GMDHによる河川の水質変動の評価について

Estimation of Water Quality Fluctuation in River

by Application of GMDH.

羽　田　守　夫

Morio HANEDA

(昭和53年10月31日受理)

1　は　じ　め　に

環境問題は，人類が解決を迫られている課題の一つであるが，その複雑さの故に現代の科学的手法では十分に対処し得ない問題が数多く含まれている。これらの問題の多くは，変数が数多く，しかもそれらが相互に関連しており，また構造が一般に非線形であるために数学的モデルの適用が困難なシステムと言える。

このような複雑なシステムの一つの解決法として提案されたのがIvakhenenkoによる発見的自己組織化の原理に基礎をおくGMDH (Group Method of Data Handing) である。GMDHは，これまでに大気汚染の予測や水質変動の予測及び河川流量の予測等に広く応用され，十分に満足の行く結果が得られている。本稿でも，河川の水質変動の評価にGMDHを応用し，水質変動のメカニズムを理解する上で有用な結果が得られた[1]のでその概要を述べることにする。

2　GMDHの概要[2]

人間は，問題に直面した時，過去の経験等に基づく判断基準に従って，その問題の中から有用な情報を選択しつつTry and Errorで解決策を探求していくのが普通である。発見的自己組織化 (Heuristic Self Organization) とは，このような人間が問題解決において通常用いている方式を原理としている。この応用例として動植物の品種改良が挙げられる。品種改良では，次のような原則がある。

(1)　1つの品種の中に取入れたい形質を持つ原種を2つ選ぶ。

(2)　品種完成までは数世代を要する。

(3)　品種改良には最適な世代が存在し，これを越すと逆に品種は退化する。

これを品種改良の仮説と呼び，発見的自己組織化のアルゴリズムも，この原則に基づいている。

今，一つのシステムを考え，出力をy，多変数入力をx_1, x_2, ……, x_Nとする。このシステムについて離散的な時刻においてk個の観側値が得られている時，これらのデータを用いて時刻tにおける出力$y(t)$を推定することを考えよう。

$$y(t)=F\ (x_1,\ x_2,\ x_3,\ \cdots\cdots,\ x_N) \qquad \cdots\cdots\cdots\cdots(1)$$

そのためには，(1)式の決定関数Fを求めることが必要である。このシステムをBlack Boxとし，システムを非線系と考えて応答関数h_nを用いてVolterra型に級数展開すると次のようになる。

$$y(t)=\sum_{n=0}^{N}\iint\cdots\int_0^\infty h_n(\tau_1,\tau_2,\cdots,\tau_n)\prod_{j=1}^{n}x(t-\tau_j)\cdot d\tau_1\,d\tau_2\cdots d\tau_n \qquad \cdots\cdots\cdots(2)$$

ここで$N=1$, $h_0=0$とすると

$$y(t)=\int_0^\infty h_1(\tau_1)x(t-\tau_1)d\tau_1 \qquad (3)$$

となり通常の1階のたたみこみ積分を表わす。即ちこのシステムは，N階までのたたみこみ積分の総和によって表現され，線形システムと順次高次の非線形性を持つシステムとが並例的に結合されたものとなる。Nを適当に取れば，所望の精度でシステムを近似することができる。

以上を離散型に直すと次のようになる。

$$y(t)=\sum_{\tau_1=1}^{M}h_1(\tau_1)x(t-\tau_1)+\sum_{\tau_1=1}^{M}\sum_{\tau_2=1}^{M}h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)+\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad (4)$$

(4)式において$x(t-\tau_1)=x_1$, $x(t-\tau_2)=x_2$, ……

とし，$h_1(\tau_1)=a_i$，$h_2(\tau_1,\ \tau_2)=a_{ij}$，……とすると

$$y(t)=\sum_i a_i x_i+\sum_i\sum_j a_{ij}x_ix_j+\sum_i\sum_j\sum_k a_{ijk}x_ix_jx_k+\cdots\cdots \tag{5}$$

これは，kolmogorov-Gavor の多項式で，(1)式の関数 F は，(5)の多項式により近似できる。

(5)式は，このままでは複雑で応用上困難があるので，次のような二変数の基礎関数を考える。

$$f\ (x_i,\ x_j)=a_0+a_1x_i+a_2x_j+a_3x_ix_j+a_4x_i{}^2+a_5x_j{}^2 \tag{6}$$

$$f\ (x_i,\ x_j)=a_0+a_1x_i+a_2x_j+a_3x_ix_j \tag{7}$$

$$f\ (x_i,\ x_j)=a_0+a_1x_i+a_2x_j \tag{8}$$

(1)式が高次の多項式であれば，基礎関数によるいくつかの部分記述式によって置きかえることができる。即ち

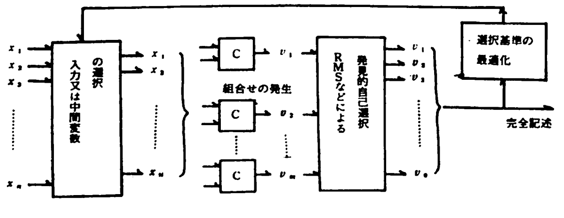
$$v_1=f\ (x_1,\ x_2),\ v_2=f\ (x_1,\ x_3),\ \cdots\cdots v_m=f\ (x_{n-1},\ x_n) \tag{9}$$

$$w_1=f\ (v_1,\ v_2),\ w_2=f\ (v_1,\ v_3),\ \cdots\cdots w_p=f\ (v_{m-1},\ v_m) \tag{10}$$

ここに $m={}_nC_2$，$p={}_mC_2$，……

入力変数 $x_1, x_2,$ ……, x_n から中間変数 $v_1, v_2,$ ……, v_m 及び w_1，w_2，……，w_p と変換するにつれ近似度は高まり複雑になっていく。また，基礎関数として一般に(6)式を用いるが，非線形で低次の場合は(7)式を，線形であれば(8)式を用いれば良い。

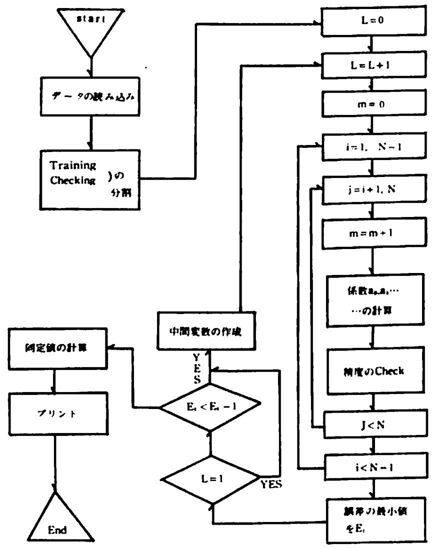
最終的に得られる最適な近似度を持つ多項式は，出力 y と中間変数 v; w, ……との二乗平均誤差が最小になる中間変数とし，これを出力 y の完全記述式という。GMDH の基本構造を図—1に示した。



図—１　GMDHの基本構造

実際に適用するに当って，正則性を高めむだな計算を省くために次の二操作を行う。一つは，データを二つのグループに分割し，一方で中間変数の係数を決定し他方でその精度を調べることである。前者を Training 操作，後者を Checking 操作という。他の一つは，中間変数の数を入力変数の数と同じ程度の大きさにしてむだな計算を省くことである。これは中間変数の発生が二個の変数の組合わせによっているので，層を通過する回数が増す程膨大な数になりこれを防ぐためである。

以上の計算の流れを，図—2のフローチャートに示した。



図—２　GMDHアルゴリズムのフローチャート

3　雄物川の水質問題への適用

GMDH は，関係があると思われるがその因果関係が明らかでない多くの変数の中から，重要な変数を発見して，最適な複雑さを持つ多項式を作り出す機能を持っている。従って，環境問題——水質汚濁問題などの本質的に複雑なシステムを持つ問題の解決に適用することができる。

3—1　評　　価　　方　　法

雄物川の水質が，流量，降雨量，気温などの水文気象因子により大きく影響されることは，これまでの調査から明らかに認められる[3]。このような河川について，時間的な水質変動は，t を短期間として周期成分を除いて考えると一般に次のように表わすことができる。

$$W\ (t)=W_T(t)+W_R(t) \tag{11}$$

ここに，$W\ (t)$：t 日の水質，$W_T(t)$：傾向成分，$W_R(t)$：ランダム成分

$W_T(t)$は，河川の持つ物理的な流送作用や溶解機構

などにより持たらされる成分である。これを t の関数型で表現できるものとし，水質と流量との基礎的関係から次の式を用いた。

$$W_T(t)=a\cdot Q(t)^b \tag{12}$$

ここに，$Q(t)$：t 日の流量，a，b：係数

a，b については，これまでの調査を基に，年間を通した全ての季節のデータから求めた回帰係数を使用した。

これに対し $W_R(t)$は，全てランダムなものと考えることもできないが，不規則に変動する成分を表わす。河川の水質評価や予測には，$W_T(t)$と同様に $W_R(t)$の構造を明らかにすることも重要である。そこで $W_R(t)$については，統計的にこの構造を求めることにし，前述の GMDH を適用してみた。この際，関係のある変数として，流量，降雨量，気温の三つを選び，それぞれの当日を含めた数日間の資料を使用した。即ち

$$W_R(t)=f\ (Q(t,\ t-1,\ t-2),\ R\ (t,\ t-1,\ t-2)\ T\ (t,\ t-1,\ t-2),\ W\ (t-1,\ t-2)) \tag{13}$$

ここに，R：降雨量，T：気温

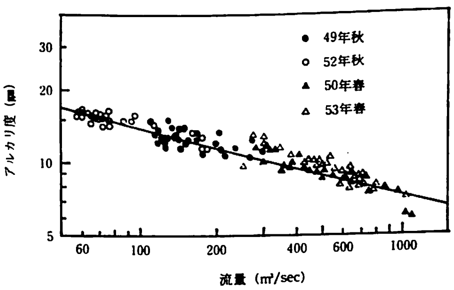
基礎関数としては，非線形システムと考えて(6)及び(7)式を用い，比較検討した。また，最適構造式の決定は，残差の大小のみによらずなるべく簡単な構造式を用いることにした。これは，構造式が複雑になる程 Training Data への適合性は良くなるが，これを用いた評価や予測は Over fitting して逆に悪くなる傾向があるためである。

3－2　結果と考察

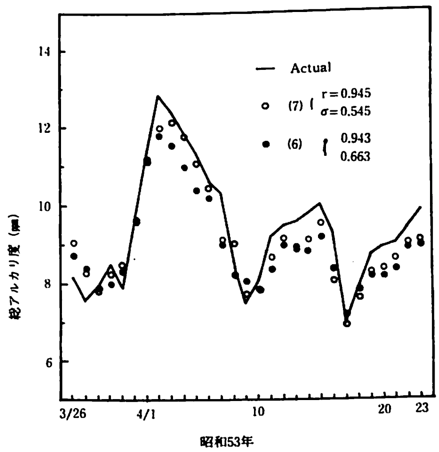
1）総アルカリ度

表－1に，調査期間内の水文気象資料をまとめて示した。

図－3にアルカリ度と流量との関係を示した。これによると，アルカリ度については，年度毎の相違もあまり認められず流量との関係で水質を求めることが可能と思われる。用いた回帰直線の方程式は次の通りである。



図－3　アルカリ度と流量



図－4　アルカリ度の評価

$$W_T(t)=49.7\cdot Q(t)^{-0.278} \tag{14}$$

秋期は，河川の流量が比較的安定する時期で，アルカリ度もほぼ(14)式を満足して増減しておりランダム成分の評価も必要ないと考えられる。これに対し春期は，53年の値が50年と比べて弱干高い傾向を示した。従ってこの傾向を構造式で表現できるかどうかを検討するために，50年のデータで構造式を決め，これに53年の水文因子を当てはめてアルカリ度の評価を行なってみた。結果を図－4に示す。用いた構造式は次の通りである。

表－1　水文因子の季節別平均値

季節	水文因子	流量 (m³/sec) min	max	av	sd	降雨量 (mm/日) 降雨日数 (日)	max	av	気温 (°C) min	max	av	sd
春期	50年	281	1,100	567	208	17	22.6	3.1	−0.76	13.5	7.16	3.66
	53年	253	1,046	544	184	20	12.0	2.4	−0.47	12.5	5.37	3.05
秋期	49年	107	303	171	53.5	22	33.3	5.6	0.83	14.6	8.67	3.33
	52年	59.2	174	83.2	33.6	17	21.1	3.1	1.40	16.3	10.5	3.60

$W_R(t)=1.15+0.178W(t-2)-2.79T(t-2)-0.0398W(t-2)T(t-2)-0.0480W(t-2)^2+1.50T(t-2)^2$ (15)

$W_R(t)=-0.0734+0.323W_1(t)+0.492W_2(t)+0.854W_1(t)W_2(t)$ (16)

ここに $W_1(t)=0.287+0.624W(t-1)-0.368T(t-2)-0.100W(t-1)T(t-2)$

$W_2(t)=0.983-0.0270Q(t-2)-1.12T(t-1)+0.0271Q(t-2)T(t-1)$

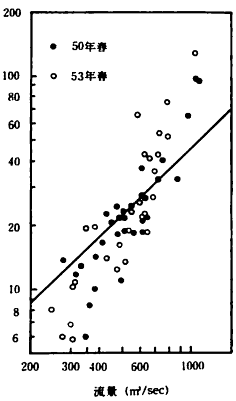
ただし流量は1/100，降雨量と気温は1/10してある。

図—4より，評価値は，53年の弱干高い水質を必ずしも十分に表現できてはいないが，両式とも水質変動によく追随して変動しており，50年のデータによる構造式が53年にも比較的良く当てはまることが認められる。

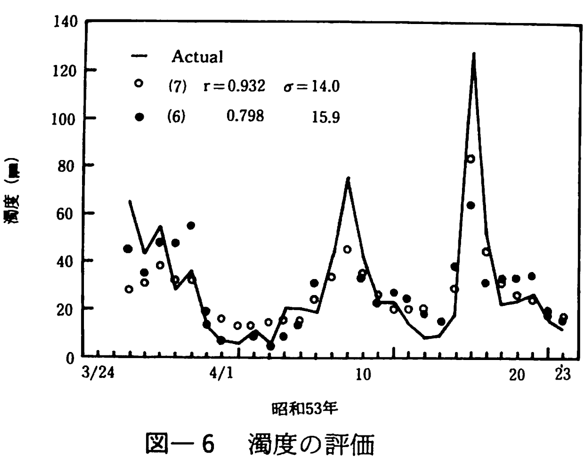
2）濁　度

流量と濁度との関係を，春期について図—5に示す。図の回帰直線の方程式は次の通りである。

$W_T(t)=0.0349\cdot Q(t)^{1.04}$ (17)



図—5　濁度と流量



図—6　濁度の評価

図から，53年のデータが(17)式によく当てはまっていることが認められる。図—6には，50年の構造式に53年のデータを当てはめた結果を示す。この時のGMDHの構造式は次の通りである。

$W_R(t)=-1.57+0.883W_1(t)+0.375W_2(t)+0.0150W_1(t)W_2(t)+0.00903W_1(t)^2+0.0658W_2(t)^2$ (18)

ここに $W_1(t)=-41.1+11.9Q(t)+4.79T(t-2)-10.1Q(t)T(t-2)+0.0662Q(t)^2+19.6T(t-2)^2$，$W_2(t)=-1.03+8.98T(t-2)-2.70R(t-1)+19.0T(t-2)R(t-1)-10.9T(t-2)^2-3.49R(t-1)^2$

$W_R(t)=-3.38+0.594W_1(t)-0.537W_2(t)+0.0323W_1(t)W_2(t)$ (19)

ここに $W_1(t)=-10.7-3.91Q(t-2)+4.29Q(t)+0.374Q(t-2)Q(t)$

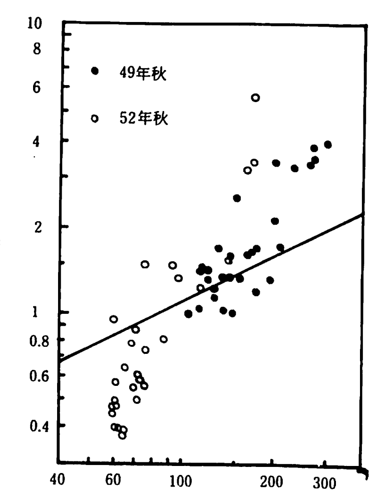
$W_2(t)=-30.9+5.99Q(t)-0.693W(t-2)+0.0749Q(t)W(t-2)$

これによると，この期間の最大値に対する評価は約半分と良くないが，これを除けば水質変動によく合致しており，極大や極小値への追随性もかなり良くなっている。従って，水文因子による水質の評価がかなり有効なことを示していると思われる。また，構造式には，流量の他に降雨量と気温が選択されており，融雪期の流出機構を考える上でこれらの変数が重要なことを示している。

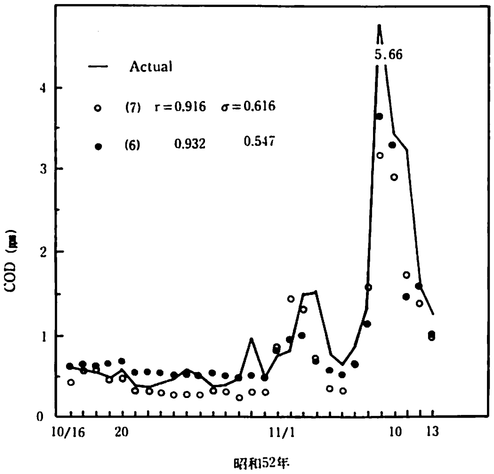
3）COD

図—7に，CODと流量との関係を示した。図の回帰直線式は次の通りである。

$W_T(t)=0.100Q(t)^{0.523}$ (20)



図—7　CODと流量



図—8　CODの評価

これによると，52年の秋期は流量がかなり少なく，この時の COD は(20)式を大きく外れている。また，降雨時の極値はほぼ49年と同程度であった。

他の項目と同時に，GMDH を当てはめて検討してみたが，用いた構造式は次の通りである。

$W_R(t)=-0.122+1.12W_1(t)-0.655W_2(t)-5.29W_1(t)W_2(t)+2.46W_1(t)^2+3.09W_2(t)^2$　(21)

ここに $W_1=-1.52+0.171Q(t-2)+1.66Q(t)-0.681Q(t-2)Q(t)+0.0739Q(t-2)^2+0.174Q(t)^2$, $W_2(t)=-0.994+0.707Q(t)+0.410W(t-2)-0.617Q(t)W(t-2)+0.0714Q(t)^2+0.142W(t-2)^2$

$W_R(t)=0.483+1.59W_1(t)+1.91W_2(t)+2.91W_1(t)W_2(t)$　(22)

ここに $W_1(t)=0.686-0.501Q(t-1)+0.146Q(t)+0.0321Q(t-1)Q(t)$

$W_2(t)=-0.645+0.249Q(t)-0.438T(t-1)-0.155Q(t)T(t-1)$

図—8に結果を示したが，これによると評価値は，降雨後のピーク値に対してはあまり良くないが，特に降雨のない時のかなり小さな値に良く追随して変化しており，全体として良く水質変動を表わしていると思われる。従って，秋期については，流量範囲が49年と52年とではかなり異なったが，水質の流出構造は類似していると思われ，このような方法が水質変動の評価にかなり有効であると思われる。

融雪期については，50年と53年とでは変動の傾向が異なり，水質評価値が多少ずれる傾向が認められた。従って，水質評価には，傾向成分の把握も大きな意味を持つわけで，更に検討を加えたい。

表—2　各水質の相関係数と残差の標準偏差

		53 Spr. GMDH (XY)	53 Spr. GMDH (X^2Y^2)	52 Aut. GMDH (XY)	52 Aut. GMDH (X^2Y^2)
Tur.	r	0.932	0.798	0.885	0.893
	σ	14.0	15.9	2.78	2.72
SS	r	0.901	0.921	0.974	0.694
	σ	44.9	42.0	2.66	7.50
COD	r	0.495	0.449	0.916	0.932
	σ	1.60	1.64	0.616	0.547
Alka.	r	0.945	0.943		
	σ	0.545	0.663		
Hard.	r	0.782	0.140		
	σ	1.86	4.67		
Chlo.	r	0.879	−0.128		
	σ	1.07	3.67		
NO_3-N	r	0.877	0.840		
	σ	0.064	0.089		
BOD	r	0.073	0.054		
	σ	1.48	1.49		

なお，各水質について，相関係数と残差の標準偏差をまとめて表—2に示した。

以上，水質や季節によって，GMDH による水質評価がかなり良く合う場合とまだ弱干不十分な場合とあることが認められたが，雄物川のような人為汚濁の少ない河川については，水文，気象条件が類似していればかなりの程度の水質評価が可能であり，基本的に有効な方法であると思われる。

4 ま と め

年度は異なるが，同じ月日の1ヶ月の水質調査データを基に，水質を傾向成分とその他の成分に分けそれぞれに関数型を与えて一方から構造式を導き，これを他方に当てはめて水質評価を行なった。結果を要約すると次の通りである。

1）季節や水質により，傾向成分が同じような場合とかなり異なる場合があり，水質評価には，この傾向成分の把握がかなり重要である。水質によっては，この成分のみで水質評価を行なうことも可能である。

2）ランダム成分の評価は，水文，気象条件が類似している構造式を用いればかなり有効である。が水質毎の変動特性を十分に把握し，それぞれの水質に合った方法が必要である。

3）GMDH の基礎関数の違いは，水質評価にかなり大きな影響を持っている。一般に非線型の(7)式の方が二次多項式より良い結果を与えたが，水質の流出構造を知る上で興味が持たれる。複雑な原因から成る河川の水質を評価するの

に，GMDH などの統計的方法を用いて構造式を決めてゆくという方法は重要と思われ，今後共水質や季節による相違について検討を進めてゆくつもりである。

謝　　辞

本研究については，進藤康幸，牧野一及び三浦一行の諸君の援助を得た。ここに記して謝意を表します。

参　考　文　献

1）羽田守夫，水文因子による河川の水質変動の評価について，第6回環境問題シンポジウム講演論文集，pp71～77，1978年8月

2）田中雅史他，GMDH の概要と適用例，農業土木学会誌，第45巻第8号，pp539～545，1977

3）羽田守夫，雄物川の水質の変動特性とそのモデル化，土木学会論文報告集，No.268，pp73～81，1977