

수질 관측지점 유사성 측정방법 연구

권세혁^{1,a}, 이요상^b

^a한남대학교 비즈니스통계학과, ^bK-water연구원 수자원연구소

요약

환경가치가 높아짐에 따라 하천 수질에 대한 관심의 증대로 수질측정망 연구가 최근 활발히 진행되고 있으나 입지환경의 지리적 특성이나 유입량, 유출량, 유량, 유속과 같은 하천 특성 중심 연구이다. 본 연구에서는 상대적으로 연구가 미미한 관측지점의 수질 유사성을 측정하는 방법으로 수질의 시계열 패턴을 고려할 수 있는 상관계수행렬 방법을 제안하고 기존의 주성분점수를 이용한 방법과 비교하였다. 용담댐에서 2년간 조사된 수질관련 데이터를 이용하여 두 방법에 대한 실증분석을 실시하여 관측지점의 지리적 특성에 의해 분류된 결과와 본 연구에서 제안된 방법에 의해 관측지점 유사성을 측정하여 얻은 군집결과가 더 일치함을 보였다.

주요용어: 수질측정망, 유사성, 군집분석, 상관계수, 다차원척도법.

1. 서론

경제성장에 따른 오염배출량의 급격한 증가로 국내 하천의 수질 오염이 심화됨에 따라 1977년부터 주요 수질오염 물질에 대한 농도규제를 시작하면서 정부의 제도적인 수질관리가 시작되었다. 그리고 원수 수질에 대한 관심과 환경보전에 대한 국민의 기대가 커짐에 따라 국가는 수질측정망을 운영하게 되었다. 수질측정망은 수질보전대상 공공수역에 대한 수질현황을 종합적으로 파악하여 대상 하천의 수질변화 추세를 파악하고 적절한 수질 보전 정책수립을 위한 기초 데이터를 확보하기 위해서 운영되고 있다. 1970년대 후반부터 수질측정망에 대한 정기적 조사가 이루어지기 시작하여 1993년에는 1,348개, 2008년에는 1,476개 지점으로 확대 운영되고 있다.

건설교통부 수문관측 매뉴얼 (건설교통부, 2004)에는 수질측정을 위한 채수 지점을 선정할 수 있다고만 적고 있을 뿐 지점 선정 방법이나 절차에 대한 내용은 없다. 그리하여 수질측정망 구성 시 수질 관리자의 경험과 주관적 판단에 의해 수질관측 지점들이 선정되고 있는 실정이다. 이로 인하여 수질이 유사한 지역들이 중복 측정되거나 수질 패턴이 상이한 지점이 누락되는 경우가 발생하게 되어 수계 관리를 위해 요구되는 다양한 수질 정보를 효율적으로 생산하는데 한계와 문제점을 내포하고 있다.

측정환경의 복잡성과 측정되어야 할 수질 항목의 다양성으로 인하여 수질 관련 데이터 수집은 쉽지 않다. 또한, 수질은 대상 하천의 환경변화에 영향을 받으므로 현재 상태에서 적절히 구성된 측정망도 시간이 지나면서 부적절한 측정망으로 변화될 수 있다. 이러한 문제로 수질측정망의 설계와 평가는 쉽지 않으며, 특히 규모가 크고 복잡한 유역일 경우에는 더욱 어렵게 된다. 이로 인하여 수질측정망 연구는 측정 목적, 입지환경, 지류 수, 유입량과 유출량, 유량과 유속 등을 고려하는 모형이나 방법론 중심으로 진행되어 왔다. 박수영 등 (2007)은 유전자 알고리즘을 활용하여 낙동강 수계에 대한 과학적인 수질측정망 설계방법을 제시하였고 경민수 등 (2006)은 시뮬레이션 데이터와 수질모형을 이용한 최적

본 논문은 한남대학교 2009 학술연구비 지원 결과임.

¹ 교신저자: (306-791) 대전시 대덕구 오정동 한남대학교 비즈니스통계학과, 교수. E-mail: wolfpack@hnu.ac.kr



그림 1: 용담호 관측지점

화 기법을 적용하여 수질관측망을 구성하였다. 왕수균 등 (2005)은 수질특성 파악, 새로 도입된 물관리 제도 및 기술에 대한 지원이 가능한 한강수계 수질측정망 설계기법을 제시하였고 권상우와 유철상 (2001)은 엔트로피 방법을 이용한 대청댐 유역의 수질측정망 평가방법을 제시하였다.

수질관측망을 효율적으로 구축하려면 관측지점의 수질 항목들의 패턴을 고려한 관측지점의 수질 유사성(similarity)을 측정하는 연구도 병행되어야 한다. 그러나 이런 접근방법 연구는 최근 Kazi 등 (2009)에 의해 발표되었을 뿐이다. Kazi 등 (2009)은 5개 관측지점에 대한 수질 관련 36개 측정치를 활용하여 주성분 점수를 얻고, 이를 이용하여 관측지점에 대한 군집분석을 실시하여 유사 관측지점을 판단하였다. 그러나 이 접근 방법에서는 수질 월별 패턴 변화를 고려하지 못하는 문제를 안고 있다. 본 연구에서는 수질의 시계열 패턴을 고려할 수 있는 상관계수행렬을 활용하여 관측지점 유사성을 측정하고 유사 관측지점을 다차원척도법에 의해 분류하는 접근방법을 제안하였다.

서해안 지역의 안정적 용수공급과 수자원의 효율적 개발을 목적으로 금강 상류에 유역변경식 댐인 용담댐을 2001년에 준공하였고, 담수 후 용담댐 수질이 어느 정도 안정기에 접어든 2005년부터 수질 항목들을 월별로 조사하기 시작하였다. 이를 본 연구의 실증데이터로 사용하여 두 방법을 비교·분석하였다. 2절에서는 실증 데이터로 사용된 용담호 2년간 데이터 특성을 기술하였고, Kazi 등 (2009) 방법과 본 연구에서 제안된 방법을 적용한 실증분석 결과와 비교는 3절에 정리하였다. 그리고 4절에는 결론을 보였다.

2. 실증 데이터

2000년 12월에 완공되어 2001년부터 담수를 시작한 용담댐은 초기 담수에 의한 수질영향이 사라진 2005년부터 저수지 수질이 안정기에 접어든 것으로 보인다. 그림 1은 계획홍수위 기준으로 36.2km²인 용담댐 저수지 지도에 수질관측망(2005년 기준) 관측지점 10개와 하천 흐름을 화살표로 표

시한 것이다. 관측지점은 댐 앞 1개(1번) 지점에서부터 시작하여 주하천 유입 지류에서 2개(2, 3번) 지점, 저수지 분류에서 6개(4, 5, 6, 7, 9, 10번) 지점 그리고 진안천 지류에 1개(8번) 지점이었으며 관측된 수질 항목은 19개였다. 분석에 사용된 실증 데이터는 2005년, 2006년에 1월부터 12월까지 월 1~4회 씩 2개의 관측지점 상층, 하층에서 측정된 7개 주요 수질항목에 대한 관측치이다 (한국수자원공사, 2005, 2006).

용담호의 수심은 갈수기인 3월경에는 저수량이 가장 적어 댐 앞 30여m 정도로 낮아지며 강우기가 지난 후인 8월경에는 최대로 저수되어 최대 50여m에 이른다. 수온 특성을 살펴보면 동절기인 1, 2월에는 전층이 4℃ 정도이다가 기온이 상승하면 표층부터 수온이 상승하여 하절기인 7월에 25℃, 8월에는 30℃ 이상까지 표층 수온이 상승하는 것으로 조사되었다. 9월 이후부터는 대기온도와 함께 댐의 수온도 떨어지게 되면 수심이 낮은 상류부터 전도현상이 발생하여 수심이 가장 깊은 댐 앞에서는 이듬해 1월경에 완전한 전도가 발생한다.

2년간 용담 저수지의 수질 특성은 분석·정리하면 다음과 같다. 물의 투명 정도를 나타내는 SD (Secchi Depth)의 경우 하천의 주요 흐름이 이루어지는 지점(1, 2, 4, 5, 6, 7번)들은 봄철에 투명도가 가장 좋게 나타나고 하절기에 투명도가 감소하는 전형적인 저수지 수질변화 패턴을 유지한다. 반면, 지류유입 하천에 영향을 많이 받는 3, 8, 9번 지점은 연중 투명도가 크게 개선되지 않았다. SS(Suspended Solid, 부유물질) 농도는 저수지에서는 평수기에 모든 지점에 걸쳐 큰 변화를 보이지는 않으나 강우기에 유입되는 탁수가 장기간 저수지에 머무는 동안에는 다소 높게 나타났다. CODMn(Chemical Oxygen Demand), T-N(Total Nitrogen) 그리고 Chl-a(Chlorophyll-a)은 일반 댐에서 관측될 수 있는 계절별 변화를 나타내었다. T-P(Total Phosphorus)와 PO₄-P(인)도 거의 유사한 계절별 수질변화를 나타냈으나 하절기에 SS에 부착되어 유입되는 특성으로 인하여 SS처럼 하절기에 높게 측정되었고 하류로 갈수록 급격히 농도가 낮아지는 현상을 나타내고 있다. 이처럼 용담댐에서 측정되는 수질 관련 관측치들은 지리적 차이는 물론 시계열 패턴 변화도 상이함을 보이고 있음을 알 수 있다.

3. 관측지점 유사성 분석

Kazi 등 (2009)는 2년(2005년, 2006년)동안 5개 각 관측지점에서 36개 수질 항목을 3번 반복 측정하여 얻은 데이터를 활용하여 실증 데이터로 활용하였다. 5개 관측지점을 개체로 하고 36개 수질항목에 대한 관측치(각 관측지점에 대하여 연도, 관측지점, 반복 관측치의 평균)를 측정변수로 하여 주성분 분석을 실시하여 3개(총 변동 기여율 97.6%) 주성분점수(Principal Component Score)를 얻었다. 이 주성분 점수들을 이용하여 계산된 관측지점 간 유클리디안 거리를 유사성 값으로 활용하여 Ward 연결방법에 의한 군집분석을 실시하여 관측지점들을 군집화 하고 설명하였다.

Kazi 등 (2009) 방법은 관측지점의 유사성 값을 계산할 때 앞 절에서 언급한 수질 항목들의 시계열 특성 및 차이를 고려하지 못하는 문제점을 갖고 있다. 이에 대한 해결책으로 본 연구에서는 각 수질의 월별 패턴의 차이를 고려할 수 있는 통계량인 상관계수를 활용한 접근방법을 제안하였다. 각 수질별로 관측지점 간 상관계수행렬들을 구하고, 이들 각 상관계수 행렬에서 대응원소들의 기중평균으로 단일 상관계수행렬을 구한다. 산술평균 대신 기하평균을 사용한 이유는 상관계수의 값은 퍼센트(%)로 측정된 변수와 유사한 속성을 갖고 있기 때문이다. 단일 상관계수행렬을 유사성 행렬로 활용하여 다차원 척도법(Multi-Dimensional Scaling)을 적용하여 성분 좌표를 얻고 좌표 간 유클리디안 거리를 관측지점 유사성 측정치로 하여 관측지점을 군집화 하는 방법을 제안하였다.

2절에서 기술한 용담호 데이터 중 일부 데이터에 Kazi 등 (2009) 방법과 본 연구의 제안방법을 적용하여 실증분석하고 비교하였다. 수집된 데이터 중 관측지점 10번은 수심이 낮아 관측위치 중 일부만 관측되는 시점이 많아 제외하고, 3개 관측위치(상층, 중층, 하층) 중 수심에 따라 측정이 불가능한 중

표 1: 주성분분석 결과

수질 항목	PC1	PC2
SD	-0.27	-0.36
SS상	0.30	-0.11
SS하	0.29	-0.06
CODMn상	0.13	-0.76
CODMn하	0.28	-0.02
T-P상	0.30	-0.11
T-P하	0.30	-0.02
PO4-P상	0.29	-0.20
PO4-P하	0.29	0.05
T-N상	0.29	-0.12
T-N하	0.29	0.14
Chl-a상	0.29	0.07
Chl-a하	0.25	0.44
고유치	10.84	1.33
변동기여율 (%)	83.40	10.20
누적기여율 (%)	83.40	93.60

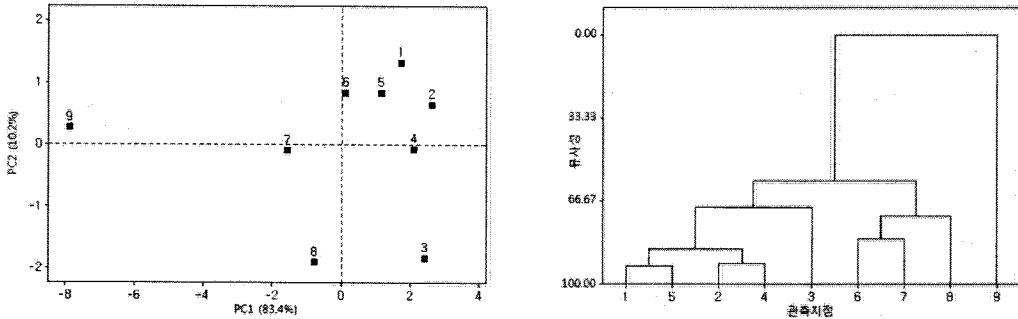


그림 2: 주성분 산점도와 완전연결법에 의한 덴드로그램

층을 제외하였다. 그리고 19개 수질 측정치 중 주요 7개 지표(SD, SS, CODMn, T-P, PO4-P, T-N, Chl-a)만을 이용하여 유사성을 계산하였다. 수질 관련 변수에 대한 분석 결과 동일 수질에서도 상층과 하층 관측지 패턴이 서로 상이하어 개별적인 측정변수로 사용하였다. 그러나 주요 수질 변수 7개, 관측지점 2개의 결합조건인 14이지만 투명도 측정항목인 SD의 경우에는 상층만 측정하기 때문에 측정변수는 총 13개이다. 그러므로 실증데이터에서 개체는 9개의 관측지점이고 측정변수는 13개 수질관련 항목이다.

Kazi 등 (2009)의 접근방법에 의해 관측지점의 유사성을 분석한 결과는 다음과 같다. 수질항목의 측정단위가 상이하므로 13개 수질항목 변수들의 상관계수행렬을 이용하여 고유치를 구하고 그에 대응하는 고유벡터를 선형계수로 얻어 표 1에 정리하였다. 관측치는 2년, 12개월, 반복 관측치들의 평균이다. 제1 주성분까지의 누적 원 변수 변동 기여율이 93.6%이므로 군집분석에 사용할 주성분 점수는 2개 성분에 대해서만 얻었다. 그림 2는 관측지점이 표시된 주성분점수의 산점도와 제1, 제2 주성분 점수로부터 유클리드 거리를 계산하고 이를 관측지점의 유사성으로 측정하여 완전연결방법(complete linkage)에 의해 얻은 덴드로그램(Dendrogram)이다. 덴드로그램에 의하면 관측지점 (1, 2, 4, 5), (6, 7, 8) 그리고 3과 9는 개별적으로 분류되 9개의 관측지점은 4개의 군집으로 분류하는 것이 적절해 보인다.

표 2: 관측지점 단일 상관계수행렬

관측지점	1	2	3	4	5	6	7	8	9
1	1.00	0.76	0.43	0.84	0.82	0.67	0.63	0.53	0.26
2		1.00	0.56	0.73	0.73	0.63	0.55	0.60	0.33
3			1.00	0.35	0.52	0.54	0.51	0.57	0.50
4				1.00	0.83	0.75	0.69	0.56	0.32
5					1.00	0.83	0.73	0.59	0.42
6						1.00	0.87	0.58	0.57
7							1.00	0.66	0.54
8								1.00	0.69
9									1.00

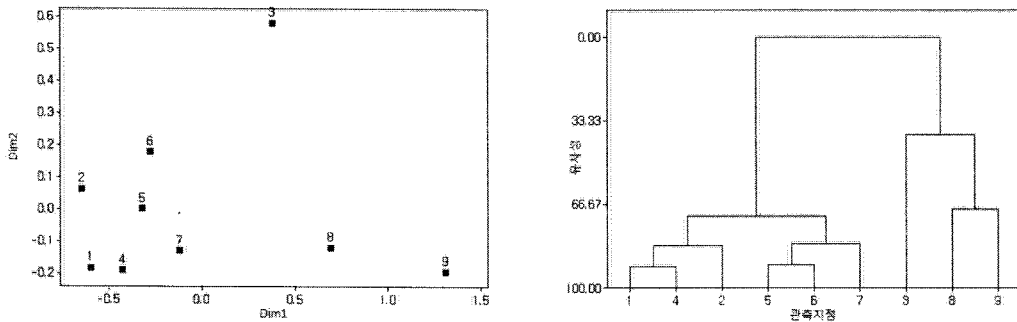


그림 3: 다차원척도법 산점도와 완전연결법에 의한 덴드로그램

표 3: 유사성 군집 결과

방법	관측지점	유출지점			저수지 특성			유입지점		
		1	2	4	5	6	7	3	8	9
Kazi 등 (2009) 주성분 점수		□	□	□	□	○	○	△	○	▽
상관계수 유사성 행렬		□	□	□	○	○	○	▽	△	△

Kazi 등 (2009) 접근방법은 관측치를 시계열 상에서 통합화(agggregation) 한 평균을 관측치로 사용하여 수질항목들의 시계열 패턴 변화를 고려하지 못하고 있다. 그러므로 본 연구에서는 수질항목의 시계열 변화를 고려할 수 있는 통계량인 상관계수를 이용하여 관측지점의 유사성을 계산하는 방법을 제안하였다. 13개 수질 항목 각각에 대하여 관측지점간 상관계수를 구하고, 각 상관계수행렬에서 대응하는 13개 원소들의 기하평균 값을 계산하여 단일 상관계수행렬을 얻었다(표 2 참조).

표 2의 상관계수행렬을 유사성행렬로 하여 다차원척도법을 적용하여 관측지점의 상대적 위치에 대한 좌표를 얻었다. 성분 1과 성분 2의 좌표 값을 이용하여 관측지점 간 유클리디안 거리를 계산하고 이를 유사성으로 활용하여 완전연결법에 의해 관측지점에 대한 계층적 군집분석을 실시하였다. 그림 3은 다차원척도법 산점도(스트레스 값 = 2.1%)와 완전연결법에 의한 덴드로그램이다. 덴드로그램을 활용하여 유사한 관측지점 군집화 하면 (1, 2, 4), (5, 6, 7) 그리고 3, (8, 9)으로 나눌 수 있다.

표 3은 Kazi 등 (2009) 방법과 본 연구에서 제안한 방법에 의해 군집화된 결과를 정리한 것이다. 하천은 지리적 특성에 의해 하천이 유입되는 지점, 저수지 특성을 갖는 지점 그리고 유출되는 지점으로 나눌 수 있는데 이 지리적 특성은 수질들의 측정치와 밀접한 관계가 있다. 표 3에서 볼 수 있듯이 본 연

구에서 제안한 상관계수를 이용한 유사성 측정방법에 의한 군집분석 결과가 지리적 특성의 관측지점 그룹과 보다 더 일치함을 알 수 있다. 이는 제안된 접근 방법이 수질의 시계열 변화 패턴을 고려하여 관측지점의 유사성을 측정하였기 때문으로 판단된다.

4. 결론

수질관련 변수들을 이용하여 관측지점간의 유사성을 측정하는 방법으로 상관계수행렬 접근방법을 제안하고 용담댐 수질 데이터를 활용하여 기존의 주성분점수를 활용한 방법과 비교·분석하였다. 본 연구에서는 수질별 관측지점 상관계수행렬을 이용하여 유사성을 측정함으로써 수질의 시계열 변화까지 고려하여 관측지점 간 유사성을 측정할 수 있는 방법을 제안하였다. 각 수질들의 상관계수행렬로부터 대응하는 원소들의 기중평균을 계산하여 얻은 관측지점 간 단일상관계수행렬을 관측지점 유사성행렬로 하여 다차원척도법을 적용한다. 다차원척도법의 성분 좌표를 군집변수로 하여 계층적 군집분석 적용하여 유사한 관측지점 분류하였다.

실증분석 결과 Kazi 등 (2009)이 제안한 주성분점수 방법은 수질 데이터의 시계열 패턴을 고려하지 못하여 관측지점에 대한 유사성 군집결과가 하천의 지리적 특성과 일치하는 정도가 낮았다. 하천의 지리적 특성인 하천 유입 지점, 저수지 특성, 하천 유출 지점의 분류와 본 연구에서 제안한 상관계수행렬 활용 접근방법의 군집분석 결과와 일치함을 보였다.

본 연구결과는 하천에서 수질 유사지역에 대한 정보는 수질 유사지점 중복 측정이나 상이지점의 누락을 방지하여 최소의 관측으로 하천 전체의 수질을 평가할 수 있는 최적 수질측정망 구축에 기초 자료로 활용될 수 있다. 그리고 제안된 연구방법을 활용해 새롭게 구축된 측정망은 저수지 수질현황 파악을 정확하고 용이하게 하며 장래 수질보전정책 수립을 위한 기초자료로 활용이 가능할 뿐만 아니라 측정된 자료는 다양한 연구나 업무에 활용 될 수 있을 것으로 판단된다.

참고 문헌

- 건설교통부 (2004). 수문관측 매뉴얼
- 경민수, 김상단, 김형수, 박석근 (2006). 통계적 기법을 이용한 경안천 유역의 수질 측정망 구성, <대한토목학회논문집>, **26**, 291-300.
- 권상우, 유철상 (2001). 엔트로피 방법을 이용한 수질측정망의 평가, <대한토목학회논문집>, **21**, 501-509.
- 박수영, 왕수균, 최정현, 박석순 (2007). 유전자 알고리즘을 이용한 낙동강 유역의 수질 측정망 설계에 관한 연구, <한국물환경학회지>, **23**, 697-704.
- 왕수균, 나은혜, 박석순 (2005). 한강수계 수질측정망 개선을 위한 목적 지향 설계 방안에 관한 연구, <대한환경공학회지>, **27**, 453-460.
- 한국수자원공사 (2005, 2006). 용담댐 일원 하천유량측정 등 수문조사보고서
- Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, R. A., Baig, J. A. and Shah, B. Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study, *Ecotoxicology and Environmental Safety*, **72**, 301-309.

Similarity of Sampling Sites by Water Quality

Sehyug Kwon^{1,a}, Yosang Lee^b

^aDepartment of Statistics, Hannam University; ^bPrincipal Researcher, K-water Institute

Abstract

As the value of environment is increasing, the water quality has been a matter of interest to the nation and people. Research on water quality has been widely studied, but focused on geographical characteristic and river characteristics like inflow, outflow, quantity and speed of water. In this paper, two approaches to measure the similarity of sampling sites by using water quality data are discussed and compared with two-years empirical data of Yongdam-Dam. The existing method has calculated their similarities with principal component scores. The proposed approach in this paper use correlation matrix of water quality related variables and MDS for measuring the similarity, which is shown to be better in the sense of being clustering which is identical to geographical clustering since it can consider the time series pattern of water quality.

Keywords: Water quality, similarity, clustering analysis, correlation coefficient, Multi-Dimensional Scaling.

This work is supported by Hannam University Research Fund of 2009.

¹ Correspondence author: Professor, Department of Statistics, Hannam University, Daejeon 306-791, Korea.
E-mail: wolfpack@hnu.ac.kr