【解　説】

計測と制御
Vol. 14, No. 2
(昭和50年2月)

GMDH（発見的自己組織化法）と複雑な系の同定・予測

池田　三郎*・樬木　義一*

はじめに

過去200年，科学技術の発展は主として科学の専門分化に伴い，簡単な系や，より単純な要素に細分化できる系を対象として，そこでの法則性を見出すことによって行なわれてきた．しかしながら，現在，われわれが解決をせまられている重要な問題，たとえば

（a）生物（人間）の大脳皮質系の活動
（b）人間社会の経済活動
（c）環境問題
（d）生態系の諸問題

など，これらは単に変数が多いという意味だけでなく，動的で相互に結合しあっているので，「複雑さ」を避けることはできず，従来の個別的，専門科学の方法で，数理科学的取扱いのできる物理モデルとして作成することは非常に困難である．仮りにできたとしても，現実の計画制御のために実用に供することはほとんど不可能なものが多い．最近まで科学はこのようなシステムの研究をさけ，単純なものに還元できるようなシステムにその焦点をしぼってきた．

「複雑さ」を取扱う方法の中で重要でかつ一般的なものはサイバネティクスである．サイバネティクスはWienerによって「生物と機械における通信と制御の一般的科学」と定義された[1]．サイバネティクスは，その本質的な複雑さによって現在のところ，われわれの攻撃の外側にいる生物的，心理的，社会的，経済的な諸問題を解決するための有効な方法を与えてくれるように思われた．しかしながらその後，サイバネティクスに関する非常に多くの研究がなされたにもかかわらず，真に複雑な系や，多くの変数と制約をもった制御問題は

（i）定式化の困難さ
（ii）高次元の呪い

のために，実用となる程度に有効に解決された例はあまりない．サイバネティクスを電子計算機とその応用に関する面に限定したとき，**図1**にこれまでの研究対象（左半面）と今後の研究対象（右半面）が模式的に与えられている．現在のサイバネティクスはWienerの最初の思想が忘れられ，図1の右半面のような問題を解くための方法論は，より固有で，より高度に専門化し，その専門家の狭い範囲内でしか手がとどかなくなっている．サイバネティクスの特別の効用はきわめて複雑ではあるが，無視するにはあまりにも重要なシステムを科学的に取扱う方法を提供する点にある．この意味でサイバネティクスは複雑なシステムに対して，その系に固有の方法でなく，一般的でかつ，柔軟な方法の提供を要求されている[2]．

A. G. Ivakhnenko[3~14]によって開発されたGMDH (Group Method of Data Handling) は，著者によれば，まさにサイバネティクスの一般性の回復の要請に答えるものであり，発見的自己組織化 (Heuristic Self-Organization) の原理に基づいて本質的に複雑なシステム，すなわち

a）非常に多くの変数とパラメータの存在（高次元）
b）相互の関係が非線形
c）原因と結果，入力と出力の関係を見出すことが

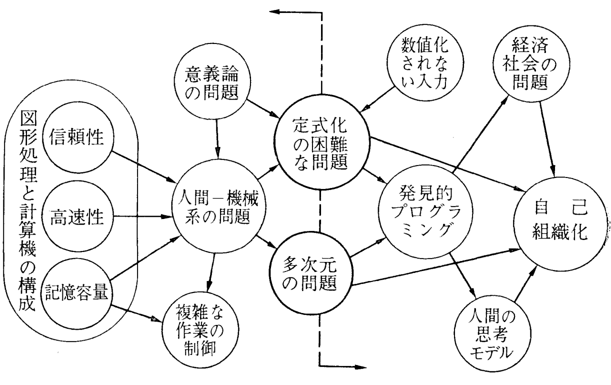

図1　工学的サイバネティクスにおける基本的なテーマ

* 京都大学　工学部

原理的にも実際的にも不可能な系を取扱う一般的方法である．

ここではわが国ではまだ充分に普及していないGMDHの基本的な思想とそのアルゴリズムについて解説する．さらに内外の研究ならびに応用例をいくつか挙げ，最後に著者らが河川の流量予測問題に適用した例を述べる．

1.　複雑な非線形系の同定・予測

今日まで，入出力データに基づく線形系の同定・予測問題に関しては多くの研究がなされているが[16,17]，非線形系の問題に関しては構造が与えらてパラメータ推定に帰着される場合を除いてほとんど考察されていない．わずかに非パラメトリックな一般的方法として，Volterra 級数による同定方法[18,19]および定常確率過程に対する Kolmogorov-Gabor の多項式表現を基礎にした Gabor らの Universal Nonlinear Filter の手法がある[20]．この方法は後述するようにRosenblattの多層パーセプトロンの理論と共にGMDHの概念的な基礎を形成している．

しかしながら Volterra 級数や高次の多項式による非線形系の同定・予測問題は推定すべき係数およびそのために必要なデータの量が莫大なものとなり，多変数に対しては計算量ならびに行列計算の安定性という面で実用に供するのは困難である．そのうえ，複雑で多変数な系のモデリングにおいては常に

i) 非常に多くの変数のうちで，どの変数がモデルの目的に“適切”であるのか．(“適切”な変数の選択)

ii) どの変数が冗長であり，“非適切”なものなのか（雑音）

iii) どのようなモデル構造がとられるべきか（非線形の種類）

iv) モデルのパラメータをどのようにして決めるのか．

v) 新しいデータ(モデリングに用いられていない)にどの程度適合させるのか．

といった問題が生じる．

Ivakhnenko の提案する Heuristic Self-Organization によるアルゴリズム (GMDH) は上記のような複雑な非線形系の「定式化の困難さ」と「次元の呪い」を解決する有力な手法であり，その特徴は

a) 少ない入出力データで複雑で多変数，非線形系の同定・予測が可能

b) 多変数のわりには計算量が少なくてアルゴリズムが安定（従来の確率的予測法と較べて）

c)「最適な複雑さ」をもつ数学的な記述が可能(数値的な精度の面で)

にあるといえる．

2.　Heuristic Self-Organization の原理と GMDH

微分方程式に基礎をもつ現代制御理論は入出力関係がはっきりと規定できないような複雑なシステムを取扱うことはできず，実際的で複雑な現象を単純化した場合にのみ適用可能となる．したがって，現象の定性的な関係を研究するには適しているが，正確な定量的な記述を得ることは非常に困難である．

このような問題に対しては人類はこれまでその永い経験によって驚くほどうまく解決してきた．たとえば動植物の品種改良による農畜産業の生産の増大，複雑な経済システムの確立と運営などにおいて人類は豊富な経験をもっている．サイバネティクスにおける「複雑さ」の処理はまず第1に，数学的な定式化の記述をもっていない上述のようなシステムの制御に対する人類の豊富な経験を利用しなければならない．すなわち，種の進化，生存競争，自然淘汰などに基づく「品種改良」の原則をサイバネティクスにおける情報処理とモデリングに用いることによって「複雑さ」を取扱うことが可能となる[5]．これが Ivakhnenko の提唱する Heuristic Self-Organization の基礎となっている．

2.1 「品種改良」の仮説とパーセプトロン[5]

植物の種子の品種改良では，まず最初の世代で必要とされている特性をもっていると考えられる種子（最初の Heurisics）が大量に蒔かれ，そして交配され(最初の組合せの発生)，最初の収穫物が成熟する．つぎにこの収穫物のうち，われわれの要求によりよく合致する部分が選択される(最初の閾値的自己選択)，そしてこの一定の割合の種子が再び蒔かれ，交配される(第2の組合せの発生)．さらに二番目の収穫から播種用に一定の割合の種子が選ばれる（第2の閾値的自己選択)．以下同様に行なわれる．

このような品種改良においては

1)　各世代で播種用に選ばれる種子の最適な一定の割合が存在する．すなわち，その割合の増加，減少は品種改良を悪化させたり遅らせたりする．

2)　1つの世代で品種改良を完成させることはできず，数世代を必要とする．

3)　長期の品種改良は植物の退化をもたらす．

4)　品種改良の問題が複雑であればあるほど，それだけ多くの淘汰の世代が必要である．

という原則が経験的に確立している[5].

これらの原則は数学的な厳密さの意味ではいまだ証明されていないので「品種改良の仮説」としておくが，この仮説を利用して Heuristic Self-Organization の数学的アルゴリズムを作成することが可能である．

脳の知覚モデルとしての Rosenblatt のパーセプトロン[21]のアルゴリズムはこの「品種改良の仮説」と類似している．（**図 2** 参照）すなわち，入力信号の確率的組合わせの発生と閾値による有用な情報の自己選択，「有害」な組合わせの淘汰および多層構造がそれである．かくして，Heuristic Self-Organization のシステムをつぎのように定義することが可能となる．

『多層または階層構造をもっていて，その各層で有用な情報の積分的あるいは閾値的自己選択が用いられるシステムである．その発見的規範 (Heuristics) に基づいた自己選択を有効なものにするために，1つのまたは若干のランダムな変数の組合わせ発生器が用いられ，その結果，各層毎にシステムの変数による記述の複雑さが増大する．』

2.2 GMDH の構造

GMDH のアルゴリズムは思想的には Heuristic Self-Organization を実現しているパーセプトロンを継承したもので，その主要な改良点は**図 3** に示すように

（a） 入力のスカラー積であるパーセプトロンの確率的組合わせに対して植物の種の交配による改良と同様に，2つの変数の非線形の組合わせを多量に発生させる．この組合わせを初等アルゴリズム (Elementary algorithms) と呼ぶ．この組合わせによって各層で部分記述 (Partial description) あるいは中間変数が生成される．

（b） 解（モデルあるいは予測式）の正則化 (Regularization) を考慮した種々の Heuristics が導入されている[10]．すなわち与えられた入出力データを中間変数の構造を決定するための training データといくつかの中間変数の中からよりよい中間変数を選択するために使われる checking データとに分割する．分割の方法に種々の Heuristics が用いられる．

（c） 各層での積分作用としての閾値的自己選択に，過去の経験を応用する多様な Heuristics を導入して望ましい結果へと導いてゆく．

（d） 閾値および正則化の方法の最適化によって，最適な複雑さをもつ最終的な記述 (Complete description) を得る．

ことにある．

特に(b)項の正則化の概念は非常に重要である．正則化は通常数値解析に用いられているが[22,23]，ここでは最終的なモデルのユニーク性と安定性を確保するために用いられている．その物理的な意味はたとえば予測式においては，入出力データによる over-fitting を避けて新しい入出力データに対する適度な滑らかさをもつ構造を定めることにある．

さらに(c)項の Heuristics としてはつぎのようなものに分類されている．

① 初等的アルゴリズム（組合わせの発生）

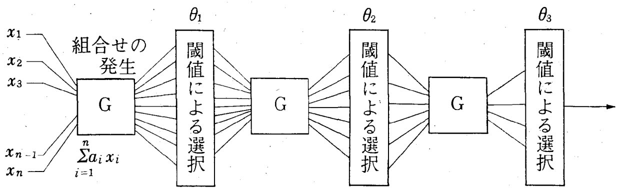

図 2 パーセプトロンの構造[5]

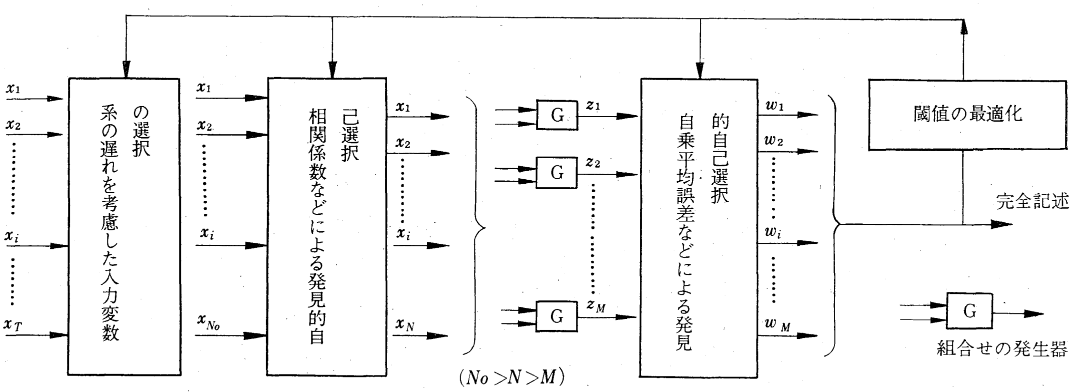

図 3 基本的 GMDH のアルゴリズムの構造

与えられた入力信号から中間変数（部分記述）への非線形変換.（たとえば多峰性関数を同定するための2次多項式，3次多項式や，周期関数を同定するための三角関数の非線形結合など）

② 自己選択のための規範
類似の問題の解の経験に基づく，有用な情報の選択のための規範.（たとえば，最小2乗法，相関係数など）

③ 積分作用
複雑な系の各要素に関する情報を個々別々に利用しないで，初等アルゴリズムをもつ各要素全体に作用させる.（たとえば，各層での閾値(threshold)的サンプリングなど）

3. GMDH のアルゴリズム

3.1 初等アルゴリズム（基礎関数）

発見的自己組織化(Heuristic Self-Organization)のシステムは，前節の結果を要約すれば，複雑な自然現象の過程において，その過程の不完全な情報のもとに変数のランダムな結合（仮説）を発生し，発見的規範(Heuristic Criterion)を用いて多層構造で最良のものを閾値的に自己選択してゆく制御技術の1つであるといえよう.

したがって，GMDH のアルゴリズムは，中間変数のための初等アルゴリズム（基礎関数）を何にするかによって，いくらでも存在することになり，ここに，複雑なシステムの問題に応じた基礎関数の選択という Heuristics が入ってくる．いい換えれば非常に柔軟な一般的アルゴリズムを形成していることになる.

最も基本となる基礎関数は，完全記述として定常確率過程の多項式近似である Kolmogorov-Gabor の多項式[20]

$$y=a_0+\sum_i a_i x_i+\sum_i\sum_j a_{ij}x_i x_j+\sum_i\sum_j\sum_k a_{ijk}x_i x_j x_k+\cdots \quad (1)$$

を用いたときには，つぎのような2次多項式となる.

$$z_k=a_0+a_1x_i+a_2x_j+a_3x_ix_j+a_4x_i^2+a_5x_j^2 \quad (2)$$

あるいは

$$z_k=a_0+a_1x_i+a_2x_j+a_3x_ix_j \quad (3)$$

このほかにも，気象統計でよく用いられる直交チェビシェフ多項式を用いた残差多項式

$$w=y_1(x_1)+z_1(x_2)+v_1(x_3)+\cdots$$

周期関数，Bayes 関数，論理関数などが考えられている[3].

(2)式の2次多項式を用いた場合，各層でのアルゴリズムは，一種の Quadratic Regression Analysis になっていて，第 n 層を通過すれば，2^n 次の多項式ができあがる．一般に(1)式の係数 a_{ni} を定めるためには，非常に多量の入出力データが必要であるが，(2)式の係数を定めるには，たかだか6個あればよい．この意味で，GMDH のアルゴリズムは少ない入出力データで，高次の多項式近似を可能にする．しかも入出力データの Training と Checking への分割という正則化(Regularization)を用いることによって，回帰モデルの最適な選択を可能にしている.

3.2 多項式による基本的アルゴリズム

Ivakhnenko の提案する基本的な GMDH のアルゴリズムの手順を以下において概略説明しよう.

$x_i\,(i=1,2,\cdots,N,\cdots)$ を入力変数，y を出力変数とし，入力と出力の間にはつぎのような非線形関係

$$y=f(x_1,x_2,\cdots,x_N,\cdots) \quad (4)$$

があるとする．入出力関係 f の同定はつぎの手続きで行なわれる.

ステップ1: y に関係する入力変数の Prehistory（時定数あるいは y に影響を与える期間）を定めて，おのおのをすべて入力変数とする．さらに必要があればデータの normalization などの前処理を行なう.（入力変数は $x_i\,(i=1,2,\cdots,T)$ であるとする.）

ステップ2: 出力 y と各入力変数の相関をとり，相関係数の大きいものを“有用な”入力変数として残し，相関係数の小さいものを“有害な”入力変数として捨てる.（この結果，入力変数をならびかえて，$x_i\,(i=1,2,\cdots,N;\ N<T)$ となった.

ステップ3: 原データを training データと Checking データの2組に分割する．分割の方法は通常 training と checking を交互にとる場合，分散の大きいほうを training に，小さいほうを checking にする場合などが考えられている.

ステップ4: 入力変数 $x_k\,(k=1,2,\cdots,N)$ の2個の組合わせ x_i, x_j に対して，中間変数 $z_k\,(k=1,2,\cdots,N(N-1)/2)$ を

$$z_k=a_0+a_1x_i+a_2x_j+a_3x_i^2+a_4x_j^2+a_5x_ix_j \quad (5)$$

によって作る．ただし，(5)式右辺における係数 $a_l\,(l=0,1,\cdots,5)$ は，training データに基づく2乗平均誤差

$$\varepsilon^2_k=\overline{(y-z_k)^2} \quad (6)$$

を最小にするように決定される．なお，(6)式における“ー”はサンプル平均を意味する．

ステップ 5：ステップ4においてtrainingデータに基づいて決定された係数を使って，checkingデータを(5)式により変換し，(6)式の2乗平均誤差を変換されたchecking データに対して計算する．この2乗平均誤差を小さくするものから順にM個中間変数を採用し，残りは捨てる．

ステップ 6：$x_i=z_i$，$x_j=z_j$ としてステップ4へへゆき，つぎの中間変数を得る．以下ステップ 4〜ステップ6を繰返す．

ステップ 7：ステップ5において checking データに対して計算された2乗平均誤差の最小のものが前回の最小な2乗平均誤差を越えた場合に，計算を停止する．前回までに計算された(5)式の2次式をつぎつぎと代入することにより，完全記述fが定まる．

これらの手順につけ加えて，各種の Heuristics（相関係数による選択の数，各層での選択の数や閾値，原データの分割の方法）の最適化を行なうことにより，全体として入出力関係の最適な同定が行なわれる．

以上のアルゴリズムをフローチャートで示せば**図4**のようになる．

4. 応用例

4.1 ソ連における研究

GMDH の提案者 A. G. Ivakhnenko はソ連のキエフにある，ウクライナ共和国科学アカデミーのオートメーション部門の責任者で，1968年以降，彼を中心としたグループが，Heuristic Self-Organization に関するあらゆる可能性について数多くの研究を行なっている．それらの結果は主として「Soviet Automatic Control」（英語版，以後 SAC と略記する）という定期刊行物に掲載されている．彼らの研究内容と研究対象がいままでどの方向に寄せられていたかを知るために，主要な論文の題目を以下に掲げておこう．

1) 「The Group Method of Data Handling—A Rival of the Method of Stochastic Approximation」[6] SAC, No. 3 (1968)
小麦の作付面積の予測を例題として GMDH の紹介を行なっている．

2) 「The Group Method of Data Handling in Pattern Recognition and Decision Problems」[7] SAC, No. 5 (1968)
多層のパーセプトロン型構造を用いて，有用な情報の heuristic selection とパラメータの適応的な fitting の代わりをする最小2乗法について研究している．

3) 「Group Handling of Data in Identification of the Static Characteristic of a Multi-Extremal Plant」[8] SAC, No. 2 (1969)
複雑なシステムへの“品種改良の仮説”の適用について解説し，静的な多峰性の関数の同定問題を GMDH で行なっている．

4) 「Group Method of Data Handling in Optimal Control with Accumulation of Information」[9] SAC, No. 5 (1969)
時間遅れをもつような複雑なプラントに対する Control の構成に GMDH を適用する場合について研究している．

5) 「Regularization of Decision Functions in the Group Method of Data Handling」[10]；SAC, No. 2 (1970)
多項式記述を用いる GMDH への正則化の適用について考察している．すなわち(a)部分記述の選択において ill-condition 行列をもつようなものの排除，(b)分散値の大きさによる原データの training と checking への分割，(c)人工的に滑らかな関数の内挿点を増加する方法，

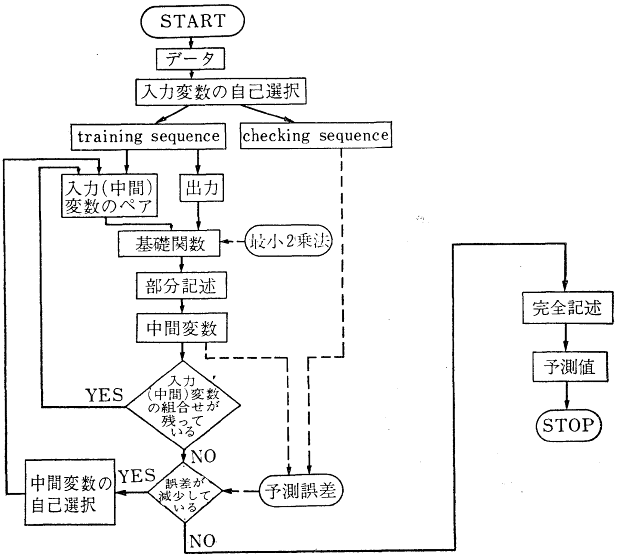

図 4 GMDH のアルゴリズムのフローチャート

などについて考察している．

6) 「Multilayer Statistical Decision Theory」[15) SAC, No. 1 (1971)

多層構造をもつ GMDH へ統計的決定理論を持込む場合について考察している．

7) 「Mathematical Simulation of Complex Ecological Systems」[11)]; SAC, No. 4 (1971)

GMDH を貯水池のバクテリヤの量の予測（1年先）に応用した例について述べている．（非常に複雑な生態系で，測定量が少ない場合の例題）

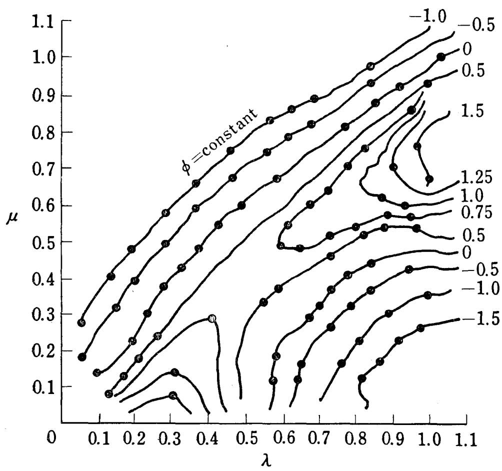

図 5　双峰性関数 ϕ の静特性[3)]

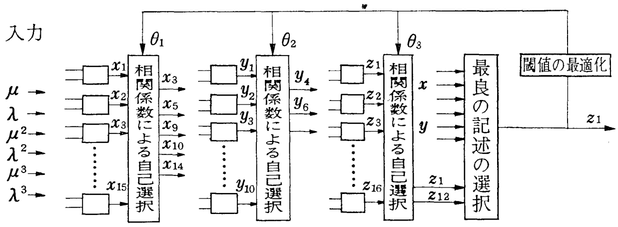

図 6　双峰性関数の同定における GMDH のアルゴリズム[3)]

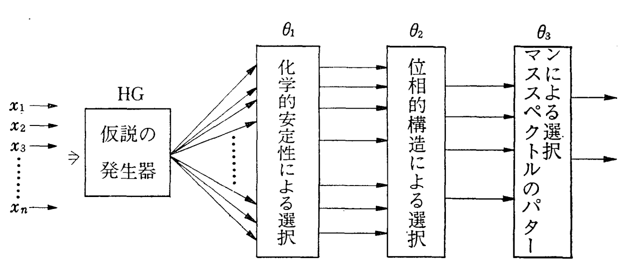

図 7　質量分析データによる分子構造決定のアルゴリズム[15)]

8) 「Problems of Simulation of Complex Systems and of Applied Mathematical Statistics」[12)]; SAC, No. 6 (1971)

回帰式に代表される数理統計学と発見的自己組織化法 (GMDH) との関連について考察し，回帰多項式の最適な「複雑さ」とユニーク性について述べている．

9) 「Structual Identification of Differential Equations by a Small Number of Observations Using Self-Organization Methods」[13)]; SAC, No. 2 (1972)

構造の定まった微分方程式の係数を少ないデータから GMDH によって決定する問題を (a) ニュートン力学の法則，(b) イギリス経済の数学的モデルを例題として考察している．

10) 「Unique Construction of Regression Curve Using a Small Number of Points」[14)]; Part 1, SAC, No. 5 (1972); Part 2, SAC, No. 5 (1973)

多項式記述をもつ GMDH の数学的な基礎について考察している．〈品種改良の仮説〉を数理統計学および代数学の立場からいくつかの定理にしている．

つぎに GMDH のアルゴリズムを理解するのに役立つ文献 8) の簡単な例題を示しておこう．

4.2　多峰性の関数の同定問題[8)]

いま図 5 に示すような 2 つの極値丘 (bimodal) をもつ 2 変数の関数 ϕ

$$\phi=f(\mu,\lambda) \cdots\cdots\cdots\cdots\cdots\cdots (7)$$

を同定することを考えよう．いま 70 個の (μ,λ) の組に対するデータ φ の値が与えられていたとしよう．そこでつぎのような Heuristics を導入して $f(\mu,\lambda)$ を決定する．

① 入力変数の選択の Heuristics：
$\mu, \lambda, \mu^2, \lambda^2, \mu^3, \lambda^3$
の 6 つを入力変数と考える．

② 各層での自己選択の Heuristics；各層で threshold の閾値を一定値 $\theta_1, \theta_2, \theta_3, \cdots$ とし，中間変数と ϕ との相関係数が閾値を越えた場合につぎの層へ進む．

③ 初等関数：2 次の多項式 (5) 式を採用する．

④ 原データの分割：60 点を training データに 10 点を checking データにする．

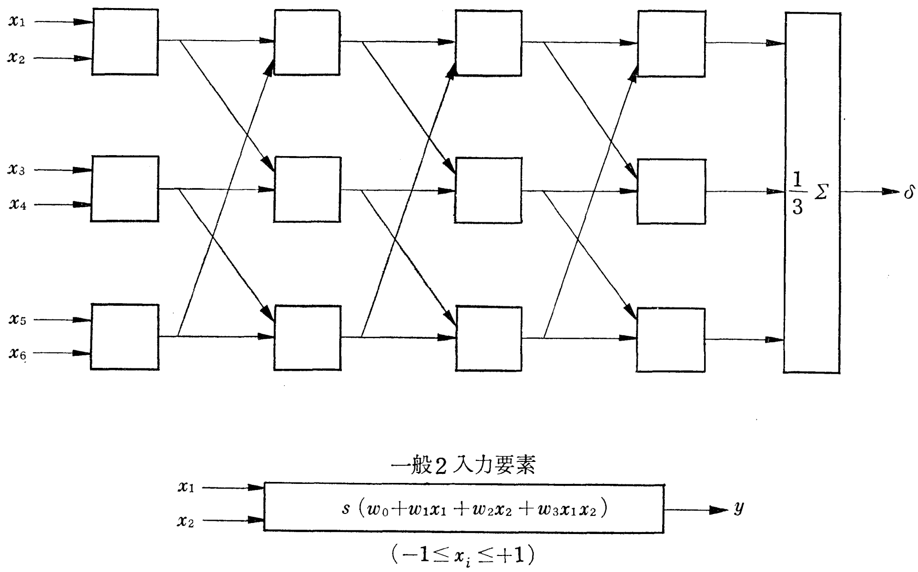

図 8　12 の要素からなる Adaptive Transformation Network[26]

以上の Heuristics を用いて 3.2 節の GMDH のアルゴリズムの手続きを図式にしたのが**図 6** に示してある．

たとえば，$\theta_1=0.84$ のとき，第 1 層への入力変数 6 個のうちの任意の 2 個の組合わせ $x_1, \cdots, x_{15}$ のうち，つぎの 5 個が選択される．

$$x_3=-0.17+7.24\lambda-12.69\mu^2-11.38\lambda^2-14.15\mu^4+31.23\lambda\mu^2,\quad k=0.8669$$

$$x_5=-0.94+8.47\lambda-18.68\mu^3-9.86\lambda^2-10.31\mu^6+31.98\lambda\mu^2,\quad k=0.8824$$

$$x_9=1.12-8.43\mu-10.11\mu^3+20.36\mu^2+4.15\mu^6-9.11\mu^4,\quad k=0.9058$$

$$x_{10}=0.33+3.65\lambda^2-3.79\mu^2-9.06\lambda^4-13.2\mu^4+22.46\lambda^2\mu^2,\quad k=0.8417$$

$$x_{14}=0.97+0.71\mu^2-4.36\mu^3+118.2\mu^4+34.34\mu^6-112.7\mu^5,\quad k=0.8646$$

ここに k は checking データによる相関係数である．

以下同様にして，第 2 層では

$$y_4=0.019+0.26x_9+0.62x_5+0.007x_9^2-0.13x_5^2-0.18x_5x_9,\quad k=0.9712$$

$$y_6=0.069+0.19x_9+0.71x_3+0.019x_9^2-0.21x_3^2+0.22x_9x_3,\quad k=0.9702$$

$$y_8=\cdots$$

を得る．第 3 層では

$$z_1=0.11+0.3y_6+0.7y_4-1.52y_6^2-1.5y_4^2+3.06y_6y_4,\quad k=0.9815$$

$$z_{12}=0.14+0.43y_4+0.54y_8-2.33y_4^2-2.36y_8^2+4.69y_4y_8,\quad k=0.9807$$

$$z_{16}=\cdots$$

を得る．第 4 層ではもはや k の値は第 3 層より悪くなるので，以下の結果から最適な解として

$$\phi=z_1=\varphi(x_3, x_5, x_9) \quad\cdots\cdots\cdots(8)$$

なる x_3, x_5, x_9 に関する 6 次の多項式を得る．

4.3　その他の国での応用例

発見的自己組織化の概念はアメリカの Stanford 大学で質量分析のデータにより有機化合物の分子構造を決定する問題に用いられている[15]．**図 7** に示すように，仮説の発生器 HG で分子構造に関するすべての可能な変体 (variant) を発生する．そのあとで実験化学者の Heuristic Criterion, すなわち(ⅰ)化学的な安定性，(ⅱ)トポロジ的構造，(ⅲ)質量分析の結果のパターンへの適合性などによって構造式が自己選択されてゆく．

一方，Ivakhnenko らの研究とほぼ同じ頃に R. L. Barron および A. N. Mucciardi らによって，GMDH と思想的には同じ立場に立つ手法「Adaptive Transformation Network (ATN)」が提案され[25,26]，Adaptronics Inc. というソフトウェア会社によって，種々の実際問題への適用が行なわれている．彼らの方法は主として多峰性関数の超曲面近似を適応的なアプローチで行なうことにあるが，

① クラスタ分析を用いたデータの前処理
② ネットワーク・エレメントの設計とその結合法
③ ネットワークを training する手法—Guided

Accelerated Random Search (GRAS) に特徴がある．GMDH ではネットワークの training (仮説の発生) に最小2乗法による2次多項式のパラメータ決定が用いられたが，ATN ではランダムサーチとグラジェント法とを併用した GRAS によってパラメータが決定される．図8にそのネットワークの例を示してある．

J. Duffy と M. Franklin はアメリカの農業地帯の河川における硝酸 (NO_3) レベルのモデルを作成するのに GMDH を応用している[27]．彼らはまず生態学的，土壌学的な立場から，この問題を取上げようとしていた．しかしながら複雑で，モデルの構造が現象論的にもはっきりしない場合に，データによる経験的な仮説を作成するのに GMDH は非常に有効であると強調している．

わが国における GMDH の研究はまだほとんど行なわれていないが，昭和47年に本学会誌で高橋安人教授による簡単な紹介が行なわれ[28]，その後，井原二郎氏によって，世界モデルの同定に応用されている[29]．そこでは，世界の人口が

$$P_n=f(P_{n-1}, P_{n-2}, E_n, E_{n-1}) \quad \cdots\cdots(9)$$

P_n: n 時点における世界人口

E_n: n 時点における1人当りの世界消費エネルギ

で表わされるものとして GMDH を適用している．その際，予測式の安定性を増すために，フィルタ付きの基礎関数

$$\left.\begin{array}{l} y=b_0+b_1z_i+b_2z_i^2 \\ z_i=wx_i+(1-w)x_j \end{array}\right\} \quad \cdots\cdots(10)$$

x_i, x_j: 入力変数　　w: 重み係数

を提案し，良好な結果を得ている．

そのほか，遠藤，畑中氏らによっても，非線形制御系への同定問題に対してモデルの構造およびパラメータの安定性という立場から GMDH の修正が提案されている[30]．また広域パターンの同定問題への適用も行なわれている[31]．

著者らも環境システムような複雑でモデル構造のはっきりしないシステムのモデリングと予測に GMDH を応用し，良好な結果を得た[32,33]．現象論的な考察によるモデルとサイバネティクス的な方法によるモデルとの比較のために，その例について次節で述べることにする．

4.4 河川の流出量予測への適用[32,33]

河川の水量の予測は，水質源の有効利用および水質汚染制御にとって重要な問題である．このために，水文学の立場から，多くの研究が行なわれ，低水量時の

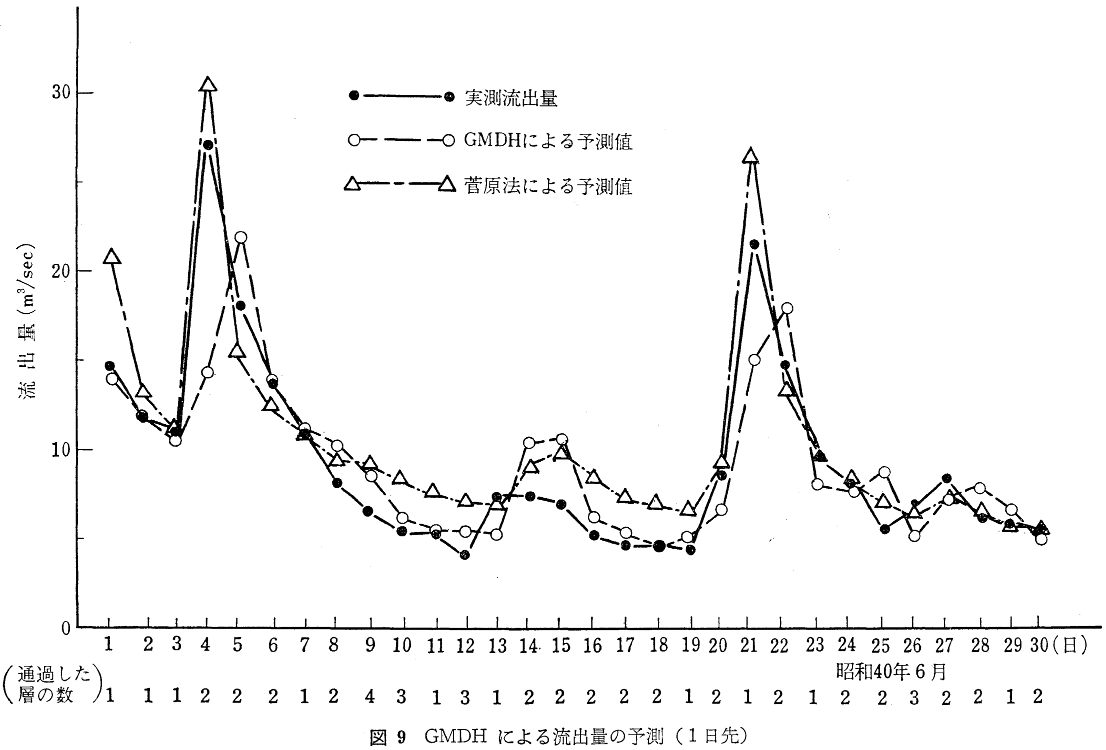

図 9　GMDH による流出量の予測（1日先）

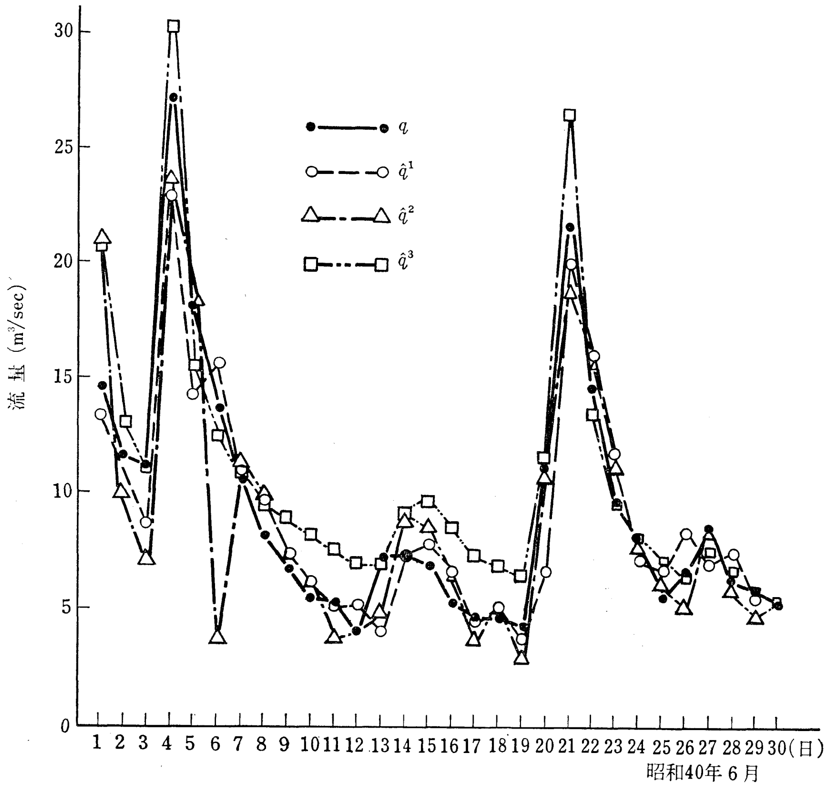

図 10 修正 GMDH による流出量予測（1日先）

流出量予測モデルとして，菅原法（直列三段のタンクモデル），西原法（流域貯水量，蒸発ポテンシャルなどによる回帰モデル）および貯留関数法（洪水流出モデルを基礎にした回帰モデル）などが知られている[34]．ここでは GMDH による入力データの時系列が少ない場合の流出量予測を行なって，上記の方法との比較を行なうことにする．

データとして毎日の烏川流域（利根川上流）での降水量 (mm)，水位 (m)，実測流出量 (m^3/sec) が昭和40年5月1日～6月30日までの2カ月間が与えられている[34]．このとき，GMDH アルゴリズムの適用により1日先の流出量の予測を与えられたデータから行なった．

前節で述べた多項式近似による GMDH アルゴリズムの適用の手続はつぎのようである．

i）降水量を r，水位を h，流出量を q で表わすとき，k 日目の流出量 q_k の予測のために，入力変数として，

$$r_{k-1}, \cdots, r_{k-10}\,;\ h_{k-1}, \cdots, h_{k-10}\,;\ q_{k-1}, \cdots, q_{k-10}$$

の30個を用いた．

ii）予測式の構造決定のためのデータ使用期間は，予測されるべき日の前30日間．

iii）30個の入力変数のうち，流出量との相関係数の大きいものから10個選択．

iv）training データと checking データへの分割は，奇数番目を前者へ，偶数番目を後者へという方法．

v）中間変数は，流出量との2乗平均誤差の小さいものから順に5個選択．

なお，最小2乗法適用の際に，逆行列をとるべき行列が正則でないかあるいはそれに近ければ，その組合わせは捨てる方法をとっている．昭和40年6月1日～30日までの1カ月間の流出量予測を行なった結果をグラフに示したのが**図9**である．予測の評価の指標を

$$J=\left\{\frac{1}{30}\sum_{k=1}^{30}(q_k-\hat{q}_k)^2\right\}^{1/2},\quad \hat{q}_k\text{：予測値}$$

としたとき，各方法に対して，

J(GMDH)=3.21，J(菅原法)=2.23

J(西原法)=3.10，J(貯留関数法)=5.19

であった．図9からわかるように，流出量の急激な増加に対する追従性能が，GMDH の場合やや悪いがこれは比較的雨の少ない5月のデータに基づいて予測式の構造が決定されたためであろう．菅原法，西原法，貯留関数法などに比べずっと少ない 30 日間のデータ

に基づいて求められた GMDH による予測がほぼ同等の精度で行なわれていることは，利用できるデータが少ない場合における GMDH の有効性を示しているといえる.

つぎに急激な増加に対する追従性を改善するために，入力変数の選び方の Heuristics および中間変数生成のための(4)式の修正を試みよう.

そこでつぎのような自己組織化によるアルゴリズムを提案する.

① 各入力変数 x_i の多項式
$$z_i=a_0+{}_1x_i+a_2x_i^2+\cdots+a_nx_i^n \quad \cdots(11)$$
を用いて，出力 y と各 z_i との 2 乗平均誤差が最小になるように(2)式の係数 $a_j\ (j=1,2,\cdots n)$ を定めたとき，その 2 乗平均誤差が小さくなるものから順にM個“有用な”入力変数として採用する．ここに，n, M はそれぞれ，適当に定められた正整数である.

② ①で選択された入力変数 $x_i\ (i=i_1,\cdots,i_M)$ と出力 y の原データを training データと checking データとに分割する.

③ 今，k 回目の中間変数 $w_i^{(k)}\ (i=1,2,\cdots,N)$ が得られているとする．($k=1$では，$w_j^{(1)}=x_{ij}$, $j=1,2,\cdots,M$) まず，中間変数 $w_i^{(k)}$ を次式によって変換する.
$$\hat{w}_i^{(k)}=b_0+b_1w_i^{(k)}+b_2(w_i^{(k)})^2 \quad \cdots\cdots\cdots(12)$$
ただし，(12)式右辺の係数 $b_j\ (j=0,1,2)$ は，出力 y と変換された $\hat{w}_i^{(k)}$ との 2 乗平均誤差が，training データに対して，最小となるように定められる.

④ 変換された $\hat{w}_i^{(k)}\ (i=1,2,\cdots,N)$ 相互の結合を
$$w_l^{(k+1)}=c_0+c_1\hat{w}_i^{(k)}+c_2\hat{w}_j^{(k)}+c_3(\hat{w}_i^{(k)})^2+c_4(\hat{w}_j^{(k)})^2+c_5\hat{w}_i^{(k)}\hat{w}_j^{(k)} \quad \cdots\cdots\cdots(13\text{-}1)$$
または，
$$w_l^{(k+1)}=d_0+d_1\hat{w}_i^{(k)}+d_2\hat{w}_j^{(k)}+d_3\hat{w}_i^{(k)}\hat{w}_j^{(k)} \quad \cdots\cdots\cdots(13\text{-}2)$$
によって行ない．$k+1$ 回目の中間変数 $w_l^{(k+1)}$ $(l=1,2,\cdots,N(N-1)/2)$ を得る．(13-1), (13-2) 式右辺の係数 $c_i\ (i=0,1,\cdots,5)$, $d_i\ (i=0,1,2,3)$ は，それぞれ，出力 y と中間変数 $w_l^{(k+1)}$ との 2 乗平均誤差が，training データに対して最小となるように定められる.

⑤ (12)および(13-1)（または，(13-2)）式によって，checking データに対する中間変数を生成し，その中間変数と出力 y との 2 乗平均誤差の小さいものから順にN個選択する.

⑥ $k\to k+1$ として，③へゆく．以下，⑤において，checking データに対する中間変数と出力との 2 乗平均誤差が前回のものを越えた場合に計算を停止する．このとき，前回までの(12), (13-1)（または，(13-2)）から中間変数を消去することによって，最適な複雑さをもつ関数fが決定される.

上述の手順において，

(i) 入力変数として，(11)式によらずに，
$$q_{k-1},\cdots,q_{k-4},r_{k-1},\cdots,r_{k-6}$$
と指定した．出力変数は q_k である.

(ii) 原データの偶数番目を training データへ，奇数番目を checking データへ.

(iii) $N=5$ として，流出量の 1 日先の予測を行なった結果が**図 10** に示されている．ただし，$\hat{q}^1$ は(13-1)，$\hat{q}^2$ は(13-2)によったものであり，$\hat{q}^3$ は菅原法による予測値である．なお，本方法では予測のために過去 30 日間のデータが用いられている．この場合，予測の評価は
$J(\hat{q}^1)=1.86$，$J(\hat{q}^2)=2.68$，$J(\hat{q}^3)=2.23$
であった．また予測式の例として6月5日の $\hat{q}^1$ の構造を**表 1** にのせておく.

表 1　6月5日の予測式の構造 (q')

入力変数	$x_1=q_1$，$x_2=r_1$，$x_3=r_2,\cdots$
第 1 層(部分記述)	$w_1^{(1)}=-0.7113+1.1809x_1+0.0148x_2+0.0220(x_1)^2+0.0024(x_2)^2+0.0101x_1x_2$
	$w_2^{(1)}=2.0088+0.4540x_2+0.0706x_3-0.00065(x_2)^2+0.00055(x_3)^2-0.0103x_2x_3$
	$w_1^{(1)}=-0.00008+1.0000w_1^{(1)}-4.0\times10^{-7}(w_1^{(1)})^2$
	$w_2^{(1)}=0.0380+0.9922w_2^{(1)}+0.00017(w_2^{(1)})^2$
第 2 層(部分記述)	$w_1^{(2)}=-0.3488+0.8171w_1^{(1)}+0.2833w_2^{(1)}-0.0084(w_1^{(1)})^2-0.0240(w_2^{(1)})^2+0.0304w_1^{(1)}w_2^{(1)}$
予測式(完全記述)	$q^1(t)=w_1^{(2)}$

お　わ　り　に

発見的自己組織化 (GMDH) の方法は，いまだ十分にその数学的論理づけが完成していない．しかし多変数で，その物理的な構造が明確でなく，かつそれに関する情報量（入出力データ）も少ない複雑なシステムの同定・予測モデルの作成に非常に有効な手段を提供

するものである．GMDH は複雑なシステムのモデリングを専門科学の狭い領域の人達だけでなく，より広汎な人々にまで可能ならしめ，かつ実際的に役立つモデルを作成する一般的方法論の確立をめざしたものとして非常に興味深い．この方法は複雑なシステムの物理的あるいは現象論的な考察に基づくモデルを作成するための仮説の作成という役割もまた果すことができ，今後の発展の可能性を大いに秘めていると思われる．

参考文献

1) N. Wiener: Cybernetics, John Wiley (1961)（サイバネティクス，岩波書店 (1962)）
2) W.R. アシュビー：サイバネティクス入門（篠崎ほか訳），宇野書店 (1967)
3) A.G. Ivakhnenko: Heuristic Self-Organization in Problems of Engineering Cybernetics; Automatica, **6**, 207/219 (1970)
4) A.G. Ivakhnenko: Polynomial Theory of Complex Systems; IEEE Trans., SMC-1, 364/378 (1971)
5) A.G. Ivakhnenko: Heuristic Self-Organization Systems in Engineering Cybenetics. Kiev, Tekhnika Press (1971) (in Russian)
6) A.G. Ivakhnenko: The Group Method of Data Handling, A Rival of the method of Stochastic Approximation; Soviet Automatic Control, **13**-3, 43/55 (1968)
7) A.G. Ivakhnenko et al: The Group Method of Data Handling in Pattern Recognition and Decision Problems; Soviet Automatic Control, **13**-5, 31/41 (1968)
8) A.G. Ivakhnenko et al: Group Handling of Data in Identification of the Static Characteristic of a Multi-Extremal Plant; Soviet Automatic Control, **14**-2, 30/37 (1969)
9) A.G. Ivakhnenko: The Group Method of Data Handling in Optimal Control with Accumulation of Information; Soviet Automatic Control, **14**-5, 57/60 (1969)
10) A.G. Ivakhnenko and Yu. V. Koppa: Regularization of Decision Functions in the Group Method of Data Handling; Soviet Automatic Control, **15**-2, 28/37 (1970)
11) A.G. Ivakhnenko, Yu. v. Koppa & M.M. Todua: Mathematical Simulation of Complex Ecological Systems; Soviet Automatic Control, **4**-4, 15/26 (1971)
12) A.G. Ivakhnenko: Problems of Simulation of Complex Systems and of Applied Mathematical Statics; Soviet Automatic Control, **4**-6, 1/6 (1971)
13) A.G. Ivakhnenko: Structual Identification of Differential Equations by a Small Number of Observations Using Self-Organization Method; Soviet Automatic Control, **5**-2, 31/44 (1972)
14) A.G. Ivakhnenko and P.I. Koval'chuk: Unique Construction of Regression Curve Using a Small Number of Points; Soviet Automatic Control, Part 1: **5**-5, 26/32 (1972), Part 2: **6**-5, 11/28 (1973)
15) O.H. Ivakhnenko, Ye. I. Spynu et al: Multilayer Statistical Decision Theory; Soviet Automatic Control, **4**-1, 37/47 (1971)
16) K.J. Astrom and P. Eykhoff: System Identification a Survey; Automatica, **7**, 123/162 (1971)
17) G.E.P. Box and G.M. Jenkins: Time Series Analysis-Forecasting and Control. San Francisco, Holden-Day (1970)
18) P. Eykhoff: Systems Identification. London, Wiley-Interscience (1974)
19) R.A. Roy and J. Sherman: A Learning Technigue for Volterra Series Representation; IEEE Trans. Auto. Cont., **AC-12**, 761/764 (1967)
20) D. Gabor, W.P.L. Wilby & R. Woodcock: A Universal Non-Linear Filter Predictor and Simulator which Optimizes Itself by a Learning Process; IEE Proc. **108**, Part B, 422/433 (1961)
21) F. Rosenblatt: Principle of Neurodynamics. New York, Spartan Books (1962)
22) A.N. Tikhonov: Regularization of incorrectly formulated problems; Dokl. Acad. Nauk, SSSR (Soviet Mathematics in English Edition), **151**, 1035/1038 (1963)
23) L.S. Kirillova and A.A. Piontkovskii: Incorrect Problems in Optimal Control Theory (Survey); Automation & Remote Cont., 2910, 1553/1563 (1968)
24) N.O. Ivakhnenko: A Self-Organizing Heuristic Program for Solving Mass-Spectrography Problems; Soviet Automatic Control, **14**-4, 66/67 (1969)
25) R.L. Barron: Adaptive Transformation Networks for Modeling, Prediction and Control; Proc. IEEE/ORSA Joint Conf. on Major Systems, Anaheim, CA, 254/263, Oct. (1971)
26) A.N. Mucciardi: Neuromine Nets as the Basis for the Predictive Component of Robot Brains; in Cybernetics, Artificial Intelligence and Ecology. 159/193, New York, Spartan Books (1972)
27) J. Duffy and M. Franklin: A Case Study of Environmental System Modeling with the Group Method of Data Handling; Proc. Joint Auto. Conf. of AACC, 101/111 (1973)
28) 高橋安人：個体数の力学と制御(Ⅲ)；計測と制御，**11**-10, 23/32 (1972)
29) 井原二郎：GMDH の世界人口モデルの同定への一応用；第 15 回自動制御連合講演会，391/392 (1972)
30) 遠藤明男，畑中 浩：GMDH による非線形制御系の同定；制御理論シンポジウム (SICE), 93/98 (1974)
30) 田村坦之，青谷年永：対話型 GMDH による広域パターンの同定；第 13 回 SICE 講演会，367/368 (1974)
32) 池田三郎，藤重悟，槌木義一：河川流量予測への GMDH の応用；システム制御におけるモデリングとシミュレーション・シンポジウム（日本自動制御協会), 59/62 (1974)
33) 藤重 悟，池田三郎．槌木義一：自己組織化法による非線形系の同定とその河川流量予測への応用；第 13 回 SICE 講演会，377/378 (1974)
34) 建設省，関東地方建設局，利根ダム統合管理事務所：利根川上流部における低水時流出解析報告書；(1967)