

권세혁\* · 오현승\*\*†

\*한남대학교 경상대학 비즈니스통계학과

\*\*한남대학교 공과대학 산업경영공학과

## Construction of Observational Locations for Measuring Water Quality in the River Area

S. H. Kwon\* · H. S. Oh\*\*†

\*Department of Statistics, Hannam University

\*\*Division of Industrial and Management Engineering, Hannam University

The methods for constructing network of observational locations for measuring water quality in water reservoirs have been widely proposed, but they had some limitations to be applied to river areas, which lie in awkward clustering and finding representative observational locations among locations within each clustering. In this paper, a statistical approach to detect anomaly locations which were significantly different in important measurements for the water quality from the previous locations and construct observational network with them was proposed. Anomaly was detected with the sampling distribution of each primary principal component score, sum of primary PCs, or sum of residual PCs. The empirical study with the data of Nakdong Dam for guiding how to use our proposed approach and showing limitations of previous studied was described.

**Keywords** : Anomaly, Observation Network, Principal Component Analysis, River Area, Water Quality

### 1. 서 론

1977년부터 시작된 하천과 상수원의 수질오염에 대한 정부의 제도적 관리는 수질보존 및 관리대상 공공수역에 대한 수질 현황을 측정하기 위한 수질 측정망을 운영하고 있다. 그러나 건설교통부 수질관측 매뉴얼에는 수질 관측 지점 선정방법에 대한 언급은 없어 수질 관리자의 주관적 판단이나 경험에 의해 한동안 수질 관측망이 구성되고 있었다. 수질환경의 급격한 변화와 측정항목의 증가 및 다양성은 주관적인 방법으로 구성된 수질 관측망으로는 정확

하고 신뢰할만한 수질 측정은 불가능하게 되어 객관적인 방법에 의한 수질 측정망 구성의 필요성이 대두되었다.

수질 측정망 구성 연구 초기에는 지리적 환경이나 지류 수, 유입량과 유출량, 유량과 유속 등과 같은 관측 대상 유역의 특성을 고려하여 연구가 진행되었다. 유전자 알고리즘을 활용하여 하천 수계에 대한 수질 측정망 설계 방법을 제시한 박수영 등[4]의 연구, 시뮬레이션 데이터와 수질모델을 이용한 최적화 기법을 적용한 경민수 등[1]의 연구가 있다. 왕수균 등[5]은 수질특성 파악, 새로 도입된 물 관리제도 및 기술에 대한 지원이 가능한

수질측정망 설계기법을 제시하였고 권상우 등[2]은 엔트로피 방법을 이용한 수질측정망 평가방법을 제시하였다.

통계적 기법을 활용한 수질 측정망 연구는 수질 관형 항목의 특성을 활용하여 관측지점 간 수질 유사성을 측정하고 측정된 유사성의 크기에 의해 관측망을 구성하는 접근방법이 제안되었다. 권세혁 등[3]은 측정된 주요 수질항목들에 주성분분석을 적용하여 관측지점들의 주성분점수를 구하고, 유클리디안 거리를 이용하여 관측지점 간 유사성을 측정하여 군집분석에 의해 최적 수질망을 구성한 방법을 제안하였다. 이 방법은 수질 데이터의 횡단적 분석으로 수질의 시간적 변화를 고려하지 못하는 한계가 있어 이요상 등[6]은 관측지점의 수질 시계열 패턴을 고려할 수 있는 상관분석을 활용하여 관측지점 유사성을 측정하고 다차원척도법으로 유사 관측지점을 분류하여 최적 수질망을 구성하는 방법을 제안하였다.

두 방법은 수질이 지리적 특성에 의한 결정되는 저수지의 수질 측정망 구성에 적용할 수 있으나 상류 수질이 하류 수질에 영향을 주는 하천의 경우에는 수질 측정망 구성에 활용하기는 적합하지 않다. 본 연구에서는 주성분에 의한 이상점 진단 방법을 활용하여 수질의 변화가 발생하는 지점을 진단하고, 이 지점들을 관측지점으로 하여 수질 측정망을 구성하는 방법을 제안하였다. 제 2장에서는 기존연구의 한계와 이상점 진단방법론을 기술하였고 제안한 방법론 활용을 제시하기 위하여 실시한 낙동강 수계 데이터 실증분석은 제 3장에 보였다. 또한 실증분석에서는 기존 연구방법의 한계를 보였으며, 제 4장에는 향후연구방향과 연구결과를 정리하였다.

## 2. 연구방법

### 2.1 기존연구 한계

관측지점 간 수질 유사성을 활용하여 수질 측정망을 구성하는 기존에 제안된 연구에서는 관측지점들을 수질이 유사한 군집으로 나누고, 각 군집의 대표지점(군집의 수질 중심에 위치한 관측지점)을 관측 대상지점으로 선정하여 수질 측정망을 구성한다. 이 방법은 초기 관측지점 선정 시 객관적인 기준보다는 지리적 환경이나 유속, 유량과 같은 물의 흐름 특성에 의존할 수밖에 없어 관측망 구성의 신뢰성이 저하되게 된다. 그리고 초기 관측되는 지점들을 어디로 선정하느냐에 따라 군집의 결과가 달라져 구성된 수질 측정망이 초기 관측지점 선정에 의존하는 한계를 지니게 된다.

수질 유사성과 군집분석을 활용하여 수질 측정망을 구성하는 기존 제안 방법을 하천 유역 수질 측정망 구성

하는 경우 다음의 절차를 따른다. 수질 측정망을 구성하려는 하천지역을 상류 시작점부터 하류 종단점까지 일정 간격으로 나누고 간격의 시작점이나 끝점의 초기 관측지점들로 설정한다. 초기 관측지점들에서 수질 관련 항목들을 측정하고 각 지점들 간 수질 유사성을 계산한다. 계산된 수질 유사성에 의해 관측지점들을 군집화 하고 각 군집 내의 관측지점 수질 관측치의 중심에 가장 가까운 지점을 관측지점으로 하여 수질 측정망을 구성한다. 그러나 이렇게 관측망을 구성하면 상류의 관측지점과 중류나 하류의 관측지점이 수질 유사 군집화 되기도 하고 군집을 대표하는 지점은 간격의 중앙에 위치한 관측지점이 선정되는 등의 한계를 지니게 된다.

### 2.2 이상점 진단 방법

본 연구에서는 주성분을 활용하여 이상점을 진단하는 방법을 활용하여 하천 유역 수질 관측망을 구성하는 접근방법을 제안하였다. 관측대상인 하천유역을 상류 시작점부터 하류 종단점까지 동일 간격으로 나누어 구성된 초기 관측지점들에서 수질이 유의적으로 변하는 관측지점들로 관측 수질망으로 구성하는 방법이다. 임의의 관측지점의 수질이 이전 다른 관측지점들과 유의적 다르다는 것은 수질 특성 측면에서 이상 관측지점을 의미하기 때문이다.

주성분분석(principal component analysis)은 원 변수들의 선형결합을 이용하여 얻은 주성분변수들을 이용하여 원 변수들의 공분산 구조를 설명하는 다변량 분석방법이다. 개체 수가  $n$ , 수질 관련 측정변수가  $k$ 개인 데이터 행렬  $X_{n \times k}$ 의 공분산 행렬  $S_{k \times k}$ 로부터 고유치  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ 을 얻고 이에 대응하는 고유벡터  $(e_1, e_2, \dots, e_k)$ 을 얻는다. 각 고유벡터를 전치하여 얻은 벡터들로 구성된 행렬  $(L)$ 을 선형계수로 하여 주성분 변수( $Y_{n \times k} = L_{n \times n} X_{n \times k}$ )을 얻는다.

주성분변수는 서로 독립이고 제일 주성분, 제이 주성분, 이렇게 순서대로 원변수의 변동을 더 많이 설명하게 된다. 주성분변수의 개수는 원변수들의 차수와 동일하고 고유치 값의 크기가 1 이상, 혹은 원변수 변동 설명비율이 80% 수준인 주성분 변수들을 주요 주성분, 그리고 나머지 주성분변수를 잔차 주성분이라 한다. 주요 주성분은 관측지점의 수질 측정변수들의 구조적 관계를 설명하므로 주요 주성분들에 의해 진단되는 이상 관측지점은 다른 관측지점과 수질들이 유의적으로 다르다는 것을 의미한다. 잔차 주성분들에는 수질 측정변수들의 구조적인 관계 중 주요 주성분에 의해 설명되지 못한 변동이므로 잔차 주성분에 의해 진단되는 이상 관측지점도 수질의 유의한 차이가 있다고 판단할 수 있다.

대표본 이론에 의해 각 주요 주성분의 분포는  $\chi^2$ -분포를 따르므로  $m$ 개 주요 주성분의 합도  $\chi^2$ -분포를 따른다.  $i$ 번째 관측지점의  $j$ 번째 주성분을  $y_{ij}$ 라 하면 아래 식의 타원은  $100(1-\alpha)\%$  신뢰영역이다. 신뢰영역 밖에 놓이는 관측지점은 이상 관측지점으로 다른 관측지점들과 관측된 수질들이 유의적으로 다르다고 할 수 있다[7].

$$\left\{ y_{ij}; \sum_{l=1}^m \frac{y_{il}^2}{\lambda_l} \leq \chi_{\alpha}^2(m) \right\} \quad (1)$$

잔차 주성분에 대해서도 동일하게 아래 식에 의해 계산된 신뢰영역 밖에 놓이는 관측지점들로 수질 관측망을 구성하면 된다.

$$\left\{ y_{ij}; \sum_{l=m+1}^k \frac{y_{il}^2}{\lambda_l} \leq \chi_{\alpha}^2(k-m) \right\} \quad (2)$$

수질 측정망을 구성하는 관측지점의 개수는 신뢰수준의 크기에 의해 결정하게 되는데, 신뢰수준을 크게 하면 신뢰 영역이 좁아져 보다 많은 관측지점을 측정하게 된다.

### 3. 실증분석

#### 3.1 데이터 설명

낙동강 댐 유역의 일부 구간인 상주부터 구미까지 유역을 4km 등간격의 200개 구간으로 나누고(<그림 1> 참고) 각 구간의 종단지점을 초기 관측지점으로 하여 수질 관련 주요 측정항목 11개(T2, Gen1, ISS1, PO4, NH4, NO3, BOD, DO, TN, TP, CHLA)를 측정하였다(2008년 측정 데이터). 상류지역의 처음 27개 관측지점들은 수질의 불안정으로 인하여 실증분석 데이터에서 제외하고 173개 관측지점 데이터를 활용하여 분석하였다.

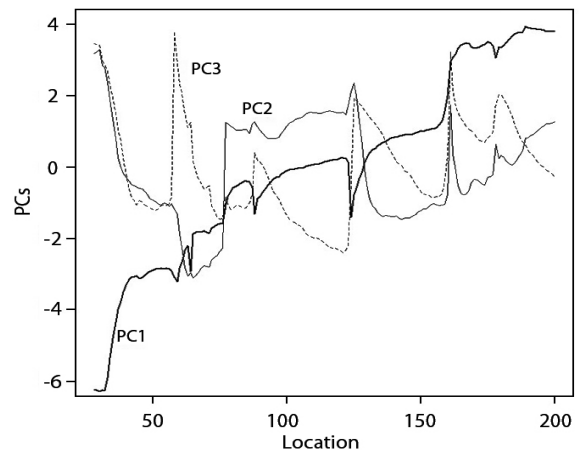


<그림 1> 낙동강 댐 유역

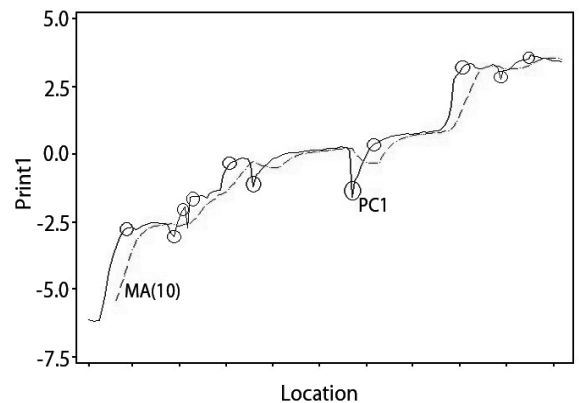
#### 3.2 주성분 점수 구하기

제일 주성분의 고유치는 5, 74, 11개 수질 측정항목들의 총변동에 대한 설명 기여율은 52.2%, 제이 주성분은 고유치 2.09 변동 기여율 19.0%, 그리고 제삼 주성분은 고유치 2.00 변동 기여율은 18.2%로(주요 주성분들의 누적 변동 기여율 89.3%) 주요 주성분의 개수는 3개이다. 제 2절에서 보인 것처럼 각 주요 주성분은 자유도 1인  $\chi^2$ -분포를 따르고 주요 주성분의 합은 자유도 3인  $\chi^2$ -분포를 따른다. 그리고 잔차 주성분의 합은 자유도 8인  $\chi^2$ -분포를 따른다. <그림 2>는 주요 주성분점수들과 관측지점 순차적 산점도이다. 이를 보면 각 주요 주성분의 값이 변하는 변곡점의 관측지점이 서로 다를 수 있다.

개별 주요 주성분에 의해 이상 관측지점을 진단하는 방법은 다음과 같다. 주성분 점수의 변화가 발생하는 변곡점의 관측지점들로 수질 관측망을 구성하거나, 이동평균 예측값과 관측값의 차이인 잔차가 일정 크기 이상인 관측지점들을 관측 수질망을 구성하면 된다. <그림 3>은



<그림 2> 관측지점과 주요 주성분점수 산점도



<그림 3> 관측지점과 주기 10 이동평균 예측치와 제일 주성분 산점도

주기 10의 이동평균 예측치와 제일 주성분 관측치의 순차적 산점도이며 잔차의 크기가 일정 수준 이상인 관측지점을 원으로 표현하였는데, 이들로 수질 측정망을 구성할 수 있다.

관측지점들의 수질 유사성 측정과 군집분석을 개념을 활용하여 수질 측정망을 구성하는 기존의 연구 방법들의 한계를 보이기 위하여 이요상 등[6]의 제안 방법을 적용하여 낙동강 유역의 수질 측정망을 구성하였다. 주성분들의 유클리디언 거리와 평균연결법을 활용한 개체군집 결과인 덴드로그램(Dendrogram, <그림 4> 참고)에 결과에 의하면 관측지점 39~57까지와 관측지점 65~76까지 수질이 유사한 관측지점(관측지점 58~64는 하나의 다른 군집)으로 군집되었다(군집의 개수 10개 사용). 군집분석 결과의 군집 이름을 표현한 제일 주성분, 제이 주성분과 관측지점 산점도(<그림 5> 참고)를 보면 수질 유사 그룹으로 군집화 된 관측지점들은 여러 관측 구간에 속하고 군집 내의 대표 관측지점도 구간 내 가운데

위치가 아닌 구간 시작이나 끝 부분에 위치하고 있음을 알 수 있다.

### 3.3 이상 관측치 판별

주성분은 서로 독립이고  $\chi^2(1)$ -분포를 따르므로 개별 주요 주성분을 이용하여 수질의 특성이 유의적으로 다른 관측지점, 즉 이상 관측지점을 판별할 수 있다. <표 1>은 3개 주요 주성분에 의해 진단된 이상 관측지점이다. 신뢰수준은 90%를 사용하였다.

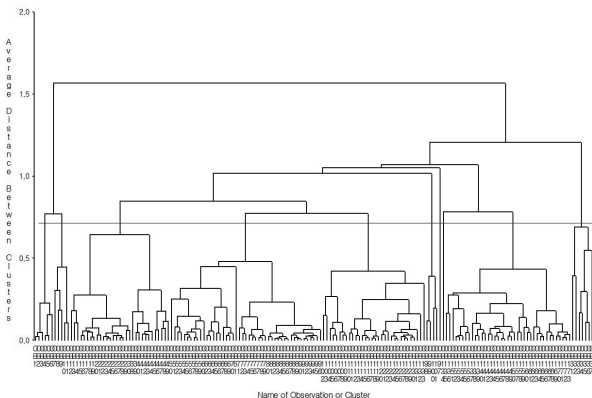
<표 1> 수질차이 유의 관측지점(주요 주성분)

주성분	제일	제이	제삼
관측지점	28~36	28~32 61~74	28~33 58~59 161

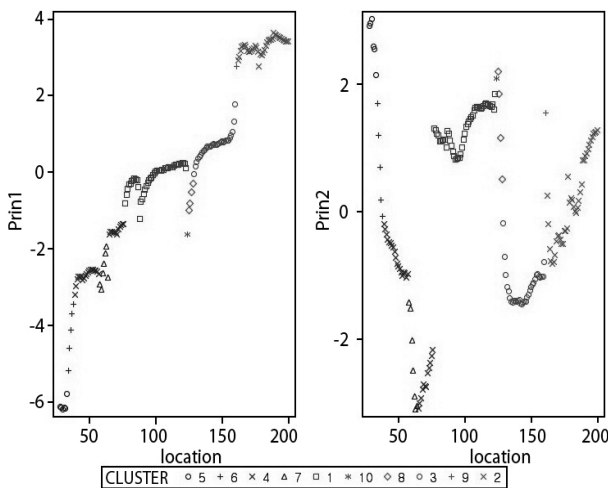
주요 주성분들의 합과 잔차 주성분들의 합은 각각  $\chi^2(3)$ , 외  $\chi^2(8)$ 을 따르므로 이를 이용하여 이상 관측지점을 진단한 결과는 <표 2>에 요약하였다(유의수준 10%). <그림 5>는 주요 주성분과 잔차 주성분을 활용한 90% 신뢰영역을 표현한 것이다. 90% 신뢰영역을 벗어난 이상 관측지점, 즉 수질 측정망으로 구성되는 관측지점은 34, 58, 64, 78, 88, 127, 162, 그리고 200번째 관측지점이다.

<표 2> 수질 차이 유의 관측지점

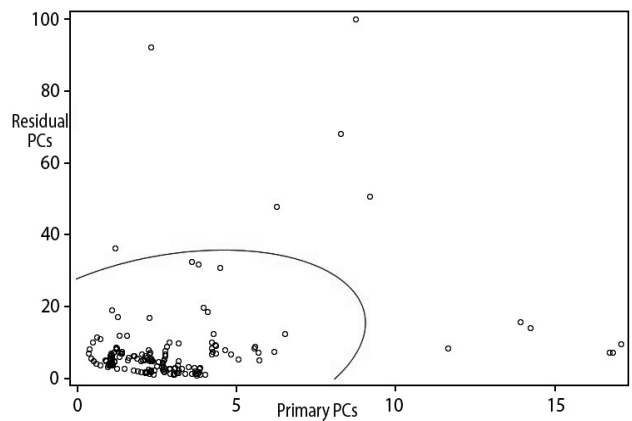
주성분	주요	잔차
관측지점	28~34 58, 59, 64 64, 161	32, 34 58, 64 75~78, 88 124~127 134~135



<그림 4> 평균결합방법 덴드로그램



<그림 5> 관측지점과 주요 주성분점수 산점도(군집결과)



<그림 6> 주요주성분과 잔차 주성분 산점도(90% 신뢰수준 타원)

#### 4. 향후 연구 및 결론

관측지점의 수질 유사성 측정과 군집화 기법에 의해 수질 측정망을 구성하는 기존 연구방법들은 저수지 유역에 적용 가능하나 상류 수질이 하류로 전달되는 하천 지역의 수질 측정망을 구성하는데 한계가 있음을 제 2절에서 기술하였다. 본 연구에서는 주성분의 이상점 진단 방법을 활용하여 이상 관측지점으로 수질 측정망을 구성하는 방법을 제안하였다. 낙동강 댐 유역 173개 관측지점 11개 수질 데이터를 이용하여 기존 연구들의 한계를 보였으며 제안된 접근방법을 활용하여 수질 측정망을 구성하는 방법을 보이기 위한 실증분석을 실시하였다. 수질 관측망은 주요 주성분들 각각의 변곡점에 위치한 관측지점,  $\chi^2$ -분포 신뢰구간을 벗어나는 관측지점들이나, 이동평균법 예측치와의 잔차가 일정 크기 이상인 관측지점들로 구성하였다. 잔차 주성분의 합의 신뢰영역을 밖에 위치한 관측지점들이나 잔차 주성분 합과 주요 주성분 합의 신뢰 타원 영역을 벗어난 관측지점들로 수질 측정망을 구성하는 방법을 실증분석에서 보였다.

본 연구의 실증분석에서는 대상 유역 전체 관측지점 수질 데이터를 한꺼번에 분석하여 얻은 수질 이상 관측지점들로 수질 측정망을 구성하였다. 이전 관측지점들과 수질이 유의적으로 바뀌는 관측지점을 판별하여 이들로 수질 측정망을 구성할 필요가 있을 때는 상류 첫 관측지점부터 일정 개수의 관측지점들을 활용하여 이상 관측지

점을 판별하여 수질 측정망을 구성하면 된다. 처음 10개 관측지점 수질 데이터를 이용하여 이상 관측지점을 판단한 결과 37번째 관측지점이 제일 주성분에 의해 이상 관측지점으로 진단되므로 이를 관측망에 포함하면 된다.

#### 참고문헌

- [1] 경민수, 김상단, 김형수, 박석근; “통계적 기법을 이용한 경안천 유역의 수질 측정망 구성”, 대한토목학회논문집, 26(3B) : 291-300, 2006.
- [2] 권상우, 유철상; “엔트로피 방법을 이용한 수질측정망의 평가”, 대한토목학회논문집, 21(5B) : 501-509, 2001.
- [3] 권세혁, 이요상; “수질 관측지점 유사성 측정방법연구”, 한국통계학회논문집, 17(1) : 39-45, 2010.
- [4] 박수영, 왕수균, 최정현, 박석순; 유전자 알고리즘을 이용한 낙동강 유역의 수질 측정망 설계에 관한 연구, 한국물환경학회지, 23(5) : 697-704, 2007.
- [5] 왕수균, 나은혜, 박석순; 한강수계 수질측정망 개선을 위한 목적 지향 설계 방안에 관한 연구, 대한환경공학회지, 27(5) : 453-460, 2005.
- [6] 이요상, 권세혁, 이상욱, 반영진; “저수지 최적 수질 측정망 구축시스템 개발 및 적용”, 한국수자원학회논문지, 44(4) : 295-304, 2010.
- [7] Johnson, R. A. and Wichern, D. W.; Applied multivariate statistical analysis, Prentice Hall, 2007.