発見的自己組織化法による水質予測

東京大学　正員　　市川　新

1. 研究の目的

水質現象を決定する因子の数は多く，かつそれぞれの因子が与える影響は複雑で簡単な関数形で表現することが出来ない。このような現象に対して数学的モデル，ないし物理的モデルを用いて表現する試みが数多くなされてきたが，それが実際河川に適用され工学的に利用しうる情報をえるのは，きわめて限られたケースであった。水文学における流出機構の解析においても，数多くのモデルがあるが，その実用性が限定されていたり，それを用いるために，係数等を試行錯誤で決定せねばならないのと同じである。水質現象で，流出機構より悪い条件は，このようなモデルを適用するのに適したフィールドがえられにくいことである。実際河川においては，農業等による取水，支川，排水等の流入，地下への伏流といった水量収支の解明が困難であることと，河床の勾配等による河川断面が一様でなく，水理条件が一様でないことと，これらの条件がなんらかの形で克服されたとしても，モデルの検証に必要な水質データがえられないことがあげられる。水質データの収集に技術と労力も費用がかかるためである。

近年の水質汚濁の進行と共に水質の予測が各方面で要求され，いろいろな提案がなされているが，今の所決定的な手段が提案されていない。その場合に要求されるモデルは，①単純な構造であること，②ソフトな構造であること，③少ないデータで予測出来ることが要求されている。ここで紹介する発見的自己組織化法（ＧＭＤＨ法）は，これら要求されている条件は十分に満足しており，水質予測の分野での利用の可能性が高い。本研究ではＧＭＤＨ法の精度，限界をあきらかにせんとするものである。

2. 対象河川とデータ

対象河川としては多摩川をとり，調布堰（河口から１３Ｋｍ）二子橋（同１７Ｋｍ）川原橋（同２８Ｋｍ）関戸橋（同３４Ｋｍ）日野橋（同４０Ｋｍ）の各点をとり，東京都水道局，建設省をはじめとする各機関での水質観測データを用いた。実際には，流量，ＢＯＤ，ＣＯＤデータである。

3. ＧＭＤＨ法の構造

ＧＭＤＨ法（Goup Method of Data Handling）はＩＶＡＫＨＮＥＮＫＯによって開発された方法でそのフローシートを図－１に示す。第１段のステップとしてデータを集める。ここでは，予測しようとする水質項目とそれに関連あると思われるデータを集めてくる。このデータの中には，データを変換したり，データ同志の積，比をとったりしたものでもよい。又，水質データが，ｎ日前のデータに支配されるとすればｎ日前迄のデータを集めることが出来る。第２段のプロセスは，このデータ群の中から，Ｎコのデータを選択する。Ｎとしては任意であるが，本研究の場合１４～２０とした。

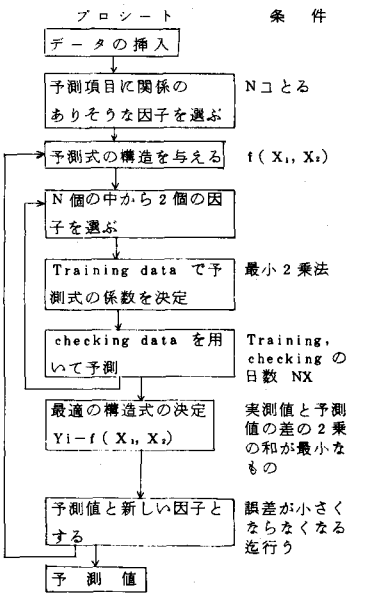
ここでＮをいくつにするかは任意に定められるが，計算等の容量から規制されると思われる。本研究で用いた東大大型計算機センター（ＨＩＴＡＣ８８００／８７００）では，Ｎ＝５０迄とってもあまり問題とはならなかった。

第３段においてあらかじめ予測式の構造を与えておく，この予測式に用いる変数の数，方程式は任意であり，この定め方によって予測精度，予測時間が異なる。② なおここで予測された値を新しい変数として計算するようにループを設けるのが普通であるので，その点を考慮して関数系を定める必要がある。本研究で用いたのは，次式を用いている。

$$y=f(x_1, x_2)=a_0+a_1 x_1+a_2 x_2+a_3 x_1^2+a_4 x_2^2+a_5 x_1 x_2 \quad \cdots\cdots\cdots(1)$$

この式の場合中間表現（この場合ｙに相当する）を新しい変数をとり，この関数系に入れると，元の因子でいえば，４次式を用いたのと同じ結果となる。しかしながら，２次式では，未知数の数は６であり，最小限６コのデータがあれば，未知数がもとめられる。この未知数をとくためには〔６×６〕の行列式の逆行列を

求めることにより，未知数を決定することが出来る。実際には６組のデータ毎に１つの係数が決定されることになる。この場合どの６個のデータを選ぶかによって係数が異なることになり通常は６以上のデータをとり最小２乗法により係数を決定するデータ数をｍとすると（ｍ，６）の行列の逆行列をとることになる。ＧＭＤＨ法は，この制限を逆に利用としているものである。というのはＡ日の水質構造を決定するために，その前のｍコのデータ（これをトレーニングデータという）で構造（係数）を決定し（Ａ＋ｐ）日後の水質構造は，その日の前のｍコのデータによって決定する。このように，最新の情報で常に水質構造を修正しながら，予測を行うという柔構造にすることが可能である。水質は，季節毎に発生パターンが異なるので，冬に決定された水質構造を夏に適用することは，あまり意味がないかもしれない。それよりも，その時期に適した構造で常に修整しながら，フィットさせていく方式が，このような複雑なシステムに対して有効であると考えられる。データの数ｍとして，ここでは１５～３５迄とってみた。前述のような季節的パターンからか，水質データの変動性の故か，あきらかでないが，ｍを大きくとっても必ずしもよい結果がえられていない。③



このようにして，水質構造（係数）が定められるが，それは，(1)式中の（x_1，x_2）という２つの因子に対して決定されるので，この場合，${}_NC_2$　通りの水質構造が出来ることになる。この中で，どの構造式を選ぶかということが問題となる。第４段のプロセスは，因子の選定となる。この場合は，実測値と上で求めた構造式によって決定される予測値との誤差を求めその平方和が最小となるものをもって中間表現とする。このとき対応させるデータのことをチェッキングデータという。前述のトレーニングデータとチェッキングデータのとり方によっても決定される中間表現が異なってくる。ここでいくつか試みている方法は，１日おきにトレーニングとチェッキングデータをとる方法，前半と後半に分離する方法，対象とするトレーニングチェッキングデータの平均値を求め平均値からの差の大きいものｍコをトレーニングデータにし，平均値に近いものを，チェッキングデータにする方法等を試みているが，今の所，どれが最適であるかを断定することが出来ていない。例えば，１日おきに交互にトレーニング，チェッキングデータをとった場合，偶数日をトレーニングとしたときと，奇数日をトレーニングにしたときとで，水質構造の係数は勿論，因子すらも異なってくる。④

第５段は，このようにして決定された中間表現のうち，最小２乗誤差の小さいものＪ組をえらび，それらを新しい因子として再び同じ操作をくり返す。すなわち今後は ${}_JC_2$ 通りの組み合せで，最適中間表現を求める。この第Ｒ次の中間表現と第（Ｒ－１）次の中間表現の予測誤差の平方和を比較し，もしＲ次の中間表現の平方和が小さくなっているときには，更に改良すべき余地が残されているものとして，上述の演算をくり返す。もし，Ｒ次の中間表現が，（Ｒ－１）次の中間表現よりも，平方和が小さくなっていないときには，それ以上の改良が行なわれないものとして（Ｒ－１）次の中間表現をもって最終表現とする。本研究ではＪを５～１５迄とってみたがあまり大きな差がなかったのでＪ＝５で十分と思われる。

ＧＭＤＨ法は，以上のようなシステムとなっているが，ここで，設計法としては、①因子とその数．②構造式，③トレーニング，チェッキングデータの選定，④判定条件（誤差の評価法）が問題となる。これらの問題について，実際の適用から検討を行ってみる。（④を参照されたい）

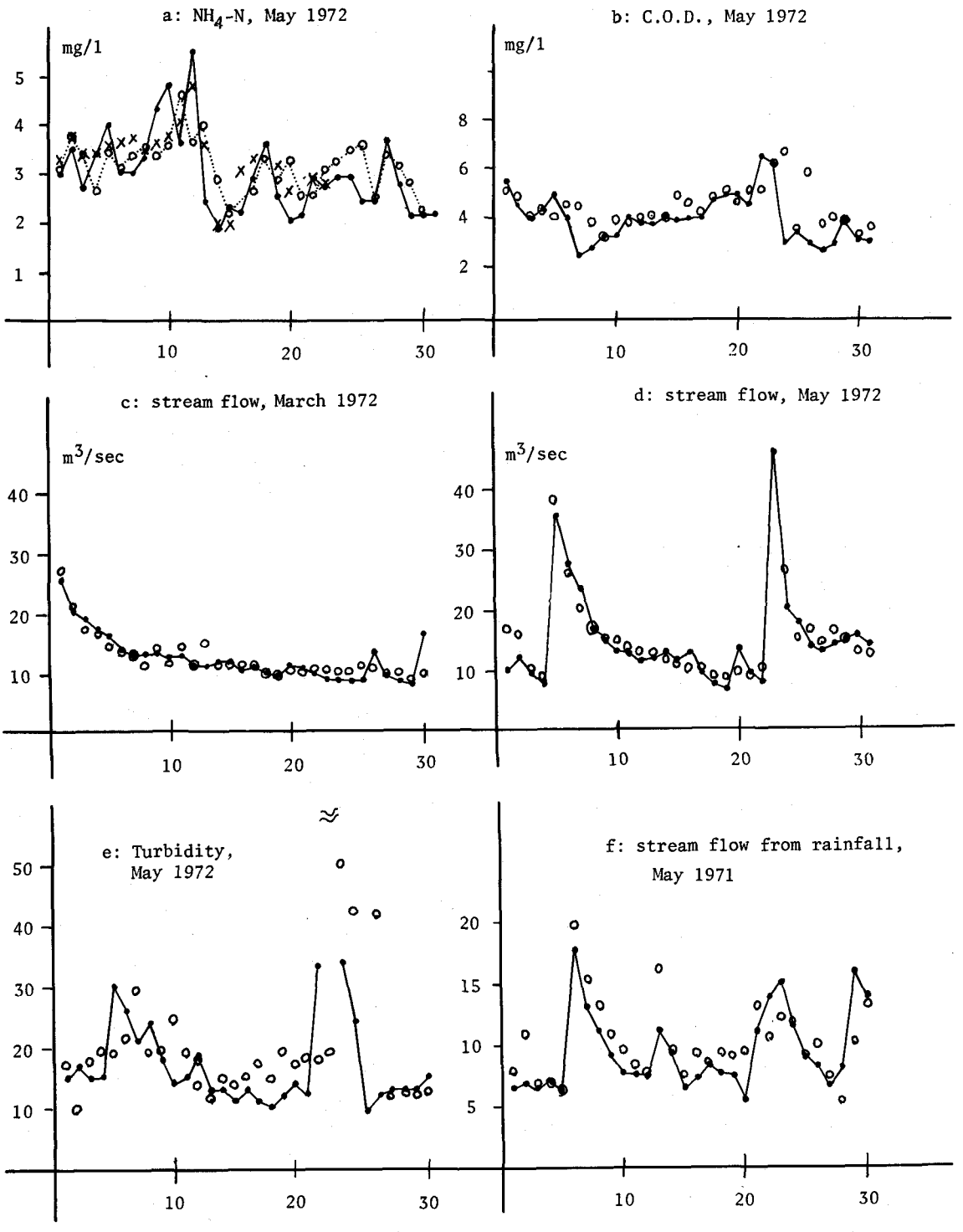
4.　結果と考察

①予測対象項目：予測を行うためには，それに対応しうるだけの水質データがなければならない。その面

Fig—2 Some Examples of Prediction by G.M.D.H. Method

• observed value

o predicted value



から現在における制限が発生してくる。ここで対象としたのは流量，ＢＯＤ，濁度，NH_4－Ｎの各項目であり，その結果を図－２に示した。なお各予測項目とその条件は次の通りである。

	予測項目	場所	年月	因子	摘要
Ⓐ	NH_4－Ｎ	調布堰	72年5月	調布堰　流量，COD，ELC，NH_4－Ｎ	○×はトレーニングとチェッキングの入れかえ
Ⓑ	ＣＯＤ	〃	72年5月	同上	
Ⓒ	河川流量	調布堰	72年3月	上流（川原橋，浅川，小河内，関戸橋）	
Ⓓ	〃	〃	72年5月	同上	
Ⓔ	濁度	〃	72年5月	Ⓐに同じ	
Ⓕ	流量	〃	71年5月	降水量，水位，流量	

② 因子とその数：　水質に関しては，十分なデータが入手出来ないため，手許にある水質データを利用するしか方法がなく，従って因子の数も限られたものになってしまった。この因子の選定においては，物理現象を把握しておくとよい。その例は，図２－ｃ，ｄと図－ｆの２つの比較である。（ｃ，ｄ）の方は，上流の当日の流量がわかっているときにその日の流量を推定する方法であり，上流の流量によって下流流量が規定されるため，流量の予測がきわめて精度の高いものとなっている。(C)は流量の減衰部の予測であるのに対し，(d)は予測期間に２つの大きなピークが発生している場合の予測であるが，この図では，この予測はこれら２つのピークをよくフォローしている。一方(f)は，降雨量水位等から流量を求めたのであるが，これらの因子が必ずしも物理的に定量的に関係を表現出来ないために，予測値も大きく変動している。すなわち，降雨量のうちどれだけが有効降雨量としてきいているのかがあきらかでないからである。なお降雨量も，上中下流域にわけてそれぞれの平均降雨量を因子とするとかなりよい近似がえられた。②

濁度について色々な因子の組合せで推定を行ったのであるが，いずれもよい結果がえられていない。これは濁度の発生機構が複雑で，同じ降雨量，流量でも，汚泥の堆積状態，流量，流速の変化にともない発生濁度が大きく異なるためと考えられる。それ以外の水質項目については，一応この程度のデータで十分予測出来ると思われる。ただ，不要なデータもあると思われるので因子数をもう少しへらすことが可能と思われる。

ここでは，水質・流量とも時系列的にとり扱っているが，この関係について十分な考察がえられていない。最終表現にえらばれる因子の中に，４日前のデータが入ることがある。４日前のデータということは，別の表現をすれば，データ同志無関係であるということである。勿論流量の減衰部の水質を対象とする場合は，このような時系列的な取扱い方は必ずしも有意義な方法とは考えられない。このことは，水質の予測として前日迄のデータを利用して求める方法と，当日のデータを利用して水質を推定する場合と２通り行ったが，後者の方がはるかに精度がよかった。③このことから１日前のデータを使用してもよい結果をえられないのに４日前のデータから予測出来ることは不可能と考えられる。

④ 構造式：　構造式ないし関数形として任意の形をとることが可能である。ここで２元２次形式をとった理由は，特別存在しないが，もっとも一般的な形としてとったものである。しかしながら，中間表現が２次３次となるにつれて，もとの因子の４乗・８乗という項が出てくることになる。このように高いベキ数が，表われかつ，予測精度がよいということは，逆の表現をするなら，その因子はあまり大きな影響を与えないということと同義なのである。すなわち，最適解がえられないことを意味しているのである。それ故，２次３次の中間表現が問題となるような場合には，次数が高くならないような工夫が必要である。一つの案として，(2)式の５元一次式を行ってみたが，(1)式との間に有意な差はえられなかった。

$$y=f(x_1, x_2, x_3, x_4, x_5)=a_0+a_1x_1+a_2x_2+a_3x_3+a_4x_4+a_5x_5 \cdots\cdots\cdots\cdots (2)$$

⑤ トレーニング，チェッキングデータの選択：　この選び方に特定のルールが存在するわけではない。それ故，いろいろな方法を試みたが，これも，どれが最適という解はえられていない。例えば，トレーニングとチェッキングを取りかえても，結果が近々異なり，よくなるときも悪くなるときもある。

⑥ 判定条件：　ここではすべてチェッキングデータにおける予測値と，実測値の差の平方和が最小となる

時を最適な解としている。しかしながら水質，流量いずれにせよ極値といわれる降水時の極大値がとくに大きく，そのときの誤差が，最小のものが最適となってしまうことがある。その結果，実測値の変動をフォローするというよりは，一種の平均値のように，一定の値となっているときが，平方和が最小となるという結果がえられることがある。図２－ｂのＣＯＤもそれに近いケースである。１つの考え方としてこのような特殊なときを除外して予測を行うという方法が考えられる。この場合前述の時系列的な考え方とどう調整するかが１つの問題となってくる。

いろいろな判定基準も考えられるが，どれといって決め手になるものは考えられていない。

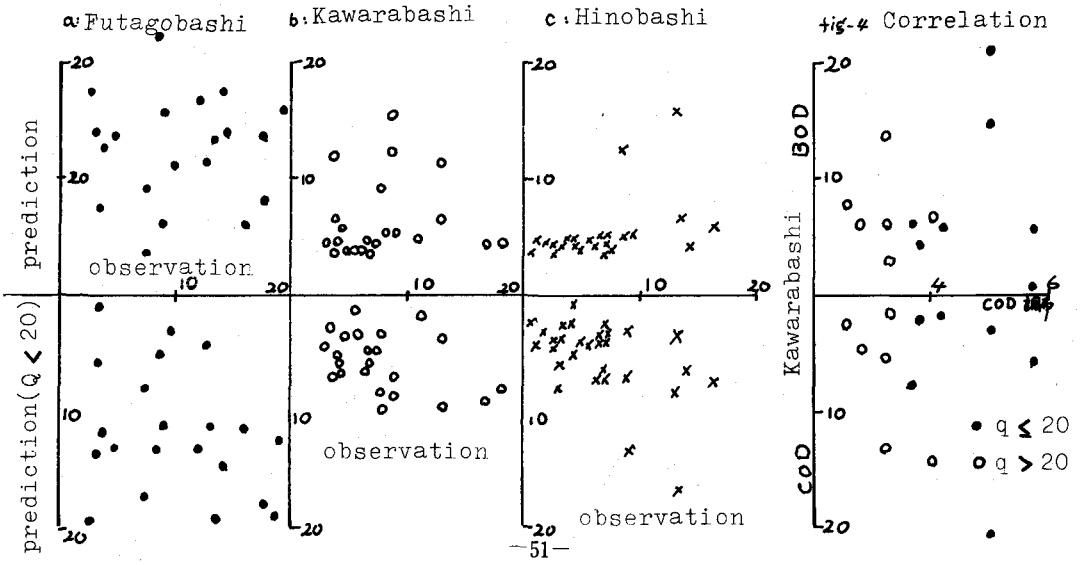
4. ＧＭＤＨ法の応用

本来のＧＭＤＨ法による水質予測の１例を示したが，次にその応用としてのＢＯＤ予測についてのべることにする。ＢＯＤは，分析に５日間かかることと，上述の指標のように毎日のデータがなく，週１回程度のデータしか存在しない。それ故にこそ，その日のうちに他の水質指標からＢＯＤについて精度高い予測を行なうことが要求されるのである。

ＢＯＤのデータが連続的にないことから上述のような方法をとることが出来ない。ここでは，ＢＯＤを多変数表示することとする。通常の多変量表示においては，因子と構造式をあらかじめきめておいて，１群のデータからその係数を決定する方法がとられている。ここではこの変数と構造式をＧＭＤＨ法で決定しようとするものである。このようにすることにより変数の数を多くとり，その中からもっとも適した因子と構造式が決定出来ることになる。ＢＯＤの特性から，その構造を簡単に表現出来ないが，本法を適用することにより，従来の多変量解析によりえられなかったような予測法がえられるのではないかと期待された。

対象として，多摩川の上流部の川原橋，日野橋におけるＢＯＤ，ＣＯＤ流量データと，調布堰におけるＣＯＤ，流量データの５つから１４因子を作製し，約３年間のデータ約７０で構造式を決定し，そのときの予測値を求めた。すなわちこの予測値は，チェッキングに行なわれた予測値に相当するものである。図－３は，川原橋におけるＢＯＤ値とその他の指標の関係図を示したものである。この図をみるとＢＯＤがどの指標に関係があるのが，あきらかでないので次のような因子をとり上げて，多変数解析を行った。

調布堰，川原橋とも流量，同２乗，同逆数，ＣＯＤ，同２乗，負荷重及び両地点のＣＯＤ，流量の比その１例を図－４に示したが，よい結果をえられていない。その理由は，実測されているＢＯＤ値そのものに誤差が含まれていること，流量の変動等により，極値が発生し，その極値によって全体の構造がくずれてしまうこと，極値の出方が各回毎に異なる等があげられよう。３節の流量で示したように，時系列的な予測なら全体的なバランスが求められるのであるが，ここでは１日の予測誤差をみることになるため，結果として誤差が大きくなっているものと考えられる。



for Kawarabashi:

$Y = 4.87 + 0.02\,X_4 - 18.88\,X_7^2 + 0.26\,X_4 \cdot X_7$ 〃〃 $Q < 20$

$Y_3 = 4.85 + 2.60\,X_1 + 0.64\,X_3 - 0.16\,X_1^2 + 0.02\,X_3^2\ 0.26\,X_1 \cdot X_3$

$Y5 = 4.65 + 0.04\,X_3 + 0.03\,X_3^2$

$Y = 0.21 + 0.58\,Y_3 + 0.48\,Y_5 - 0.05\,Y_3^2 - 0.04\,Y_5^2 = 0.09\,Y_3 \cdot Y_5$

for Futagobashi:

$Y1 = 22.1 + 0.55\,X_3 - 3.20\,X_5 + 0.10\,X_5^2 - 0.01\,X_3 \cdot X_5$ 〃〃〃〃 $Q < 20$

$Y_5 = 8.67 + 0.55\,X_3 - 0.22\,X_{13}$

$Y = 2.25 - 0.01\,Y_1 + 1.53\,Y_5 - 0.03\,Y_1^2 - 0.23\,Y_5^2 + 0.20\,Y_1 \cdot Y_5$

$Y = 0.74 + 0.02\,X_4 + 85.7\,X_7 - 101\,X_7^2 + 0.64\,X_4 \cdot X7$

そこで極値をとりのぞくために，流量に制限を加えた。すなわち，下流の調布堰での流量が２０m³以上の日を除いて，構造式をきめて予測を行った。その結果を図－５に示した。結果からみるとデータに制限のない場合よりも誤差が少なくなっているが，必ずしも十分とはいえない。例えば川原橋では，全データで予測を行うと，一部をのぞいて平均的な値が求められているにすぎないが，流量を制限することにより，予測値はかなり実測値に近い値となっている。しかし関戸橋ではほとんど改良された結果があらわれていない。二子橋のデータでみるかぎり，バラツキが多く，特異な点が多いが，実測ＢＯＤが２０mg／l 以下の場合にかぎってみるとかなりよい結果がえられている。

構造式については各地点毎に異なり定性的に関数表示は行なえない。表に各地点の構造式に採用された因子を示す。

6. 考察と今後の課題

いくつかのケースについてＧＭＤＨ法の適用を行い，その汎用性，とくにＢＯＤの予測における限界と精度の検討を行ってきた。考えられるケースは無数にあり，修整法も限りないと思われるが，ここではほんの１部しか行なっていない。これまでの結論をしいて求めるとするならば，データに信頼度がある場合で，かつ同質のデータである場合には，かなりの精度で予測することが可能と考えられる。同質のデータの定義はむつかしいが，同質のデータのみが与えられれば予測する意味もなくなるであろう。ＢＯＤで示したような流量２０m³／sec というのが１つの目安ともなり，それにより予測精度を大巾に上げることが可能となったのは，１つの収穫であると考えられる。

今後の課題として，ＧＭＤＨ法を用いてシミュレーションを行ってみたい。その方法が確立されれば，不足した情報を補完することが可能となり，モータリングの設計に導入することにより，合理的な情報網をうちたてることが可能となろう。

本研究は，東大・京大の各計算機センターで計算を行ない，データは東京都水道局はじめ多くの機関から拝借した。ここに関係各位に厚く御礼申し上げます。又，本研究の一部は，文部省特定研究「環境汚染制御」の研究費によった。

ＧＭＤＨに関する参考文献として次のものを上げる。

① 池田三郎：　ＧＭＤＨによる環境システムの同定と予測　文部省特定研究環境汚染シンポ（1974）

② 池田三郎：　ＧＭＤＨと複雑な系の同定予測　計測と制御　vol 14　№2（1975）

③ 市川　新：池田三郎：　環境汚染制御の方法論の多摩川水系の水質問題への適用　文部省特定研究環境汚染シンポ（1975）

④ 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　1976（投稿中）