

다변량 분석법을 이용한
소양강댐 상류 유역의 하천 수질 평가

Evaluation of Water Quality on the Upstreams of the
Soyanggang Dam by using Multivariate Analysis

최한규* 백호선** 허준영***
Choi, Han-Kyu Baek, Hyo-Sun Heo, Joon-Young

Abstract

The object of this study is to evaluate the factors affecting the water quality and to propose the influence of dominant factor quantitatively.

The correlation analysis was performed to know the correlation among the water quality items. As a result of partial correlation analysis, it was shown that the water quality items are affected by the rainfall item directly.

The factor analysis was performed to grasp some number of factors on each point for deducing the items of similar variable characteristics. The four points were divided into different factor groups. It was grasped that $\text{NH}_3\text{-N}$ and $\text{NO}_3\text{-N}$ items have different variable characteristics after comparing the items.

The Multiple regression analysis can decrease the number of observation. In the deduced multiple regression formula, it was shown that the rate of T-N, $\text{NH}_3\text{-N}$ and $\text{NO}_3\text{-N}$ in the independent variable took about 60% among all the regression formulas.

키워드 : 수질, 다변량분석, 인자분석, 다중회귀분석

Keywords : water quality, multivariate Analysis, factor analysis, multiple regression analysis

1. 서론

1.1 배경 및 목적

하천유역을 오염시키는 발생원으로는 가정하수,

오수와 같은 점오염원(point source)이 가장 큰 부분을 차지하고 있지만, 산림, 농경지, 도로의 침식, 가축 분뇨 등과 같은 비점오염원(nonpoint source)도 상당부분을 차지하고 있다. 비점오염원에서의 유출수는 많은 양의 부유물질, 유기물질 등을 함유하고 있으며 강우시 하천으로 유입되어 수질에 영향을 미친다. 그러므로 수질오염문제를 보다 근본적으로 해결하고 하천의 종합적인 수질관리를 위해서는 점오염원뿐만 아니라 비점오염원에 대한 합

* 강원대학교 토목공학과 교수, 공학박사
** 강원대학교 대학원 토목공학과 박사과정
*** 강원대학교 대학원 토목공학과 석사과정

리적인 관리 방안이 필요하다. 비점오염원은 강우시 유출되는 것이 많으므로 개현성을 포함한 정량화에는 많은 노력이 필요하다. 강우시 유출되는 비점오염원의 부하가 수질에 끼치는 영향은 하수처리율이 향상되고, 경제활동수준이 증가하고, 토지이용이 고밀도화 될수록 높아지고 있다 또한, 급격한 산업발달과 사회생활 수준의 향상은 물 사용량의 현저한 증가를 수반하게 되었으며, 이로 인해 수자원의 이용 효율 증대 및 양질의 수자원 확보에 대한 관심이 더욱 높아지고 있는 실정이다. 특히, 강우로 인한 비점오염원의 하천으로의 유입은 하천 수질 악화에 큰 영향을 미치고 있다. 또한 강우에 의한 하천 수질반응은 그 반응 속도가 무척 빠르다

이러한 하천의 오염문제를 해결하기 위해 수많은 관측망이 하천에 설치되어 많은 양의 데이터가 수집·보관되고 있으나, 이에 대한 통계학적 특성을 파악하고, 이를 응용하고자 하는 노력은 상대적으로 적은 실정이다.

본 연구의 목적은 강원도 상류지역의 하천에서 강우시 하천 수질의 변동 특성에 영향을 미치는 인자들을 평가하여 수질 변수를 예측하고, 지역적인 수질변동 형태를 구명하며, 각 항목간의 통계학적 특성치를 파악하여 하천 수환경에 영향을 미치는 인자들을 분류하여, 지배인자의 영향력을 정량적으로 제시함으로써 수환경의 효율적인 관리를 위하는데 있다. 내린천, 북천, 수입천, 인북천 지점의 자료를 대상으로 다변량분석법을 적용하여 강우와 하천 수질항목간의 특성을 연구하였다.

본 연구에서 적용한 다변량분석에 의한 수질해석 기법은 수자원의 효율적인 관리를 위한 수환경의 장기적인 경향을 파악하고, 이러한 정보는 정책입안자 등 공공기관에서의 정책 수립에 효과적인 기여를 할 수 있을 것으로 사료된다.

본 연구에서 사용한 자료는 소양강댐 상류지역인 내린천지점의 2000년 7월 ~ 9월, 북천지점의 2000년 5월 ~ 11월, 수입천지점의 2000년 5월 ~ 12월, 인북천지점의 2000년 5월 ~ 2001년 10월의 자료를 이용하였다.

1.2 연구 방법

사회가 점차 도시화, 현대화되면서 인간의 활발한 활동을 야기하게 되고, 그러한 인간의 활동에 의한 비점오염원의 부하량은 양적 팽창과 더불어 질적 팽창을 거듭하면서 오염의 복잡성을 유발하고 있다. 또한 농촌 지역에서는 농약 사용량의 증가로 인한 하천의 수질 오염 문제가 심각히 대두되고 있다. 이러한 수환경의 오염문제를 해결하기 위해

최근 수십년 동안 많은 측정 자료가 수집·보관되고 있으나, 이에 대한 통계학적 특성을 규명하고자 하는 노력은 상대적으로 미약한 실정이다. 본 연구에서는 강우에 의한 비점오염원들의 하천유입으로 인한 강우 항목과 수질 항목들 간의 통계학적 특성을 파악하여 각 수질 변수를 예측하고 변동 특성이 유사한 항목들을 분류하여, 보다 수월한 수환경 정책 수립에 도움을 주고자 한다. 이를 위하여 다변량분석법중의 다중상관분석, 부분상관분석과, 인자분석, 다중회귀분석을 시행하였다.

통계처리를 위하여 SAS (Statistical Analysis System) 6.12 통계프로그램과 SPSS (Statistical Packages for Social Science) 10.0 통계프로그램을 이용하였다.

각 관측지점의 24시간 선행 강우량과, 유량 등의 강우항목들과 BOD, COD, TEMP, T-N, T-P, SS, NH₃-N, NO₃-N 등 8항목의 수질인자들간의 상관성을 파악하기 위하여 다중상관분석을 실시하였고, 강우항목의 영향을 완전히 배제한 부분상관분석을 하여 피어슨 상관계수(Pearson Correlation)를 구하였다. 유사한 변동 특성을 가진 인자들을 추출하여 인자 그룹으로 분류하기 위하여, 원자료와 원자료에 상용대수를 취한 값을 바탕으로 인자분석을 하였다. 그리하여 각 지점에 대해서 공통된 특성을 가지는 항목별로 각 그룹을 이룰 수 있게 하였다. 이를 바탕으로 각 지점별로 예측하고자 하는 수질항목들을 종속변수로 하고 그 외의 항목을 독립변수로 하였다. 각 독립변수의 유의도를 하나씩 검정하여 유의도 0.05 이상인 독립변수 항목은 배제하는 방법으로 최적의 다중회귀식을 도출하였다

2. 기본 이론

2.1 상관분석

둘 이상의 변량 사이에 존재하는 어떤 상호 종속적인 관계나 그 관계의 강도를 통계적으로 분석하는 방법을 상관분석(correlation analysis)이라 한다.

모상관계수(population correlation coefficient)가 ρ 인 이변량 모집단으로부터 얻은 크기 n 의 확률표본을 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 이라 할때, X 와 Y 의 분산에 대한 불편추정량은 각각

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (1)$$

이고 마찬가지로 X 와 Y 의 공분산에 대한 불편추정량은

$$S_{XY}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (2)$$

이다.

또한 모상관계수는 모공분산 $\sigma_{XY} = Cov(X, Y)$

와 모표준편차들 σ_X, σ_Y 에 대하여

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

이므로 ρ 를 추정하기 위한 표본상관계수(sample correlation coefficient) r 은 다음과 같이 정의한다.

$$r = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3)$$

여기서, S_{XY} 는 표본으로부터 얻어진 X 와 Y 의 공분산이며, S_X 는 X 의 표본표준편차, S_Y 는 Y 의 표본표준편차이다.

부분상관계수(partial correlation coefficient)는 두

변수간의 순수한 상관성을 찾기 위해 사용되는 방법으로 식 (4)와 같이 나타낼 수 있다.

$$\begin{aligned} r_{X_1 X_2 \cdot X_3} &= \frac{r_{X_1 X_2} - r_{X_1 X_3} r_{X_2 X_3}}{\sqrt{(1 - r_{X_1 X_3}^2)(1 - r_{X_2 X_3}^2)}} \\ r_{X_1 X_3 \cdot X_2} &= \frac{r_{X_1 X_3} - r_{X_1 X_2} r_{X_2 X_3}}{\sqrt{(1 - r_{X_1 X_2}^2)(1 - r_{X_2 X_3}^2)}} \\ r_{X_2 X_3 \cdot X_1} &= \frac{r_{X_2 X_3} - r_{X_1 X_2} r_{X_1 X_3}}{\sqrt{(1 - r_{X_1 X_2}^2)(1 - r_{X_1 X_3}^2)}} \end{aligned} \quad (4)$$

여기서, $r_{X_i X_j \cdot X_k}$ 는 X_k 의 영향력을 제거한 X_i 와 X_j 의 부분상관계수이고, $r_{X_i X_j}$ 는 X_i 와 X_j 의 상관계수이다.

순위로 평가하는 것이 용이한 경우에 상관계수 대용으로 쓰이는 것으로 순위상관계수가 있다. 다음과 같이 정의된다.

$$r_s = 1 - \frac{6 \sum_{i=1}^n (X_i - Y_i)^2}{n(n^2 - 1)} \quad (5)$$

확률변수 X 가 평균이 μ 이고 분산이 σ^2 인 정규분포를 따를 때 간단히 $N(\mu, \sigma^2)$ 으로 표현하며, 이 분포형을 표준정규분포의 확률변수 z 로 표현하면 $z(0, 1)$ 이 되며 이러한 과정을 표준화라 하며 다음과 같다

$$z = \frac{X - \mu}{\sigma} \quad (6)$$

여기서 z 는 표준정규 확률변수, X 는 확률변수, μ 는 평균, σ 는 표준편차를 의미한다.

2.2 인자분석

인자분석(factor analysis)은 상호 연관된 다변량 확률변수들간의 내부적 상호의존관계를 그 저변에 내재하는 가설적 개념인 소수 몇 개의 공통인자를 통해 재현, 해석하고자 하는 통계학적 분석방법이다.

각 관찰 개체가 $P(\geq 2)$ 개의 변수 X_1, X_2, \dots, X_P 에 관해 측정되어 P 개의 원소를 가진 각각의 다변량 관찰벡터 $X = (X_1, X_2, X_3, \dots, X_P)$ 를 형성하고, P -변량 변수벡터 X 는 모집단 평균벡터 $\mu = (\mu_1, \mu_2, \dots, \mu_P) = E(X)$ 와 모집단 공분산행렬 $\Sigma = [\sigma_{ij}] = Var(X)$ 가진다고 하면 공통 인자분석 모형은 다음과 같다.

$$\begin{aligned} X_1 - \mu_1 &= \lambda_{11} F_1 + \lambda_{12} F_2 + \lambda_{13} F_3 \dots \lambda_{1m} F_m + \epsilon_1 \\ X_2 - \mu_2 &= \lambda_{21} F_1 + \lambda_{22} F_2 + \lambda_{23} F_3 \dots \lambda_{2m} F_m + \epsilon_2 \\ &\vdots \\ X_P - \mu_P &= \lambda_{P1} F_1 + \lambda_{P2} F_2 + \lambda_{P3} F_3 \dots \lambda_{Pm} F_m + \epsilon_P \end{aligned} \quad (7)$$

행렬을 이용하여 이들을 전체적으로 표현하면 다음과 같다.

$$X_i - \mu_i = \lambda_{ij} F_j + \epsilon_i \quad (8)$$

여기서, X_i : 확률변수의 벡터

μ_i : 평균벡터

λ_{ij} : j 번째 인자에 대한 i 번째의 부하

F_i : 공통인자(common factor)

ϵ_i : 특수인자(specific factor)

변수들 사이의 모집단 공분산 행렬은 다음과 같이 표현된다.

$$\begin{aligned} Var(X_i) = \sigma_{ii} &= [\lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{im}^2] + \phi \\ &= h_i^2 + \phi_i \end{aligned} \quad (9)$$

$$\begin{aligned} Cov(X_i, X_j) = \sigma_{ij} &= \sigma_{ji} \\ &= \lambda_{i1} \lambda_{j1} + \lambda_{i2} \lambda_{j2} + \dots + \lambda_{im} \lambda_{jm} \quad i \neq j \\ &= l_{ij} \end{aligned} \quad (10)$$

식(9)에서 h_i^2 은 X_i 의 공통분산(common var

iance)이다. ϕ_i 는 특수분산(specific variance)이다. 각 변수들의 요인부하를 구한 인자행렬(factor matrix)을 얻은 후 적재값 l_{ij} 들의 속성을 살펴 인자들의 의미를 해석하고자 할 때 처음 구한 인자(unrotated factor)는 해석하기 난해할 수도 있으므로 뚜렷한 특징이 나타나도록 인자의 축을 회전하게 된다. 이때 인자들을 회전하는 방식에는 직교해법(orthogonal solution)과 사각해법(oblique solution)이 사용된다.

\hat{L} 를 $p \times m$ 을 갖는 행렬이라면 다음 식이 성립된다.

$$\hat{L}^* = \hat{L}T \quad (11)$$

여기서, \hat{L}^* 은 회전시킨 인자부하이고, $TT' = T'T = I$ 이며, T 는 $m \times m$ 의 직교행렬이고, T' 은 행렬 T 의 역행렬이며, I 는 단위행렬이다.

2.4 다중회귀분석

독립변수가 한 개인 경우 즉, $p=1$ 이면 단순회귀 모형이라 하고, 독립변수가 두 개 이상이면 다중회귀 모형이라 한다. 변수들간의 선형관계를 나타내는 회귀모형은 다음과 같다.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon \quad (12)$$

여기서, Y_i : 종속변수

X_{ip} : p 개의 독립변수

ϵ : 모형의 예측오차를 나타내는 오차항
위 식을 행렬을 사용하여 나타내면,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{N1} & X_{N2} & \dots & X_{Np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_p \end{bmatrix} \quad (13)$$

또는

$$Y = X\beta + \epsilon \quad (14)$$

Y 는 n 차원의 벡터이며 $Y = (y_1, y_2, y_3, \dots, y_n)$ 이며, β 는 $(p+1)$ 차원의 벡터로 $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_p)$ 로 정의되며, X 는 $n(p+1)$ 차원의 행렬로 다음과 같이 표현된다.

다.

$$X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix} \quad (15)$$

오차항을 뜻하는 ϵ 는 n 차원의 벡터로 최소제곱법에 의한 오차 제곱합은 다음과 같다.

$$\begin{aligned} S(\beta_0, \beta_1, \dots, \beta_p) &= \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} \\ &\quad \dots \beta_p X_{ip})^2 \\ &= (Y - X\beta)'(Y - X\beta) \end{aligned} \quad (16)$$

행렬과 벡터의 미분을 이용하면 최소제곱추정량은

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (17)$$

이다. 잔차벡터는

$$\hat{\epsilon} = Y - \hat{Y} = [I - X(X'X)^{-1}X']Y \quad (18)$$

으로 계산된다. 한편, \hat{Y} 와 $\hat{\epsilon}$ 의 직교성에 의하여

$$Y'Y = (\hat{Y} + \hat{\epsilon})'(\hat{Y} + \hat{\epsilon}) = \hat{Y}'\hat{Y} + \hat{\epsilon}'\hat{\epsilon} \quad (19)$$

가 성립한다. 결정계수 R^2 은 다음 식에 의해 계산된다.

$$R^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (20)$$

3. 유역현황 및 자료 조사

3.1 대상유역현황

본 연구는 소양강댐 상류지역의 소하천에서의 강우 및 유량과 수질항목간의 통계학적 특성치를 분석하여, 수질변수를 예측하고 유사한 변동특성을 지닌 수질항목들을 도출해서 하천 수환경에 미치는 지배인자들을 분석, 정량화하여 수환경 정책 수립에 기초하기 위한 것으로서 연구대상지역과 자료는 지난 2년간 실측하고, 조사한 지역과 자료이다.

본 연구의 대상유역은 소양강댐 상류 유역의 본류 3개지점(인북천, 내린천, 북천)과 양구 수입천을 대상으로 하였다.

소양강은 총 유로연장 166.2km이며 한강 수계를 형성하는 북한강의 최대 지류이며 오대산 부근에서 발원하여 남으로 유하여 춘천시 동북방향에서 북한강과 합류한다.

소양강댐 유역의 연평균 강우량은 1,179mm이며 유역으로부터 연평균 유입량은 21억6천7백만m³이다. 댐 유역의 토지이용 현황은 논 37.5km², 밭

106.7km², 임야 및 산지 2,410.8km², 대지 6.4km², 그리고 기타 141.4km²로 전체 유역면적의 89.2%가 임야 및 산지이다.

인북천 지역의 현황을 보면, 인북천은 휴전선 이북의 인제군 최북단에서 발원하며 유역면적은 577.92km²이며, 유로연장은 68.6km이다. 본 연구대상 지역은 원통수위국 유역으로 유역면적 525.6km²이며, 유로연장은 63km이다.

내린천 수위국은 소양강 지류인 내린천에 위치하고 있다. 내린천의 유역면적은 1,075km²이고 유로연장은 92.4km이며 내린천 수위국의 유역면적은 1,059.6km², 유로연장 86.2km로서 소양강댐 유역면적의 39.2%에 해당한다.

북천은 301.3km²의 유역면적을 가진 인북천의 제1지류로 소양강댐 유역면적의 11%에 해당하는 준용하천이다. 북천 유역의 평균 하상 경사는 1/129 정도를 이루고 있으며, 토지이용현황을 보면 총 면적 중 14.31km²가 농경지이고 임야는 277.87km², 대지는 8.28km²를 기타 9.08km²를 차지하고 있어 대부분이 산지로 이루어져 있음을 알 수 있다.

양구 방산 수입천 유역은 우리나라 제 1 하천인 한강의 제 3 지류로 양구군 양구를 상무릉리 지점에서 국가하천인 양구 서천에 유입되며, 유역면적은 334.0km², 유로연장은 50.06km이며, 유역의 경사는 구간별로 완급을 달리하고 있으나 1/80 정도의 급경사를 이루고 있으며 토지이용 상태를 살펴보면 논 25.2km², 밭 13.4km², 임야 250.4km², 기타 45km²를 이루고 있다. 전체적인 유역 현황은 그림 1과 같다.



그림 1 연구대상지역 유역도

3.2 강우 및 수질 자료

본 연구 대상 지역에 대한 우량 자료는 다음 표 1의 관측소의 자료를 이용하였으며, 수질 자료는 다음의 측정 장소에서 직접 측정하였다

표 1 각 유역별 우량관측소

하천	우량관측소	수질측정장소
인북천	서화관측소	원통수위국
내린천	창촌관측소	내린천수위국
북천	용대관측소	북천수위국
수입천	서화관측소	각시교수위국

4. 다변량 분석 및 고찰

4.1 강우항목 및 수질항목 통계분석

강우에 따른 유량과 수질항목간의 수환경에서의 변동 특성을 분석하기 위하여 강우항목 및 수질항목 등으로 기초자료를 구성하여 분석하였다.

수질 항목의 변동 특성을 파악하기 위하여 강우 자료는 각 지점의 도달시간을 고려하여 수질측정 선행 24시간부터의 강우자료를 입력하였다.

분석대상 자료로는 수질항목 BOD, COD, TEMP, SS, T-N, T-P, NH₃-N, NO₃-N 이고 강우자료로는 강우량과 유량을 사용하였다

4.2 항목간 상관분석

강우시 비점오염원의 하천 유입으로 인한 강우항목과 수질항목간의 상관성을 분석하기 위하여 상관분석을 하여 다중상관계수 및 부분상관계수를 산정하였다 각 지점에 대한 결과는 각각 표 2 ~ 9 에 정리하였다.

각각 지점에 대한 상관계수(Simple Correlation)와 강우항목을 통제한 부분상관계수(Partial Correlation)을 비교해보면, 내린천 지점의 경우 COD 항목은 부분상관계수가 증가함을 보이고 있고, BOD, NO₃-N항목은 감소함을 보이고 있다.

북천 지점은 BOD 항목의 부분상관계수가 증가했고, 수입천 지점의 부분상관계수는 BOD, COD, SS, T-N, T-P, NH₃-N, NO₃-N항목이 전반적으로 감소되었다. 인북천 지점의 경우 모든 항목에서 부분상관계수의 감소현상을 보였다. 기타 다른 항목은 지점마다 서로 상이한 증감을 보이고 있다 이는 강우항목이 여하간에 수질항목에 영향을 미치고 있다고 판단된다.

각 항목의 상관분석으로부터 강우항목들과 수질항목들의 상관성은 하나의 변수와의 관계라기보다는 여러 항목들간의 복합적인 영향으로 지점마다 그 특성이 상이한 결과를 보여주고 있다.

최한규, 백효선, 허준영

표 2 내린천 지점 다중상관분석

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N	유량	강우량
BOD	Pearson Correlation	1	0.6208	0.7971	0.5596	0.8332	0.7426	0.8836	0.5770	-0.4314	-0.1248
COD	Pearson Correlation		1	0.6416	0.8932	0.8974	0.8759	0.5186	-0.0815	0.0954	0.0690
TEMP	Pearson Correlation			1	0.5441	0.8358	0.7871	0.8759	0.5585	-0.3953	-0.1574
SS	Pearson Correlation				1	0.8347	0.7670	0.4916	0.0245	-0.0115	-0.0255
TN	Pearson Correlation					1	0.9696	0.8186	0.3268	-0.2991	-0.1569
TP	Pearson Correlation						1	0.7819	0.2627	-0.2726	-0.1586
NH ₃ -N	Pearson Correlation							1	0.7256	-0.6872	-0.1926
NO ₃ -N	Pearson Correlation								1	-0.7662	-0.5486
유량	Pearson Correlation									1	0.3742
강우량	Pearson Correlation										1

표 3 내린천 지점 부분상관분석 및 증감

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N
BOD	Pearson Correlation		0.7367	0.7547	0.6165	0.8221	0.7249	0.8966	0.4951
COD	Pearson Correlation	-		0.7438	0.9001	0.9786	0.9466	0.8085	0.0029
TEMP	Pearson Correlation	+	+		0.5874	0.8192	0.7696	0.9109	0.4768
SS	Pearson Correlation	-	+	+		0.8714	0.7944	0.6713	0.0159
TN	Pearson Correlation	-	+	-	+		0.9673	0.8945	0.1525
TP	Pearson Correlation	-	+	+	+	-		0.8623	0.0660
NH ₃ -N	Pearson Correlation	+	+	+	+	+	+		0.5240
NO ₃ -N	Pearson Correlation	-	+	-	-	-	-	-	

표 4 북천 지점 다중상관분석

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N	유량	강우량
BOD	Pearson Correlation	1	0.8933	0.2862	0.8985	0.7042	0.8848	0.1699	-0.0605	0.5202	0.6527
COD	Pearson Correlation		1	0.1990	0.9334	0.7031	0.9569	0.0915	-0.0634	0.7381	0.8615
TEMP	Pearson Correlation			1	0.1178	0.5685	0.1933	0.0998	0.5691	-0.2183	-0.0150
SS	Pearson Correlation				1	0.6701	0.9322	0.0816	-0.1930	0.6699	0.7341
TN	Pearson Correlation					1	0.6414	0.3084	0.5409	0.1382	0.4165
TP	Pearson Correlation						1	0.0546	-0.1988	0.7784	0.7814
NH ₃ -N	Pearson Correlation							1	0.2853	-0.2936	-0.1081
NO ₃ -N	Pearson Correlation								1	-0.4840	-0.0945
유량	Pearson Correlation									1	0.8246
강우량	Pearson Correlation										1

표 5 북천 지점 부분상관분석 및 증감

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N
BOD	Pearson Correlation		0.8694	0.4035	0.8347	0.6669	0.8748	0.3260	-0.0415
COD	Pearson Correlation	-		0.4882	0.8730	0.8594	0.9326	0.4324	0.1525
TEMP	Pearson Correlation	+	+		0.2733	0.5703	0.5418	-0.0387	0.4770
SS	Pearson Correlation	-	-	+		0.7271	0.8579	0.3261	-0.0905
TN	Pearson Correlation	-	+	+	+		0.8352	0.2880	0.5562
TP	Pearson Correlation	-	-	+	-	+		0.4204	0.1134
NH ₃ -N	Pearson Correlation	-	+	-	+	-	+		0.0246
NO ₃ -N	Pearson Correlation	+	+	-	+	+	+	-	

다변량 분석법을 이용한 소양강댐 상류 유역의 하천수질 평가

표 6 수입천 지점 다중상관분석

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N	유량	강우량
BOD	Pearson Correlation	1	0.9591	0.0934	0.9472	0.8798	0.9316	0.2417	0.1900	0.5060	0.7151
COD	Pearson Correlation		1	0.1146	0.9753	0.9142	0.9427	0.2077	0.2446	0.6314	0.7331
TEMP	Pearson Correlation			1	0.0832	0.0897	0.1297	-0.3576	0.2315	-0.0571	0.0707
SS	Pearson Correlation				1	0.9397	0.9819	0.1458	0.1368	0.6434	0.6939
TN	Pearson Correlation					1	0.9473	0.2301	0.3637	0.6995	0.7299
TP	Pearson Correlation						1	0.1670	0.1824	0.5581	0.7016
NH ₃ -N	Pearson Correlation							1	0.4528	-0.0608	0.3036
NO ₃ -N	Pearson Correlation								1	0.1951	0.5787
유량	Pearson Correlation									1	0.4998
강우량	Pearson Correlation										1

표 7 수입천 지점 부분상관분석 및 증감

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N
BOD	Pearson Correlation		0.9289	0.0910	0.9138	0.7631	0.8548	0.1070	-0.3746
COD	Pearson Correlation	-		0.1583	0.9405	0.7617	0.8712	0.1078	-0.2981
TEMP	Pearson Correlation	-	+		0.1125	0.1418	0.1587	-0.4439	0.2232
SS	Pearson Correlation	-	-	+		0.8424	0.9715	0.0330	-0.4444
TN	Pearson Correlation	-	-	+	-		0.9055	0.1991	-0.0366
TP	Pearson Correlation	-	-	+	-	-		0.0207	-0.3644
NH ₃ -N	Pearson Correlation	-	-	-	-	-	-		0.3366
NO ₃ -N	Pearson Correlation	-	-	-	-	-	-	-	

표 8 인북천 지점 다중상관분석

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N	유량	강우량
BOD	Pearson Correlation	1	0.9546	0.4648	0.8832	0.9178	0.9373	0.4370	0.7638	0.8929	0.9216
COD	Pearson Correlation		1	0.2768	0.9702	0.9810	0.9959	0.2763	0.7065	0.9280	0.9262
TEMP	Pearson Correlation			1	0.2244	0.2174	0.2552	0.1945	0.3016	0.2070	0.2350
SS	Pearson Correlation				1	0.9750	0.9672	0.1149	0.6527	0.8801	0.8408
TN	Pearson Correlation					1	0.9764	0.2684	0.7044	0.8770	0.8739
TP	Pearson Correlation						1	0.2440	0.6877	0.9278	0.9202
NH ₃ -N	Pearson Correlation							1	0.6690	0.3619	0.4684
NO ₃ -N	Pearson Correlation								1	0.7611	0.7457
유량	Pearson Correlation									1	0.9753
강우량	Pearson Correlation										1

표 9 인북천 지점 부분상관분석 및 증감

		BOD	COD	TEMP	SS	T-N	T-P	NH ₃ -N	NO ₃ -N
BOD	Pearson Correlation		0.7457	0.6559	0.6381	0.6303	0.6546	-0.0207	0.3126
COD	Pearson Correlation	-		0.2022	0.9555	0.9339	0.9713	-0.3942	-0.0057
TEMP	Pearson Correlation	-	-		0.1198	0.0508	0.1487	0.0552	0.2262
SS	Pearson Correlation	-	-	-		0.9476	0.9104	-0.4486	-0.0528
TN	Pearson Correlation	-	-	-	-		0.9046	-0.2544	0.1160
TP	Pearson Correlation	-	-	-	-	-		-0.4530	-0.0820
NH ₃ -N	Pearson Correlation	-	-	-	-	-	-		0.7704
NO ₃ -N	Pearson Correlation	-	-	-	-	-	-	-	

4.3 인자분석

강우에 의한 수질의 변동 특성은 편차가 매우 크고, 비선형적 관계를 가지므로, 원자료와 원자료에 상용대수를 취한 자료에 대해 인자분석을 실시하였다. 인자 개수의 결정을 위해서 고유값이 1 이상인 인자를 선정하였다. 각 지점에 대한 고유값 및 비율은 표 10 ~ 13에 나타내었다. 표 10 ~ 13에서 보면 전체 분산의 누가 설명비율을 나타내는 누가값이 내린천 지점을 제외한 북천, 수입천, 인북천 지점에서는 상용대수자료가 원자료보다 큰 값을 나타낸다. 인자의 설명력이 높은 누가값이 큰 인자그룹을 선택하였으며, 내린천 지점과 인북천 지점에서 인자의 수를 2개로, 북천, 수입천 지점에서는 인자의 수를 3개로 결정함이 타당할 것으로 판단된다.

표 10의 내린천 지점의 인자분석 결과를 보면, 인자의 개수가 2개이고, 각각 수질항목과 강우항목으로 구분되어 있음을 알 수 있다.

표 11의 북천 지점에서는 인자의 개수가 3개로 결정되었으며, 제1인자그룹에는 BOD, COD, SS, T-N, T-P, 유량, 강우량, 제2인자그룹에는 TEMP, NH₃-N이 형성되었으며, 제3인자그룹에는 NO₃-N 항목이 포함되었다.

표 12 수입천 지점은 인자의 수가 3개로 제1인자그룹에 BOD, COD, TEMP, SS, T-N, T-P, 제2인자그룹에 NH₃-N, NO₃-N, 강우량항목이, 제3인자그룹에는 유량항목이 포함되었다.

표 13 인북천 지점의 인자 개수는 2개이며 제1인자그룹에는 BOD, COD, TEMP, SS, T-N, T-P, NO₃-N, 유량, 강우량항목이 제2인자그룹에는 NH₃-N 항목이 포함되었다. 4개지점 모두 유사한 변동 특성을 지니는 항목은 BOD, COD, T-N, T-P 항목이다.

표 11 북천 지점 고유값 및 비율

	원자료		상용대수자료		
	인자1	인자2	인자1	인자2	인자3
BOD	0.9136	0.1414	0.9410	0.1174	-0.3082
COD	0.9915	0.0181	0.9835	-0.1594	0.0016
Temp.	0.2068	0.7549	0.6524	0.7277	0.1301
SS	0.9535	-0.0376	0.9795	-0.0449	-0.1941
T-N	0.6984	0.6570	0.9742	-0.1119	0.1942
T-P	0.9769	-0.0609	0.9411	-0.0521	-0.3158
NH ₃ -N	0.0608	0.5170	0.2074	0.8870	-0.3329
NO ₃ -N	-0.0947	0.8696	0.2192	0.2108	0.9420
유량	0.7509	-0.5711	0.5295	-0.8333	-0.0511
강우량	0.8482	-0.2204	0.9005	-0.0172	0.4345
고유값	5.5077	2.4252	6.2550	2.1119	1.4766
누가값	55.0771	79.3287	62.5505	83.6695	98.4358

표 12 수입천 지점 고유값 및 비율

	원자료			상용대수자료		
	인자1	인자2	인자3	인자1	인자2	인자3
BOD	0.9398	-0.0800	-0.0853	0.9365	0.0062	-0.0954
COD	0.9710	-0.0992	-0.0508	0.9655	0.0540	0.1227
Temp.	0.1011	-0.3700	0.8735	0.7976	0.2019	-0.5655
SS	0.9657	-0.1884	-0.1241	0.9761	-0.1554	0.1318
T-N	0.9687	-0.0245	-0.0112	0.8778	0.1405	0.3373
T-P	0.9549	-0.1521	-0.0598	0.9948	-0.0184	0.0768
NH ₃ -N	0.2601	0.8611	-0.2062	-0.5081	0.6294	0.4588
NO ₃ -N	0.3677	0.6563	0.5839	-0.2434	0.9244	0.2086
유량	0.6816	-0.1893	-0.1225	0.3257	-0.3525	0.8745
강우량	0.8207	0.2728	0.1845	0.6482	0.7516	-0.1206
고유값	5.9602	1.4935	1.2245	6.0400	2.0278	1.5143
누가값	59.6016	74.5362	86.7810	60.3999	80.6782	95.8209

표 10 내린천 지점 고유값 및 비율

	원자료		상용대수자료	
	인자1	인자2	인자1	인자2
BOD	0.8972	-0.0785	0.9253	-0.1720
COD	0.7723	0.6144	0.9164	0.3309
Temp.	0.9010	-0.0587	0.8734	-0.2466
SS	0.7385	0.5066	0.8465	0.1315
T-N	0.9677	0.2271	0.9837	0.0211
T-P	0.9219	0.2455	0.9630	0.1774
NH ₃ -N	0.9282	-0.2781	-0.8261	-0.3038
NO ₃ -N	0.5421	-0.7944	0.2656	-0.9258
유량	-0.4739	0.7417	-0.4842	0.7461
강우량	-0.2351	0.5586	0.1988	0.6270
고유값	5.9803	2.3263	6.0974	2.1482
누가값	59.8033	83.0662	60.9744	82.4566

표 13 인북천 지점 고유값 및 비율

	원자료		상용대수자료	
	인자1	인자2	인자1	인자2
BOD	0.9733	0.0719	0.9924	0.1126
COD	0.9814	-0.1707	0.9974	-0.0711
Temp.	0.3334	0.4164	0.9355	0.3584
SS	0.9327	-0.3150	0.8921	-0.3578
T-N	0.9579	-0.1957	0.9834	-0.1396
T-P	0.9724	-0.2055	0.9825	-0.0987
NH ₃ -N	0.4208	0.8192	-0.0558	0.9897
NO ₃ -N	0.8122	0.4080	0.9927	-0.0722
유량	0.9533	-0.0645	0.9841	-0.0199
강우량	0.9561	0.0264	0.9362	0.3467
고유값	7.4142	1.2299	8.4130	1.4088
누가값	74.1420	86.4414	84.1300	98.2185

4.4 다중회귀분석

강우에 의한 변동성이 큰 수질항목들을 파악하고, 강우와 수질항목들간의 관계식을 제시하기 위해 인자분석의 결과로부터 다중회귀분석을 실시하였다. 각 수질 항목을 종속변수로 놓고 그 외 항목들을 독립변수로 배치하였다. 독립변수로 제시된 항목들의 유의도를 파악하여 유의도가 0.05이상인 항목은 배제하고 다중회귀분석을 실시하였다. 유의도가 0.05이상인 경우에는 통계치가 모수치를 대표함에 있어 오차가 5%이상을 의미하므로 종속변수에 미치는 영향이 무의미하여 배제하였다.

지점별 다중회귀식은 표 14에 제시하였으며, 회귀식은 식 (14) $Y = X\beta + \epsilon$ 이다.

여기서, X : 독립변수
 β : 회귀계수
 ϵ : 절편

각 회귀식의 상관계수는 복천 지점에서 두 식, 수입천 지점에서의 한 식을 제외하고 모두 0.766 ~ 0.999 이므로 식의 타당성은 있다고 볼 수 있다.

표 14 지점별 다중 회귀식

지점	다중회귀식	R
내린천	$BOD = 0.734 + 37.503 \cdot (NH_3-N)$	0.884
	$COD = 0.876 + 6.402 \cdot (T-N) - 6.854 \cdot (NO_3-N) + (4.67E-03) \cdot Q$	0.990
	$SS = -17.151 + 10.539 \cdot COD$	0.893
	$T-N = 1.152 + 0.274 \cdot BOD + (3.15E-03) \cdot SS + 4.123 \cdot SS$	0.995
	$T-P = -0.181 + 0.142 \cdot (T-N)$	0.970
	$NH_3-N = (-4.1E-02) + (9.354E-03) \cdot BOD + (3.221E-03) \cdot Q + (-2.9E-05) \cdot TEMP$	0.980
	$NO_3-N = 1.616 + (-6.9E-04) \cdot Q$	0.766
복천	$BOD = 0.554 + (1.246E-02) \cdot SS$	0.898
	$COD = -1.08 + 62.966 \cdot (T-P) + (5.34E-02) \cdot (T-N) + 1.6 \cdot 강우량$	0.986
	$SS = -10.424 + 6.268 \cdot COD$	0.933
	$T-N = (-6.0E-02) + (8.96E-02) \cdot COD + (9.22E-03) \cdot SS + 1.003 \cdot (NO_3-N) + (-1.0E-02) \cdot 강우량$	0.984

지점	다중회귀식	R
복천	$T-P = (1.119E-03) + (3.086E-03) \cdot COD + (3.832E-02) \cdot (T-N) + (-3.7E-02) \cdot (NO_3-N) + (6.839E-05) \cdot Q$	0.982
	$NH_3-N = (1.097E-02) + 0.319 \cdot T-P + (-8.8E-05) \cdot Q$	0.538
	$NO_3-N = 0.179 + (4.68E-02) \cdot TEMP$	0.569
수입천	$BOD = 0.511 + (8.296E-02) \cdot COD + (-1.2E-03) \cdot Q$	0.968
	$COD = -3.234 + 2.665 \cdot BOD + (5.47E-02) \cdot SS + (-1.611) \cdot (T-N) + (-15.139) \cdot (T-P) + 6.375 \cdot (NO_3-N)$	0.994
	$SS = 25.826 + 11.489 \cdot COD + 473.372 \cdot (T-P) + (-75.468) \cdot (NO_3-N) + 0.214 \cdot 강우량$	0.998
	$T-N = -0.123 + 4.409 \cdot (T-P) + 1.036 \cdot (NO_3-N) + (3.649E-03) \cdot Q$	0.985
	$T-P = (-1.1E-02) + (-8.8E-03) \cdot COD + (1.21E-03) \cdot SS + (6.31E-02) \cdot (T-N) + (-5.3E-04) \cdot Q$	0.995
	$NH_3-N = (1.672E-02) + (-2.0E-03) \cdot TEMP + (3.729E-02) \cdot (NO_3-N)$	0.656
	$NO_3-N = 0.380 + (5.488E-02) \cdot COD + (-4.0E-03) \cdot SS + 0.293 \cdot (T-N) + (2.984E-03) \cdot 강우량$	0.926
	$BOD = (-4.7E-02) + 0.109 \cdot COD + (3.55E-02) \cdot TEMP + 6.71 \cdot (NH_3-N)$	0.989
	$COD = 0.743 + 2.256 \cdot BOD + (-5.7E-02) \cdot TEMP + (7.04E-03) \cdot SS + 15.551 \cdot (T-P)$	0.999
	$SS = -86.533 + 90.087 \cdot (T-N) + (-1709.419) \cdot (NH_3-N) + 0.286 \cdot Q$	0.992
$T-N = 0.628 + 0.258 \cdot COD$	0.981	
$T-P = (-5.1E-02) + (3.493E-02) \cdot COD$	0.996	
	$NH_3-N = (-2.2E-02) + (-2.7E-04) \cdot SS + (2.441E-02) \cdot (T-N) + (3.259E-02) \cdot (NO_3-N)$	0.898
	$NO_3-N = -0.201 + 0.341 \cdot BOD + 10.023 \cdot (NH_3-N)$	0.850

5. 결론

1. 강우항목이 수질항목에 미치는 영향을 파악하기 위하여 다중상관분석과 부분상관분석을 실시하였다. 강우항목의 영향을 통제한 부분상관분석의 결과로부터 각 수질 항목의 부분상관계수가 증감과 감소를 보여주었으며, 이는 강우항목이 수질항목에 복합적인 영향을 미친다고 판단된다.
2. 인자분석을 실시하여 비슷한 변동특성을 갖는 항목별로 분류하여 그룹을 나누었다. NH₃-N, NO₃-N 항목이 전반적으로 기타 항목에 비해 상이한 변동특성을 갖는 것으로 파악됐다. 4개 지점 모두 유사한 변동특성을 지니는 항목은 BOD, COD, T-N, T-P 항목이다.
3. 다중회귀분석을 실시함으로써 대표성을 지니는 항목을 파악할 수 있었고 관측항목수를 줄일 수 있었다.
4. 다중회귀식에서 독립변수로서 N 계열 (T-N, NH₃-N, NO₃-N)이 포함되어 있는 비율이 전체 회귀식 중 약 60%를 차지하는 것으로 나타났다. 이는 4개 지점의 유역 특성에 기인하는 것으로 농약살포와 토사유출 등으로 인한 비점오염원의 유입이 하천수환경에 영향을 크게 미치는 것으로 판단된다.

참고문헌

[1] 김기영, 전명식. SAS 인자분석. 자유아카데미, 1993

[2] 김기영, 전명식. 다변량 통계자료분석. 자유아카데미, 1997

[3] 백경원, “한강하류부에서의 확성론 및 통계학적 기법에 의한 수질분석”. 박사학위논문. 홍익대학교, 1995

[4] 백경원, 정용태, 한건연, 송재우. “한강 하류부 수질의 통계학적 해석”. 한국수자원학회지, 제 29권, 제2호, pp.179 ~ 190, 1998

[5] 윤상운, 이태섭. 실용통계학. 자유아카데미, 1996

[6] 왕창근 외 5인 공역. 수질데이터의 통계적 해석. 동화기술, 2000

[7] 이기훈, 김연형. 통계자료분석. 자유아카데미, 1997

[8] 이용환, 백경원, 한건연, 송재우. “인자분석법에 의한 강우와 하천 수질변동 특성에 관한 연구”. 대한토목학회논문집, 제18권, 제II-3호, pp.263 ~ 276, 1988

[9] 정상만, 임경호, 최정현. “금강지류 유역에서의 유출량과 오염부하량의 상관관계 분석” 한국수자원학회지 제33권, 제5호. pp.527 ~ 536, 2000

[10] 한국수자원공사, 소양강댐관리단. 소양강댐 유량측정등 수문기초조사, 2000 ~ 2001

[11] 황입구, 윤태훈. “하천오염인자의 통계적 특성” 한국수문학회지. 제14권, 제4호. pp.19 ~ 26

[12] Abahams,A.D. "Factor analysis of drainage basin properties: evidence for stream abstraction accompanying the degradation of relief." *Water Resorces Research Vol. 8, No. 3,, pp.624 - 633, 1972*

[13] Alexander B. *Statistical factor analysis and related method.* John Wiley & Sons, Inc. 1994

[14] Barring, L. "Regionalization of daily rainfall in Kenya by means of common factor analysis" *Journal of Climatology, Vol. 8, No.4., pp.371 ~ 389, 1988*

[15] Beaudon, P. and Rousselle, J. "A study of space variation of precipitation by factor analysis." *J ournal of Hydrology, Vol. 59, No. 1-2, pp.123 ~ 138, 1982*

[16] Gatz, D. F. "Associations and mesoscale spatial realtionships among rainwater constituents." *Journal of Geophysical Researsh, Vol. 85, No. C10, pp.5588-5598, 1980*

[17] Hordon, R. M. "A factor analysis of selected water quality variables in central New Jersey during 1960 ~ 1969." Paper No.H58, presented at 53rd Annual Meeting of the American Geophysical Union, Hydrology Section, April 17-21