

## 다변량분석법을 이용한 금강 유역의 수질오염특성 연구

김미아<sup>†</sup> · 이재관 · 조경덕\*

국립환경과학원 환경진단연구부

\*서울대학교 보건대학원 환경보건학과

### Evaluation of the Geum River by Multivariate Analysis: Principal Component Analysis and Factor Analysis

Mi-Ah Kim<sup>†</sup> · Jae-kwan Lee · kyung-Duk Zoh\*

Department of Environmental Diagnostics Research, National Institute of Environmental Research

\*Department of Environmental Health, School of Public Health, Seoul National University

(Received 19 October 2006, Accepted 28 November 2006)

#### Abstract

The main aim of this work is focus on the Geum river water quality evaluation of pollution data obtained by monitoring measurement during the period 2001-2005. The complex data matrix 19 (entire monitoring stations)\*13 (parameters), 60 (month)\*13 (parameters) and 20 (season)\*13 (parameters) were treated with different multivariate techniques such as factor analysis/principal component analysis (FA/PCA). FA/PCA identified two factor (19\*13) classified pollutant Loading factor (BOD, COD, pH, Cond, T-N, T-P, NH<sub>3</sub>-N, NO<sub>3</sub>-N, PO<sub>4</sub>-P, Chl-a), seasonal factor (water temp, SS) and three Factor (60\*13, 20\*13) classified pollutant Loading factor (BOD, COD, Cond, T-N, T-P, NH<sub>3</sub>-N, NO<sub>3</sub>-N, PO<sub>4</sub>-P), seasonal factor (water temp, SS) and metabolic factor (Chl-a, pH). Loadings of pollutant factor is potent influence main factor in the Geum river which is explained by loadings of pollutant factor at whole sampling stations (71.16%), month (52.75%) and season (56.57%) of main water quality stations. Result of this study is that pollutant loading factor is affected at Gongju 1, 2, Buyeo 1, 2, Gangkyeong, Yeongi stations by entire stations and entire month (Gongju 1, Cheongwon stations), April, May, July and August (buyeo 1) by month. Also the pollutant Loading factor is season gives an influence in winter (Gongju 1, buyeo 1) from main sampling stations, but Cheongwon characteristic is non-seasonal influenced. This study presents necessity and usefulness of multivariate statistic techniques for evaluation and interpretation of large complex data set with a view to get better information data effective management of water sources.

**keywords** : Factor analysis (FA), Geum river, Principal component analysis (PCA)

## 1. 서 론

하천수질은 유역 내 각종 오염원이 지천을 통해 유입되어 하천의 수리·수문학적 특성과 기후, 수체 내 생물학적인 영향으로 측정지점과 시간의 흐름에 따라 변동하는 경향이 있어 수질에 영향을 미치는 인자는 매우 다양하며 이러한 영향으로 도출된 오염지표 또한 여러 종류이다. 그러나 현재까지 선행되어온 대부분의 연구에서는 이러한 복잡하고 다양한 정보에 대하여 주관적인 오염평가와 산술적인 수질해석으로 인하여 평이한 유역의 오염 상태만을 설명하여 왔으며, 우리나라 유역관리에 있어 하천의 일반적인 오염원이 대부분 유기오염원일 것이라는 판단은 과거와 현재에도 계속되고 있다. 그러나 현실적으로 여러 수질 오염원이 존재하고 있으며, 또한 이 논리를 정확하게 뒷받침하는

근거자료로 제시할 수 있는 방법론이 있음에도 이를 활용하지 못하고 연구자의 주관과 경험에 의지하여 해석하고 일반화하는 경우가 대부분이었다. 이에 본 논문은 이점을 극복하여 금강유역의 주요오염원을 밝힘으로써 과거의 주관적이고 불확실한 수질자료의 해석에 대한 오류를 줄이고자 한다. 이에 다 변수 혹은 다 개체에 관하여 정보의 손실 없이 현상을 단순하고 명료하게 설명하여, 주된 오염영향인자를 쉽게 산출함으로써 합리적이고 신뢰성 있는 수질해석 자료를 지원하기 위하여 다변량 분석방법을 이용하여 과학적 수질관리의 기초를 마련하고자 한다.

다변량 분석법은 크게 주성분분석, 요인분석, 군집분석, 정준상관분석 등 여러 방법이 있다. 현재까지, 국내뿐만 아니라 국외에서도 여러 차례 위 분석방법을 이용하여 수질 자료의 평가를 실시해왔다. 그 예로 군집분석과 인자분석을 이용하여 금강수질을 평가하였고(김, 2002) 주성분분석에 의하여 도시호수의 수질평가(김, 2003)와 신 등(1998)은 낙동강하구해역의 수질을 평가하였다. 또한, 요인분석을 통하

<sup>†</sup> To whom correspondence should be addressed.  
miah00@me.go.kr

여 낙동강 수질을 평가하였으며(박 등, 2001), 유 (2002)는 한강수질에 영향을 미치는 요인들을 분석하기도 하였다. 외국논문을 살펴보면 New Jersey Passic 강에 대하여 주성분 분석을 이용하여 공간적, 시간적 변화를 분석하였으며(Karim et al., 2003), 주성분분석과 요인분석을 통해 지표 수 수질측정지점의 수질 평가에 이용하였으며(Ying, 2005), 호소의 영양상태를 파악하기 위하여 Bernard 등(2004)은 관련 수질 항목들의 평가와 관리를 위하여 주성분분석을 이용하였다. Petersen 등(2001)은 Elbe 강의 시간에 따른 영양물질농도와 관련된 수질에 영향을 미치는 요인사이의 공통성을 파악하기 위하여 주성분분석을 이용하는 등 많은 연구가 국·내외적으로 수행되어졌다.

금강은 타 하천에 비하여 오염원이 적은 수계로써 수량과 수질 면에서 안정적인 편이었으나, 지형적으로 대청댐 이하는 경사가 낮아 유속이 지체되면서 수질이 저하되고 있으며 금강으로 유입되는 지천은 50여 개에 이르고, 금강 수계에 직접적인 영향을 주고 있는 대전시 하수가 유입되는 갑천과 청주시에서 배출되는 도시하수가 미호천을 통하여 유입되고 있으며(임, 1999), 공주, 부여, 강경과 같은 중소도시에서 배출되는 생활하수와 주변 공단에서 배출되는 폐수의 영향을 직접 받고 있고(김, 2002), BOD농도는 3등급을 나타내어 다른 유입 하천에 비하여 오염상태가 심하고(환경부, 2001), 최근 신형정수도 건설계획으로 인하여 예상되는 개발의 영향으로 용수수요 증가와 오염부하량의 증가로 인하여 보다 과학적이고 합리적인 수질관리의 필요성이 요구되고 있다.

## 2. 연구내용 및 방법

### 2.1. 연구대상 하천의 개요

금강유역은 동경 126° 40' 25"~128° 03' 53", 북위 35° 34' 47"~37° 03' 03" 사이인 우리나라 중앙부 서쪽에 위치하며 충청남·북도의 약 절반과 전라북도의 대략 1/4정도를 차지하고 경상북도 일부와 경기도 극히 일부를 포함하고 있는 우리나라 제3의 유역으로서, 그 유역면적은 9,912.15 km<sup>2</sup> 정도이고, 유로연장은 397.79 km이다.

Fig. 1에 금강유역의 수질측정지점을 상·하류로 구분하여 나타내었다. 유역의 북쪽으로는 우리나라에서 가장 큰 한강유역과 접해 있고, 동쪽은 소백산맥을 경계로 낙동강유역과 접하며, 남쪽은 섬진강 및 만경강 유역이, 서쪽으로는 서해와 접해 있다. 또한 유역의 북서쪽에는 안성천·삼교천 유역과 접하고 있으며, 금강유역의 동·서장은 약 120 km 정도이고 남·북장은 약 160 km이다.

금강은 침식계곡을 따라서 발달하여 있으며, 풍화와 침식 작용이 활발하여 낮은 경사의 지구릉지가 발달하였고 분지형 지형에서 나타날 수 있는 수지상 수계가 발달하여 있다. 금강의 상류는 대부분 변성암으로 구성되어 있으며 중류와 하류는 중생대 화강암류가 주로 분포되어 있다. 본류 중류부에는 토사의 유출이 많아 천정천 현상을 보이고 있다. 금강 유역의 61.2%가 임야로 지목분류상 가장 큰 면적을



Fig. 1. Location of sampling stations in the Geum river.

차지하며, 시가지지역은 약 4.6%, 농경지 면적은 전체 유역 면적의 27.4%에 해당된다(김, 2006). 또한, 유역 상류부에는 용담댐, 중류부에는 대청댐 하구부에는 하구 독 등의 주요 시설물이 설치되어 유역 내 치수 및 이수, 하천환경 개선의 주요 기능을 담당하고 있다. 또한 2005년 기상관측소 월별 강수량측정결과 평·갈수기에 9.5~184.8 mm, 홍수기에는 202.5~408.5 mm를 나타내고 있어 장마가 시작되면서 유량의 급격한 증가를 알 수 있다(금강수계관리위원회, 2005). 월별 평균유량의 변화는 2월에서 5월까지 크게 변화가 없었고, 장마철인 7월에서 9월의 유량증가로 인하여 대청댐 하류의 유량이 급격히 증가하는 경향을 보이고 있다. 또한, 대청댐하류는 금강본류 유량의 60% 이상을 차지하기 때문에 유량변화에 지대한 영향을 미치고 있으며, 대청댐조정지의 방류량은 5월에는 크게 변화가 없으나 6월에 39.16 m<sup>3</sup>/s로 유입량의 증가와 장마에 대비하여 방류량을 증가시킨 것으로 판단되고 홍수기인 7~9월에 평균 183.69 m<sup>3</sup>/s를 기록하고 있다. 10월 이후는 약 29 m<sup>3</sup>/s로 평·갈수기 방류량을 나타내고 있다(금강수계관리위원회, 2005).

### 2.2. 주성분분석 및 인자분석

본 연구에서는 다변량분석법 중 주성분과 요인분석을 실행하여 유역의 오염상태를 파악하고자 한다. 다변량분석은 2개 이상의 변수들의 관계를 동시에 분석할 수 있는 모든 통계적인 기법을 말하는 것으로 이들 중 여러 개의 변수들이 내포된 다변량 자료를 분석하는 일은 매우 복잡하고 힘든 경우가 많다. 이러한 상황에서 원 변수들의 몇 개 일차결합을 통해 간단한 구조를 갖도록 자료를 축약하게 되면 공분산행렬을 최대한 설명하는 몇 개의 일차결합을 산출하게 되는데 이것을 주성분이라 부르며, 이러한 주성분을 구하고 해석하여 원 자료를 축약하는 방법을 주성분분석이라 한다. 주성분분석은 그 자체가 목적이라기보다는 일단 자료

를 간단히 하여 추후에 추가적인 다른 통계적 절차를 적용하기 위한 사전분석이라 볼 수 있다. 또한 자료에 대한 통계적 모형이나 어떤 특별한 가정을 필요로 하지 않으며 공분산 행렬의 고유값과 고유벡터에 의해 주성분이 결정된다는 점에서 수학적인 절차라고 볼 수 있다.

인자분석은 통계모형을 가정하고 공통인자의 추정을 위한 몇 가지 가정을 필요로 하며 인자 해석을 위한 인자회전 등의 절차를 수행한다는 점에서 주성분분석에 비해 정교한 통계적 방법이라고 볼 수 있다. 인자분석은 p(임의의 변수 수)차원 변수들간의 상호의존 구조를 나타내는 공분산이나 상관행렬에서 m개(단, m<p)의 '공통인자(commom factor)'를 추출하고 이들을 해석하여 원래 변수들이 나타내는 복잡한 구조를 쉽게 파악하기 위한 자료축약(data reduction)기법이다. 공통인자란 원 변수들의 저변에 내재된 잠재변수로서 관찰할 수 없는 인자를 뜻한다. 인자분석에 의해 추출된 공통인자들은 적당한 기준에 의해 회전시킴으로써 공통인자를 훨씬 더 쉽게 해석할 수 있게 해준다. 이와 같은 일련의 과정을 통해 원 변수들의 복잡한 차원을 저차원으로 축약하여 자료의 구조를 파악할 수 있게 해준다. 더 나아가 각 개체들에 대한 공통인자 점수를 계산하고 이를 이용하여 다른 통계기법에 활용할 수도 있다(최 등, 2003).

아래 식은 인자분석의 기본적인 모형으로 다음 식과 같이 나타내었다.

m-인자모형 (m-factor model) :  $m \ll p$

$$x = L f + \epsilon \tag{1}$$

$p \times 1 \quad p \times m \quad m \times 1 \quad p \times 1$

여기서  $x = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$ ,  $L = \begin{pmatrix} l_{11} \cdots l_{1m} \\ l_{21} \cdots l_{2m} \\ \vdots \\ l_{p1} \cdots l_{pm} \end{pmatrix}$ ,  $f = \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix}$ ,  $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix}$

변수  $F_1, \dots, F_m$ 은 공통인자(commom factor)를 나타내고 (평균 0, 분산 1이며 독립), 변수  $\epsilon_1, \dots, \epsilon_p$ 는 변수 고유의 특정인자(specific factor)를 나타낸다(평균 0, 분산  $\psi_1^2, \dots, \psi_p^2$ ,  $\psi_j^2 = \sigma_j^2 - (l_{j1}^2 + \dots + l_{jm}^2)$ 이며 독립). 그리고  $l_{11}, \dots, l_{1m}, l_{21}, \dots,$

$l_{2m}, \dots, l_{p1}, \dots, l_{pm}$ 은 각 변수와 인자를 연결해주는 파라미터 (=인자부하)이다(허 등, 2004).

### 3. 결과 및 고찰

#### 3.1. 대상유역의 수질특성 및 상관분석

금강수계의 수질분석은 5년간(2001~2005)의 수질자료를 바탕으로 금강수계 전체 수질측정망(19지점)의 연간 수질항목별 평균자료, 환경부에서 정한 수질측정망 주요지점인 공주1, 대청댐, 옥천, 청원 측정소에 대하여 월별평균(5개 주요지점의 5년간 1~12월 평균), 계절별 평균(5개 주요지점의 5년간 봄:3~5월/여름:6~8월/가을:9~11월/겨울:12~2월 평균) 자료로 나누어 총 3가지 주제로 분석을 각각 실시하였다. 수질자료는 국립환경과학원의 물환경정보시스템 (환경부, 2006)을 이용하였으며, 분석 항목은 수온, pH, DO, SS, BOD, COD, TN, TP, PO4-P, Nitrate, Ammonia nitrogen, Cond, Chlorophyll-a로 총 13개이다.

전 측정소에 대한 정규성 검정(Shapiro-wilk)결과 정규분포하지 않으므로 변수변환을 시도하여 재 검정을 실시하였으나 같은 결과를 나타내었다. 대부분 다변량 분석은 모수적 데이터에 대하여 이용하는 것으로 알려져 있으나, 실제로 크게 두 가지 방법으로 나타낼 수 있다. 첫 번째는 정규분포에 대해 자료를 분석하는 방법이며, 두 번째 방법은 비정규분포에 대해서도 분석이 가능한 데이터를 변환시키지 않고 있는 그대로 두고 탐색적으로 정보를 도출하는 방법이다(허 등, 2004). 또한, Bernard Parient 등(2004)은 주성분분석에 대해서 데이터의 어떤 가정도 필요하지 않는 비모수 방법이라고 나타내었으며, 덧붙여 주성분 분석은 흔히 통계적 모형에서 포함되는 오차와는 관계가 없기 때문에 관측치들이 다변량 정규분포를 하는 경우에 주성분은 보다 의미 있겠으나, 이러한 정규성의 가정이 반드시 요구되는 것은 아니다(한, 1997). 그러므로 본 연구에서는 자료의 변환 없이 원 자료를 그대로 표준화 시킨 후 분석하였다.

금강 수계의 수질을 전 측정소에 대하여 Table 1에 상관분석 결과를 나타내었으며, Fig. 2에서 TN, TP, PO4-P, NH3-N, NO3-N, COD, BOD에 대하여 2001년에서 2005년까지의 평균을 각 측정소별로 상류에서 하류 방향으로 나타내었다.

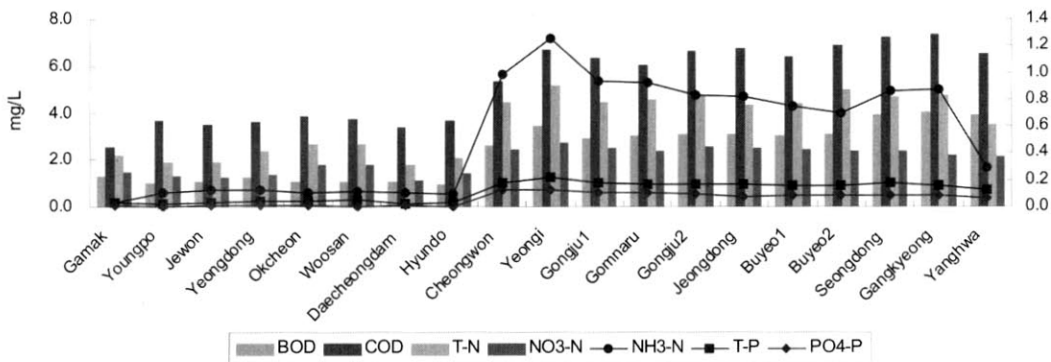


Fig. 2. Comparison of water quality mean values of each sampling stations from 2001 to 2005.

**Table 1.** PCA general approach: correlation matrix of total sampling stations (a), main sampling stations by month (b) and season (c)

	Variable	NH <sub>3</sub> -N	NO <sub>3</sub> -N	pH	COD	Cond	BOD	SS	DO	T-N	T-P	Water temp	Chl-a	PO <sub>4</sub> -P
(a)	NH <sub>3</sub> -N	1.000												
	NO <sub>3</sub> -N	0.916**	1.000											
	pH	0.120	0.107	1.000										
	COD	0.861**	0.892**	0.405	1.000									
	Cond	0.917**	0.962**	0.247	0.950**	1.000								
	BOD	0.819**	0.860**	0.395	0.958**	0.946**	1.000							
	SS	0.838**	0.874**	0.218	0.893**	0.865**	0.817**	1.000						
	DO	-0.490*	-0.310	-0.058	-0.248	-0.211	-0.031	-0.335	1.000					
	T-N	0.946**	0.968**	0.193	0.935**	0.963**	0.897**	0.919**	-0.355	1.000				
	T-P	0.972**	0.971**	0.176	0.921**	0.977**	0.902**	0.871**	-0.339	0.977**	1.000			
	Water temp	0.124	0.090	0.712**	0.221	0.224	0.347	0.044	0.181	0.144	0.153	1.000		
	Chl-a	0.777**	0.843**	0.390	0.931**	0.915**	0.909**	0.879**	-0.129	0.885**	0.857**	0.277	1.000	
	PO <sub>4</sub> -P	0.970**	0.938**	0.136	0.868**	0.943**	0.853**	0.808**	-0.386	0.953**	0.978**	0.158	0.814**	1.000
	(b)	NH <sub>3</sub> -N	1.000											
NO <sub>3</sub> -N		0.683**	1.000											
pH		-0.138	-0.089	1.000										
COD		0.580**	0.654**	0.087	1.000									
Cond		0.823**	0.895**	0.068	0.736**	1.000								
BOD		0.586**	0.597**	0.101	0.919**	0.717**	1.000							
SS		-0.104	0.134	-0.190	0.458**	0.010	0.266*	1.000						
DO		0.360**	0.126	-0.079	-0.243	0.264*	-0.140	-0.623**	1.000					
T-N		0.881**	0.898**	-0.059	0.757**	0.940**	0.752**	0.033	0.219	1.000				
T-P		0.811**	0.880**	-0.171	0.794**	0.895**	0.779**	0.236	0.003	0.918**	1.000			
Water temp		-0.537**	-0.341**	0.184	0.074	-0.435**	-0.032	0.567**	-0.911**	-0.440**	-0.224	1.000		
Chl-a		0.242	0.325*	0.418**	0.762**	0.464**	0.705**	0.267	-0.199	0.418**	0.416**	0.138	1.000	
PO <sub>4</sub> -P		0.790**	0.877**	-0.204	0.623**	0.876**	0.588**	0.090	0.139	0.885**	0.923**	-0.326*	0.254	1.000
(c)		NH <sub>3</sub> -N	1.000											
	NO <sub>3</sub> -N	0.745**	1.000											
	pH	-0.212	-0.253	1.000										
	COD	0.621**	0.701**	0.062	1.000									
	Cond	0.879**	0.921**	-0.118	0.767**	1.000								
	BOD	0.698**	0.654**	-0.019	0.959**	0.756**	1.000							
	SS	-0.074	0.208	0.088	0.626**	0.136	0.535*	1.000						
	DO	0.343	0.103	-0.260	-0.277	0.238	-0.170	-0.714	1.000					
	T-N	0.923**	0.924**	-0.217	0.783**	0.961**	0.800**	0.150	0.176	1.000				
	T-P	0.847**	0.912**	-0.241	0.826**	0.936**	0.833**	0.321	0.002	0.948**	1.000			
	Water temp	-0.528*	-0.323	0.355	0.091	-0.423	-0.023	0.632**	-0.953**	-0.405	-0.221	1.000		
	Chl-a	0.216	0.367	0.372	0.843**	0.466*	0.770**	0.661**	-0.393	0.412	0.488**	0.310	1.000	
	PO <sub>4</sub> -P	0.831**	0.923**	-0.345	0.654**	0.919**	0.65**	0.138	0.129	0.907**	0.954**	-0.323	0.292	1.000

\*\* Correlation is significant at the 0.01 level (2-tailed), \* Correlation is significant at the 0.05 level (2-tailed).

Fig. 2에서 대부분 하류에 위치한 강경, 성동, 연기, 정동 등에서 높은 오염 수치를 보이고 있으며 특히, COD와 BOD의 경우 강경, NH<sub>3</sub>-N, NO<sub>3</sub>-N, TN, TP, PO<sub>4</sub>-P는 연지에서 가장 높은 수치를 나타내었다. 강경은 금강 수계의 최 하류 부분에 위치하며 여러 지류를 거치면서 중소도시의 생활하수와 축산분뇨, 토지이용에 따른 비점오염원 등의 오염물질이 유하함에 따라 유기물에 의한 높은 농도를 나타냈으며, 연기는 주요 하천 지류인 갑천과 미호천의 영향을 받고 있다고 볼 수 있다. 갑천은 대전시 도시 하수의 특성을, 미호천은 농촌과 도시적 오염특성이 함께 나타나

기 때문에 다른 측정소에 비해 영양염류에 의한 높은 오염상태를 보이고 있어 위와 같은 결과를 나타내었다. 또한, 연기 측정소가 위치한 연기군의 경우 시가인구보다 비시가 인구가 높은 분포를 나타내고 있어, 도시보다 농촌지역의 오염특성을 결과적으로 더 반영하고 있다고 판단된다.

수질인자의 단순관련성을 알아보기 위하여 13개 수질항목에 관하여 Table 1에서 전체 측정소, 주요지점의 월별, 계절별 단순상관분석을 실시하였다. 전체 측정소에 대하여 유기오염 지표인 BOD와 COD는 (r=0.958), T-N(r=0.897),

T-P( $r=0.902$ ), Chl-a( $r=0.915$ )와  $PO_4$ -P( $r=0.853$ )로 높은 상관성( $p<0.01$ )을 나타내고 있으며, DO와 수온은 각 관련항목에 대하여 낮은 연관성을 보이고 있으나, pH와 수온의 상관성이 높게 분석되었다.

주요지점인 공주1, 대청댐, 부여1, 옥천, 청원을 중심으로 평균 월별, 계절별로 상관분석을 실시한 결과 월별 상관관계 분석 시 BOD와 COD는 ( $r=0.919$ ), T-N( $r=0.752$ ), T-P( $r=0.779$ ), Chl-a( $r=0.705$ )와  $PO_4$ -P( $r=0.588$ ), pH와 Chl-a( $r=0.418$ ), 수온과 SS( $r=0.567$ ), 계절별로 BOD와 COD는 ( $r=0.959$ ), T-N( $r=0.800$ ), T-P( $r=0.833$ ), Chl-a( $r=0.770$ )와  $PO_4$ -P( $r=0.650$ ), 수온과 SS( $r=0.632$ )로 유의수준 0.01에 해당하는 높은 상관성을 나타내었다.

### 3.2. 주성분분석 및 요인분석

금강수계 전체 수질측정망(19지점)의 연간 수질항목별 평균자료와 환경부에서 정한 수질측정망 주요지점인 공주1, 대청댐, 옥천, 청원 측정소에 대하여 5년간 월별, 계절별 평균자료로 나누어 원 자료를 표준화 한 후 분석을 실시하였으며, 측정소별 자료 행렬은 19(수질측정망수) $\times$ 13(수질항목수)의 형태를 갖으며, 5개 주요 측정망의 5년간 월별자료 행렬은 60(월별수) $\times$ 13(수질항목수), 계절별 자료 행렬은 20(계절수) $\times$ 13(수질항목수)의 차원으로 구분하여 분석하였다.

주성분분석을 이용하여 자료 추출결과 측정소별, 월별, 계절별로 13개의 요인의 고유치와 수질에 미치는 기여율을 나타내었으며, 산출된 주성분에 대하여 그 차원을 축소시켜 보다 정확한 해석을 위하여 차원 수를 결정하기 위해, 추출된 13개의 최소요인 중 직교회전 방법인 배리맥스(VARIMAX) 회전을 이용하여 요인수를 결정하였다.

이때 추출된 차수를 결정하는데 여러 방법이 있으나, 본 연구에서는 요인을 설명하는 분산의 크기를 나타내는 고유치(eigen value)를 기준으로 이 고유치가 1이상일 경우에 하나의 요인이 변수1개 이상의 분산을 설명하며, 고유치가 1이하인 경우 1개의 요인이 변수 1개의 분산을 설명할 수 없으므로 고유치가 1.0 이상이 되는 요인을 추출하는 방법과 함께, Fig. 3의 Scree plot으로 variance가 급격히 작아지기 직전의 차수까지만을 고려한 그래프를 통해 산출하였다. 그 결과 전체 측정소별로는 2개의 주성분, 주요지점의 월별, 계절별 자료는 모두 3개의 주성분이 추출되었다. 덧붙여, 13개의 항목간의 상관계수를 통해서 구한 각 축의 고유값과 누적 점유율을 Table 2에 나타내었다. 누적점유율은 각 요인의 고유치에 대한 누적기여율을 설명하는 것으로 이때, 주성분 전체 수질변동에 대하여 측정소별로 86.01%, 월별 87.37%, 계절별 91.22%를 설명하고 있다.

요인분석 실행 전 분석에 대한 타당성을 알아보기 위하

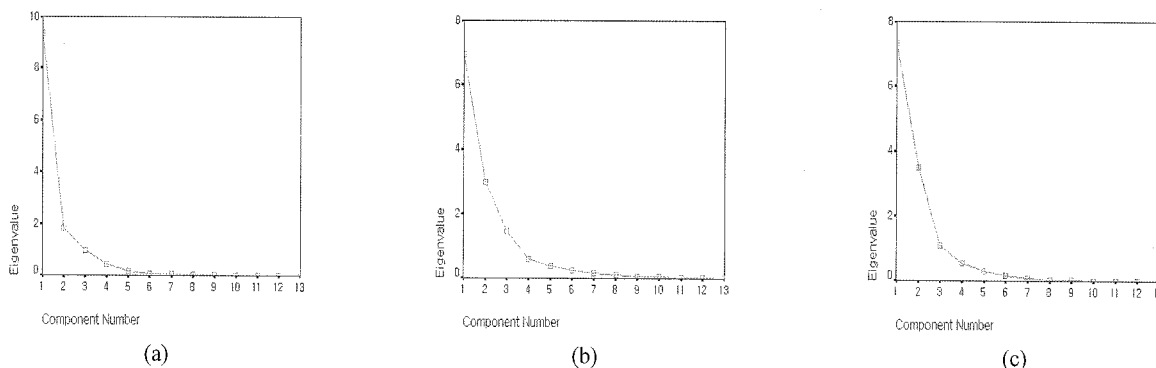


Fig. 3. Scree plot of total sampling stations (a), main sampling stations by month (b) and season (c).

Table 2. Initial eigenvalues explain of factor

Component	Total			% of Variance			Cumulative %		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
1	9.358	6.949	7.325	71.986	53.455	56.350	71.986	53.455	56.350
2(a)	1.824	2.951	3.467	14.028	22.697	26.666	86.014	76.152	83.016
3(b,c)	0.972	1.458	1.067	7.481	11.218	8.208	93.495	87.370	91.224
4	0.453	0.582	0.531	3.482	4.474	4.085	96.977	91.845	95.309
5	0.158	0.374	0.300	1.215	2.877	2.307	98.192	94.721	97.616
6	0.081	0.250	0.152	0.625	1.920	1.172	98.817	96.642	98.787
7	0.072	0.170	0.093	0.551	1.307	0.715	99.368	97.949	99.502
8	0.040	0.105	0.031	0.305	0.809	0.238	99.674	98.759	99.739
9	0.023	0.053	0.014	0.177	0.410	0.108	99.850	99.168	99.847
10	0.009	0.043	0.008	0.066	0.333	0.064	99.916	99.501	99.911
11	0.006	0.036	0.007	0.045	0.274	0.055	99.962	99.775	99.966
12	0.004	0.018	99.913	0.034	0.138	99.958	99.995	0.004	99.995
13	0.001	0.011	100.000	0.005	0.087	100.000	100.000	0.001	100.000

Total sampling stations(a),main sampling stations by month(b) and season(c)

여 Kaiser-Meyer-Olkin(KMO)과 Bartlett's test를 실행하였다. KOM test는 측정할 샘플에 대한 변수의 크기가 적당한가를 나타내는 척도로 1에 근접한 값일수록 요인분석의 타당성이 높고 0.5이하일 경우 분석이 적당하지 않다. Bartlett's test는 변수 간 상관행렬이 단위행렬(대각선이 1이고 나머지는 0인 행렬)인지 아닌지를 검정하는 것을 말하며, 각 변수가 서로 상관성이 존재함을 나타낼 때 유의적 관계라 하며 이때 0에 가까울수록 유의성이 높다(Bernard et al., 2004; 이, 2006).

본 연구의 KOM test는 전체 측정소 0.690, 주요지점의 월별 0.751, 계절별 0.617로 분석되었고 Bartlett's test는 모두 0.000(유의수준<0.05)을 나타내어 상관행렬이 단위행렬임을 기각하여 변수 간 상관이 있음을 나타냄으로써, 두 분석 모두 조건에 만족하여 요인분석이 가능함을 설명해주고 있다.

요인분석은 관측된 변수들에 영향을 미치고 있는 숨어있는 공통인자를 찾아내는 데 목적이 있으므로 그 결과 주성분분석으로 산출된 차원수를 직교회전이후 차원축소 수를 산출하여 더 뚜렷하게 각 영향 요인을 규정한다. 한편, 추출된 요인에 대한 명명화는 요인분석에서 핵심적으로 중요하면서도 연구자의 주관성과 전문성에 크게 영향을 받는 부분이다(허 등, 2004).

본 연구의 영향요인은 상관관계와 요인분석 결과를 바탕으로 크게 3가지로 구분하여 명명화하였다. 제 1요인은 오염부하요인으로 유기물에 의한 BOD, COD, Cond, 영양염류인 T-N, T-P, NH<sub>3</sub>-N, NO<sub>3</sub>-N, PO<sub>4</sub>-P, 제 2요인은 수계의 계절적 요인으로 SS와 수온으로 정하고 제 3요인은 물질대사 요인인 조류증식에 의한 Chl-a와 pH로 각각 구분하였다. 단 전체 측정소별로 산출된 요인은 2개를 나타내므로, 3요인인 조류와 pH는 계절적 영향보다 제1요인인 유기오염물질의 영향과 밀접한 관련이 있기 때문에 두 요인을 결합하여 설명하였다.

그 결과 Table 3에 주성분으로 산출된 인자를 회전 전·

후의 요인점수로 나타내었다. 전체 측정소의 제1요인은 오염부하물질로 연기와 갑천, 미호천에서 발생하는 생활하수에 의한 유기오염물 유입과 그와 동시에 축산농가와 토지이용에 따른 영양염류에 의한 오염물의 변화로 인하여 특히 하류에 위치한 연기, 공주1,2, 곰나루, 경동, 부여1,2, 성동, 강경, 양화 등에서 상류 측정소에 비해 월등히 조류증식에 의한 영향을 크게 받고 있음을 알 수 있다. 또한 2요인은 조류성장으로 인한 결과로 증가되는 pH와 유기오염에 의해 생산된 2차 오염물질과 계절적요인인 수온에 영향을 받는 것으로 나타나고 있다. 덧붙여 월별, 계절별의 1요인도 유기오염물질과 영양염류에 의한 오염부하인자로 대전, 청주, 공주, 부여의 도시에서 유출되는 도시지역의 생활하수가 갑천과 미호천을 거쳐 연기 측정소에 도달함에 따라 도시적 특성과 농가의 축산폐수 및 토지이용에 따른 유기오염물질과 영양염류 유출로 오염도가 극대화함에 따라 수질관리에 있어서 가장 중요한 변수임을 위 분석결과를 통해 분명히 알 수 있다. 요인 2는 계절적 요인으로 수온과 특히 강우 시 물리적 영향의 증가에 기인하여 유출되는 SS로 설명되었다.

요인 3은 Chl-a와 pH로 하류에 위치한 수질측정망 주요 지점인 청원과 공주1로 인구수증가와 전·답, 축산농가 같은 토지 이용면에서 영양염류로 인한 오염이 증가되어 Chl-a농도의 상승과 함께 그에 관련한 변수인 pH의 영향으로 이어져 상류에 위치한 대청댐과 옥천 측정소에 비해 특히 높은 요인 부하량을 기록하고 있다.

Table 4는 직교회전이전·후의 고유값과 누적점유율을 비교하여 나타내었다. 전 측정소는 가장 큰 순서대로 2축, 월별, 계절별로 각각 3축까지 고려하는 것이 타당함을 알 수 있다. 분석 결과 회전 전·후 모두 동일한 개수의 축이 산출되었으며, 측정소별로 제1요인은 71.16%, 제2요인은 14.86%, 월별로 제1요인은 52.75%, 2요인은 21.43%, 3요인은 13.19%, 계절별로 56.57%, 24.17%, 10.80% 순으로 금강수계는 유기오염물질과 영양염류에 의한 오염이 가장 크

Table 3. Rotated factor matrix extracted from principal component analysis

Variable	(a)		(b)			(c)		
	Factor1	Factor2	Factor1	Factor2	Factor3	Factor1	Factor2	Factor3
NH <sub>3</sub> -N	0.961	-0.080	0.853	-0.313	-0.046	0.883	-0.315	-0.029
NO <sub>3</sub> -N	0.970	-0.045	0.904	-0.057	-0.039	0.917	-0.022	-0.156
PH	0.193	0.846	-0.158	-0.086	0.891	-0.196	0.112	0.911
COD	0.947	0.216	0.820	0.376	0.351	0.854	0.407	0.261
Cond	0.975	0.132	0.938	-0.195	0.170	0.971	-0.150	0.062
BOD	0.903	0.337	0.793	0.234	0.383	0.863	0.303	0.220
SS	0.920	-0.001	0.194	0.853	-0.126	0.271	0.882	0.080
DO	-0.372	0.395	0.114	-0.919	-0.099	0.110	-0.941	-0.065
T-N	0.991	0.017	0.968	-0.155	0.069	0.981	-0.106	-0.037
T-P	0.989	0.013	0.976	0.092	-0.037	0.978	0.096	-0.101
Watertemp	0.116	0.890	-0.334	0.871	0.165	-0.325	0.893	0.138
Chl-a	0.895	0.280	0.453	0.273	0.732	0.531	0.517	0.569
PO <sub>4</sub> -P	0.963	-0.023	0.922	-0.068	-0.153	0.921	-0.054	-0.251

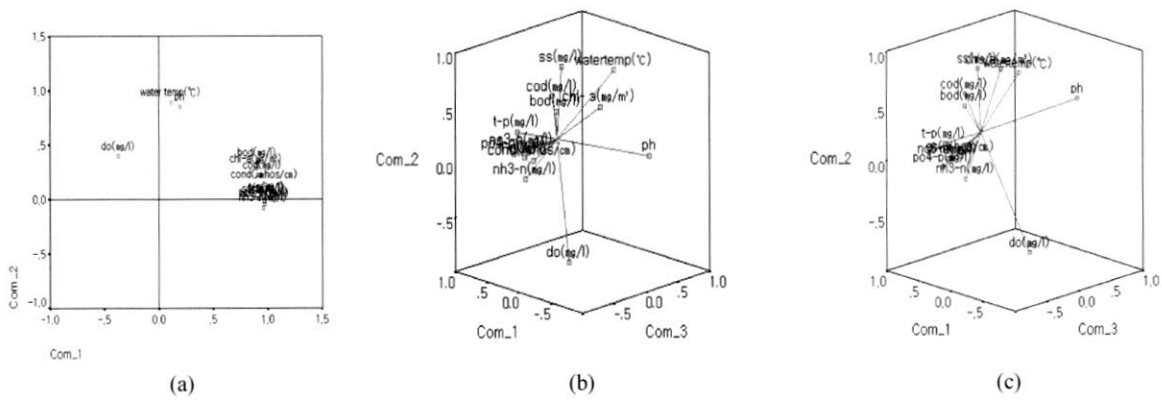
Rotation Method: Varimax with Kaiser Normalization

Total sampling stations (a), main sampling stations by month (b) and season (c)

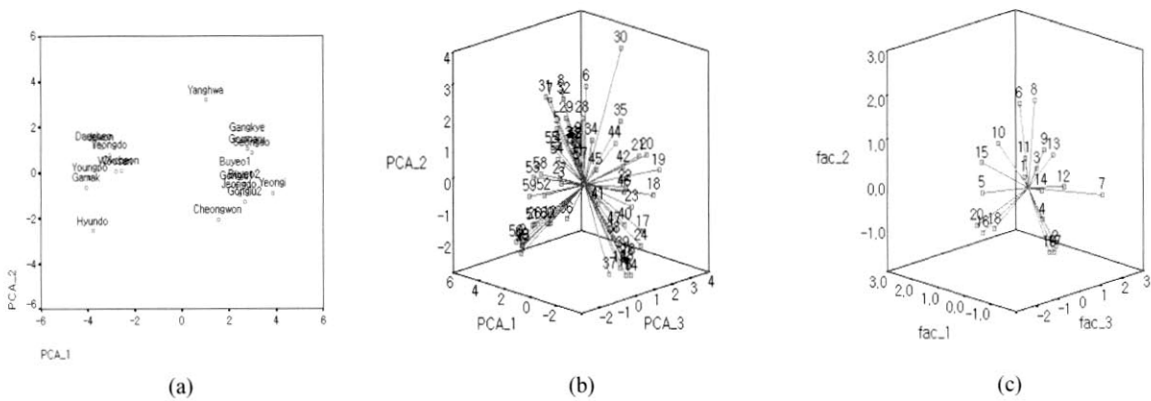
**Table 4.** Comparison between varimax rotation before and after matrix of factor loading

Sample	Component	Extraction sums of squared loadings			Rotation sums of squared loadings		
		Total	%of Variance	Cumulative %	Total	%of Variance	Cumulative %
(a)	1	9.358	71.986	71.986	9.251	71.160	71.160
	2	1.824	14.028	86.014	1.931	14.855	86.014
(b)	1	6.949	53.455	53.455	6.858	52.751	52.751
	2	2.951	22.697	76.152	2.786	21.431	74.182
	3	1.458	11.218	87.370	1.714	13.188	87.370
(c)	1	7.325	56.350	56.350	7.315	56.266	56.266
	2	3.467	26.666	83.016	3.141	24.165	80.431
	3	1.067	8.208	91.224	1.403	10.793	91.224

Total sampling stations (a), main sampling stations by month (b) and season (c)



**Fig. 4.** Component plot of total sampling stations (a), main sampling stations by month (b) and season (c).



**Fig. 5.** Scatter plot of total sampling stations (a), main sampling stations by month (b) and season (c)

(b) :1~12 (1~12 month of Gongju1), 13~24 (1~12 month of Daecheongdam), 25~36 (1~12 month of buyeo1), 37~48 (1~12 month of Okcheon), 49~60 (1~12 month of Cheongwon)

(c) :1~5 (Spring-Gongju1, Daecheongdam, buyeo1, Okcheon, Cheongwon order), 6~10 (Summer-on the same order in spring), 11~15 (Fall-on the same order in spring), 16~20 (Winter-on the same order in spring)

며 계절적 요인과 수체 내의 생물학적인 물질대사에 의한 요인 순으로 영향을 받고 있음을 설명해주고 있다. Fig. 4은 인자분석에 사용된 13개의 변수들에 대한 플롯이고 Fig. 5는 전체 수질측정지점과 주요지점의 월별, 계절별 개체플롯이다. Fig. 4(a)의 그림에서 제 1요인 축에 유기오염물과 영양염류에 해당하는 지표가 대부분 몰려있으며, 그 외 계절적 영향을 받는 수온과 DO는 2축에 분산되어 나타나고 있다. 월별, 계절별로 살펴본 주요측정지점의 그림인 Fig. 4(b)와 (c)도 유기오염물과 영양염류에 의한 것이 1요인

축에 표시되어 있고 그 외 계절적 영향을 받는 수온과 SS는 2축에, pH는 3축에 분산되어 있음을 알 수 있다.

측정소별로 살펴본 Fig. 5(a)에 따르면 유기오염물질에 영향을 받는 측정소는 대부분 하류에 위치한 연기, 부여1,2, 공주2, 강경, 양화 등에서 제 1요인에 의한 영향이 컸으며, 양화는 1요인보다 계절적 요인인 제 2요인에 영향을 더 받고 있는 지점으로 표시되었다. Fig. 5(b)는 주요측정지점을 기준으로 월별 평균을 내어 나타낸 결과 숫자 1~10번대와 50~60번대 개체인 공주1과 청원측정소로 연중 유기오염에

영향을 많이 받고 있으며, 28, 29번인 부여1의 4, 5월과 31, 32번에 해당하는 부여1의 7, 8월이 유기오염과 영양염류에 의한 오염이 가장 큰 요인으로 분석되었다. 이는 하류에 위치한 지역적 특성과 함께 4, 5월의 유량이 적어 수체내의 오염물질의 누적과 함께 자정작용이 활발하지 못한 것으로 판단되며, 7, 8월은 다른 달에 비해 강우로 인한 유량은 풍부하나(금강수계관리위원회, 2005), 비점오염원에 의한 오염물질의 다량 유입으로 위와 같은 분석결과를 나타낸 것으로 보인다. Fig. 5(c)의 유기오염물질지표와 영양염류에 의해 오염이 가장 심한 곳은 10(청원-여름), 15(청원-가을), 5(청원-봄), 20(청원-겨울), 16(공주-겨울), 18(부여-겨울)로 계절적으로 겨울에 오염이 가장 심하였고 청원측 정소는 계절에 관계없이 높은 오염된 상태를 보이고 있다. 이는 주 오염지류인 갑천의 영향을 받아 도시하수의 연중 유입이 원인으로 생각되며, 갑천과 미호천, 논산천을 제외하고는 주요 지천이 농촌유역 특성이 주를 이루며(임, 1999), 금강은 갑천, 미호천을 제외한 소하천의 유입에 의한 영향보다는 상류로부터 유입되어 오는 본류에 걸쳐 영향을 받아(한국토지공사 등, 2006) 위와 같은 분석결과를 나타낸 것으로 판단된다.

#### 4. 결론

금강수계의 수질 특성을 평가하기 위하여 2001~2005년 까지의 수질자료를 바탕으로 총 13개 오염지표(수온, pH, DO, SS, BOD, COD, TN, TP, PO<sub>4</sub>-P, Nitrate, Ammonia nitrogen, Phosphate, Chlorophyll-a)에 대하여 금강 본류 전체 측정소(19개 지점)별로 분석을 실시하였고 주요수질측정 지점 5개(옥천, 대청댐, 청원, 공주1, 부여1)에 대하여 5년간의 월별평균, 계절별평균 자료로 구분하여 주성분분석과 요인분석을 실시한 결과 측정소별로는 2개의 주성분, 주요 지점의 월별, 계절별 모두 3개의 주성분이 추출되었으며, 요인분석 시 제 1요인은 오염부하요인(유기오염물: BOD, COD, Cond, 영양염류: T-N, T-P, NH<sub>3</sub>-N, NO<sub>3</sub>-N, PO<sub>4</sub>-P), 제 2요인은 계절적 요인(수온, SS), 제 3요인은 물질대사요인(Chl-a, pH)으로 명명화하여 분석하였다. 그 결과 3가지 개체 모두 유기오염물질과 영양염류에 해당하는 오염부하 요인에 의한 영향이 가장 크며, 계절적 요인과 수체 내의 생물학적인 물질대사에 의한 요인 순으로 분석되었다.

본 연구의 결과를 다변량분석법을 이용하여 도출함으로써 자료의 확실성을 바탕으로 향후 수계관리에 있어 주 오염지류에 대한 연중관리와 함께 특히 하류지역의 4~5월과 7~9월의 오염부하 관리에 더욱 힘써야 하며, 계절적으로 동절기의 환경기초시설의 철저한 이용, 관리가 필요한 것으로 나타났다.

#### 참고문헌

- 금강수계관리위원회, 금강 수역별 유량관측망 설치 및 운영 보고서 (2005).
- 김건하, 금강유역 비점원에서 발생하는 미생물 오염 및 수질에 대한영향, *한국물환경학회지*, **22**(3), pp. 504-512 (2006).
- 김영주, 주성분분석에 의한 도시호수의 수질평가에 관한 연구, *환경관리학회지*, **9**(2), pp. 197-203 (2003).
- 김종구, 통계분석 기법을 이용한 금강수계의 수질평가, *한국환경과학회지*, **11**(12), pp. 1281-1289 (2002).
- 박해식, 박정길, 송미경, 백경훈, 신성교, 통계분석법을 이용한 낙동강 수질특성 평가, *한국물환경학회지*, **17**(6), pp. 693-701 (2001).
- 신성교, 박정길, 송교육, 주성분 분석법을 이용한 낙동강 하구 해역의 수질 평가, *한국환경관리학회지*, **7**(2), pp. 171-176 (1998).
- 유호식, 한강수질에 영향을 끼치는 요인들의 통계분석, *대한환경공학회지*, **24**(12), pp. 2139-2150 (2002).
- 이재우, 논고-수출선행지수 개발 및 적합성 분석, *수은 해외경제*, pp. 4-17 (2006).
- 임창수, 금강유역 14개 관측점의 수질자료를 이용한 수질의 다변량 분석, *한국환경과학회지*, **8**(3), pp. 331-336 (1999).
- 최용식, 정관모 공저, *SAS를 활용한 다변량 분석 기법과 응용*, 자유아카데미, pp. 36-95 (2003).
- 한국토지공사, 국립환경과학원, 행정중심복합도시 환경·생태분야 조사연구 중간보고서 (2006.7.).
- 한원식, 농업생명과학을 위한 통계적 방법, 자유아카데미, pp. 632-645 (1997).
- 허명희, 양경숙공저, *SPSS 다변량자료분석*, SPSS 아카데미, pp. 3-65 (2004).
- 환경부, 물환경정보시스템, <http://water.nier.go.kr/weis>, 국립환경과학원 (accessed Jul. 2006).
- 환경부, 중부권의 젓줄 금강 살리기-대청호등 금강수계 물관리 종합대책 (2001).
- Bernard, P., Antoine, L. and Bernard, L., Principal Component Analysis: an Appropriate Tool for Water Quality Evaluation and Management-application to a Tropical Lake System, *Ecological Modeling*, **178**, pp. 295-311 (2004).
- Karim, B. and Taha, F. M., Using Principal Component Analysis to Monitor Spatial and Temporal Changes in Water Quality, *Journal of Hazardous Materials*, **B100**, pp. 179-195 (2003).
- Petersen, W., Bertino, L., Callies, U. and Zorita, E., Process Identification by Principal Component Analysis of River Water-quality Data, *Ecological Modeling*, **138**, pp. 293-213 (2001).
- Ying, O., Evaluation of River Water Monitoring Stations by Principal Component Analysis, *Water Research*, **39**, pp. 2621-2635 (2005).