の趣旨からすると、下限値として何らかの物理的根拠のある値を用いるべきであり、とりあえずこの手法により非負の値を求めることにする。

さらに、上限値の推定値として、降水量のデータに関しては、日本で記録された1min~1yearの既往最大地点降水量を図上にプロットし、それらのプロット点を包絡するような線(降水量—継続時間曲線)を描くことにより統計的に推定したPMP(可能最大降水量; Probable Maximum Precipitation, 可能最大1日雨量1576mm, 同3

日雨量2988mm)⁵⁾を、流量に関しては、PMF(可能最大洪水; Probable Maximum Flood)を用いる。ここでは便宜的に、国内外の流域におけるPMFを既往最大値の1.5倍、2倍とした。今後、流量についてもより物理的根拠の明白な上限値を用いる必要がある。

3. 適合度評価

まず、あてはめた分布モデルの確率密度関数とヒストグラムを同一のグラフ上に描き、その適合度を目視により評価する。さらに、適合度を定量的に評価してモデルの優劣を客観的に評価するのに、4種の評価規準(SLSC, MLL, AIC, COR)⁶⁾を用いる。

図-1 に大津市における74年間(1912-1985)の年最大1日降水量、図-2 に岐阜市における100年間(1893-1992)の年最大3日降水量、そして図-3、図-4 にSt. Mary川における59年間(1916-1974)の年最大日流量に各分布をあてはめた結果を示す。

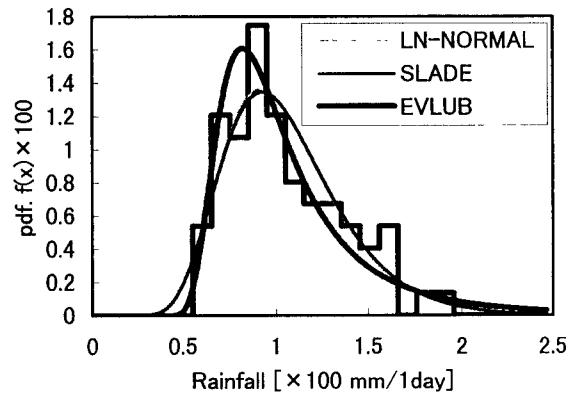


図-1 大津市における年最大1日降水量の確率密度関数とヒストグラム

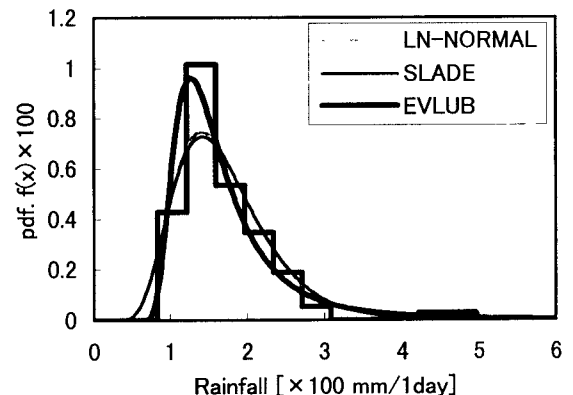


図-2 岐阜市年最大3日降水量の確率密度関数とヒストグラム

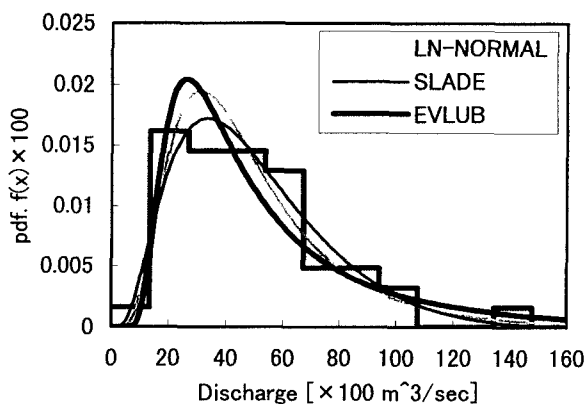


図-3 St. Marys川における年最大日流量の確率密度関数とヒストグラム(上限値=既往最大値×1.5)

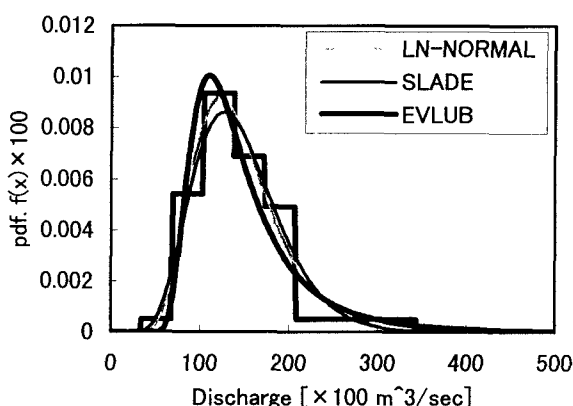


図-4 St. Marys川における年最大日流量の確率密度関数とヒストグラム(上限値=既往最大値×2)

さらに、各分布をいくつかの水文データに適用した場合のSLSCによる適合度評価の結果を表-1に示す。ほとんどの場合、EVLUB分布の適合度がSlade型4母数対数正規分布より悪い。しかし、岐阜市の年最大3日降水量やSt. Marys川の年最大日流量のデータを用いた場合には、EVLUB分布の適合度がSlade型4母数対数正規分布を上回っている。

4. 与える上下限値の違いによる影響評価

ここまで“水文頻度解析に物理的意味のある上下限値

を導入する”ことの意義を述べてきた。しかしながら，“上下限値”の推定法は数多く存在し、用いる分布モデルやデータセットとの組み合わせにより、その適合度は様々な影響を受ける。そこで、上・下限を様々な値に設定したときのEVLUB分布とSlade型4母数対数正規分布の適合度の違いを、彦根市の99年間(1894-1992)の年最大1日降水量データ(min.=40.8mm、max.=596.9mm)を用いて調べてみる。

(1) 与える上限値の違いによる影響評価

用いたデータの既往最大値が596.6mmであることを考慮して、下限値は一定のまま、上限値を順に、597, 600, 610, 625, 650, 700, 800, 1000, 1400mm/dayに設定する。そして、各場合に得られる確率密度関数をグラフ上に描き、さらに各種適合度評価規準の値を求める。そのうち、597, 650, 1400mm/dayとした場合のEVLUB分布とSlade型4母数対数正規分布グラフ(図-5)と、その際のあてはめた分布のSLSCの変化を図-6に示す。Slade型4母数対数正規分布は、上限値が小さくなる、つまり、上限値が既往最大値に近づくほど適合度が明らかに悪くなる。特に、700mm程度以下になると急激に悪くなっている。それに対し、EVLUB分布では、上限値が700mmに上側から近づくにつれて適合度が少しずつよくなり、700mmをこえると適合度が急激に悪くなる。

他のデータセットでの適用結果も同様であり、Slade型分布は、上限値を既往最大値に徐々に近づけていくと、極端に適合度が悪くなるという性質がある。

(2) 与える下限値の違いによる影響評価

(1)の場合とは逆に、上限値は一定のまま、下限値を順に0, 5, 10, 15, 20, 25, 30, 35, 40mm/dayに設定する。そして(1)の場合と同様に、グラフを作成し、適合度評価基準を求める。そのうち、0, 30, 40mm/dayとした場合のグラフ(図-7)と、その際のあてはめた分布のSLSCの変化を図-8に示す。上限値を変えたときは異なり、Slade型4母数対数正規分布は、下限値を0から30mm程度まで大きくしていくと、徐々に適合度がよくなる。35mm付近から急に適合度が悪化していく。EVLUB分布は、下限値が大きくなる、つまり、下限値が既往最小値に近づくほど適合度が悪くなる。

表-1 SLSCによる適合度の比較評価

	大津			彦根			今津			敦賀			岐阜			M川		S川		St. Marys		B.Ussury	
	1日	2日	3日	1日	2日	3日	1日	2日	3日	1日	2日	3日	1日	2日	3日	*1.5	*2	*1.5	*2	*1.5	*2	*1.5	*2
LN-NORMAL	○	○	◎	◎	○	○	△	◎	○	△	○	○	○	○	○	×	×	○	○	○	○	△	△
SLADE	○	○	◎	◎	○	○	△	◎	○	△	○	○	○	○	△	△	△	○	○	×	△	×	×
EVLUB	△	×	×	×	×	×	△	×	×	△	×	△	×	△	△	×	×	×	×	○	△	△	△

ただし、◎: SLSC≤0.02, ○: 0.02<SLSC≤0.03, △: 0.03<SLSC≤0.04, ×: SLSC>0.04

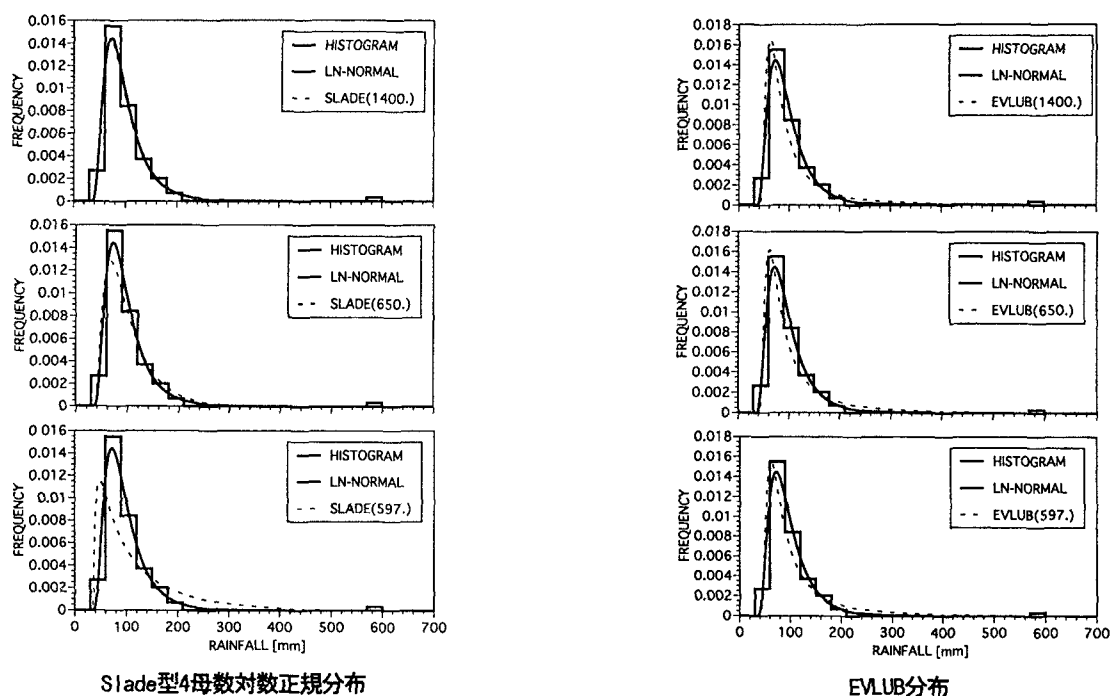


図-5 与える上限値の違いによる影響評価

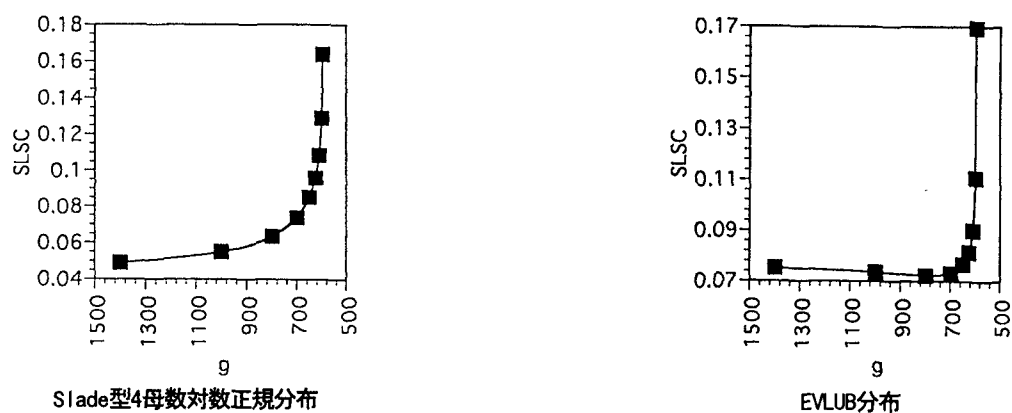


図-6 上限値の変化に伴うSLSCの変化

5. ひずみ係数と適合度の関係

右側に長い裾をもつヒストグラムは、右にひずんでいる、または正にひずんでいる分布といわれる。分布のひずみ具合を示す指標としてひずみ係数 (SKEW) :

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 / \hat{s}^3 \quad (5)$$

を用いる。ここに、 $\hat{s}$ は標準偏差の推定値を表す。どのような場合にEVLUB分布の適合度がSlade型4母数対数正規分布より良くなるかを調べるため、図-9 に様々なデータのひずみ係数とあてはめた分布のSLSCの関係を示した。

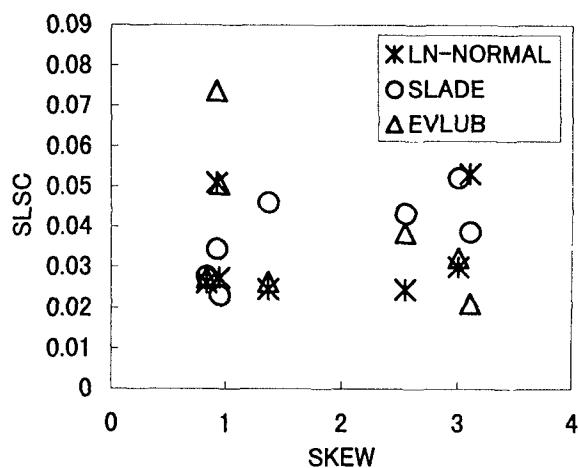


図-9 ひずみ係数とSLSCの関係

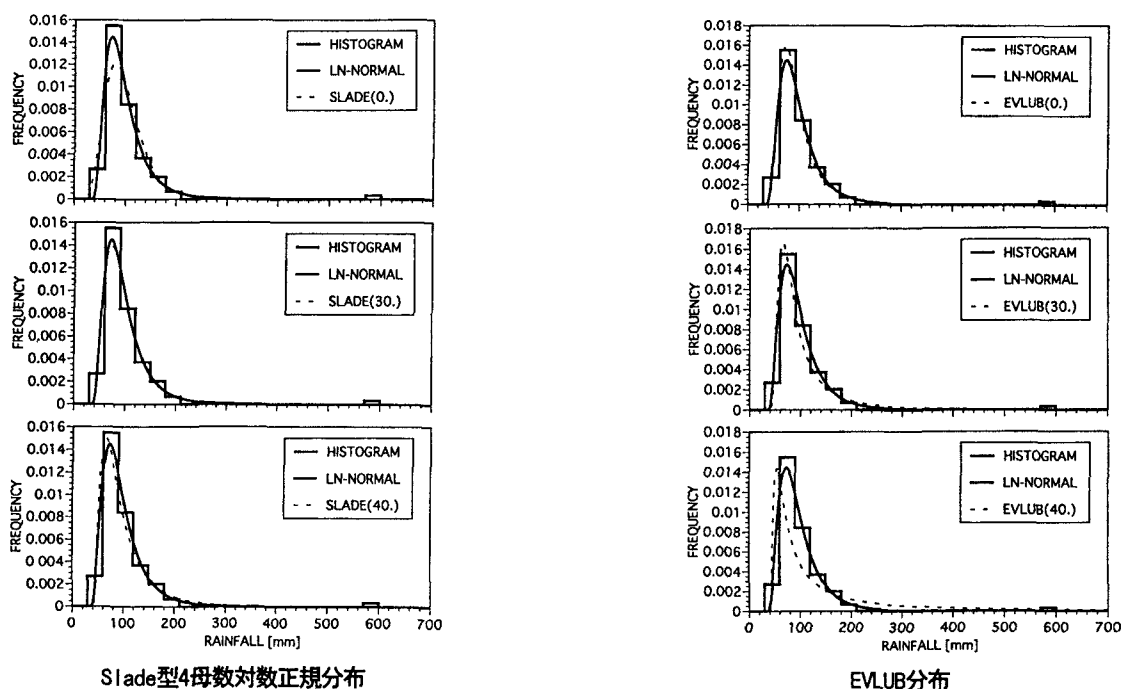


図-7 与える下限値の違いによる影響評価

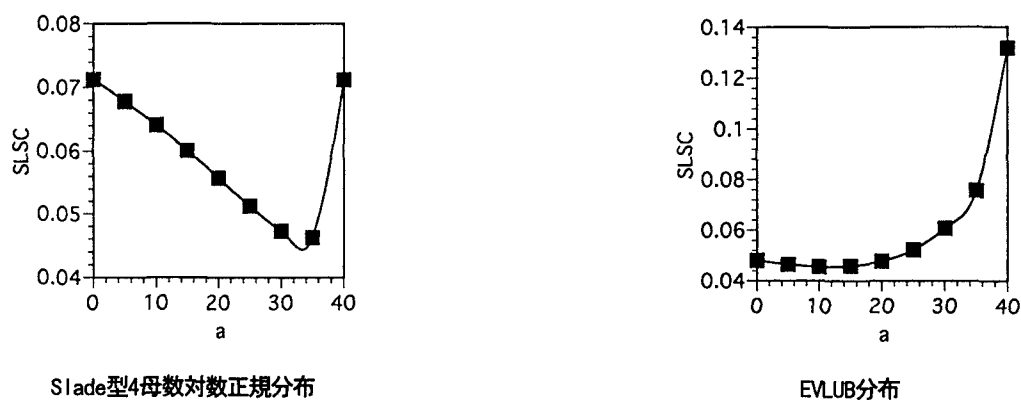


図-8 下限値の変化に伴うSLSCの変化

この図から、SLSCから見た適合の良い分布とひずみ係数の関係が見て取れる。すなわち、ひずみ係数が1より小さくなると、Slade型を用いる方が適合度が良くなり（SLSCが小さくなり）、ひずみ係数が1より大きいと、EVLUB分布の方がSlade型分布よりも適合度が良くなる傾向が読み取れる。

6. 確率水文量の変動性による評価

年最大値などの極値水文量を取り扱う場合、データと分布全体の適合度も重要であるが、分布の裾部分の形状や適合度がより重視される。したがって、データの蓄積が進んでも、確率水文量の推定値が大きく変動しないような確率分布モデルが実用上望ましい。こうした考え方

に基づき、確率水文量の変動性をモデル評価の基準とする。

確率水文量の変動性（推定精度）を調べるために、リサンプリング手法（本研究では bootstrap法）を適用する。リサンプリング手法とは、簡単にいうと、現在手元にある一組のデータセット（標本）から、部分的にデータを抽出したり、繰り返して元の標本と同じデータ個数だけ抽出したりという操作を反復して、多数のデータセットを作り出し、元の標本から得られる統計量の偏倚を補正したり、統計量の推定誤差を求めたりする手法である⁸⁾。bootstrap法とは、N個のデータを持つデータセットから重複を許してN個取り出すという、この操作を独立に多数回(B回)繰り返す。そして、得られたB個のbootstrap標本から求められる平均値をbootstrap推定値、分散をbootstrap誤差とする。本研究

では、この繰り返し回数Bを1500回とする。

両側有界の分布を採用することにより、外挿部（分布の両裾部）の不安定性の軽減、すなわち、確率水文量の推定精度の向上が期待できる。表-2 にはbootstrap法により求めた100年確率水文量の推定精度とその推定精度（bootstrap誤差 σ_B および変動係数 CV）を示した。

表-2 確率水文量のbootstrap推定値とbootstrap誤差、変動係数 (B=1500回)

100年確率流量 [m ³ /sec]		$\hat{x}_{100}$	σ_B	CV
B.Ussury	LN-NORMAL	5832	905	0.16
	SLADE	5167	710	0.14
	EVLUB	7918	730	0.09
S川	LN-NORMAL	15626	1988	0.19
	SLADE	12231	1006	0.09
	EVLUB	16496	990	0.06
St Marys	LN-NORMAL	29719	2688	0.09
	SLADE	27469	2173	0.08
	EVLUB	34035	2389	0.07
100年確率降水量 [mm]		$\hat{x}_{100}$	σ_B	CV
大津1日	LN-NORMAL	229.2	23.8	0.10
	SLADE	203.5	12.4	0.06
	EVLUB	287.1	24.6	0.09
岐阜3日	LN-NORMAL	429.7	62.7	0.15
	SLADE	402.4	42.7	0.11
	EVLUB	462.8	37.5	0.08

EVLUB分布を用いた場合には、他の分布を用いた場合と比較して、確率水文量の値が大きくなることがわかる。確率密度関数のグラフが右にひずむ傾向があることから考えても、EVLUB分布を用いた場合の確率水文量は、過大評価される可能性が高い。推定精度については、岐阜3日降水量のデータに対して、3母数対数正規分布の場合mmであったのが、両側有界のEVLUB分布を用いると37.5mmになる。つまり、bootstrap誤差でみると推定精度が40%向上している。Slade型4母数対数正規分布と比較しても12%向上している。S川のデータに対しても同様に、3母数対数正規分布の場合1988m³/secであったのが、両側有界のEVLUB分布を用いると990m³/secになり、bootstrap誤差でみると推定精度が50.2%も向上している。50, 200, 300年の場合にも同様の向上がみられる。

7. 結語

これまで水文頻度解析の分野であまり用いられてこなかった両側有界の確率分布、すなわちSlade型4母数対数正規分布とEVLUB分布を用い、それらがどのようなデータセットに有用であるかといった特性を明らかにできた。さらに、リサンプリング手法を用いてT年確率水文量の変動を定量的に評価し、上限値を導入した確率分布を用いると、T年確率水文量の変動が小さくなることを例証した。

上下限値を持つ分布は、物理的合理性を有するとともに、確率水文量推定精度を向上させることができる。すなわち、実用的合理性も併せ持つといえる。特に、小標本（データが少ない場合）に対して有望な方法である。今後の課題として上下限値の推定精度を上げる必要がある。PMPやPMFの推定手法について、さらに検討していきたいと考えている。

参考文献

- 1) 宝 馨：水資源システムにおける確率論的モデルと手法の評価に関する研究、京都大学学位論文, pp.13-14, 1989.
- 2) Takara, K., T. Takasao & K. Tomosugi: Possibility and necessity of paradigm shift in hydrologic frequency analysis, Proc. of Int'l Conf. on Water Resources and Environment Research, Kyoto, Japan, Vol. 1, pp. 435-442, 1996.
- 3) Kanda, J.: A New Extreme Value Distribution with Lower and Upper Limits for Earthquake Motions and Wind Speeds, *Theoretical and Applied Mechanics*, Vol. 31, University of Tokyo Press, pp. 351-354, 1981.
- 4) 岩井重久・石黒政儀：応用水文統計学, 森北出版, pp. 64-83, 1970.
- 5) 水文・水資源学会編：水文・水資源ハンドブック, 朝倉書店, p. 230, 1997.
- 6) 宝 馨・高埴琢馬：水文頻度解析における確率分布モデルの評価基準, 土木学会論文集, 第393号/II-9, pp. 151-160, 1988.; 宝・高埴, 合田良実の討議に対する解答, 土木学会論文集, 第405号/II-11, pp. 267-272, 1989.
- 7) 岩井重久・石黒政儀：応用水文統計学, 森北出版, p. 102, 1970.
- 8) 宝 馨：水資源システムにおける確率論的モデルと手法の評価に関する研究, 京都大学学位論文, pp. 68-69, 1989.

(1998. 9. 30受付)