水工学論文集，第43巻，1999年2月

誤差を含む歴史洪水データの確率洪水評価への導入シミュレーション

APPLICATION OF MONTE CARLO METHOD ON FLOOD FREQUENCY ANALYSIS USING HISTORICAL FLOOD DATA WITH ESTIMATION ERROR

庄 建治朗[1]・岩崎 誠一郎[2]・長尾 正志[3]・冨永 晃宏[4]
Kenjiro SHO, Seiichiro IWASAKI, Masashi NAGAO and Akihiro TOMINAGA

[1]正会員 工修 名古屋工業大学助手 工学部システムマネジメント工学科
（〒466-8555 名古屋市昭和区御器所町）
[2]学生会員 名古屋工業大学大学院 工学研究科社会開発工学専攻
[3]フェロー 工博 名古屋工業大学教授 工学部社会開発工学科（〒466-8555 名古屋市昭和区御器所町）
[4]正会員 工博 名古屋工業大学教授 工学部社会開発工学科（〒466-8555 名古屋市昭和区御器所町）

When the systematic observed hydrological record is not long enough, the historical flood information is useful to increase the precision of flood-quantile estimation. Historical flood data, which can be used in flood frequency analysis by adjusted-moment method or maximum likelihood method, often contain much larger error than systematic observed data. Thus it is important to evaluate the effect of the error on the precision of flood-quantile estimators. Monte Carlo simulation by using Gumbel distribution shows that flood-quantile estimator contains positive bias when the standard error exceeds a certain level (in this study, approximately 1/10 to 1/8 of the average for total sample). Moreover, when it exceeds 1/5 to 1/4 of the average, the improvement of the precision cannot be confirmed for the calculated flood-quantile estimators.

Key Words : *Flood frequency analysis, historical flood information, estimation error, Monte Carlo simulation*

1. はじめに

治水計画において計画基準年に対応する確率水文量を推定する際、過去の観測データを標本に流域の極値水文現象の従う確率分布を同定するが、標本数が十分でない場合には同定された分布の不確定性が大きくなるという問題が生ずる。その解決法の一つとして、観測時代以前の歴史洪水に関する資料の利用が考えられる。歴史洪水資料から水文量を再現する場合、値そのものが推定可能な場合ばかりでなく、ある値(閾値)を超えたかどうかのみが分かる場合、ある範囲内にあることのみが分かる場合など、限定的な情報しか得られない場合も多い。こうした不完全なデータを近年の観測データとともに確率分布モデルに導入する方法についてはこれまでにも研究がなされており、歴史洪水資料の利用による標本数の増加がモデルの信頼性向上に寄与することも Monte Carlo シミュレーションにより確かめられている[1),2),3),4)]。しかし、歴史時代の水文量は、地層や樹木年輪等に間接的な形で残されている洪水の痕跡や、多分に筆者の主観を含んだ古記録・古文書等から推定されるため、値そのものが推定可能な場合でもその推定値には観測データに比べ非常に大きな誤差が含まれている可能性があり、また閾値を超えたかどうかといった情報についてもその中には誤った情報が含まれている可能性も考えられる。歴史洪水資料の信頼性が低い場合、こうした誤差の大きいデータや誤った情報が多く含まれるため、標本数の増加に伴い却って同定された分布の信頼性が低下することも考えられるが、従来のシミュレーションではそうした影響は考慮されていない。本研究では、Monte Carlo シミュレーションに誤差発生の過程を取り入れることにより、歴史時代のデータの誤差が確率分布モデルの信頼性に及ぼす影響を評価し、どの程度までの誤差がモデルの信頼性向上に有効であるか、またその際の適切な標本数や母数推定法について検討を加えた。なお本研究では、極値水文量は Gumbel 分布に、歴史データの推定誤差は正規

分布に従うものとし、データには定常性、独立性を仮定している。

2. 歴史洪水資料の確率分布モデルへの導入法

近年の観測データとは異なり、歴史時代の洪水に関する情報は特に大規模な洪水が生起した年についてのみ得られるのが通常である。またその場合にも水文量が値として推定できる場合ばかりでなく、洪水がある一定規模以上であったということのみが分かる場合など、限定的な情報しか得られないことも多い。まず本章では、こうした特性を持つ歴史時代のデータを近年時代のデータとともに確率分布モデルに導入する手法について、概略を述べる。ここで近年時代とは、毎年の水文量が観測データによって正確に与えられている期間のことである。他方、一定規模(閾値)を超える洪水が生起した場合のみ水文量が推定可能であり、それ以外の年の水文量は閾値以下であるという情報しか得られない期間を歴史時代と定義する。歴史時代の期間中でも、時代の新旧により得られるデータの質・量が異なるため、実際には歴史時代をさらに細分化して閾値等の設定を変化させる必要があるが、ここでは標本数の増加に伴うモデルの信頼性の変化を見るのが目的であるので、単純化して歴史時代は単一の期間とした。

なお本研究では、極値水文量が従う分布として、洪水頻度分析において多く用いられる Gumbel 分布を仮定する。Gumbel 分布とは、確率密度関数及び分布関数がそれぞれ、

$$f(x) = \alpha \cdot \exp[\{-\alpha(x-\mu) - \exp\{-\alpha(x-\mu)\}]$$

$$F(x) = \exp[-\exp\{-\alpha(x-\mu)\}]$$

で与えられる極値分布である。母数μ, αは、平均値をm、標準偏差をσとすると、

$$\left.\begin{aligned} \mu &= m - 0.455005\sigma \\ \alpha &= 1.28255/\sigma \end{aligned}\right\} \qquad (2.1)$$

で与えられ、積率法や最尤法などにより推定される。

(1) 積率法による母数推定

近年時代の期間ををs年、水文量の値を$x_i(i=1,2,\cdots,s)$、歴史時代の期間をH年、閾値をUとする。また歴史時代について水文量が閾値を超えた回数をkとし、その値を$y_i(i=1,2,\cdots,s)$とする。

歴史時代において、水文量が閾値を超えない場合、それらの値は未知であるため、その平均値は近年時代において水文量が閾値を超えなかったときの平均値と同じと仮定し、さらに算定された平均値に対する分散も同じであると仮定する。

ここで近年時代のデータのうち、歴史時代の閾値を超えないものの数をn_{xy}とし、その値をx_{yi} $(i=1, 2,\cdots, n_{xy})$とすると、歴史時代での水文量の平均と分散は、

$$M_1 = \frac{1}{H}\cdot\left[(H-k)\frac{1}{n_{xy}}\cdot\sum_{i=1}^{n_{xy}} x_{yi} + \sum_{i=1}^{k} y_i\right],$$

$$\sigma_1^{\ 2} = \frac{1}{H_1}\cdot\left[(H-k)\frac{1}{n_{xy}}\cdot\sum_{i=1}^{n_{xy}}(x_{yi}-M)^2 + \sum_{i=1}^{k}(y_i-M)^2\right]$$

と表現できる。ここにMは全期間にわたっての平均値であり、

$$M = (sM_0 + HM_1)/(s+H). \qquad (2.2)$$

また全期間にわたっての分散は、

$$\sigma^2 = (s\sigma_0^{\ 2} + H\sigma_1^{\ 2})/(s+H) \qquad (2.3)$$

となる。但しM_0及びσ_0は近年時代のデータに対する平均値及び分散であり、

$$M_0 = \frac{1}{s}\sum_{i=1}^{s} x_i\,,\quad \sigma_0^{\ 2} = \frac{1}{s}\sum_{i=1}^{s}(x_i - M)^2\ .$$

式(2.2)、(2.3)を式(2.1)に代入すると母数μ, αが計算される。

(2) 最尤法による母数推定

一般に最尤法では次の尤度関数

$$L = \prod_{i=1}^{n} f(x_i;\theta)$$

を基礎として母数推定を行う。但し確率密度関数fの形は既知とする。データ$\{x_1, x_2, \cdots, x_n\}$が与えられたとき$L$は$\theta$のみの関数となり、全ての母数$\theta$に対して、

$$L(\hat{\theta}) > L(\theta)$$

が成立するような$\hat{\theta}$を最尤推定値という。いま母集団として Gumbel 分布を考えているので、母数はμとαの2つであり、

$$\frac{\partial}{\partial\theta_i}L(\theta_1,\theta_2) = 0 \qquad (i=1,2)$$

をθ_1, θ_2 (α及びμがそれぞれθ_1, θ_2に対応)について解くことによりそれらの最尤推定量が求められる。

以上が一般的な最尤法による母数推定法であるが、歴史時代を含む水文データには、尤度関数として次の4通りの扱い方が考えられる。

a) 値そのものが既知である場合

近年時代のデータ全てと、歴史時代のデータのうち値そのものが推定可能なものがこの場合にあたる。このときの水文量の値を$z_i(i=1,2,\cdots,s)$、母数をθ_1, θ_2とすると、尤度関数は次のようになる。

$$L_1 = \prod_{i=1}^{s} f(z_i;\theta_1,\theta_2)$$

b) 閾値以下であることだけが分かっている場合

歴史時代において、洪水記録が存在しないあるいは信頼性の低い記録しかない場合、洪水規模がある一定以上

なら必ず記録が残されていると仮定し、その年の水文量はそのレベル（閾値）に達していないと考える。閾値 U_j 以下であることが分かっているデータの数を n_j(j=1, 2,⋯, k) とすると、尤度関数は次のようになる。

$$L_2 = \prod_{j=1}^{k} F(U_j;\theta_1,\theta_2)^{n_j}$$

c）閾値を超えたことだけが分かっている場合

閾値を T_j、閾値を超えたことが分かっているデータの数を m_j(j=1, 2,⋯, k) とすると尤度関数は次のようになる。

$$L_3 = \prod_{j=1}^{k} [1 - F(T_j;\theta_1,\theta_2)]^{m_j}$$

d）ある幅の範囲内にあることが分かっている場合

歴史時代のデータの中には、信頼性が低くある程度の誤差を考慮すべき場合や、ある上限値と下限値の間の値をとることだけが分かっている場合がある。上限値を Y_i、下限値を Z_i (i=1, 2,⋯, n) とすると、尤度関数は次のようになる。

$$L_4 = \prod_{i=1}^{n} [F(Y_i;\theta_1,\theta_2) - F(Z_i;\theta_1,\theta_2)]$$

歴史洪水資料を加味したときに得られるデータは、上述の 4 つのケースの組み合わせであると考えられる。従って、尤度関数もこの 4 つを組み合わせて、

$$L = L_1 L_2 L_3 L_4$$

となる。

最尤法を用いることにより、上述の a）～ d）のそれぞれの種類の情報を取り込むことが可能である。また、こうした手法によって複数個の閾値を設定することが容易であるという利点があり、積率法よりも汎用性がある手法であるといえる。

3．シミュレーションによるモデルの検証

本章では、前章で提示した各種の母数推定法について Monte Carlo シミュレーションにより比較評価する。なお評価基準には平方根平均平方誤差 RMSE(Root Mean Square Error) を用いる。

(1) 母数推定モデル

本章では、最尤法を用いた場合と積率法を用いた場合について、確率分布モデルの母数の推定およびそれによって算定される適当な再現期間に相当する確率水文量の推定に関して、その再現効果について比較・検討する。具体的には、歴史時代の情報の与え方によって以下に示す 4 つのモデルを考える。

a）**最尤法 Case1(MLE1)**

近年時代 k_0 年間の定量データ x_{0i}(i=1, 2,⋯, k_0) の他、歴史時代 h_1 年間で、ある閾値 U を超えた場合のみ、データ値 x_{1i} (i=1, 2,⋯, k_1) そのものが既知であると考える。このときの尤度関数は次のようになる。

$$L = \prod_{i=1}^{k_0} f(x_{0i}) \cdot \prod_{i=1}^{k_1} f(x_{1i}) \cdot F(U)^{h_1 - k_1}$$

b）**最尤法 Case2(MLE2)**

近年時代 k_0 年間のデータ x_{0i} (i=1, 2,⋯, k_0) の他、歴史時代 h_1 年間で、閾値 U を超えた場合のみ、データ値 x_{1i} (i=1, 2,⋯, k_1) を中心に±R の幅を持つ範囲内に存在すると仮定する。これは情報として与えられた値をある幅の範囲内に存在すると考え、情報が与えられない場合は閾値以下としたモデルである。このときの尤度関数は次のようになる。

$$L = \prod_{i=1}^{k_0} f(x_{0i}) \cdot \prod_{i=1}^{k_1} [F(x_{1i} + R) - F(x_{1i} - R)] \cdot F(U)^{h_1 - k_1}$$

c）**最尤法 Case3(MLE3)**

近年時代 k_0 年間のデータ x_{0i} (i=1, 2,⋯, k_0) の他、歴史時代 h_1 年間で、閾値 U を超えた場合のみ、データ値そのものはわからない、あるいは非常に信頼性が低いため値そのものは無視するが、閾値を超えた回数が k_1 回であることのみが分かっていると仮定する。これは情報として与えられた値をある閾値以上であると考え、情報が与えられない場合は閾値以下としたモデルである。このときの尤度関数は次のようになる。

$$L = \prod_{i=1}^{k_0} f(x_{0i}) \cdot F(U)^{h_1 - k_1} \cdot [1 - F(U)]^{k_1}$$

d）**積率法(MOM)**

この場合は前章で提示したモデルをそのまま適用することとする。

(2) Monte Carloシミュレーション

前出 a）～ d）のそれぞれのモデルについて、誤差を含む歴史時代のデータを用いることによる母数および確率水文量の推定精度への影響を以下の手順により評価・検討する。

Step1：母集団を想定する。母集団の確率分布と母数を仮定する。

Step2：母集団から大きさ N の標本を抽出する。仮定した分布にしたがう乱数を N 個発生させる。

Step3：誤差を発生させる。即ち仮定した誤差の分布に従う乱数を歴史時代の年数と同じだけ発生させ、Step2 で発生させた乱数のうち歴史時代分のものに加え合わせる。近年時代については Step2 で発生させた乱数をそのまま使用する。

Step4：Step3 で得た標本に対して、それぞれの推定法により母数を推定し、確率水文量を算定する。

Step5：step2～4 を M 回繰り返し、それぞれの推定法によって推定した母数及び確率水文量の平均値と RMSE を算定する。

各々の Step について、本研究では琵琶湖の洪水頻度

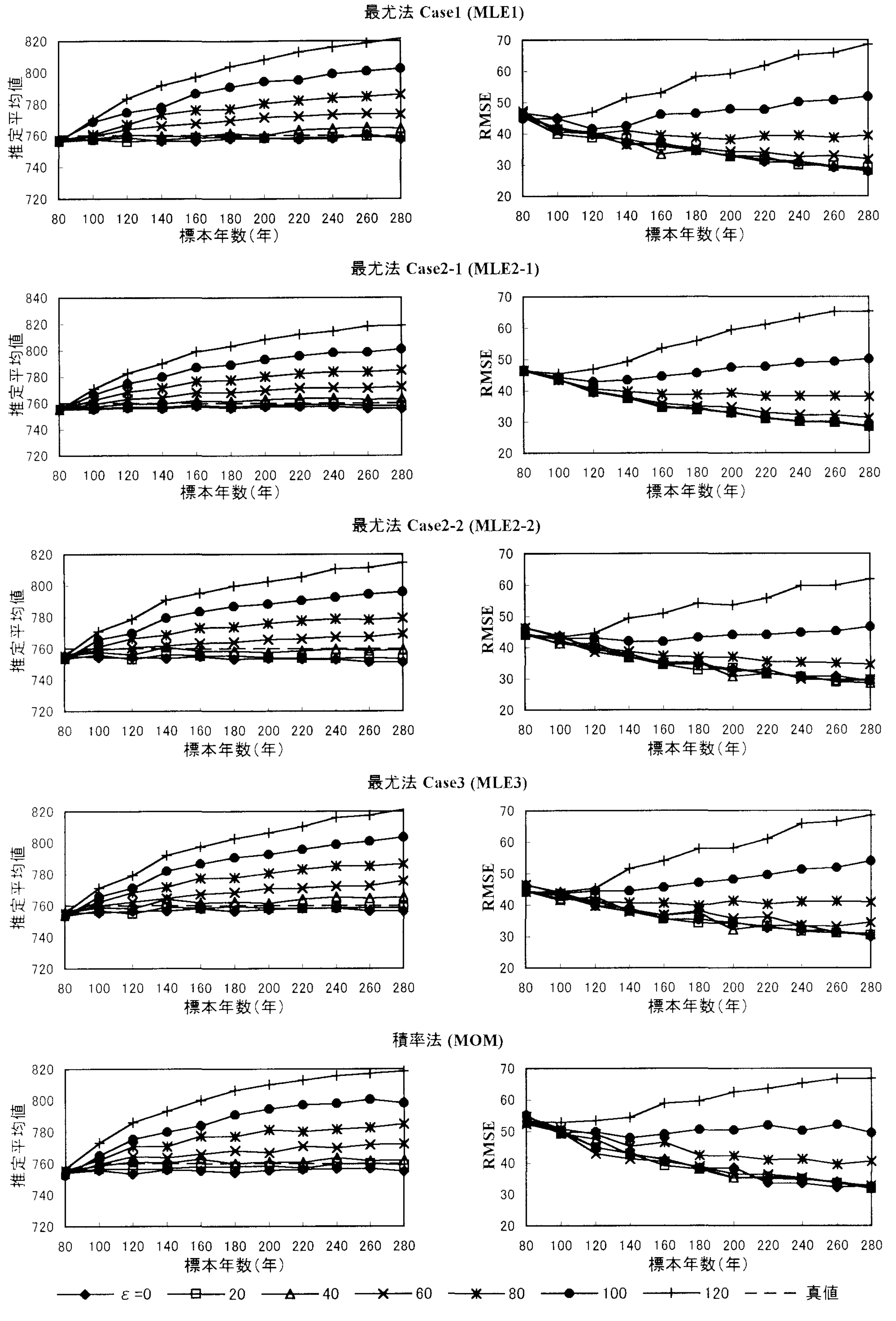

図-1　シミュレーション計算による100年確率水文量の推定平均値（左）とRMSE（右）

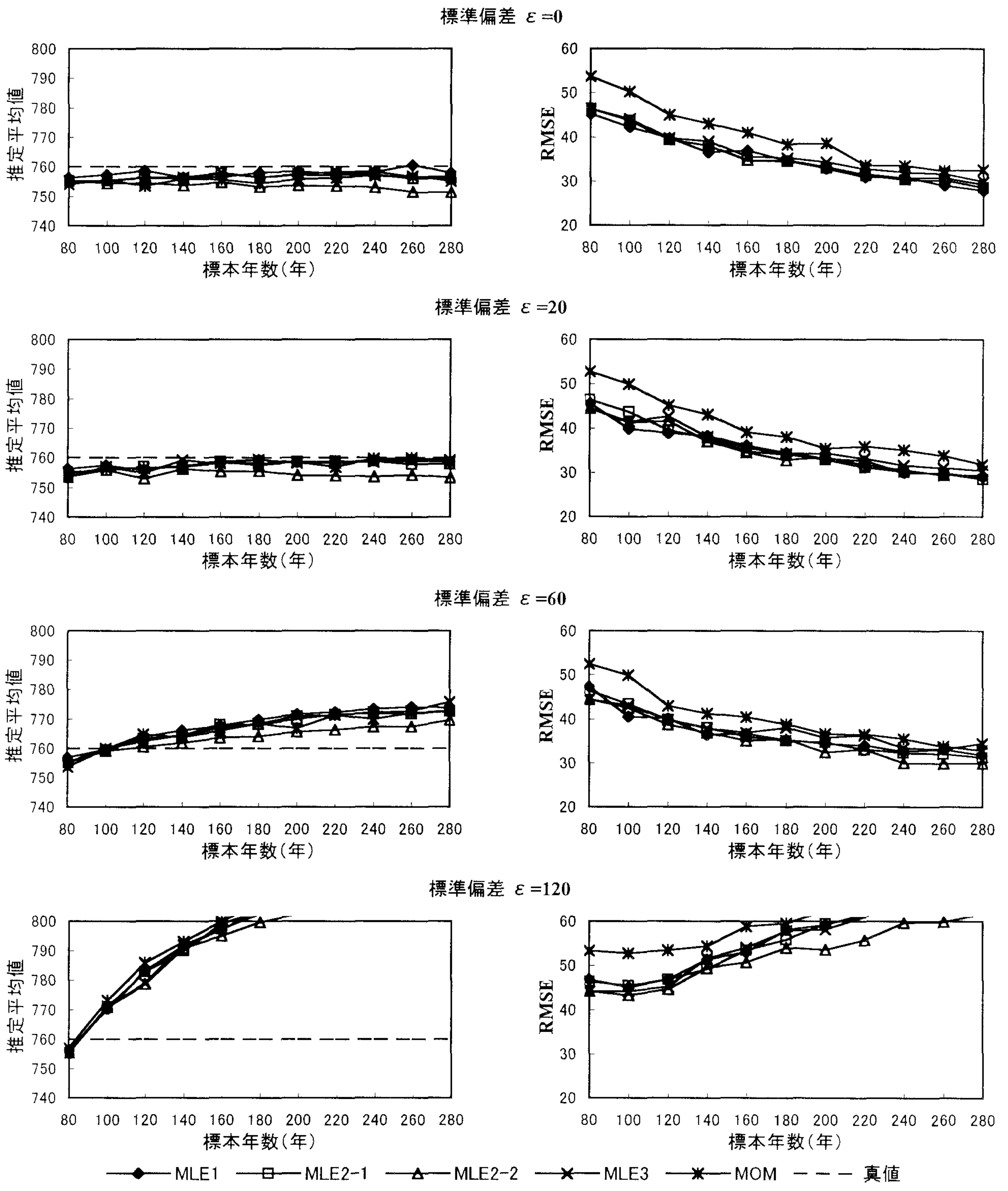

図-2　シミュレーション計算による100年確率水文量の推定平均値（左）とRMSE（右）（誤差ごとに比較）

分析を想定し、以下のような条件設定で計算を行う。

Step1

本研究では、母集団は Gumbel 分布に従うと想定している。母数は任意に設定してもよいが、年最大琵琶湖流域平均 30 日雨量の近年データから得た実際の母数に近い値として、本研究ではμ=300, α=0.01 と設定する。

Step2

本研究の主旨は、近年時代の標本に誤差を含む歴史時代のデータを付加していくことにより、母数及び確率水文量の推定精度がどのように変化するかを調べることにある。琵琶湖流域において各種の水文観測データが整備されるのが 1912 年であること、歴史洪水に関してある程度の質・量を伴う情報が得られるようになるのがほぼ

18 世紀初頭以降であることを考慮し、ここでは近年時代を 80 年間、歴史時代を 200 年間と設定する。故に標本数Nは、80 から順次増加させ最大 280 とする。

標本は、母数μ=300, α=0.01 をもつ Gumbel 分布に従う乱数発生によって作成する。

Step3

Step2 で発生させた乱数のうち、まず近年時代分 80 個については誤差を付加せずこのまま母数推定に用いる。次に歴史データの推定誤差は、本研究では平均値 0、標準偏差εの正規分布に従うと仮定する。即ち、歴史時代については、正規分布 $N(0, \varepsilon^2)$ に従う乱数を$(N-80)$個発生させ、これを Step2 で発生させた乱数に加えたものを母数推定に用いる標本とする。ここで標準偏差εは、様々な誤差が母数及び確率水文量の推定精度に及ぼす影響を見るため、0 から 120 まで 20 ごとに変化させた 7 通りのケースについて検討する。

Step4

Step3 で作成した標本を前節で提示した a) ～ d)のそれぞれのモデルに適用し、母数及び確率水文量の推定を行う。ここでは確率水文量の再現期間を 100 年とし、歴史時代の閾値は 500 と設定する(このとき閾値を超える標本の数は全標本数のおよそ 1/8 となる)。また、b) の最尤法 Case2 におけるRの値は 20、60 の 2 つの場合を想定し、それぞれ最尤法 Case2-1(MLE2-1)、最尤法 Case2-2(MLE2-2) とする。

Step5

上述の Step で得られた計算値の真値に対する RMSE を算定するために、 Step2～4 の繰り返し計算を行う。この繰り返し計算の回数が少ないと正確な RMSE の比較ができないので、本研究では繰り返し回数を 1000 回とした。

4. 結果とまとめ

前節の Step に従って、各ケースについて母数μ,α及び 100 年確率水文量 $Q(100)$ の推定値の平均値と RMSE を算定した。そのうち $Q(100)$ について、標本数の増加に伴う平均値と RMSE の変化を様々な標準誤差εについて計算した結果を **図-1** に示す。また **図-2** は、ε=0, 20, 60, 120 の 4 通りについて、$Q(100)$ の平均値及び RMSE をモデル間で比較したものである。

図-1 及び **図-2** から見出されることをまとめると次のようになる。

1) ε=0～40 では標本数増加とともに着実に $Q(100)$ の推定精度は向上していく。従って歴史時代のデータは標準誤差が 40(即ち本研究の場合、Gumbel 分布の母数μ(これは分布の平均値に近い値であると考えてよい)の 1/8 程度)であることが望ましい。このとき、データは誤差を含まないデータとほぼ同様に考えることが可能である。
2) ε=60～80 程度では、はじめの 40～60 年分については歴史時代の標本を付加することにより RMSE は低下するが、それ以上標本数を増加させてもあまり改善は見られない。
3) εが 60 以上のとき、歴史時代の標本を加えるに伴い正の偏りが大きくなり、この傾向はεが大きくなるほど強くなる。誤差を含むデータを多く用いると確率水文量が過大に評価される傾向があるという点については留意する必要がある。
4) **図-2** のε=20, 60 のケースを見ると、εに応じた幅(R) を持たせた最尤法 Case2 モデルが他の推定法に比べて偏り、RMSE ともにやや小さくなっている。従って、およその誤差の大きさが推測できる場合には、それに応じた幅 (R) を持たせた最尤法 Case2 (MLE2) を適用するのが好ましい。
5) 標準誤差が 100 程度(母数μの 1/4 程度)以上になると、歴史時代の標本を付加するほどモデルの信頼性は低下する。このような場合には、近年時代のデータのみを用いた最尤法による推定が最も信頼性が高い。
6) 積率法 (MOM) は計算が容易であり短時間に解が得られるという利点はあるものの、推定精度は最尤法に比べ明らかに低い。また閾値を超えたかどうかという情報のみを用いる最尤法 Case3(MLE3) は他の最尤法のケースに比べ、RMSE が大きくなる。

以上で明らかなように、歴史洪水資料を現実の洪水頻度分析に利用可能にするためには、誤差を平均的なデータ値の 1/10～1/8 程度以下にすることが望ましい。歴史時代について質の高いデータを得るよう努める必要がある。また、本研究では歴史時代を単一としたり、誤差の分布として平均値 0 の正規分布を仮定するなど単純な条件の下で計算したが、今後はさらに現実の流域に即したシミュレーションも試みたい。

参考文献

1) Hosking, J. R. M. and Wallis, J. R.: Paleoflood hydrology and flood frequency analysis, *Water Resour Res.*, **22**(4), pp.543-550, 1986.
2) Stedinger, J. R. and Cohn, T. A.: Flood frequency analysis with historical and paleoflood information, *Water Resour. Res.*, **22**(5), pp.785-993, 1986.
3) Cohn, T. A.and Stedinger, J. R.: Use of historical information in a maximum-likelihood framework, *J. Hydrol.*, 96, pp.215-223, 1987.
4) 池淵周一・前田勝：歴史洪水資料を利用した計画降雨算定手法, 京都大学防災研究所年報, **34**B-2, pp.103-125, 1991.

(1998. 9. 30受付)