

GA を用いたタンクモデル同定に関する統計的検討

STATISTICAL STUDY FOR TANK MODEL IDENTIFIED BY GENETIC ALGORITHM

鈴木 誠¹・百田博宣²・神野健二³・河村 明⁴
Makoto SUZUKI, Hironobu MOMOTA, Kenji JINNO and Akira KAWAMURA

¹正会員 工博 清水建設(株)和泉研究室 (〒100 東京都千代田区内幸町 2-2-2)

²正会員 工博 清水建設(株)技術開発センター (〒105-07 東京都港区芝浦 1-2-3)

³正会員 工博 九州大学教授 工学部建設都市工学科 (〒812 福岡区東箱崎 6-10-1)

⁴正会員 工博 九州大学助教授 工学部建設都市工学科 (〒812 福岡区東箱崎 6-10-1)

The tank model is useful for runoff analysis since it can represent non-linear stream flow behavior. It is difficult to properly identify a lot of parameters of the model from observed data. The genetic algorithm (GA) is a search procedure based on the mechanism of natural genetics and is efficient for the global optimization. On the other hand, the uncertainty of the identified parameters, which may significantly affect the prediction of the stream flow, should be taken into account.

This paper describes the statistics of the identified parameters using the GA. The statistics is evaluated by the bootstrap method to simulate the data fluctuation. Furthermore, the information criterion EIC which is derived from the bootstrap method is introduced and applied to select the best tank model.

Key Words: runoff analysis, tank model, genetic algorithm, bootstrap, extended information criterion

1. 概要

山岳地域にトンネルなどを掘削すると、湧水などが生じることがある。湧水地域では、このような地下構造物の構築にあたり、河川流量などの環境に与える影響を検討することが重要となってくる。タンクモデルは流域をいくつかの貯留型容器の結合で表現する流出解析法である。菅原らの提案している直列4段タンクモデル¹⁾は、簡単な構造で非線形の挙動が表現できることから、長期流出解析に広く用いられている。一般に、このモデルは1つの流域流出量から16個のパラメータを推定しなければならず、モデルの同定には経験が必要となることから、自動化プログラムが検討された^{2),3)}。

タンクモデルの同定を、流域流出量の誤差が最小となるような非線形関数の最適問題に置き換え、パラメータを推定しようとする研究は古くから行われてきた。小林・丸山⁴⁾はPowellの共役方向法を、渡辺⁵⁾はNewton法を、永井・角屋^{6),7)}はPowell法、DFP法に基準化形式を導入したSP法、SDEP法を提案している。また、安永⁸⁾は拡張カルマンフィルタを用い、逐次推定を試みている。これらの手法は、局所的な最適解に陥りやすいことから、大域的探索法の

研究がなされるようになってきた。たとえば、Wang¹⁰⁾は概念的雨水流出モデルにGAを、田中丸¹¹⁾はタンクモデルにGAを導入し、適用性を検討している。Duran¹²⁾やSorooshian¹³⁾はSimplex法にGAに類似した概念を取り入れたSCE-UA法を提案し、概念的雨水流出モデルに適用しており、これらの研究の概要は田中丸の論文¹⁴⁾で紹介されている。

タンクモデルのパラメータ推定は、逆問題と考えると解の一意性や連続性のない非適切問題となることが考えられる。そこで、本研究では環境評価を目的として、数学的な最適解にこだわらず、工学的な利用を考えた最適解を求めるため、主観性の少ない解析手法の精度と有効性の検討を行う。解析手法として、流域流出量の最小誤差となるような目的関数は局所的な解が複数存在する多峰性の最適問題となることから、大域的探索を得意とする遺伝的アルゴリズム(GA)を用いる。さらに、ブートストラップ法¹⁵⁾によりGAによる推定パラメータの不確定性を評価し、情報量規準を用いたタンクモデルの最適な段数の設定についても考察する。

2. 解析モデルとパラメータ推定

(1) 解析手法

降雨量と流出量の観測値から、その現象を表現するための流出解析モデルを設定するとき、モデルのパラメータを求めることが重要となる。特に、タンクモデルは非線形挙動を説明できるためよく用いられる流出解析モデルであるが、パラメータを推定することが難しい。そこで、なるべく主観的な知識がなくても、ある程度の推定が行えることを目標とする。大域的探索を得意とするGAは、離散値を扱うことから精度を上げようとするパラメータのビット列を多くする必要があり、組み合わせ数が増えることにより計算時間が増加する。ここでは、GAの設定条件にはあまり言及せず、ある条件のもとで行われたGAによる推定パラメータがどのくらいばらつくのかを検討する。

(2) タンクモデル

流出解析モデルとして、図-1に示す直列4段タンクモデルを用いる。図中、 r ：降雨量 (mm/hour)， h_i ：タンク水深 (mm)， a_i ：タンク側方流出孔係数 (1/hour)， b_i ：タンク下方浸透孔係数 (1/hour)， c_i ：タンク側方流出孔高さ (mm) を表わす。

次にタンク側方からの流出量を $Q_1 \sim Q_5$ (mm/hour)，タンク下方からの浸透量を $I_1 \sim I_3$ (mm/hour)とすると、

$$\begin{aligned} Q_1(t) &= a_1(h_1(t) - c_1) U(h_1(t) - c_1) \\ Q_2(t) &= a_2(h_1(t) - c_2) U(h_1(t) - c_2) \\ Q_3(t) &= a_3(h_2(t) - c_3) U(h_2(t) - c_3) \\ Q_4(t) &= a_4(h_3(t) - c_4) U(h_3(t) - c_4) \\ Q_5(t) &= a_5 h_4(t) \\ I_1(t) &= b_1 h_1(t) \\ I_2(t) &= b_2 h_2(t) \\ I_3(t) &= b_3 h_3(t) \end{aligned} \quad (1)$$

となる。ここに、 $U(x)$ は次式の単位階段関数である。

$$U(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

各タンクの連続式は次式のように表現できる。

$$\begin{aligned} dh_1/dt &= r(t) - Q_1(t) - Q_2(t) - I_1(t) \\ dh_2/dt &= I_1(t) - Q_3(t) - I_2(t) \\ dh_3/dt &= I_2(t) - Q_4(t) - I_3(t) \\ dh_4/dt &= I_3(t) - Q_5(t) \end{aligned} \quad (2)$$

また、総流出量 $Q(t)$ は次式で表わされる。

$$Q(t) = \sum_{i=1}^5 Q_i(t) \quad (3)$$

(3) 目的関数の設定

誤差評価関数は次式に示すような種々の関数が使

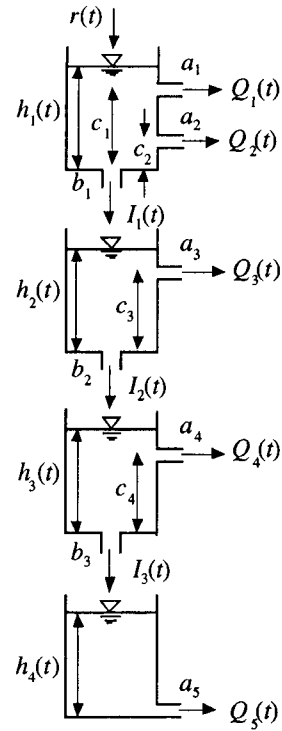


図-1 タンクモデル

われてきた。最初は最小二乗誤差基準であり、2番目は対数をとった値の最小二乗誤差基準である。3番目は X^2 誤差基準で、最後は相対誤差基準である。結果的には、2番目と最後は同等の評価を与える。

$$\begin{aligned} J_{LS} &= \frac{1}{M} \sum_{i=1}^M (Q_c(i) - Q_o(i))^2 \\ J_{LL} &= \frac{1}{M} \sum_{i=1}^M (\log Q_c(i) - \log Q_o(i))^2 \\ J_{XS} &= \frac{1}{M} \sum_{i=1}^M \frac{(Q_c(i) - Q_o(i))^2}{Q_o(i)} \\ J_{RE} &= \frac{1}{M} \sum_{i=1}^M \frac{|Q_c(i) - Q_o(i)|}{Q_o(i)} \end{aligned} \quad (4)$$

ここに、 Q_o ：観測流出量， Q_c ：計算流出量， M ：データ数を示す。

流出量のようなピークを有するような関数では、最初の最小二乗誤差基準より他の評価基準の方が特徴を捕らえやすい。ここでは、2番目の対数最小二乗誤差基準を用い、制約条件などのペナルティ関数は考慮しない。

(4) 遺伝的アルゴリズム (GA)

GAにより、目的関数 $f(x_1, x_2, \dots, x_n)$ ， $x_i^{\min} \leq x_i \leq x_i^{\max}$ ， $i=1, \dots, n$ を最小とするパラメータベクトル $\mathbf{x} (=x_1, x_2, \dots, x_n)$ を探索する方法について述べる。GAでは、 $\mathbf{x}$ を0と1のビットを用いて N 個の記号列によって表現する。すなわち、連続量を扱う場合は、 x_i の探索範

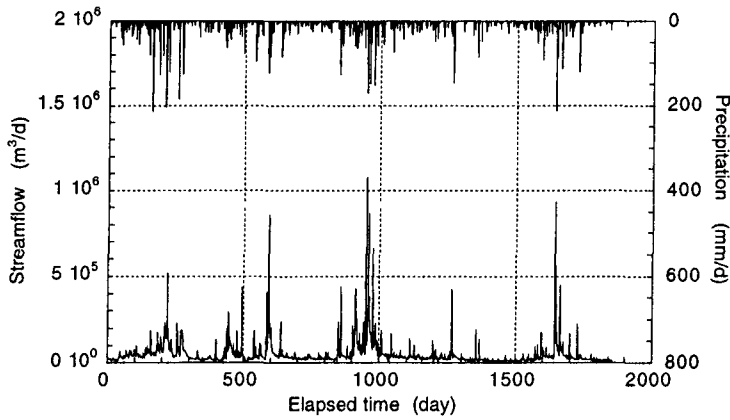


図-2 観測された日流出量と日降水量

囲を $2L$ 個の点に離散化して、各点に 0 から $2L-1$ までの L ビットの 2 進コードを n 個連結した $N=n \times L$ ビットのビット列で表現される。このようなビット列を遺伝子とみなしたパラメータベクトル (個体) に基づいて、選択、交叉、突然変異の 3 種類の遺伝的操作からなるアルゴリズムで、目的関数を最小にする個体を探索する。

まず、最小化する目的関数の値が小さい個体は適用度が高いと定義する。選択とは、この適用度の大きさに応じた選択確率に従って、子孫を残すための親をランダムに選択するという操作である。本研究では、選択に際してエリート保存戦略といわれる適用度の大きい個体を無条件で次世代に残す方法を採用する。交叉は、選択された 2 個体を選び、その特定部位を交換して新しい個体を作る操作である。突然変異は、ある小さな確率で選び出した個体の任意のビット値をある確率で反転させるものである。

3. 適用事例

(1) 対象地域

解析に用いる資料は、M ダム流域の流域平均日降水量で解析期間は 1991 年から 5 年間とした。図-2 に日流出量と日降水量を示す。降雪は流出量に影響するほどなく、本研究では洪水流出量に焦点をあてたことから蒸発散は考慮していない。この日流出量に 4 段タンクモデルを適用し、初期タンク水深を含めた 16 個のパラメータを推定する。

(2) 解析条件

GA による探索のため、既往の文献¹⁴⁾より 16 個のパラメータの探索範囲として、表-1 に示すような下限値および上限値を設定する。この探索範囲は、これまでの 4 段タンクモデルの適用結果を参考に定められており、実用上十分な大きさと考えられる。

これらの探索範囲を、GA では $L=7$ として各パラ

表-1 パラメータの探索範囲

Parameter	Upperbound
a_1, a_2	0.635
a_3	0.254
a_4	0.0127
a_5	0.0127
b_1	0.635
b_2	0.254
b_3	0.127
c_1	127.0
$c_2 \sim c_4$	63.5
$h_{1,0} \sim h_{4,0}$	127.0

表-2 解析結果

Parameter	Estimated value	Parameter	Estimated value
a_1	0.174	c_1	108.0
a_2	0.119	c_2	14.9
a_3	0.179	c_3	41.2
a_4	0.0189	c_4	11.4
a_5	0.0001	$h_{1,0}$	41.7
b_1	0.203	$h_{2,0}$	16.9
b_2	0.491	$h_{3,0}$	97.2
b_3	0.0139	$h_{4,0}$	754.0
σ_e	0.3105		

メータ p_i を次のように離散化する。

$$p_i = \Delta p_i z_i$$

$$z_i = (p_i - p_i^{\text{lower}}) / (p_i^{\text{upper}} - p_i^{\text{lower}}) \quad (5)$$

$$\Delta p_i = (p_i^{\text{upper}} - p_i^{\text{lower}}) / (2^7 - 1)$$

ここに、 z_i : 整数、 $0 \leq z_1, z_2, \dots, z_{16} \leq 2^7 - 1 = 127$, Δp_i : p_i の離散化幅を表わす。

ここでは、GA の個体数 1,000、交叉率 0.6、総試行回数 100,000 (世代数) と設定し、探索結果への影響を検討する。突然変異の発生確率は 0.01 を採用した。

(3) 解析結果

GA による解析結果を表-2 に示す。また、対数最小二乗誤差 J_{LL} は $\sigma_e = 0.3105$ であり、GA の設定パラメータにも改善の余地は残るものの、大きな傾向は変わらないと考えられる。推定結果を観測結果の一部と対比すると図-3 のようになる。

観測データの対数と推定データの対数の差 (対数誤差) は図-4 のようになり、目的関数で定義したように正規分布を示し、最小二乗推定量と扱うことができると考えられる。ここでは、タンクモデルからの流出量の単位となっており、平均値 -0.022、分散値 0.101

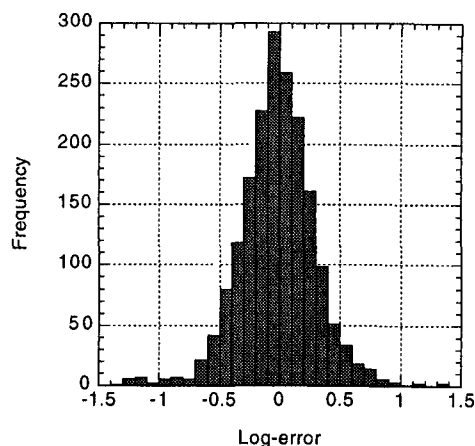
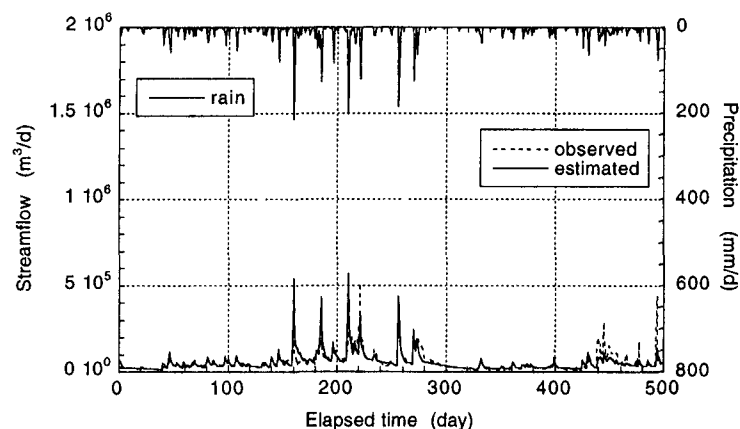


図-4 対数誤差のヒストグラム

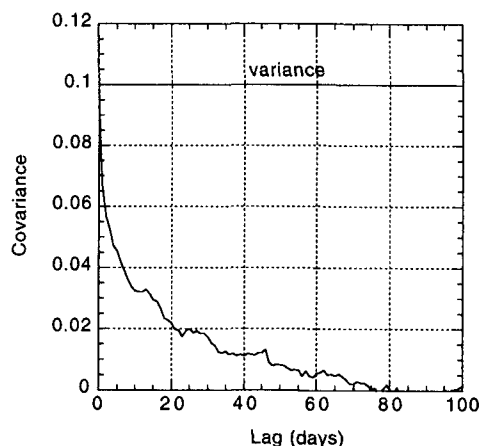
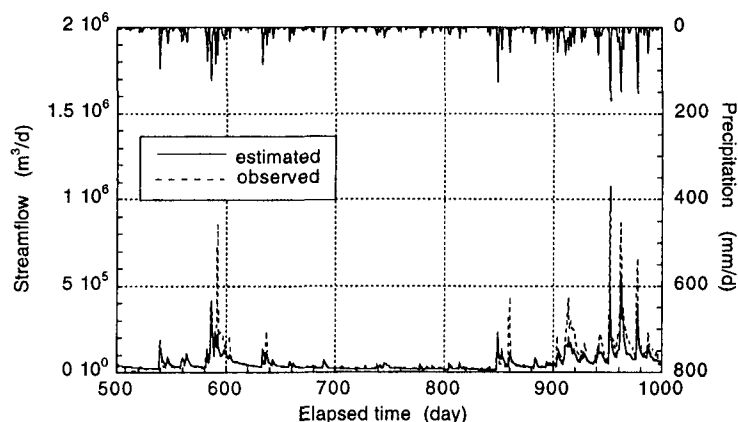


図-3 観測結果と推定結果の一部

図-5 対数誤差の自己相関性

である。また、対数誤差の自己相関関数は図-5のようになり、白色性には問題が残ったが、モデルを複雑にしないためホワイトノイズとしてブートストラップ法による検討を行う。

互いに相関を有していることが考えられる。特に、初期タンク水深 $h_{i,0}$ とタンク側方流出孔高さ c_i などは相対的なものであるから、この傾向が想定される。しかし、表-3 にパラメータ間の相関係数を示すが、強い相関は見られないことがわかった。

(4) パラメータの不確定性

4. モデル選択基準

推定されたタンクモデルの誤差は、流出量から定義された目的関数で評価される。ここでは、この誤差を $N(0.0, \sigma_e^2)$ の正規分布と仮定し、GAにより推定されたモデルパラメータの不確定性をパラメトリック・ブートストラップ法により算定する。ブートストラップ法は、推定された流出量の経時変化にモデル誤差をモンテカルロ法により付加したブートストラップ標本を作成し、GAによるパラメータ推定から、そのばらつきを評価することができる。

モデルの選択基準でよく知られているAICは、最大対数尤度をパラメータ数から補正した平均対数尤度を用いる評価指標であり、これを最小とするモデルが適切であると判断する。しかし、AICは最尤推定量に基づいた指標であることから、それ以外の手法では使えないという問題がある。これを拡張したEIC(Extended Information Criterion)はブートストラップ法により、データによる対数尤度の偏りを補正する評価指標であり、計算時間は必要であるが適用性は広い¹⁶⁾。式で表現すると次のようになる。

図-6に結果の一部のヒストグラムを示す。初期タンク水深 $h_{i,0}$ は均一にばらついており、どこでも最適パラメータになりえる。結論として、多くのパラメータは探索範囲の影響を強く受けており、推定されたパラメータが一意的でない解となっている。

$$\text{EIC} = -2 \log f(X|\hat{\theta}) + 2E_x \left[\log f(X^*|\hat{\theta}^*) - \log f(X|\hat{\theta}^*) \right] \quad (6)$$

ブートストラップ法で求められたパラメータは、

ここで、 X^* はブートストラップ標本を、 $\hat{\theta}^*$ は推定され

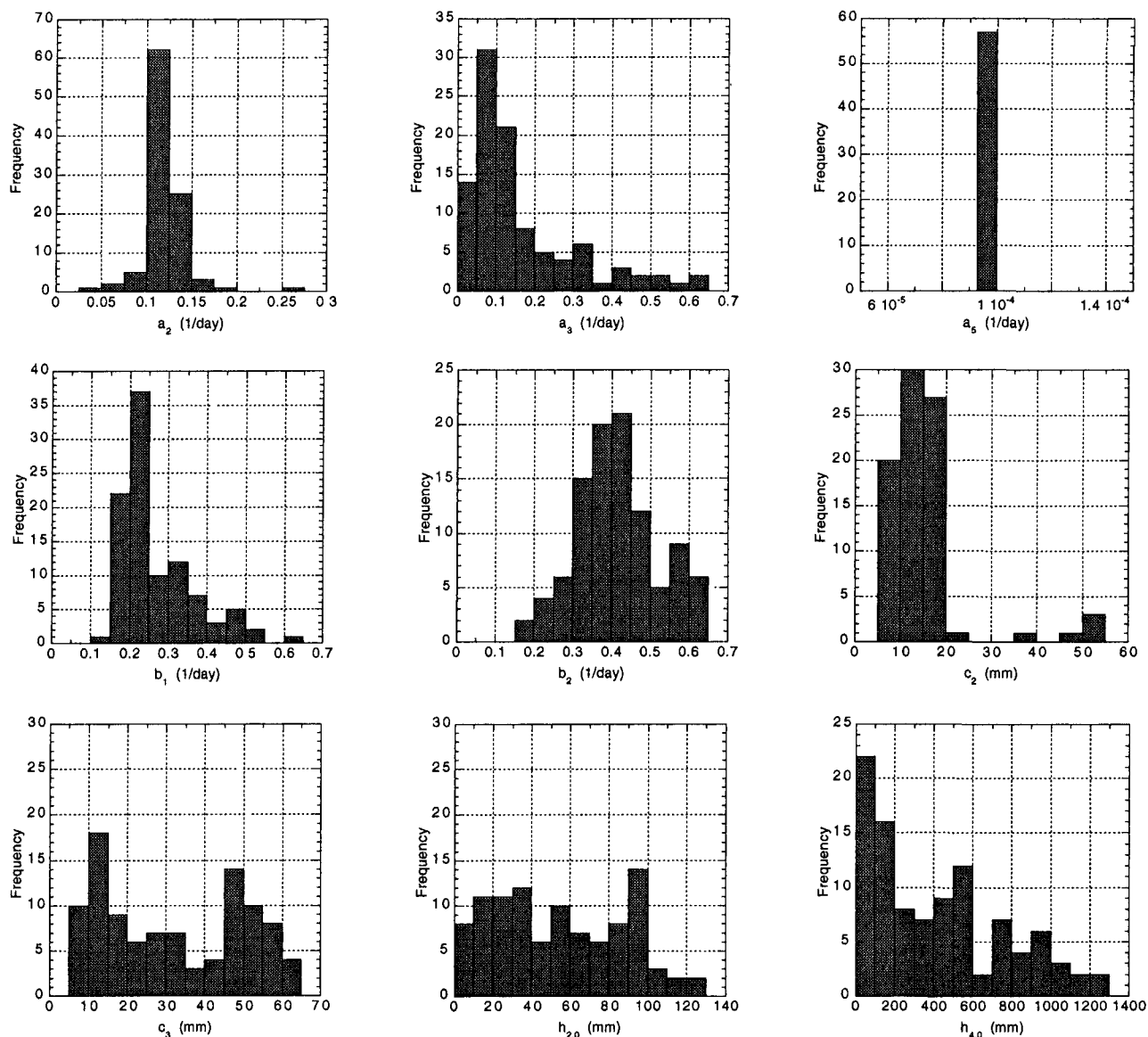


図-6 パラメータのばらつきの一部

表-3 パラメータ間の相関性

	$h_{1,0}$	$h_{2,0}$	$h_{3,0}$	$h_{4,0}$	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	c_1	c_2	c_3	c_4
$h_{1,0}$	1.000	-0.226	-0.271	0.024	-0.045	0.017	0.063	-0.136	-0.078	-0.211	0.034	-0.194	0.017	0.128	0.204	0.151
$h_{2,0}$	-0.226	1.000	-0.375	-0.045	-0.041	-0.090	0.020	-0.126	0.109	-0.031	-0.080	-0.119	-0.049	0.198	-0.018	0.252
$h_{3,0}$	-0.271	-0.375	1.000	-0.094	0.034	-0.006	0.064	-0.152	0.144	-0.112	0.164	-0.079	0.149	-0.055	0.038	0.377
$h_{4,0}$	0.024	-0.045	-0.094	1.000	0.000	-0.213	0.044	-0.117	-0.256	-0.141	-0.008	0.103	-0.024	-0.028	0.139	0.043
a_1	-0.045	-0.041	0.034	0.000	1.000	0.232	0.111	0.162	0.250	0.150	-0.110	-0.051	0.409	-0.192	-0.071	0.000
a_2	0.017	-0.090	-0.006	-0.213	0.232	1.000	0.084	0.546	0.498	0.440	-0.236	0.068	0.198	-0.226	-0.171	-0.075
a_3	0.063	0.021	0.064	0.044	0.111	0.084	1.000	0.224	0.195	-0.201	0.225	0.043	0.096	0.097	0.460	0.186
a_4	-0.136	-0.126	-0.152	-0.117	0.162	0.546	0.224	1.000	0.687	0.318	-0.241	0.419	-0.165	0.017	0.066	-0.015
a_5	-0.078	0.109	0.144	-0.256	0.250	0.498	0.195	0.687	1.000	0.274	-0.166	0.239	-0.097	0.119	0.028	0.463
b_1	-0.211	-0.031	-0.112	-0.141	0.150	0.440	-0.201	0.318	0.274	1.000	-0.337	0.074	-0.113	-0.259	-0.678	-0.219
b_2	0.034	-0.080	0.164	-0.008	-0.110	-0.236	0.225	-0.241	-0.166	-0.337	1.000	-0.138	0.083	0.094	0.139	0.090
b_3	-0.194	-0.119	-0.079	0.103	-0.051	0.068	0.043	0.419	0.239	0.074	-0.138	1.000	0.074	-0.145	0.228	-0.307
c_1	0.017	-0.049	0.149	-0.024	0.409	0.198	0.096	-0.165	-0.097	-0.113	0.083	0.074	1.000	-0.654	0.079	-0.053
c_2	0.128	0.198	-0.055	-0.028	-0.192	-0.226	0.097	0.017	0.119	-0.259	0.094	-0.145	-0.654	1.000	0.230	0.255
c_3	0.204	-0.018	0.038	0.139	-0.071	-0.171	0.460	0.066	0.028	-0.678	0.139	0.228	0.079	0.230	1.000	0.178
c_4	0.151	0.252	0.377	0.043	0.000	-0.075	0.186	-0.015	0.463	-0.219	0.090	-0.307	-0.053	0.255	0.178	1.000

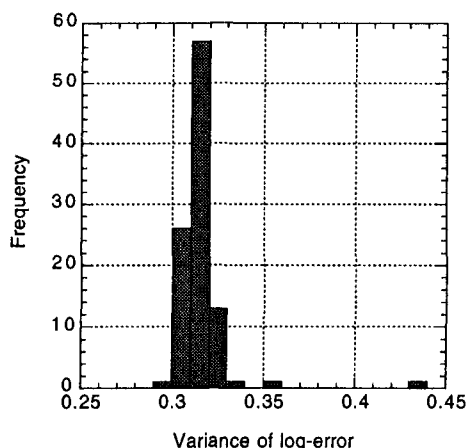


図-7 4段タンクモデルの対数誤差分散のばらつき

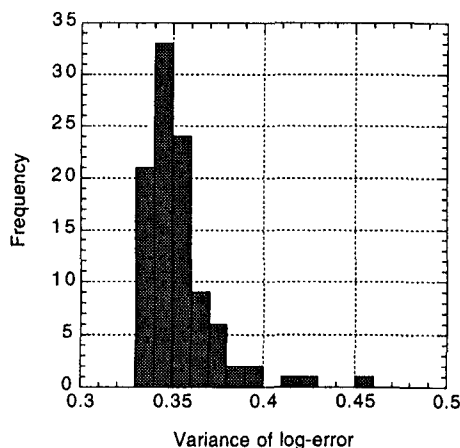


図-8 3段タンクモデルの対数誤差分散のばらつき

たパラメータベクトルを示している。

EICを用い、図-1で示した4段タンクモデルと3段目のタンクをはずしたパラメータ数12の直列3段タンクモデルの比較を試みる。4段タンクモデルと3段目のタンクモデルの目的関数のヒストグラムを図-7と図-8に示す。4段タンクモデルの方がばらつきは小さいことがわかる。推定された3段モデルを観測結果は4段モデルより上式右辺第1項の推定誤差が大きく、第2項の補正項は小さくなった。最終的に、EICでは4段モデルが498.87に対し、3段モデルは1002.15と大きくなり、4段モデルが適切であると選択された。

5. まとめ

Mダム流域を対象にGAを用いた流出解析モデル同定とその統計的不確定性の評価を行った。また、流域の流出解析モデルに直列4段タンクモデルを適用し、そのパラメータを先験的情報を用いずに推定することを試みた。結果として、多峰性で、解の一意性や連続性のない非適切問題であるが、工学的な利用を考えた最適解およびその不確定性を求めることはできることがわかった。問題点として、同定されたパラメータのばらつき、目的関数の定義については今後さらに検討する必要がある。

参考文献

- 菅原正巳：流出解析法，共立出版，1972。
- 菅原正巳・尾崎睿子・渡辺一郎・勝山ヨシ子：タンクモデルの構造を自動的に定める計算機プログラムの開発（第1報），国立防災科学技術センター研究報告，Vol.15，pp.43-89，1977。
- 菅原正巳・尾崎睿子・渡辺一郎・勝山ヨシ子：タンクモデルの構造を自動的に定める計算機プログラムの開発（第2報），国立防災科学技術センター研究報告，Vol.20，pp.157-216，1978。
- 小林慎太郎・丸山利輔：Powellの共役方向によるタンクモデル定数の探索，農業土木学会論文集，Vo.65，pp.42-47，1976。
- 渡辺和好・館谷清・松本賢治・星清：タンクモデル定数の最適化手法の改良，第33回水理講演会論文集，pp.55-60，1989。
- 永井明博・角屋睦：流出モデルの最適化手法，京都大学防災研究所年報，Vo.22 B-2，pp.209-224，1979。
- 永井明博・角屋睦：タンクモデルの最適同定法に関する基礎的検討，京都大学防災研究所年報，Vo.23 B-2，pp.239-247，1980。
- 安永武英・神野健二・河村明：宅地開発にともなう農業用溜池への流入量の変化について，第36回水理講演会論文集，pp.629-634，1992。
- 角屋睦・田中丸治哉・永井明博・金口正幸：愛知川上流域への長短期流出両用モデルの適用と実時間洪水予測，水資源研究センター研究報告，pp.45-60，1989.7。
- Wang, Q.J. : The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models, Water Resource Research, Vol.27, No.9, pp.2467-2471, 1991。
- 田中丸治哉：遺伝的アルゴリズムによるタンクモデル定数の同定，京都大学防災研究所年報，Vo.36 B-2，pp.231-239，1993.4。
- Duan, Q., Sorooshian, S. and Gupta, V.K. : Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, Water Resource Research, Vol.28, No.4, pp.1015-1031, 1992。
- Sorooshian, S., Duan, Q. and Gupta, V.K. : Calibration of Rainfall-Runoff Models : Application of Global Optimization to the Sacramento Soil Moisture Accounting Model, Water Resource Research, Vol.29, No.4, pp.1185-1194, 1993。
- 田中丸治哉：タンクモデル定数の大域的探索，農業土木学会論文集，Vo.178，pp.103-112，1995.8。
- Efron, B. : Bootstrap methods: Another look at the jack-knife, Annals of Statistics, Vol.7, No.1, pp.1-26, 1979。
- 北川源四郎・石黒真木夫・坂元慶行：情報量規準AICとEIC，電子情報通信学会技術研究報告，Vol.92，No.503，pp.49-62，1993。

(1997. 9. 30 受付)