

発見的自己組織化法による水質予測

PREDICTION OF WATER QUALITY BY HEURISTIC SELF-ORGANIZATE

市川 新*・池田 三郎**

By Arata ICHIKAWA and Saburo IKEDA

1. はじめに

人間はいうにおよばずあらゆる動物は、ある状況に対しある判断を行って行動している。この判断は、無意識に行われるものもあり、熟慮して行う場合もあるが、いずれにせよ、従来の経験とある価値基準に基づき、いくつかの代替案の中から最善と思われるものを選び行動している。なかにはある状況下で行動の選択の余地がなく、あるきまった行動しかとれない場合もあるが、状況が流動的でかつ代替案が多いときには、過去の経験と知識に基づき代替案の中から1つの案が選択される。その判断が正しいかどうかは、過去の経験の量と質、さらにそれを組立てる能力によってきまる。この判断を行うためには、状況ないし現象をできるだけ精ちに分析を行い正しい認識を行うように努力するのが常であるが、状況や現象を完全な形で認識できるとはかぎらず、また不可能な場合もある。全体からみれば不可能な場合の方がはるかに多いのである。さらにわれわれには時間的制約もあり認識が不完全なままならぬかの決断を行っているし、行わざるをえないのである。

近代科学の進歩とともに状況ないし現象の解明が進み、いままで経験で判断してきたものでも機械的な判断に委ねるようになったものが数多くある。さらに従来複雑な現象と思われていて、数学的予測を行うことが困難であった場合に対してもモデル化が可能となり、われわれの判断そのものを示してくれたり、判断のための材料を提供してくれるようになった。たとえば天気予報の予測はあたるとはかぎらないが、われわれの生活にきわめて有用な情報をもたらしてくれるのである。

このような科学の進歩とともに、われわれに有用な情報ないし、判断基準を示してくれるものが多くなったけ

れども、いまだそれらの手におえない問題がかぎりなく残されている。われわれ人間活動の規模が大きくなるにつれて、このようなわれわれの行動の判断根拠を与えられていない問題に対しても、早急に解決せねばならなくなったのである。たとえば、環境問題・自然の摂理・経済現象・人間の行動などがあげられよう。これらの問題に共通している特徴は、それぞれに関係する因子が多く、かつそれらが互いに関係しあっていること、さらにその相互関係が簡単にモデル化できないことである。

このような問題に対して、従来行われてきた物理的モデルによって現象を記述するというパラダイム (Paradigm) を適用することは不可能であり、新しいパラダイムが必要となってくる。N. Wiener のサイバネティクス (Cybernetics) は、このような問題の解決に有効な新しいパラダイムということができよう。このパラダイムの有効性が、わずかの間に急速に多くの分野での適用の可能性をうんだのである。このパラダイムは簡単にいうならば、入力と出力というわれわれの感知できる情報から、両者の関係を求めるところにその特徴を有し、その間に存在する変換系 (システム) の完全記述は行わずに、ブラックボックスとして、未知のまま有用な情報を与えるものである。それゆえこのパラダイムは、上述したような人間の判断法にきわめて近いものであり、その中で先駆的、かつもっとも進んだものであったといえることができる。

しかしながら、サイバネティクスが新しいパラダイムとして登場し、多くの研究者によってその実用化と精選化が行われてくると、中山 茂¹⁾が述べているように、パラダイムのもつ生の問題に対する活力が失われ、抽象化とそれに伴う理論化、さらには特定の分野への適用に伴う固有のテクニックが、より重視されるようになり、それにふりまわされて人間の判断法に近いというナイーブな姿がうすれてきたのである。これはサイバネティクスの責任というよりは、すべてのパラダイムに共

* 正会員 工修 東京大学助教授 工学部都市工学科

** 工博 京都大学助手 工学部数理工学教室

通する必然であろう。

サイバネティクスは有力な武器として、複雑な系に対し、適用が試みられてきたが、今日までのところその結果は必ずしも満足いくものになっていない。そこでサイバネティクスが当初めざしていた人間の判断のプロセスを再度見直し、その思考法にそったアルゴリズム (Algorithm) を確立し、複雑な系に対し適用しようとするのが、本論文の目的である。ここでは、この人間の判断法を発見的自己組織化法 (Heuristic Self-Organization) と名付け、そのアルゴリズムとして A.G. Ivakhnenko²⁾ が提唱した GMDH 法 (Group Method of Data Handling) を水質問題に適用し、その汎用性とその精度、さらに修正することによってどの程度汎用性が増加するかについて検討を加えるものである。

2. 発見的自己組織化法とそのアルゴリズム

発見的自己組織化法の内容を簡単に説明する例として品種改良をあげる。穀物にせよ、動物にせよ、いままで行ってきた品種改良のプロセスを現時点でさかのぼって追跡することはできるし、それを通して最適解をうるための最短距離を求めることが可能であるが、その出発点においては、今日までの全プロセスを完全に計画していたわけではない。それぞれ途中の段階である価値基準 (寒さに強い、早い、多収である) に基づいて取捨選択を行って今日に至っている。この発展過程をモデル化し、ダイアグラムに示したのが図-1である。

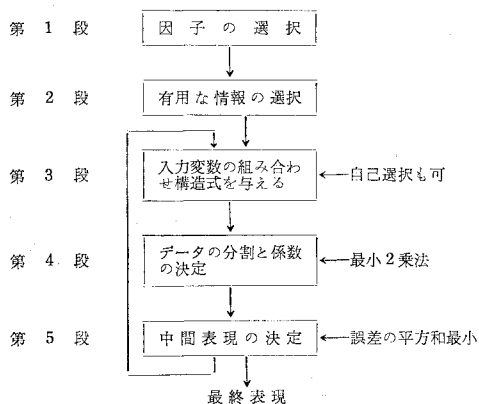


図-1 発見的自己組織化法のフローシート

第1段の因子の選択とは、穀物でいえば収量の多い品種、寒さに強い品種、病虫害に強い品種など目的に近いものをもつ品種を選択することに相当する。この場合因子として数世代前の種子にさかのぼることもある。これを数学的にかくと、次式で示される。

$$Y = f(X_1, X_2, \dots, X_n, \dots) \dots \dots \dots (1)$$

ただし、Y はでき上がった品種、X_i は選択された品種

現象が複雑ならば n は大きくなり、因子が多いことを示す。この Y を完全に表現することが不可能である場合には、上の因子の中からいくつかの因子を選んで、

$$\hat{Y} = f(X_p, X_{p+1}, \dots, X_{p+m+1}) \dots \dots \dots (2)$$

という推定式で表現する。すなわち m 個の因子 (m < n) で Y を推定するということである。X_{p+j} の因子の中には、時間おくれが因子として入ることがあるのは、すでに述べた通りである。

推定をより正しく行うためには、因子の数を多くするのも1つの考え方であるが、問題を複雑にしかつ計算の手間ないし実験の規模を大きくするので、ある種の制限が必要となってくる。この決定に際しては現象の観察を十分に行うことが、効果的な選択の必須の条件となる。この場合、必要に応じて変数の正規化や変換が行われることがある。後述の BOD 予測の例では変換された量を新しい因子としている。

第2段の操作は、第1段でリストアップされた因子の中から有用な入力変数と、無用ないし関係のあまりない因子との分類の操作である。有用か否かの因子の分別の判断基準はヒューリスティックな規範であり、この手法の1つのポイントとなるものである。規範としていろいろ考えられるが、予測項目と相関係数の大きな因子を有用な入力変数とすることがその1つであり、しばしば用いられている。また、物理的ないし化学的モデルが用いられる場合にはモデルに使用される因子をそのまま入力変数とすることがある。流量予測の場合に雨量を入力変数とするのもその一例である。周期性のある現象の場合には、時系列データを因子として用いることもある。これらの判断がそれぞれ規範となるのである。目的とする対象の性質が明らかであればあるほど、因子の数をへらすことが可能となる。

第3段は入力変数を組み合わせて新しい因子ないし中間変数を作製することである。品種改良の例でいえば、種子の交配に相当するものであるが、数学的にはその結合関数をつくることに相当する。一般性をもたせるために、出力の完全記述として式 (3) で示すような Kolmogorov-Gabor の多項式がとられるが、ここで用いた GMDH 法では式 (4) を繰り返し用いることにより、一般的な N 因子の多項式表現を行っている。

$$z_i = b_0 + \sum_i a_i x_i + \sum_i \sum_j a_{ij} x_i x_j + \sum_i \sum_j \sum_k a_{ijk} x_i x_j x_k + \dots \dots \dots (3)$$

$$\hat{z}_1 = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2 + \dots \dots \dots (4)$$

式 (3) でとり上げた因子が N 個の場合、1次の項までで (N+1) 個、2次の項までで 1/2(N+1)(N+2) 個、3次の項までで 1/6(N+1)(N+2)(N+3) 個の未知数

があり、これを解くことはきわめて大変な作業となる。因子が 4 で 4 次の項までとると未知数は 70 となり、 70×70 の行列計算を行わねばならず、因子がふえるにつれ計算は飛躍的に増加する。一方式 (4) で中間表現を用いる場合には、各段での未知数は 6 であり、 (6×6) の行列で係数を決定できる。さらに求められた係数を用いてえられる中間表現を新しい変数として上述の計算を繰り返し行くと、はじめの因子で式 (3) の 4 次の項までとったのと同じ結果がえられる。式 (4) で示される中間変数の総数は、因子数 N の場合 $N C_2$ となるが、ここで規範を設けて選別を行うと、計算はさらに簡略化される。このことは任意の 2 つの種子を交配させることにより、多数の新しい種子を作り、その中から劣性のものを除外し、優性のもの同志を再びかけ合わせていくプロセスのアルゴリズムと考えられる。

第 4 段は、データを経験用と結果の予測の適合性を調べるためのものとの 2 つに分割することである。第 3 段で述べたように式 (3) を解くためには、6 個のデータがあればよいが現象が複雑であるために、6 個のデータから求められた構造式が再現性をもつとはかぎらない。そこで普遍性をもたせるために、 NX 個 ($NX > 6$) のデータを用い最小 2 乗法で係数をきめるアルゴリズムを採用することにする。この係数決定のために用いられるデータを、ここではトレーニングデータとよぶ。トレーニングデータから決定された構造式がどの程度の予測精度をもつのかを調べ、かつ $N C_2$ 個の中間変数のうち、次の因子として残されるべき中間変数を選ぶために、1 部のデータを使用する。これに用いられるデータを、ここではチェックデータとよぶ。この有用な中間変数の選択を行うためには、別の規範が必要となってくる。ここで用いたアルゴリズムは、予測値と実測値との差の平方和を予測誤差とみなしその小さいものを選ぶこととしている。このようなデータの分割が本法の特徴の 1 つであってトレーニングデータへのオーバーフィッティングを避け、予測の精度を上げることが可能となるのである。ただし、データの分割をどのようにするかが問題であり、後述するように分割法により、得られる結果が大きく異なることがある。

第 5 段は、このようにしてできた中間表現の中から予測値として最も近いものを決定することである。決定されたものをここでは最終表現という。

3. 水質予測問題への適用

発見的自己組織化法について説明を行ってきたが、各段で採る規範によりアルゴリズムが異なる。A.G. Ivakhnenko とそのグループは、この発見的自己組織化法によ

る変数の組合せ計算法を GMDH (Group Method of Data Handling) と名付け、その適用の可能性と理論的検証を行っている。その例として、① 小麦の作付面積の予測³⁾、② 貯水池内のバクテリア量の予測⁴⁾、③ 品種改良の仮説の改良⁵⁾、がある。R.L. Barron, A.N. Mucciarioli らによっても同様な考え方によって ATN 法 (Adaptive Transformation Network) が提案され、その実施例が数多く報告されている⁶⁾。

ここでは、この新しいパラダイムを水質と流量の予測同定問題に適用した。これら 2 つの現象は、複雑でかつ簡単なモデルでは記述しにくいことから、取り扱いが困難な問題である。このような問題に発見的自己組織化法の適用が可能なのか、適用限界があるのか、適用条件をかえることによりどの程度汎用性を高めることができるのかを検討してみる。これらの検討は、すべて試行錯誤的であって、結果論として適用可能であるか否かが求められるにすぎず、その評価法によっても結論が異なってくる。そこでは、求められた結果の数値的な精度そのものが問題となることももちろんであるが、パラダイムそのものからくる限界性ならびにデータの中に含まれている情報量の大きさを考慮して、われわれに有用な情報を与えるかどうかで判断せねばならない。注意しておかねばならないことは、本法は非線型関数 f の構造を多項式表現式 (4) で、与えられた少ないデータから定量的に記述し多くの代替案の中から、たしからしいものを 1 つ選ぶ方法であるが故に、求められた表現法は、現象の因果関係の完全記述とは本質的に異なることである。現象の完全記述やパラメーターの推定、同定そのものを本法に期待することは、間違いであると考えられる。

本法は相対的評価ともいべき、ある目的に対し実用的ないし工学的目的に対し有用な情報をうるところにその中心がある。流量予測の場合には、最大流量を推定することに目的がある。水質予測の場合には、目的を何にするかという問題もありより複雑である。ここでは水質調査の観測網・回数などのモニタリングシステムの合理的設計と必要な場所と時間の水質情報をオンラインで得ることを目的としている。

なお、本計算は東大大型計算センター (HITAC 8700/8800) で行ったので、記憶容量・計算時間 (30 日間の予測で約 75 秒位) とともに、制限因子とはならなかった。

4. 適用例とその考察

多摩川流域の水質・流量・降雨量のデータを用いた。多摩川水系の概況とその水質についての概略はすでに報告^{7),8)}しているのでここでは簡単にのべる。流域は 図-2 に示したように、細長く本川流路延長 124 km、流域

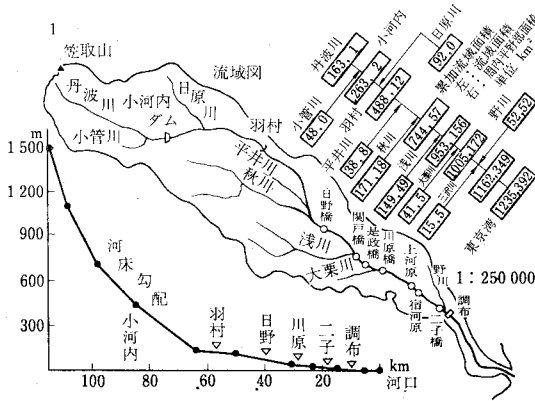


図-2 多摩川流域図

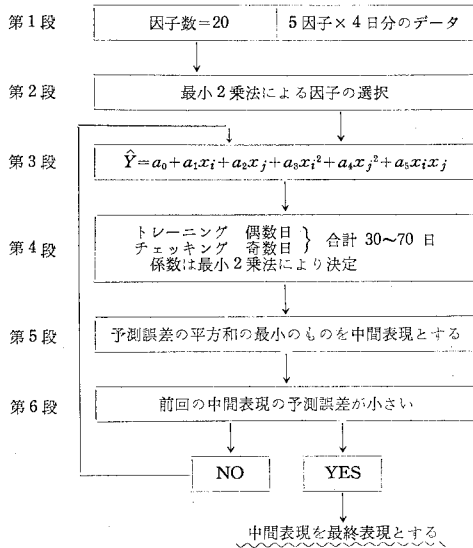


図-3 水質予測に適用した規範

面積1235 km²の細長い、青梅市より上流は山地であり、流域面積にしろる山地面積は大きい。人口の集中地区は、平地部にかぎられ、青梅市より下流と、浅川流域である。一方羽村地点で、洪水期と灌漑期を除いて溢流量がないため、流量からみると羽村より上流は、本川とはいえず、浅川ないし秋川が本川に相当する。資料は雨量(気象庁・建設省)、流量(建設省・東京都水道局)、水質(東京都水道局・同公害研究所・川崎市水道局・建設省)など

表-1 予測項目と入力

予測項目	入力
1. NH ₄ -N	各地点での濁度, COD, 伝導度, 流量, NH ₄ -N
2. COD	
3. 濁度	その地点での流量, 濁度
4. 流量	上流地点での流量 各点での降水量
5. BOD	上下流での流量, COD等

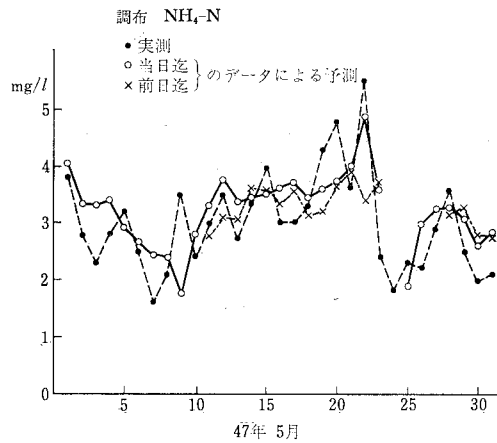
による。図-3にここで用いた規範を示した。これを図-1

に對比することにより本法の構造があきらかとなるものと思う。

表-1に本研究で行った予測項目とそれに用いたデータの一覧表を示した。

(1) NH₄-N の予測

NH₄-Nは溶解性物質であり、沈殿や掃流作用に影響されることが少なく、流量の変動に伴う希釈効果が水質濃度に大きな影響を与える因子と考えられる。日濃度の変化を調べた結果⁹⁾によると、年間ないし月間最大値と平均値を比べるとその差は他の水質指標に比べると小さく比較的安定した指標である。予測結果を図-4に示した。前日から4日前までの表に示した各因子を入力因子として、過去30日分のデータをトレーニングおよびチェック用データとして予測を行った。この方法による予測結果では、どうしても1日のおくれが生じている。降雨に伴う流量および水質の変化に関する最近の研究の成果⁹⁾によると、水質の急激な変化は、短時間で現われ、日単位のデータでその変化を追跡することが困難で、不連続なデータとなっていることが多い。このことは、前日までのデータで、翌日の水質を予測することができないわけで、そのことが予測結果に示されている。そこで当日のNH₄-N以外のデータ、すなわちこの場合は、当日の流量、COD、電気伝導度、濁度を入れて予測してみるとおくれはなくなり、予測の精度は、大幅に改善された。なおトレーニングとチェックのデータを入れか



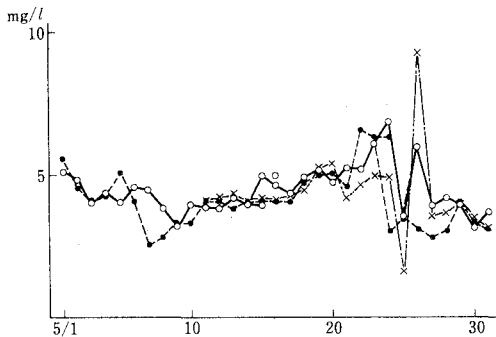
入力因子		
NH ₄ -N	○	×
COD	前日から4日前	前日から4日前までのデータ
濁度	当日から3日前	
電気伝導度 流	までのデータ	

図-4

えて予測しても、結果には大きな差は認められなかった。この場合に最終表現に選択された因子が多かったのは電気伝導度であり、流量はあまり選択されていなかった。電気伝導度も同じ溶解性物質を示す指標の1つであり、かつ電氣的に簡単に連続的に測定できるため、これからNH₄-Nが予測できることは好都合であり、この手法の実際問題への適用の可能性があることを示すものでもある。流量について考えてみると、流量の出水時における変化に比しNH₄-Nの変動が小さかったことによるものと考えられる。

(2) COD の予測

CODは有機物量の指標の1つであり、溶解性のものと、浮遊性のものの両方あり、NH₄-Nとはその特徴が異なる。水質 hidro グラフの研究⁹⁾によると、流量の増大時に急激に増加し、短時間で急激に減衰する。模式的に言えば、増加分は浮遊性ないし、堆積物中のCOD成分であり、減衰部は溶解性のそれと考えられるがその厳密な区分は行われていない。



実測値	●	
入力因子	○	×
COD, 流量, NH ₄ -N, 濁度, COD, 電気伝導度	前日から4日前までの因子 当日から3日前までの因子	前日から4日前までの因子

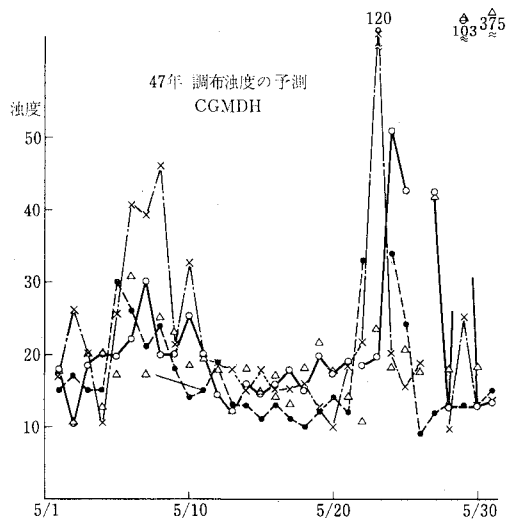
図-5 調布堰 COD の予測 (1972年5月)

図-5に予測結果を示した。結論を言えば、NH₄-Nと同様に当日の資料を入れて予測を行うとトレーニングのデータ数が15であってもきわめてよい予測結果がえられた。しかしながら流量の変動するとき、特に出水後の流量の減衰するとき、予測値と実測値に大きな差が生じた。これは、流量の減衰とCODのそれとが異なることと、最大流量時のCOD値の出かたがさまざまで、トレーニング期間を長くとっても、学習しきれないことが原因と考えられた。なお、最終表現に選択された因子はNH₄-Nとは異なり、対象予測期間ごとに因子が異なっている。このことは、発見的自己組織化法においては、

有用な情報を、その場に応じて採用していることを示すものであり、本法がきわめて柔軟な構造をもっていることを示し、従来の方法の多くが、あらかじめパラメータを決めておくのと著しい対照をなしている。CODとNH₄-Nとの因子の相関性をみてみると、NH₄-Nの予測の最終表現にはCODは30日の予測期間のうち2回だけえらばれているが、CODの最終表現にはNH₄-Nは1度もえらばれていない。

(3) 濁度の予測

濁度については、2通りの予測を行ったが図-6にみられるように良好な予測結果はえられていない。第1の方法は、流量と濁度の2因子で予測するもので、第2の方法は、COD, NH₄-Nと同じ水質因子からの予測法である。濁度は非出水時には10~30であるが、出水時には10~100倍となる変動の大きなかつ流量に従属する指標であるので、流量からの予測を試みた。しかしながら実際には、出水時がトレーニングやチェックングデータに入ると、それにより解が不安定となることにより極値の予測が困難となる。そのうえこのアルゴリズムでは、予測値と実測値の差の平方和を最小となるような規範を使っているのので、極値の発生する日の予測誤差が全体の評価の中での比重を大きくすることとなり、極値の



実測値	●		
予測値	○	×	△
入力因子	濁度	前日から4日前までのデータ	前日から10日前までのデータ
	流量	前までのデータ	当日から9日前までのデータ
因子	COD 電気伝導度 NH ₄ -N	当日から3日前までのデータ	使用せず

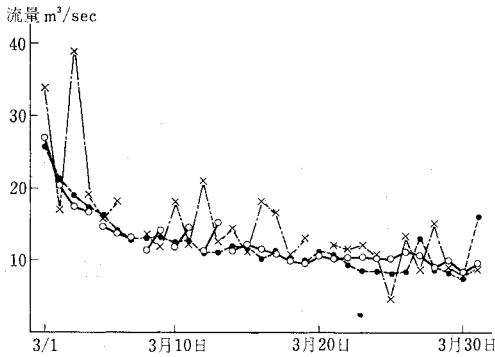
図-6 調布堰濁度の予測 (1972年5月)

予測がよいものが最終表現として選ばれ、非出水時の日の予測値は評価の対象となくなる。このようなシステムに対しては、規範を修正する必要があると考えるが、現在のところ適当なものがえられていない。1つの考え方として、差の絶対値のかわりに実測値と予測値の比を最小にする規範もあろう。ただし、この問題は規範をかえればよいというだけの問題よりも、出水時の濁度のハイドログラフの特性や、他の水質指標との関係を明らかにすることにより、初めて最適な規範ができてくるもの

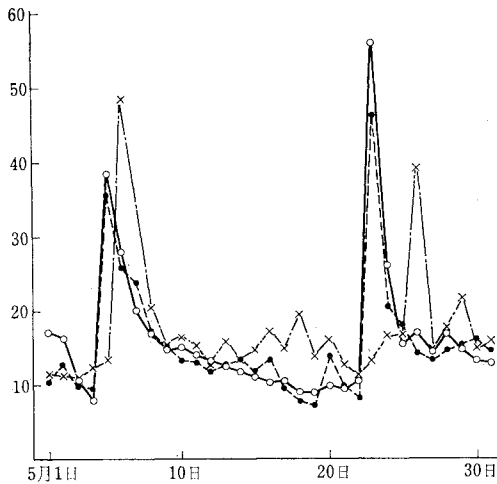
と思い、当面は濁度の特性を把握することが緊急の課題と考えている。

(4) 流量の予測

流量の予測には、上流の本支川の定点4箇所の流量データを因子として最下流の予測を行ってみた。結果は図-7(a), (b)に示す通り非常によい予測結果がえられ、その結果を実用化することが可能と考えられる。流量の場合47年5月に2回ピーク(最大日流量40.0m³/sec)あるときと、47年3月の1回ピークで減衰部の場合と、2度予測を行ってみたがいずれもよい結果である。この場合も上流のデータとして当日の資料をつかうことが不可欠であり、最終表現にも予測地点にもっとも近い流量観測所(川原橋)が因子として選ばれ、その比重が大きかった。これは、流量が流域面積に關係するという常識的な關係を示すものである。なお次いで大きな比重を置いていた因子は、支川秋川の当日の流量であった。降水量から流量を求めるアルゴリズムを作り予測した。ここでは全流域の平均降水量と、上中下流の3ブロックにわけてのおおの平均降水量を求め、この4つを入力因子とした。図-8は上中下流の平均降水量からの予測結果であるが、満足いく予測結果がえられたものと考えている。一方、全流域平均降水量のみで予測してもよい結果はえられなかった。これは多摩川の流域の最大高度が1800mであり、そのため上中下流の気象条件が異なることと、羽村での取水の影響があるため、これらの条件を加味せざるをえないことを示すものであろう。



(a) (47年3月)

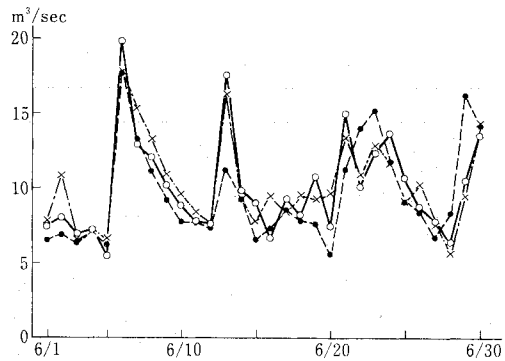


(b) (47年5月)

実測値		●	
予測値		×	○
入力因子	調布堰	前日から4日前までのデータ	前日から4日前までのデータ
	秋川原村内		当日から3日前までのデータ

(a, b 共通)

図-7 調布堰流量予測



実測値		●	
予測値		○	□
入力因子	調布流量	前日から4日前までのデータ	同左。但しトレーニングとチェックデータを入れかえている。
	上流降雨量	当日から3日前までのデータ	
	中流降雨量		
	下流降雨量		
	流域平均降雨量		

図-8 調布堰流量予測(1971年5月)

(5) BOD の予測

BOD の測定には 5 日間かかることから、ルーティンの作業の指標とはならないため日データはほとんどはじめられていない。それゆえ、いままでのべてきたような水質・流量とは異なり十分な資料がない。われわれが入手しえた BOD の実測値は、各機関で行っている河川調査の BOD 試験結果である。幸いなことに、多摩川の BOD データが豊富¹⁹⁾にあり、ほぼ 1 週間に 1 回程度行われているのでこれを利用することとする。

ここではこれらの資料を用いて BOD の予測を行ってみた。すなわち、BOD 試験の行われている地点の BOD、流量、COD 値と、日データのある調布堰の COD、流量の 5 つを入力因子とした。これらのデータは日デー

タではないので、時系列データとして扱うことができず、各調査日ごとの独立したデータとみなさざるをえない。そしてその日の各因子と、その変数を新しい因子として多変量解析を行うアルゴリズムにかえて、発見的自己組織化法を適用した。従来の多変量解析においては、因子はあらかじめ固定されているわけであるが、ここでは 14 個の因子をとり上げ、その中から自己選択するところに本法の特徴がある。ここでとり上げた 14 個の因子は次のとおりである。

- 調布地点の COD, 同 2 乗, 同負荷量
- 調布地点流量, 同 2 乗, 同逆数
- 対象地点 COD, 同 2 乗, 同負荷量
- 対象地点流量, 同 2 乗, 同逆数
- 調布地点と対象地点との COD および流量の比

表-2 BOD の最終表現

A. 川原橋	
第 1 層での	$X_1 = 4.39 + 0.15$ (川原橋 COD) $+ 0.16$ (川原橋流量)
	$X_2 = 5.10 + 0.04$ (川原橋 COD) ²
	$X_3 = 4.85 + 2.60$ (調布 COD) $+ 0.64$ (川原橋 COD) $+ 1.65$ (調布 COD) ²
中間表現	$X_4 = 4.54 + 0.04$ (川原橋 COD) ²
	$X_5 = 4.65 + 0.04$ (川原橋 COD)
第 2 層での	$Y_1 = -0.01 + 1.22 X_3$
	$Y_2 = -0.32 + 1.06 X_5$
最終表現	$Z = 0.21 + 0.58 Y_1 + 0.48 Y_2 - 0.03 Y_1^2 - 0.04 Y_2^2 + 0.09 Y_1 \cdot Y_2$
B. 関戸橋	
最終表現	$Z = 2.24 + 1.30$ (関戸橋 COD) $+ 0.07$ (調布 COD 負荷量) $- 0.07$ (関戸橋 COD) ²

実測値と予測値 (川原橋 BOD)
440812~450804

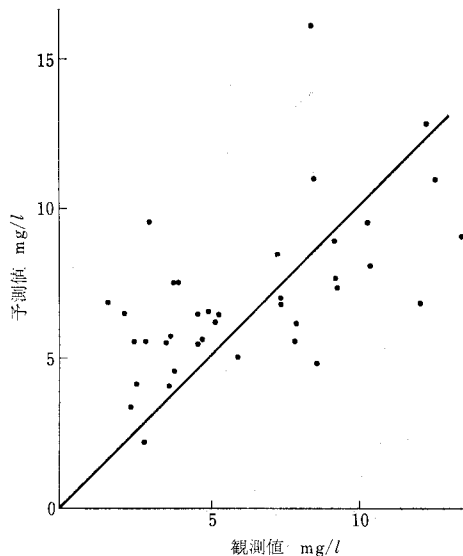


図-9 (a)

実測値と予測値 (是政橋 BOD)
430123~460705

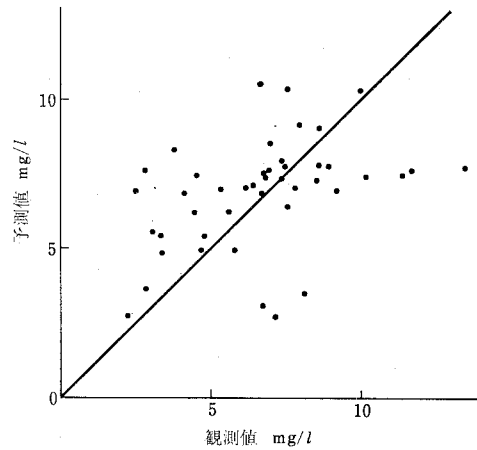


図-9 (b)

実測値と予測値 (関戸橋 BOD)
430117~440902

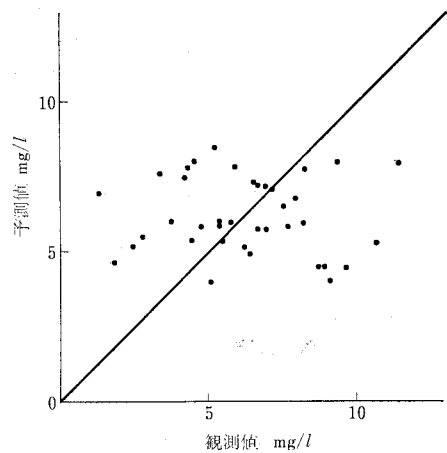


図-9 (c)

これら 14 因子で中間表現をつくり、約 3 年間 60~80 回の河川調査結果をトレーニングおよびチェックングデータとして構造式を決定した。ここで、取り上げた因子は、有機物指標と流量のみであるが、他の水質項目を因子としてとり上げることによって予測結果を改善することが可能であろう。今後に残された課題である。

予測結果を図-9 (a), (b), (c) に示す。結果は必ずしもよくない。これはここで用いたアルゴリズムそのものが欠陥があるのか、あるいは用いたデータの信号成分と雑音成分の比 (S/N) が悪いのか、それとも両方が影響しあっているのか、これだけの例からは判断できない。BOD 値は測定のみずかしい指標の 1 つであり、同じ水でも人により機関により、その結果が異なってくるもので、用いたデータそのものの精度がどの程度なのかを検討しなければならない。実際に BOD 値と他の水質因子とを比較してみても、説明のつきにくいような BOD 値となっている例がいくつかみられる。

このようにして求めた BOD の最終表現の一例を表-2 に示したが、COD が密接な関係をもっているというきわめて常識的な結論となっているが、個々の係数、次数など微妙な差がみとめられた。特にトレーニングの期間を長くすると最終表現はかなり安定し、実用的に本法を適用することが可能であると思われた。

5. アルゴリズムに関する考察

(1) 因子の数

ここで取り上げた因子は、BOD の場合を除き、予測項目をふくめて 5 種類、それぞれ 3 日前までのデータを取り計 20 個とした。水質項目としては、他にデータがないため、これ以上ふやすことはできない。期間として 4 日とっているが、水質指標の自己相関係数をとってみると、4 日前の場合にはきわめて小さく、ほぼ独立な値であることがわかる。それゆえ 4 日前のデータを用いることすら長すぎるものと考えている。そのかわりに、逆数や対数をとった変換形を因子とした方がよい結果がえられるのではなかろうかと考える。さらに BOD の予測の例で行ったように、別の地点のデータを利用することが必要と思われる。流量の予測では、上流のデータにより、よい結果がえられている。1 つの問題点として考えねばならないことは、水質データが独立なのか、時系列的なものかという点である。流量予測の場合、特に減衰部では、5~10 日間ほど指数関数に近似できるため時系列データとみなせるが、出水時は必ずしも連続とはいえない。水質の場合、減衰がきわめて早く、流量ほどの時系列性は無い。この水質データの特徴を生かすアルゴリズム

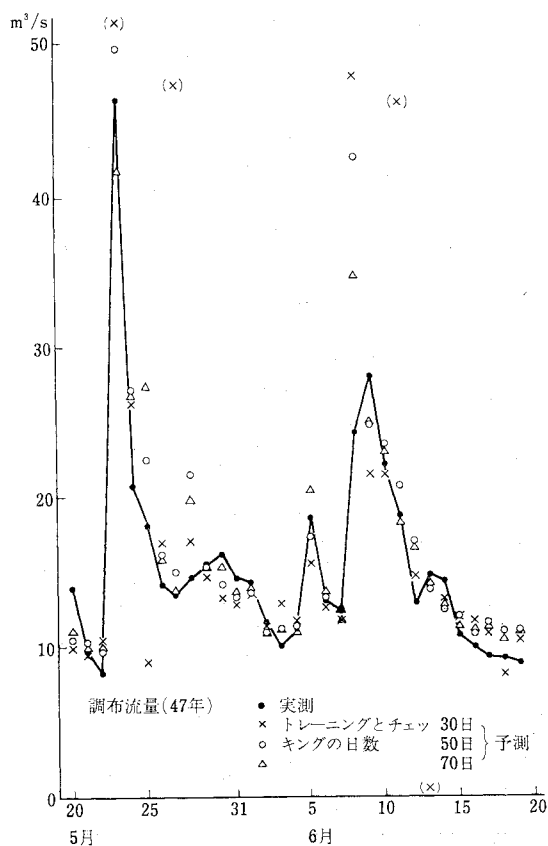


図-10(a) トレーニング日数の差による予測値の差

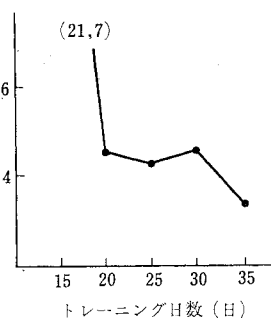


図-10(b) トレーニング日数と R.M.S.E.

ムを早急にたてねばならないが現在のところ結論はえられていない。

(2) トレーニングとチェックングデータ

すでに述べたように、理論的には構造式の決定のためには、未知数の数だけのデータがあれば十分であるが、データのとりかたが異なるごとに、係数が異なるのが例であるので、データ数を多くとり最小 2 乗法で未知数を決定するアルゴリズムをとっている。その結果解は安定

する。一方われわれの求めているのは、できるかぎり少ない情報から有用な情報をうることであり、トレーニングの日数が少ないことが望ましい。図-10は流量予測の際のトレーニングの日数と予測誤差の関係を示したものである。この例では、20日のトレーニング期間（チェックも入れると、2倍の40日となる）が必要なことがわかる。なおトレーニング期間を長くしすぎると予測誤差も大きくなることがある。これは構造が季節的に変化するためであり、年間に通用する構造式を求めるよりも季節ごとの構造を求める方が現実的であると思われる。

表-3に流量予測の場合の最終表現に選択された因子を示したが、日ごととトレーニングの日数がわかることにより、選択された因子が異なるのは、興味ある事実である。このことは、現象が複雑であることと、そのような構造にもかかわらず精度の高い予測が可能であることは、いかにこの発見的自己組織化法が柔軟な構造であることを証明するものである。

表-3(a) 47年6月の流量予測の結果一例(単位: m³/sec)

日付	実測値	予測値			因子番号		
		30	50	70	30	50	70
1日	14.167	13.8	13.4	13.8	5,9	5,9,19	5,7,9
2日	11.493	11.2	10.7	10.8	5,19,20	5,9,19	5,7,9
3日	9.865	13.0	11.3	11.2	4,5	5,6	5,7,9
4日	11.053	11.5	11.3	11.1	5,19	2,5,19	5,7,9,19
5日	18.491	15.6	17.4	20.5	5,10,16	4,5	5,7,9
6日	13.178	12.8	13.3	13.5	5,14	5,19	5,9,19
7日	12.336	11.7	11.4	11.9	5,19	5,16	5,16
8日	24.275	49.8	42.6	35.0	5,10	5,19	5,9,19
9日	28.087	21.6	24.9	25.1	5,19	5,16	5,16
10日	22.118	21.5	23.5	23.1	5,17	2,5,8,20	5,20

表-3(b) 因子番号

	当日	前日	前々日	3日前	4日前
調布		1	2	3	4
川原橋	5	6	7	8	
秋川	9	10	11	12	
羽村	13	14	15	16	
小河内	17	18	19	20	

(3) 規 範

この計算で採用している規範は2箇所、1つは構造式の係数決定の際の最小2乗法と、ほかには、中間表現の決定の際の実測値と予測値の差の平方和を最小にすることである。多くの場合、この規範で十分であるが、流量や濁度のように極値が特に大きい場合には解が不安定となり、ここで用いた規範を修正することが必要となろう。水質予測の目的がなにかによっても異なるが低水流量の時の水質の予測の比重が大きいためそれに応じたアルゴリズムを開発する必要がある。BODの予測の場合に

は、調布流量が20 m³/secを越える日を除くことにより、予測精度が上昇したのは、1つの成功の試みといえよう。しかしながらBOD予測の場合、時系列データでないからこそ、特定のデータの除外が可能であるが、時系列データの場合には、特定のデータを除くことは問題であるし、たとえ可能としてもその欠落させたデータをどのようにとりあつかうかという新たな問題が出てくる。出水時の高水流量時を完全に除くか、別のデータでおきかえるなどの方法が考えられるが、いずれも一長一短があり、今後検討すべき課題である。

(4) 予測式の物理的意味

予測式の最終表現から、その物理的意味を吟味することは可能であるが、このモデルは統計的であって物理的でないで、その意味には一定の限界がある。現実には、最終表現の中に相関度の高い因子が選択される確率が大きい。流量やNH₄-Nの予測のところでみたように、予測項目にハードな関係を有するものが選択され、水質学・水文学などであきらかにされてきた関係があらわれている。しかしながら、この発見的自己組織化法によってえられる関係式は、与えられた入出力データにかくされている定量的な因果関係を抽出するものであり、地点ごと・日ごとに因果関係が異なるというソフトな構造となっている。このことは、まさに水質とか、環境という現象への本法の適用の有効さを証明するものであり、他の方法にない利点である。しかしだからといってこの方法の単に機械的な適用が可能であるが、それはさけるべきであり、その適用には、現象に対する深い洞察が必要と考える。

6. 結論：発見的自己組織化法の水質・水量問題への適用の可能性

本手法について、A.G. Ivakhnenkoが提案したアルゴリズムを若干修正して、水質・流量問題に適用したが、その結論を要約すると以下ようになる。

(1) 変動幅の少ない因子(ここではCOD, NH₄-N)のような水質因子への適用は可能であり、かなりの精度が期待しうる。

(2) 予測の精度を高めるためには、当日の情報を必要とする。そのためには、たとえば電氣的に測定可能で、かつ瞬間的に測定できる指標を入力としてとり入れることが望ましい。その代表例として、本研究の例では電気伝導度をあげた。

(3) このような手法を導入することにより、水質観測網の整理統合が可能となろう。

(4) 流量については、最大40 m³/sec程度までな

ら十分予測しえた。特に上下流地点に定点観測点がある場合には、その予測精度はよくなる。さらに最終表現の構造式はかなり安定しており、1年前の構造式からの予測も可能である。たとえば現在の流量測定は $H-Q$ 曲線を用いている例が多いが、それにかわるものとして、この手法により季節ごとに構造式を求めておいて、他地点の観測結果から外挿することが可能と思われた。

(5) BOD についても同様なことがいえ、その地点ないし上下流地点の水質・流量指標からその日のうちに予測することが可能となった。

(6) 以上の事実から水質観視システムのオンライン化が可能となろう。

(7) 予測された結果をもとにして、新しく次の予測を行うダイナミックな予測法の開発が今後の重要な課題と思われる。

7. おわりに

本研究は文部省特定研究「環境汚染制御」の研究の一環として行われたものである。本研究を行うためには東京都水道局・川崎市水道局・建設省京浜工事事務所など多くの機関の方々のご協力と、東京大学、京都大学の研究室、および各計算センターの協力なしには行いえない

かつし、ケーススタディへの適用は不可能であった。ここにすべての関係された方々に深く御礼申し上げる次第である。

参考文献

- 1) 中山 茂：歴史としての学問 (1974).
- 2) A.G. Ivakhnenko : Heuristic Self-Organization in Problems of Engineering Cybernetics Automatica, 6, 207/219 (1970).
- 3) A.G. Ivakhnenko : The Group Method of Data Handling, Soviet Automatic Control 13-3, 43/55 (1966).
- 4) A.G. Ivakhnenko et al. : Mathematical Simulation of Complex Ecological Systems; Soviet Automatic Control, 4-4, 15/26 (1971).
- 5) A.G. Ivakhnenko et al. : The Group Method of Date Handling in Identification of The Static Characteristic of Multi-Extremal Plant; Soviet Automatic Control, 14-2, 30/37 (1969).
- 6) 池田三郎・榎木義一：GMDH と複雑な系の同定・予測、計測と制御, 14-2, 11/21 (1975).
- 7) 市川 新：水理学研究試論, 第 9 回衛生工学研究討論会 (1973).
- 8) 市川 新：水文学的手法による汚濁流出機構の解析, 第 18 回水理講演会 (1974).
- 9) 市川 新：水質ハイドログラフに関する研究, 第 19 回水理講演会 (1975).
- 10) 市川 新：環境の定量化, 地球科学, Vol. 8 (1975).

(1975.2.25・受付)