

GMDH による水質予測モデルの構成

田 中 雅 史*

Construction of Prediction Models of the Water Quality in a
Stream with the Application of GMDH

Masafumi TANAKA

Faculty of Agriculture, Mie University

Abstract Models for predicting the quality of water in a stream are constructed with the group method of data handling (GMDH), which is said to be useful for identification of complex systems. This method is an empirical heuristic self learning algorithm using information on the system.

In this paper, starting from linear prediction models which describe dispersion phenomena of a solute matter in a stream, the structure and parameters of the models are estimated using numerically simulated data and water quality (tracer concentration) data experimentally obtained. With simulated data it is shown that GMDH is applicable to these models, though it is difficult to find an exact model structure in the case of noisy data. A close correspondence of predicted values with observed values in experimental data indicates that this method is useful for real time water quality management in a stream. Several important features of algorithms are also discussed from the view point of the successful application of GMDH to the water quality system.

I. ま え が き

流水系における溶解物質や浮遊物質の伝バ拡散過程を表現し、流下に伴う水質変化を予測するモデルには拡散式に基礎をおく物理モデルとフィルター理論に代表される非物理モデルがある。物理モデルはモデル中のパラメーターのもつ物理的意義が明確であり、変化過程の本質的な見通しが立つという点では計画段階のシミュレーションによる予測や評価においては有用であろう。しかしながら、モデル中の拡散係数や分散係数のようなパラメーターが現状では十分評価出来ない点や現実の複雑な種々の要因がモデル化により除外されているといった点で不満足である。一方、非物理モデルではパラメーターのもつ意味は不明確なものの、複雑な要因を総合的に変数やパラメーターの形でとり入れる余地を残している。よって、現存する水質監視システムの管理を行う際にはむしろ有利な面もあり、モニタリングステーションからえられる生の情報をベースとすれば、より現実に即したモデル構成が可能である。また、いずれも何らかの形でパラメーターの推定を行うという点では両モデル共そう隔りはないと考えられる。

本文では水質管理の立場から、水質変化の予測モデルを非物理モデルである線形予測モデルとして GMDH に

より構成する。ここで採用した GMDH の手法はモデルの構造とパラメーターを同時に推定するという意味で従来の方法と異なっている。この手法により筆者らも木曾川の水質データを解析し一応の成果をえたが¹⁾、最近では種々の分野で基礎的、応用的研究が行われそれと共に手法のもつ欠陥が指摘されアルゴリズムの改良が行われてきている²⁾³⁾。ここでは、モニタリングステーションからえられる水質の情報をベースに比較的短時間の将来の水質を予測するためのモデルを構成することを念頭において、まずシミュレーションデータにより GMDH によるモデル構成が可能かどうか、また適用に当たっての注意すべき諸点について検討する。ついで、開水路流れの幾分簡単化された分散実験からえられたトレーサーの濃度変化のデータに適用する。

II. 予測モデルの構成

流れ系における水質変化過程を表現する代表的モデルである移流分散方程式は次式で与えられる⁴⁾。

$$\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = \frac{\partial}{\partial x} E \frac{\partial c}{\partial x} + \phi - \psi \dots\dots\dots (1)$$

ここに、 c : 流路断面について平均化した水質濃度、 u : 平均流速、 ϕ : 排出項、 ψ : 自浄作用等による反応項、 E : 混合、拡散に関するパラメーターで

$$E = E_D + K_x + D \dots\dots\dots (2)$$

により構成される。ここに、 E_D : 分散係数、 K_x : 乱流

*三重大学農学部

キーワード: 環境の予測, 汚染濃度, 集中制御装置システム, 多変量解析, プログラミング手法

拡散係数, D : 分子拡散係数であり, 通常 $E_D \gg K_x \gg D$ により $E \approx E_D$ とみなせる。

いま排出源を表す強制項 ϕ に対し,

$$\phi(t, x) = f(t) \cdot g(x) \dots (3)$$

と時間項と空間項の積の形で表現し, 反応項については

$$\phi = h(x) \cdot c \dots (4)$$

なる一次反応で表す⁵⁾。(2) 式の

$$c(t, a) = c(t, b) = 0 \dots (5)$$

$$c(0, x) = 0 \dots (6)$$

なる境界条件および初期条件のもとでの解は放物形偏微分方程式の一般理論より基本解 $U(t, x, \xi)$ を用いて,

$$c(t, x) = \int_0^t d\tau \int_a^b U(t-\tau, x, \xi) f(\tau) \cdot g(\xi) d\xi \dots (7)$$

と表現できる⁵⁾⁶⁾。特定の地点 $x=d$ における水質濃度の時間的変化を

$$c(t, d) = y(t) \dots (8)$$

とすれば, (7) 式から

$$y(t) = \int_0^t d\tau \int_a^b U(t-\tau, d, \xi) f(\tau) \cdot g(\xi) d\xi \dots (9)$$

をえる。

$$K(t, \tau) = \int_a^b U(t-\tau, d, \xi) g(\xi) d\xi \dots (10)$$

とおけば,

$$y(t) = \int_0^t K(t, \tau) f(\tau) d\tau \dots (11)$$

となり, 基本解をえることにより任意の排出源に対応する地点 d における水質濃度の時間的変化が求められる。しかしながら, 基本解をえることは一般には困難である。そのかわりに (11) 式において, $f(t)$ と $y(t)$ を観測して積分核 $K(t, \tau)$ を逐次求めて予測を行うことが考えられる。

また, 水質変化過程のモデルとして変化過程を連結した混合槽により表現するモデルがある。田村はこのようなモデルとして分布ラグモデルを提案している⁷⁾。分布ラグモデルでは第 i リーチの水質濃度 C_i の時間的変化は次式で表される。

$$\begin{aligned} \frac{dC_i}{dt} = & a_i(t) \cdot C_i + a_{i-1} \int_0^t \phi_{i-1}(\tau) \cdot C_{i-1}(t-\tau) d\tau \\ & + b_i u_i(t) + e_i \quad (i=1, \dots, n) \dots (12) \end{aligned}$$

ここに, a_i , b_i は i 番目の槽の容量, 流量および反応係数に關係するパラメーター, ϕ_i : 分布ラグ係数, u_i : i 番目の槽の排出に關する項, e_i : システムノイズ。

(11) 式や (12) 式を離散的な形で書けば, k 時刻の水質濃度はそれぞれ次の式で表される。

$$Y_k = \sum_{i=0}^N K_i^k \cdot C_{k-i} \dots (13)$$

$$C_i(k) = A_{0i} C_i(k) + \sum_{j=1}^N A_{i-1,j} \cdot C_{i-1}(k-j)$$

$$+ u_i(k) + e_i(k) \dots (14)$$

(13) 式や (14) 式はいずれも線形予測モデルの形をしており

$$y_k = a_0^k + \sum_{i=1}^N a_i^k \cdot x_i \dots (15)$$

と表現出来る。よって, 水質濃度の観測値を用いて, 次数 N (モデル構造) やパラメーター値を決定することにより将来の水質濃度を予測するモデルが構成出来る。なお, 上述のモデルは最もらしい考慮のもとでえられたものであり現実の現象を表現しているのか否かについても検討する必要がある。ここでは, これらの構造推定とパラメーター推定の問題を GMDH の適用により検討する。

GMDH のアルゴリズムには提案者の A. G. Ivakhnenko 自身のものでも数種類, 他の研究者の改良を含めれば統一した方法はないといってよい。最も基本的な初等アルゴリズムについては池田, 楳木により詳しく解説されており⁸⁾, 筆者らも 2, 3 の適用例を示した⁹⁾。以後の展開上, アルゴリズムについて簡単に述べておく。システムの入力変数 (x_1, x_2, \dots, x_m) と出力変数 y の間に

$$y = f(x_1, x_2, \dots, x_m) \dots (16)$$

なる関数関係を仮定する。入力変数のうちには実際は y と無関係の変数があっても良い。具体的な (16) 式の式形, すなわちシステムの完全表現式を決定するために一連の入力変数と出力変数のデータの組を準備し, これらのデータを Training Data (T. D) とよばれる部分表現式 (中間変数)

$$y_k = f(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j \dots (17)$$

$$(k=1, 2, \dots, m(m-1)/2)$$

の係数値を最小 2 乗法で決定する組と, 部分表現式の出力変数への適合の度合を判定するための Checking Data (C. D) の組に分割する。T. D を用いて入力変数のすべての組合せに対し, (17) 式の係数値を決定し, この中間表現式により C. D に対する中間変数 y_k を求める。ついで, C. D に対する平均 2 乗誤差

$$\varepsilon_k^2 = (y_k - \hat{y})^2 \dots (18)$$

を計算し, ε_k^2 の小さい y_k を M 個選択し 1 回目の計算を終える。これら M 個の中間変数の組合せに対して再び (17) 式を適用して以下同様の計算を繰返し, 各回の最小の平均 2 乗誤差が前回のそれより大きくなった時に計算の終了とする。最も平均 2 乗誤差の小さい式をもってシステムの完全表現式とする。すなわち, この方法は一度にシステムの表現式を作るのではなく, システムに關係すると考えられる入力変数の 2 個づつの組合せで試行錯

誤により精度の良いモデルを構成する多層化構造をもつアルゴリズムとなっている。

なお、部分表現式としては(17)式よりも完全2次式

$$y_k = a_0 + a_1x_1 + a_2x_j + a_3x_i^2 + a_4x_ix_j + a_5x_j^2 \quad \dots\dots(19)$$

を用いるのが一般的であり、(19)式によれば、層化に伴って式形の非線形性が高くなる。このことは GMDH 法の1つの特徴ともなっているが、本文では線形モデルをまず対象としているので(17)式を採用する。GMDH を適用するに当たって、

- 1) データの T.D と C.D への分割法
- 2) 各層で選択する中間変数の個数
- 3) 部分表現式の形
- 4) 入力変数の入力方法

といった点が問題となる。精度の良いモデルをえるためには上述の諸点に対する配慮が必要であり、かつアルゴリズムの欠陥もこれらの点から派生している。一方、筆者の考えでは、GMDH によるモデル構成における手法上の利点は忠実に情報に依存して、かつ簡単な計算から情報の示唆するモデルの構造をもとの情報よりも整理した形で提示するところにあると思われる。このような考えに基づいて、ここでは出来るだけ初等アルゴリズムに近い形で適用し、入力方法の改良を除いては大きな改変は行わなかった。

III. シミュレーションデータの検討

1. 解析の方針

水質予測モデルの構成を行う前に GMDH で正しいモデルが見つかるか否かを検討する。このため、入力としては後に使用する実験データからえられたトレーサー濃度の時間的変化の系列 $X_i(i=1, \dots, N)$ を用い、これ

に既与のフィルター

$$Y_k = \sum_{i=1}^7 \alpha_i X_{k-i} \quad \dots\dots(20)$$

をかけて出力を作成する。ここに N はデータ数、 α_i : 7 項からなるフィルターの係数値で、その値は Table 1 の値を採用した。また、Fig. 2 にはその形を、Fig. 1 にはデータの入、出力の性状の1例を示す。

問題は『生成された出力変数と入力変数の系列をもちいて、既与のフィルターの次数および正しい係数値を探すこと』にある。入力変数は当時刻より11単位時間前までの12個を用いる。すなわち、 k 時刻の出力変数 y_k を

$$y_k = f(x_k, x_{k-1}, \dots, x_{k-11}) \quad \dots\dots(21)$$

とおく。よって、入力変数のうち、5変数は実際の出力変数と無関係である。なお、今後、構造のみを表す時には簡単に x_{k-i} を単に x_i と書く。

データの総数は50個とし、T.D と C.D への分割は奇数番目のデータを前者に、偶数番目のデータを後者に入れ交互にとる。各層での中間変数の選択は、2, 3 の予備的検討により4個とした。入力方法については、2層以下の入力変数を中間変数のみから構成するのではなく Fig. 3 に示すように常にもとの入力変数を保持し、選択される中間変数と共に入力する。この理由は必要な変数が1層で選択されないことにより構造が見付からない場合があるためで、後に層化のプロセスにおいて考察する。

2. 解析結果

データ性状の異なる11ケースについて行った結果によれば、11ケースのうち8ケースは精度の差こそあれ10～12層前後の計算で正しいフィルター形に収束した。2ケースについては正しい変数は見つけたが係数値が異なりパラメーター推定が出来なかった。残りの1ケースについては1変数が脱落して構造推定も出来なかった。これ

らの結果を Table 1 にのせる。No.

1 は早い層で既与のフィルターに収束する例であり、係数値を小数第3位で四捨五入すれば正しい値となっている。No. 2 と No. 3 はフィルターの係数値が既与の値と異なるケースであり、No. 4 は変数 x_6 が脱落したため他のフィルター形に収束している。No. 2～4 ではいずれも最も寄与の大きい変数 x_2 の係数値を正しい値よりも大きく評価していることが注目される。No. 2～4 のケースについて全データを T.D と C.D にして再度計算を行った結果、No. 4 についてはやはり見付くせなかった。

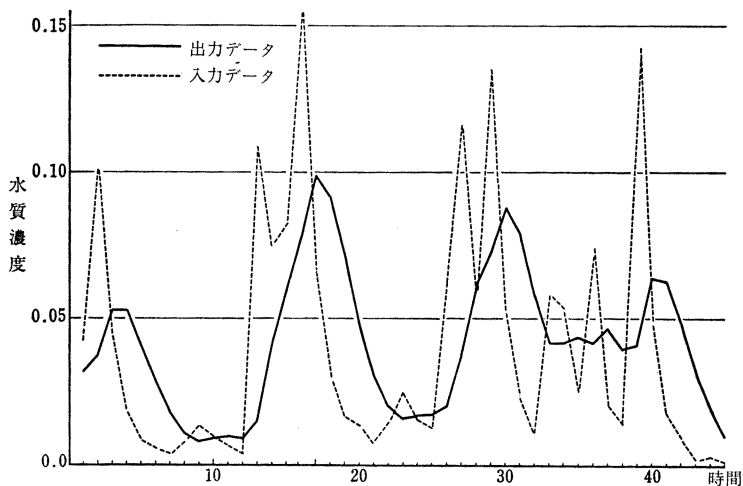


Fig. 1 シミュレーションデータの性状の一例

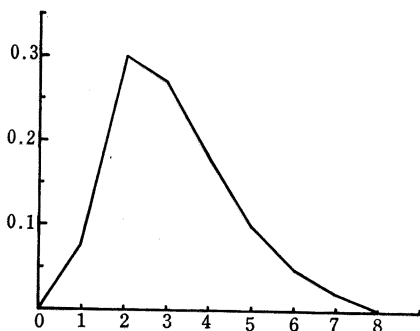


Fig. 2 フィルターの形

一方, No. 1 についてデータ数を25と元の半分に減らしこれらを T. D と C. D に分割して計算した結果 (No. 5) は前と同様正しいフィルター形に収束した。これらの結果から, GMDH は入, 出力変数の性状に左右され, いわば情報に内在する“質”に依存し, “量”によるのではないことがわかる。

Table 1 シミュレーションデータによる計算結果
—GMDHによりえられるフィルターの係数値—

データNo.	No.1	No.2	No.3	No.4
α				
α_1	0.08	0.0792	0.0732	0.0507
α_2	0.30	0.2992	0.3380	0.3632
α_3	0.27	0.2701	0.2207	0.2075
α_4	0.18	0.1803	0.1694	0.2076
α_5	0.10	0.1010	0.1681	0.1112
α_6	0.05	0.0491	0.0097	0.0327
α_7	0.02	0.1996	0.0174	0.0271
定数項	0.0002	0.0004	0.0006	0.0001

次に層化のプロセス, すなわち探求の過程について検討する。層化のプロセスは上述のようにデータ性状により各ケースごとにすべて異なるがその典型的な例をFig. 4 に示す。図における数字は変数の番号を表し, 丸印はその層の計算が新しい変数の導入にあることを示し, 四角印は係数の調整を行う計算であり, 矢印はその変数の番号に相当する係数値の調整方向を示す。同図より層化のプロセスをたどると, まず第1層においては数値的に出力に対し最も寄与の大きい入力変数 x_2 や x を中心とした選択がなされる。次いで第2層ではこれらの変数から作られた中間変数の組合せにより精度を上げる。この時には元々ちがった入力変数により構成された中間変数が結合し, 同一変数で構成された互によく似た中間変数は選択されない。第3層以下では係数値の調整と新しい入力変数の導入が繰返し行われるが, 層化が進むと係数の調整のみを行う様になる。

また, Fig. 5 は C. D に対する各層における平均2乗誤差を示したものである。正しいフィルター形に収束す

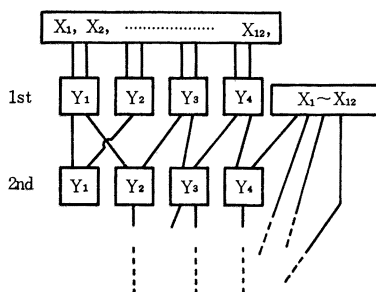


Fig. 3 GMDH入力方法

る場合には層化と共に誤差が減少するのに対し, 見付けだせないデータについてはある程度の層化で誤差の減少率が低くなることがわかる。

これらの層化のプロセスの特徴をまとめると次の様になる。

(1) 構造推定とパラメーター推定はそれぞれ違った層で行われ, 新しい変数の導入は前層までの変数群による構造で最大限の精度をあげて後に行われる。すなわち, 構造推定→パラメーター推定→調整のサイクルを繰返す。

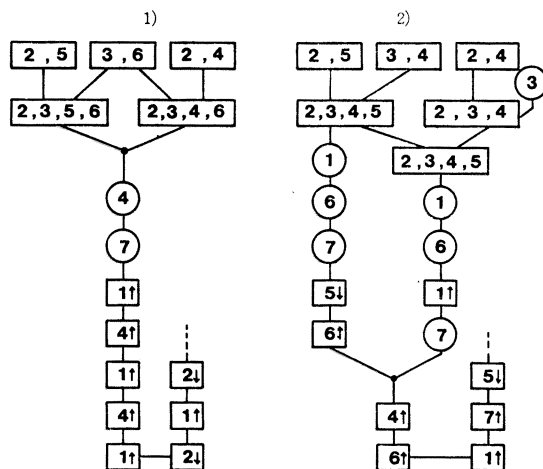


Fig. 4 層化のプロセスの例

1) データ No. 5 2) データ No. 1

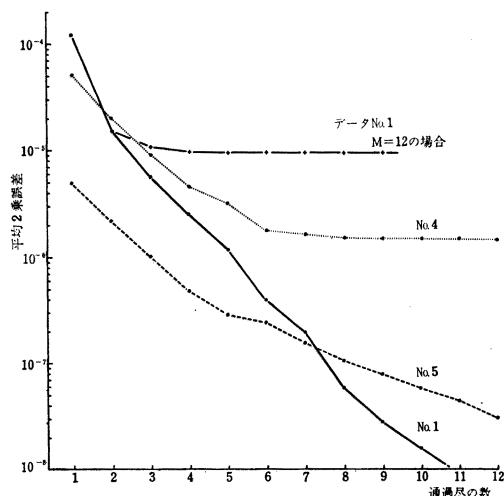


Fig. 5 層化による誤差の減少

(2) 数値的に寄与の大きい変数から導入されるので、早い層における表現式では数値的に寄与の大きい変数の係数値は正しい値よりも更に大きく評価している。したがって、新しい変数の導入により係数値は下がる傾向にある。

(3) (2) と関連し、出力変数に対し寄与の小さい入力変数は第1層や第2層で選択される機会が少なくなるため、2層以下で中間変数のみで入力変数を構成すると最終的に正しい推定が行われないことがある。すなわち、変数入力の方法はその変数の出力に対する寄与と関係しているがあらかじめ解らない。よって、ここで採用した様に常にもとの入力変数を保持する方がよい。たとえば No. 1 のケースについて選択数を入力変数の数と同じにして中間変数のみから構成した場合、Fig. 5 に示したように4層で収束にむかう変数 x_7 が脱落して正しいフィルター形は見付け出せなかった。

(4) 層化が進むにつれて、表現式の構造が固定化されて一方方向に進展していく。このため、新しい変数の導入による効果は既に出来上っている構造全体におよぶため係数値の改変、すなわちパラメーター推定は数層にわけて行われ、改良の数値の大きい自体も小さくなる。

(5) 早い層においては、その層の中間変数のうち最小の平均2乗誤差の構造が次層で選択されるとは限らないし、また最終表現式に含まれるわけではない。むしろ、特定の入力変数の出力変数に対する寄与が大きい場合、早い層で良い近似をえるが、正しい探索方向をむいていないと有害となる場合すらある。この例が Fig. 5 に示した No. 4 の場合である。

(6) 各層における中間変数の構成は層の進化につれて、次の様に行われる。

$$Z = a_1 Y_i + a_2 Y_j \dots \dots \dots (22)$$

$$a_1 + a_2 \approx 1.0$$

ここに、 Z は新しい層における中間変数、 Y_i 、 Y_j は前層における中間変数、 a_1 、 a_2 は係数。係数値の和は1に近く2つの中間変数に重みをつけ、かつ定数項は無視しうる程小さくなる。このことは以前にも指摘したが¹⁾、計算途中において係数値の変化から大よその計算内容や探索が正しい方向にわかっているかが判断出来る。

従来、上述のようなGMDHの層化のプロセスについての考察は十分行われていない様であるが、今後のアルゴリズムの改良点についてもいくつかの問題を提示している。

以上の検討はいずれも、元の情報に観測誤差などのノイズを含まない場合であった。実際の観測値には誤差が含まれている。そこで、No. 1 のデータについて原データにそれぞれ平均値ゼロ、標準偏差 $=1.0 \times 10^{-3}$ 、 $5.0 \times$

10^{-3} 、 1.0×10^{-2} の正規ノイズを付加した場合の3ケースについて検討した。なおノイズのレベルを原データの変動の標準偏差との比で示せば、それぞれ4.2%、21%、42%である。解析結果を Table 2 に、層化における平均2乗誤差を Fig. 6 に示す。

図には比較のために誤差のない場合ものせてある。いずれの場合も正しい変数は選択したが係数値については最も付加ノ

Table 2 データに観測誤差がある場合
えられるフィルターの係数値

誤差 レベル α_i	標準偏差 $S=0.001$	$S=0.005$	$S=0.01$
α_1	0.08	0.0855	0.1035
α_2	0.30	0.2956	0.2969
α_3	0.27	0.2725	0.2607
α_4	0.18	0.1837	0.2012
α_5	0.10	0.0948	0.0807
α_6	0.05	0.0484	0.0450
α_7	0.02	0.0165	0.0173
定数項	≈ 0	-0.0009	-0.0013

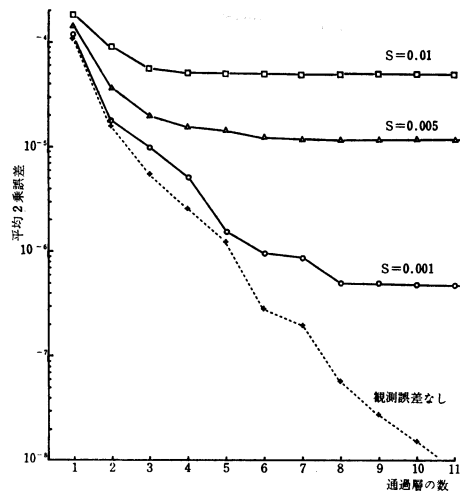


Fig. 6 データに観測誤差がある場合の層化による
誤差の減少 (データ No. 1)

イズの小さい場合を除いて正しい値とは差があり、定数項についてもゼロとはみなされない。また、Fig. 6 によれば層化のプロセスにおいてもノイズのあるデータではそのレベルに応じて改良が進展しないことがわかる。すなわち、層化が進むにつれて本来のノイズレベルと選択基準となる平均2乗誤差のレベルが同程度となると、より以上の改良は望めない。

なお、ここでの問題の設定は既与のフィルター形を見出しうるかということにあったが、単なる推定に限れば、いずれの場合も十分満足すべきものであることはいうまでもない。

IV. 実験データの検討

1. 実験概要と解析の方針

実験室で行った分散実験データをもとに予測モデルを構成する。実験は次の様に行われた。実験水路として、長さ 21.6 m, 幅 50 cm, コウ配 1/1000 の長方形断面水路を用い、水路上流端から 10 m の 1 地点とそれより適当な距離だけ離れた地点の 2 測定にトレーサー（塩水）濃度測定用の電導度計をそれぞれ水路中央、1/2 水深の位置に設置する。水路の路床は 5 mm 格子の金網を張付けた状態と粒径 5 mm~10 mm の砂利を一面に敷詰めた状態の 2 条件とした。所定の流量のもとで上流側測点よりさらに上流の任意断面において原濃度ほぼ 1.5‰ の塩水をビニル製トイで流水断面に一樣に注入する。実験時の水利条件、測定地点の距離間隔、路床条件を Table 3 に挙げる。トレーサーの塩水の塩分濃度変化による電導度の変化を測点 1 と測点 2 の 2 台の電導度計で検出し計録計に出力する。出力記録を 0.1 sec ごとに数値化し、検定曲線にしたがって塩分濃度値に変換する。これらのデータをさらに 2 sec ごとに平均化して解析のためのデータとする。作成されたデータの解析の単位時間スケールはえられた実験曲線のパターンを考慮して適当に設定したもので本来の分散現象のスケールとは無関係である。よって、II で検討したシミュレーションデータと実験データは第 1 に観測誤差があること、第 2 に解析の単位時間が必ずしも現象の物理的關係を明確にする上で必要とする解析時間に一致していない点で異なる。Fig. 7 はこのようにしてえられた実験データの 1 例である。なお、バックグラウンドの塩分濃度はわずかであるが差引いてある。

GMDH の適用の方針はⅢとほぼ同様である。1 実験からえられた 50 個の全データのうち、20 個に GMDH を適用し 20 個のデータすべてを T. D と C. D にしてモデルの同定を行

Table 3 解析データの試験条件

Ex. No	流量 l/s	平均水深 cm	流下距離	路床条件
1—(1)	4.99	4.85	200.	金網張り
1—(2)	"	"	300.	"
1—(3)	"	"	400.	"
2—(1)	9.03	6.24	200.	"
2—(2)	"	"	400.	"
2—(3)	"	"	600.	"
3—(1)	6.95	6.22	200.	砂利敷
3—(2)	"	"	400.	"
4—(1)	9.13	7.0	200.	"
4—(2)	"	"	400.	"

った。残り 30 個のデータは単にえられたモデルの予測精度の検討のみに用いた。各層での選択数は 4 個とし、部分表現式は (15) 式を前提として (17) 式の線形式を用いた。実験データは実験の設定条件から明らかなように定常流のもとで行われている。一般には同定すべき構造とパラメーターは汚染源のパターンの変化、流況変化、あるいは流水中の反応による物質変化に対応して非定常性をもっているが、ここでは汚染源ないし上流の濃度変化に対応する変化のみを考える。よって、問題は、上述の制約のもとで上流地点のトレーサー濃度パターンから下流地点の濃度パターンを予測するモデルを同定し、予測モデルの精度と (13) 式や (14) 式の水質予測モデルの可能性を検討することにある。

2. 解析結果とその考察

入、出力関係の表現において次の 2 つのモデルを設定する。

$$y_k = f(x_k, x_{k-1}, \dots, x_{k-11}) \dots (23)$$

$$y_k = f(x_k, x_{k-1}, \dots, x_{k-11}, y_{k-1}) \dots (24)$$

(23) 式は入力変数を当時刻より 11 単位時間前までの上流地点のトレーサー濃度値で構成した式で、(13) 式を予測モデルの前提においている。以後これを M-1 と表す。(24) 式は入力変数を上流地点の測定値以外に下流地点の当時刻より 1 時間前の測定値を入れて構成した式で、(14) 式を予測モデルの前提としている。これを M-2 として表す。

Table 4 に妥当と考えられるモデルの精度に関する解析結果を挙げる。表における精度は平均 2 乗誤差 ε^2

$$\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots (25)$$

で示す。ここに、 n は同定データもしくは予測データの

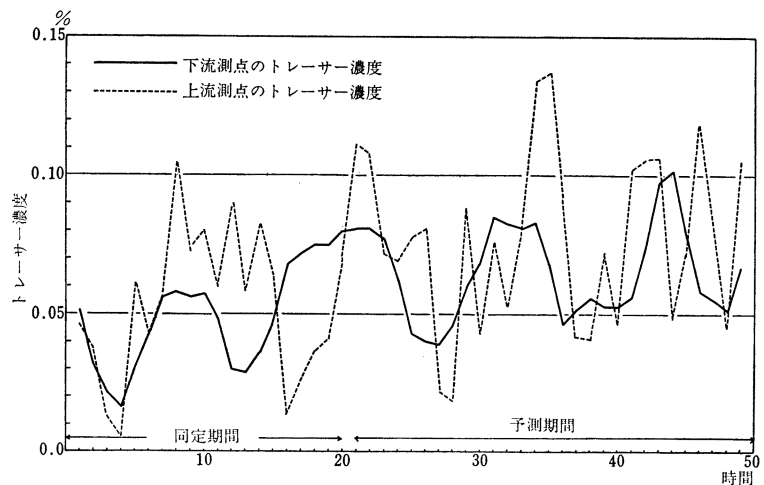


Fig. 7 解析データの一例 データ No.2 (3)

個数, y_i : 観測値, \hat{y}_i : 完全表現式, すなわち予測モデルによる同定ないし予測値。

相関係数は

$$r = \frac{\sigma_{Y\hat{Y}}}{\sigma_Y \cdot \sigma_{\hat{Y}}}, \quad \sigma_{Y\hat{Y}} = \frac{1}{n} \sum Y_i \cdot \hat{Y}_i \quad \dots\dots\dots (26)$$

で与えられる。ここに, Y_i, \hat{Y}_i はそれぞれ観測値, 予測値からそれぞれの平均値を差引いた値で, $\sigma_{Y\hat{Y}}$ は共分散値, $\sigma_Y, \sigma_{\hat{Y}}$ はそれぞれ観測値, 予測値の分散値, すなわち,

$$\sigma_Y^2 = \frac{1}{n} \sum_{i=1}^n Y_i^2, \quad \sigma_{\hat{Y}}^2 = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i^2 \quad \dots\dots\dots (27)$$

G はモデル感度を表すパラメーターで,

$$G = \sigma_{\hat{Y}} / \sigma_Y \quad \dots\dots\dots (23)$$

で与えられる。これらの精度の評価にしたがえばM-1, M-2共に相関係数もかなり高く, 感度パラメーター G もほとんど1.0に近く観測データの変化パターンにモデルによる変化が追従していることがわかる。また, 測点間の距離が長くなると, すなわち流れの中での分散時間が長くなるとモデルの精度は悪くなる。これは, 分散過程は応答関係のみから見れば平滑化過程であるので距離が長くなるとノイズを含む情報からは正しい構造やパラメーターの同定が困難となるためであろう。平均2乗誤差については同定精度の方が予測精度よりも良く, 同定データに対し *over fit* している。よって, 同定データに過剰に適合しすぎるという欠陥は克服されていない。Fig. 8~10 は No. 2 による観測値と予測値を比較した例であり, 観測値と予測値の差はピーク付近や低い値の所で大きい。これらが平均2乗誤差に反映しているのであろう。また,

Table 4 モデルとその予測精度

データ No.		同定精度	予測精度	相関係数	G
モデル 1	1—(1)	6.873×10^{-6}	1.637×10^{-5}	0.991	0.944
	1—(2)	7.067×10^{-5}	3.368 "	0.979	0.957
	1—(3)	1.910 "	1.729×10^{-4}	0.980	0.868
	2—(1)	5.553 "	1.006 "	0.986	0.839
	2—(2)	3.316 "	1.738×10^{-5}	0.992	1.001
	2—(3)	2.320 "	4.761 "	0.931	0.884
	3—(1)	1.223×10^{-6}	8,567 "	0.990	0.963
	3—(2)	7.136×10^{-5}	4.652 "	0.942	1.058
	4—(1)	5.493 "	2.188 "	0.989	1.024
	4—(2)	3.481 "	5.493 "	0.975	1.018
モデル 2	1—(1)	9.700×10^{-6}	1.835×10^{-5}	0.990	0.952
	1—(2)	3.238×10^{-5}	1.681 "	0.990	0.960
	1—(3)	1.967 "	6.014 "	0.989	0.910
	2—(1)	3.167 "	4.515 "	0.992	0.925
	2—(2)	1.507 "	2.383 "	0.990	1.001
	2—(3)	1.932 "	1.639 "	0.974	0.949
	3—(1)	9.559×10^{-6}	2.877 "	0.996	0.969
	3—(2)	1.118×10^{-5}	1.649 "	0.982	1.061
	4—(1)	3.789 "	3.378 "	0.983	1.002
	4—(2)	1.640 "	2.777 "	0.989	1.029

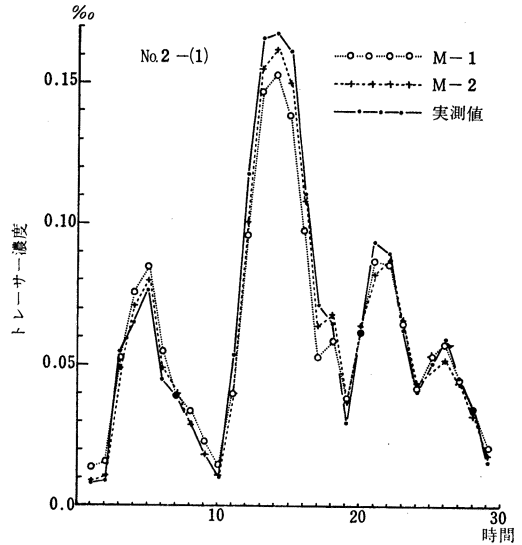


Fig. 8 予測結果 (1)

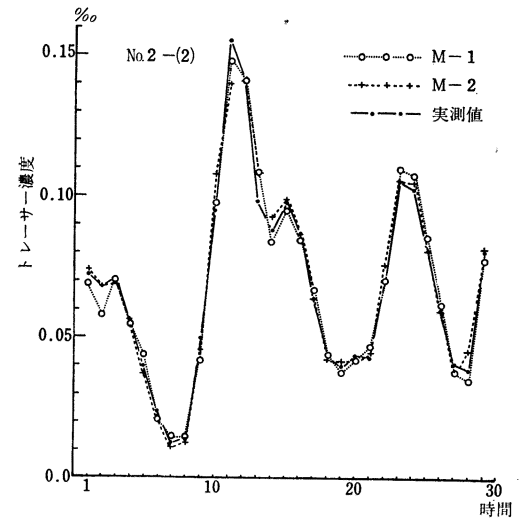


Fig. 9 予測結果 (2)

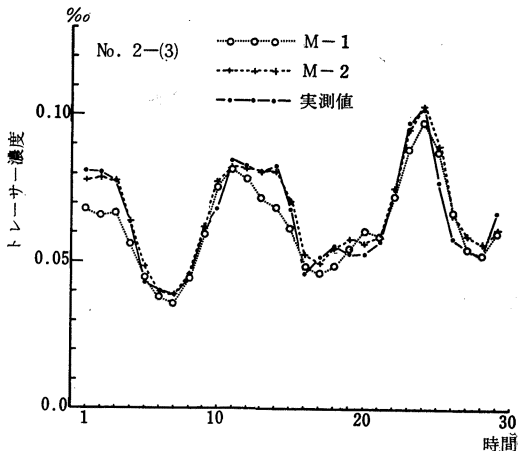


Fig. 10 予測結果 (3)

M-1 と M-2 を比較すると、一般に M-2 の方が M-1 に比して精度が良いが測点間の距離が短かいと大差がなく、距離が長くなると M-1 よりも優れている。これは変化パターンからすれば、下流測点の1時間前の値の導入により主として、濃度変化の上昇部分よりも低減部分について M-1 よりも予測の改良がなされるためである。

次に、GMDH の層化のプロセスにおける精度の改良の状況を示すと Fig. 11~12 の様になる。なお、図では $\sqrt{\epsilon^2}$ で示してある。図で丸印は M-1 に三角印は M-2 に対応し、それぞれのモデルに対して同定精度と予測精度の変化を示してある。図より実験データにおいてはシミュレーションデータの場合層化の進展がなく、かつ層化に伴う大きな誤差の減少がないことがわかる。この原因の1つはえられたモデルの精度は良好であることを考慮すると、実際予測モデルが少数の入力変数で構成されるためである。すなわち、実験データの方がⅢで与えたフィルタよりも少数の項で構成されているのであろう。また、精度は同定精度の方が予測精度よりも高く、かつ同定精度が層化につれてわずかながらも小さくなるのに対して予測精度は多くの場合4層を超すと逆に劣化して劣化する。すなわち、層化につれて、モデルの構造が数値的に適合しすぎて予測モデルとしての一般性を失う。この意味で、現実観測誤差をもつシステムでは最小の平均2乗誤差をもつ構造が必ずしも最適構造とはいえない。したがって、GMDH の適用に際しては対象とする現象についての何らかの Heuristics (発見的規範) の導入や観測誤差レベルを考慮に入れた選択規範の導入が必要であろう。

さて、同定精度と予測精度の差異を構造推定とパラメータ推定の立場から考察する。同定データ、予測データについて、それぞれ最良の精度のモデル、いいかえると完全表現式を構成する入力変数群は多くの場合一致している。中には一致しない場合もあるが予測精度の良い完全表現式の方が物理的観点からみても妥当な形式であるのは興味深い。No. 1 (3) のデータについて、データを2分して同定、予測の役割を交換すると、前半のデータを同定データとした場合

$$y_k = 0.011 + 0.1558 x_{k-9} + 0.1759 x_{k-8} + 0.3238 x_{k-7} + 0.1083 x_{k-6} \dots \dots \dots (i)$$

なる完全表現式がえられ、後半データからは

$$y_k = 0.011 + 0.1515 x_{k-9} + 0.2419 x_{k-8} + 0.3255 x_{k-7} + 0.1452 x_{k-6} \dots \dots \dots (ii)$$

なる完全表現式がえられる。この例では式を構成する変数は同じであるが、 x_{k-4} と x_{k-6} のパラメータ値が異なっている。もとのデータのもつ観測誤差のレベルは前半、後半はほぼ同様と考えられるので結局同定精度と予測精度の差は係数値が誤差のある同定データに適合し過ぎるためと考えられる。すなわち、パラメータ推定の立場からすれば一意的モデルはえられない。この意味で、原データに誤差があれば GMDH においても他の数値ベースの非物理的方法と同様、少くとも初等アルゴリズムではその欠陥はまぬがれてない。これらの検討から対象とした分散実験データについての予測モデルは構造はほぼ固定し、パラメータは若干変化することがうかがえる。

次にえられた予測モデルの構造、すなわち完全表現式について物理的な観点から検討する。えられた完全表現式を Fig. 2 と同様フィルタの形で横軸に時間、後軸に係数値をとり Fig. 13 に示す。横軸の数値は遅れ時間を示し縦軸は係数値である。完全表現式中の定数項については省略してある。また、参考のため、Fig. 14 に瞬間源の実験からえられた分散曲線の形を No. 3 (2) の場合について示してある。

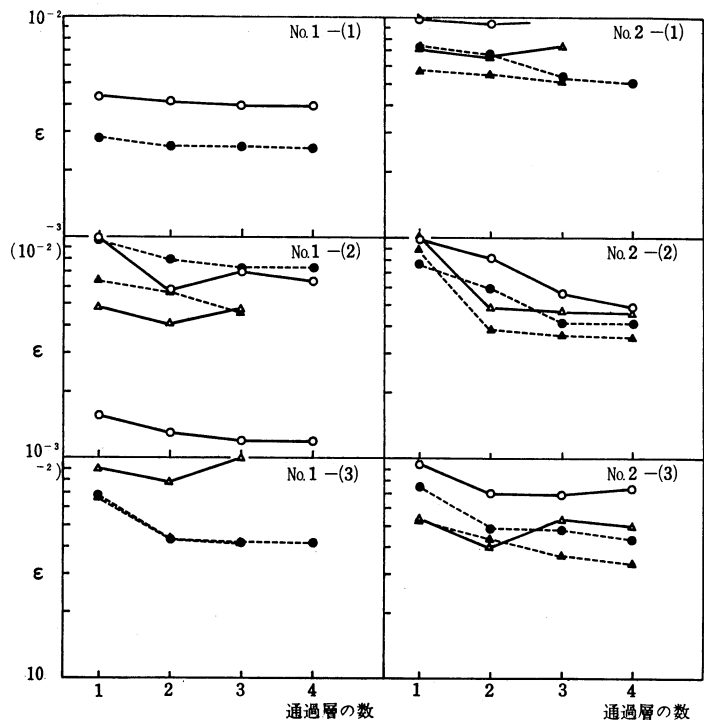


Fig. 11 層化による誤差の減少 (1)
丸印: M-1, 三角印: M-2, 白抜き: 予測

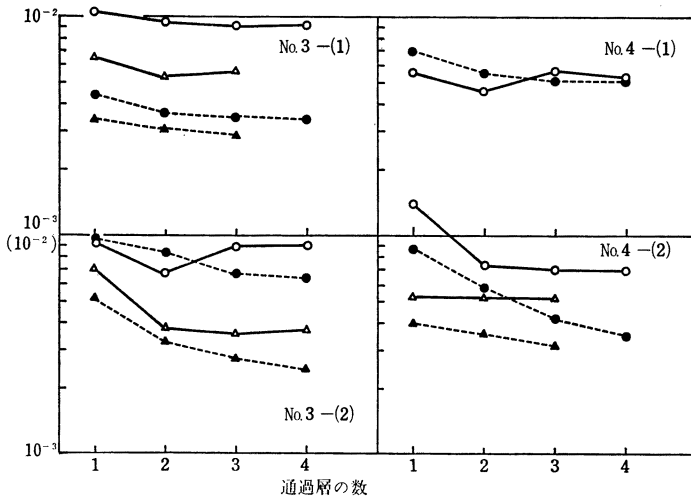


Fig. 12 層化による誤差の減少 (2)

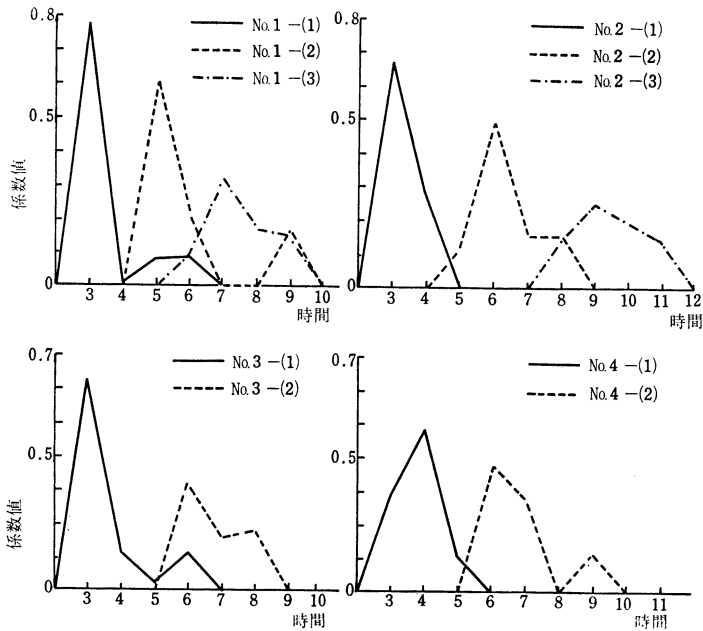


Fig. 13 予測モデルの係数値

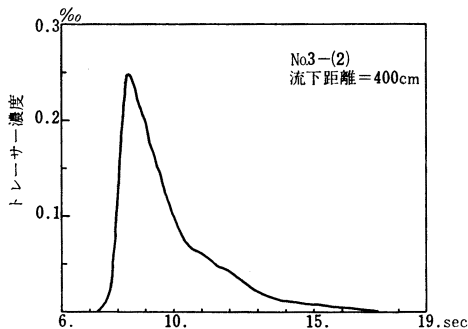


Fig. 14 瞬間源による分散曲線の例

Fig. 14 に示すような形が物理的意味をもつ構造であろう。このような物理的に妥当かつ一意的な構造が存在すると仮定して、Fig. 13 を見れば、次の様な点が指摘できる。第1に特定の変数の寄与が大きいこと、第2にNo. 1 (3), No. 3 (1), No. 4 (2)に見られるように1度ゼロの項となった後に裾の部分に小さな係数値の項が表れ、いわば2山型となることである。第1の点は前にも述べたように解析の単位時間スケールに比し分散現象の拡がりが大きくないことと、IIIで述べた問題点(2)により関係の大きい入力変数のパラメーター値をさらに大きく評価することによって考えられる。第2の点については、1つはそれぞれのデータの変化パターンの特徴を反映したものともいえるが、主として次の事情によるものと考えられる。GMDHにおいて、(17)式を中間表現式とした場合、層化の過程では2つの中間変数を基底として出力変数を2次元空間に射影することに相当している。2つの基底の直交条件は表立っては表されていないが相互に独立性のある中間変数が選択される機会が多い。このため、選択状況によっては出力変数に対し、中程度の寄与をする入力変数が捨たるか、あるいは主要な表現式が構成されて後に導入されるために必ずしも物理性という意味では正しい構造がえられない。以上の状況は原データに誤差がなくとも層化の過程で観測誤差の選択規範に対する影響がなく、かつ解析の単位時間スケールが現象の物理性を明らかにしうる程のスケールであれば解消されるであろう。しかしながら、2つの条件を満足することは不可能である。

一方、情報処理の点からすれば、結果は上流地点での観測情報を集約して下流地点での濃度変化を精度よく予測出来ることを示している。その過程において、(13)式や、(14)式で表される分布遅れモデルが予測のための物理モデルとして妥当なことが示唆されている。

V. あとがき

水質管理の立場から、えられる観測情報をもとに

GMDH を適用して水質予測モデルを構成した。その結果、流れの分散過程の物理的な表現式から導かれる線形予測モデルが有用であることが示された。しかし、GMDH による予測式の構造やパラメーターは物理モデルの構造やパラメーター自体を表現しているのではないことが明らかになった。GMDH は数値をベースにして情報を予測に役立つように集約することにより良好な結果をえる方法として位置付けるべきである。

また、アルゴリズムの最も大きな特徴である層化のプロセスについて検討し、その問題点を指摘して適用上注意すべき諸点について考察した。とりわけ、予測の意味での最良のモデルは同定による最良のモデルと一致しないことが重要な点である。いまのところアルゴリズムを複雑化するよりも解析対象についての知識や Heuristics を利用する方が得策と考えている。

今後の発展としては、流れの非定常性や非一様性に対する配慮、直接的には構造とパラメーターの非定常性を考慮する必要がある。そのためには、ここに述べた GMDH を逐次的に適用するか、ある程度構造のみを固定してパラメーターをカルマン・フィルターや他の適応制御方式により更新することが考えられる。これらの点については現実の流れの観測値への適用を含めて検討中である。

最後に、本研究の遂行に際し、日頃ご指導、ご鞭撻をいただいている本学松下玄教授に深く感謝いたします。なお、本研究の一部は文部省科学研究費（一般B、代表者 松下玄）の交付を受けたのでここに記し謝意を表します。

引用文献

- 1) 田中雅史, 板垣博: 水質データへのGMDHの適用—木曾川の水質の時間的変動特性—, 農土論集 **74**, pp. 32~39 (1978)
- 2) Ivakhnenko, A. G.: Present State of the Theory of Computer-Aided Self-Organization of Mathematical Models (Survey), Soviet Automatic Control, 8-5, pp. 18~26 (1975)
- 3) 田村坦之, 近藤正: 最近のGMDHの方法論と応用, オペレーションズ・リサーチ **2**, pp. 104~111 (1978)
- 4) 合田健編: 水質工学(応用編), 丸善, pp. 245~255 (1976)
- 5) 池田三郎: 分布定数系の同定理論とそのアルゴリズム, 計測と制御, 16-10, pp. 38~47 (1977)
- 6) 伊藤清三: 偏微分方程式, 培風館 (1977)
- 7) 田村坦之: 分布ラグモデルによる河川水質のモデリングとラグ係数の逐次推定, システムと制御 **20-5**, pp. 253~261 (1976)
- 8) 池田三郎, 榎本義一: GMDH (発見的自己組織化法) と複雑な系の同定・予測, 計測と制御 **14-2**, pp. 185~195 (1975)
- 9) 田中雅史, 福岡順, 鳥居清司: GMDH (発見的自己組織化法) の概要と適用例, 農土誌 **45**(8), pp. 539~545 (1977)

[1978. 6. 22. 受稿]

[この論文に対する公開の質疑あるいは討議(4,000字以内, 編集委員会あては1979年12月25日まで受け付けます)]

水田作土の収縮挙動より見た工学的性質に関する研究

河野 英一

水田作土の収縮挙動およびそれと工学的性質との関連を検討した。シロカキ状態のペースト土の収縮は、流動沈下、正規収縮および残留収縮というパターンになる。収縮挙動とコンシステンシーとの関係を明らかにできた。また、 q_e -含水比曲線において、水平収縮開始時の含水比で q_e が急上昇することが明らかとなった。さらに、水田作土のキ裂は水平収縮開始時と発生の指標と考えることができる。(農土論集 第81号)

キーワード

水田、土壌構造、収縮(土壌)、フッターベルグ限界、土の乾燥

集中豪雨による改良山成畑の崩壊について

福岡 喜弘・宮崎 毅・井上 久義

昭和51年17号台風に伴う集中豪雨により、造成後10年未満の改良山成畑が多数崩壊した。崩壊の実態と原因を明らかにするため、当時の降雨状況、災害現場の地形地質、農地の造成経歴、土層構成と土壌の物理性などを精査した。その結果、谷を埋めた盛土が崩壊しやすかったこと、盛土の下に原生林の植生を埋め込んでいるためスベリ抗抵が小さいこと、その植生を通して浸透水が集中しやすいことなどを推定し、これに応じた防災対策を指摘した。(農土論集 第81号)

キーワード

土石流、山崩れ、刈払い、土壌断面、粘着力、内部摩擦角、安定係数

局所磁場からみた吸着水の特性

佐藤 裕一

土壌吸着水中の水素原子核が、その周辺に存在する、他の水素原子核の核スピによって与えられる磁場の大きさを測定した。その磁場は局所磁場と称する。吸着水分子層数によって、吸着水中水素原子の受ける局所磁場に相違のあることから、吸着水を大別することができる。さらに、局所磁場の温度変化から、吸着水分子の運動・停止温度がわかった。また、吸着水と氷との磁気的特性を比較し、両者に相違のあることもわかった。(農土論集 第81号)

キーワード

土壌水、pF、分子、電磁波

GMDHによる水質予測モデルの構成

田中 雅史

水質管理の立場から、観測情報をもとに比較的短時間の将来の水質を予測するモデルをGMDHを適用して構成した。流れの分散方程式から帰着される線形予測式を対象にし、シミュレーションデータと分散実験データを用いてモデルの構造推定とパラメーター推定を行い良好な予測モデルをえた。GMDHによる予測法を、数値をベースにして情報を集約し予測に役立てる方法として位置付け、有効なモデル作成上注意すべき問題点を指摘した。(農土論集 第81号)

キーワード

環境の予測、汚染濃度、集中制御装置システム、多変量解析、プログラミング手法

降雨確認後の予備放流

佐藤 政良

日本の多目的ダムにおける予備放流方式は降雨量予測に基づいて運用されることになっているが、検討の結果、予測と実績値の適合性はきわめて悪いことが分かった。そこで、降雨を確認した後、流出の遅れを利用して実施する新しい予備放流方法の原理と手順を提唱し、岐阜県・横山ダムと山形県・荒沢ダムで、過去の洪水に対して予備放流操作のシミュレーションを行い、提唱した方法の有効性を検証した。(農土論集 第81号)

キーワード

水利用計画、多目的ダム、予備放流量、雨量観測施設、有効雨量

サイホン式排液器について

村上 康蔵

従来の越流式排液器では、栽培槽の底敷ビニルシートから漏液の心配があるところから、市販の塩ビ管によるサイホン式排液器を考案してこの問題を解決するとともに、この排液器に必要な諸元の決定方法と、タン液深の制御、槽内の養液の入れ替わり、全量排液等についても越流式と同様な機能をもつことを示したものである。(農土論集 第81号)

キーワード

内水位、排水量、サイホン余水吐、水耕栽培、流量係数