

学習の方法による流出予測に関する研究

北海道大学工学部 学生員 ◎ 神田 格

同 上 正 員 藤田睦博

1 はじめに

最近になり、流出系の同定、予測に学習の手法が採り入れられている。この代表的なものは、カルマン理論及びGMDH理論である。カルマン理論は、流出予測の際のオンライン性に優れ、GMDH理論は非線型な現象に適している。ここでは、まず、GMDH理論を流出系に適用するにあたっての問題点をシュミレーション法及び、実測資料を用いて吟味したものである。

2 GMDH理論¹

A.G. Ivakhnenkoにより提唱された手法で、システムの構造が複雑で高次の非線形性をもつ系について発見的自己組織化法(Heuristic Self-Organization)を適用することにより、少数のデータで高次の多項式近似を行なって、系の構造式を決定しようとするものである。

一般化されたたたみ込み積分式(Vollterra Series)

$$q(t) = \int_{-\infty}^{\infty} h_1(\tau) r(t-\tau) d\tau + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) r(t-\tau_1) r(t-\tau_2) d\tau_1 d\tau_2 + \dots \quad (2.1)$$

が、降雨-流出系に成立しているものとする。 $q(t)$, $r(t)$ は、それぞれ流量、雨量を表わす。又物理的実現性の条件より、各核は次式を満足しているものとする。

$$h_1(t) = h_2(t_1, t_2) = \dots = 0 \quad t, t_1, t_2, \dots < 0 \quad \dots \quad (2.2)$$

(2.1)式を差分化すると

$$q_t = a_0 + \sum_{i=1}^N a_i r_i + \sum_{i=1}^N \sum_{j=1}^N a_{ij} r_i r_j + \dots \quad \dots \quad (2.3)$$

更に(2.3)式を変形して

$$q_t = f(r(t-\Delta t), r(t-2\Delta t), \dots, r(t-m\Delta t), q(t-\Delta t), q(t-2\Delta t), \dots, q(t-n\Delta t)) \quad (2.4)$$

を考える。発見的自己組織化法は、(2.3)式の関数形を次のような手順で決めるものである。

step 1 n と m を決める。実際には相関係数より推定する。

step 2 原データをtrainingデータとcheckingデータに区分する。この場合、区分法に特別のルールはなく、データの前半、後半としてもよいし、交互に区分してもよい。

step 3 $x_1=r(t-\Delta t), x_2=r(t-2\Delta t) \dots, x_N=q(t-N\Delta t)$, ($N=n+m$) の N 個の入力変数についてtrainingデータを用いて中間変数 $Z_K^{(1)}$ よりなる次式を求める。

$$Z_K^{(1)} = a_{K0} + a_{K1}x_1 + a_{K2}x_2 + a_{K3}(x_i)^2 + a_{K4}(x_j)^2 + a_{K5}x_i x_j \quad \dots \quad (2.5)$$

ただし $i \neq j$, $i, j < N$, $K=1, 2, \dots, N(N-1)/2$

(2.5)式より、係数 a_{Kl} $l=0, 1, \dots, 5$ を最小二乗法で求める。

step 4 checkingデータを用いて誤差の二乗平均値 $\epsilon_K^2 = (q_t - Z_K^{(1)})^2$ を求め、 ϵ_K^2 の小さい方から M 個の中間変数を選ぶ。最小の ϵ_K^2 を $\epsilon_K^{2(1)}$ とする。

step 5 選び出された M 個の中間変数を $X_i = Z_i^{(1)}$, $X_j = Z_j^{(1)}$, $i, j(M)$ において step 3 にもどる。 n 回目の ϵ_K^2 の最小値を $\epsilon_K^{(n)}$ とおくと、 $\epsilon_K^{(n)} > \epsilon_K^{2(n-1)}$ になるまで step 3 ~ step 5 をくり返し、 $\epsilon_K^{2(n-1)}$ を与える $Z_K^{(n-1)}$ をもって構造式とする。

ここでは(2.5)式において中間変数を二次式で与えたが、必ずしも二次式でなくともよい。

3 修正GMDH理論

前述のGMDH理論を用いて、構造式の既知であるデータによって検討した結果を示す。特に、中間変数のとり方を問題とした。まず、適当に構造式を仮して q_t, r_t を定めて、次に構造式を隠してGMDH法により構造式を推定した。採用した構造式を次に示す。次数の最高は3次で、比較的線形項のウェイトを大きくしてある。

$$q_t = 0.5r_{t-1} - 0.6r_{t-3} + 0.3q_{t-1} - 0.4q_{t-2} \\ + 0.05r_{t-1}q_{t-3} - 0.04r_{t-2}q_{t-4} + 0.06r_{t-5}q_{t-1}q_{t-5} + 0.009 \quad (3.1)$$

r_t に[0~1]の矩形乱数を発生させて60個の q_t, r_t をデータとして、前半の30個を交互にtraining, checkingにとり、後半の30個のデータの適合度で計算法の良否を判定した。計算法としては、次の3ケースを採用した。

Case 1 前章のGMDH理論と全く同じ。

Case 2 Case 1を行ない構造式を推定する。この構造式と $r_{t-1}, r_{t-2}, r_{t-3}, \dots, r_{t-N}$ から Q_t を求めて、これを $\hat{Q}_t$ とおく。次式で表わされる $\tilde{Q}_t$ を用いて、再びGMDH理論を

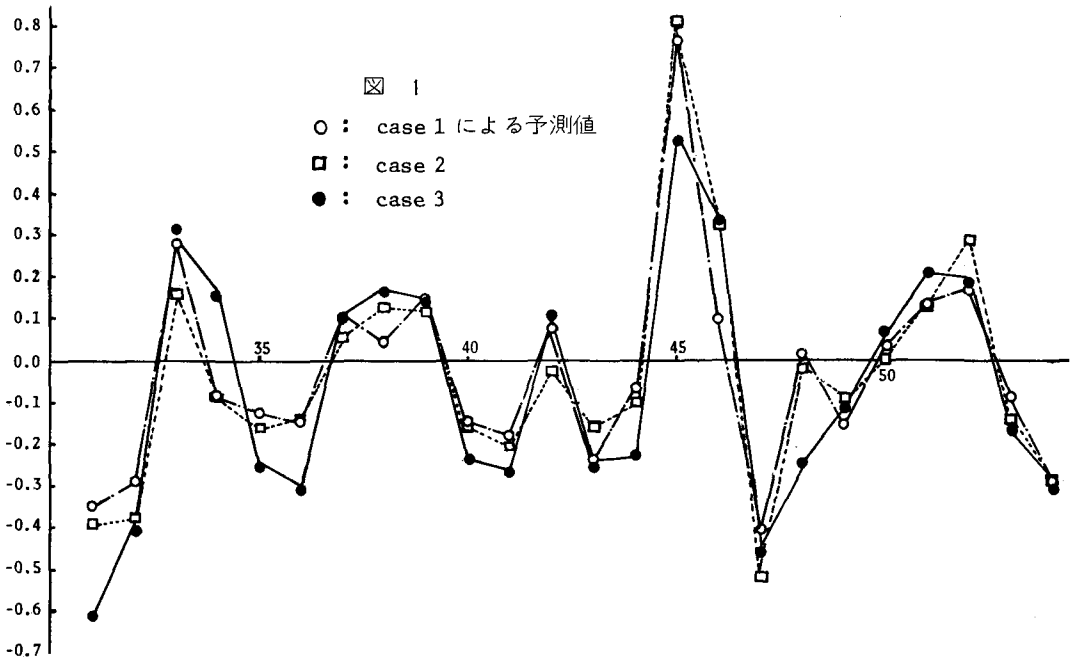
$$\tilde{Q}_t = \hat{Q}_t - Q_t \quad (3.2)$$

適用し、ここで中間変数を3次式とする。すなわち

$$\tilde{Q}_t = a_{K0} + a_{K1}x_i + a_{K2}x_j + a_{K3}(x_i)^2 + a_{K4}(x_j)^2 + a_{K5}x_i x_j \\ + a_{K6}(x_i)^3 + a_{K7}(x_i)^2 x_j + a_{K8}x_i (x_j)^2 + a_{K9}(x_j)^3 \quad (3.3)$$

Case 3 Case 2と同じであるが、1回目では中間変数を1次式、2回目では2次式としたものである。

各ケースについて、得られた構造式を次に示す。ただし、項数が多いので2次以上の項を省略した。ケース1では、2回目で構造式が決定された。ケース2については、初回で2回目に、次回で1度で構造式が定まった。ケース3では、初回が5回目、次回が1度で求まった。



Case 1

$$q_t = 0.195r_{t-1}^{-0.301}r_{t-3}^{+0.108}q_{t-2}^{+0.398}(r_{t-1})^2 - 0.166(r_{t-3})^2 + 0.048(q_{t-2})^2 - 1.149r_{t-1}r_{t-3}^{-0.722}r_{t-1}q_{t-2}^{+0.092}r_{t-3}q_{t-3}^{+0.003} \quad (3.4)$$

Case 2

$$\begin{aligned}
q_t = & 0.195r_{t-1} - 0.301r_{t-3} - 0.957r_{t-4} + 1.118r_{t-5} + 0.108q_{t-2} + 0.395(r_{t-1})^2 - 0.166(r_{t-3})^2 \\
& + 2.190(r_{t-4})^2 - 1.461(r_{t-5})^2 + 0.048(q_{t-2})^2 - 1.149r_{t-1}r_{t-3} - 0.722r_{t-1}q_{t-2} \\
& + 0.092r_{t-3}q_{t-2} - 1.458r_{t-4}r_{t-5} - 0.008
\end{aligned} \quad (3.5)$$

Case 3

$$q_t = 0.509r_{t-1} - 0.548r_{t-3} + 0.0005r_{t-5} + 0.272q_{t-1} - 0.358q_{t-2} + 0.013(r_{t-1})^2 - 0.016(r_{t-5})^2 - 0.011r_{t-1}r_{t-5} + 0.009 \quad (3.6)$$

この例では、ケース3の計算法が(3.1)式に近い係数を同定している。(3.1)式の係数及び変数をかえた他の計算例では、ケース1,2の場合でも与えた構造式を近似する結果が得られる時もあるが、その場合でも常にケース3の計算法がケース1,2と同等もしくはそれ以上の解が得られている。すなわち、中間変数は、最初から高次の式を用いるよりも1次,2次と順に次数を上げていく方が効果的であると思われる。

4 実測資料による解析

1967年～1971年の多度志における日流量及び沼田、幌加内、朱鞠内における日降雨量資料を用いた計算例を示す。夏期の台風性の洪水予測を考えているので、8、9月を対象して、降雨量に関しては3地点の雨量資料より流域平均雨量を用いた。

表1は、8,9月の日流量,流域平均雨量を用いて前述のケース3の計算方法で構造式を求めたものである。構造式が複雑で、各々の月の構造式を図表化したものである。すなわち、第一行は線形項を表わし、第二行以下は非線形項を示す。表中の数字は、表2に示すように年度を表わす。又、カッコは各項の係数のオーダーを示し、記号との対応を表3に示す。横軸と縦軸に対応している変数の積の項が表われており、年度及び係数のオーダーが数字とカッコで与えられる。

表によると比較的同じ場所に印が固まっており、年度異なっても月毎に同様な流出特性を示している。予測問題を考える時、9月の予測をするには8月のデータより構造式を推定して9月に適用するよりも、前年度の9月のデータを用いて構造式を同定してそれを用いて今年度の9月に適用する方がよいこ

表 2

1	:	1967
2	:	1968
3	:	1969
4	:	1970
5	:	1971

表 3

(1次項係数)

無印: 10^{-2}

() : 10^{-1}

 $\{ \} : 10^0$

(2次項係数)

無印： 10^{-4}

() : 10^{-3}

 $\{ \} : 10^{-2}$

表 1

8月	R_{E-2}	R_{E-2}	R_{E-3}	R_{E-4}	R_{E-5}	R_{E-6}	R_{E-7}	Q_{E-7}	Q_{E-2}	Q_{E-3}	Q_{E-4}	Q_{E-5}	Q_{E-6}	Q_{E-7}	Q_{E-8}
	1 2 (3)	1 (2)	(2)	1 (2)	3			(1) (3)	(2)	(1) 3	(1) (3)	3	2		1
	(2) (3) (4)	4	(2) 4	(2)	(3) (4)	4		(4)	(2) (4)	4	(2) (3) (4)	(3) 4	(2) 4	3	(3)
		(1)		1						(1)	(1)				R_{E-1}
			(2) (4)	(2) 4	(4)	(4)			(2) (4)	(4)	(2) 4	(4)	(2) (4)		R_{E-2}
				(2)	(4)				(2) (4)	(1) 4	(1) (2)	4	(2) 4		R_{E-3}
					(3) (4)	4			(4)	(3) (4)	(3) (4)		3	(3)	R_{E-4}
									4	(4)	4				R_{E-5}
															R_{E-6}
Q_{E-8}															R_{E-7}
Q_{E-7}		(1)	5												Q_{E-1}
Q_{E-6}		(1)	(5)	(1) (4) (5)											Q_{E-2}
Q_{E-5}	3	1	3 (5)	1 (4) (5)	2 (3) (4) (5)				(2) (4)	(4)	(2) (4)	(4)			Q_{E-3}
Q_{E-4}		(1)		(1)	(1) (2)	(1) (2)			(1) 4	(1) 4	(4)	2			Q_{E-4}
Q_{E-3}					(2)	(2)	(2)				(1) (2) (3) (4)	(3) (4)	2 (4)	3	(3)
Q_{E-2}															Q_{E-5}
Q_{E-1}															Q_{E-6}
R_{E-7}				(4)	(4)			4							Q_{E-7}
R_{E-6}	3	1	3 (5)	(1) (4) (5)	1 (3) (4) (5)	(1)			4		1 (3) (4) (5)				R_{E-1}
R_{E-5}		1		1	2	(1) (2)					1				R_{E-2}
R_{E-4}					(2)	(2)	(2)					2	(2)		R_{E-3}
R_{E-3}							(2)								R_{E-4}
R_{E-2}				(4)	(4)				(4)	(4)					R_{E-5}
R_{E-1}	(3)	1	3 (5)	(1) (5)	1 (2) (3) (5)	(1) (2)	(2)				1 (3) (5)	1 (2)	(2)		(1) (2) (3) (5)
1			5	3 (4) (5)	(3) (4)	(1) 2	(2)	(1) 2 (5)	1 (2) (5)		1 (3) (4) (5)	2		(1) 2 (4) (5)	(2) (3) (4) (5)
	Q_{E-8}	Q_{E-8}	Q_{E-7}	Q_{E-6}	Q_{E-5}	Q_{E-4}	Q_{E-3}	Q_{E-2}	Q_{E-1}	R_{E-7}	R_{E-6}	R_{E-5}	R_{E-4}	R_{E-3}	R_{E-2}

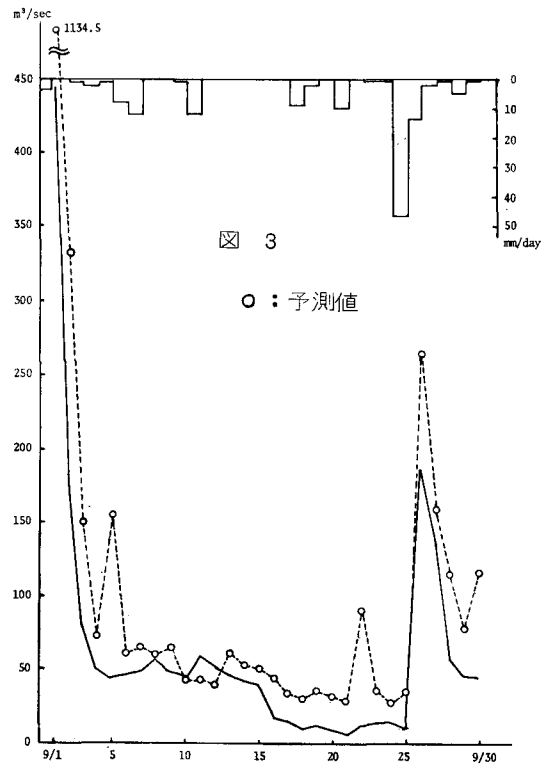
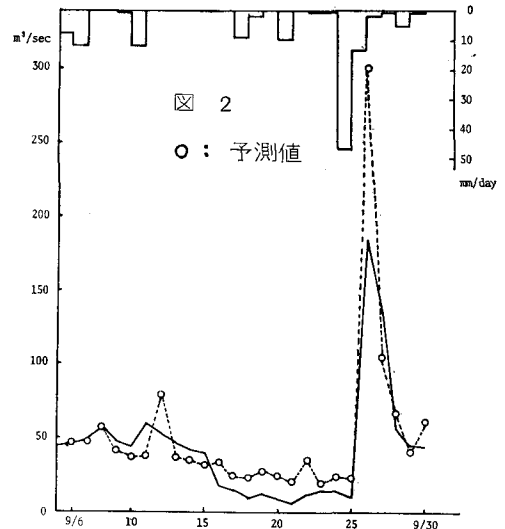
とを示唆している。特にこの傾向は、9月について強く表われている。常にこのようなことが成立するとは限らないが、予測とは一種の外挿であるから構造式が大きく変化している場合には予測の精度が低く、事前の検討が必要なのは当然である。

図2は、1969年8月5日～9月5日までのデータで構造式を同定し、9月6日～9月30日までの流量を予測(1日だけ将来)をしたものである。又、図3は1968年9月のデータで構造式を同定し、1969年9月の流量を予測したものである。いずれもケース3の計算法を用いたものである。図2を得るために用いた構造式は

$$\begin{aligned}
 q_t = & 0.129r_{t-1} + 0.025r_{t-5} + 0.360q_{t-1} \\
 & + 0.010q_{t-3} + 0.023q_{t-4} + 0.016q_{t-5} \\
 & + 0.00702(r_{t-1})^2 - 0.00706(r_{t-5})^2 \\
 & - 0.0209(q_{t-4})^2 - 0.00271(q_{t-5})^2 \\
 & + 0.0275r_{t-1}r_{t-5} + 0.0243r_{t-5}q_{t-4} \\
 & - 0.0474r_{t-1}q_{t-4} - 0.00874r_{t-5}q_{t-5} \\
 & + 0.0171r_{t-1}q_{t-5} - 0.00024r_{t-5}q_{t-7} \\
 & + 0.00047r_{t-1}q_{t-7} - 0.00160r_{t-5}q_{t-8} \\
 & + 0.00313r_{t-1}q_{t-8} \\
 & + 0.0150q_{t-4}q_{t-5} - 0.00015q_{t-5}q_{t-7} \\
 & + 0.00042q_{t-4}q_{t-7} - 0.00099q_{t-5}q_{t-8} \\
 & + 0.00276q_{t-4}q_{t-8} + 1.848 \quad (4.1)
 \end{aligned}$$

図3を得るために用いた構造式は

$$\begin{aligned}
 q_t = & -0.163r_{t-1} + 0.027r_{t-2} - 0.017r_{t-4} \\
 & + 0.565q_{t-1} + 0.057q_{t-2} + 0.140q_{t-3} \\
 & - 0.079q_{t-4} + 0.00510(r_{t-1})^2 \\
 & + 0.00274(r_{t-4})^2 + 0.0758(q_{t-3})^2 \\
 & + 0.0242(q_{t-4})^2 + 0.00025(q_{t-5})^2 \\
 & + 0.0310r_{t-1}r_{t-4} - 0.0289q_{t-3}r_{t-4} \\
 & + 0.00078r_{t-4}r_{t-5} + 0.00231r_{t-5}q_{t-4} \\
 & + 0.0044r_{t-1}r_{t-5} - 0.0041q_{t-3}r_{t-5} \\
 & + 0.0163r_{t-4}q_{t-4} + 0.00023r_{t-5}q_{t-5} \\
 & - 0.163r_{t-1}q_{t-3} - 0.0859q_{t-3}q_{t-4} \\
 & + 0.00164r_{t-4}q_{t-5} + 0.00488q_{t-4}q_{t-5} \\
 & + 0.0921r_{t-1}q_{t-4} - 0.00866q_{t-3}q_{t-5} \\
 & + 0.00929r_{t-1}q_{t-5} + 0.101 \quad (4.2)
 \end{aligned}$$



GMDH理論は、 q_t, r_t に関して非線形な構造式を同定するが各項の係数が固定している。一種の非線形ブラックボックス問題として流出系をとらえているわけである。一方、カルマン理論は非線形な系にも適用できるが、非線形項の変数の組合せは別途に決めてやらねばならない。ただし、状態変数は時間が経過して情報量が増える毎に変化し、この意味で変係数の構造式を同定することになる。したがってGMDH法で非線形項の変数の組合せを決定し、これにカルマン理論を適用すると両者の長所利用できるはずである。次のような構造式を仮定する。(3.1)式よりも非線形項のウエイトを大きくしている。

$$q_t = 0.05r_{t-1} - 0.03r_{t-2} + 0.04q_{t-1} - 0.02q_{t-2} + 0.5(r_{t-1})^2 - 0.3r_{t-1}q_{t-1} + 0.2r_{t-1}q_{t-2}r_{t-2} + 0.009 \quad (4.3)$$

前に述べたようにケース3の計算法で構造式を同定し、この構造式とカルマン理論を組合せて予測した結果とカルマン理論だけを用いて予測したものを図4に示す。図5は、GMDH法だけによる予測値である。カルマン理論だけの場合、構造式として非線形項を省略した次式を用いた。

$$q_t = \sum_{i=1}^5 a_i r_{t-i} + \sum_{i=1}^5 b_i q_{t-i} \quad (4.4)$$

上式と(4.3)式を非較すると非線形項がないので、図4に示すようにカルマン理論だけを用いた計算値は他の手法に比較して精度が悪かった。GMDH理論とカルマン理論を併用した手法とGMDH理論だけを用いた結果は大差がない。しかし、系の性質が時間的に変化することは実際の現象でも多くみられ、例えば流出率の変化などが考えられる。実測の資料を用いた計算例では、現在充分な結果を得ていないが、カルマン理論とGMDH理論のCOUPLINGを否定するものではなく今後とも検討したい。

参考文献

- 1) 市川新, 池田三郎: 発見的自己組織化法による水質予測, 土木学会論文報告集246号, 1976
- 2) 日野幹雄: 水文流出系予測へのカルマンフィルター理論の適用, 土木学会論文報告集, 221号, 1974

