

欠測データを含むデータの解析が可能な多階層型 ニューラルネットワークシミュレーション (CQSAR) 法を 用いた河川の上流および中・下流を示す水質パラメータの抽出 - 東京多摩川の水質データ (1994 ~ 2002) を用いて -

神部 順子^a, 袁 えん^b, 長嶋 雲兵^{c,d*}, 青山 智夫^b

^a 大東文化大学外国語学部, 〒 175-8571 板橋区高島平 1-9-1

^b 宮崎大学工学部, 〒 889-2192 宮崎市学園木花台西 1-1

^c 産業技術総合研究所計算科学研究部門, 〒 305-8568 つくば市梅園 1-1-1 中央第二

^d 科学技術振興機構戦略的創造研究推進事業 (JST-CREST), 〒 332-0012 川口市本町 4-1-8

*e-mail: u.nagashima@aist.go.jp

(Received: April 11, 2006; Accepted for publication: June 23, 2006; Published on Web: August 23, 2006)

1994 年から 2002 年までの東京多摩川の水質データ (Tables 3–5) に関し, 上流および中・下流を示す水質パラメータを抽出することを試みた. 2000 年からのデータは欠測を含み, その欠測データはもはや再測定不能である. そこで, 欠測データを含むデータの解析が可能な多階層型ニューラルネットワークシミュレーション (CQSAR) 法を適用した.

水質検査項目の 12 種類のパラメータ (Table 2) のうち, 河川の清浄度を示すパラメータとして溶存酸素 (DO) と全磷 (T-P) が重要であることが判った. さらに上流および中・下流の分類の検定で, 第 6 地点を上流と中流の境界とすべきであることが判った (Table 8, Figure 4, Figure 5). DO は全領域で水質汚染が進んでいることを示している (Figure 4) が, T-P は中流と下流で水質改善がなされていることを示した (Figure 5).

水質汚染という観点からみると, 多摩川では上流から中流域の水質の変化に比べて中流から下流域の水質変化の方が緩やかであることが判った. このことにより多摩川の水質保全には, 中流域での汚染原因を取り除くことが重要であることが示唆された.

キーワード: Water pollution, Tamagawa River, Compensation quantitative structure-activity relationship (CQSAR), Dissolved Oxygen(DO), Total phosphorus(T-P)

1 はじめに

近年, 様々な環境保全問題が注目されている. 特に河川の水質は, 人の生活に密着する重要な問題である. 先に我々は第 1 報 [1] において, 1997 年から 1999 年の 3 年間の 17 地点, 12 種類の水質データを, 主成分分析とニューラルネットワークシミュレータ Neco[2–7] に実装された 3 層パーセプトロン型ニューラルネット

ワークの入力パラメータの偏微分係数解析 [8, 9] を用いて解析した. 17 地点の場所とサンプルの分析機関を Table 1 に示した. また, 17 のサンプル採取地点の海からの距離を大まかに捉えるために, 測定地点の概略を Figure 1 に示した. 12 種類の水質データの内容は Table 2 に示した.

主成分分析により得られた第一主成分と第二主成分を用いてサンプルの採取地点を Kahunen-Loeve プロッ

ト (K-L Plot) すると, その平面上で採取地点を上流, 中流および下流域に分類することができた. 第一主成分では上流と中流・下流域の2つのグループに分けられる. 第二主成分では上流・中流と下流域の2つのグループに分けられる. 第一主成分の分離が第二主成分による分離に比べ, より鮮明である. このことより多摩川では上流から中流域の水質の変化に比べて中流から下流域の水質の変化の方が緩やかであることが言えた. そのため多摩川の水質保全には, 中流域での水質汚染原因を取り除くことが有効であることが判った.

このように水質に関する化学的パラメータデータを用いた多変量統計解析の手法を用いることで, 河川の

環境保全の指針が得られることが示唆された. しかしながら, 第一主成分と第二主成分は, 12種類のパラメータの複雑な線形結合となっており, 主成分分析では, 12種類のパラメータのうち上流・中流・下流の分類に大きな寄与をするパラメータの抽出はできなかった.

偏微分係数解析からは Cl⁻, COND, NH₄-N などの微係数が大きいことが判った. これらの変数は主に上流・中流と下流を分類するものである. また, これらに続く絶対値の大きさを持つ T-P, NO₃-N, PO₄-P は, 中流と下流および上流を分ける変数となっており, 上流を分類するパラメータは, pH と DO であることが判った.

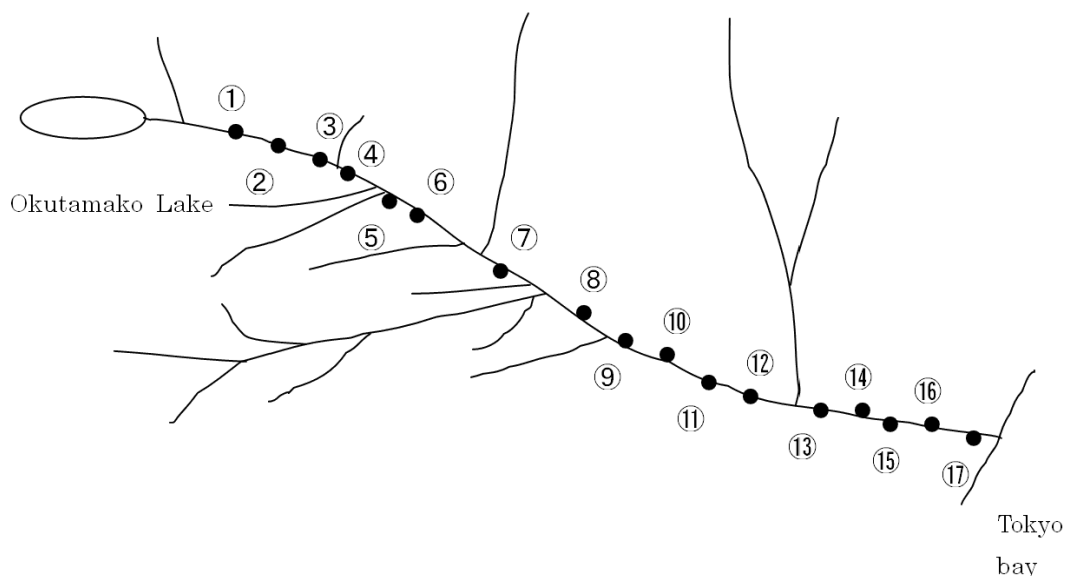


Figure 1. Schematic Map of Sampling Points along the Tamagawa River, Tokyo, Japan. See Table 1 for details of sampling points.

Table 1. Sampling points and agency of analysis

Sampling Points	Agency	Sampling Points	Agency
1 和田橋	東京都	10 多摩川原橋	建設省
2 調布橋	建設省	11 多摩水道橋	建設省
3 羽村堰	東京都	12 砧下取水点	東京都
4 永田橋	建設省	13 第三京浜多摩川橋	建設省
5 拝島原水補給点	東京都	14 調布取水点	東京都
6 拝島橋	建設省	15 田園調布堰上	建設省
7 日野橋	建設省	16 六郷橋	建設省
8 関戸橋	建設省	17 大師橋	建設省
9 是政橋	建設省		

Table 2. Abbreviations and explanation of chemical index of water analysis

測定項目	略称	備考
上流・中流・下流の指標	pos.	上流:1, 中流:2, 下流:3.
水素イオン濃度	pH	海水や石灰岩の影響を受けて pH が高くなる場合がある.
溶存酸素量	DO	量が少ない方が河川は汚染されている.
生物化学的酸素要求量	BOD	好気性微生物の有機養分分解の際に消費する酸素消費量.
化学的酸素要求量	COD	酸素を消費する物質の含有量を示す指標.
全窒素	T-N	生物の生育にとって欠くことのできない栄養塩分.
全リン	T-P	生物の生育にとって欠くことのできない栄養塩分.
塩化物イオン	Cl ⁻	海水の影響を受けない順流域の人為的汚染状況の指標.
アンモニア性窒素	NH ₄ -N	し尿による汚染にはアンモニウムイオンが検出される.
亜硝酸性窒素	NO ₂ -N	硝酸イオンの還元あるいはアンモニアの酸化によって生ずる.
硝酸性窒素	NO ₃ -N	一般の淡水には 0.1 ~ 1mg / l 以上含まれる.
リン酸性リン	PO ₄ -P	生活排水の影響を受けた河川のリン濃度は大きい.
電気伝導率	COND	清澄な河川では低く, 汚濁の進んだ場合逆に高い値を示す.

本研究では, 東京都環境局の「公共用水域および地下水の水質測定結果」[10] に掲載されている 1994 年から 2002 年までの多摩川の水質データをすべて用いた解析をおこなった. 残念なことに 2000 年から 2002 年までのデータには欠測があり, 従来の方法では十分な解析が困難であった.

そこで青山, 長嶋によって新たに開発された, 多階層型ニューラルネットワーク(複数)で欠測データを含む学習を行い, その因果関係を解くアルゴリズム(以下, CQSAR)[11, 12] を適用し, 12 種類の水質パラメータから, 上流および中・下流の分類に重要な役割を果たすパラメータの抽出を行い, そのパラメータの解析を試みた. 具体的には 12 種類の水質パラメータを説明変数とし, 上流・中流・下流(pos.)を目的変数として, 3 層の階層型ニューラルネットワークのネットワークの重みを解析することで, 上流および中・下流の分類に重要なパラメータの抽出し, そのパラメータの経年変化や位置的变化を解析する.

2 データ

1994 年から 1999 年までの多摩川のデータは 17 地点 (Figure 1), 12 種類のパラメータが完全に揃って

いるが, 2000 年から 2002 年までのデータには欠測がある.

Table 3 に 1994 年から 1996 年の 3 年分のデータを示した. Table 4 に 1997 年から 1999 年の 3 年分のデータを, Table 5 に 2000 年から 2002 年のデータをそれぞれ示した. Table 5 の陰の部分は, 欠測であることを示している.

3 欠測データを含むデータの解析法 (CQSAR)

階層型ニューラルネットワークを用いた多変量解析は, 因果関係が不明で説明変数と観測値が非線形関係にあるような現象の分析に用いられる. 化学研究における主な用途は薬理活性解析, 環境問題解析などであるが, それらには, しばしば欠測問題が存在する. 多変量解析では欠測は単に説明変数の減少に止まらず, 複数の説明変数の中の一部が欠測であっても, それ以外の説明変数の該当部分を使用できなくなる悪影響を派生する. そのため多変量解析が実施不可能になる場合がある.

Table 3. Values of chemical index of water analysis from 1994 to 1996

1994.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	CI-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	8.1	12.5	0.5	0.9	0.53	0.014	2	0.01	0.002	0.48	0.003	90
(2)調布橋	1	7.7	13	0.5	1.1	0.62	0.017	2	0.01	0.02	0.59	0.01	99
(3)羽村橋	1	8.2	13.8	0.5	0.9	0.55	0.007	2	0.01	0.002	0.48	0.006	91
(4)永田橋	1	7.7	13.1	0.7	1.6	0.65	0.019	3	0.01	0.02	0.57	0.011	118
(5)拝島原水補給点	1	8.5	13	0.5	1.2	1.2	0.014	5	0.01	0.008	1.03	0.006	127
(6)拝島橋	2	7.7	13.1	0.9	1.8	1.12	0.025	5	0.01	0.02	1	0.021	149
(7)日野橋	2	7.6	11.3	3.7	4.7	5.81	0.645	31	1.5	0.34	3.53	0.624	314
(8)関戸橋	2	7.5	11.3	4.3	5.3	6.87	0.682	33	1.5	0.33	4.51	0.608	353
(9)是政橋	2	7.5	11.4	4.9	5.6	8.07	0.752	36	1.1	0.31	6.11	0.704	371
(10)多摩川原橋	2	7.5	9.4	6.2	7.2	9.66	0.988	51	3.4	0.74	4.7	0.862	475
(11)多摩水道橋	2	7.8	11.6	4.8	5.9	7.99	0.705	38	1.4	0.35	5.69	0.652	380
(12)砧下取水点	2	7.5	10.3	8.7	6	8.57	0.845	40	1.45	0.304	5.9	0.72	327
(13)第三京浜多摩川橋	2	7.3	9.1	5.1	6.4	8.91	0.635	42	1.8	0.31	6.18	0.556	404
(14)調布取水点	2	7.3	8.5	4.8	6.5	8.71	0.725	39	1.63	0.271	5.96	0.642	324
(15)田園調布堰上	2	7.3	8.5	4.6	6.2	8.22	0.637	40	1.9	0.31	5.44	0.556	395
(16)六郷橋	3	7.9	11	3.8	6.8	7.87	0.727	1040	2.6	0.3	4.29	0.565	3900
(17)大師橋	3	7.8	11.8	2.8	5.8	7.26	0.654	1000	2.3	0.29	3.97	0.544	3610
1995.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	CI-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	7.7	12.5	0.5	1.2	0.49	0.006	2	0.01	0.004	0.47	0.003	87
(2)調布橋	1	7.8	12.5	0.5	1.3	0.62	0.018	2	0.01	0.003	0.53	0.005	86
(3)羽村橋	1	7	13.1	0.5	1.1	0.5	0.006	2	0.01	0.003	0.43	0.003	92
(4)永田橋	1	8.1	12.6	0.8	2.1	0.64	0.018	2	0.01	0.003	0.52	0.007	107
(5)拝島原水補給点	1	8.1	15	0.5	1.7	1.03	0.01	6	0.01	0.007	0.95	0.003	139
(6)拝島橋	2	8.1	12.8	0.8	2.3	1.06	0.034	8	0.01	0.01	0.89	0.009	163
(7)日野橋	2	7.7	9.8	5.1	7.9	9.58	1.03	46	1.64	0.588	6.74	0.972	404
(8)関戸橋	2	7.6	10.5	7.2	8.2	8.83	0.668	42	1.74	0.251	6.37	0.662	416
(9)是政橋	2	7.7	10.5	7.2	8.2	8.64	0.666	42	1.5	0.397	6.11	0.657	418
(10)多摩川原橋	2	7.7	8.9	8.9	9.2	12.4	1	46	5.18	0.775	5.73	0.958	531
(11)多摩水道橋	2	7.8	10.5	7.6	8.2	11.1	0.906	48	3.7	0.5	6.42	0.864	468
(12)砧下取水点	2	7.2	11	4.6	7.3	9.53	0.809	48	1.97	0.444	6.54	0.729	439
(13)第三京浜多摩川橋	2	7.7	9.4	7.4	7.7	10.1	0.785	48	2.91	0.275	6.34	0.735	476
(14)調布取水点	2	6.8	9.5	3.9	7.3	9.49	0.757	48	2.22	0.308	6.09	0.672	442
(15)田園調布堰上	2	7.8	9.9	5	8	9.21	0.72	49	2.58	0.3	5.87	0.695	469
(16)六郷橋	3	7.5	9.2	2.5	8.7	15.1	0.598	6160	6.04	0.314	8.12	0.568	18600
(17)大師橋	3	7.6	7.9	2.8	7.4	14	0.432	11300	4.9	0.251	8.31	0.392	30600
1996.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	CI-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	7.3	11.9	0.5	1.1	0.73	0.013	2	0.01	0.002	0.62	0.003	113
(2)調布橋	1	7.8	12.6	0.7	2.2	0.82	0.022	2	0.04	0.004	0.62	0.011	123
(3)羽村橋	1	7.3	13.1	0.5	0.8	0.65	0.009	3	0.01	0.002	0.64	0.003	124
(4)永田橋	1	7.8	12.5	0.9	2.3	0.93	0.023	3	0.02	0.005	0.74	0.018	153
(5)拝島原水補給点	1	7.9	12.8	0.5	1.7	1.27	0.014	2	0.01	0.006	1.18	0.003	157
(6)拝島橋	2	7.9	12.5	1.2	2.4	1.77	0.035	6	0.07	0.015	1.42	0.018	178
(7)日野橋	2	8	10.2	5.6	7.1	3.89	0.874	40	1.48	0.321	1.31	0.825	378
(8)関戸橋	2	7.9	10	7.8	7	6.66	0.854	46	1.38	0.269	4.28	0.825	403
(9)是政橋	2	7.9	10.3	6.4	7.1	8.02	0.783	45	0.99	0.246	6.08	0.758	417
(10)多摩川原橋	2	7.3	8.3	10	8.7	11.4	0.897	44	3.18	0.602	5.54	0.747	389
(11)多摩水道橋	2	7.4	9	7.5	7.9	9.76	0.834	42	3.21	0.488	5.81	0.695	366
(12)砧下取水点	2	7.4	10.2	5	5.5	9.13	0.565	42	1.52	0.322	6.3	0.483	392
(13)第三京浜多摩川橋	2	7.4	9.4	8.3	7.6	9.89	0.668	42	2.05	0.39	6.74	0.556	361
(14)調布取水点	2	7.4	9.2	4.7	5.5	8.46	0.488	40	1.36	0.233	5.77	0.427	376
(15)田園調布堰上	2	7.4	9.2	4.7	7.2	9.29	0.652	40	1.72	0.366	6.65	0.573	352
(16)六郷橋	3	7.6	11.9	8.1	9.8	9.89	0.735	5410	2.49	0.44	3.85	0.339	14200
(17)大師橋	3	8	10.2	9.6	6.5	6.8	0.415	11600	0.58	0.32	2.76	0.204	29000

多変量解析において欠測を取り扱う方法はいくつか知られている。

方法1: 欠測のあるデータを除く。

これは、一番簡単で広く使われている方法であるが、欠測以外の情報を捨ててしまうことになる。情報自体が少ない場合、情報の損失の影響は大きい。

方法2: 欠測データに他のデータの平均値を当てはめる。

方法1を用いた場合の最大の欠点である「情報の損失」は回避されるが、欠測データとして本来あるべきデータと当てはめられる平均値の差が大きい場合、解析結果に悪影響を与える場合が多い。

方法3: 欠測データを何らかの方法で予測し、それを

欠測データの代わりに用いて解析に用いる。

この方法の解析結果は欠測データの予測精度に依存する。通常は方法2よりよい結果を与える。この方法は欠測データ予測精度が解析結果に大きく反映する。

方法4: 欠測を持つデータに対する統計モデルを仮定して、欠測データ周辺の類似性から欠測データを予測する。

この方法としては、EM(Expectation Maximization)法 [13, 14] が良く知られている。EM法は不完全データの問題を完全データの枠組みで逐次的に解く方法であり、その操作は以下の通りである。

Table 4. Values of chemical index of water analysis from 1997 to 1999

1997.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	Cl-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	7.6	12.3	0.5	1.2	0.77	0.008	2	0.01	0.002	0.72	0.003	108
(2)調布橋	1	7.7	12.4	0.5	1	1.02	0.011	2	0.01	0.002	0.96	0.005	106
(3)羽村橋	1	7.8	13.5	0.5	1.4	0.93	0.014	3	0.01	0.002	0.83	0.010	118
(4)永田橋	1	8.0	12.1	0.5	1.6	1.35	0.011	4	0.02	0.005	1.20	0.007	134
(5)拝島原水補給点	1	8.6	15.5	0.5	1.2	2.24	0.008	6	0.01	0.007	2.03	0.003	179
(6)拝島橋	1	8.2	12.8	1.1	1.6	2.67	0.011	7	0.02	0.008	2.53	0.009	155
(7)日野橋	2	7.8	10.5	2.0	4.7	7.19	0.557	28	0.52	0.178	6.01	0.506	284
(8)関戸橋	2	7.7	10.4	2.8	5.1	6.64	0.496	28	0.80	0.227	5.21	0.462	296
(9)是政橋	2	7.8	11.0	1.8	4.9	6.13	0.596	30	0.59	0.205	4.95	0.572	316
(10)多摩川原橋	2	7.1	8.8	5.0	8.1	9.24	0.763	41	0.67	0.589	7.47	0.635	341
(11)多摩水道橋	2	7.4	10.6	4.3	7.2	8.34	0.663	37	0.72	0.407	6.47	0.546	318
(12)砧下取水点	2	7.7	11.9	3.3	5.6	8.58	0.633	35	0.02	0.350	6.56	0.571	376
(13)第三京浜多摩川橋	2	7.4	10.7	4.2	6.5	7.68	0.603	36	0.53	0.367	6.53	0.500	317
(14)調布取水点	2	7.7	11.2	2.7	5.6	7.48	0.540	36	0.05	0.261	6.62	0.479	378
(15)田園調布堰上	2	7.5	11.1	3.6	6.4	7.64	0.610	38	0.51	0.343	6.23	0.481	315
(16)六郷橋	3	7.4	9.6	2.0	6.3	8.13	0.514	3400	2.16	0.353	4.79	0.425	8670
(17)大師橋	3	7.5	9.3	2.1	5.2	6.42	0.372	8220	1.12	0.281	3.38	0.302	18100
1998.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	Cl-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	7.3	11.9	0.5	0.9	0.68	0.008	1	0.01	0.002	0.62	0.005	97
(2)調布橋	1	8.0	12.1	0.5	1.3	0.61	0.019	2	0.05	0.05	0.57	0.005	101
(3)羽村橋	1	7.2	12.6	0.5	0.9	0.65	0.012	2	0.01	0.002	0.59	0.006	104
(4)永田橋	1	8.1	12.3	0.6	1.7	0.75	0.026	2	0.05	0.05	0.59	0.006	125
(5)拝島原水補給点	1	7.3	12.7	0.5	0.9	1.37	0.012	5	0.01	0.004	1.05	0.007	135
(6)拝島橋	1	8.0	12.2	0.7	1.7	1.10	0.029	3	0.05	0.05	0.88	0.011	148
(7)日野橋	2	7.5	9.7	2.8	6.5	7.64	0.894	43	0.96	0.26	6.02	0.801	391
(8)関戸橋	2	7.3	8.9	2.7	5.5	8.34	0.598	37	0.79	0.19	6.79	0.508	379
(9)是政橋	2	7.6	9.7	2.0	6.3	8.91	0.756	41	0.59	0.19	7.68	0.669	427
(10)多摩川原橋	2	7.3	9.8	3.1	7.3	9.80	0.891	50	0.43	0.28	8.74	0.780	379
(11)多摩水道橋	2	7.4	10.0	3.8	6.5	9.55	0.856	44	0.45	0.34	8.03	0.764	348
(12)砧下取水点	2	7.5	10.6	1.7	5.1	9.08	0.636	39	0.15	0.161	8.24	0.581	384
(13)第三京浜多摩川橋	2	7.5	10.5	2.8	6.2	9.13	0.687	43	0.59	0.26	7.92	0.618	355
(14)調布取水点	2	7.2	9.3	6.8	6.2	8.18	0.665	39	0.65	0.167	6.17	0.353	392
(15)田園調布堰上	2	7.6	10.9	2.6	6.1	8.96	0.710	43	0.48	0.24	7.85	0.648	353
(16)六郷橋	3	7.6	9.6	11	8.9	8.54	0.741	11200	0.48	0.16	3.29	0.143	27400
(17)大師橋	3	7.9	10.0	4.7	5.3	5.86	0.314	13000	0.41	0.15	3.11	0.258	31400
1999.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	Cl-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	7.4	12.3	0.5	1.0	0.62	0.008	1	0.01	0.002	0.56	0.003	95
(2)調布橋	1	7.8	12.8	0.5	1.3	0.62	0.011	2	0.05	0.005	0.55	0.008	106
(3)羽村橋	1	7.3	12.9	0.5	1.1	0.61	0.024	2	0.01	0.002	0.49	0.003	104
(4)永田橋	1	7.9	12.9	0.5	1.5	0.65	0.012	2	0.05	0.005	0.56	0.008	116
(5)拝島原水補給点	1	7.2	13.9	0.6	1.1	1.06	0.010	3	0.01	0.003	0.94	0.005	131
(6)拝島橋	1	7.9	12.7	0.5	1.4	1.11	0.022	3	0.05	0.005	1.02	0.010	148
(7)日野橋	2	7.5	10.1	3.1	6.3	6.95	0.816	36	0.83	0.276	5.44	0.72	382
(8)関戸橋	2	7.5	10.8	4.2	6.9	7.79	0.713	35	0.99	0.348	6.09	0.634	397
(9)是政橋	2	7.5	10.7	2.4	7.0	8.06	0.770	39	0.57	0.260	6.68	0.707	413
(10)多摩川原橋	2	7.3	10.1	3.5	7.0	9.15	0.829	41	0.42	0.327	7.85	0.708	464
(11)多摩水道橋	2	7.5	10.7	2.4	6.3	9.43	0.837	41	0.43	0.294	7.62	0.731	448
(12)砧下取水点	2	7.3	10.9	1.4	5.1	8.93	0.677	39	0.17	0.180	8.20	0.659	380
(13)第三京浜多摩川橋	2	7.5	11.2	2.0	5.5	8.91	0.685	40	0.41	0.207	7.70	0.568	445
(14)調布取水点	2	7.4	9.7	1.5	5.0	9.23	0.663	36	0.28	0.183	7.78	0.634	386
(15)田園調布堰上	2	7.7	11.5	1.7	5.5	8.79	0.733	40	0.26	0.204	7.31	0.641	454
(16)六郷橋	3	7.4	9.8	2.6	6.3	8.71	0.484	4210	3.04	0.242	4.48	0.388	14600
(17)大師橋	3	7.6	9.0	2.3	4.8	4.74	0.270	10203	0.72	0.184	2.41	0.218	34100

(1) 手元にある観測データ Y_{obs} だけでは解を導くことが困難な問題に対して、解決が容易なレベルまでデータを完全化し、(欠損データ部分 Y_{mis} が存在するとして)問題を定式化する。(2) パラメータの推定値 I を導くことが目的であれば Y_{mis} に暫定的な値を埋め込み I_0 を求める。(3) I_0 を用いて Y_{mis} を改良する。(4) I_n の値が収束するまで前の2つのステップを繰り返す。

ここで我々の注目する化学、薬理学、環境科学データについて Y_{mis} の存在は確かである。そのため、操作(1)、(2)は可能である。しかし(3)、(4)は問題を多く含む。たとえば、そのような結果を繰り返して観測値を補正することは物理、化学の分野では禁止されている。このような禁制は薬理活性問題の化合物の置換基物性値と生理学的活性の測定の間接性を考えれば当然である。逆に統計学では「確率論的モデルを仮定し、そ

の現れた現象を観測データ」と考えている。そのような場合には、(3)、(4)は適用可能である。言い換えると化学、薬理学、環境科学データについては、ナイーブなEM法の適用はもともと禁じられている。

つまり薬理活性問題や環境問題では、現象に関係がある複数種類の測定値を用い(仮定)、それらを説明変数として現象の再現を試みる。この際、関係の無い説明変数の適用、過剰のパラメータの採用、不適切な関数などの存在など手段が適切でないならば測定値は高精度で再現されても、現象について得られる情報は少ないか全く無いということは広く知られている。そのため、このようなシミュレーションの有意性判定は現象の総合的な知識が必要である。そのため、EM法をそのまま薬理活性問題、環境問題に適用することは難しい。

Table 5. Values of chemical index of water analysis from 2000 to 2002

2000.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	Cl-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	8.2	12.3	0.5	1.1	1.21	0.01		0.01	0.002	0.64	0.003	133
(2)調布橋	1	7.5	12.2	0.5	0.8	0.71	0.013	1.24	0.05	0.005	0.61	0.008	101
(3)羽村橋	1	8.2	12.6	0.5	1.2	0.99	0.01		0.01	0.002	0.75	0.003	141
(4)永田橋	1	7.8	12.1	0.5	1	0.89	0.017	2.34	0.05	0.005	0.78	0.009	114
(5)拝島原水補給点	1	9.2	14.1	0.5	0.9	1.49	0.013		0.01	0.006	1.1	0.006	204
(6)拝島橋	1	8	11.8	0.8	1.1	1.8	0.015	5.24	0.05	0.006	1.65	0.009	158
(7)日野橋	2	7.5	10	2.8	5.2	5.97	0.657	25.9	0.48	0.309	4.7	0.451	315
(8)関戸橋	2	7.4	9.7	2.5	2.5	7.1	0.591	29.1	0.58	0.272	5.72	0.416	329
(9)是政橋	2	7.4	10.3	2.6	5	7.55	0.6	29.1	0.43	0.22	6.49	0.431	350
(10)多摩川原橋	2	7.3	9.1	3.7	6.5	7.62	0.682	35.8	0.37	0.222	6.8	0.56	350
(11)多摩水道橋	2	7.4	9.7	3.1	6.1	7.39	0.637	36	0.25	0.196	6.72	0.551	347
(12)砧下取水点	2	7.6	11.8	2	5.4	7.45	0.654		0.08	0.187	5.97	0.612	471
(13)第三京浜多摩川橋	2	7.5	10.1	2.7	5.6	7.51	0.526	35.3	0.38	0.162	6.58	0.467	344
(14)調布取水点	2	7.5	10.3	2	5.8	12.6	0.564		0.21	0.19	5.6	0.518	585
(15)田園調布堰上	2	7.6	10.6	2.1	5.6	7.18	0.545	34.7	0.24	0.154	6.61	0.467	342
(16)六郷橋	3	7.4	9.3	1.7	6.3	7.83	0.535	2330	2.25	0.186	5.1	0.452	7670
(17)大師橋	3	7.6	8.6	1.9	4.9	6.23	0.411	8410	1.6	0.154	3.88	0.349	24600
2001.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	Cl-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	8	11.9	0.5	1.2	1.08	0.019		0.01	0.005	0.67	0.008	150
(2)調布橋	1	7.5	12.2	0.5	1.2	0.74	0.019		0.05	0.005	0.64	0.015	
(3)羽村橋	1	7.9	12	0.5	1.2	1.41	0.017		0.01	0.003	0.77	0.009	151
(4)永田橋	1	7.7	12.2	0.5	1	0.84	0.013		0.05	0.005	0.76	0.01	
(5)拝島原水補給点	1	7.8	12	0.5	1.1	1.87	0.023		0.03	0.004	1.66	0.016	213
(6)拝島橋	1	7.9	12.2	0.5	1.2	1.74	0.025	3.45	0.05	0.005	1.64	0.022	160
(7)日野橋	2	7.6	10.2	2.5	4.8	5.75	0.451		0.89	0.187	4.28	0.343	
(8)関戸橋	2	7.5	10.2	1.6	4.2	7.07	0.54		0.6	0.209	5.9	0.447	
(9)是政橋	2	7.5	10.3	1.7	5	7.03	0.577		0.41	0.177	6.06	0.467	
(10)多摩川原橋	2	7.4	9.8	3.6	6.5	7.07	0.678	34.1	0.53	0.658	5.43	0.569	344
(11)多摩水道橋	2	7.4	9.9	2.2	5.5	6.95	0.547		0.36	0.394	5.91	0.489	
(12)砧下取水点	2	7.7	9.3	1.5	5	6.34	0.524		0.17	0.286	3.98	0.465	438
(13)第三京浜多摩川橋	2	7.5	10.4	1.9	5.1	6.99	0.518		0.22	0.308	6.01	0.457	
(14)調布取水点	2												
(15)田園調布堰上	2	7.5	10.6	1.8	4.9	7.07	0.477	32.4	0.28	0.273	6.01	0.43	332
(16)六郷橋	3	7.4	9.5	1.7	5.7	7.85	0.488	2730	2.71	0.304	4.69	0.421	7970
(17)大師橋	3	7.6	8.6	1.9	4.8	5.38	0.323	8930	1.22	0.194	3.16	0.267	24900
2002.Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	Cl-	NH4-N	NO2-N	NO3-N	PO4-P	COND
(1)和田橋	1	7.4	12.2	0.5	1.5	0.65	0.026		0.01	0.002	0.6	0.014	106
(2)調布橋	1	7.9	12.2	0.5	1.7	0.66	0.033		0.05	0.005	0.57	0.022	
(3)羽村橋	1	7.4	12.4	0.5	1.2	0.71	0.024		0.01	0.002	0.65	0.02	118
(4)永田橋	1	7.8	12.2	0.5	1.4	0.81	0.027		0.05	0.005	0.68	0.017	
(5)拝島原水補給点	1	7.9	12.4	0.5	1.2	1.65	0.014		0.01	0.002	1.55	0.011	156
(6)拝島橋	1	7.8	12.2	0.6	1.7	1.58	0.026	3.9	0.05	0.005	1.45	0.012	16
(7)日野橋	2	7.4	10.4	2.3	4.4	5.42	0.475		0.57	0.195	4.3	0.42	
(8)関戸橋	2	7.4	10.7	1.8	3.8	6.04	0.48		0.45	0.164	5.15	0.43	
(9)是政橋	2	7.5	10.6	1.9	4.1	6.47	0.581		0.28	0.129	5.75	0.528	
(10)多摩川原橋	2	7.3	9.5	4.2	6.3	7.93	0.671	32.6	0.93	0.704	6	0.513	33
(11)多摩水道橋	2	7.3	10.1	2.4	5.1	7.2	0.546		0.89	0.372	5.65	0.485	
(12)砧下取水点	2												
(13)第三京浜多摩川橋	2	7.4	10.2	1.7	4.4	7.2	0.432		0.6	0.256	5.96	0.386	
(14)調布取水点	2												
(15)田園調布堰上	2	7.5	10.3	1.4	4.4	6.83	0.461	29.1	0.6	0.231	5.72	0.4	31
(16)六郷橋	3	7.3	9.1	1.4	4.7	5.73	0.329	5672	1.05	0.191	3.61	0.277	1950
(17)大師橋	3	7.5	8.9	1	4.7	5.34	0.282	8248	0.69	0.174	3.21	0.26	2420

Shaded squares indicate defective data.

CQSAR[11, 12]は、方法3つまり欠測データを何らかの方法で予測し、それを欠測データの代わりに用いて解析に用いる方法に基づくアルゴリズムであり、実測の50%以上にも及ぶ欠測の悪影響を取り除き、多変量解析を可能にする階層型ニューラルネットワークを用いた新しい手法である。以下では方法1とその精度を比較する。

3.1 CQSARのアルゴリズム

CQSARのアルゴリズムは、大まかに2つのステップから構成される。CQSARの第1ステップは、欠測

部を各説明変数および観測値を独立に補完することである。この時、補完法の選択は主に観測数に依存し、観測値の分布と連続的変化の有無に依存する。我々の経験から、観測数が5個未満ならば線形最小2乗法を用いる。観測数が5個以上7個未満で現象の非線形性が明らかな場合、例えば現象が単調で次第に飽和するような場合ならば、単一sigmoid関数、また、極値の存在が明らかで、その部分の確定が重要な薬理学分野の問題の場合は、可変絶対値関数 fitting で補完する。ただし環境科学問題のように過去の測定データが利用できる場合は、その測定データを関数の離散表現として新たにパラメータを付加した関数 fitting で補完する。

観測値の数が7以上で説明変数の変化が連続的である場合、ニューラルネットワークと等差数列を使う補完法が良い。この補完法は現象の支配方程式を要求しないので容易に使用でき、さらに精度も優れている。

第2ステップは第1ステップで完全化したデータのニューラルネットワークを用いた定量的構造活性相関QSARである。この際、補完データには誤差が相当量含まれている可能性が高いので、その悪影響を除くため、通常の学習に用いられるバックプロパゲーション学習法と再構築学習法 [15] などの simulated annealing 法の併用が望ましい。

CQSAR の適用制限条件は「観測数がある程度（経験的には7以上）無いと精度が保証できない」ことである。またその偏微分係数の精度を要求するのならば、経験的に11以上の観測数が必要である。したがって広範囲の問題に適用可能である。

3.2 CQSAR の精度

簡単な例を用いて、CQSAR の精度を示す。CQSAR のみならず計算法の適用限界を数値的に調べる場合、誤差のない理想的なデータが必要である。環境および化学分野で一般的な測定に類似の初等関数の等間隔サンプリングにより、その種のベクトルを生成した。説明変数には、 $x^{1/2}$, x , x^2 , $(x - a)^2$, $(x - a)(x - b)(x - c)$ の

五種類の関数値を与え、観測値に $b - (x - a)^2$ の関数を選んだ。ここで $0 < x, a, b, c < 1$ である。ニューラルネットワークでは増加と減少は学習上同等であるので、増加のみを選んだ。サンプリング数は環境および化学分野で頻繁に見出される測定値数 {18} とした。

欠測のパターンを Figure 2 に示す。パターン A は、前半の6データに欠測が存在する。パターン B は中央部の6データ、パターン C は後半の6データに欠測が存在する。パターン D はすべてのデータに欠測が存在する。パターン D の欠測データの割合は 17% である。パターン E は、すべてのデータに欠測が存在し、50% のデータが失われている。パターン D および E では、欠測データがある場合に広く用いられている方法 1：欠測のあるデータを削除する方法は適用できない。

用いた階層型ニューラルネットワークの構造は3層型であり、ニューロン数はバイアスニューロンを含み、それぞれ 6, 4, 1 である。中間層のニューロン数 4 は、欠損のないデータを再構築学習法を用いて学習させ、10 から減少させることによって得た。中間層のユニットを 3, 4, 5 としたニューラルネットワークの情報量基準 AIC [16] はそれぞれ、95.2, 73.3, 92.1 であり、この場合 4 が適当であることが判る [17]。

本例での CQSAR による欠測データの補完は、データ数が7以上であるので、ニューラルネットワークと等差数列を使う補完法 [11] を用いておこなった。

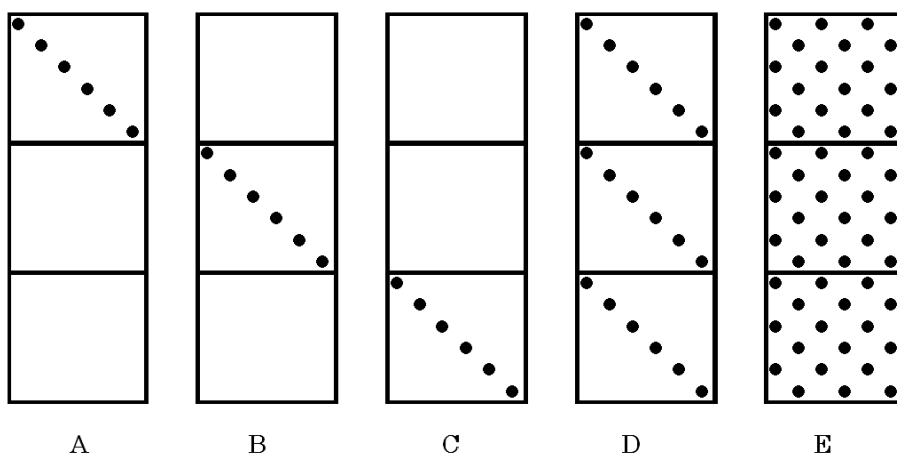


Figure 2. Patterns of defects.

Table 6. Averaged standard deviation of error(ASDE) and AIC for each defects pattern

Pattern of defects	CQSAR		Approach1:remove defects	
	ASDE (ratio*)	AIC(ratio*)	ASDE	AIC
A: defects in the first block	0.0151(8.6)	74.2(1.18)	0.1301	87.6
B: defects in the center	0.0046(1.4)	67.3(1.06)	0.0063	71.6
C: defects in the terminal	0.0343(3.2)	75.5(1.11)	0.1107	84.2
D: dispersed defects of 17%	0.0132(∞)	71.1(∞)	- impossible -	
E: dispersed defects of 50%	0.0195(∞)	74.4(∞)	- impossible -	

* ratio=Approach1/CQSAR

Table 7. Ratios of defect parts and scopes of two methods

years	defects	indexes/observation points	approach-1	CQSAR
1994	0.0000	0.0000/0.0000	practicable	practicable
1995	0.0000	0.0000/0.0000		
1996	0.0000	0.0000/0.0000		
1997	0.0000	0.0000/0.0000		
1998	0.0000	0.0000/0.0000		
1999	0.0000	0.0000/0.0000		
2000	0.0245	0.2941/0.0833		
2001	0.1471	0.7059/1.0000	impossible	
2002	0.2010	0.7059/1.0000		

leave-one-out 法および CQSAR で予測された値を学習したニューラルネットワークの AIC を計算し学習の検定を行った。leave-one-out 法での誤差の平均標準偏差と AIC を Table 6 に方法 1 の結果とともに示した。誤差の平均標準偏差をみると、パターン A では、CQSAR は方法 1 に比べ 10 倍程度性能が良いことが判る。パターン B, C でもわずかであるが良い。CQSAR も方法 1 も、パターン B に比べ A, C が悪いのは、それらが外挿に対応しているからである。パターン D, E に関して、CQSAR がパターン A とほぼ同等の精度を持っていることは驚きである。言うまでもなくパターン D, E に対し、方法 1 は適用できない。

CQSAR で補完を実施したニューラルネットの AIC は、方法 1 よりも小さな値を示しており、より妥当な学習をしていることがうかがえる。また、AIC の値が小さいほど誤差の平均標準偏差が小さく、両者には何らかの相関があるように見える。

4 CQSAR の多摩川の水質データへの適用

欠測データを含む多摩川の水質データについて、欠測を含むデータを除く方法 1 と CQSAR の適用可能性を比較した。Table 7 に各年度のデータの欠測の割合と適用可能性を示した。Table 7 の indexes と observation points の意味を説明するためにデータの構造の概要を Figure 3 に示した。

1994 年から 2000 年までのデータではすべてのデータが存在するため方法 1 でも学習可能である。ところが 2001 年と 2002 年では垂直方向にすべてのデータを除かなければならないので方法 1 の適用は不可能であった。一方、CQSAR では、2000 年から 2002 年においても欠測データを補完するため、学習可能であった。

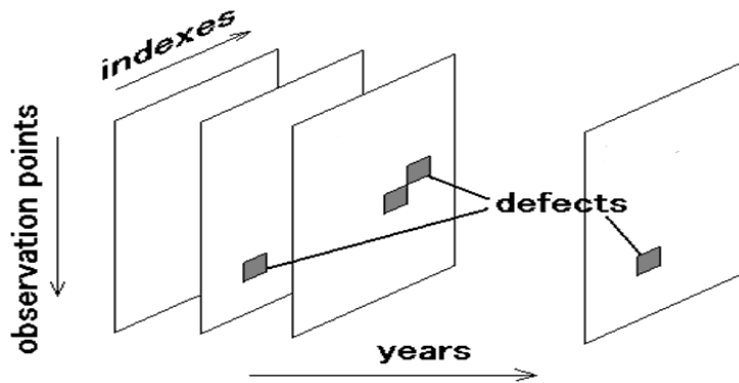


Figure 3. Structure of data.

5 多摩川のデータでの上・中・下流の3つの分類の検定

Tables 3-5 に示した測定データを見れば下流の 16-17 地点の CI と COND が極端に大きいことから，中流と下流の分類を与えるパラメータは CI と COND であることが直ちに判る．ところが第 1-6 地点を上流，7-15 を中流とする分類を与えるパラメータは明白ではない．そのため，CQSAR を用いて，1) 第 1-7 地点を上流，8-15 を中流，16-17 を下流とした場合，2) 第 1-6 地点を上流，7-15 を中流，16-17 を下流とした場合，3) 第 1-5 地点を上流，6-15 を中流，16-17 を下流とした場合の 3 つの場合について学習の誤差および AIC を指標にその分類の妥当性を評価した．

それぞれのネットワーク構造は再構築学習法により最適化し，AIC を用いて確認した．この時，再構築学習は 10 万回行い，50 回ごとに 0.05 以下の結合を消去した．消去はそれぞれの場合，合計 2000 回行われネットワーク構造の最適化が行われた．また最適化された中間層ニューロン数に 1 を足した(または引いた)ネットワークで学習を行い，AIC を計算して中間層ニューロン数の妥当性を検証した．例えば 2) の場合，再構築学習法では中間層ニューロン数が 2 となったので，中間層のニューロン数を，1, 2, 3 として学習を行いそれぞれの AIC を計算すると，66.2, 28.2, 32.6 となり，中間層ニューロン数 2 が妥当であることが判った．

最終的なネットワーク構造は，1) の場合が (12, 3, 1)，2), 3) の場合が (12, 2, 1) であった．

Table 8. Conformation of assignment to upper region

Assignment of upper region	Max. error	average error	AIC
1)	0.0770	0.00704	34.2
2)	0.0026	0.00098	28.2
3)	0.0044	0.00111	28.5

それぞれの場合の学習誤差の最大と平均および AIC を Table 8 に示す．1-6 を上流とする場合が，一番エラーの少ない学習をしており，AIC も一番小さい．この結果より Tables 3-5 にあるように，第 1-6 地点を上流，7-15 を中流，16-17 を下流とすることは妥当であることが判る．

6 CQSAR による学習と再構築学習法を用いた結果

先の偏微分係数解析を用いた解析 [1] は，上流・中流・下流を分ける明確なパラメータ抽出はできなかった．CQSAR と再構築学習法を用いて学習を行うと，ネットワーク構造が最小化され中間層ニューロン数が 2 で十分であることが判った．通常のバックプロパゲーション学習法では 5 ニューロン以下では学習が上手くいかず，6 ニューロン必要であった．これが前回の論文 [1] で偏微分係数で特定のパラメータが抽出できなかった理由である．言い換えれば，本例の場合は，適切な学習には再構築学習法のようなシミュレーテッドアニーリング法とバックプロパゲーション法の組み合わせが必要であるということである．

Table 9. Weight matrix of optimized neural network by reconstruction learning

The first layer – the second layer													
	pH	DO	BOD	COD	T-N	T-P	Cl ⁻	NH ₄ -N	NO ₂ -N	NO ₃ -N	PO ₄ -P	COND	
1	0	-18	0	0	0	35	0	0	0	0	2	0	
2	0	22	0	0	0	-38	0	0	0	0	-5	0	

The second layer – the third layer		
	1	2
1	43	50

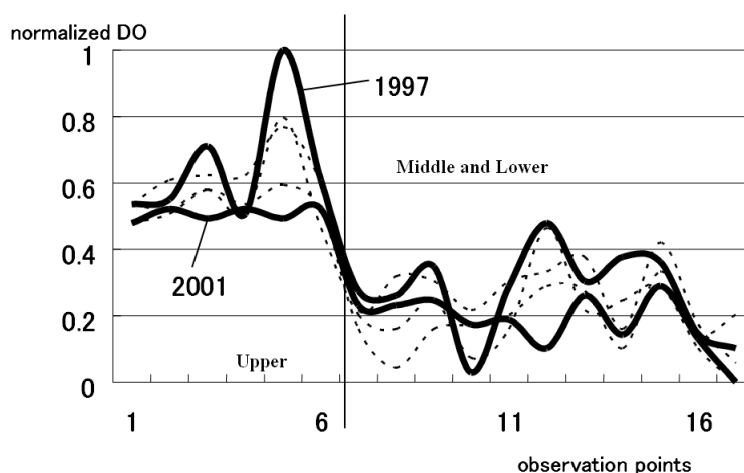


Figure 4. The changes of DO-index at 17 observation points between 1997 and 2001.

最適化されたニューラルネットワークの結合行列を Table 9 に示した。再構築学習法による結合の淘汰により，上流と中・下流を分ける主なパラメータとして，DO（溶存酸素量）と T-P（全リン）および PO₄-P（リン酸性リン）が抽出された。ただし PO₄-P の影響は小さい。以下 DO と T-P について考察する。

Figure 4 に 1997 年から 2001 年までの各測定地点での DO の変化を示す。ここでは，データ中の最大値（第 5 地点，1997 年）を 1.0，最小値（第 17 地点，1997 年）を 0.0 としてスケールしてある。DO は値が大きいほど清浄であることを示す。これを見ると確かに第 6 地点と第 7 地点の間にターニングポイントがあることがわかる。第 5 地点を見ると 1997 年で非常に清浄であった水質が 2001 年にかけて徐々に汚染が進んでいることが判る。反対に第 10 地点は清浄さが回復している。第 3 地点は 1997 年から清浄化が進んだが，2001 年にはまた 1997 年レベルまで汚染が元に戻ってしまった。逆に第 8 地点では，1997 年から汚染が進み，2001 年にはまた 1997 年レベルまで清浄化が回復した。

この第 3 地点と第 8 地点，第 10 地点を除くと 1997

年から 2001 年にかけてほぼすべての流域において水質汚染が進んでいることが判った。

また各観測地点の T-P 変化を Figure 5 に示す。T-P は値が大きいほど汚染が進んでいることを示している。ここでも最大値と最小値を 1.0, 0.0 としてスケールしてある。これを見ると DO と同様第 6 地点と第 7 地点の間にターニングポイントがあることが判る。

ここで判ることは T-P は上流すなわち第 1 地点から第 6 地点に関しては全く情報を与えないが，中・下流域で値が大きくなっている。中流域および下流域の T-P の変化は，1997 年から 1999 年にかけて汚染が進み，その後汚染は減少し，2001 年では 1997 年位までは回復していることを示している。

新聞の記事（2004 年 3 月 21 日朝日新聞）によると「羽村堰の流水効果で多摩川の水質は年々良くなっている」というコメントは，T-P を見る限り正しいようであるが，DO ではそのような傾向は見られない。今後堰の清浄化の効果を見積もるための方法の開発が必要である。

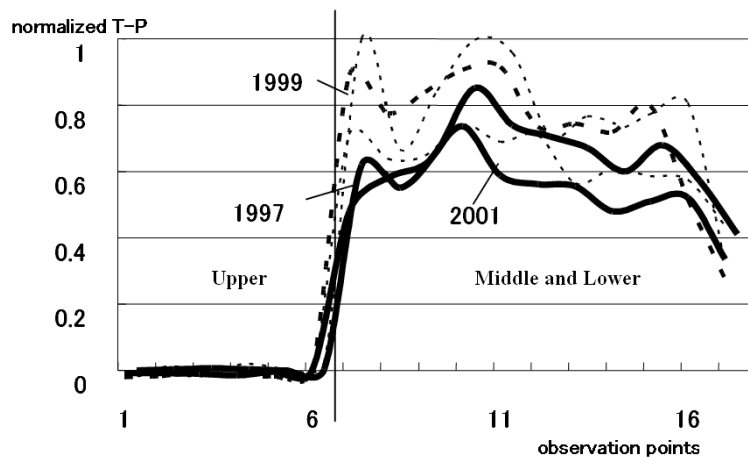


Figure 5. The changes of T-P-index at 17 observation points between 1997 and 2001.

水質汚染という観点からみると、多摩川では上流から中流域の水質の変化に比べて中流から下流域の水質変化の方が緩やかであることが判った。このことにより多摩川の水質保全には、中流域での汚染原因を取り除くことが重要であることが示唆された。

7 まとめ

欠測データがあり、十分な解析が困難であった多摩川の水質データに、欠測データを推測して補完する方法である CQSAR を適用して解析を行った。ニューラルネットワークによる多摩川にデータでの上流および中・下流の分類の検定で、第 6 地点を上流と中流の境界とすべきであることが判った。

CQSAR では、欠測データを補完し、学習するため、欠測を持つ 2000 年から 2002 年のデータ解析も可能となった。

本研究により水質検査項目の 12 種類のパラメータのうち、主なパラメータとして DO と T-P が抽出された。

DO はすべての流域において水質汚染が進んでいることを示した。T-P は上流に関して全く情報を与えないが、中・下流で浄化傾向にあることを示した。T-P の結果は朝日新聞の記事の記述にある「羽村堰の流水効果で多摩川の水質は年々良くなっている。」を一部支持した。

今後堰の効果を見積もるための河川モデルの構築が必要である。

水質汚染という観点からみると、多摩川では上流か

ら中流域の水質の変化に比べて中流から下流域の水質変化の方が緩やかであることが判った。このことにより多摩川の水質保全には、中流域での汚染原因を取り除くことが重要であることが示唆された。

参考文献

- [1] 神部順子, 福田朋子, 長嶋雲兵, 青山智夫, *J. Chem. Software*, **8**, 27 (2002).
- [2] 井須芳美, 長嶋雲兵, 細矢治夫, 青山智夫, *J. Chem. Software*, **2**, 76 (1994).
- [3] 井須芳美, 長嶋雲兵, 細矢治夫, 大島茂, 坂本曜子, 青山智夫, *J. Chem. Software*, **3**, 1 (1996).
- [4] Isu, Y., Nagashima, U., Aoyama, T., Hosoya, H., *J. Chem. Info. Comp. Sci.*, **36**, 286 (1996).
- [5] 藤谷康子, 小野寺光永, 井須芳美, 長嶋雲兵, 細矢治夫, 青山智夫, *J. Chem. Software*, **4**, 19 (1998).
- [6] 田島澄恵, 松本高利, 田辺和俊, 長嶋雲兵, 細矢治夫, 青山智夫, *J. Chem. Software*, **6**, 115 (2000).
- [7] 福田朋子, 田島澄恵, 斎藤久登, 長嶋雲兵, 細矢治夫, 青山智夫, *J. Chem. Software*, **7**, 115 (2001).
- [8] Aoyama, T., Ichikawa, H., *Chem. Pharm. Bull.*, **39**, 372 (1991).
- [9] Aoyama, T., Wang, Q., Zhu, H., Nagashima U., *IPSJ-SIG notes*, **2000-HPC-84**, 7 (2000).

- [10] 公共用水域および地下水の水質測定結果, 東京都環境局 (1994-2002 年度) .
- [11] Bitou, K., Yuan, Y., Aoyama, T., Nagashima, U., Proceedings of International Conference on Control Automation and Systems 2003, CD-ROM (TA06-04), 10.22-25, 2003.
- [12] 青山智夫, 神部順子, 長嶋雲兵, *J. Comp. Chem. Jpn* (2006), 投稿中.
- [13] 渡辺美智子, 山口和範, *EM アルゴリズムと不完全データの諸問題*, 多賀出版, ISBN4-8115-5701-8.
- [14] 渡辺美智子, Web 資料:
[http://stat.eco.toyo.ac.jp/~michiko/em/emohp\(0\)/emohp.ppt](http://stat.eco.toyo.ac.jp/~michiko/em/emohp(0)/emohp.ppt)
- [15] Aoyama, T., Ichikawa, H., *Chem. Pharm. Bull.*, **39**, 1222 (1991).
- [16] Akaike, H., *IEEE Trans. on Automatic Control*, **AC-19**, 716 (1974).
- [17] 栗田多喜夫, 電子情報通信学会論文誌, **J73-D-II**, 1872 (1990).

Extraction of Chemical Parameters Characterizing the Upper, the Middle and Lower Stream by CQSAR - The Case of the Tamagawa River, Tokyo, Japan -

Junko KAMBE^a, Yuan YAN^b, Umpei NAGASHIMA^{c,d,*} and Tomoo AOYAMA^b

^aFaculty of Foreign Language, Daito Bunka University
1-9-1 Takashimadaira, Itabashi, Tokyo 175-8571, Japan

^bFaculty of Engineering, University of Miyazaki
Gakuenkhanadai Nishi, Miyazaki 889-2192 Japan

^cResearch Institute for Computational Science, National Institute for Advanced Industrial Science and Technology
1-1-1 Umezono, Tsukuba, Ibaraki 305-8565 Japan

^d Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST)
Kawaguchi Center Building, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012 Japan

*e-mail: u.nagashima@aist.go.jp

We attempted to extract chemical parameters characterizing the upper, middle and lower stream of the Tamagawa River in Tokyo, Japan, by Compensation Quantitative Structure-Activity Relationships (CQSAR) method [11, 12] for analysis of incomplete data by perceptron type neural network with three layers.

Dissolved Oxygen (DO) and Total Phosphorus (T-P) were extracted from 12 chemical parameters (Table 2) at 17 points (Table 1) along the main stream of the Tamagawa River (Figure 1) in Tokyo, Japan, for 1997-2001 [10].

DO indicated that water pollution became worse at all points of the Tamagawa River (Figure 4) but T-P shows that water pollution is changed for the better in the middle and the lower stream (Figure 5).

This result suggests that the water pollution is more drastic in the middle stream of the Tamagawa River than the lower stream. The water pollution in the middle stream should be decreased for keeping the Tamagawa River clean.

Keywords: Water pollution, Tamagawa River, Compensation quantitative structure activity relationship (CQSAR), Dissolved oxygen (DO), Total phosphorus(T-P)