

다변량 통계분석을 이용한 저수지증발량 영향인자에 관한 연구

A Study of Influence Factors for Reservoir Evaporation Using Multivariate Statistical Analysis

이경수*, 광성현**, 서용재***, 류시완****

Kyungsu Lee, Sunghyun Kwak, Yong Jae Seo, Siwan Lyu

요 지

지구온난화로 인해 세계 곳곳에서 기온상승이 관측되고 있으며, 이는 전지구적 기후시스템의 변화를 보여주는 대표적인 예이다. 온도를 비롯한 강수량, 풍속, 증발량 등의 기상학적, 수문학적 인자들이 각각 서로에게 영향을 주고 받으며 복잡하게 변화할 것이고, 그 변화폭도 점점 커질 것이다. 증발에 영향을 미치는 인자들은 크게 세 가지로 나뉘는데, 태양복사에너지, 온도, 바람, 기압, 습도와 같은 기상학적인자, 증발표면의 특성인자 그리고 수질인자로 분류할 수 있다. 증발에 영향을 주는 인자들은 예전부터 알려져 있지만 이들 간의 복잡한 상호작용에 대해 정확히 이해하기는 쉽지 않다. 본 연구에서는 댐유역의 증발량에 영향을 미치는 기상인자 파악을 위해 2008년부터 2016년까지 관측된 낙동강수계 내 안동댐과 남강댐의 기상자료(기온, 강수량, 풍속, 상대습도, 기압, 일사량, 일조시간, 전운량)를 이용한 변화를 분석하였으며, 다변량 통계기법인 요인분석을 통해 증발량과 상관성이 높은 인자들을 분류하였다. 안동댐과 남강댐 공통적으로 증발량과 기온, 기압이 같은 요인으로 분류되고 높은 상관성을 보였으며, 강수량, 일조시간, 일사량, 전운량이 같은 요인으로 분류되었다. 국내의 증발량 측정지점에 대한 추가적인 분석과 영향인자를 이용한 다변량회귀식과 인공신경망 통해 증발량 미측정 지점의 증발량 산정이 가능할 것으로 판단된다.

핵심용어 : 저수지증발량, 요인분석, 영향인자, 기상자료

1. 서론

최근 지구온난화로 인해 세계 곳곳에서 기온상승이 관측되고 있으며, 이는 전지구적 기후시스템의 변화를 보여주는 대표적인 예이다. 온도를 비롯한 강수량, 풍속, 증발량 등의 기상학적, 수문학적 인자들이 각각 서로에게 영향을 주고 받으며 복잡하게 변화할 것이고, 그 변화폭도 점점 커질 것이다. 증발에 영향을 미치는 인자들은 크게 세가지로 나뉘는데, 태양복사에너지, 온도, 바람, 기압, 습도와 같은 기상학적인자, 증발표면의 특성인자 그리고 수질인자로 분류할 수 있다. 증발에 영향을 주는 인자들은 예전부터 알려져 있지만 이들 간의 복잡한 상호작용으로 인해 정확히 이해하기는 쉽지 않다. 정대일 등(2009)은 1960년부터 2007년까지 관측된 국내 기상관측소의 팬증발량과 기상인자들에 대한 변화 분석을 통해 팬증발량의 감소현상을 확인하였으며, 임창수 등(2011)은 국내 5개 댐유역에 대해 물수지분석을 통해 산정된 증발산량과 다변량회귀식으로 산정된 증발산량의 비교를 통해 타당성을 검토하였다. 기존 연구는 제한된 연구지역에 대해서 제한된 기간의 자료를 이용하고 있으며, 다양한 지형적 조건을 고려한 증발산 산정방법들의 평가 연구는 미흡한 실정

* 정회원 · 창원대학교 친환경플랜트FEED공학과과정 박사과정 · E-mail : kidhan@changwon.ac.kr

** 정회원 · 창원대학교 토목공학과 박사과정 · E-mail : grestddk@gmail.com

*** 비회원 · 창원대학교 친환경플랜트FEED공학과과정 석사과정 · E-mail : sjlbsid@gmail.com

**** 정회원 · 창원대학교 토목환경화공융합공학과 교수 · E-mail : siwan@changwon.ac.kr

이다. 본 연구에서는 증발량에 영향을 미치는 기상 인자 중 주요 영향인자 파악을 위해 다변량 통계분석 중 하나인 탐색적 요인분석을 이용한 분석을 실시하였다.

2. 본론

2.1 탐색적 요인분석

요인분석은 여러 변수들 간의 공분산과 상관관계 등을 이용하여 변수들 간의 상호관계를 분석하고, 그 결과를 토대로 문항과 변수들 간의 상관성 및 구조를 파악하여 여러 변수들이 지닌 정보를 적은 수의 요인으로 묶어서 나타내는 분석기법이다. 요인분석을 실시하면 여러 변수들에 대한 정보가 몇 개의 핵심 내재 요인으로 간추려진다. 이렇게 되면 정보를 좀 더 쉽게 이해할 수 있으며, 추가 분석을 진행하기도 쉽다. 그러나 이와 같은 요인분석의 장점에도 불구하고 산출된 요인이 임의성을 띠고 있을 경우에는 요인해석이 어려울 수 있으므로, 간추려진 분석 결과에 대한 타당성과 신뢰성 검증에 주의를 기울여야 한다. 요인분석은 표본 전체의 변수에 공통적인 요인이 있다고 가정하고 이 요인을 찾아내어 각 변수가 어느 정도 영향을 받고 있는지 그 정도를 산출하기도 하고 그 집단의 특성이 무엇인가를 설명하려는 다변량 통계분석 방법이다.

요인의 개수를 정하기 위해서는 Kaiser 판단(eigen-value)과 Scree Plot을 이용한 방법이 있다. Kaiser 판단은 변수들의 상관 관계가 0이면(관계가 없으면) 상관관계수 행렬(R)은 항등 행렬 I 이다. 이 경우 원 변수의 개수와 주성분의 개수가 같아지고 주성분의 분산은 모두 1 이므로 각 주성분이 가지는 분산 평균도 1 이다. 그러므로 R 로부터 구한 고유치가 평균인 1 이상이 되어야 한다는 판단 하에 고유치가 1 이상인 것으로 요인의 개수를 정한다. Scree Plot은 총 변동 80 %에 연연하지 않고 주성분 분산 설명 변동의 크기(고유치)가 앞의 값과 비교하여 큰 차이를 보이며 작은 값이 평준화되는 현상을 보일 때, 평준화되기 직전의 값으로 요인의 개수로 정한다.

요인 분석에서 요인의 부하 값은 요인(공통 개념)과 원 변수의 상관관계 정도를 나타내는 크기로 해석될 수 있으므로 부하 값에 의해 원 변수를 그룹화 한다. 그러나 하나의 원 변수에 부하 값이 큰 요인이 2개 이상 존재하거나, 인자의 크기가 0을 중심으로 ±의 작은 값이 있는 경우 부하 값으로 변수를 그룹화 하는 것은 불가능하다. 요인 회전은 각 요인이 상대적으로 큰 부하 값을 갖도록 요인을 회전(rotate)하는 것으로 QUARTIMAX rotation, OBLIQUE rotation, PROMAX rotation 방법 등이 있는데, 가장 많이 사용되는 것은 직교 회전방법인 VARIMAX rotation 방법이다. 이는 Kaiser(1958)가 제안한 것으로 간단한 구조의 측정치로 요인 행렬의 각 열 내의 부하 제곱의 분산의 합을 제한하고 이 분산을 최대화 하는 회전 방법이다(권세혁, 2004).

2.2 연구 방법

본 연구에서는 2008부터 2016년까지 관측된 낙동강 수계의 안동댐과 남강댐 기상자료를 이용하여 탐색적 요인분석을 실시하였다. 분석에 사용된 기상 인자는 기온, 강수량, 풍속, 상대습도, 기압, 일조시간, 일사량, 전운량, 증발량 총 9항목이며, 표 1과 같이 표기하였다. 요인의 수는 고유값(eigenvalue)이 1 이하를 기준으로 판단하거나 누적값(Cumulative)이 0.8 이상이 되는 것으로 결정하였으며, 각 변수별 공통성(Communality) 값으로부터 분류된 요인에 의해 변수의 변동정도를 확인하였다. 그리고 불확실성을 줄여주기 위해 직교회전(Varimax Rotation)한 성분 값으로 요인분석을 실시하였다.

표 1 기상인자 항목 및 표기

| 항목 | 기온 | 강수량 | 풍속 | 상대습도 | 기압 | 일조시간 | 일사량 | 전운량 | 증발량 |
|----|----|-----|----|------|------|------|------|-------|------|
| 표기 | T | P | WS | RH | Atm. | Rad. | Sun. | Cloud | Eva. |

2.3 연구 결과

표 2 고유값(Eigenvalue) 및 누적값(Cumulative)

| No. | 안동댐 | | 남강댐 | |
|-----|-------|-------|-------|-------|
| | 고유값 | 누적값 | 고유값 | 누적값 |
| 1 | 3.801 | 0.422 | 3.254 | 0.362 |
| 2 | 2.705 | 0.723 | 2.605 | 0.651 |
| 3 | 1.012 | 0.835 | 1.133 | 0.777 |
| 4 | 0.724 | 0.916 | 0.768 | 0.862 |
| 5 | 0.287 | 0.948 | 0.558 | 0.924 |
| 6 | 0.216 | 0.972 | 0.227 | 0.950 |
| 7 | 0.111 | 0.984 | 0.211 | 0.973 |
| 8 | 0.080 | 0.993 | 0.160 | 0.991 |
| 9 | 0.064 | 1.000 | 0.083 | 1.000 |

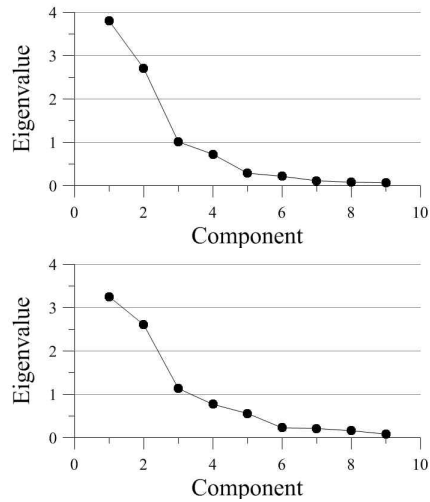


그림 1 Scree Plot 안동댐(위), 남강댐(아래)

표 3 회전된 요인분석 결과(안동댐)

| | Factor1 | Factor2 | Factor3 | 공통성 |
|-------|---------|---------|---------|-------|
| Cloud | 0.840 | 0.173 | -0.257 | 0.802 |
| RH | 0.715 | 0.144 | -0.540 | 0.824 |
| P | 0.663 | 0.119 | 0.178 | 0.486 |
| Sun. | -0.685 | 0.669 | 0.134 | 0.935 |
| Rad. | -0.877 | 0.203 | 0.264 | 0.880 |
| T | 0.177 | 0.918 | -0.239 | 0.932 |
| Eva. | -0.498 | 0.806 | 0.155 | 0.922 |
| Atm. | -0.359 | -0.844 | -0.081 | 0.848 |
| WS | -0.102 | 0.025 | 0.937 | 0.889 |

표 4 회전된 요인분석 결과(남강댐)

| | Factor1 | Factor2 | Factor3 | 공통성 |
|-------|---------|---------|---------|-------|
| Rad. | 0.901 | 0.023 | 0.213 | 0.859 |
| Sun. | 0.784 | 0.469 | 0.246 | 0.896 |
| P | -0.640 | 0.122 | 0.145 | 0.446 |
| Cloud | -0.694 | 0.327 | -0.003 | 0.588 |
| T | -0.097 | 0.923 | -0.123 | 0.876 |
| Eva. | 0.469 | 0.682 | 0.266 | 0.756 |
| Atm. | 0.273 | -0.848 | -0.093 | 0.803 |
| WS | -0.021 | 0.117 | 0.943 | 0.903 |
| RH | -0.569 | 0.461 | -0.575 | 0.866 |

안동댐과 남강댐의 경우 고유값(Eigenvalue)과 Scree Plot을 고려한 결과 각각 3개의 요인으로 분류되었으며, 표 2와 그림 1에 나타나있으며, 요인을 회전하여 각 요인별로 분류된 결과가 표 3과 표 4에 나타나 있다. 각 인자별 부하(loading) 값의 의미는 각 요인이 원 변수를 설명하는 정도(크기)를 나타내며 요인은 변수들의 내재된 관계에서 공통 부분에 해당된다. 그러므로 각 요인에서 부하 값의 절대 값이 큰 변수들을 선택하여 그룹화 하였다(음의 부호는 동일 개념의 반대 척도).

안동댐의 경우 첫 번째 요인으로 전운량(Cloud), 상대습도(RH), 강수량(P), 일조량(Sun.), 일사량(Rad.) 이 같은 요인으로 분류되었고, 두 번째 요인으로 기온(T), 증발량(Eva.), 기압(Atm.)이 분류되었고, 세 번째 요인으로 풍속(WS)이 분류되었다. 첫 번째 요인으로 분류된 Cloud, RH, P와 Sun., Rad.는 음의상관성으로 분류되어 서로 반대의 변동특성을 나타낸다. 두 번째 요인의 T, Eva.와 Atm. 역시 음의 상관성을 보이며, 세 번째 요인으로 분류된 WS은 단일 인자로 분류되어 다른 인자들과 영향이 크지 않음을 나타낸다.

남강댐의 경우 3개의 요인으로 분류되었으며, 첫 번째 요인으로 Rad., Sun., P, Cloud가 분류되었고, 두 번째 요인으로 T, Eva., Atm., 세 번째 요인으로 WS와 Atm.가 분류되었다. 남강댐의 경우 RH가 WS와 음의상관성으로 같은 요인으로 분류되었으나 요인1의 부하량이 -0.569로 요인3의 부하량과 크게 차이가

없는 것으로 안동댐과 다르게 WS의 영향을 더 받는 것으로 나타났다. P의 경우 공통성 값이 두지점 모두 0.486(안동댐), 0.446(남강댐)으로 0.5미만의 값으로 나타나 분류된 요인으로는 P의 영향을 판단하기 어려움을 의미한다.

3. 결론

본 연구에서는 댐유역의 증발량에 영향을 미치는 기상인자 파악을 위해 2008부터 2016년까지 관측된 낙동강수계 내 안동댐과 남강댐의 기상자료(기온, 강수량, 풍속, 상대습도, 기압, 일사량, 일조시간, 전운량)를 이용한 변화를 분석하였으며, 다변량 통계기법인 요인분석을 통해 증발량과 상관성이 높은 인자들을 분류하였다. 안동댐과 남강댐 공통적으로 증발량과 기온, 기압이 같은 요인으로 분류되고 높은 상관성을 보였으며, 강수량, 일조시간, 일사량, 전운량이 같은 요인으로 분류되었다. 국내의 증발량 측정지점에 대한 추가적인 분석을 통해 같은 요인으로 분류되는 공통인자 사이의 관계를 파악하고 주요 영향인자를 이용한 다변량회귀식과 인공신경망 분석을 통해 증발량 미측정 지점의 증발량 산정이 가능할 것으로 판단된다.

감사의 글

이 연구에 참여한 연구자(의 일부)는 「BK21플러스 사업」의 지원비를 받았음

참고 문헌

1. 정대일, 강제원 (2009). 증발량 관련 기후인자와 팬증발량의 변화 분석, 한국수자원학회 논문집, 제42권, 제2호, pp. 117-129.
2. 임창수, 임가희, 윤세의 (2011). 유역 물수지를 이용한 연 실제증발산 산정에 미치는 수문기후 영향 연구, 한국수자원학회 논문집, 제44권, 12호, pp. 915-928
3. 권세혁 (2004). (SAS활용)다변량분석, 자유아카데미 pp. 89~101