流出モデル評価への情報量規準の導入

京都大学工学部　正員　高棹琢馬　京都大学工学部　正員　宝　馨
鹿島建設　正員　今村　宏　京都大学大学院　学生員○楠橋康広

1. 緒　言：　本研究では，情報量規準(AIC)[1),2)]を導入することにより，出力の適合度だけでなく，パラメタの個数をも考慮できる流出モデルの評価法を提案する。階層的構造を持つモデルに本手法を適用して，その最適モデル次数を決定した一例を示す。

2. 基本的な考え方：　流出モデル評価問題が未解決である主な理由として，(i)出力の適合度のみを評価基準とした場合，パラメタ数の多い複雑なモデルが最良のモデルとなる；(ii)観測データに系統的誤差がある；(iii)使用目的，物理性の考慮の度合が異なる集中パラメタモデルと分布パラメタモデルを同列に並べて評価しようとしてきた：などが考えられる。本研究では，(i)'出力の適合度とパラメタ数の両方を評価できるAICをモデルの評価基準とする；(ii)'想定流域を分布パラメタモデルによって記述し，所要の規模・パターンの降雨を与えて流出計算を行ない，これに確率変動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；(iii)'こうして得た入出力関係を最もよく表すことのできる集中パラメタモデルをAICを用いて評価する：すなわち，図1のように分布型モデルと集中型モデルとを分離し，前者は後者より上位レベルにあるものとして扱うのである。

3. モデル構造とAIC：　本研究では，流出現象を物理的かつ確率的であるという認識に立つことが基本的な前提となる。流出系内の真の物理的現象が，m個の物理量（状態量）とn個のパラメタ$\theta_1, \cdots, \theta_n$で説明でき，真の確率的現象が$\varepsilon_k$で表わされるとすると，時刻$k$における系の出力$y_k$は，真の物理的現象を与える関数を$f(\cdot)$として，次式で表わされる。

$$y_k = f(x_1(k), \cdots, x_m(k); \theta_1, \cdots, \theta_n) + \varepsilon_k \qquad (1)$$

実際には，$f(\cdot)$の完全な記述は不可能であり，種々の仮定を設けてy_kを記述しようとしたものが「モデル」である。すなわち，モデルiの状態量の数を$m(i)$，パラメタ数を$n(i)$として，

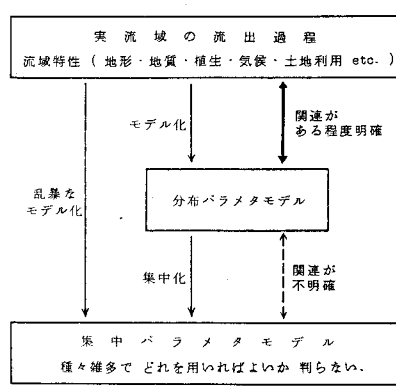

図1

$$y_k = g_k^i(x_1(k), \cdots, x_{m(i)}(k); \theta_1, \cdots, \theta_{n(i)}) + e_k^i \qquad (2)$$

で真の現象を記述しようとするのである。$g^i(\cdot)$はモデルiの関数形で系の平均的挙動を表わし，e_k^iは平均値からの偏差を表わす。ただし，$E\{e_k^i\} = 0$　である。

(2)式のような構造を持つ確率(統計的)モデルの評価には，AICが有力である。AICは，

$$\mathrm{AIC} = -2 \times \log(\text{最大尤度}) + 2 \times (\text{モデルの自由パラメタ数}) \qquad (3)$$

で定義される。赤池の方法は，AICを最小とするモデルが最適なモデルであるとするものであり，時系列解析の分野で多用されている。

Takuma TAKASAO, Kaoru TAKARA, Hiroshi IMAMURA, Yasuhiro KUSUHASHI

流出モデルの評価基準としてAICを適用するには，評価しようとする流出モデルのパラメタが最尤法により求められていることが前提となる。ε_k^iが正規分布に従うとき，最尤法と最小二乗法が等価であることは容易に導かれ，モデルパラメタ数が$n(i)$のときの情報量規準を$AIC(n(i))$とおくと，σ^2が未知の場合，次式で与えられる。

$$AIC(n(i)) = M\log\hat{\sigma}^2 + 2(n(i)+1) + M(\log 2\pi + 1) \qquad (4)$$

ここに，$$\hat{\sigma}^2 = \frac{1}{M}\sum_{k=1}^{M}\{y_k - g_k^i(\cdot)\}^2 \qquad (5)$$

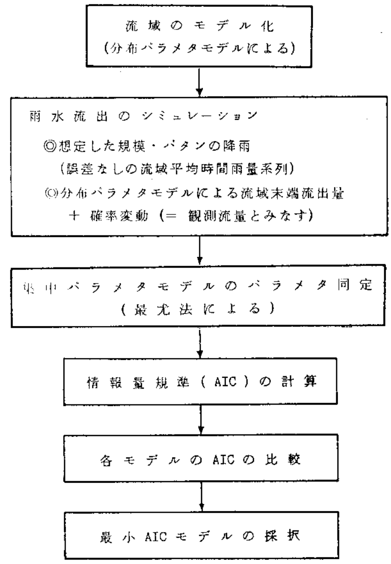

図 2

4. 流出モデル評価の手順： 集中型モデルの評価の手順は図2に示すとおりである。すなわち，系統的誤差のない観測データを2節で述べた方法によりシミュレートし，得られた入出力データを用いて最尤法により各種の集中型モデルのパラメタを同定する。こうして同定された各種モデルのAICを算定し，AICが最小のモデルを最良のモデルとして採択する訳である。

5. 階層的構造を持つモデルの最適次数の決定： 本節では，図3のような集中型モデルに本手法を適用する。仮想流域(図4)において，分布型モデルによりシミュレートされた入出力データを用いて，パラメタ同定を行なった。モデル次数(タンクの個数)は，N=1,…,5とした。図5に，モデル次数と誤差二乗和とAICの関係を示す。次数を増やせば増やすほど適合度はよくなるが，AICはN=3で最小となる。すなわち，タンク数3段のモデルが最適なモデルである。このように，本手法によれば，出力の適合度だけでなく，モデル次数(パラメタ数)をも同時に評価することができる。

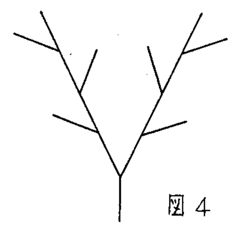
図4

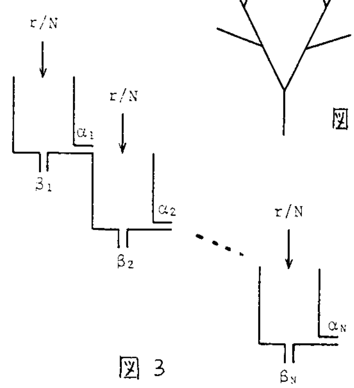

図 3

6. 結　論： 本研究では，出力の適合度のみでなく，パラメタ数をも同時に評価できる方法を提案し，階層的構造を持つ集中型モデルへの簡単な適用例を示した。本手法は，AICを導入することにより，情報量という同一の尺度でモデルの良否を比較するので，異なる形式のモデル同士を比較評価することももちろん可能である。[3),4)]

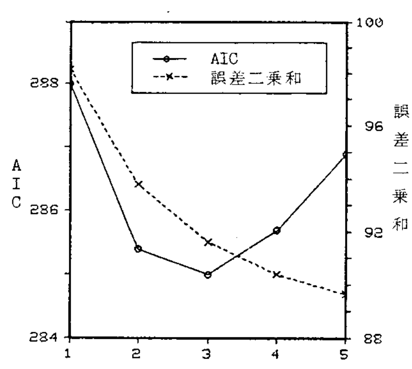

図 5

［参考文献］

1) 赤池：情報量規準とは何か，数理科学 No.153, 1976.

2) 坂本・石黒・北川：情報量統計学，共立出版，1983.

3) 高棹・宝・今村・楠橋・安藤：昭和59年度関西支部年講，II-40.

4) 高棹・椎葉・宝：京大防災研年報，第27号B-2, 1984(投稿中).