

主成分分析法とニューラルネットワークを用いた 河川の上流・中流・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を用いて -

神部 順子^a, 福田 朋子^{b, c}, 長嶋 雲兵^{d*}, 青山 智夫^e

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

^b 日本女子大学家政学部生活芸術学科, 〒 112-8681 文京区目白台 2-8-1

^c (株)ベストシステムズ, 〒 305-0035 つくば市松代 4 丁目 15-2-1-204

^d 産業技術総合研究所先端情報計算センター, 〒 305-8561 つくば市東 1-1-1

^e 宮崎大学工学部, 〒 889-2192 宮崎市学園木花台西 1-1

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

(Received: February 28, 2001; Accepted for publication: October 10, 2001; Published on Web: December 7, 2001)

河川の上流・中流・下流を示す水質パラメータを抽出するために、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」(1997年から1999年)[4]に掲載の多摩川の17地点の水質に関するデータを主成分分析と新たにJavaを用いてニューラルネットワークシミュレータNeco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[2, 3]を用いて解析した。

主成分分析の結果から、上流・中流・下流の分類は第一主成分と第二主成分の主成分得点を組み合わせることで可能であることがわかった。しかしながら、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であることがわかった。

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わかった。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 Contamination, Principal Component Analysis, Differential Coefficients Analysis of Input Parameter for Neural Network, Upper Stream, Middle Stream, Lower Stream

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。本研究では、川の水の汚染度すなわち上流、中流および下流域を端的に示す化学的パラメータを抽出すること、すなわち河川の水質を記述する多くの複雑なパラメータから、上流・中流・下流を記述するパラメー

タを探すことを目的として、ケモトリックス[1]としてもっとも広く用いられている化学的多変量統計解析の一つである主成分分析を用いてパターン認識を行い、そのパターンがどのような化学的パラメータによるのかを検討した。主成分分析は線形解析に基づくものであるため、さらに非線形解析を基にするニューラルネットワークを用いた解析として、青山が開発した

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 [2, 3] を行い、パラメータの挙動を解析した。偏微分係数解析は、ニューラルネットワークによる分類に対する入力変数の感受性を調べるもので、入力変数の分類への影響の度合いをみることができる。

本研究で行った東京都の多摩川の水質汚染に関する水質データを用いた主成分分析では、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。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いえた。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わかった。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的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。

本報告では、線形解析手法である主成分分析と K-L Plot および非線形解析手法である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を用いて、17の測定地点を上流・中流・下流域の3つに分類するパラメータを抽出することを試み、さらにそれらのパラメータの挙動を解析したので報告する。

非線形解析手法である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プログラムは、新たに Java を用いて実装した。

2 解析データ

解析に用いた河川の水質の観測データは、東京都環境局が発行した河川の水質測定データ集「公共用水域および地下水の水質測定結果」(1997-1999年度) [4]のうちから、源流から河口までのデータが比較的均質にそろっている多摩川の多摩湖から多摩川本流を通じて東京湾に注ぐ17のサンプル採取地点での水質分析データ3年分を用いた。17地点の場所とサンプルの分析機関を Table 1 に示す。また、17のサンプル採取地点の海からの距離を大まかにとらえるために、測定地点の概略を Figure 1 に示す。

「公共用水域および地下水の水質測定結果」データ集では、測定点1の和田橋から測定点6の拝島橋までを上流域、測定点7の日野橋から測定点15の田園調布堰までを中流域、測定点16と17の六郷橋と大師橋を下流域としてあるため、それを上流、中流、下流の指標としてそのまま用いた。「公共用水域および地下水の水質測定結果」には、水質を示す化学的パラメータとして55種類のパラメータが掲載されている。しかし、本研究では17地点のすべてに測定結果があり、かつ17地点ですべてが同じ値ではない12種類の化学的パラメータについて解析をおこなった。

「公共用水域および地下水の水質測定結果」(1997-1999年度)には各年度の月単位のデータが掲載されているが、本報告では1997年度から1999年度にかけて、もっともデータのそろっている2月のデータを用いた結果を報告する。(8月のデータを用いてもほぼ同様な結果を得ているが、本論文では、紙面の都合上8月のデータを用いた解析結果の記述は割愛する。)

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		

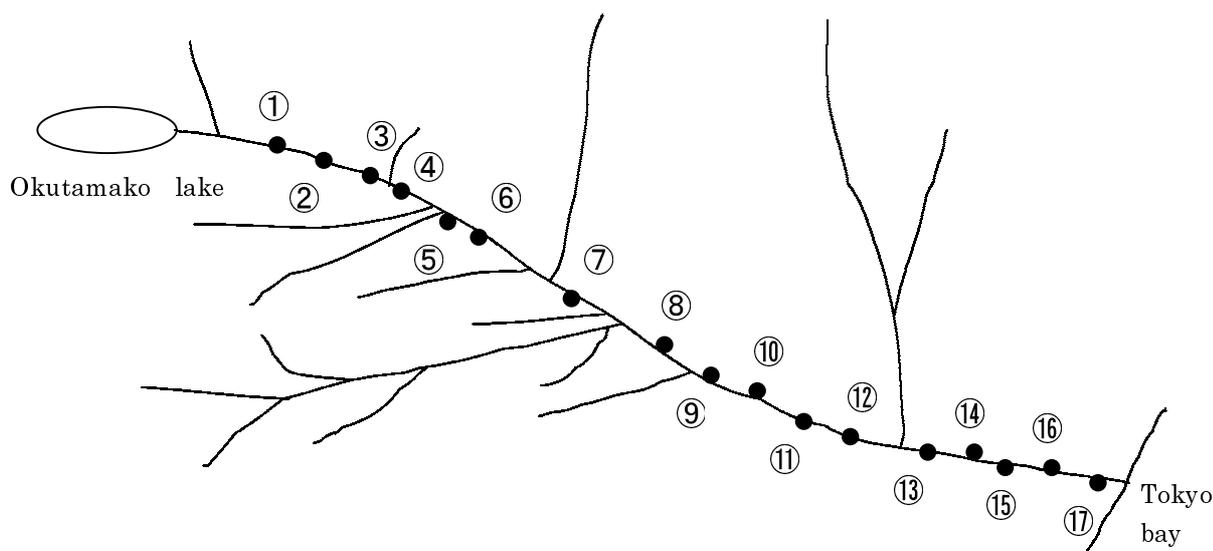


Figure 1. Schematic Map of Sampling Points along Tamagawa River, Tokyo, JAPAN

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 ~ 1 mg / 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。

Table 2 にパラメータの説明と略号、Table 3 に解析に用いたデータを示した。Table 2 および Table 3 には、上流、中流、下流の指標を pos として併記した。

主成分分析には、多変量解析プログラム Excel 統計 2000(Windows98 版)[5] を用い、ニューラルネットワークには、我々が開発している分子の構造活性相関解析

のためのニューラルネットワークシミュレータ: Neco (NEural network simulator for structure-activity COrrrelation of molecules)[6-11] を用いた。入力変数の偏微分計算は新たに実装した。すべてパーソナルコンピュータ (SONY VAIO PCG-Z505VR/K, CPU: mobile Pentium III 800MHz, memory:256MB) を用いて行った。

Table 3. Values of chemical index of water analysis (1997-1999)

1997 Feb.	pos.	pH	DO	BOD	COD	T-N	T-P	Cl ⁻	NH ₄ -N	NO ₂ -N	NO ₃ -N	PO ₄ -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) 羽村堰	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 ⁻	NH ₄ -N	NO ₂ -N	NO ₃ -N	PO ₄ -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) 羽村堰	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 ⁻	NH ₄ -N	NO ₂ -N	NO ₃ -N	PO ₄ -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) 羽村堰	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

3 主成分分析の結果

本節では、先に説明した河川の水質検査データを用いた主成分分析の結果をもとに、上流・中流・下流域を区別するパラメータの抽出を行う。

上流、中流、下流の指標 Pos. を除いた 12 パラメータを用いた主成分分析によって得られた主成分に対し

て、累積寄与率が 70% を越えるまでの主成分について議論することとする。今回の解析結果では、第一主成分の寄与率が 61.0% であり、第二主成分の寄与率は 20.1% であった。第三主成分の寄与率は、7%、第四主成分は、6% である。

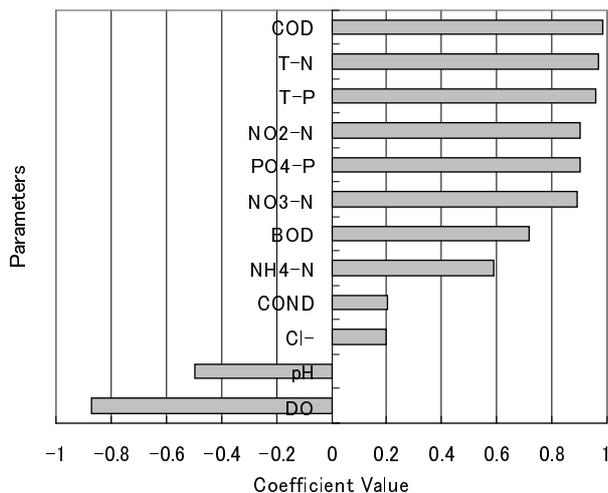


Figure 2. Sorted coefficients in the first principal component

Figure 2 と Figure 3 に 3 年分のデータに対する第一主成分と第二主成分の主成分負荷量を示した。第一主成分は、COD・BOD・T-N・T-P・NH₄-N・NO₃-N・NO₂-N・PO₄-P の 8 パラメータが正の大きな寄与をしており、pH と DO は大きく負の寄与をしている。第一主成分は 12 パラメータの全てにわたってほぼまんべんなく寄与している。ただし Cl⁻ と COND の寄与は他のパラメータに比べ小さい。

逆に第二主成分に対しては Cl⁻ と COND のパラメータが大きな寄与をしており、他のパラメータの寄与は小さい。つまり、第二主成分は Cl⁻ と COND の寄与を代表しているといえる。BOD、や NH₄-N および PO₄-P また NO₃-N も寄与するが、それらの影響は Cl⁻ と COND より小さい。

第一主成分と第二主成分における 17 地点の主成分得点を Table 4 に示した。また Figure 4 には、この得点を用いた K-L プロットを示した。Table 4 および Figure 4 をみると、上流域 (1 から 6)、中流域 (7 から 15)、下流域 (16、17) が第一主成分と第二主成分によって分類されていることがわかる。第一主成分では、上流と中流・下流の分類が可能であり、中流と下流は分離できていない。他方第二主成分では上流・中流と下流がだまかに分離されている。

より詳細にみると第二主成分の主成分得点が 1 以上の領域が下流域を示しており、第一主成分が -3 以下で

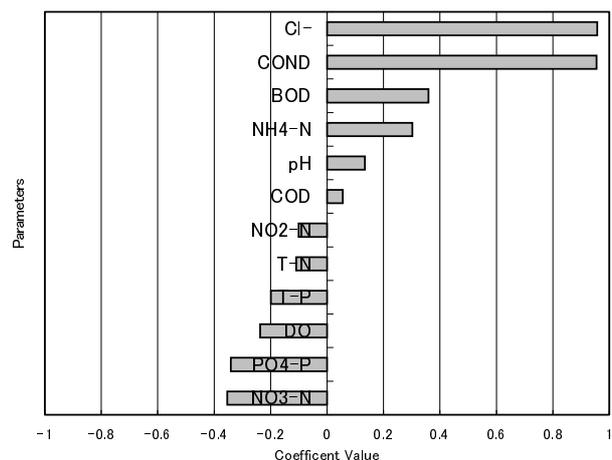


Figure 3. Sorted coefficients in the second principal component

第二主成分が 0 付近が上流域、第一主成分が正で第二主成分が負のものが中流域となっている。つまり、上流・中流・下流の分類は第一主成分と第二主成分の主成分得点を組み合わせることで可能であることがわかった。第二主成分から Cl⁻ と COND が上流・中流・下流をだまかに分類するうえで重要なパラメータであることが示唆されるが、しかし、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いことがわかる。Figure 4 では、1997 年に比べ 1998 年は汚染が進み、1999 年にわずかに汚染の度合いが改良されたことを伺うことができる。特に下流域の六郷橋 (16 番目) の 1998 年のデータは、さらに下流の大師橋のそれと同じ位になり、1999 年に回復していることをみることができる。

これらの結果を水質汚染の観点から見ると、第一主成分によって上流と中流・下流がよく分離されることから、多摩川は上流域から中流域の水質汚染の度合いに比べ、中流域から下流域の水質汚染の度合いは小さいということがいえる。言い換えると上流から中流へ移る段階 (地点 6 と 7 の間) で流入する支流の汚染を除去する等、中流域での水質汚染原因を取り除き中流域での汚染原因を少なくすることが多摩川の環境保全を行う上で重要であることが示唆された。

Table 4. REGR factor score of the first and second principal components

Position	pos	1997		1998		1999	
		First	Second	First	Second	First	Second
1	1	-3.416	-0.059	-3.2	-0.102	-3.35	-0.106
2	1	-3.475	-0.069	-3.5	0.0834	-3.65	-0.027
3	1	-3.742	-0.142	-3.29	-0.205	-3.4	-0.198
4	1	-3.434	0.0385	-3.51	0.1037	-3.7	-0.006
5	1	-4.469	-0.229	-3.25	-0.234	-3.46	-0.36
6	1	-3.34	-0.035	-3.33	0.0643	-3.55	-0.034
7	2	0.6631	-0.587	2.547	-0.7	2.139	-0.615
8	2	0.8876	-0.296	1.956	-0.471	2.411	-0.486
9	2	0.5345	-0.607	2.042	-0.869	2.058	-0.926
10	2	4.2193	-0.662	3.152	-1.157	2.886	-1.005
11	2	2.5541	-0.562	3.009	-1.014	2.326	-1.143
12	2	1.3003	-0.992	1.282	-1.128	1.468	-1.326
13	2	2.0611	-0.578	2.104	-0.883	1.423	-1.038
14	2	0.8614	-0.816	2.303	0.065	1.665	-1.07
15	2	1.6896	-0.625	1.849	-0.974	1.197	-1.138
16	3	2.4256	1.4307	2.974	5.4314	2.58	2.532
17	3	1.2656	3.1877	0.675	5.4827	0.558	5.0519

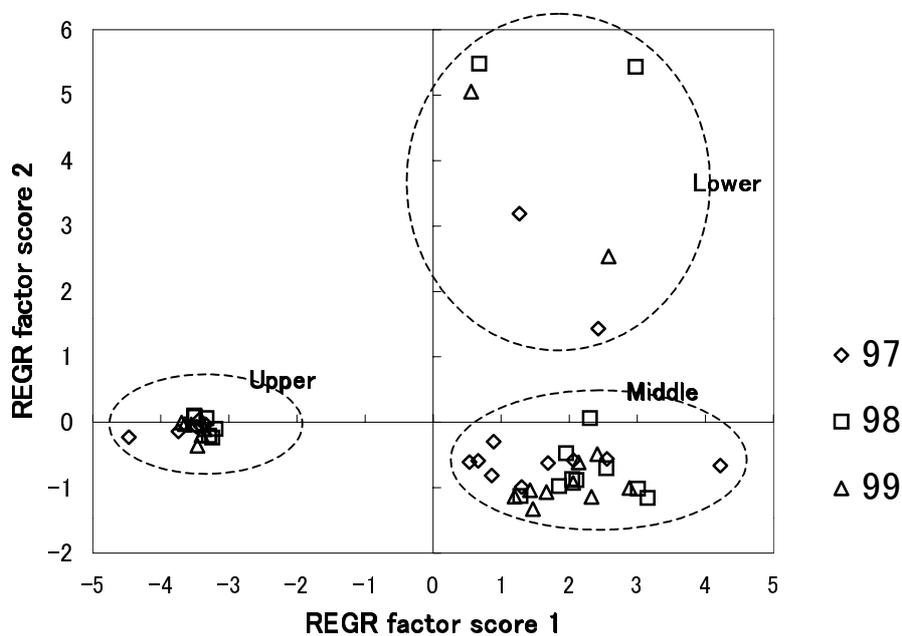


Figure 4. the K-L plot of component score

Differential Coefficients(1997-1999)

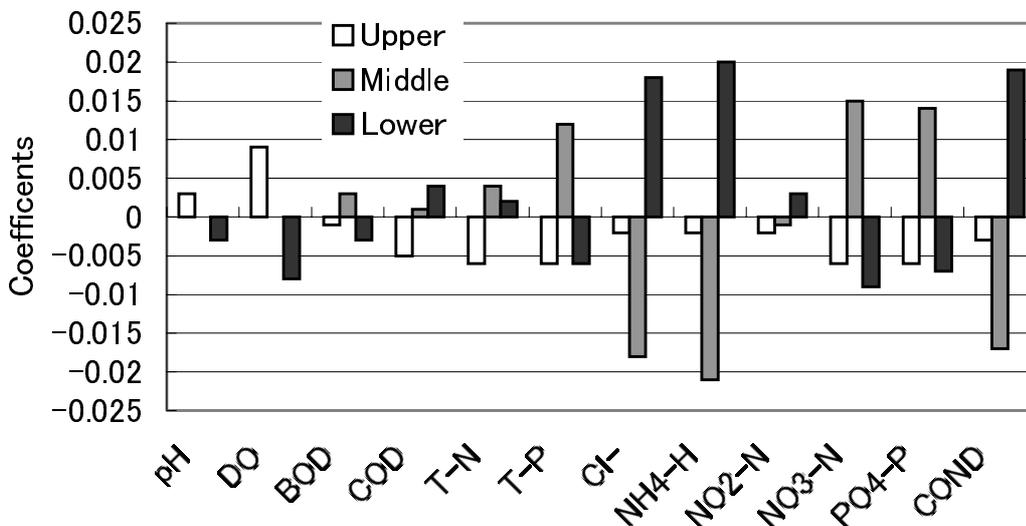


Figure 5. Averaged differential coefficients (1997-1999)

4 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の結果

青山の開発した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法 [2, 3] を新たに我々が開発中のパーセプトロン型ニューラルネットワークシミュレータ Neco[6-11] に実装した。

本研究で用いた三層パーセプトロン型ネットワークは、入力層、中間層、出力層ニューロン数がそれぞれ13,12,3である。入力層には12種類の測定項目と常に1.0の値のバイアスを入力し、出力層は上流、中流、下流を001, 010, 100とそれぞれ、表現するものとした。中間層ニューロン数は、3年分のデータに対して再構築学習法を用いて最適化した。学習の収束のための閾値は、誤差の最大値が0.005とした。

偏微分係数解析は、入力パラメータの感受性を明らかにするものである。これにより、個々のパラメータが分類にどのように寄与するのかをみることができる。

3年間の51データを用いた偏微分係数の平均値をそれぞれ Figure 5 に示した。微係数の値は小さく、これらの変数は単独ではどれもあまり大きな影響を上流・中流・下流の分類に与えないことが示唆される。

ただ主成分分析から示唆されたように、Cl-, COND, NH4-N など微係数が大きいことが判る。しかしなが

ら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ことがうかがえる。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。このように偏微分係数解析は入力パラメータの分類に対する寄与を詳細にみることができる。

さらに、Figure 6 に各年度ごとのデータを用いた微係数の平均値を示した。化学的パラメータの上流・中流・下流への影響の大きな傾向は、1998年と1999年がほぼ同じであり、1997年はNH4-Nの重要度が1998年、1999年に比較して低くなっている。また1998年と1999年とは中流と下流で符号が異なっており、NH4-Nは1997年は上流と中流を分類するパラメータであったが、1998年と1999年では中流と下流の分類に関するパラメータとなっていることがわかる。これは、1998年にNH4-Nに反映する汚染が中流域で急激に進み下流域に流出していることを示しており、1999年にわずかにNH4-Nに反映する汚染は解消されている傾向が見て取れるが、まだ完全には除去されていないことがわかる。

また、Figure 5 に示す全体の傾向と、Figure 6 の1998年の傾向がよく似ていることは、1998年の中流域でのNH4-Nに反映する汚染が大きなものであったことを示唆している。

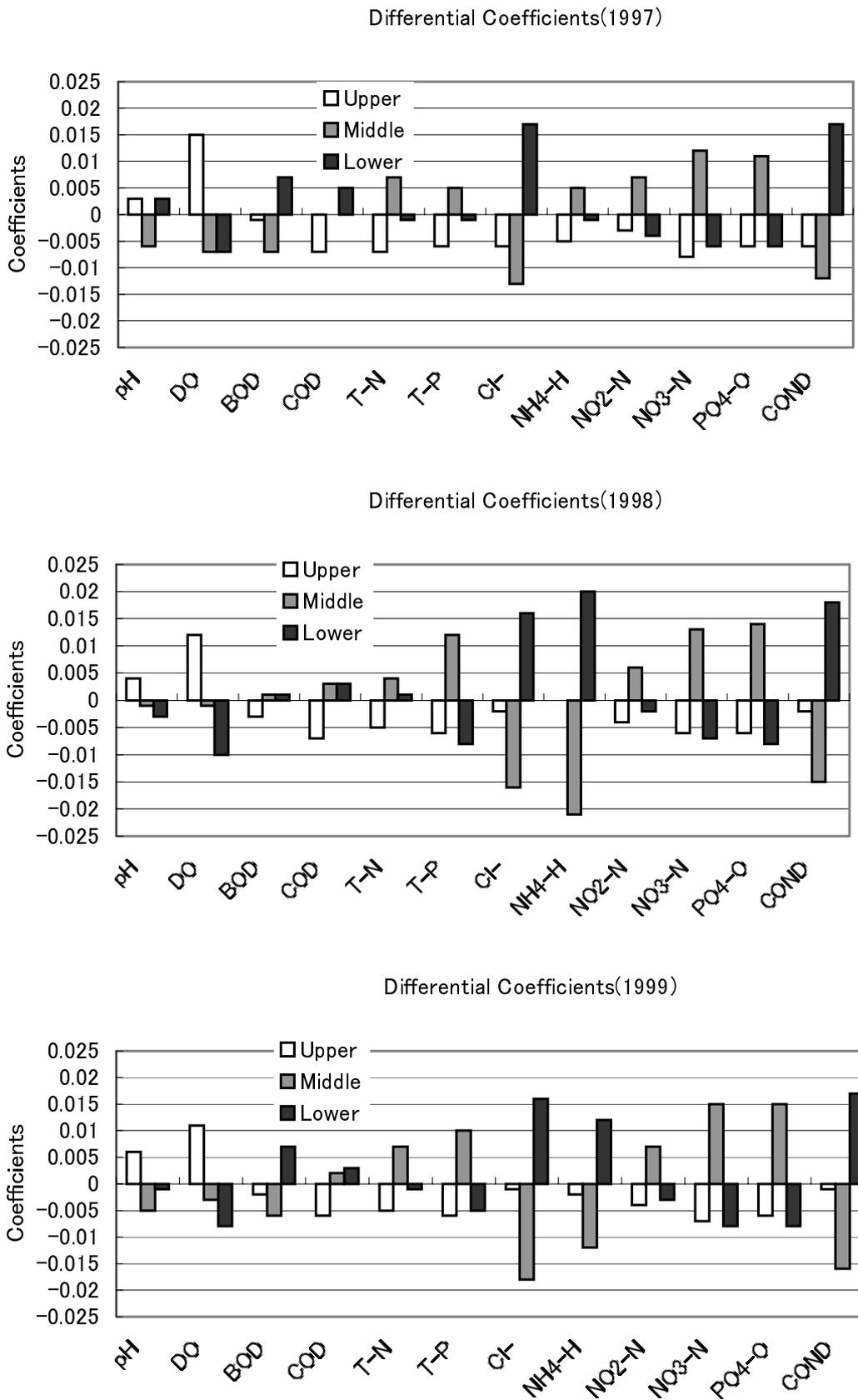


Figure 6. Averaged Differential Coefficients of 1997, 1998 and 1999

5 まとめ

多摩川の17地点における水質検査データ3年分を用いた主成分分析の結果と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の結果をもとに、上流・中流・下流域を区別するパラメータの抽出を試みた。

各地点でのデータがそろっている12種類のデータを用いると第一主成分および第二主成分を座標軸とした主成分得点のK-L Plotを用いて、上流・中流・下流域をパターン認識的に分類することが可能であった。第二主成分において大きな寄与をするCl⁻およびCONDが上流・中流・下流を大まかに分類するうえで重要なパラメータであることが示唆されたが、各流域を分けるのに大きな寄与をするパラメータの抽出はできなかった。

ニューラルネットワークを用いた偏微分係数解析の結果では、12種類の入力パラメータの微係数の平均値は小さく、これらの変数は単独ではどれもあまり大きな影響を上流・中流・下流の分類に与えないことが示唆された。ただCl⁻, COND, NH₄-Nなど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であることがわかった。

多摩川の水質保全のためには、中流域での汚染原因を取り除き、そこでの汚染の程度を減少させることが必要であることが示唆された。

いつも有益な御議論をさせていただいている、お茶の水女子大学の細矢治夫教授、田島澄恵氏 産総研の松本

高利博士、豊橋技大の船津公人助教授に深く感謝する。

参考文献

- [1] 宮下芳勝, 佐々木慎一, ケモメトリックス - 化学パターン認識と多変量解析 -, 共立出版 (1995).
- [2] Aoyama, T., Ichikawa, H., *Chem. Pharm. Bull.*, **39**, 372 (1991).
- [3] Aoyama, T., Wang, Q., Zhu, H., Nagashima, U., *IPSJ-SIG notes*, **2000-HPC-84**, 7 (2000).
- [4] 公共用水域及び地下水の水質測定結果, 東京都環境局 (1997-1999年度のデータを用いた。)
- [5] Excel 統計 2000, 社会情報サービス http://www.ssri.com/software/ex2000/e_license.html
- [6] 井須芳美, 長嶋雲兵, 細矢治夫, 青山智夫, *J. Chem. Software*, **2**, 76 (1994).
- [7] 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).
- [8] Isu, Y., Nagashima, U., Aoyama, T., Hosoya, H., *J. Chem. Info. Comp. Sci.*, **36**, 286 (1996).
- [9] 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).
- [10] 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).
- [11] 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).

Extraction of Chemical Parameter Characterizing the Upper, Middle and Lower Stream by Principal Component Analysis and Neural Network

– The Case of Tamagawa River, Tokyo, Japan –

Junko KAMBE^a, Tomoko FUKUDA^{b, c}, Umpei NAGASHIMA^{d,*} and Tomoo AOYAMA^e

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

^bDepartment of Life Arts, Faculty of Home Economics, Japan Women's University
2-8-1 Mejirodai, Bunkyo-ku, Tokyo 112-8681, Japan

^cBestsystems Co. Ltd., 4-15-2-1-204 Matsushiro, Tsukuba, Ibaraki 305-0035, Japan

^dTsukuba Advanced Computing Center, National Institute for Advanced Industrial Science and Technology
1-1 Higashi, Tsukuba, Ibaraki 305-8565 Japan

^eFaculty of Technology, Miyazaki University
Gakuenkhanadai Nishi, Miyazaki 889-2192 Japan

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

We attempted to extract chemical parameter characterizing the upper, middle and lower stream by the principal component and the analysis of differential coefficients of input parameter for perceptron type neural network with three layers. The analysis of differential coefficients of input parameter for perceptron type neural network was developed by Aoyama [2] and was newly equipped into a neural network simulator Neco. The data used are 12 chemical parameters at 17 points along the main stream of the Tamagawa river in Tokyo, Japan, for 1997-1999 [3].

The K-L plot of the first and second principal components (Figure 4) well divides 17 points into three groups corresponding to the three regions: upper, middle and lower streams, respectively. From results of the analysis of differential coefficients of input parameter for perceptron type neural network, Cl⁻, COND and NH₄-N have relatively large differential coefficients and divide middle and lower streams. DO and pH are large in upper stream of Tamagawa river (Figure 5). The first principal component classifies well two groups: upper and middle-lower streams on the K-L plots. This result suggests that the water contamination is more drastic in the midstream of Tamagawa river than downstream. The water contamination in midstream should be decreased for keeping Tamagawa river clean.

Keywords: Water Contamination, Principal Component Analysis, Differential Coefficients Analysis of Input Parameter for Neural Network, Upper Stream, Middle Stream, Lower Stream