

다변량 분석을 이용한 국내 강설의 권역화 연구

A Study on the Regionalization of Snowfall using Multivariate Analysis in Korea

이정식* · 신창동** · 김병철***

Lee, Jung Sik · Shin, Chang Dong · Kim, Byung Chul

요 지

국내의 경우 매년 발생하는 홍수에 대해서는 많은 분석과 대비를 하고 있지만, 겨울철 강설에 대한 연구는 미비한 실정이다. 본 연구에서는 기상청 산하 57개 관측지점의 30년 이상의 강설 자료를 수집하고 다변량 분석을 실시하였다. 국내 지역의 강설 동질성을 검정하기 위하여 주성분분석과 군집분석을 실시하였으며 본 연구의 수행으로 얻어진 결과를 요약하면 다음과 같다. 첫째, 본 연구에서는 지금까지 선행된 타 연구에 비하여 보다 많은 강설지점에 대한 다양한 강설특성자료를 이용함으로써 권역화 분석에 대한 타당성과 정확성을 확보할 수 있었다. 둘째, 주성분 및 군집분석으로 관측지점간의 수문학적 동질성 검정을 합리적이고 효율적으로 분석할 수 있었으며, 군집분석의 결과로부터 지형학적 영향을 고려한 강설의 권역화는 국내의 강설지역을 5개의 권역으로 분류하였다. 셋째, 각 강설성분의 권역별 평균값으로부터 각 권역별 강설의 양적 특성을 분석하였으며, 여러 가지 강설 성분을 종합적으로 이용하여 분석된 본 연구의 결과는 향후 재난방지 계획을 위한 강설의 수문학적 분석에 많은 도움을 주리라 사료된다.

핵심용어 : 강설, 동질성 검정, 다변량 분석, 권역화

1. 서론

최근 지구 온난화로 인한 기온상승은 기상이변을 일으키는 원인이 되고 있으며, 기상이변은 여름철 홍수와 겨울철 폭설을 발생시켜 많은 피해를 입히고 있어 인적 및 물적 피해가 나날이 증가하고 있다. 이러한 기상이변 상황은 국내도 예외는 아니어서 2008년에는 강원도에 115cm라는 폭설이, 2010년 최근에는 전국적인 폭설로 인하여 주요 도시 및 시설의 기능이 마비되는 등 극심한 피해가 발생하였다. 국내에서는 급증하고 있는 자연재해에 대비하여 수년간 지속적으로 많은 연구와 투자가 진행되고 있으며 일부 성과도 나타나고 있으나, 현재 계획되고 추진중인 재해방지계획 수립 및 설계의 대부분은 여름철 재해를 대비한 것으로서 최근에 들어 증가하고 있는 겨울철 폭설에 대비한 연구와 계획은 미비한 실정이다.

현재까지 강설에 관해 국내에서 진행된 연구로는 이훈과 이태영(1994)이 수치모형을 이용하여 영동지방의 폭설요인을 분석한 바 있고, 강수만 등(2005)이 위성영상을 이용하여 우리나라의 적설분포 및 적설심을 추출하였으며, 국립방재연구소(2005)에서 2004년의 폭설피해를, 문영일 등(2006)이 적설 및 융설을 고려한 장기유출을 분석하였다. 강설에 관한 연구는 예전부터 진행되었으나, 최근에 들어와 도시지역의 확대와 도로망의 발달, 농업기술의 개발에 따른 동절기 농작물의 경작 등 주변 여건과 결합하여 폭설피해에 따른 사회적 손실이 여름철 수해에 못지않음이 인식되어 그 필요성이 새로이 제기되고 있다.

따라서 본 연구에서는 국내 57개 지점들에 대하여 강설과 관련된 15개의 특성변수들로 다변량 분석을 실

* 정회원 · 금오공과대학교 토목환경공학부 교수 · E-mail : jungsik@kumoh.ac.kr

** 정회원 · 금오공과대학교 건설기술연구소 연구원

*** 정회원 · 금오공과대학교 대학원 토목공학과 석사과정

시하고, 그 결과를 토대로 전국에 대한 권역화를 수행하였다. 연구 진행은 강설특성변수들을 다변량 분석방법 중의 자료축약기법인 주성분 분석을 통하여 각 변수들이 갖는 정보들을 소수의 독립적인 새로운 변수(주성분)들로 생성, 군집 분석으로부터 이들의 유사성 정도를 해석하고 수문학적 및 지형학적 요인을 고려하여 동질성을 갖는 권역으로 분류하였다. 본 연구를 통하여 국내 강설의 공간적 분포도를 파악할 수 있도록 하여, 강설과 관련된 재난방지 계획 및 주요 구조물의 설계에 보다 합리적인 기준을 제시하고자 한다.

2. 기본 자료 및 분석이론

2.1 기본자료

수문자료의 확률빈도를 통계적으로 해석하는 경우 기록년수는 일반적으로 30년 이상이므로 본 연구에서는 국내의 수문관측 실정을 감안하고 관측소의 분포를 고려하여 최소 30년 이상의 기록년수를 가지고 있는 57개 지점의 강설량을 이용하였다. 본 연구의 대상인 57개 관측소와 기록년수는 표 1과 같다. 다변량 분석을 위한 강설특성자료는 강설의 양적 및 지리적 요인 등 모두 15개의 변수를 적용하였으며, 강설 특성별 변수는 표 2와 같다.

표 1. 관측지점의 기록년수 및 특성

지점	기록기간	관측년수	지점	기록기간	관측년수	지점	기록기간	관측년수
속초	'68 - '08	41	광주	'65 - '08	44	임실	'72 - '08	37
대관령	'71 - '08	38	부산	'65 - '08	44	정읍	'72 - '08	37
춘천	'66 - '08	43	통영	'68 - '08	41	남원	'73 - '08	36
강릉	'65 - '08	44	목포	'65 - '08	44	순천	'72 - '08	37
서울	'65 - '08	44	여수	'65 - '08	44	장흥	'72 - '08	37
인천	'65 - '08	44	완도	'72 - '08	37	해남	'71 - '08	38
원주	'71 - '08	38	진주	'70 - '08	39	고흥	'72 - '08	37
수원	'65 - '08	44	강화	'72 - '08	37	영주	'72 - '08	37
서산	'68 - '08	41	양평	'72 - '08	37	문경	'72 - '08	37
울진	'71 - '08	38	이천	'72 - '08	37	영덕	'72 - '08	37
청주	'67 - '08	42	인제	'73 - '08	36	의성	'73 - '08	36
대전	'69 - '08	40	홍천	'73 - '08	36	구미	'73 - '08	36
추풍령	'65 - '08	44	제천	'73 - '08	36	영천	'72 - '08	37
안동	'73 - '08	32	보은	'72 - '08	37	거창	'72 - '08	37
포항	'65 - '08	44	천안	'72 - '08	37	함천	'73 - '08	36
군산	'68 - '08	41	보령	'72 - '08	37	밀양	'73 - '08	36
대구	'65 - '08	44	부여	'73 - '08	36	산청	'72 - '08	37
전주	'65 - '08	44	금산	'72 - '08	37	거제	'73 - '08	36
울산	'65 - '08	44	부안	'73 - '08	36	남해	'72 - '08	37

표 2. 강설특성별 변수명 및 변수설명

강우특성	변수명	개수	변수설명
강설량	Xnsnow	1	최심신적설량
	Xsnow	1	최심적설량
	Xmnsnow	5	1, 2, 3, 11, 12월 최심신적설량
	Xmsnow	5	1, 2, 3, 11, 12월 최심적설량
지리적 요인	Xcor	1	경도(radian)
	Ycor	1	위도(radian)
	Zcor	1	해발(m)

2.2 주성분분석

주성분분석은 다변량으로 이루어진 자료로부터 소수의 특징적인 변수를 찾아서 결합한 후에 자료를 분석하는 기법이다. 즉, 해석하고자하는 다차원의 자료를 포함된 정보의 손실을 가능한 한 적게 해서 2차원 또는 3차원의 자료로 축약하는 방법이다.

주성분분석은 원래의 변수들에 조합상수들을 곱해주어 변환하되, 그렇게 변환된 변수들을 모두 더해서 얻어지는 인위적 변수의 분산이 최대가 될 수 있는 조합상수를 찾아내는 기법이며, 분산이 최대인 인위적 합산값이 바로 주성분이다. 주성분의 분산이 최대가 되어야 하는 이유는 주성분이 원래의 변수들이 가지고 있던 정보를 최대로 함축하여야 원변수들을 대신해서 사용될 수 있는 의의를 획득하기 때문이다.

식 (1)과 같은 선형모형으로 집약된다. X 는 원변수들의 행렬이며, 행은 n 개의 관찰단위, 열은 m 개의 변수로 이루어진다. p 는 변환상수 또는 주성분계수의 행렬로서 변수의 개수인 m 개의 행과 주성분의 개수인 k 개의 열로 이루어진 주성분점수 행렬이다.

$$PC = Xp \tag{1}$$

식 (1)에서 주성분점수들로 이루어진 행렬 PC 는 각 관측단위와 주성분들의 관계정도를 나타내며, 이 행렬의 각 열의 분산을 고유값이라 한다. 또한 주성분계수들의 행렬 p 는 원변수들과 주성분들의 관계 정도를 나타내며, 이 행렬의 각 열을 고유벡터라 한다.

2.3 군집분석

군집분석은 자료의 구조를 아는 유효한 수단으로서 그 자료를 구성하고 있는 관측대상을 어떠한 기준에 의하여 분류하는 방법을 말한다. 개체의 군집화를 위해서는 각 개체들이 가지고 있는 여러 특성들을 종합하여 각 개체간의 유사성거리를 측정하게 되는데, 유사성 거리는 두 개체간의 속성이 유사할수록 가깝게, 두 개체간의 속성이 상이할수록 멀게 측정되어야 한다. 이와 같은 유사성 거리의 측정방법에는 여러 방법이 있지만 그 중 주로 유클리드 거리 또는 유클리드 제곱거리가 이용된다. 유클리드 거리는 기하학적으로 두 점간의 물리적인 직선거리를 의미하는 것으로 통계적으로 두 개체의 변수값 차이를 제곱하여 합산한 값을 제곱근해서 다차원 공간에서 통계적 직선 최단 거리를 나타내는 유클리드 거리는 식 (2)과 같다.

$$d(O_1, O_2) = \sqrt{\sum_{j=1}^m (X_{1j} - X_{2j})^2} \tag{2}$$

여기서 d 는 개체사이의 평균거리이며, O_i 는 i 번째 개체, X_{ij} 는 i 번째 개체의 j 번째 변수의 값이다.

3. 적용 및 결과분석

3.1 주성분분석

주성분분석은 다변량으로 이루어진 자료로부터 소수의 특징적인 변수를 찾아서 결합한 후에 자료를 분석하는 기법이다. 추출된 각 변수의 값들은 그 단위나 범위, 규모 등이 각기 다르기 때문에 주성분분석을 하기 전에 각 변수의 평균이 0, 분산이 1이 되도록 표준화시켜 변수들의 척도를 동일하게 하여 변수간의 차이를 소거하였다.

표준화된 원변수의 상관계수행렬을 이용하여 원변수의 수만큼 주성분을 산출하였으며, 각 주성분의 고유값은 표 3과 같다. 중요주성분의 선택은 고유값이 1이상인 성분만을 추출하는 최소고유값 기준과 분산비율법에 의한 누적분산비가 80%를 넘는 성분까지 취한다는 경험적 기준(박상우 등, 2003)을 적용하여 2개의 중요주성분이 선택되었다. 고유값은 각 주성분의 정보량을 나타내는 주성분 점수의 분산으로 제1주성분은 고유값이 10.551 분산비가 70.343로 원변수에 대한 정보량을 가장 많이 함축하고 있으며, 이후로 갈수록 정보축약의 정도가 줄어드는 것을 알 수 있다.

표 3. 중요주성분의 고유값과 분산비

성분	초기 고유값		
	전체	% 분산	% 누적
1	10.551	70.343	70.343
2	2.047	13.646	83.989
3	0.904	6.029	90.018
4	0.534	3.559	93.577
5	0.427	2.850	96.427

표 4. 관측대상간의 유클리디안 거리행렬

	속초	대관령	춘천	강릉	서울	인천	원주
속초	-	5.342	0.806	0.938	1.198	1.334	0.947
대관령	5.342	-	6.105	5.058	6.521	6.649	6.276
춘천	0.806	6.105	-	1.598	0.421	0.544	0.479
강릉	0.938	5.058	1.598	-	1.888	2.028	1.467
서울	1.198	6.521	0.421	1.888	-	0.141	0.470
인천	1.334	6.649	0.544	2.028	0.141	-	0.596
원주	0.947	6.276	0.479	1.467	0.470	0.596	-

3.2 군집분석

주성분분석에서 산출한 2개의 주성분 점수를 이용하여 군집분석을 실시하였으며, 다음의 결과들을 이용하여 그룹화를 실시하였다. 표 4는 대상지점의 일부인 7개 관측지점간의 유클리디안 거리행렬을 나타낸 것으로서, 거리행렬의 계수는 상이성의 크기를 나타내기 때문에 숫자가 작을수록 유사성이 높다고 해석할 수 있다. 따라서 인천과 서울은 0.141로 유사성이 높으며, 인천과 대관령은 6.649로 상대적으로 유사성이 낮다고 할 수 있다. 군집분석에 의한 강설에 대한 권역 구분은 표 5와 같으며, 크게 5개 권역으로 구분되었다. 구분된 지역은 강원도지역을 I 권역, 강원도일부 및 경북북부 지역을 II 권역, 중부권 지역을 III 권역, 서해안 및 호남지역을 IV 권역, 경남북 지역을 V 권역으로 구분하였다.

표 5. 군집분석을 이용한 권역화

권역	지점수	관 측 지 점
I	3	속초, 대관령, 강릉
II	9	춘천, 원주, 추풍령, 인제, 홍천, 제천, 보은, 영주, 문경
III	18	서울, 인천, 수원, 청주, 대전, 안동, 군산, 강화, 양평, 이천, 천안, 보령, 부여, 금산, 부안, 정읍, 의성, 구미
IV	17	서산, 광주, 통영, 목포, 여수, 완도, 임실, 남원, 순천, 장흥, 해남, 고흥, 거창, 산청, 거제, 남해, 전주
V	10	울진, 포항, 대구, 울산, 부산, 진주, 영덕, 영천, 합천, 밀양

4. 결론

본 연구는 증가일로에 있는 겨울철 기상이변에 대비하여 기상청 산하 주요 관측지점 57개소에 대하여 강설 자료를 수집하여 상호간의 통계적인 동질성 검정을 통하여 국내 강설에 대하여 권역화를 시행한 것으로서, 본 연구의 수행으로 다음과 같은 결과를 얻었다.

- (1) 강설과 관련된 여러 변수들을 활용한 다변량 분석을 수행하여, 강설의 양 및 지리적으로 신뢰도 있는 동질지점들을 분류할 수 있었다.
- (2) 강설의 양 및 지리적 요인 등 모두 15개의 변수를 적용하여 국내를 동질성이 인정되는 5개 권역으로 구분할 수 있었으며, 강원도지역을 I 권역, 강원도일부 및 경북북부 지역을 II 권역, 중부권 지역을 III 권역, 서해안 및 호남지역을 IV 권역, 경남북 지역을 V 권역으로 구분하였다.
- (3) 본 연구를 통하여 국내 강설의 공간적 분포도를 파악할 수 있었으며, 강설과 관련된 재난방지 계획 및 주요 구조물의 설계에 보다 합리적인 기준을 적용하는데 도움이 되리라 생각된다.

감 사 의 글

본 연구는 국토해양부가 출연하고 한국건설교통기술평가원에서 위탁시행한 건설기술혁신사업(08기술혁신 F01)에 의한 차세대홍수방어기술개발연구단의 연구비 지원에 의해 수행되었습니다.

참 고 문 헌

1. 강수만, 임혁진, 권형중, 김성준 (2005), NOAA/AVHRR 위성영상을 이용한 적설분포 및 적설심 추출, 한국수자원학회 학술발표회 논문집, 한국수자원학회, pp.948~952.
2. 국립방재연구소 (2005), 2004 중부지방 폭설피해 조사.
3. 기상청 (1961~2008), 기상연보.
4. 박상우, 전병호, 장석환 (2003), 다변량 분석기법에 의한 지점강우의 권역화 연구, 한국수자원학회논문집, 한국수자원학회, 제36권, 5호, pp. 879~892.
5. 문영일, 김종석, 오태석, 차영일 (2006), 적설 및 용성을 고려한 장기유출분석에 관한 연구, 한국수자원학회 논문집, 한국수자원학회, 제39권, 5호, pp. 1767~1771.
6. 이훈, 이태영 (1994), 영동지방의 폭설요인, 한국기상학회 논문집, 한국기상학회, pp.1~22.