

## 公開講演会要旨

## 河川流量時系列の解析

統計数理研究所 尾 崎 統

(昭和63年11月4日, 統計数理研究所 講堂)

## 1. はじめに

統計ソフトウェアという統計データの平均や分散を計算するコンピュータプログラムを想像する人が多いかもしれない。確かにデータの平均と分散はデータを特徴付ける為の基本的なパラメータであるが、現代の統計学では、データの中に更に複雑な構造の内在したものがあれば、それが何であれ、その段階にまで立ち入り、その構造の特性も含めて特徴付けるということまでやる。本講ではそのような統計ソフトウェアの一例として、河川流量予測に関する貯蔵モデル及びその計算ソフトウェアに関連する計算環境について述べてみたい。

## 2. タンクモデル

降雨量から河川の流量を予測する問題は農業、土木など水資源に関連する諸分野において重要な問題であり、種々の手法が考えられてきたが、現在最も一般的でよく用いられる方法はタンクモデルによる方法であろう。タンクモデルの典型的な例は図1のような構造になっている。降った雨はタンクの底や壁の穴から流れ出て、順次、下のタンクに行き最終的に河川流量となる。タンクの数、穴の数、穴の大きさ、穴の位置を変えることによって、それぞれの河川の雨量と流量の間の非線形形でダイナミックな関係を表現しようというものである。例えば、最下段のタンクの穴を底面より少し高くしておけば少量の降雨では流量が増加しないという、ある種の非線形現象が再現できる。また、最上段のタンクの側孔を大きくしておけば大雨の時は水が大量に流れ出て、河川の水位が引くのも速いといった現象を再現できるようになっている。

このような極めて柔軟な構造を持つタンクモデルも柔軟であることからくる難点がある。そ

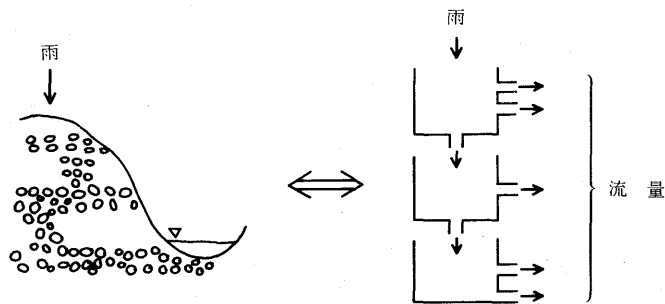


図1. タンクモデル

れは、タンクの数、穴の数、穴の位置、穴の大きさなどをどうやって決めるかの問題である。タンクモデルの決定は、普通、タンクモデルに精通した熟練者が河川流域の地形、地質等を考慮に入れてモデルを構成し、データを使って予測のパフォーマンスをチェックしながら少しずつ手を加えて改善していく。このような試行錯誤法は非常に時間のかかる作業であり、しかも普通の人にとってうまくいくかどうか保証のない方法である。このような場合、普通の人にも容易にできるよううまいモデリング法はないものかと考えるのが自然の流れである。一つの流れはタンクモデルのモデリングに習熟した熟練者のモデル導出の試行錯誤のプロセスを、そのままエキスパートシステムにプログラム化してしまうことである。実際、このような方向の研究は現在のように人工知能関連の言語やシステムが盛んになる以前に、早くからタンクモデルの研究者の間で始められている(菅原他(1977))。もう一つの流れはタンクモデルの統計モデル化である。タンクモデルと同等な数理モデルを考え、そのモデルのパラメータを統計的に推定する形に問題を組み立てることができるなら、統計学における最尤法、AICなどの手法によって誰でも容易に流量予測モデルにたどりつくことができる。

### 3. 貯蔵モデル

タンクモデルに対応する数理モデルには次のような貯蔵モデルと呼ばれるものがある。

$$(3.1) \quad \dot{x}(t) = f(x(t)) + r(t)$$

$r(t)$ は入力変量で降雨量に相当し、 $x(t)$ は出力変量で流量に相当する。 $f(x)$ は一般に非線形な関数である。この式によって定義される $x(t)$ がエルゴード性などの望ましい性質を持つためには $f(x)$ や $r(t)$ がある種の性質を持つことが必要とされるが、ここでは深く立ち入らない。 $f(x)$ が図2bのような線形関数の時、図2aのような単純なタンクモデルに同等になる。タンク内の水位を $z(t)$ とすると水位のダイナミックスを規定する式は

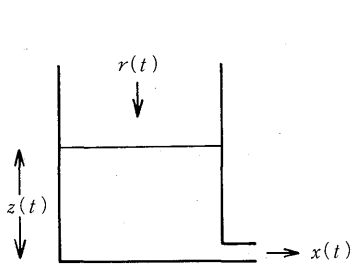


図 2a.

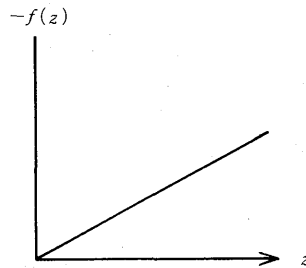


図 2b.

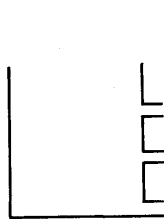


図 3a.

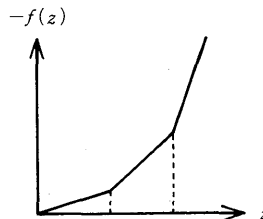


図 3b.

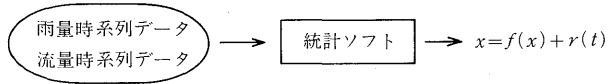
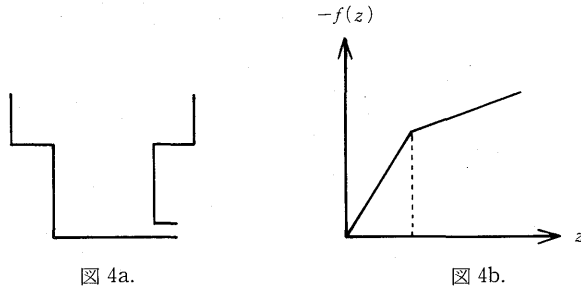


図 5.

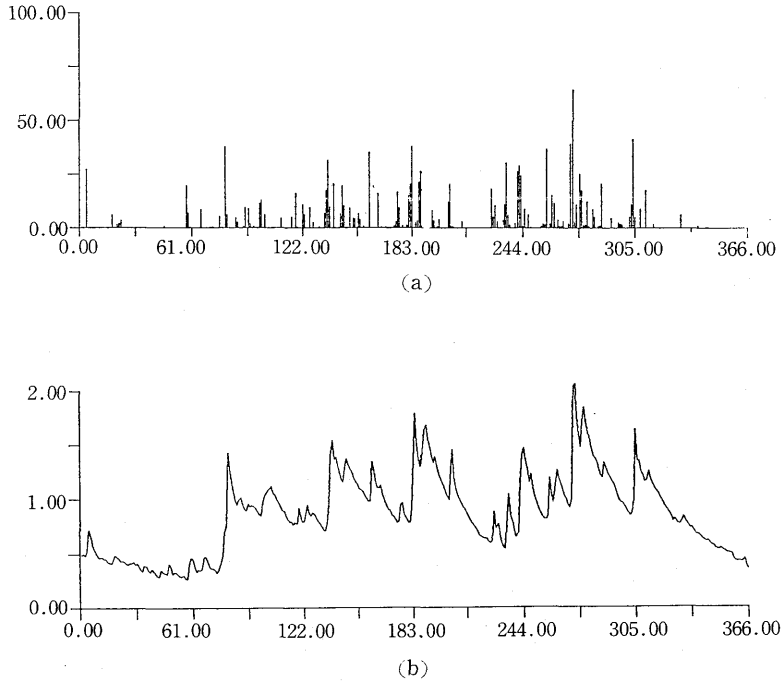


図 6. (a) 雨量, (b) 流量データの例

$$\dot{z}(t) = -az(t) + r(t)$$

流量  $x(t)$  は水位に比例する量として

$$x(t) = Cz(t)$$

となる。図 3a, 図 4a のようなタンクの場合は  $f(x)$  はそれぞれ図 3b, 図 4b のようになる。

この貯蔵モデルのパラメータを統計的方法によって図5のように過去のデータから推定することができれば、どんな未熟で未経験な人でも熟練者と同じか、時にはそれ以上のモデルが得られることになる。このような数理モデルの構造を推定するコンピュータプログラムは、まさに、ソフトウェアである。統計ソフトウェアという定義は人によって少しずつ違っているようであるが、本講ではこのようなソフトウェアを統計ソフトウェアと呼ぶことにする。このソフトウェアの具体的内容にはここでは立ち入ることができない。詳しくは Ozaki (1985) を参照されたい。

この貯蔵モデルの推定の場合、データ  $x(t)$  がダイナミックな性質を持っていること、 $f(x)$  が非線形であることなどが絡んで非線形最適化計算を行なわねばならず、大量のデータでパラメータ数も多い場合には大型の高速計算機なしでは殆ど不可能である。図6はこの統計ソフトウェアを適用した流量と雨量のデータ、図7は統計数理研究所の HITAC-M280H を使ってこのデータから推定した貯蔵モデルの関数  $f(x)$  の図である。この二つの図の関数はそれぞれ違

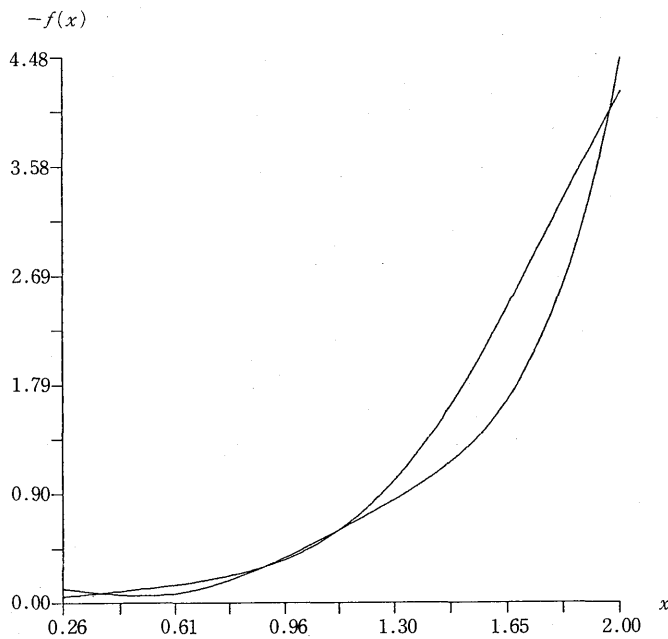


図7. 推定した関数

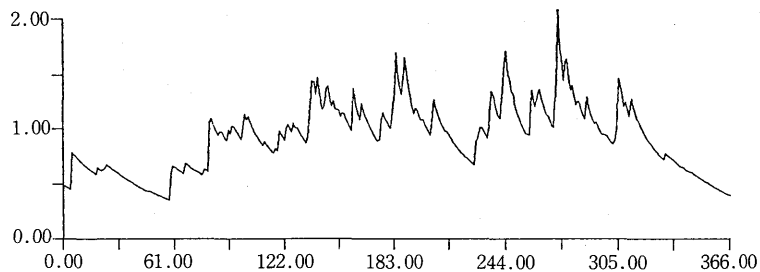


図8. 非線形モデルによるシミュレーション

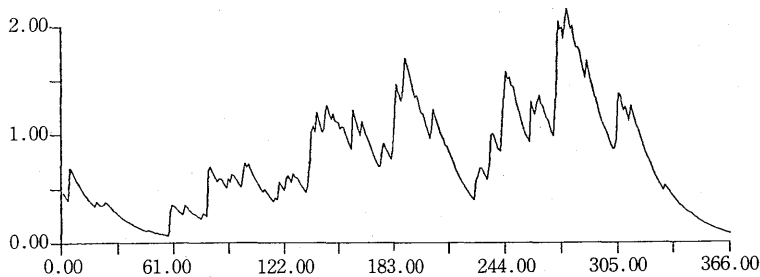


図9. 線形モデルによるシミュレーション

う非線形パラメトリックモデルによって得られた関数である。

図7はこの河川の水位が低い時と高い時で流出のパターンが大きく違うことを示している。また、そのことは推定したモデルのシミュレーションによっても、はっきり読み取ることができる。図8はこの推定した非線形貯蔵モデルに観測雨量を入力として入れた場合の流量、図9は貯蔵モデルの関数  $f(x)$  を線形関数に制限して推定した最適モデルを使って同じ観測雨量を入力として入れた場合の流量である。図9に比べて図8は水位が高い時、減衰の速度も速く、水位が低いと減衰の速度が遅くなる特徴を示している。

#### 4. 貯蔵モデルの統計ソフトウェアと関連する計算環境

データから貯蔵モデルのような数理モデルを推定する統計ソフトウェアはそれ自体大きな計算量を要するが、これが実際応用される場合を想定すると、これに関連して図10にあるようにソフトウェア使用前と使用後に二つの性格の異なった計算が必要となる。

統計ソフトウェア使用前に必要な計算は過去の膨大な観測データをコンピュータに貯え選択し、編集して統計ソフトウェアに使える形にする処理計算である。統計学者は普通、観測データ、すなわち統計データと考えがちであるが、これは実際にはむしろ全く別物と考えたほうがよい場合が殆どである。観測データは諸科学の個々の分野で計測されるが、データをとる現場の科学者やエンジニアにとっては計測すること自体が大変な大仕事であり、後の統計ソフトウェアによる処理を想定してデータを計測したりはしないのが普通である。データは観測の時間間隔が長すぎたり、短かすぎたり、また欠測値（しかも、回復不可能な大量の欠測値）があったり、異常値があったりといった様々な問題を含んでいる。このような観測データの中から適当な部分を選択し、関連の統計ソフトウェアに使えるタイプの統計データに処理編集することが行なわれる。この選択と編集の部分は計算自体は統計ソフトウェアの部分と独立しているが、現場の科学者、エンジニアと統計学者の両者の吟味検討が必要とされる重要な作業であるといえる。また、最終的に統計モデルに使われるデータは小規模でも、それを選び出し編集する過程で巨大な記憶領域を必要とするという点が多く、統計学者に見落されている点も注意しておきたい。例えば、雨量や河川の水位が一分間毎に計測されている場合、データ数は一日で1,440、一年で525,600になる。更に、計測地点が50地点あると26,280,000になる。したがって、一つのデータに4バイト使って、きっちり最大限詰め込んでも一年間のデータだけで100メガバイトを越える。これが数十年のデータとなると一つのデータセットだけで数ギガバイト必要になる。しかも、このような巨大なデータセットが数百、数千と貯蔵してあってスピードは少しくらい遅くてもいいから、いつでも容易に取り出して使える計算機システムが望まれるのである。

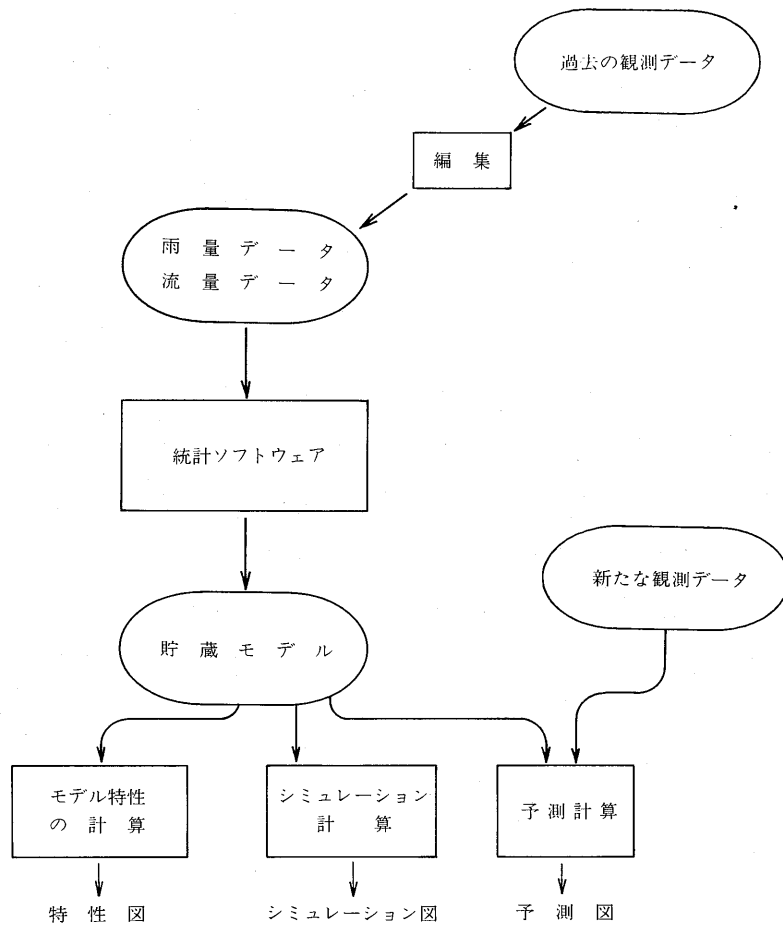


図10. 貯蔵モデルの統計ソフトウェアと関連する計算

もうひとつ、統計ソフトウェア使用後に必要な計算は以下のようなものである。統計ソフトウェアによって推定した貯蔵モデルを実際に使う場合、そのモデルのシミュレーションによる検討や、モデルを使って流量を予測し、実際の流量と比較するといったことをする。この場合に要求される計算は統計ソフトウェアの場合の科学計算と違って、得られたモデルの持つ種々の情報を“視覚化”するということがポイントになり、グラフィックス処理などのアプリケーションソフトウェアが必要となる。また、推定したモデルの確率的特性などを数式処理によって計算し、その特性をグラフにして見るといったことも必要になる。

以上述べたように統計ソフトウェアが実際に使われる場合を想定すると、使用前、使用中、使用后と三つの異なったタイプの計算が必要になる。使用前は大規模なデータの貯蔵、選択と編集計算、使用中は繰り返ループを多く含んだ大型の科学計算、使用后は科学計算の結果の視覚的利用の為のリスト处理的な計算である。過去においては、この三つの計算は全て大型計算機で行なわれるのが一般的であったが、近年、少なくとも三番目のリスト处理的計算は前二者とはっきり別れ、ワークステーションにその役割を課す方向にきているようである。しかし、これも最近は大規模計算機のオペレーティングシステムがUNIXに統一される方向にあるようで、そうならば多くのUNIX上で動く図形処理などのシステムが大型機で利用できるようになり、

以前のように全て大型機で行なうようになるのかもしれない。これらは計算機を生産するメーカーの方針に左右される部分大きい。

第一番目の大規模データの貯蔵，選択，編集計算は統計学の外の個々の諸科学分野でそれぞれが独自のやり方を工夫しているようであるが，統計学の中においてもそれを図 10 に示したように観測データから統計ソフトウェアをとおして最終的な統計情報の視覚的利用までの一貫した流れの中で捉え，将来の望ましい方向を打ち出し，具体化させていくことに努力を払う必要があるのではなかろうか。このような統計ソフトウェアが実際問題に有効に役立つトータルなシステムというものは，統計学だけが発達しても，ハードウェアだけが発達してもどうにもならない。それらを結んで全体的に機能させるソフトウェアがあって，はじめて成立するものであるが，筆者の関係した統計ソフトウェアを取り巻く計算環境は残念ながら非常にいびつで貧しく，その進歩は苛立たしいほど遅いといわねばならない。

### 参 考 文 献

- 菅原正己 他 (1977). 「タンクモデルの構造を自動的に決める計算機プログラムの開発」, 国立防災科学技術センター研究報告, 17, 43-89.
- Ozaki, T. (1985). Statistical identification of storage models with application to stochastic hydrology, *Water Resources Bulletin*, 21, 663-675.