

## 線形多変量解析を用いた徳島吉野川の水質の解析

神部 順子<sup>a</sup>, 青山 智夫<sup>b</sup>, 山内 あい子<sup>c</sup>, 長嶋 雲兵<sup>d,e,\*</sup>

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

<sup>b</sup> 宮崎大学工学部, 〒 889-2192 宮崎市学園木花台西 1-1

<sup>c</sup> 徳島大学薬学部, 〒 770-8505 徳島市庄町 1-78-1

<sup>d</sup> 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

<sup>e</sup> 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: September 11, 2006; Accepted for publication: October 3, 2006; Published on Web: December 1, 2006)

吉野川の水質浄化の指針を得るために、吉野川水系の上流から河口に至るまでの5種類の水質パラメータの変化を主成分分析とクラスター分析を用いて解析し、吉野川水系の上流から下流までの汚染度をみるための複合的な指標を得ることを試みた。吉野川の汚染の傾向を見るためには、少なくとも5つのパラメータほぼ同じ程度重要であった。

吉野川は4つの河口を持つ (Figure 1) が、水質は河口からの距離とは対応しておらず (Table 5)、最も河口からの距離が長い平和橋 (銅山川) と河口の1つである新町川の水質汚染が進んでいることがわかった (Figure 6, Table 12)。このことから、それぞれの流域の水質に影響を与えている要因を個々に探る必要があることが示唆された。

吉野川の水質浄化のためには、このような要因の解析と総合的な環境指標に基づいた解析が重要であることがわかった。

キーワード: Water Pollution, Yoshinogawa River, Chemometrics, Principal Component Analysis, Cluster Analysis

### 1 はじめに

吉野川は徳島県内最大の一級河川であり、幹線流路延長 194km のうちに数多くの支川と合流しながら流下する。この川は近年四国一きれいな川として知られている穴吹川を支流として抱える一方で、下流部には昭和 40 年代に水質汚濁で知られた新町川を派川としている [1]。吉野川流域は多くの支流の合流や分流があり、複雑である。そのため、総合的な水質の解析は困難である。

本研究では、吉野川の水質浄化の指針を得るために、吉野川水系の上流から河口に至るまでの5つの水質パラメータの変化を主成分分析とクラスター分析を

用いて解析し、吉野川水系の上流から下流までの汚染度をみるための複合的な指標を得ることを試みた。主成分分析はなるべく少ない合成変数で、なるべく多くの情報を把握するという情報の縮約を実現する方法であるため、汚染度をみるための複合的な指標を得るために用いた。それにより、測定地点のグループ化が可能であった。クラスター分析は似通った測定点あるいは変数のグループ化を行うための分析手法であり、本研究ではユークリッド距離を用いたグループ化と、主成分分析によって得られた第一主成分と第二主成分のKahunen-Loeveプロット (K-L Plot) によるグループ化の対応を確認した。解析には WindowsPC 上で SPSS ver. 11.0[2] を用いた。

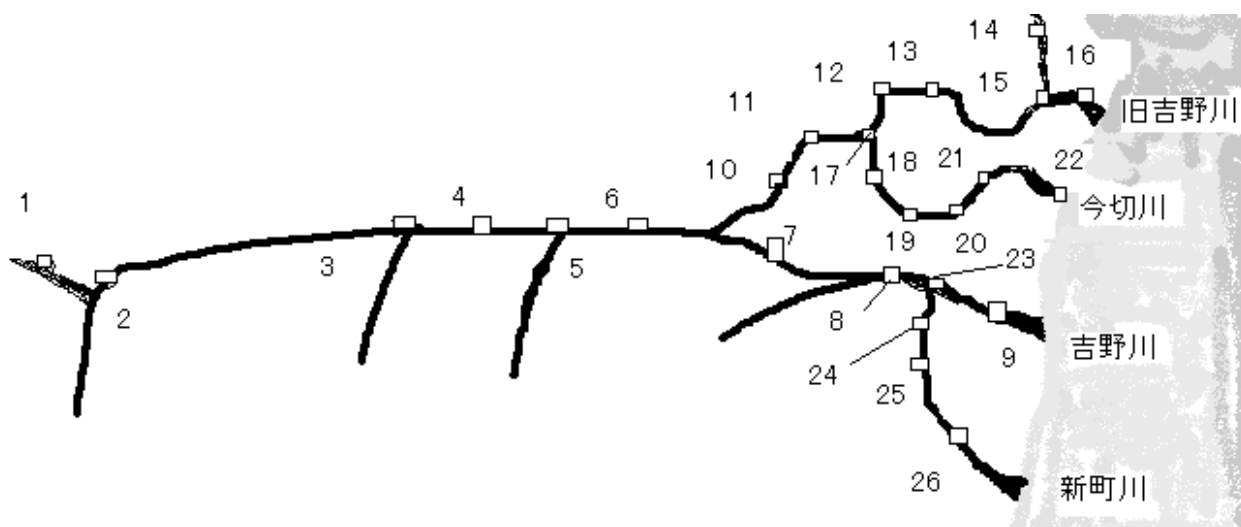


Figure 1. Schematic Map of Sampling Points along the Yoshinogawa, Tokushima, Japan. See Table 2 for details of monitoring points.

Table 1. Abbreviations and explanation of chemical index of water analysis

Chemical index (abbreviations)	Explanation
Dissolved Oxygen (DO)	Amount of dissolved oxygen that is freely available in water to sustain fish and other aquatic organisms. Lower DO level indicates higher level of water pollution.
Biochemical Oxygen Demand (BOD)	Total amount of oxygen consumed in the biological processes that break down organic matter in the water. Higher BOD level indicates higher level of water pollution.
Chemical Oxygen Demand (COD)	Mass concentration of oxygen consumed by the chemical breakdown of organic and inorganic matter. Higher COD level indicates higher level of water pollution.
Total Nitrogen (TN)	Total amount of nitrogen compounds contained in water. Can be divided into inorganic and organic matter groups, as well as into dissolved matter and particulate matter groups.
Total Phosphorus (TP)	Total amount of phosphorus compounds contained in the water. Can be divided into inorganic and organic matter groups, as well as into dissolved matter and particulate matter groups.

本研究では，国立環境研究所が提供する「国立環境研究所環境データベース」公共用水域水質年間値データファイル [3] を利用した．このデータは，全国の公共用水域において都道府県等が測定した結果を環境省がとりまとめたものである．2002 年度のデータで 26 地点（支流も含む）の 5 種類の水質パラメータについて分析した．このデータには，欠測が有り再測定不能

であるためそれらを線形関数とニューラルネットを用いて補完した．これにより，得られる情報を最大限利用することができる．

Figure 1 に吉野川 26 の測定地点の河口からの距離を大まかに捉えるために測定地点の概略を示す．吉野川は中・下流域が複雑であることがわかる．

解析に用いた 5 種類の水質データの略号と内容を

Table 1 に示す．ここで，TP は 8 地点，TN は 5 地点で欠測となっている．そこで，まず最初に 3 つのパラメータ ( BOD , COD , DO ) がそろっている 26 測定地点 3 パラメータのデータを用い，主成分分析とクラスター分析を行った．その後 TP と TN のデータを加えた 5 つのパラメータがそろっている 18 測定地点 5 パラメータのデータを用いて，同様の解析を行った．

18 測定地点 5 パラメータの結果から，TP と TN も河川の汚染度を見るのに重要であることがわかったので，さらに TP と TN の値が欠測となっている地点のデータを予測し，そのデータを用いて 26 測定地点 5 パラメータの解析をおこなった．

## 2 データ

Table 2 に 2002 年の吉野川のデータを示す．Table 2 の陰の部分には，欠測であることを示している．

## 3 26地点3パラメータ( BOD, COD, DO )を用いた線形多変量解析

Table 2 の全てのデータが揃っている 26 地点 3 パラメータ ( BOD, COD, DO ) を用いたクラスター分析によるデンドログラムを Figure 2 に示す．結合されたクラスター間のユークリッド距離の 2 乗が 2.0 以上で分類すると，26 測定地点は 4 グループに分類することができる．

グループ 1 は，河口からの距離が長い順に平和橋 (1) から吉野川水系が 4 つに分岐するまでの 6 測定地点と，共栄橋 (12) 以外の旧吉野川上流水域の 3 測定地点，合計 9 測定地点である．水質が清浄であると解釈される測定地点のグループである．

グループ 2 は旧吉野川上流の共栄橋 (12) と旧吉野川下流の大津橋 (16)，旧吉野川に流れ込む支流の撫養川 2 測定地点，今切川のうち鍋川合流点 (20) 以外の 5 測定地点，吉野川下流 2 測定地点とこの 2 つの間に流れ込む支流の鮎喰川 (8)，新町川のうち新町水門 (23)，合計 13 測定地点である．

Table 2. Values of chemical index of water analysis and distances from estuary of monitoring points in the Yoshinogawa region

Fig.1の 番号	水域名	調査地点名	BOD	COD	DO	TN	TP	河口か らの距 離	各水域河口か らの順位
1	銅山川	平和橋	1.2	1.9	9.9	0.91	0.02	94.0	吉野川河口9
2	吉野川上流	大川橋	0.7	1.4	10.0	0.29	0.01	87.0	吉野川河口8
3	貞光川	貞光	0.9	1.5	11.0	0.50	0.02	55.0	吉野川河口7
4	吉野川下流	脇町潜水橋	0.8	1.6	10.0	0.44	0.01	48.0	吉野川河口6
5	穴吹川	穴吹	0.5	0.8	10.0	0.41	0.01	45.0	吉野川河口5
6	吉野川下流	高瀬橋	0.7	1.6	9.3	0.79	0.02	19.0	吉野川河口4
7	吉野川下流	送電線下	1.7	2.6	8.9		0.05	10.5	吉野川河口3
8	鮎喰川	鮎喰	1.6	3.1	8.2	0.98	0.11	6.5	吉野川河口2
9	吉野川下流	吉野川大橋	1.1	2.5	8.6	0.50	0.05	3.5	吉野川河口1
10	旧吉野川上流	藍園橋	0.7	1.8	9.6	0.73	0.04	18.5	旧吉野川河口7
11	旧吉野川上流	市場橋	1.0	1.9	9.3	0.89	0.04	10.7	旧吉野川河口6
12	旧吉野川上流	共栄橋	1.5	1.8	9.2			8.5	旧吉野川河口5
13	旧吉野川上流	牛屋島橋	1.2	2.2	9.6	0.89	0.04	7.0	旧吉野川河口4
14	撫養川	城見橋	1.6	3.4	8.1			4.5	旧吉野川河口2
15	撫養川	大里橋	1.6	2.6	7.6	0.65	0.07	6.5	旧吉野川河口3
16	旧吉野川下流	大津橋	1.2	2.3	9.0	0.66	0.05	2.3	旧吉野川河口1
17	今切川上流	三ッ合橋	1.6	1.8	9.1			10.7	今切川河口6
18	今切川下流	鯛浜橋	1.4	2.4	8.9			7.7	今切川河口5
19	今切川上流	鯛浜堰上流側	1.7	2.6	9.9	0.93	0.06	7.5	今切川河口4
20	今切川下流	鍋川合流点	2.0	4.0	8.1			5.0	今切川3
21	今切川下流	加賀須野橋	1.5	2.5	8.6	0.88	0.06	4.0	今切川2
22	今切川下流	河口	1.5	3.1	8.0	0.89	0.06	0.0	今切川1
23	新町川上流	新町水門	1.0	2.7	7.7		0.09	7.3	新町川河口4
24	新町川上流	三ッ合橋	2.2	4.4	6.1		0.15	5.6	新町川河口3
25	新町川上流	新町橋	3.4	5.3	6.0	1.60	0.20	4.2	新町川河口2
26	新町川下流	漁連前	2.0	4.3	7.2	0.68	0.10	1.7	新町川河口1

Shaded parts indicate defect data

Dendrogram using Average Linkage (Between Groups)

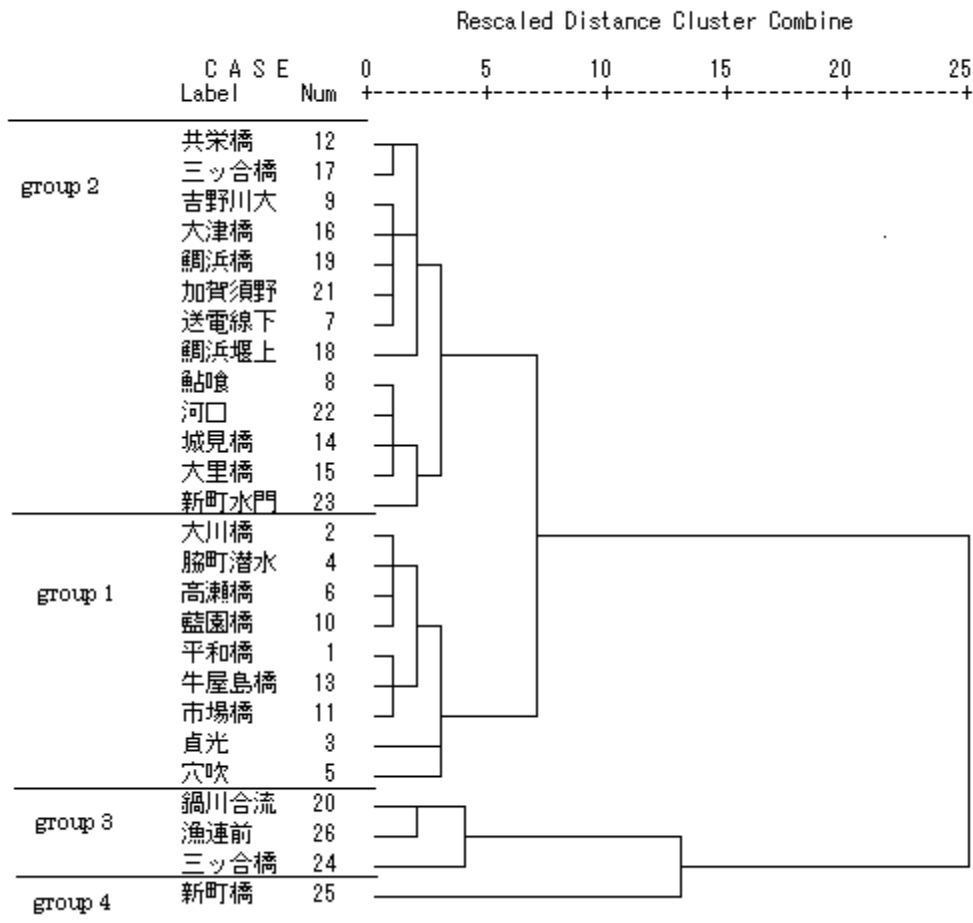


Figure 2. Dendrogram of the Yoshinogawa, 3 parameters, 26 monitoring points

グループ 3 は今切川水域の測定地点のうち最も汚染の進んでいる鍋川合流点 (20), および新町川の三ッ合橋 (24) と漁連前 (26) の 3 測定地点である。グループ 4 は新町橋 (25) のみである。新町橋 (25) は他のグループとは距離が大きく離れており、全測定地点の中で、特に水質汚染が進んでいる。

BOD と COD は値が大きいほど汚染の度合いが進んでいることを示すパラメータであり、DO は値が小さいほど汚染の度合いが進んでいることを示すパラメータである。第一主成分 (fac. 1) の係数の符号が負であるのはその関係を示している。そのため第一主成分の値が小さいほど清浄である。第二主成分は BOD と DO が大きな重みをもっている。

Table 3 に 26 地点 3 パラメータを用いた主成分分析の第一から第三主成分の固有値と寄与率および累積寄与率を示す。26 地点 3 パラメータを用いた主成分分析の寄与率は、第一主成分 (90.1%) と第二主成分 (7.54%) を合わせると 95 % を超えており、第一主成分だけで

汚染程度の記述がほぼ可能である。Table 4 に 26 地点 3 パラメータを用いた主成分分析の主成分負荷量を示す。第一主成分の係数は全ての項目がほぼ等価の重みを持つ。

26 地点 3 パラメータを用いた主成分分析の第一主成分の主成分得点の高い順に並べたものをクラスター分析によるグループ番号および各測定地点の河口からの距離を Table 5 に示す。主成分分析による第一主成分の主成分得点は、クラスター分析によるグループ番号とよく適応していることがわかる。

一方、主成分得点は河口からの距離と対応しておらず、それぞれの流域の水質の悪化に何らかの影響を与える原因が個々にあることが示唆された。特に、今切川流域、新町川流域は、河口からの距離が長い領域 (上流) の方で汚染が進んでいる。また、新町橋 (25) が最も汚染が進んでいるので、三ッ合橋 (24) と新町橋 (25) 間での汚染源の特定とその除去が急がれる。

Table 3. Eigenvalues and contribution (%) of principal components (the Yoshinogawa, 3 parameters, 26 monitoring points)

	eigenvalue	contribution rate(%)	cumulative contribution rate(%)
fac.1	2.70	90.12	90.12
fac.2	0.23	7.54	97.66
fac.3	0.07	2.34	100.00

fac1: the first principal component  
 fac2: the second principal component  
 fac3: the third principal component

Table 4. Coefficients in the first and second principal component (the Yoshinogawa, 3 parameters, 26 monitoring points)

	fac.1	fac.2
BOD	0.936	0.332
COD	0.977	0.008
DO	-9.934	0.341

fac1: the first principal component  
 fac2: the second principal component  
 fac3: the third principal component

Table 5. Score of Factor 1, groups in cluster analysis and distances (the Yoshinogawa, 3 parameters, 26 monitoring points)

Fig.1の番号	水域名	調査地点名	主成分分析による第1因子の主成分得点	クラスター分析によるグループ	河口から距離	各水域河口からの順位
5	穴吹川	穴吹	-1.468	1	45.0	吉野川河口5
3	貞光川	貞光	-1.286	1	55.0	吉野川河口7
2	吉野川上流	大川橋	-1.147	1	87.0	吉野川河口8
4	吉野川下流	脇町潜水橋	-1.021	1	48.0	吉野川河口6
10	旧吉野川上流	藍園橋	-0.894	1	18.5	旧吉野川河口7
6	吉野川下流	高瀬橋	-0.877	1	19.0	吉野川河口4
1	銅山川	平和橋	-0.660	1	94.0	吉野川河口9
11	旧吉野川上流	市場橋	-0.602	1	10.7	旧吉野川河口6
13	旧吉野川上流	牛屋島橋	-0.470	1	7.0	旧吉野川河口4
12	旧吉野川上流	共栄橋	-0.321	2	8.5	旧吉野川河口5
16	旧吉野川下流	大津橋	-0.263	2	2.3	旧吉野川河口1
17	今切川上流	三ッ合橋	-0.235	2	10.7	今切川6
9	吉野川下流	吉野川大橋	-0.137	2	3.5	吉野川河口1
19	今切川上流	鯛浜堰上流側	-0.133	2	7.5	今切川4
18	今切川下流	鯛浜橋	-0.086	2	7.7	今切川5
21	今切川下流	加賀須野橋	0.092	2	4.0	今切川2
23	新町川上流	新町水門	0.134	2	7.3	新町川河口4
7	吉野川下流	送電線下	0.155	2	10.5	吉野川河口3
8	鮎喰川	鮎喰	0.471	2	6.5	吉野川河口2
22	今切川下流	河口	0.472	2	0.0	今切川1
15	撫養川	大里橋	0.472	2	4.5	旧吉野川河口2
14	撫養川	城見橋	0.603	2	6.5	旧吉野川河口3
20	今切川下流	鍋川合流点	1.039	3	5.0	今切川3
26	新町川下流	漁連前	1.401	3	1.7	新町川河口1
24	新町川上流	三ッ合橋	1.867	3	5.6	新町川河口3
25	新町川上流	新町橋	2.892	4	4.2	新町川河口2

26地点3パラメータを用い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観測地点を K-L Plotした結果を、クラスター分析で得られたグループと対応させて Figure 3 に示す。観測地点は第一主成分に沿って左側から、一番清浄なグループ1から汚染のもっとも進んでいるグループ4まで順に並んでおり、クラスター分析によって4つに分けられたグループは、第一主成分得点の大きさとよく対応した結果となった。

#### 4 欠測値のない18地点での5パラメータを用いた解析

次に、TPとTNのデータを加えた5つの観測値がそろっている18測定地点のデータを用いて、同様の解析を行った。

Figure 4に18地点5パラメータを用いたクラスター分析によるデンドログラムを示す。

ここで、結合されたクラスター間のユークリッド距離の2乗が2.0以上になったものは、6つのグループに分類することができた。グループ1は、河口からの距離が長い順に大川橋(2)から穴吹(5)の4測定地点である。グループ2は、平和橋(1)、吉野川大橋(9)と旧吉野川のうち共栄橋(12)を除いた4測定地点である。大川橋(2)よりもより上流に位置する平和橋(1)はグループ2に分類された。これは26地点3パラメータでの分析結果とは異なっている。平和橋(1)と大川橋(2)の間に水質浄化作用のある要因があることが伺えるが、具体的にその要因が何であるかは不明である。グループ3は今切川の鯛浜堰上上流側(18)のみである。グループ4は大里橋(15)、今切川下流の加賀須野橋(21)と河口(22)、鮎喰(8)の4測定地点である。グループ5は新町川の漁連前(26)のみである。グループ6は新町橋(26)のみである。新町橋(26)は26地点3パラメータでの分析結果と同様、他のグループとは大きく距離が離れており、新町橋(26)の上流の汚染源の存在を強く示唆している。

2つのパラメータを加えることで、3つのパラメータ

では分離できなかった清浄なグループの2つ(3パラメータの場合のグループ1と2)に属する観測点がそれぞれ2つのグループに分離された。

Table 6に18地点5パラメータを用いた主成分分析の第一から第五主成分の固有値と寄与率および累積寄与率を示す。18地点5パラメータを用いた主成分分析の寄与率は、第一主成分(87.0%)と第二主成分(8.5%)を合わせて95%を超えている。この傾向は26地点3パラメータの傾向と一致している。

Table 7に18地点5パラメータを用いた主成分分析の主成分負荷量を示す。第一主成分の係数は全ての項目が等しい重みを持ち、DOの係数の符号が負である。これも26地点3パラメータの結果と一致しているが、さらにTNとTPも水質を表す成分のコンポーネントとして重要であることを示している。第二主成分の係数はTNが0.54、DOが0.34で大きな重みをもっている。26地点3パラメータの結果の第二主成分はBODとDOが大きく、傾向が異なっている。これらによりTNとTPも汚染度をみる指標として重要であることがわかる。

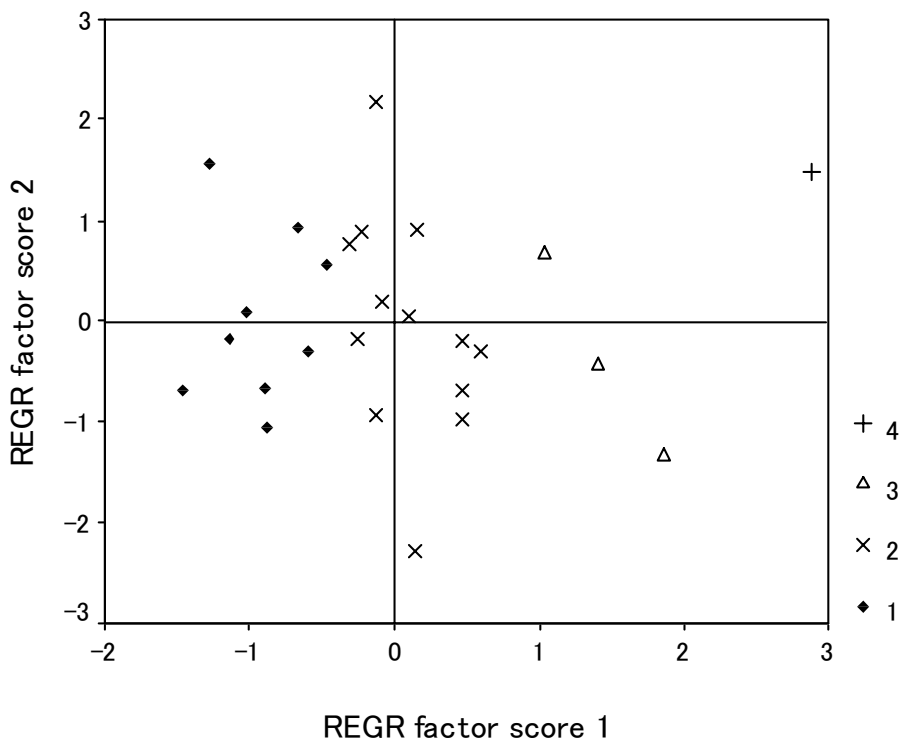


Figure 3. K-L plots by factors 1 and 2 of the Yoshinogawa, 3 parameters, 26 monitoring points

Dendrogram using Average Linkage (Between Groups)

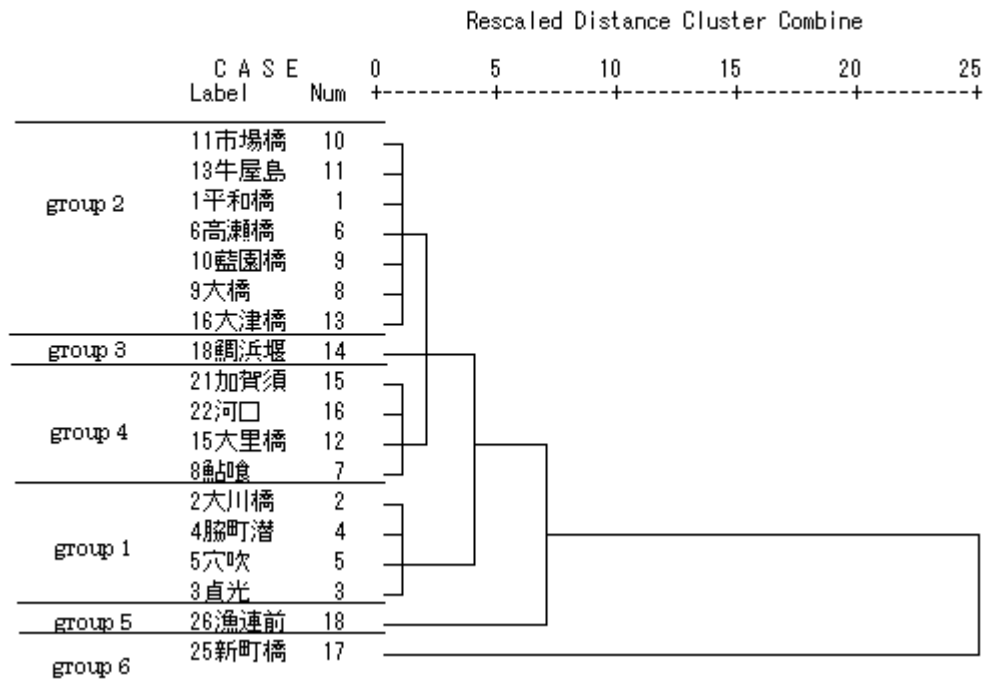


Figure 4. Dendrogram of the Yoshinogawa, 5 parameters, 18 monitoring points

Table 6. Eigenvalues and contribution (%) of principal components (the Yoshinogawa, 5 parameters, 18 monitoring points)

	eigenvalue	contribution rate(%)	cumulative contribution rate(%)
fac.1	4.35	87.04	87.04
fac.2	0.42	8.50	95.53
fac.3	0.14	2.74	98.27
fac.4	0.05	1.07	99.34
fac.5	0.03	0.66	100.0

fac1: the first principal component  
 fac2: the second principal component  
 fac3: the third principal component  
 fac4: the fourth principal component  
 fac5: the fifth principal component

Table 7. Coefficients in the first and second principal component (the Yoshinogawa, 5 parameters, 18 monitoring points)

	fac.1	fac.2
BOD	0.972	0.031
COD	0.970	-0.142
DO	-0.904	0.339
TN	0.832	0.536
TP	0.978	-0.032

fac1: the first principal component  
 fac2: the second principal component

Table 8. Score of Factor 1 and groups in cluster analysis (the Yoshinogawa, 5 parameters, 18 monitoring points)

Fig.1の 番号	水域名	調査地点名	主成分分析による 第1因子の 主成分 得点	クラス ター分析 による グループ
5	穴吹川	穴吹	-1.202	1
2	吉野川上流	大川橋	-1.088	1
3	貞光川	貞光	-0.986	1
4	吉野川下流	脇町潜水橋	-0.916	1
6	吉野川下流	高瀬橋	-0.557	2
10	旧吉野川上流	藍園橋	-0.508	2
1	銅山川	平和橋	-0.351	2
11	旧吉野川上流	市場橋	-0.233	2
9	吉野川下流	吉野川大橋	-0.161	2
13	旧吉野川上流	牛屋島橋	-0.155	2
16	旧吉野川下流	大津橋	-0.133	2
18	今切川上流	鯛浜堰上流側	0.167	3
21	今切川下流	加賀須野橋	0.267	4
15	撫養川	大里橋	0.389	4
22	今切川下流	河口	0.499	4
8	鮎喰川	鮎喰	0.798	4
26	新町川下流	漁連前	1.107	5
25	新町川上流	新町橋	3.063	6

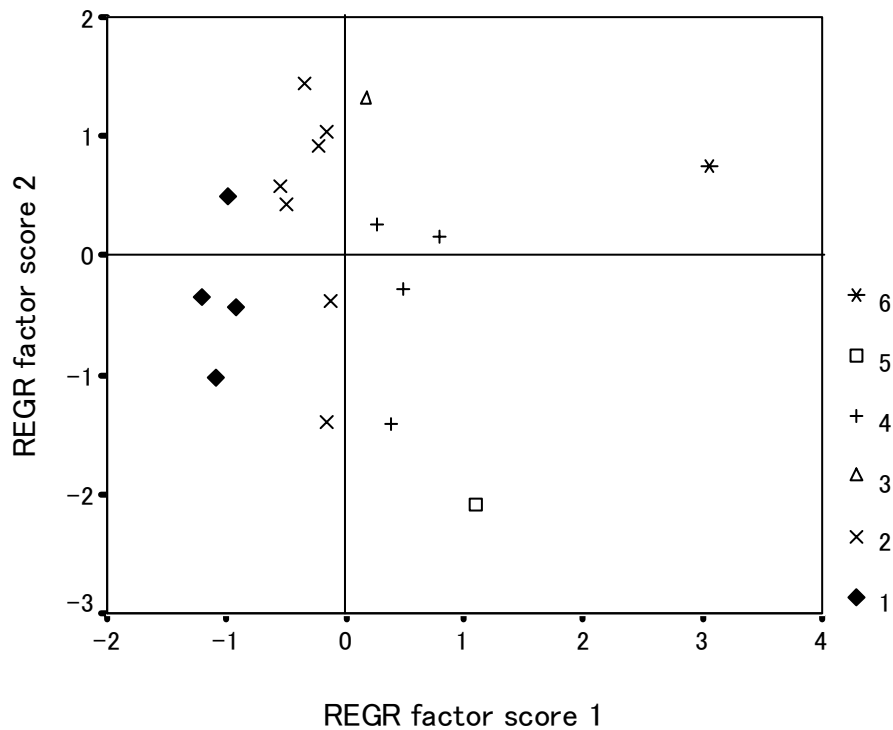


Figure 5. K-L plots by factors 1 and 2 (5 parameters, 18 points)

18 地点 5 パラメータを用いた主成分分析の第一主成分の主成分得点の高い順に並べたものとクラスター分析によるグループを Table 8 に示す。主成分分析による第一主成分の主成分得点は、クラスター分析によるグループの結果とよく適応していることがわかる。26 地点 3 パラメータと比べると、18 地点 5 パラメータは 26 地点 3 パラメータのグループ 1 が 2 つのグループ (グループ 1, 2) に分かれ、26 地点 3 パラメータのグループ 2 が 3 つのグループに分解された。これは、パラメータを増やすことにより、より精密な分類が行われたことに他ならない。

18 地点 5 パラメータを用い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 K-L Plot した結果をクラスター分析で得られたグループと対応させたものを Figure 5 に示す。6 つに分けられたグループのうち、グループ 1, 2, 5, 6 は、第一主成分得点とよく対応した結果となった。だが、クラスター分析によって分けられたグループ 3 とグループ 4 は第一主成分だけではなく、第一主成分得点と第二主成分得点の組み合わせによって、ようやく分離することができる。

5 つのパラメータで分析した結果、3 パラメータのときは異なり、最も河口からの距離が長い平和橋 (1) は、水質汚染が進んでいるグループに分類された。パラメータを増やしたことによって、より吉野川の水質の特徴が明確になったが、欠測のある測定地点を除いたことによって、吉野川全体を知ることは困難になってしまった。特に、汚染が進んでいる新町川に関しての情報が不足している。そこで欠測値を補完し、測定地点を増やしてさらに水質変化を検討する必要がある

ことがわかった。

## 5 欠測値の予測後、26 地点 5 パラメータを用いた解析

TP の値が欠測となっている 8 地点と TN の値が欠測となっている 5 地点のデータを補完した。欠測値の補完には線形関数とニューラルネットを用いて比較した。線形関数の予測の方法は、5 パラメータがそろっている 18 地点のデータで、各パラメータの 2 パラメータ間の相関係数を求め、その相関係数が高いものを採用することにした ( $TN=0.3522 \times BOD+0.3008$ ,  $TP=0.041 \times COD-0.0435$ )。このときニューラルネットワークは PSDD[4, 5] と NECO[6-11] を用いた。

Table 9 に線形関数とニューラルネットによる予測の結果を示す。TN に関して共栄橋 (12) と三ツ合橋 (17) および (24) の値が絶対値で両者は 0.5 以上異なる。TP は共栄橋 (12) と三ツ合橋 (17) および鯛浜橋 (19) の値が絶対値で 0.1 程度異なった。

TP の予測に関してニューラルネットは大きな値を予測しているが、TN と TP の観測値があるものと比べ、TP の値は大きすぎる。特に、共栄橋 (12) の  $TN=0.16$  に対して  $TP=0.11$ 、三ツ合橋 (17) の  $TN=0.38$  に対して  $TP=0.12$  であり、整合性に欠ける。そのため今回は、線形関数による予測値を採用した。

Figure 6 に 26 地点 5 パラメータを用いたクラスター分析によるデンドログラムを示す。ここで、結合されたクラスター間のユークリッド距離の 2 乗が 2.0 以上になったもので分類すると、26 測定地点は 7 つのグループに分類することができた。

Table 9. Estimation of defect data

Fig.1の 番号	水域名	調査地点名	TN(線形 関数による 予測)	TN(NNIに よる予測)	TP(線形 関数による 予測)	TP(NNIに よる予測)
7	吉野川下流	送電線下	0.90	0.77		
12	旧吉野川上流	共栄橋	0.83	0.16	0.03	0.11
14	撫養川	城見橋	0.86	0.89	0.10	0.12
17	今切川上流	三ツ合橋	0.86	0.38	0.03	0.12
19	今切川下流	鯛浜橋	0.79	0.62	0.05	0.11
20	今切川下流	鍋川合流点	1.01	1.32	0.12	0.15
23	新町川上流	新町水門	0.65	0.51		
24	新町川上流	三ツ合橋	1.08	1.70		

Dendrogram using Average Linkage (Between Groups)

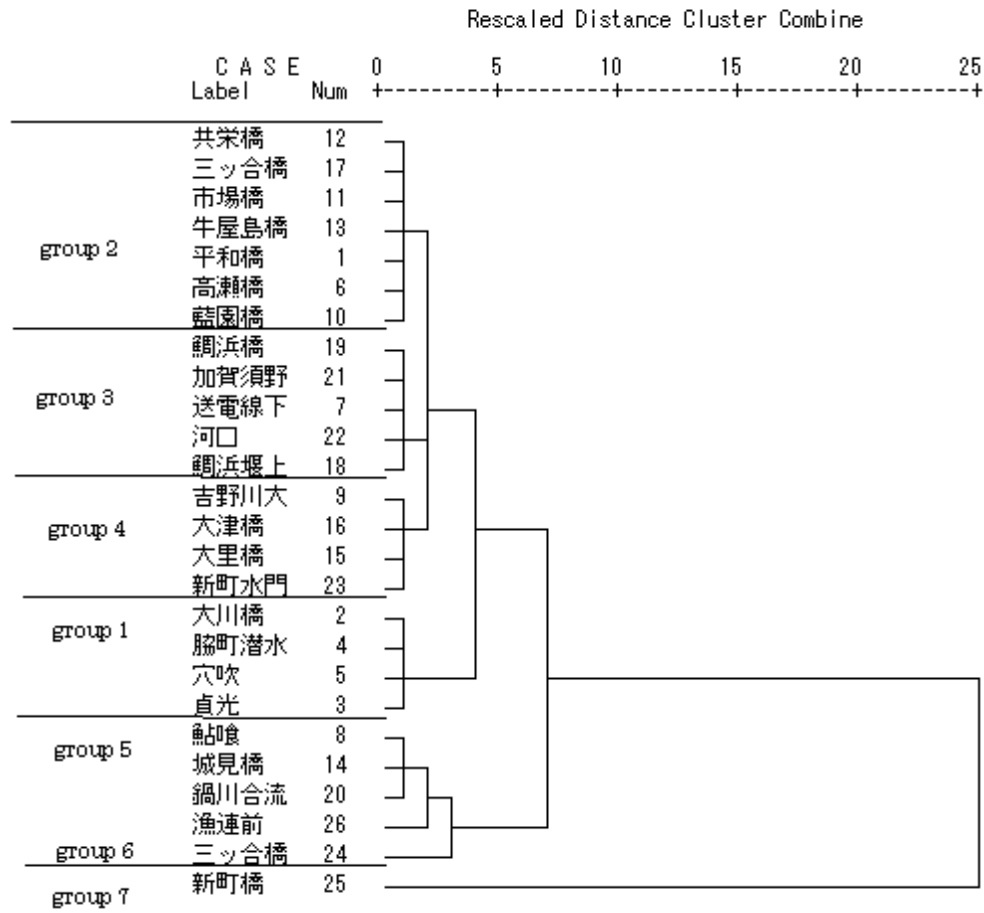


Figure 6. Dendrogram of the Yoshinogawa, 5 parameters, 26 monitoring points

グループ 1 は大川橋 (2) から穴吹 (5) の 4 測定地点である。これは、18 地点 3 パラメータで分析したときと同じである。グループ 2 は平和橋 (1) と高瀬橋 (6)、旧吉野川上流の 4 測定地点と今切川の三ッ合橋 (17) の合計 7 測定地点である。グループ 3 は、今切川のうち三ッ合橋 (17) と鍋川合流点 (20) を除いた 4 測定地点と、吉野川下流の送電線下 (7) の合計 5 測定地点、グループ 4 は大里橋 (15)、大津橋 (16)、新町水門 (23)、吉野川大橋 (9) で合計 4 測定地点である。グループ 5 は、城見橋 (14)、鍋川合流点 (20)、鮎喰 (8)、漁連前 (26) の 4 測定地点である。グループ 6 は新町川の三ッ合橋 (24)、グループ 7 は新町川の新町橋 (25) である。

26 地点 3 パラメータの結果と比べると、第一番目のグループが 2 つに、第二番目のグループが 4 つに、第 3 番目のグループが 2 つに分解された。より精密な分類が可能である。26 地点 3 パラメータを用いたクラスタ分析の結果と 26 地点 5 パラメータを用いたクラスタ分析の結果を比較すると、平和橋 (1) が異なる

グループに属することがわかる。これは平和橋 (1) の TN の値が大きいことによって生じたと考えられる。これについては鮎喰 (8) や城見橋 (14) についても同様で、TN の値が分析に加わったことでクラスタ分析の結果が異なった。また、新町川については、新町水門 (23) から新町川三ッ合橋 (24) の間と、新町川三ッ合橋 (24) と新町橋 (25) の間には何らかの汚染源があり、新町橋 (25) は吉野川水系の中で最も汚染されていることがわかる。一方、新町橋 (25) の下流にある漁連前 (26) は、新町川より清浄なグループに分類されているので、その間には河川の清浄化を強力に進める何らかの要因がある。

Table 10 に 26 地点 5 パラメータを用いた主成分分析の第一から第五主成分の固有値と寄与率および累積寄与率を示す。26 地点 5 パラメータを用いた主成分分析の寄与率は、第一主成分と第二主成分でほぼ 95% となっている。この傾向は 26 地点 3 パラメータおよび 18 地点 5 パラメータの結果と一致している。

Table 10. Eigenvalues and contribution (%) of principal components (the Yoshinogawa. 5 parameters, 26 monitoring points)

	eigenvalue	contribution rate(%)	cumulative contribution rate(%)
fac.1	4.26	85.22	85.22
fac.2	0.47	9.38	94.59
fac.3	0.14	2.77	97.36
fac.4	0.09	1.87	99.24
fac.5	0.04	0.76	100.00

fac1: the first principal component  
 fac2: the second principal component  
 fac3: the third principal component  
 fac4: the fourth principal component  
 fac5: the fifth principal component

Table 11 に 26 地点 5 パラメータを用いた主成分分析の主成分負荷量を示す。第一主成分の係数は BOD が 0.94, COD が 0.97, DO が -0.90, TN が 0.83, TP が 0.97 であり, 全ての項目が等価である。第二主成分の係数は TN が 0.53, DO が 0.35 で大きな重みを持つ。

18 地点での結果と同様, 第一主成分は 5 つのパラメータの総合で, 第二主成分は TN と DO である。18 地点での結果と同様, TN と TP も汚染度をみる指標として重要であることがわかる。

26 地点 5 パラメータを用いた主成分分析の第一主成分の主成分得点の高い順に並べたものとクラスター分析によるグループを示したものを Table 12 に示す。主成分分析による第一主成分の主成分得点は, クラスター分析によるグループの結果とは適応していないことがわかる。結合されたクラスター間のユークリッド距離の 2 乗が 3.0 以上になったもので分類すると, グループ 2 からグループ 4 は 1 つのグループにまとまる。そうすると, 主成分分析の第一主成分得点が -0.8 以上, 0.4 以下である測定地点が 1 つにまとまることになるので, 全体では 26 測定地点は 5 つのグループに分類することができる。この 5 つのグループにすれば, 第一主成分得点とクラスター分析によるグループは適応する。26 地点 5 パラメータを用い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 K-L Plot した結果をクラスター分析

Table 11. Coefficients in the first and second principal component (the Yoshinogawa, 5 parameters, 26 monitoring points)

	fac.1	fac.2	fac.3
BOD	0.94	0.16	0.24
COD	0.97	-0.15	0.12
DO	-0.90	0.35	0.20
TN	0.83	0.53	-0.17
TP	0.97	-0.13	-0.02

fac1: the first principal component  
 fac2: the second principal component  
 fac3: the third principal component

で得られたグループと対応させて Figure 7 に示す。7 つに分けられたグループは, 第一主成分得点だけではグループ 1,5,6,7 は分類可能であるが, グループ 2,3,4 はうまく分類できないことを示している。

26 地点 5 パラメータを用いた主成分分析により得られた各測定地点の第一主成分得点と河口からの距離の相関を Figure 8 に示す。回帰直線とその相関係数は, それぞれ  $y = -0.02x + 0.40$ ,  $R^2 = 0.31$  である。第一主成分得点と河口からの距離の分布に強い相関はみられない。河口からの距離 20km までに測定地点が集中しているが, この付近の第一主成分得点の分布を詳細にみると, 回帰曲線より上部に位置する測定地点は, 新町橋 (25), 三ッ合橋 (24), 鍋川合流点 (20), 漁連前 (26), 鮎喰 (8), 城見橋 (14) であり, 新町川, 今切川, 撫養川流域の汚染が進んでいることがわかる。吉野川は 4 つの河口を持つが, 河川の汚染状況を示す第一主成分得点の値と河口からの距離とは対応しておらず, 吉野川に流れ込む支流のうち, 特に最も河口からの距離が長い平和橋 (1) と河口の 1 つである新町川の水質汚染が進んでいることがわかった。それぞれの流域の水質に影響を与えている個々の要因を探る必要があることが示唆された。より詳細にみると, 新町川, 今切川, 撫養川の汚染が甚だしく, それらの汚染要因を除去することが吉野川全体の浄化につながることを示唆された。

Table 12. Score of factor 1, groups in cluster analysis and distances (the Yoshinogawa, 5 parameters, 26 monitoring points)

Fig.1 の番号	水域名	調査地点名	主成分分析による 第1因子の主成分 得点	クラスター 分析による グループ	河口から 距離	各水域河口か らの順位
5	穴吹川	穴吹	-1.460	1	45.0	吉野川河口5
2	吉野川上流	大川橋	-1.348	1	87.0	吉野川河口8
3	貞光川	貞光	-1.222	1	55.0	吉野川河口7
4	吉野川下流	脇町潜水橋	-1.155	1	48.0	吉野川河口6
6	吉野川下流	高瀬橋	-0.756	2	19.0	吉野川河口4
10	旧吉野川上流	藍園橋	-0.713	2	18.5	旧吉野川河口7
1	銅山川	平和橋	-0.524	2	94.0	吉野川河口9
11	旧吉野川上流	市場橋	-0.408	2	10.7	旧吉野川河口6
9	吉野川下流	吉野川大橋	-0.363	4	3.5	吉野川河口1
13	旧吉野川上流	牛屋島橋	-0.323	2	7.0	旧吉野川河口4
12	旧吉野川上流	共栄橋	-0.323	2	8.5	旧吉野川河口5
16	旧吉野川下流	大津橋	-0.320	4	2.3	旧吉野川河口1
17	今切川上流	三ッ合橋	-0.246	2	10.7	今切川河口6
19	今切川下流	鯛浜橋	-0.109	3	7.5	今切川河口4
18	今切川上流	鯛浜堰上流側	0.022	3	7.7	今切川河口5
23	新町川上流	新町水門	0.111	4	7.3	新町川河口4
21	今切川下流	加賀須野橋	0.118	3	4.0	今切川河口2
7	吉野川下流	送電線下	0.126	3	10.5	吉野川河口3
15	撫養川	大里橋	0.229	4	4.5	旧吉野川河口2
22	今切川下流	河口	0.361	3	0.0	今切川河口1
14	撫養川	城見橋	0.618	5	6.5	旧吉野川河口3
8	鮎喰川	鮎喰	0.675	5	6.5	吉野川河口2
26	新町川下流	漁連前	0.983	5	1.7	新町川河口1
20	今切川下流	鍋川合流点	1.105	5	5.0	今切川河口3
24	新町川上流	三ッ合橋	1.817	6	5.6	新町川河口3
25	新町川上流	新町橋	3.104	7	4.2	新町川河口2

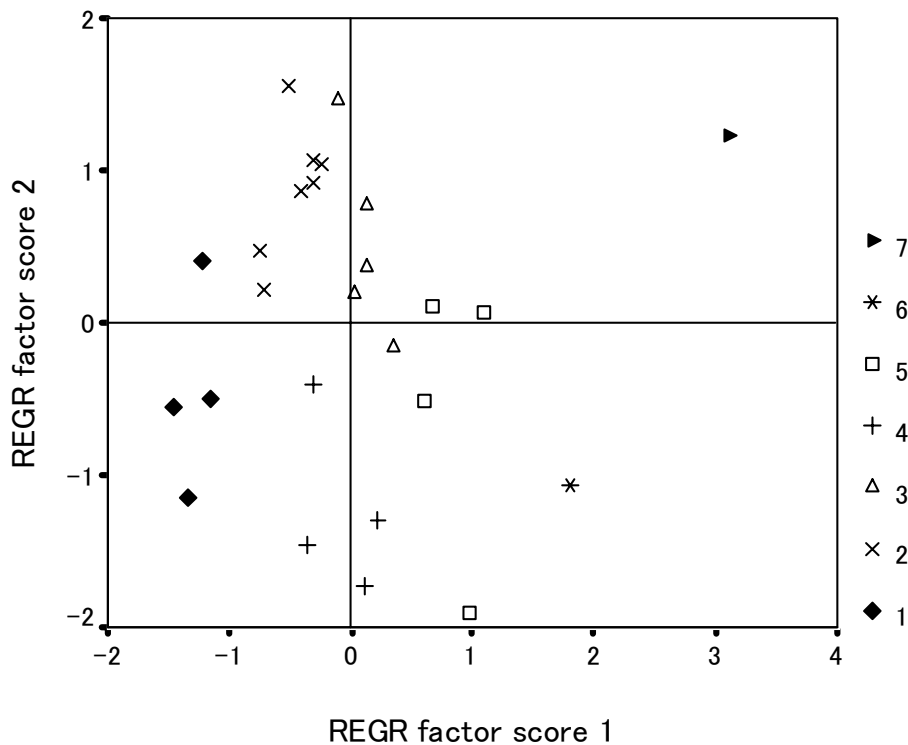


Figure 7. K-L plots by factors 1 and 2 (5 parameters, 26 monitoring points)

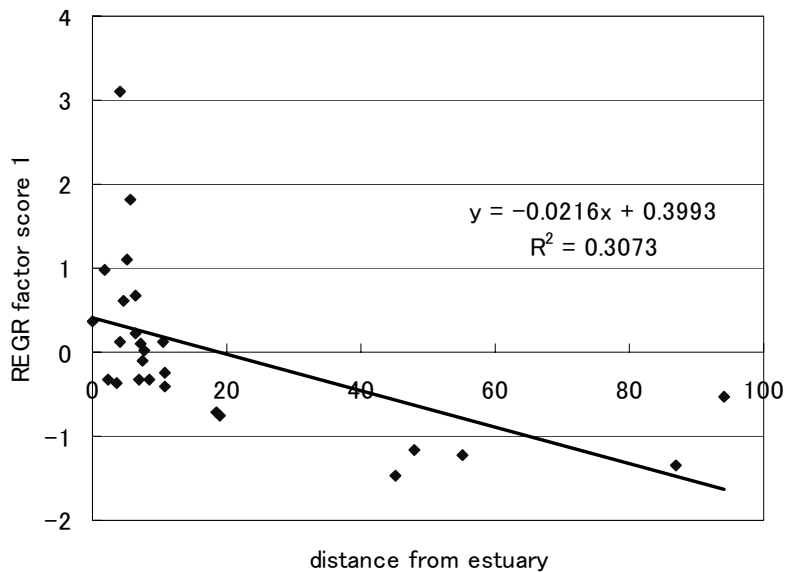


Figure 8. Relationship between distances and score of factor 1 (5 parameters, 26 monitoring points)

平成 16 年度版徳島県環境白書 [1] によると「新町川は，昭和 40 年代前半には新町橋で BOD が 30mg/l を超える汚濁がみられましたが，法令等による排水規制，下水道事業の推進，市民の清掃活動等により，現在は 3mg/l 以下にまで水質が改善されました。」とあるように，近年，新町川の水質は良くなっているとの報告がある [12]．しかし，吉野川全体の中では依然として他の測定地点とは異なった傾向を示し，さらに浄化が必要な測定地点である．今後，上記の政策施行など水質データ以外のものがどう水質データに影響を与えたかについてと，経年経過の観点からもみていく必要がある．

## 6 2002 年度吉野川と多摩川の水質データの比較

先の東京多摩川の水質データの分析結果 (1994 ~ 2002) では各年度の 2 月のデータを用いて分析を行い，多摩川の清浄度を示す主なパラメータとして DO と TP が抽出された [13]．そこで吉野川と同様の分析を 2002 年度の多摩川の水質データについても行い，比較検討する．多摩川のデータも吉野川と同様，国立環境研究所が提供する「国立環境研究所環境データベース」公共用水域水質年間値データファイル [3] を利用した．5 種類の水質データを Table 13 に示す．

Figure 9 に多摩川のクラスター分析のデンドログラ

ムを示す．ここでクラスター間のユークリッド距離の 2 乗が 2.0 以上になったもので分類すると，15 測定地点は 5 つのグループに分類することができた．グループ 1 は河口からの距離が長い順に和田橋 (1) から拝島橋 (6) の 6 測定地点である．グループ 2 は，日野橋 (7) と多摩川原橋 (10) を除いた中流の 5 測定地点である．グループ 3 は日野橋 (7) のみである．グループ 4 は下流の 2 測定地点である．グループ 5 は多摩川原橋 (10) のみである．多摩川原橋 (10) は特に汚染が進んでおり，他の中・下流の測定地点とは異なるグループに分類された．

Table 14 に多摩川の主成分分析の第一から第五主成分の固有値と寄与率および累積寄与率を示す．主成分分析の寄与率は，第一主成分 (89.9%) と第二主成分 (5.2%) で 95 % を超えている．吉野川と同じように第一主成分が重要である．ただ第二主成分と第三主成分 (4.5%) が擬似縮退しており，第二主成分とあわせて第三主成分も重要であるところが吉野川の結果と大きく異なる．第三主成分までを考慮することとすると寄与率は 99% を超える．

Table 15 に多摩川の主成分分析の主成分負荷量を示す．第一主成分は全ての項目がほぼ同じ重みを持ち，DO の符号が負である．このことは吉野川と同じ結果である．第二主成分の係数は DO と TP および BOD が大きな寄与をしており，吉野川の結果 (TN と DO) と異なる．

Table 13. Values of chemical index of water and distances in the Tamagawa

Fig.9の 番号	調査地点名	BOD	COD	DO	T-N	T-P	河口からの 距離
1	和田橋	0.50	1.21	10.99	0.75	0.015	61.8
2	調布橋	0.85	2.00	10.48	0.83	0.018	59.8
3	羽村堰	0.50	1.20	11.12	0.78	0.015	53.2
4	永田橋	0.77	1.80	10.64	0.93	0.018	51.7
5	拝島原水補給点	0.53	1.20	11.03	1.50	0.016	47.8
6	拝島橋	0.84	1.80	10.20	1.50	0.024	46.2
7	日野橋	2.05	4.30	9.03	4.70	0.380	39.8
8	関戸橋	1.42	3.70	9.12	5.20	0.400	34.6
9	是政橋	1.59	4.30	9.28	5.80	0.470	31.5
10	多摩川原橋	2.35	5.60	8.46	6.70	0.610	27.9
11	多摩水道橋	1.32	4.40	8.67	5.90	0.430	23.2
12	第三京浜多摩川橋	1.14	3.50	8.91	5.99	0.360	16.5
13	田園調布堰上	1.02	3.99	9.22	5.67	0.380	13.1
14	六郷橋	1.46	4.40	8.04	5.00	0.310	5.5
15	大師橋	1.39	4.40	7.49	4.20	0.260	2.4

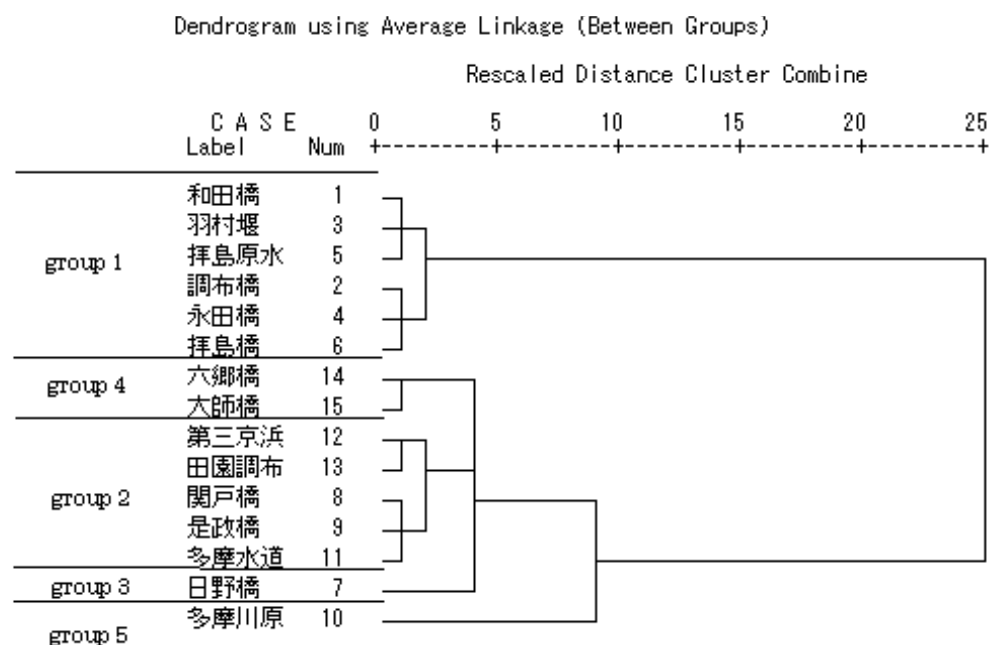


Figure 9. Dendrogram of the Tamagawa, 5 parameters, 16 monitoring points

主成分分析の第一主成分の主成分得点の高い順に並べたものとクラスター分析によるグループを示したものを Table 16 に示す。

主成分分析による第一主成分の主成分得点は、クラスター分析によるグループの結果とは適応していない。グループ 2 に属する観測地点の間にグループ 4 が割り込んでいる。とはいえ、吉野川の複雑さと比べると、主成分得点の順とクラスター分析のグループとの対応は単純である。

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 K-L Plot をクラスター分析で得られたグループ番号で表示したものを Figure 10 に示す。第一主成分と第二主成分を用いて 15 地点を K-L プロットすると、多摩川の場合、第一主成分によって上流と中・下流が分類可能であり、第二主成分によって上・中流と下流が分類可能であった。吉野川の K-L プロット ( Figures 3, 5, 7 ) と比較すると第一主成分による多摩川の上流と中・下流域の分離の良さが良くわかる。主成分分析により得られた各測定地点の第一主成分得点と河口からの距離の関係を Figure

11 に示す．回帰直線とその相関係数は，それぞれ  $y = -0.04x + 1.37$ ， $R^2 = 0.59$  である．第一主成分得点と河口からの距離の分布に強い相関はみられないが，最上流 6 地点の水質は清浄である．

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．一方，吉野川は 4 つの河口を持っているため，多摩川ほど明確な上流から下流にかけての分類はできなかった．それぞれの河口毎の浄化に向けての対策が必要である．

Table 14. Eigenvalues and contribution (%) of principal components (the Tamagawa, 5 parameters, 15 monitoring points)

	eigenvalue	contribution rate(%)	cumulative contribution rate(%)
fac.1	4.49	89.88	89.88
fac.2	0.26	5.22	95.11
fac.3	0.22	4.47	99.58
fac.4	0.02	0.33	99.91
fac.5	0.00	0.09	100.00

fac1: the first principal component  
 fac2: the second principal component  
 fac3: the third principal component  
 fac4: the fourth principal component  
 fac5: the fifth principal component

Table 15. Coefficients in the first and second principal component (the Tamagawa, 5 parameters 15 monitoring points)

	fac.1	fac.2	fac.3
BOD	0.91	0.21	-0.34
COD	0.99	-0.04	-0.04
DO	-0.91	0.42	0.07
TN	0.96	0.04	0.28
TP	0.97	0.20	0.16

fac1: the first principal component  
 fac2: the second principal component  
 fac3: the third principal component

Table 16. Score of factor 1 and groups in cluster analysis (the Tamagawa, 5 parameters, 15 monitoring points)

Fig.9 の番号	調査地点名	主成分分析による第1因子の主成分得点	クラスター分析によるグループ
3	羽村堰	-1.336	1
1	和田橋	-1.314	1
5	拝島原水補給点	-1.242	1
4	永田橋	-1.046	1
2	調布橋	-0.968	1
6	拝島橋	-0.885	1
13	田園調布堰上	0.429	2
12	第三京浜多摩川橋	0.463	2
8	関戸橋	0.529	2
15	大師橋	0.669	4
14	六郷橋	0.725	4
11	多摩水道橋	0.770	2
9	是政橋	0.780	2
7	日野橋	0.800	3
10	多摩川原橋	1.625	5

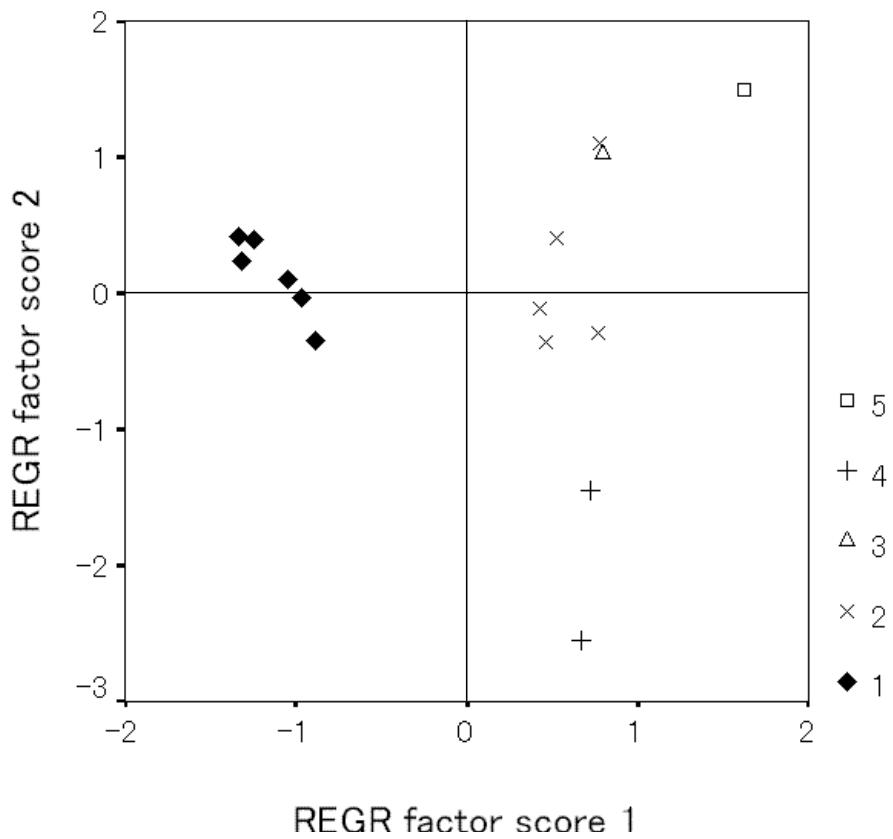


Figure 10. K-L plots by factors 1 and 2 of the Tamagawa, 5 parameters, 15 monitoring points

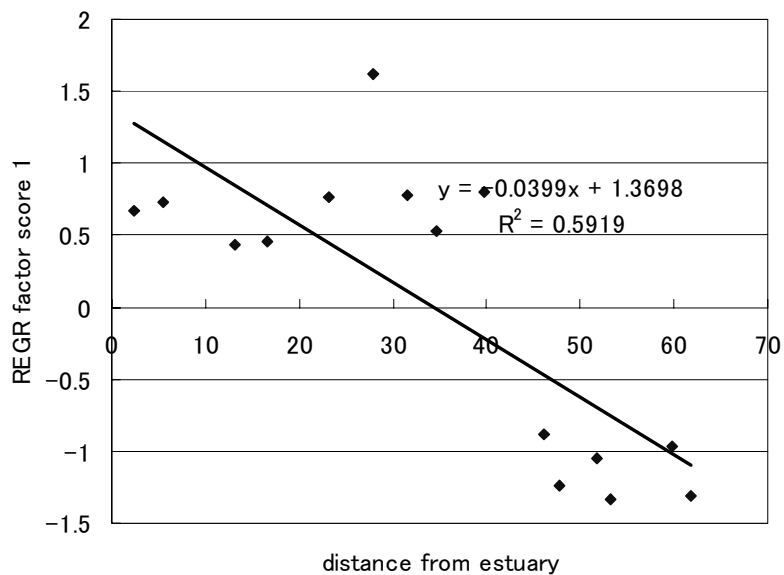


Figure 11. Relationship between distances and score of factor 1 (the Tamagawa, 5 parameters, 15 monitoring points)

## 7 まとめ

吉野川の上流から河口に至るまでの5種類の水質パラメータの変化を主成分分析とクラスター分析を用いて解析し、吉野川の上流から下流までの汚染度をみるための複合的な指標を得ることを試みた。またそれらを用いて、吉野川の水質を解析した。

5種類のパラメータの測定データには、欠測があり、それらを予測して解析を行った。その方法の妥当性を調べるために、まず、全てのデータが揃っている3つのパラメータ(BOD, COD, DO)26測定地点のデータを用い、クラスター分析と主成分分析を行った。クラスター分析により、26測定地点は4つのグループに分類された。

主成分分析の第一主成分では、3つのパラメータが同じ程度に重要であった。第一主成分の主成分得点は水質の清浄度を表現する成分であるが、その値は河口からの距離とは対応していなかった。それぞれの流域の水質に影響を与える原因が個々にあることが示唆された。

次にTPとTNのデータを加えた5つのパラメータのある18測定地点のデータを用いて、同様の解析を行った。クラスター分析により18測定地点は6つのグループに分類された。3パラメータのときと異なり、最も河口からの距離が長い平和橋(1)は、水質汚染が進んでいるグループに分類された。主成分分析の第一主成分の5つのパラメータはほぼ同じ重みであった。第二主成分はTNとDOが大きな重みを持っていた。これらから3つのパラメータ(BOD, COD, DO)だけでなく、TNとTPも汚染度をみる指標として重要であることがわかった。

解析のパラメータを5つに増やしたことによって、より吉野川の水質の特徴が明確になったが、欠測のある測定地点を除いたことによって、吉野川全体を知ることが困難になってしまった。そこで、欠測値を補完し、測定地点を増やしてさらに水質変化を検討した。

最後にTPとTNの値が欠測となっている地点のデータを線形関数によって予測した後、同様の解析を行った。クラスター分析により、26測定地点は7つのグループに分類された。

欠測を補完することで得られた情報を捨てることなく、吉野川全体の状況を知ることが可能となった。今後TNとTPの欠測が続くことが惜まれる。

吉野川と同様の分析を2002年度の多摩川の15地点5パラメータのデータについても行い、比較検討した。

クラスター分析により、多摩川は、5グループに分類できた。15測定地点は上流(1~6)と中・下流(7~15)に分類可能であった。

主成分分析の第一主成分は吉野川の第一主成分と同じく、5つの水質パラメータすべてが重要であった。第二主成分は主にDOが大きな重みをもっていた。この結果は吉野川の結果(TNとDO)とは異なっていた。

多摩川の場合、上流は清浄であるが、中・下流域の汚染の度合は河口からの距離によらなかった。吉野川の場合も河川の汚染の度合は、河口からの距離とは対応しておらず、それぞれの流域の水質に影響を与えている個々の要因を探る必要があることが示唆された。

河川の総合的な水質調査のためには、第一主成分のような個々の環境指標の線形結合であらわされる指標を用いた解析が重要であることがわかった。

いつも有益な御議論をしていただいている、徳島大学の中馬寛教授に深く感謝する。

## 参考文献

- [1] 徳島県 環境白書 平成16年度版 第2部2章2節  
<http://ourtokushima.net/kankyo/hakusho16.php>
- [2] SPSS ver. 11.0, SPSS Japan Inc..  
<http://www.spss.co.jp/>
- [3] 国立環境研究所, 国立環境研究所環境データベース, 公共用水域水質年間値データファイル  
<http://www-gis.nies.go.jp/>
- [4] Aoyama, T., Ichikawa, H., *Chem. Pharm. Bull.*, **39**, 372 (1991).
- [5] Aoyama, T., Wang, Q., Zhu, H., Nagashima, U., *IPSI-SIG notes*, **2000-HPC-84**, 7 (2000).
- [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).
- [12] 徳島市, “ 主な河川の水質 ”  
[http://www.city.tokushima.tokushima.jp/kankyo\\_hozen/gaiyo26.html](http://www.city.tokushima.tokushima.jp/kankyo_hozen/gaiyo26.html)
- [13] 神部順子, 袁えん, 長嶋雲兵, 青山智夫, *J. Comp. Chem. Jpn.*, **5**, 201 (2006).

## Analysis of the Water Quality of the Yoshinogawa, Tokushima, Japan, by Using Multivariate Analysis

Junko KAMBE<sup>a</sup>, Tomoo AOYAMA<sup>b</sup>, Aiko YAMAUCHI<sup>c</sup> and Umpei NAGASHIMA<sup>d,e,\*</sup>

<sup>a</sup>Faculty of Foreign Language, Daito Bunka University  
 1-9-1 Takashimadaita, Itabashi, Tokyo 175-8571, Japan

<sup>b</sup>Faculty of Engineering, Miyazaki University  
 Gakuen Kihanadai Nishi, Miyazaki, 889-2192, Japan

<sup>c</sup>Graduate School of Pharmaceutical Sciences, University of Tokushima  
 1-78 Sho-machi, Tokushima 770-8505, Japan

<sup>d</sup>Research Institute for Computational Science, National Institute of Advanced Industrial Science and Technology  
 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

<sup>e</sup>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](mailto:u-nagashima@aist.go.jp)

Principal component analysis, cluster analysis and neural network were applied to assess the pollution levels along the Yoshinogawa River in Tokushima, Japan, by using water quality data (Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and Total Phosphorus (TP)) measured in 2002. Because there are a few defects in the data, we estimated them by linear equation and neural network to avoid information loss.

In the Yoshinogawa River, there are four branching bays (Figure 1). Though it is usually difficult to find the relationships between the water quality of the Yoshinogawa River and the distances from their estuaries, the first principal components consisting of five parameters explained the nature of the river well.

The water quality at Heiwabashi of the Shinmachigawa River was the most contaminated in the Yoshinogawa River (Figure 6, Table 12). This suggests that there is a source of water pollution in the upper region of Heiwabashi and that there is a source to improve water quality in the lower region of Heiwabashi.

**Keywords:** Water Pollution, Yoshinogawa River, Chemometrics, Principal Component Analysis, Cluster Analysis