環境システム研究 Vol. 24 1996年10月

都市域における雨天時負荷流出モデルパラメータの同定方法

IDENTIFICATION OF RUNOFF MODELS FOR NONPOINT POLLUTANT LOAD IN URBAN AREAS

張 昇平[1]・酒井 彰[2]

Sheng Ping ZHANG[1] · Akira SAKAI[2]

ABSTRACT: In this paper, we present a parameter identification procedure for the pollutant load runoff models in urban areas. The procedure consists of two steps : the parameter estimation based on the optimization method and the observation noise separation of the input-output data with the neural network model. This procedure is effective for most of the pollutant load runoff models in urban areas and expected to give a more reliable and stable model parameter estimation.

KEYWORDS : POLLUTANT LOAD RUNOFF MODEL, PARAMETER IDENTIFICATION, OPTIMAL PARAMETER ESTIMETION, OBSERVATION NOISE, NEURAL NETWORK MODEL

1．はじめに

都市域における下水道整備は著しく進捗したものの，都市河川や公共用水域の水質は大きな改善が見られない（環境庁、1995）．これに関連する主な要因の一つとして，雨天時に流出する面源汚濁負荷量が挙げられる（国松ら、1989；酒井、1996）．

雨天時に流出する面源汚濁負荷を制御・管理するためには，負荷流出モデルを用いて負荷の挙動を把握し，負荷量を予測・評価することが必要である．従来では，雨天時負荷流出メカニズムに対する理解が不十分であったことと，観測データが非常に少なかったこともあって，構造的に簡単なモデル構成とほとんど経験的に決められたモデルパラメータが用いられてきた（和田、1990）．雨天時汚濁負荷が注目されるようになってから，信頼性の高い観測がいくつも実施されており，これらの観測結果をもとに，地域特性や雨天時負荷流出過程をより正確に考慮できるモデルの作成が可能になりつつある．しかし，モデル構造が複雑になるにつれ，モデルパラメータは数が増えると同時に，質的にも分化し，経験的にすべてのモデルパラメータを決定することはほとんど不可能になり、観測データを用いたパラメータ同定が必要になる。

観測データに基づくモデルパラメータの同定は，数学的には逆問題を解くことになるが，それに際しての最も困難な課題として，逆問題の解が唯一である保証がないことと観測データに誤差が含まれることが一般的に挙げられる．本研究では，前者に対して負荷流出モデルによる解析精度の向上を最優先するとして最適化手法を用いて対応し，後者に対してニューラルネットワークモデルを用いて観測データから誤差成分を分離する一連の同定手法を提案する。

2．負荷流出システムのモデル化

雨天時における負荷流出過程は、場（流域）の条件と、負荷流出の原動力となる降雨の流出によって支配される。降雨量と流域の条件を入力とし、流（出）量と水質を出力として捉える場合に、水文学的には“負

[1] 名城大学都市情報学部(Faculty of Urban Science, Meijo University)

[2] 株式会社日水コン(Nihon Suido Consultants Co., Ltd.)

荷流出システム”と呼ばれる。

負荷流出システムをモデル化する場合に、二つの視点が考えられる。すなわち、システムの入出力関係のみに着目する見方と、システムにかかわるすべての要素（の関係）を考慮するものである。前者はblack-box的アプローチ、後者は力学的あるいは物理学的（physical-based）アプローチと呼ばれる。モデル化の手順は通常、

1) モデル構造の決定
2) パラメータの同定
3) モデルの評価

という三つのステップからなる（宝、1990）。

モデル化しようとする実際のシステムについてわれわれは先験的な情報を持っている。すなわち、システムの本質的特性、流域の特性、物理的あるいは水文学的知識などである。これらは実験なり観測を通じても得られるものであり、また先験的情報が実験・観測の方法を規定するというフィードバックが存在する。実際のシステムに対してどのようなモデルを組み立てるか（モデル構造の決定）は、これら先験的情報と実験・観測を通じて決定される一方、対象とする問題とモデル化の目的にも依存する。従来では、雨天時における負荷流出について、先験的情報が少なく、流出メカニズムに対する理解が不十分であったことと，実験・観測データが非常に少なかったこともあって，流域の特性や使用目的を考慮してモデルを選択することはできずに、構造的に非常に簡単な、いわゆるblack-boxモデルを多用されてきた。近年、公共用水域の水質に対する雨天時汚濁負荷の影響が注目されるようになってから，面源負荷流出過程に関する信頼性の高い観測が実施されるようになり、これらの観測結果をもとに，流域特性や雨天時における負荷流出メカニズムをより正確に理解できるようになり、いわゆる物理学的モデルの作成が可能になりつつある.

モデルの構造が決定されると、モデルに含まれるパラメータの値を推定しなければならないが、このとき通常二つの方法が用いられる、すなわち、経験的な推定値を用いる方法と、所与の入出力データを用いてなんらかの目的関数を最適化することにより推定する方法である。前述のように、従来では、パラメータ推定に使用可能な信頼できる観測データ（入出力データ）が少なかったことと、モデルの構造が簡単でパラメータの数も少なかったことが原因で、モデルパラメータがほとんど経験的に決められてきた。しかし，モデル構造が複雑になるに連れ，モデルパラメータは数が増えると同時に，質的にも分化し，経験的にすべてのモデルパラメータを決定することはほとんど不可能になり、観測データを用いたパラメータ推定が必要になる。

モデル評価とは、対象となる流域の特性と解析目的に対して、構造とパラメータの与えられたモデルが適切であるかどうかを何らかの基準を用いて検証することである。ここで、重要なのは評価基準の設定である。

本研究では、一般的なモデル構造に対して、パラメータの推定に焦点を絞り、パラメータ推定に当たっての基本的な考え方と推定方法、特に、観測データに含まれる誤差の取り扱い方について提案を行う。

3．モデルパラメータの同定方法

3．1 雨天時の面源負荷流出モデル

負荷流出モデルは、現実の複雑な負荷流出システムを理想化あるいは概念化し、数学的記法によって表現したものであり、一般的に下記の四つの基礎方程式の連立解として記述される。

1)雨水の連続方程式
2)雨水の運動方程式
3)負荷の連続方程式
4)負荷の運動方程式

雨水の連続方程式と運動方程式によって定義されるのはいわゆる（雨水）流出モデルであり、その結果（連立解）として求められるのは（雨水）流（出）量である。こうして求められた流量が通常負荷の運動方程式への入力となる。雨水の流出モデルは伝統的な水文学の中心的研究課題であり、ここでは、これについての

詳細な議論を省略する。

雨水流量が雨水流出モデルによって求められたとして、負荷の連続方程式と運動方程式は一般的に次のように表される。

$$\frac{dx(t)}{dt} = \Phi(x(t), I(t), L(t), c) \tag{1}$$

$$L(t) = H(x(t), q(t), c) \tag{2}$$

ここに、$x(t)$ は時刻 t における状態量、$\Phi(\cdot)$ と $H(\cdot)$ はスカラー関数、$I(t)$ は負荷供給量、$L(t)$ は負荷流出量、c は流域条件（初期条件を含む）を表すパラメータ、$q(t)$ は雨水流出モデルからの出力である雨水流量である。

負荷流出モデルで状態量としてよく用いられるのは、堆積負荷量である。連続方程式(1)の右辺において、状態量を含まない場合もあるが、負荷の溶出、沈殿、あるいは内部生産等を考慮する必要があり、これらが状態量に依存するとした場合に式(1)の表記になる。従って、一般的な表記としては式(1)が適切であると考えている。負荷供給量 $I(t)$ については、面源堆積負荷だけでなく、雨水に含まれる負荷や、対象システム外からの流入負荷を考える場合に、ベクトル変数として扱うことになる。

上記モデルの具体的な一例として酒井（1996）が提案したモデル以下にを示す。従来からよく使用されている土研モデルなども適切な変数変換により上記の一般的な表記に書き表すことができる。

$$\frac{dx(t)}{dt} = I(t) - L(t)\,, \quad \text{subject to } x(0) = x_0 \tag{3}$$

$$L(t) = k \cdot x(t)^m \cdot q(t)^{n-1} \cdot (q(t) - q_c) \tag{4}$$

ここに、x_0 と q_c は流域の条件を表すパラメータで、それぞれ初期堆積負荷量と限界流量を表す。また、k,m,n はモデル構造が与えられたときの関数形を決定するパラメータ（モデル定数とも言う）である。

この例からわかるように、負荷流出モデルにおいて観測データにより推定すべきパラメータは、流域条件を表すパラメータとモデル構造を決定するパラメータの二種類である。

なお、負荷流出モデルの分類上、上記モデルは集中型モデルに属する。負荷流出システムは、実際分布型システムである。このとき、モデルの扱い方に関して様々な困難な問題が起きるが、モデルの表記方法だけを考える場合には、流域を分割し、式(1)と式(2)に含まれるすべての変数を空間的位置の関数と見なせばモデルの表記方法を変える必要はない。また、分布型モデルを適切な集中化（lumping）の方法によって式(1),(2)のような形式に変換できる（高棹ら、1982）。そういう意味では、式(1),(2)によって表される負荷流出モデルは一般性を有している。

３．２ パラメータ推定問題の定式化

前述のように、モデルパラメータの求め方は、モデルの構造に依存して多様である。モデルの構造が簡単な場合に、先験的情報に基づいて経験的に決められるが、モデルの構造が複雑になるにつれ、パラメータの数が増えると同時に、質的にも流域の条件を表すものとモデルの構造を規定するものに分化する。このような場合に、数学的最適化手法を用いたパラメータ同定が有効になる。

最適手法によるパラメータ同定では、まず、目的関数を決めて、次に、決められた目的関数に応じて探索方法を選ぶ必要がある。負荷流出モデルでは、実測流出負荷量（水質）と計算流出負荷量（水質）との適合度を表す関数を目的関数に採用するのが一般的である。しかしながら、適合度の評価基準が必ずしも明確にはなっていない。たとえば、流出負荷量の全時間経過をあわせるのか、それとも、総流出負荷量をあわせるのか、あるいは、ピーク水質を合わせるのかいろいろと考えられる。これは、具体的な問題に際してモデルの目的に応じて決める必要がある。たとえば、流出負荷量が放流先水質に対する短期的影響を評価しようと

する場合には、流出負荷量の全時間経過を合わせるべきであり、逆に長期的影響を把握しようとする場合には総流出負荷量を合わせたほうが望ましいといわれる。また、流出水の処理計画を考える場合には、ピーク水質およびその出現時刻を合わせなければならないことになる。

適合度指標が決定されれば、次に、目的関数の関数形を決める必要がある。最も一般的な関数形として実測流出負荷量（水質）と計算流出負荷量（水質）との絶対誤差を採用する場合と自乗誤差を採用する場合の二つが挙げられる。どちらを選ぶべきかは、明確な基準はないが、モデルの精度に対する要求、モデル同定に使用するデータの信頼性、および計算時間等を考慮して総合的に判断する必要がある。たとえば、モデルの精度に対する要求が高く、使用するデータの信頼性も高い場合に、自乗誤差を採用したほうが望ましいといわれる。しかし、自乗誤差を採用した場合に、問題によっては探索時間が長くなることはあり、特に、パラメータの数が多い場合に、最適値の探索が不十分に終わってしまうことも考えられる。

負荷流出モデルで、全時間経過にわたる実測流出負荷量（水質）と計算流出負荷量（水質）と自乗誤差を目的関数とした場合に、これを次式で表すことができる。

$$R=\sum_{t=1}^{T}\ [\hat{L}(t)-L_o(t)]^2 \tag{5}$$

ここに、R は目的関数、t は時間、T はパラメータ同定に使用するデータ期間（データ数）、$\hat{L}(t)$ は時間tにおける計算流出負荷量、$L_o(t)$ は時間tにおける実測流出負荷量である。

最適化によるパラメータ同定手法は、計算流出負荷量 $\hat{L}(t)$ を同定すべきパラメータベクトルの関数と見なし、目的関数 R をパラメータベクトルに関して最小化することである。つまり、最適化問題

$$\min_p R(p)=\min_p\sum_{t=1}^{T}\ [\hat{L}(p,t)-L_o(t)]^2 \tag{6}$$

を解くことによりモデルパラメータを求める。ここに、p は同定すべきモデルパラメータである。式(3),(4)で示された負荷流出モデルを例に言えば、$p=(x_0\ \ q_c\ \ k\ \ m\ \ n)$ となる。

上記の最適化問題の解析手法は負荷流出モデルの構造に応じて選ぶ必要がある。事例研究では再急降下法を用いた。

4．観測データ誤差の除去

流出負荷量の観測データ時系列 $L_o(t)$ が与えられた場合に、前節で述べた最適化手法によってモデルパラメータを求めることができる。言うまでもなく、得られた結果の信頼性は観測データの信頼性により決まる。つまり、観測データに誤差が含まれる場合に、最適化手法はむしろ不適切なモデルパラメータを結果として求めてしまう危険性はある。従って、観測負荷量系列から如何に誤差成分を取り除くかが推定したパラメータの信頼性を高める鍵となる。

負荷量の観測データに誤差が含まれる原因として以下のものが考えられる。

まず、観測技術による純粋な観測誤差が必ず存在する。ここで言う観測技術とは観測機器は勿論のこと、負荷の算定方法、分析方法、サンプリング方法等の違いも含む。

次に、負荷流出システム自身に含まれる誤差も存在する。これは最も本質的な誤差発生原因の一つである。具体的には、時間の経過とともに、降雨や流域条件等が変化し、負荷流出過程そのものがランダムな変動特性を示すようになる。

さらに、対象システム外からの負荷量の流入も誤差をもたらす原因として挙げられる。負荷量流入の原因としては、管渠の誤接や特殊な水理条件（逆流、越流等）が考えられる。

また、先験的情報が不足であるため、原因不明の負荷源が存在することもある。このほかに、たとえ負荷源の存在がわかっていても、モデル構造上の制約により考慮しきれない場合もある。

負荷量の観測データに誤差をもたらすこれらの原因の中に、十分な事前調査によって原因を究明し、誤差を取り除くことができるものもあるが、どうしても取り除けないものが必ず残る。そういう意味では、観測データにおける誤差の存在が本質的で、避けて通れない問題である。本研究では、ニューラルネットワークモデルにより観測データから誤差成分を分離する方法を提案する。

階層型ニューラルネットワークモデルは、入出力間の関係を学習により求めることができる。ここでは、その学習の特性を利用して観測データの誤差成分を分離する。以下その基本的考え方およびアルゴリズムについて述べる。

説明を簡単にするため、ニューラルネットワークの構造を一個の出力ユニットを持つ3層構造とし、入力層と中間層のユニット数をそれぞれN、Mとする。入力ユニットへの入力が$(I_i\ ,\ i=1,2,\cdots,N)$であるとき、中間層および出力層の各ユニットからの出力値は次のように計算される。

$$\text{中間層}:Y_j=f(X_j)\quad ,\qquad X_j=\sum_{j=1}^{N}\omega_{ij}I_i+\theta_j\ ,\qquad j=1,2,\cdots,M \tag{7}$$

$$\text{出力層}:O=f(Z)\qquad ,\qquad Z=\sum_{j=1}^{M}\omega_jY_j+\theta \tag{8}$$

ここに、$f(\cdot)$は入出力間数、ω_{ij}は入力層ユニットiと中間層ユニットjとの結合係数、θ_jは中間層ユニットjのしきい値、ω_j中間層ユニットjと出力層ユニットとの結合係数、θは出力層ユニットとのしきい値である。入出力関数$f(\cdot)$としては様々な関数形が提案されており、中でもシグモイド（Sigmoid）関数

$$f(X)=\frac{1}{1+\exp(-X)} \tag{9}$$

が最もよく使われている。

理論的には、上記のネットワークモデルによって、中間層のユニットを必要なだけたくさん使えば、ユニット間の結合係数および各ユニットのしきい値を適切に設定することによって任意の連続関数を任意の精度で近似することができる（麻生、1988）。

上記のニューラルネットワークモデルを負荷流出量に含まれる誤差の分離に適用するときに、まず、負荷流出量に非線形な自己回帰関係、

$$L(t)=r(L(t-1),L(t-2),\cdots,L(0)) \tag{10}$$

を仮定し、$(L(t-1),L(t-2),\cdots,L(0))$と$L(t)$をそれぞれニューラルネットワークモデルの入力変数と出力変数とし、その関係$r(\cdot)$を流出負荷量の観測データを用いた学習により求める。事例研究では、モデル構造を3層構造とし、入力層のユニット数を対象流域の流達時間に対応して3つ、中間層のユニット数を学習速度を参考に試行錯誤により27とした。

ニューラルネットワークモデルの学習方法として、ここでは、最もよく使用されている誤差逆伝播法（error back propagation）を採用する。以下そのアルゴリズムを示す。

T個の教師データ

$$(I_1^{(t)},I_2^{(t)},\cdots,I_N^{(t)},O^{(t)})=(L(0),L(1),\cdots,L(t-1),L(t))\ ,\ t=1,2,\cdots,T \tag{11}$$

が与えられているとする。まず、結合係数およびしきい値の初期値$\omega_{ij}^{(0)},\omega_j^{(0)},\theta_j^{(0)},\theta^{(0)}$を任意の値に設定する。これらの値を用いて式(7)、式(8)により教師データの入力$(I_1^{(t)},I_2^{(t)},\cdots,I_N^{(t)})$に対応する出力を計算し、

$\{U^{(t)}, t = 1,2,\cdots,T\}$ と記す。これを教師データの出力と比較し、次式に示す誤差関数を定義する。

$$R^{(0)} = \sum_{t=1}^{T} (O^{(t)} - U^{(t)})^2 \tag{12}$$

明らかに、R は結合係数およびしきい値の関数である。誤差逆伝播法では、誤差関数 R が最小になるように結合係数およびしきい値の初期値をを逐次修正していき、修正量を決定するのに非線形計画法の一つである再急降下法を用いる。

上記の学習法の特徴の一つはいわゆる過剰学習である。すなわち、学習回数がある限界を超えるとネットワークモデルは入出力間の本質的な関係だけでなく、教師データ（負荷量の観測データ）に含まれるノイズまで学習してしまうことである。観測データから誤差成分を分離するのにこの性質を利用する。具体的には、入出力間の関係の学習がほぼ終了し、ノイズの学習が始まる前に学習プロセスを打ち切る。このように学習回数をコントロールすることにより得られたネットワークモデルからの出力が誤差成分がほぼ除去されたものとなる。以下、学習プロセスを打ち切る基準について述べる。

まず、観測データをランダムに二つのセットに分ける。一つは教師データとしてネットワーク学習用に、もう一つは学習プロセス検証用に使用する。学習回数毎に、検証用データセットを用いて誤差関数 R の値を計算し、これを学習データセットによる誤差関数の値と比較する。最初、学習が進むに連れ、両方とも減少していく。しかし、ある時点を越えると、学習データセットによる誤差関数の値が引き続き減少するのに対し、検証用データセットによる誤差関数の値が増加に転じる。この時点を入出力間の本質的な関係の学習が終了し、誤差の学習に入っているとし、学習プロセスを停止する。

5．まとめ

本研究では、都市域における雨天時負荷流出モデルの一般的なモデル構造に対し、モデルパラメータの同定方法を提案した。本手法は、観測データを用いたパラメータ最適推定法とニューラルネットワークモデルによる観測誤差の分離方法から構成されており、従来多用されてきた経験的パラメータ推定法に比べてより信頼性の高いモデルパラメータ推定値が得られるものと考える。本手法を用いた事例研究では、流出水質（COD）の再現精度が平均 11%も上昇した。本手法の大きな特徴はモデルパラメータ同定に使用する観測データから誤差成分を分離したことにある。観測データをそのまま使用して得られたモデルパラメータと誤差を分離したデータを使用して得られたモデルパラメータを比較した事例では、パラメータの変動係数が最大 31%、平均 15%も減少し、観測データから誤差成分を分離した場合にはより安定したパラメータの推定値が得られることが示された。

負荷流出モデルパラメータの同定方法の確立は，より信頼できる負荷流出解析モデルの構築や解析精度の改善に役立つだけでなく，困難な雨天時負荷流出観測の計画作成にも示唆を与えるものとして期待される．

【参考文献】

1) 環境庁編：「環境白書」、1995。
2) 国松孝男・村岡浩爾：「河川汚濁のモデル解析」、技報堂出版、1989。
3) 酒井彰：下水道による雨天時汚濁負荷制御に関する研究、京都大学学位論文、1996。
4) 和田安彦：「ノンポイント汚染源のモデル解析」、技報堂出版、1990。
5) 宝　馨：水資源システムにおける確率論的モデルと手法の評価に関する研究、京都大学学位論文、1990。
6) 高棹琢馬・椎葉充晴・宝馨：集中型流出モデルの構成と流出予測方法、京都大学防災研究年報、第 25 号 B-2、pp.221-243、1982。
7) 麻生英樹：「ニューラルネットワーク情報処理」、産業図書、p.17、1988。