

流出モデル評価への情報量規準の導入について

高棹 琢馬・椎葉 充晴・宝 馨

EVALUATION OF RUNOFF MODELS BY AN INFORMATION CRITERION

By *Takuma TAKASAO, Michiharu SHIIBA and Kaoru TAKARA*

Synopsis

Physically-based runoff models are generally classified into two groups : distributed parameter model (DPM) and lumped parameter model (LPM). In this paper, we present a framework to evaluate LPM's, while a DPM is regarded as a prototype and used to simulate discharge data containing only random errors.

Akaike's information criterion (AIC) is incorporated into the evaluation framework ; so that our method can be used to determine the order of a hierarchical model and to compare various types of models.

Tank cascade model (TC), Sugawara-type tank model (TANK) and Kimura's storage function model (KSFM) are evaluated for several storm-flood data simulated. Though KSFM gives minimum AIC values for each data, the values of the model parameters vary with the flood.

The framework presented here is important to evaluate LPM's and establish a proper LPM.

1. 序 論

河道の設計や洪水防御施設の規模決定などの河川計画,あるいは洪水予報やダム操作,さらには水資源の計画・管理などにおいて,降水の流出計算はその基本となる。その計算の道具として,種々の流出モデルが提案されてきた。しかし,それらのうち「どのモデルが良いのか」というモデルの評価問題に対しては,明確な解答が与えられていないというのが現状である。

一口に流出モデルと言っても,そのモデルが考案された背景,用途,構造などまさに千差万別であり,このことがモデルの統一的な評価を困難にしてきた一つの要因なのである。そこで本研究では,面積数百 km²以下のいわゆる単一流域の洪水流出モデルを対象を限定し,その評価法を検討する。本論に入る前に,この章では,流出モデルの分類と評価に関する問題点を明らかにしておく。

1.1 洪水流出モデルの分類と主な用途

洪水流出モデルを分類する方法についてはいろいろな考え方がある¹⁾が,ここでは大雑把に次のように分類しておく。

洪水流出モデル	$\left\{ \begin{array}{l} \text{Black box モデル} \dots\dots\dots \text{単位図法, 時系列モデルなど} \\ \text{物理モデル} \left\{ \begin{array}{l} \text{集中モデル} \dots\dots\dots \text{貯留関数法, タンクモデルなど} \\ \text{分布モデル} \dots\dots\dots \text{Kinematic wave モデル} \end{array} \right. \end{array} \right.$

洪水流出は物理現象であるから,用いる目的が何であれ,物理的基礎をもつモデルが望ましい。物理モデルは,集中モデル (lumped parameter model) と分布モデル (distributed parameter model) とに大別でき,その用途はそれぞれ異なる。

分布モデルは、モデルを記述する独立変数として時刻の他に空間座標を含むもので、Kinematic wave 法がその例である。全流出システムをできるだけ忠実に表現しようとするものであるから、流出現象の本質的理解という点が秀れている。したがって、洪水流出のシミュレーションに有用である。反面、流域をある程度細分化して、それぞれについて斜面・河道の初期条件、境界条件、パラメータを与え雨水流を追跡してゆくこととなるので、計算が煩雑になる。

集中モデルは、モデルを記述する独立変数が時刻だけのもので、貯留関数法、タンクモデルなどがその例である。流域内の現象との結びつきに欠けるといった欠点がある反面、数理的取扱いが容易で、実時間予測や広い流域（数百～数千 km²）の流出計算において果たす役割は大きい。その用途の広さゆえ、多数の研究者や実務家が単純な構造のものから複雑なものまで、多種多様な集中モデルを提案しており、まさに「モデルの洪水」とも言える状況にある。

1.2 モデル評価問題が未解決である理由

前節で述べたように、物理モデルにも様々なものがあり、それらのモデルの評価問題は未解決である。その理由を以下に整理してみた。

(a) 従来、分布モデルと集中モデルを同列に並べて評価することもあったが、物理性の考慮の度合と用いる目的の異なるもの同士の比較そのものが無意味であったと言える。

(b) モデルの良否は、当該流域における実測の雨量・流量データを用いて判定されることが多いが、そのデータが量的・質的に不十分である。量的に不十分な例としては、当該流域のデータの不在・偏在が挙げられる。所要の規模・時間分布の降雨流出データが無いことにより、モデルの同定・検証が十分に行えないのである。質的に不十分な例として、データの精度、欠測値の存在およびその補充などにおける諸問題が考えられる。

(c) たとえ、質のよいデータがかなりの量だけ揃ったとしても、全く同一の流域条件のもとに得られたのではないから取扱いが難しい。すなわち、河川流域は、気温・植生・土壌などの季節的変化のほかに森林伐採・ダム建設・都市化や観測地点・方法の改変などにより経年的変化を示す。これらの季節的あるいは経年的変化は、データに当然反映しているはずであるから多数の洪水データを一概に取扱うことには問題がある。

(d) モデルパラメータ値の求め方が、モデル構造に依存して様々な決定の仕方がなされているので既存のモデルをそのまま比較評価することができない。なるべく同一の条件のもとに同定されたもの同士でないと比較する意味があまりないのではないかと考えられる。近年、計算機の著しい能力の向上により、数学的最適化手法を用いたパラメータ同定が頻繁に行なわれるようになった。この場合、目的関数および探索法に何を選ぶかという問題があるし、パラメータ数が多くなると計算時間がかかるうえに、最適値の探索が不十分に終了することも生じる。

(e) 評価規準は何を採用すればよいのであろうか。従来は、実測流量と計算流量の適合度を主たる規準としてきたが、モデルの目的が何であるかによって（たとえば、ハイドログラフ全体の波形、最大流量、最大流量生起時刻など、何を合わせたかによって）、適合度の評価規準が異なってくるはずである²⁾。どの評価規準を用いるにせよ、実測データに含まれる誤差が大きいと適合度そのものの意味がないし、ただ単に合えばよいという観点からすればパラメータ数の多い複雑なモデルほど良いモデルであるということになりかねない。結局、出力の適合度は、モデル評価の十分な規準とは言えないのである。

1.3 洪水流出モデル評価の方針

物理的な流出モデルは、分布モデルと集中モデルに分類できるが、実際の流出過程との対応という点から見て、前者は後者より上位にあり（Fig. 1）、また、その用途は異なる。したがって、本研究では両者を区別し、集中モデルの評価問題を取扱う。実際、次から次へと新しいモデルが開発され、乱立状態にあるのは、

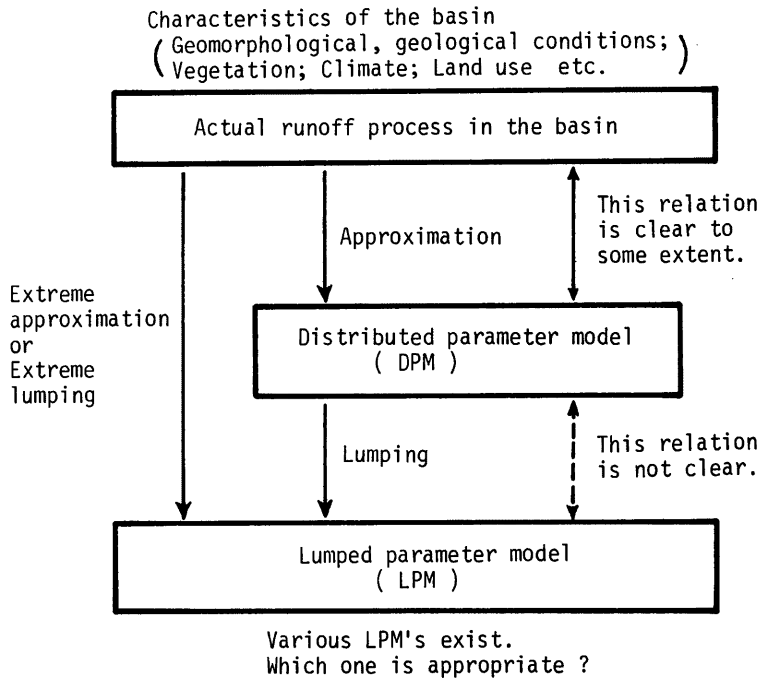


Fig. 1 Relation among actual runoff process, distributed parameter model and lumped parameter model

分布モデルではなく集中モデルの方なのである。

実測のデータをそのまま用いて、種々の集中モデルを評価することには、1.2で述べたように問題がある。そこで、実流域との対応がある程度明確な分布モデルを介して評価することとする。

すなわち、流域を分布モデルによって記述し、その分布モデルの挙動を最もよく表す集中モデルを見出すとするのである。分布モデルと集中モデルの対応の問題に限定すれば、実測データは不要となる。つまり、分布モデルに所要の規模・パタンの降雨を入力し、模擬流量を発生させて、このデータをモデル評価に用いる。このようにして、分布モデルの挙動を最もよく表す集中モデルを見出すことができたなら、それを当該流域の最良の集中モデルとする。そして、その集中モデルと実現象との乖離は、専ら分布モデルと実現象との乖離に帰されることとなる。(したがって、分布モデルによる流域の記述がどの程度の精度で可能か(つまり、分布モデルの評価)が新たな問題となってくるが、これについては別の機会に論じたい。)

分布モデルを介して(分布モデルを prototype として)集中モデルの評価を行なうという本研究の方針の利点は、実測のデータに含まれる諸々の系統的誤差、流域条件(初期条件も含む)の差異を排除できることに加え、同一の流域条件で入力条件を自由に変えられるので、種々の洪水をシミュレートすることができ、入力条件と流域特性と集中モデルの関連の把握が容易となる点である。

次章で詳しく述べるが、流出モデルは現象の物理性と不確定性を考慮できることが重要である。また、過度に構造の複雑なものは適当でない。

本研究では、こうした事情を評価できる規準として、情報量規準を導入し、モデルの適合度だけでなくモデルの簡潔さをも評価する手順のフレームワークを提示する。

2. 流出現象のモデル化

1.1において、流出モデルの分類を示したが、モデルの変換特性（変換が決定論的 (deterministic) か確率過程的 (stochastic) か）については特に言及しなかった。というのは、20年以上も前の高樺⁹⁾の指摘にもかかわらず、わが国における洪水の流出計算法は、そのような変換特性のことを特別意識することもなく、専ら決定論的に取扱われてきたからである。無論、従来の流出モデルが現象の平均的な挙動を表すために用いられてきたと解釈してもよいが、個々のモデルの確率特性（言いかえるとモデルの精度）というものについては何ら明らかにされてはこなかった。

最近10年程の間に、Kalman フィルター理論を応用した洪水予測手法（例えば、文献4),5)を見よ）が検討されるようになって、電気や自動制御の分野で用いられる確率過程的な状態空間表現による流出モデルの記述が行なわれるようになった。このとき、系の平均的挙動を従来の集中型流出モデルで表し、平均値0の確率モデルをノイズ項として付加する。このように流出モデルを stochastic に取扱うのは、Kalman フィルター等の手法を用いるための便法としてではない。流出現象そのものが本来 stochastic な力学現象であるのだという基本的認識に立つべきこと、したがって、それに沿ったモデル化なり取扱いなりをするのであるということをここでもう一度強調しておこう。

さて、一般に我々の手元に得ることのできるシステム出力（流量）の観測値の原系列が

$$Y^0 = \{y_0^0, y_1^0, y_2^0, \dots\} \dots\dots\dots(1)$$

で与えられているとする。この原系列は、ある種の系統的誤差 (systematic error, 例えば、河床変動等の河川条件の変化や、観測 (推定) 方法に起因するもの) を含んでいるのが普通である。次に、このような系統的誤差を原系列から取り除いた値 (非系統的観測値と呼んでおく) の系列を

$$Y = \{y_0, y_1, y_2, \dots\} \dots\dots\dots(2)$$

とおくことにする。この系列が、真の現象によって得られる出力を表すと考えられる。時刻 k におけるこの出力 $y_k (k=0, 1, \dots)$ のうち、平均的現象が m 個の状態量 $x_1(k), \dots, x_m(k)$ と n 個のパラメタ $\theta_1, \dots, \theta_n$ で記述でき、確率変動成分を ϵ_k (ϵ_k と $\epsilon_{k+j}, j = \pm 1, 2, \dots$ は独立) で記述できるとすれば、

$$y_k = f(x_1(k), \dots, x_m(k); \theta_1, \dots, \theta_n) + \epsilon_k \dots\dots\dots(3)$$

ここに、関数 $f(\cdot)$ は、真の平均的現象を与えるものである。実際には、 $f(\cdot)$ の完全な記述は不可能であり、何らかのモデル化が必要となる。そこで、モデル i の状態量の数を $m(i)$ 、パラメタ数を $n(i)$ として、次のようにモデル化する。

$$y_k = g_i(x_1(k), \dots, x_{m(i)}(k); \theta_1, \dots, \theta_{n(i)}) + e_k^i \dots\dots\dots(4)$$

ここに、 $g_i(\cdot)$ はモデル i の関数形で系の平均的挙動を表し、 e_k^i は平均からの偏差を表す。ただし、 $E\{e_k^i\} = 0$ である。

本研究では、分布モデルを prototype として、(3)式 $f(\cdot)$ に相当するものとし、それをモデル化 (集中化) したものが集中モデル $g_i(\cdot)$ であるとする。すなわち、対象流域を分布モデルによって記述し、所要の規模・時間分布の降雨を与えて流出計算を行ない、これに確率変動 (乱数発生法による) を付加して得られた出力を系統的誤差を含まない観測データとする。こうして得た入出力関係を最も良く表す集中モデルを見出そうとするのである。

3. 情報量規準を用いた流出モデル評価の方法

“良いモデル”であれば、(非系統的な) 観測値とそのモデルを用いた計算値の適合度は良いはずである。逆に、適合度が良いからと言って、そのモデルが“良い”とは断定できない。つまり、出力の適合度は、“良いモデル”の必要条件であって十分条件ではないのである。実際、black box 的アプローチによってもある程度の適合度は得られる。これは、降雨・流出という一組の入出力の対応が外見上鮮明で単純であるからに他ならない。

では、何をもってモデルが“良い”と言えるのであろうか。

3.1 尤度——確率モデルの適合度

stochastic な力学事象の実現値として得られた系列 Y ((2)式) が与えられたとき、その系列とモデルによる計算値との適合度は尤度 (likelihood) によって測られる。

互いに独立な M 個の確率変数 (Y_1, \dots, Y_M) の同時確率密度関数が $P(y_1, \dots, y_M|\theta)$ で与えられているものとする。 θ は密度関数を規定するパラメタのベクトルである。このとき、実現値 (y_1, \dots, y_M) は与えられたものとして固定し、 P を θ の関数と考えるとき、これを尤度関数と呼び $L(\theta)$ で表す。すなわち、

$$L(\theta) = P(y_1, \dots, y_M|\theta) \dots\dots\dots (5)$$

である。 Y_1, \dots, Y_M が独立であれば、

$$L(\theta) = p_1(y_1|\theta) \dots p_M(y_M|\theta) \dots\dots\dots (6)$$

となる。ここに、 $P_k(y_k|\theta)$ は、第 k 番目の実現値 y_k が得られるときの確率密度関数 (probability distribu-

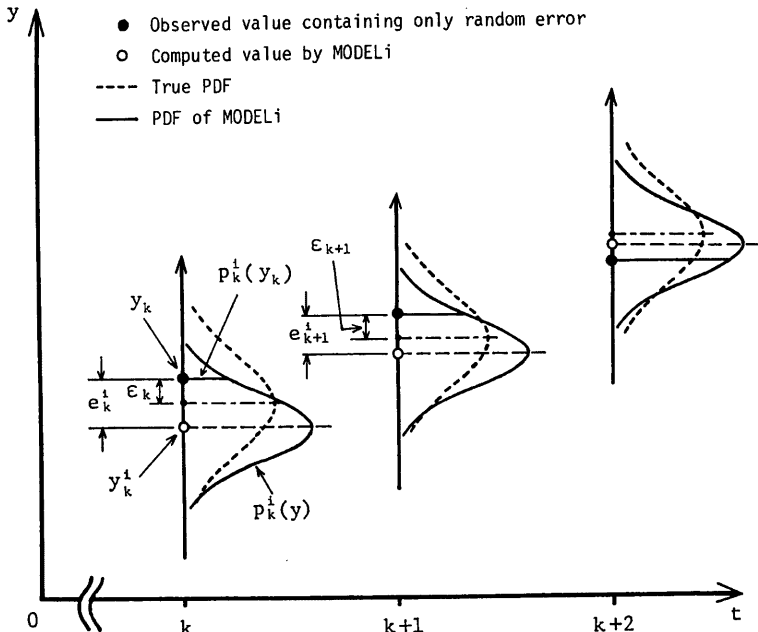


Fig. 2 Stochastic aspect of the observed and computed data

tion function, PDF) である。(3)式で与えられる y_k は、平均値 $f(x_1(k), \dots, x_m(k); \theta_1, \dots, \theta_n)$ で、その確率分布は ε_k の確率分布に従い、 ε_k は時間的に独立である。したがって、いま我々が扱おうとしている y_k の系列 Y に対しても(6)式の形で尤度を取扱うことができる。

(6)式の両辺の対数をとると、

$$l(\theta) = \log L(\theta) = \sum_{k=1}^M \log P_k(y_k|\theta) \quad \dots\dots\dots(7)$$

となり、これは対数尤度と呼ばれる。

尤度あるいは対数尤度を最大とするようにパラメタ θ を選ぶことによってモデルを同定する方法を最尤法 (maximum likelihood method) と呼ぶ。また、最尤法で求められた θ の推定量 $\hat{\theta}$ は最尤推定量と呼ばれ、この最尤推定量によって定められるモデルを最尤モデル、最尤モデルの(対数)尤度を最大(対数)尤度という。

(3)式で与えられる真の分布を(4)式のモデル i で近似する場合、モデル i による計算値 y_k^i は、

$$y_k^i = g^i(x_1(k), \dots, x_m(i)(k); \hat{\theta}_1, \dots, \hat{\theta}_{n(i)}) \quad \dots\dots\dots(8)$$

この計算値の系列を

$$Y^i = \{y_1^i, \dots, y_M^i\} \quad \dots\dots\dots(9)$$

とする。このとき、最大尤度 $L(\hat{\theta})$ を与える $P_k^i(y_k) = P_k(y_k|\hat{\theta})$, $k=1, \dots, M$ と Y, Y^i との関係を図示すると Fig. 2 のようになる。尤度が、(4)式のような確率モデル(あるいは統計モデル)の適合度となることは容易に理解されよう。

3.2 情報量規準

1.2(e)で述べたように、データとモデルの適合度は有効な評価規準とはならない。したがって、(4)式のような確率モデルの場合の尤度も同様である。尤度を大きくするには、Fig. 2 の $t=k+2$ のような状況を多く出せばよいのであるから、一連のデータのランダムな変動にも追従していくようなモデルを作ればよい。一般には、パラメタ数を増やすことによって尤度そのものは大きくできるが、いわゆる over fitting となる。

赤池⁶⁾は、確率モデルの評価規準として、次式で定義される情報量規準 (Akaike's information criterion, AIC) を提案した。

$$AIC = -2 \log(\text{最大尤度}) + 2(\text{自由パラメタ数}) \quad \dots\dots\dots(10)$$

尤度を大きくすれば AIC の値は小さくなり、自由パラメタ数を増やせば AIC は大きくなる。赤池の方法は、AIC を最小とするモデルが最も良いとするもので、適合度の向上とモデルの簡略化というトレードオフ関係の中で最適のものを見出そうとするものである。

AIC の誘導過程や適用例は、文献(7), (8), (9)に詳しい。水文流出系への適用としては、例えば、Katayama¹⁰⁾, McLead and Hipel¹¹⁾, 尾崎^{12), 13)}などがあり、いずれも時系列モデルの次数決定に AIC を用いている。

データが与えられたとき、(4)式のモデルの最尤法による同定が実現できれば、AIC は直ちに適用できるが、データ数 M とモデルの自由パラメタ数に留意する必要がある。

坂元⁹⁾は、AIC 利用上の注意事項として、データにあてはめるモデルの自由パラメタ数はデータ数を M として $2\sqrt{M}$ までにすべきだとしている。この条件を満たしていても、 M が小さい場合には AIC の適用は不適当である。というのは、AIC は M が大きいとして誘導されるからである。また、最大尤度を求める必要性から、(4)式のモデル構造 ($g_i(\cdot)$ の関数形および ε_k^i の確率分布) が複雑であれば、尤度の計算がかなり煩雑になる。

3.3 最尤法による流出モデルのパラメタ同定

一般に、流出モデルのパラメタ値は、計算値とデータとが適合するように決められる。物理モデルでは、

パラメタのうちで流域の物理的条件との対応が明確なものについては、その条件に応じた値を予め与え固定する。このとき、そのパラメタは自由パラメタではない。他のパラメタは自由パラメタであり、自由パラメタの値を動かして（調整して）所要の適合度を得ようとするのである。

その動かし方は、人間の手による試行錯誤的方法や図解法が用いられる場合が多い。このような場合、目視によって実測ハイドログラフと計算ハイドログラフの適合度を判定する。すなわち、解析者の主観的判断が介入してくる点に問題がある。

一方、ある適合度の評価規準（目的関数；たとえば誤差二乗和）を設定し、数学的最適化手法を用いてその規準を最小化（あるいは最大化）するパラメタの組合せを求める方法もある。これは、決められたアルゴリズムに従って、人間ではなくコンピュータが試行錯誤しながらパラメタ空間内の最適点を探索してゆくものである。したがって、目的関数と探索方法が異なれば、同じデータに対して異なる最適点が求まるのが普通である。目的関数と探索法が同一であっても、探索開始点の与え方によって異なる結果を得る場合もある。パラメタ空間内で目的関数は一般に多峰性をもつので、領域内をくまなく探索し、常に最適点を見出すことができる方法が望ましい。

さて、(4)式のような確率モデルを最尤法で同定する場合、目的関数は(7)式で与えられ、これを最大化する。(4)式の e_k^i が平均値 0、分散 σ^2 の正規分布に従うとすれば、

$$E\{y_k\} = g_k^i(x_1(k), \dots, x_{m(i)}(k); \theta_1, \dots, \theta_{n(i)}) \quad \dots\dots\dots(11)$$

で、

$$y_k^i \sim N(g_k^i(\cdot), \sigma^2) \quad \dots\dots\dots(12)$$

となる。このとき、 $y_k = g_k^i(\cdot) + e_k^i$ の確率密度関数は、

$$P_k^i(y_k) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\{y_k - g_k^i(\cdot)\}^2\right] \quad \dots\dots\dots(13)$$

となるので、尤度は

$$\begin{aligned} L(\theta, \sigma^2) &= \prod_{k=1}^M P_k^i(y_k) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^M \{y_k - g_k^i(\cdot)\}^2\right] \quad \dots\dots\dots(14) \end{aligned}$$

で与えられ、対数尤度は、次式で与えられる。

$$\ln(\theta, \sigma^2) = -\frac{M}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^M \{y_k - g_k^i(\cdot)\}^2 \quad \dots\dots\dots(15)$$

(15)式が最大となるようなパラメタ推定値 $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_{n(i)})$, $\hat{\sigma}^2$ を求めるのが最尤法である。 σ^2 が既知であれば、(15)式を最大化するには

$$\sum_{k=1}^M \{y_k - g_k^i(\cdot)\}^2 \rightarrow \min \quad \dots\dots\dots(16)$$

とすればよく、これが目的関数となる。 σ^2 は一般に未知であるが、この場合でも、同様にして $\hat{\theta}$ を求めた後、対数尤度を最大にするように $\hat{\sigma}^2$ を求めることにより最尤モデルを見出すことができる。

結局、 e_k^i の正規性を仮定すれば、最尤法は最小二乗法と一致し、最大対数尤度は、

$$\ln(\hat{\theta}, \hat{\sigma}^2) = -\frac{M}{2} \log 2\pi\hat{\sigma}^2 - \frac{M}{2} \quad \dots\dots\dots(17)$$

となる。ここに、

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{k=1}^M \{y_k - g_k^i(x_1, \dots, x_{m(i)}; \hat{\theta}_1, \dots, \hat{\theta}_{n(i)})\}^2 \quad \dots\dots\dots(18)$$

モデルパラメタ数 $n(i)$ のときの情報量規準 AIC ($n(i)$) は(17)式を(10)式に代入して、

$$\begin{aligned} \text{AIC}(n(i)) &= -2l(\hat{\theta}, \hat{\sigma}^2) + 2(n(i)+1) \\ &= M \log \hat{\sigma}^2 + 2(n(i)+1) + M(\log 2\pi + 1) \dots\dots\dots(19) \end{aligned}$$

となる。

3.4 集中モデル評価の手順

本研究では、「prototype としての分布モデルの挙動を最もよく表す集中モデルは何か」という観点から、各種の集中モデルを AIC を導入して評価しようとする。その手順をまとめると以下のようになる (Fig. 3)。

- ① 流域を分布モデルによりモデル化する。
- ② 想定した規模・パタンの降雨 (誤差のない流域平均時間雨量系列とみなす) を①の分布モデルに入力

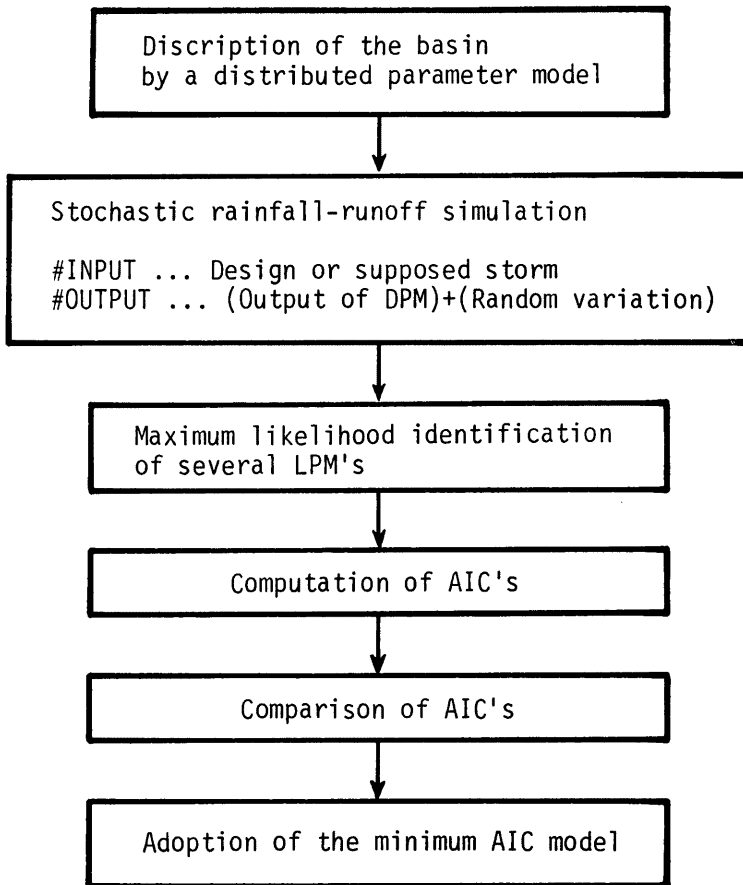


Fig. 3 Procedure of evaluating lumped parameter models

し流量データをシミュレートする。

- ③ ②で得られた流量データが観測流量の平均値に相当するとみなし、これらのデータに白色の確率変動（乱数発生により生成）を付加することによって、系統的誤差を取り除いた（言い換えれば、非系統的誤差のみを含む）観測流量をシミュレートする。
- ④ 以上のようにして得られた観測流量データと入力した雨量データを用いて、最尤法により各種集中モデルのパラメータを同定する。
- ⑤ 同定された種々の集中モデルの情報量規準（AIC）を算定し、AIC が最小のモデルを採択する。

4. 分布モデルによるデータの生成

流域で起こる真の流出現象の平均的挙動を分布モデルで記述できるものとし、真の現象の確率的挙動を何らかの確率モデルで代表させることにより、系統的誤差を取り除いた雨量・流量データを生成する。このとき、システム入力である降雨や流域の条件を設定して、種々のデータを生成することができる。

4.1 用いる分布モデルの概要

流域を斜面システムと河道システムに分け、降雨流出を2段階の力学モデルによって追跡する。すなわち、降雨はまず斜面システムモデルによって河道への横流入ハイドログラフに変換され、次に、河道システムモデルによって流域末端の河道流出に変換されるものとする。

1本の河道区分（セグメント）とその両側の矩形斜面を流域を構成する最小単位とし、河道網構造に応じてこれを組み合わせることにより、流域全体をモデル化する。河道区分毎・斜面毎にモデルパラメータは異なっておりよい。

このような分布モデルとして、ここでは高樺・椎葉のモデル¹⁴⁾を用いることとする。これは、Kinematic Wave 法を基礎とするもので、斜面ではA層と水みち（裸地域）とを考慮した地表面流・中間流モデル、河道は普通の Kinematic Wave モデルで構成される。文献¹⁴⁾では、斜面を矩形以外の形状とした場合の定式化も考慮しているが、ここでは矩形とする。

4.2 想定する流域条件

対象流域を分布モデルでモデル化することが、集中モデル評価の基本となる。実流域を対象とする場合には、この作業にかなりの労力を要する。本研究の主題は、特定の流域の分布モデルによるモデル化にあるのではなく、集中モデル評価のフレームワークを確立することにあるので、ここでは Fig. 4 に示すような2つの異なる河道網構造をもつ仮想流域を対象とした。Fig. 4 (a), (b) はそれぞれ、羽状流域 (featherlike basin, 以下 FEA と略記する)、平行流域 (parallel basin, PAL) と呼ばれる典型的な流域形状を想定している。それぞれ15本のセグメントからなり、セグメント長は 3 km、各セグメントの両側には斜面長 3 km の矩形斜面を有するものとし、250 km² 程度の面積を想定した。30個の斜面はすべて同一条件とし、斜面モデルの定数は、庄内川猪子石流域で高樺・椎葉が求めた値¹⁴⁾をそのまま用いた。河道モデルも同様である。

分布モデルとして高樺・椎葉のモデルを用いたこと、

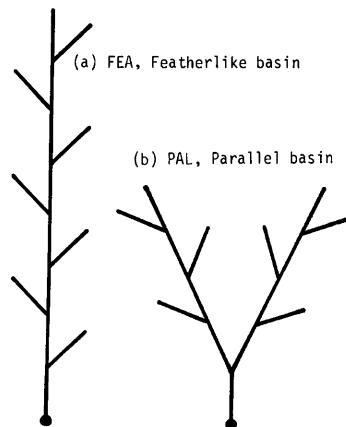


Fig. 4 Channel networks of the basins considered in this research. Rectangular slopes covered with A-layer exist on both sides of each channel segment.

全斜面および全セグメントを一様にしたこと、斜面長・セグメント長の仮定、モデル定数に猪子石流域を求めた値を用いたことなどは、あくまで便宜的なものであって、本質的ではない。実流域を対象とする場合は、その流域を最も適切に記述できる分布モデルがあればそれを用いればよいし、流域条件は各斜面・各セグメント毎に異なって然るべきである。

4.3 想定する降雨条件

流域の分布モデルによるモデル化ができると、次に入力する降雨データを設定しなければならない。過去の実測の降雨データが揃っているような流域では、計画降雨を用いるとよい。そうでない流域では、適当な規模・時間分布の降雨を想定する。

本研究では、総雨量 300 mm, 継続時間 50 hr の三角形降雨を想定し、ピーク時刻が降雨開始後10時間後、40時間後、25時間後の3種、すなわち、前半ピーク型、後半ピーク型、中央ピーク型の3種の時間分布パターンを考慮した。以下では、それらの降雨を No. 01, No. 02, No. 03 と略称する。

4.4 非系統的観測データの生成

分布モデルでモデル化された流域に、想定降雨(あるいは計画降雨)を入力することにより、降雨・流量

データを生成することができる。ただし、こうして得た流量データは、観測流量の平均値(すなわち(3)式の f の部分)に相当するものと考え、真の現象の実現値としては stochastic な変動成分 ε_k をも含んだ Y_k ((3)式)の系列 Y ((2)式)を考える。

さて、 ε_k の確率特性をどのように仮定すべきか。 Y は非系統的観測値の系列(すなわち、ランダムな変動成分のみを含む)ということであるから、 ε_k と ε_{k+j} ($j = \pm 1, \pm 2, \dots$) とは独立であり、 $E\{\varepsilon_k\} = 0$ でありさえすればよいと言える。(4)式のモデルの e_k^i を正規分布と仮定して最尤法で同定することを考えると、 ε_k も正規分布と仮定しておくのが最も素直であろう。すなわち、

$$\varepsilon_k \sim N(0, \sigma^2) \quad \dots\dots\dots (20)$$

ここに、 σ は標準偏差を表す。 ε_k は、乱数発生法によって生成し、毎正時の流量(単位は便宜上 mm/hr とする)に加えて Y_k とするのである。すなわち、流量観測時間間隔は 1 hr とし、降雨開始後100時間(降雨継続時間の2倍)分のデータをシミュレートする。 ε_k の標準偏差は、シミュレートしたヒドログラフが、見た目上、不自然に変動しない程度の値として、 $\sigma = 0.5 \text{ mm/hr}$ とした。

平行流域(PAL)に No. 01~03 の降雨を入力して得られる流量データ系列を PAL 01~PAL 03, 羽状流域(FEA)と同様にして得られるものを FEA 01~FEA 03 と呼ぶことにする。Fig. 5 に PAL 01~PAL 03 を示した。

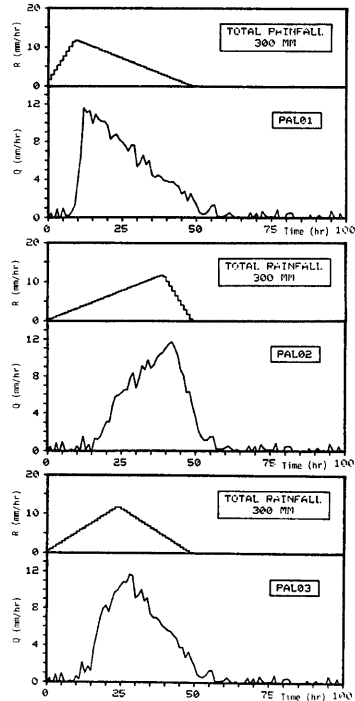


Fig. 5 Simulated rainfall-runoff data which are used to optimize parameters of lumped parameter models (in the case of PAL)

5. 適用結果と考察

前章までの展開で、モデル評価の準備はすべて整った。本章では、いくつかの集中モデルを最尤法によって同定し、AICの算定・評価を行なう。パラメタ同定に用いる最適化手法はコンプレックス法（不等号制約条件下の多変数非線形関数の最大値の直接探索法）¹⁵⁾によった。

5.1 タンクカスケード型モデルの評価

1つの典型的な例として、Fig. 6のようなタンクカスケード型のモデル（TCO(N)モデルと呼んでおく。 N ：タンク数）の評価について考える。このモデルはいわゆる線形貯水池モデルの変形で、各タンクに等分に降雨が供給され、浸透損失孔も考慮されている。タンクの個数をどんどん増やすことができるという意味で階層的構造をもつモデルである。ここで評価しようとするのは、タンク数 N をいくらにするのが適切か、ということである。

TCO(N)モデルのパラメタ数は $2N$ 個であるが、簡単のため浸透損失孔の係数は一定 ($\beta_1 = \dots = \beta_N = \beta$) として、結局 $N+1$ 個のパラメタ数をもつとした。

$N=1, \dots, 5$ の最尤モデルをデータ PAL 03 について求め、それらの誤差二乗和 (sum of square errors, SSE) と AIC とタンク数 N の関連を示したのが、Fig. 7 である。モデル次数 (パラメタ数) を増やすと適合度はよくなってゆく (誤差二乗和が小さくなってゆく) が、AIC は $N=3$ で最小値をとる。すなわち、TCO(3)モデルが PAL 03 に対して最良である。

TCOモデルの最終タンク (第 N 段目のタンク) の側方流出孔の位置と数を Fig. 8 (a), (b) のように変更して、それぞれ、TC 1(N), TC 2(N) モデルと呼ぶことにする。先程と同様の図を描いてみると、Fig. 9 のようになり、TC 1(3), TC 2(4) が最小 AIC モデルである。

このような階層的構造をもつモデルの次数をどの程度にするかという問題には、我々はしばしば会おう訳であるが、AIC はそのような場合に一つの解答を与

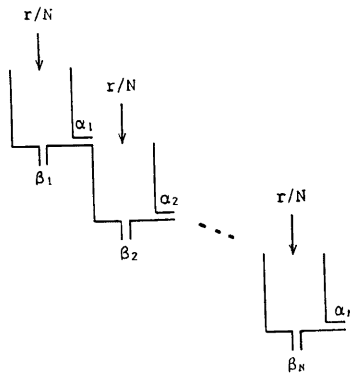


Fig. 6 Tank cascade model (TCO(N) model)

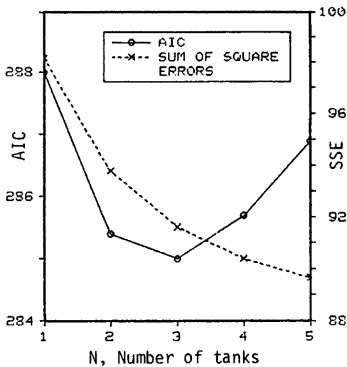


Fig. 7 Relation among AIC, SSE and N (TCO(N) model. PAL 03)

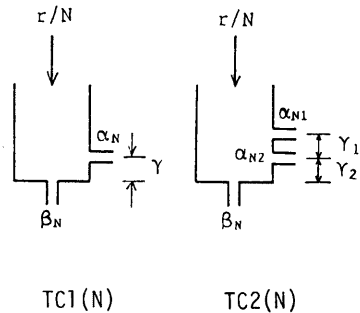


Fig. 8 Modification of the N-th tank of TCO(N)

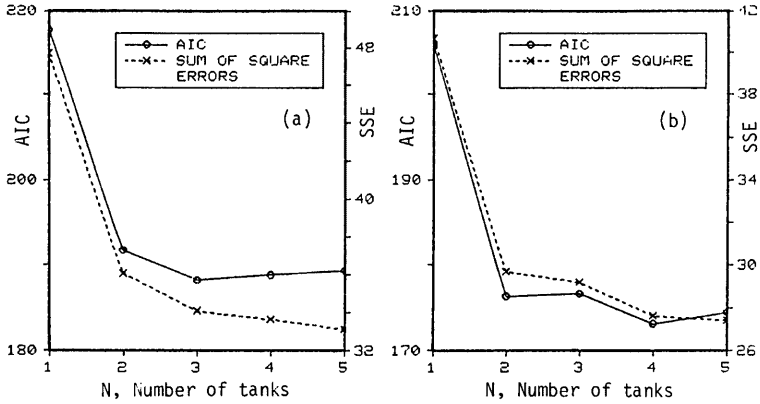


Fig. 9 Relation among AIC, SSE and N (a: TCI(N), PAL 03; b: TC2(N), PAL 03)

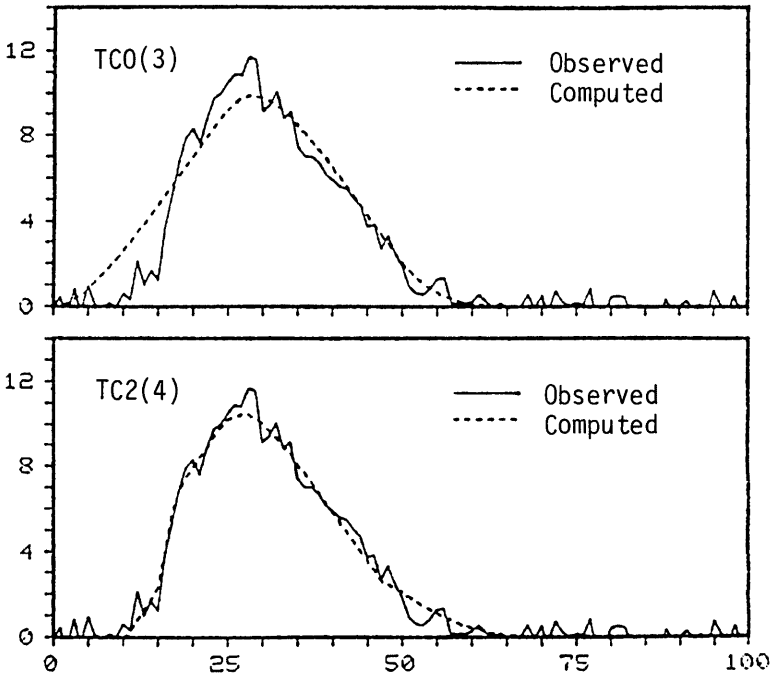


Fig. 10 Comparison between TCO(3) and TC2(4)

えてくれる有力な規準であると言える。

さて、PAL03 の場合について TC0(3), TC1(3), TC2(4) の3種のモデルを比較してみよう。AIC の値は、それぞれ285.0, 188.3, 173.2であり、TC2(4)モデルが最もよいということになる。Fig. 10 に TC0(3) と TC2(4) の再現性を示した。TC0 モデルはハイドログラフの立ち上がり部の再現が不十分であり、流出の非線形性が表現できない。TC2 モデルは TCO モデルを修正したものであるが、このモデルの複雑化（パラメタ数の増加）というデメリットは、適合度の向上というメリットで補えることが、AIC により判定された訳である。

5.2 タンクモデルの評価

菅原¹⁶⁾が提案したタンクモデルは、長期流出・短期（洪水）流出どちらにも用いられ、基本的には前者の場合は直列4段、後者の場合は直列3段とされることが多いが、流域に応じて種々のバリエーションが考えられている。

本研究では、タンクモデルのパラメタを最適化手法でまず求めねばならない。永井・角屋^{17),18)}は、既にこのような試みを行なっている。文献17)では、Powell の共役方向法を用いて 1~3 km² のかなり小さい流域の洪水流出用タンクモデルを最適化している。文献18)では、長期流出用タンクモデルの真のパラメタ値を仮定してシミュレーションデータを発生させておいた上で、上記の Powell 法や Davidon らによる共役勾配法により、最適パラメタの探索を行なっている。タンクモデルの最適化は、いずれの場合も不十分なようである。文献18)の図17, 18に指摘されているように、パラメタの組み合わせが一意に定まらないのである。これは、永井・角屋の方法に問題があるのではなく、タンクモデルの構造が以下に述べるように最適化手法になじまないものであることによる。

本研究では、Fig. 11 の4種のタンクモデルの最適化を試みるが、側方流出孔が2個以上ある場合、どの孔からの流出がハイドログラフに寄与するかによってパラメタ値が大きく変わってくる。例えば、Fig. 11 の TANK 1(2) モデルの場合、上段タンクからの流出を多とすれば、 β_1 の値を小さくして α_1 の値を大きくするとよいし、下段タンクからの流出を多とすれば、 β_1, α_2 を大きくしてやればよいのである。

このことから判るように、タンクモデルは、パラメタ数が多い上、側方流出孔が複数個あるため、パラメタ相互の依存性が強いという構造を本来もっているのである。このことが、タンクモデルの利点でもあり欠点でもあると言えるのであって、少なくとも最適化手法にはなじまないモデルであると言える。

したがって、パラメタに何らかの制約を設けて最適化を行なうこととする。まず、次のようにいくつかのパラメタを固定して、自由パラメタ数を減らす。

TANK 1(1) モデル……固定するパラメタはなし

TANK 2(1) モデル…… $\gamma_1 = 15.0$

TANK 1(2) モデル…… $\gamma_2 = 15.0, \beta_2 = 0.003$

TANK 2(2) モデル…… $\gamma_1 = 10.0, \gamma_2 = 15.0, \beta_2 = 0.003$

これにより、自由パラメタ数は TANK 1(1) モデルで3個、TANK 2(1), TANK 1(2) モデルで4個、TANK 2(2) モデルで5個となる。また、自由パラメタのうち γ は、離散的な値をとるものとし、

TANK 1(N) モデルの γ_1 は、30.0, 40.0, 50.0

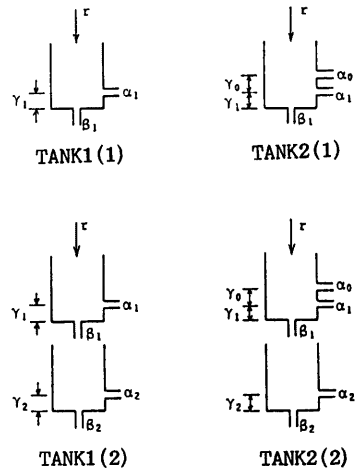


Fig. 11 Sugawara-type tank models considered in this research

TANK 2(N) モデルの γ_0 は, 20.0, 30.0, 40.0

のいずれかをとることとした ($N=1, 2$).

これらの値は, 菅原¹⁶⁾が洪水流出用タンクモデルの定数解析を実流域で得たものを参考にして決定したので, 現在の水準のタンクモデルの性質を大きく損う設定ではない。3段目のタンクを考慮しなかったのも同様で, 菅原が解析した3段の洪水流出用タンクモデルの3段目タンクの側方流出孔の係数は極めて小さくハイドログラフへの寄与は無視できる。すなわち, 2段のモデルで十分であると考えてよい。

このようにして同定した結果, PAL 03 の場合の各タンクモデルの AIC は, TANK 1(1) が220.7, TANK 2(1) が203.8, TANK 1(2) が197.8, TANK 2(2) が169.1であった。これらのうちでは, TANK 2(2) が最良であり, 前節の TC 2(4) モデル (AIC=173.2) よりも良い。

5.3 貯留関数法の評価

木村¹⁹⁾が提案した貯留関数法 (Kimura's storage function method, KSFM) は, 現在わが国で最もよく用いられる洪水流出モデルである。貯留関数法の基礎式は

$$\begin{cases} \frac{ds}{dt} = r_e(t - T_L) - q & \dots\dots\dots(21) \\ s = KqP \end{cases}$$

ここに, $r_e(t)$: 時刻 t の有効降雨強度 (mm/hr), q : 直接流出高 (mm/hr), s : 流域内雨水貯留高, K, P : 定数。有効降雨は,

$$r_e(t) = \begin{cases} f_1 r(t) & (\text{累加雨量} \leq R_{sa}) \\ r(t) & (\text{累加雨量} > R_{sa}) \end{cases} \dots\dots\dots(22)$$

ここに, f_1 : 一次流出率 ($0 < f_1 < 1$), R_{sa} : 飽和雨量 (mm); あるいは

$$r_e(t) = fr(t) \dots\dots\dots(23)$$

ここに, f : 流入係数 ($0 < f < 1$);

で, (22)式か(23)式のどちらかが用いられるが, 本研究では, 前者を採用した。この場合, パラメタは, K, P, T_L, R_{sa}, f_1 の5個となる。

コンプレックス法により最尤モデルを求め, AIC を算定したところ, PAL 03 の場合, AIC=157.0 であった。これは, 前節で求めた TANK 2(2) モデルよりも小さい値となっている。他の入力条件 (PAL 01, PAL 02) の場合も同様であった (Table 1)。羽状流域の場合も各入力条件 (FEA 01~03) のすべてにおいて, KSFM モデルが最小 AIC モデルとなった。

以上の結果, 本研究で検討した集中型モデルのうちでは, Fig. 4 の2種の仮想流域に対しては KSFM モデルが最良である。KSFM はパラメタ数も多くないし, TC, TANK 両モデルに比して計算法も簡単である上, 適合度も優れている。現在, わが国の非常に多くの流域で KSFM が用いられている事情も, 一応理解できる訳である。

しかしながら, 問題がない訳ではない。と言うのは, KSFM の場合同一流域条件であっても, 入力条件が変わると (すなわち, 洪水毎に) 最適パラメタ値が大きく変動するのである²⁰⁾。これは流出モデルにとっ

Table 1 SSE and AIC values of some models

BASIN-RAINFALL	TC 2(4)		TANK 2(2)		KSFM	
	SSE	AIC	SSE	AIC	SSE	AIC
PAL 01	25.1	159.6	30.3	176.3	24.1	153.3
PAL 02	31.6	186.6	38.7	200.8	22.8	148.1
PAL 03	27.7	173.2	28.2	169.1	25.0	157.0

て不都合なことであるので、これを解消する方向でモデルを構成してゆかねばならない。

そのようなモデルが確立されていない現状においては、例えば、次のような2段階の手順によって、対処していく必要があろう。

- ① 同定用と予測（検証）用のデータを数洪水ずつ用意しておき、個々の同定用洪水についてモデル同定（最尤法による）を実行する。それぞれの AIC を求め AIC の小さいいくつかのモデルについて、パラメタ値の変動性の検討や、それが出力に与える影響の感度分析的検討を行なって、入力条件に影響されないモデルを複数個選ぶ。
- ② 予測（検証）用データを用いて、①で選んだモデルの予測性能を比較し、最良のものを採択する。

6. 結 語

単一流域（面積数百 km² 程度）における洪水流出モデルの評価法を提示した。本手法の特徴を挙げると、

- ① 分布モデルと集中モデルを区別し、分布モデルを介して当該流域における最良の集中モデルを見出すとする；したがって実測データが不要である；
- ② 情報量規準（AIC）を導入したので、モデルの適合度だけでなく、モデルの簡潔さをも評価できる；などである。

本研究では、階層的構造をもつモデルの最適次数の選択、異なる構造をもつモデル相互の優劣の比較といった問題に対する適用例を示した。貯留関数型のモデルは一組の入出力データに対する適合性がよいことが判ったので、Prasad²¹⁾ や星²²⁾の貯留関数モデルの評価も今後試みてゆきたいと考えている。

実流域に本手法を適用する際に問題となるのは、分布モデルによる流域の記述であろう。十分な精度で記述できれば、流域条件が変化しても、それに応じて分布モデルのパラメタを合理的に変更することができるはずであるから、実測データの蓄積を待たずに直ちに新条件での最適モデルが求められることになる。従来、どちらかと言えば集中モデルの開発・利用に力が注がれてきたが、こうした意味でも分布モデルは重要であり、よりよい分布モデルの構築を図ってゆく必要がある。

集中モデルの構築について一言付け加えておくと、従来、局所的な物理現象に基づいて導かれた集中モデルをいきなり広い流域に適用したり、入出力データのみに基づいて black box 的アプローチによって集中モデルを構成したりすることが多かったようである。一組の入出力関係に対しては、このような方法で得たモデルであってもパラメタ値を適当に動かせば、かなりの適合度が得られることは事実である。しかしながら、このようなモデルは必ずしも一般性をもたず、5.3の最後に述べたように洪水毎にパラメタ値が大きく変動することが多い。既存のモデルの評価だけでなく、新たな集中モデルが考案された場合にも、この種の問題に対するチェックの方法として、本研究で示したようなフレームワークは重要であると考えられる。

本研究を進めるにあたり、大学院学生 今村宏氏（現鹿島建設）、楠橋康広氏の協力を得た。また、数値計算には京都大学大型計算機センター FACOM M-382 を利用した。ここに記して、謝意を表する次第である。

参 考 文 献

- 1) 高樺琢馬：流出系の同定と予測，水工学に関する夏期研修会講義集，A-2，土木学会水理委員会，1982。
- 2) 高樺琢馬・室 馨・今村 宏・楠橋康広：出力の適合度による流出モデルの評価，土木学会関西支部年次学術講演概要，II-34，1983。
- 3) 高樺琢馬：出水現象の生起場とその変化過程，京都大学防災研究所年報，第6号，1963，pp. 166-180。
- 4) Chiu, C-L. (ed.): Application of Kalman Filter to Hydrology, Hydraulics, and Water Resources, Proc. of AGU Chapman Conf., Univ. of Pittsburgh. Penn., U.S.A., 1978。

- 5) 宝 馨・高棹琢馬・椎葉充晴：洪水流出の確率予測における実際的手法，第28回水理講演会論文集，土木学会水理委員会，1984，pp. 415-422.
- 6) Akaike, H.: A New Look at the Statistical Model Identification, *IEEE Trans. on Automatic Control*, Vol. AC-19, No. 6, 1974, pp. 716-723.
- 7) 赤池弘次：情報量規準 AIC とは何か，数理科学，No. 153, 1976, pp. 5-11.
- 8) 赤池弘次：統計的検定の新しい考え方，数理科学，No. 198, 1979, pp. 51-57.
- 9) 坂本慶行・石黒真木夫・北川源四郎：情報量統計学，共立出版，1983.
- 10) Katayama, T.: Application of Maximum Likelihood Identification to River Flow Prediction, *Real-Time Forecasting/Control of Water Resource Systems*, Selected Papers from an IIASA Workshop, Oct. 18-21, 1976, Ed. Wood, E.F., PERGAMON PRESS, 1980, pp. 63-82.
- 11) McLeod, A.I. and K.W. Hipel: Preservation of the Rescaled Adjusted Range, 1. A Reassessment of the Hurst Phenomenon, *Water Resources Research*, Vol. 14, No. 3, 1978, pp. 491-508.
- 12) Ozaki, T.: On a Model of a Nonlinear Feedback System for River Flow Prediction, *Water Resources Research*, Vol. 16, No. 1, 1980, pp. 225-231.
- 13) 尾崎 統：水文システムの統計モデル，数理科学，No. 213, 1981, pp. 37-44.
- 14) 高棹琢馬・椎葉充晴：地形パターン関数を導入した洪水流出モデル，第26回水理講演会論文集，土木学会水理委員会，1982，pp. 217-222.
- 15) 上田光三郎：不等号制約条件下の多変数非線形関数の最大値の直接探索法（コンプレックス法），京都大学大型計算機センター利用の手引き——プログラムライブラリ編——，1980，pp. 71-78.
- 16) 菅原正巳：流出解析法，共立出版，1972.
- 17) 永井明博・角屋 睦：洪水流出モデルの適用比較，京都大学防災研究所年報，第21号B-2，1978，pp. 235-249.
- 18) 永井明博・角屋 睦：流出モデル定数の最適化手法，京都大学防災研究所年報，第22号B-2，1979，pp. 209-224.
- 19) 木村俊見：貯留関数法，土木技術資料，第3巻第12号，1961および第4巻第1, 4, 5, 6, 7号，1962.
- 20) 宝 馨・高棹琢馬：貯留関数法のパラメタ変動特性と同定法の数値実験的検討，土木学会第39回年次学術講演会講演概要集，第2部，1984.
- 21) Prasad, R.: A nonlinear hydrologic system response model, *Proc. of ASCE*, Vol. 93, No. HY 4, pp. 201-221.
- 22) 星 清：一般化貯留関数法とそのパラメタ同定，昭和55年度文部省科研（一般C）報告書，確率流量の算定に関する総合的研究，代表・藤田睦博，1981.