

침수위험순위 결정을 위한 요인분석에 관한 연구

Predicting Flood Risk Assessment Classified by Factor Analysis

이의훈* · 최현석** · 김중훈***

Lee, Eui Hoon · Choi, Hyeon Seok · Kim, Joong Hoon

1. 서 론

수자원공학에서 침수위험을 예측하는 것은 많은 연구자들의 관심 대상이다. 본 논문에서는 이러한 침수위험을 예측하기 위해서 산술평균방법, 가중평균방법, 주성분 분석방법을 사용하여 침수위험에 영향을 주는 인자들간의 상관관계분석을 실시하였다. 표준화를 통한 산술평균, 상관계수를 이용한 가중평균과 주성분 분석을 이용해서 위험순위를 예측하였다.

본 연구에서는 재해연보 및 도시계획현황에 근거하여 침수위험예측을 적용하기 위한 기초연구로서, 침수에 영향을 주는 각 인자들을 바탕으로 산술평균방법, 가중평균방법을 실시하여 비교하였다. 산술평균방법은 간단한 방법이지만 각 인자들에 동일한 가중치를 주는 문제가 있고, 가중평균방법의 경우 각 인자들에 각각 다른 가중치를 주기는 하나 많은 변수로 인해 상관계수가 복잡하고 다중공선성(multicollinearity)문제가 발생하는 단점이 있다. 이런 문제점들을 보완하기 위해 각 인자들마다 다른 가중치를 주고 상관관계에 따라 변수를 묶어 변수의 수를 줄일 수 있는 주성분 분석방법을 사용하였다.

2. 분석을 통한 침수위험순위 평가

침수위험순위를 예측하기 위해서 재해연보 및 도시계획현황 등의 자료를 통해 홍수피해액, 침수면적, 행정구역면적, 불투수면적, 인구, 인명피해, 유수지용량 등의 변수를 얻어낼 수 있었고 이를 산술평균방법과 가중평균방법, 주성분 분석방법을 이용하여 분석하였다. 여러 개의 변수를 모두 사용하여 다중회귀분석을 하는 경우 변수들 사이의 높은 상관관계에 의해 앞에서 언급한 다중공선성문제를 야기시킬 수 있다.(권세혁, 2008). 이러한 문제를 해결하기 위해, 본 논문에서는 주요 분석방법으로 주성분 분석을 통해 주성분 변수를 얻어 이를 설명변수로 이용함으로써 다중공선성 문제를 해결하였다. 주성분 분석이란 여러 변수를 하나 또는 기존의 변수의 개수보다 더 적은 변수를 찾아내는 분석으로, 변수의 수를 줄여가면서 상관관계가 높은 성분들을 파악하고 이를 어떻게 그룹화 할 것인지를 분석하는 것이다. 주성분 분석에서 유의해야할 점은 상관관계수행렬에서 일반적으로 고유값이 1이상인 주성분과 설명력이 80%이상인 주성분 변수를 선택해야 한다는 점이다. 따라서, 고유값과 고유벡터의 값이 1이상이며 주성분 분산의 설명력이 80%가 넘어야한다. 주성분 값이 크면 원 변수의 영향이 큰 것이므로 이 원리를 바탕으로 주성분 값이 큰 변수를 파악할 수 있다.

* 고려대학교 건축사회환경공학과 박사과정 · 공학석사 · E-mail : hydrohydro@naver.com** 고려대학교 건축사회환경공학과 석사과정 · 공학사 · E-mail : chseokuo@gmail.com*** 정회원 · 고려대학교 건축사회환경공학과 교수 · 공학박사 E-mail: jaykim@korea.ac.kr - 교신저자

표 1. 침수위험 인자 분석 일부분

		홍수피해액 (천원)	침수면적 (km^2)	행정구역면적 (km^2)	불투수면적 (km^2)	평균인구(명)	유수지용량 (m^3)
서울시	강남구	3,425,644	0.55	39.55	23.17	523,984	216,00
	강동구	6,270,297	39.11	24.58	12.54	468,200	323,950
	강북구	9,164,591	0	23.58	9.39	348,593	0
	강서구	1,664,434	102.00	41.40	15.18	523,370	407,766
	관악구	20,329,088	0	29.57	11.92	526,259	8,000
	광진구	3,764,616	0	17.05	11.58	374,651	109,375
경기도	가평군	54,838,880	2.62	83.89	5.17	50,559	21,000
	고양시	130,597,240	62.55	194.77	37.80	763,544	256,700
	과천시	1,622,083	0.02	35.81	3.36	64,550	0
	광명시	13,036,429	2.74	38.65	11.07	333,570	202,610
	광주시	61,607,976	4.67	172.26	11.97	161,788	31,800
	구리시	25,016,384	1.00	33.29	6.97	168,994	890,038
부산광역시	강서구	143,826,938	69.05	241.98	81.93	55,321	0
	금정구	5,053,543	0	65.18	11.21	274,685	379,510
	기장군	40,455,719	0.08	221.60	20.89	75,863	188,110
	남구	49,349,756	0	42.51	30.05	297,365	6,104
	동구	954,370	0	11.59	8.64	118,024	124,600

2.1 산술 평균방법을 이용한 계산결과

각 변수별 측정단위가 서로 다르고, 값의 차이가 크기 때문에 변수를 표준화하여 분석하였다.

표 2. 산술평균에 의한 침수위험순위 일부분

	지 역	실제홍수피해액 (천원)		지 역	산술평균점수
1	경기도 파주시	240,716,297	1	경기도 파주시	17.63731
2	경기도 연천군	208,717,408	2	경기도 평택시	14.73375
3	경기도 포천시	157,879,403	3	경기도 화성시	13.22650
4	부산광역시 강서구	143,826,938	4	경기도 김포시	12.92484
5	경기도 고양시	130,597,240	5	부산광역시 강서구	6.35695
6	경기도 양주시	127,157,751	6	경기도 고양시	6.15607

표 2에서 실제홍수피해액에 의한 순위와 산술평균점수로 구한 순위가 차이가 나는 것을 볼 수 있다. 이는 앞에서 언급한 산술평균의 문제점 중 하나인 모든 변수들에 동일한 가중치를 주는 문제점 때문이다. 이러한 문제점을 보완하기 위하여 가중평균을 이용한 분석을 2.2에서 언급하겠다.

2.2 가중 평균방법을 이용한 계산결과

변수에 가중치를 부여할 때, 객관적인 방법으로 상관계수가 높은 변수를 그룹화하여 가중치를 설정하였다. 산술평균에서의 문제점이었던 유사한 능력을 지닌 항목을 상관계수를 통해서 그룹화하였다. 표 3은 변수의 상관계수 일부분을 나타낸 것이다.

표 3. 가중평균에 의한 상관계수 일부분

	홍수피해액	침수면적	행정구역면적	불투수면적	평균인구	유수지용량
홍수피해액 (<i>p-value</i>)	1	0.570 (0.000)	0.434 (0.000)	0.278 (0.018)	-0.247 (0.037)	-0.084 (0.485)
침수면적 (<i>p-value</i>)	0.570 (0.000)	1	0.356 (0.002)	0.375 (0.001)	-0.120 (0.314)	0.014 (0.910)
행정구역면적 (<i>p-value</i>)	0.434 (0.000)	0.356 (0.002)	1	0.579 (0.000)	0.127 (0.287)	-0.154 (0.195)
불투수면적 (<i>p-value</i>)	0.278 (0.018)	0.375 (0.001)	0.579 (0.000)	1	0.343 (0.003)	-0.057 (0.632)
평균인구 (<i>p-value</i>)	-0.247 (0.037)	-0.120 (0.314)	0.127 (0.287)	0.343 (0.003)	1	0.182 (0.125)
유수지용량 (<i>p-value</i>)	-0.084 (0.485)	0.014 (0.910)	-0.154 (0.195)	-0.057 (0.632)	0.182 (0.125)	1

표 4. 가중평균에 의한 침수위험순위 일부분

	지 역	실제홍수피해액 (천원)		지 역	가중평균점수
1	경기도 파주시	240,716,297	1	경기도 파주시	16.81180
2	경기도 연천군	208,717,408	2	경기도 평택시	14.64393
3	경기도 포천시	157,879,403	3	경기도 화성시	13.24026
4	부산광역시 강서구	143,826,938	4	경기도 김포시	12.80181
5	경기도 고양시	130,597,240	5	부산광역시 강서구	6.55152
6	경기도 양주시	127,157,751	6	경기도 고양시	6.05993

가중평균에 의한 순위와 실제 피해액에 의한 순위가 많이 다른 것을 알 수 있었다. 이는 측정변수가 너무 많고 변수들 간의 높은 상관관계수에 의해 발생하는 다중공선성 때문이라 여겨진다. 상관관계수의 크기만으로 변수들을 분류하는 것은 쉽지 않으므로 변수의 수를 정량적으로 축약하는 주성분 분석이 필요하다고 여겨지며 2.3에서 활용하도록 하겠다.

2.3 주성분 분석방법을 통한 순위축정결과

모든 변수를 이용하여 상관관계수의 크기를 분류하는 것은 쉽지 않으므로 이러한 문제를 해결하기 위해서 주성분 분석을 통해 변수를 축약하였다. 이를 통해 변수의 개수를 줄일 수 있었으며, 일반적으로 고유치가 1 이상이고 누적 설명력이 80%이상인 주성분을 선택한다.

표 5. 각 변수들의 고유치와 누적설명력 일부분

	고유치(eigen value)	누적설명력	비고
1	11.847	0.39490	
2	3.425	0.50906	
3	2.722	0.59978	
4	1.879	0.66241	
5	1.517	0.71297	
6	1.447	0.76119	
7	1.172	0.80028	
8	1.018	0.83420	

표 6. 주성분 분석에 의해 얻어진 고유벡터의 일부분

	$Prin_1$	$Prin_2$	$Prin_3$	$Prin_4$	$Prin_5$	$Prin_6$	$Prin_7$	$Prin_8$
침수면적	0.600	0.292	0.382	-0.377	-0.032	0.052	-0.302	-0.134
행정구역면적	0.335	-0.308	0.706	-0.104	-0.046	-0.370	0.160	-0.023
불투수면적	0.68	0.033	0.843	0.016	-0.212	0.218	0.271	-0.143
침수면적비	0.667	0.308	0.064	-0.385	0.060	0.086	-0.299	-0.115
불투수면적비	-0.432	0.352	-0.137	0.213	-0.095	0.583	0.062	-0.078
평균인구	-0.233	0.313	0.298	0.365	-0.394	-0.318	0.309	-0.152
공공시설피해	0.955	-0.158	-0.091	-0.039	-0.064	0.042	0.088	0.150
건물침수	0.824	0.107	-0.049	0.160	0.142	0.231	-0.018	-0.156

표 5에서는 8개의 변수로 축약이 되었으며, 주성분 변수가 83%의 누적설명력을 가지고 있다. 선택된 주성분 변수에 의해서 고유벡터를 표 6에 나타냈다. 이 고유벡터를 바탕으로 주성분 변수로 축약할 수 있었다.

표 7. 주성분 분석에 의한 침수위험순위 일부분

	지역	실제홍수피해액 (천원)		지역	주성분 점수순위
1	경기도 파주시	240,716,297	1	경기도 파주시	11.09727
2	경기도 연천군	208,717,408	2	경기도 연천군	8.92206
3	경기도 포천시	157,879,403	3	경기도 양주시	7.96572
4	부산광역시 강서구	143,826,938	4	경기도 김포시	7.78473
5	경기도 고양시	130,597,240	5	부산광역시 강서구	3.94064
6	경기도 양주시	127,157,751	6	경기도 고양시	3.85928

4. 결론 및 토의

본 논문에서는 침수위험순위에 영향을 미치는 변수를 사용하여 산술평균, 가중평균, 주성분 분석을 실시하였다. 침수위험순위를 예측하는데 있어서 주성분 분석을 실시한다면 비교적 정확한 결과를 얻을 수 있다.

감사의 글

본 연구는 국토교통부 물관리연구사업의 연구비지원(13AWMP-B066744-01)에 의해 수행되었습니다.

참고문헌

1. 권세혁(2008) <다변량 데이터 분석과 활용>, 자유아카데미, 서울.
2. 성우현(1998) <응용다변량분석>, 탐진출판사, 서울.
3. 소방방재청(1995~2006) <재해연보>
4. 국토해양부(2003) <도시 침수피해 방지를 위한 효율적 실행방안 연구>
5. 김우철(1996) <일반통계학>, 영지문화사, 서울.
6. 국토해양부, 한국토지공사(2006,2007) <도시계획현황>