한국대기환경학회지 제 22 권 제 6 호
J. KOSAE Vol. 22, No. 6 (2006) pp. 808~820
Journal of Korean Society for Atmospheric Environment

분산주성분 분석을 이용한 고등학교교실 내 오염패턴분류에 관한 연구

Classification of Pollution Patterns in High School Classrooms using Disjoint Principal Component Analysis

장 철 순 · 이 태 정[1] · 김 동 술*
경희대학교 환경 · 응용화학대학 대기오염연구실 및 환경연구센터
[1]경희대학교 산학협력기술연구원

(2005년 11월 16일 접수, 2006년 7월 27일 채택)

Choul-Soon Jang, Tae-Jung Lee[1] and Dong-Sool Kim*
College of Environment & Applied Chemistry and
Center for Environmental Studies, Kyung Hee University
[1]*Industrial Liaison Research Institute, Kyung Hee University*

(Received 16 November 2005, accepted 27 July 2006)

Abstract

In regard to indoor air quality patterns, the government introduced various polices that were about managing and monitoring quality of indoor air as a major assignment, and also executed "Indoor Air Quality Management Act" which was presented in the May, 2004. However, among the multi-usage facilities controlled by the Act, the school was not included yet.

This study goal was to investigate PM10 pollution patterns of the high school classrooms using a pattern recognition method based on cluster analysis and disjoint principal component analysis, and further to survey levels of inorganic elements in May, June, and September, 2004. A hierarchical clustering method was examined to obtain possible objects in pseudo homogeneous sample classes by transformation raw data and by applying various distance. Following the analysis, the disjoint principal component analysis was used to define homogeneous sample class after deleting outliers. Then three homogeneous patterns were obtained as follows: the first class had been separated and objects in the class were considered to be sampled under semi-open condition. This class had high concentration of Ca, Fe, Mg, K, Al, and Na which are related with a soil and a chalk compounds. The second class was obtained in which objects were sampled while working air-conditioners and was identified low concentration of PM10 and elements. Objects in the last class were assigned during rainy day. A chalk, soil element and various types of anthropogenic sources including combustions and industrial influenced the third class. This methodology was thought to be helpful enough to classify indoor air quality patterns and indoor environmental categories when controlling an indoor air quality.

Key words : Cluster analysis, Disjoint principal component analysis, SIMCA, High school, IAQ

*Corresponding author.
Tel : +82-(0)31-201-2430, E-mail : atmos@khu.ac.kr

1. 서 론

현대인은 다양한 실내공간에서 하루 80~90% 이상을 생활하고 있으나, 건축자재의 화학물질 사용 증대 및 에너지 절감을 위한 건물 밀폐화 등으로 실내공기 오염은 더욱 악화되고 있다. 정부는 실내 공간에 대해 새집증후군의 원인인 실내공기 오염을 줄이고, 이의 적정 관리를 위해 2003년 "다중이용시설 등의 실내공기질 관리법(이하, 실내공기질관리법)"을 재정·시행하는 등 실내공기질 관리를 최우선 정책과제로 삼아 다양한 정책을 도입·추진하고 있다. 또한 다중이용시설의 시설별 관리대책, 공동주택의 새집증후군 방지대책, 친환경 건축자재의 사용 확대를 위한 대책 등 실내공기 관리에 대한 종합적인 청사진을 담은 "실내공기질 관리 기본계획"을 2004년 수립하였다. "실내공기질 관리법"에 의한 관리 대상 시설에는 지하역사, 의료기관, 도서관 등을 포함한 17개 시설군 및 신축공동주택을 대상으로 하였으나 한정된 공간에 다수의 학생이 장시간 생활하고 있는 학교시설은 포함되어있지 않다(환경부, 2003). 학교시설에 대한 공기질 관리는 교육인적자원부에서 "학교보건법"을 근거법령으로 관리하고 있으며, 2005년 3월에 개정되어 2006년 1월부터 시행되고 있는 동법에 의하면 기존 학교시설의 공기질 관리항목을 CO_2와 미세먼지 2종에서 VOC와 포름알데히드를 포함하여 총 12종으로 확대하여 유지기준을 마련하였다. 또한 신축학교에 대해서는 "실내공기질관리법"의 규제에 의한 오염물질 방출 건축자재의 사용을 제한하고 있다(교육인적자원부, 2005). 그러나 지금까지의 학교 공기질 연구결과에 의하면 CO_2의 경우 동절기 조사 대상시설 모두 환경기준을 초과하고 있으며, 미세먼지 역시 공기질 기준의 1.7~3.7배까지 초과하는 높은 농도를 보였다(신은상 등, 2002). 또한, 최근 실시된 대도시 50개 교육시설의 오염실태 조사 자료에 의하면 조사시설의 약 22%에서 환경부가 정한 VOC 공기질 기준을 초과하는 것으로 보고된바(임종한, 2005), 학교시설 공기질 개선을 위한 다각적이고 집중적인 연구가 요구된다.

교실의 공기질 상태는 위치, 면적, 층수, 문의 개폐상태, 학생 수, 실외오염도, 냉난방기 운영상태, 측정계절 및 기상상태 등에 의해 실내 오염도는 큰 차이를 보일 수 있다. 따라서 다양한 외부 환경 및 물리적 패턴에 따른 오염도 변화를 분석함으로써 학교 공기질의 적정 관리를 위한 합리적인 대책과 관리방안을 마련할 수 있다. 이러한 패턴분류를 위해서는 사전정보가 전혀 없거나 거의 없는 경우, 각각의 측정자료, 즉 개체들을 유사한 특성을 가진 군집(또는 패턴)으로 분류하여 군집들 간의 관계를 파악할 수 있는 군집분석법(cluster analysis)과 군집화 작업이 올바르게 수행되었는지 여부를 정량적으로 평가 할 수 있는 주성분분석법(principal component analysis) 또는 인자분석법(factor analysis)과 같은 다변량 통계(multivariate statistics) 방법을 이용할 수 있다. 이와 같이 패턴 분류를 통한 오염농도 해석에 관한 연구는 김동술과 김형석(1990)이 서울시내 지하상가 공기질 관리에 응용한바 있으며, 남보현 등(2002)이 다양한 실내 환경 중 PM10 오염 분류에 적용한 바 있다.

본 연구에서는 오염 패턴분류를 통한 학교시설 공기질 관리방안 기틀 마련을 위해 경기도에 위치한 4개 고등학교 교실에서 온·습도 및 교실여건 등 물질적 환경조건과 함께 직경 10㎛ 이하의 입자상 오염물질인 PM10을 채취하여 19종의 무기원소를 분석하였다. 분석된 PM10 및 원소자료는 군집분석법을 이용하여 유사한 특성을 지닌 군집으로 분류하였으며, 분류된 군집을 확률적으로 검증하기 위해 분산주성분 분석법(disjoint principal component analysis)을 이용하였다. 최종적으로 확률적 검증을 통해 이상치를 제외한 순수군집(homogeneous cluster)인 패턴(pattern) 유형에 따른 PM10 오염도 특성을 해석하였다. 본 연구에서 사용된 패턴별 학교 공기질 특성을 파악하기 위한 상기 방법론은 학교를 포함한 실내공간의 공기질 적정관리를 위한 기초연구로서 유용하게 적용될 수 있을 것으로 판단된다.

2. 연구 방법

2.1 실험 방법

본 연구는 2004년 5월부터 9월까지 경기도 수원시와 안산시에 위치한 4개 고등학교에서 수행하였다. 그림 1은 4곳의 시료채취 대상 고등학교의 위치를 보여주고 있다. 지점 A는 바닷가에서 멀지 않은

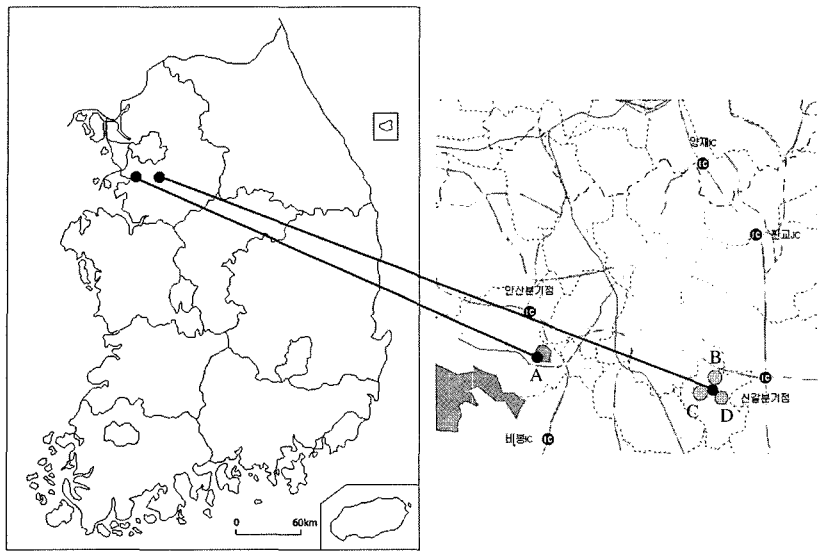

Fig. 1. Location of sampling sites.

주거지역에 위치하고 있으며, 학교 앞에는 큰 도로가 있지만 교통량은 많지 않다. 주위에 자동차와 관련(타이어, 튜닝회사, 정비회사)된 소규모 상가들이 위치하고 있다. 지점 B는 대학 근접지역으로 매우 발달한 상업지대로 일일 교통량과 유동인구가 상당히 많은 편이며, 인근에 공단이 위치하고 있다. 또한, 학교 뒤로 녹지가 조성되어 있다. 지점 C는 수원 중심가 부근의 발달된 상업지역에 위치하고 있고, 학교 앞 도로는 교통량이 매우 많고, 상습적으로 정체되는 구간이다. 지점 D는 전형적인 아파트 단지 내에 위치하고 있어 출퇴근 시간을 제외하고는 교통량은 많지 않다.

시료 채취는 교실과 외기의 영향을 조사하기 위해 실외에서 동시에 실시하였다. 교실에서의 시료 채취기는 교실 뒤쪽 중앙에 설치하였다. 시료채취 시간은 학생들의 활동시간인 08:00시부터 16:00까지 8시간 동안 수행하였다. 실외의 시료 채취는 교실이 위치한 해당 건물의 옥상 중앙에 설치하여 수행하였다. 채취된 총 시료 수는 실내 27개, 실외 11개로 각각 독립된 자료를 얻었다. 조사된 학교 운동장은 모두 흙으로 이루어져 있으며, 실내화를 착용하였으나 건물 바깥 일정거리에서 많은 학생이 실내화를 신고 활동하였다. 각 시료채취 시 특이사항으로 문의 개폐여부, 온 · 습도, 학생 수, 냉난방기 사용유무, 외부의 기상상태 등을 조사하였다. 연구지역의 현황은 표 1에 나타내었다.

PM10 시료채취는 미국 Airmetric사의 mini-volume portable sampler를 사용하였으며, 채취 유량은 5.0 L/min으로 8시간 동안 측정하였다. 시료 채취에 사용한 여지는 미국 Corning Costa사의 직경 47 mm, pore size 0.2 μm의 membrane filter를 사용하였으며, 여지는 채취 전 · 후에 전자 데시케이터에 24시간 보관하였다. 농도는 시료채취 전 · 후의 무게차를 채취 유량으로 나누어 계산하였다. 채취된 미세먼지의 무기원소 분석을 위하여 전처리를 수행하였다. 전처리 방법은 미국 EPA에서 고시한 microwave 전처리법인 Questron (U.S.A., Questron Co., Model Q-15 MicroPrep)을 이용한 질산, 염산 전처리법을 수행하였다. 이 장치는 용매를 고온으로 진공 · 가압하는 장치로서, 고압 하에서 산분해가 가능하므로 신속한 가열이 가능하여 전처리 시간을 단축시킬 수 있다. 또한 전처리 과정에서 사용되는 용매인 산의 소모가 적으며, 전처리 과정에서 발생될 수 있는 오염을 최소화시킬 수 있다. 그 밖에 200℃ 이상의 고온에서 분해가 가능하며, 용기당 1.0~2.0 g의 소량으로 시료의 전처리가 가능하다. 채취한 여지를 PFA liner에 넣은 후 질산 7.0 mL와 염산 3.0 mL를 넣고 녹인 다음 파워 5와 4에서 각각 5분씩 전처리하였다. 전처리

Table 1. Brief conditions of sampling sites.

ID	Date	Temperature (℃)		Humidity (%)	Reference	Place	Vol. of classroom (m^3)
		Min.	Max.				
1 2 3	04-May-06	8.9	22.3	48.5	-	A 2nd floor	216.0
4 5	04-May-11	13.6	33.3	51.9	-		
6 7	04-Jun-03	10.9	22.4	75.9	A/C condition working		
8 9 10	04-Jun-16	11.0	22.1	46.3	-	B 3rd floor	171.5
11 12 13	04-Jun-17	17.9	22.9	80.9	Rainy		
14 15	04-Jul-08	20.3	25.0	83.9	Rainy		
16 17 18	04-Sep-10	15.9	26.8	61.8	Air condition working	C 3rd floor	162.8
19 20	04-Sep-13	17.8	24.3	65.1	-		
21 22	04-Sep-14	16.4	28.1	72.1	Air condition working		
23 24	04-Sep-22	12.0	24.5	68.9	-	D 2nd floor	171.0
25	04-Sep-23	15.8	25.3	68.6	-		
26 27	04-Sep-24	14.9	27.1	68.1	Air condition working		

가 끝난 시료는 유도결합프라즈마 원자방출분광법 (inductive coupled plasma-atomic emission spectrometry; ICP-AES) (DRE ICP, Leeman Labs Inc.)을 이용하여 Al, Mn, Ti, V, Cr, Fe, Ni, Cu, Zn, As, Se, Cd, Ba, Ce, Pb, Na, Mg, K, Ca 등 총 19종의 무기원소 분석을 실시하였다. 시료 전처리 과정은 그림 2와 같으며, 사용된 ICP-AES의 분석조건은 표 2에 제시하였다.

무기원소 분석에 이용된 ICP-AES의 검출한계는 시료의 채취에 사용된 여지의 바탕시험(blank test)에 대한 3σ 방법으로 산출하였으며, Al 0.056, Mn 0.000, Ti 0.002, V 0.001, Cr 0.012, Fe 0.013, Ni 0.201, Cu 0.005, Zn 0.003, As 0.008, Se 0.044, Cd 0.002, Ba 0.014, Ce 0.089, Pb 0.132, Si 0.236 mg/L로 계산되었다.

2. 2 군집분석 과정

군집분석법은 각 개체(object)의 유사도(similarity)를 측정하여 유사성이 높은 대상 집단을 분류하고, 같은 군집에 속한 개체들의 유사성과 다른 군집에 속한 개체간의 상이성을 규명하는 통계분석법으로, 대상들을 분류하기 위한 명확한 기준이 존재하지 않거나 기준이 밝혀지지 않은 상태에서 다양한 특성을 지닌 대상들을 집단으로 분류하는데 사용되는 기법이다. 군집분석은 군집의 개수, 내용, 구조 등이 완전히 알려지지 않은 상태에서 개체 사이의 거리 또는 비유사도에 근거하여 군집을 형성하고, 형성된 군집의 특성을 파악하며, 군집들 간의 관계를 분석하는 것으로 집단간 분산(between-group variance)을 최대화 시키며, 집단내 분산(within-group variance)을 최

Table 2. Analytical conditions ICP-AES.

RF Power[a] (kW)		1.0
Flow Rate (L/min)	Coolant	19.0
	Auxiliary	0.5
	Carrier	1.0
Uptake Rate (mL/min)		0.8
Nebulizer Pressure (psi)	Hilderbrand grid nebulizer (Ultrasonic nebulizer)	45.0
Spray chamber	Scotty type	
Spectrometer		750 mm focal length 1,800 groove/mm 0.2 nm/mm

[a]RF Power: radio frequency power

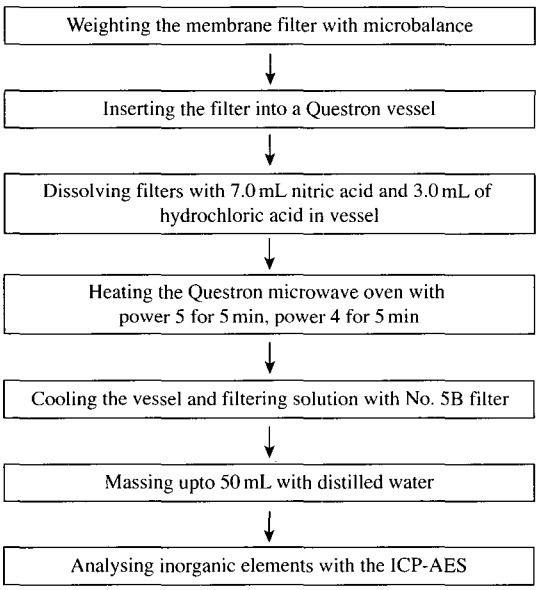

Fig. 2. The pretreatment procedure for analyzing inorganic elements.

소화시키는데 목적이 있다(김기영과 전명식, 1994; Dorling *et al.*, 1992). 즉, 두 개체사이의 거리를 기준으로 공간에서 비유사도를 측정함으로써, 동일한 패턴을 가진 개체를 규명하는 것이며, 거리 또는 비유사도가 작은 두 개체는 동질집단(homogeneous class)으로 결정된다(Hopke, 1991).

일반적으로 군집분석은 크게 위계(hierarchical) 분석법과 비위계(non-hierarchical) 분석법의 2가지 종류로 나눌 수 있다. 비위계분석법은 군집의 수를 이미 알고 있거나 가정할 수 있을 때, 개체들을 정해진 특정한 군집으로 최적 분배하는 방법이다. 위계분석법은 한 군집이 다른 군집에 포함되지만, 군집간 중복이 허용되지 않고 계보형식의 나뭇가지와 같은 구조를 취하는 방식의 분석법이다. 위계분석법은 비위계분석법에 비해 비교적 알고리즘이 간단하고 계산시간이 상대적으로 짧으며, 전체 군집들 간의 구조적인 관계를 수상도(dendrogram)를 통해 2차원상의 공간에 간단히 표현 할 수 있다는 장점이 있다(김기영과 전명식, 1994; Massart and Kaufman, 1983). 위계분석법은 다시 응집위계(agglomerative hierarchical) 분석법과 분산위계(divisive hierarchical) 분석법으로 세분된다. 응집위계분석법은 가장 가까운 군집들이 나무가 가지를 치듯이 하나의 큰 군집을 만들면서 끝을 맺는 방법이고, 반면에 분산위계분석법은 응집위계분석법의 역과정으로, 모든 개체를 포함하는 하나의 큰 군집으로 시작하여 한 개의 개체가 한 개의 군집을 만들 때까지 분열되는 분석법이다.

특정 응용에 맞는 올바른 비유사도의 선택은 위계분석법 응용에서 매우 중요하다. 각 대상들의 비유사도를 계산하는 방법에는 주로 거리측정법을 이용한다. Hopke (1976)는 환경관련 자료로 부터 해석 가능한 군집 분류를 위해서는 유클리디안 제곱거리를 사용한 바 있다. 즉, 군집 속에서 개체 간 거리의 제곱의 합이 최소증가를 보일 때, 이미 군집화된 자료를 해석하는데 가장 유용함을 보였다. 본 연구에서는 비유사도를 측정하기 위해서 일반적으로 많이 사용하는 방법으로는 유클리디안 거리(Euclidian distance), 유클리디안 제곱거리(squared Euclidian distance), 시티블록거리(city-block distance) 등 세 가지 방법을 이용하여 모델링하였으며, 분류된 군집의 자료해석이 가장 용이한 방법을 선택하였다. 세 가지 비유사도 거리 계산 방법은 다음 식과 같다.

$$\text{유클리디안 거리}: D_{jk}=\sqrt{\sum_{i=1}^{m}(X_{ji}-X_{ki})^2} \quad (1)$$

$$\text{유클리디안 제곱거리}: D_{jk}=\sum_{i=1}^{m}(X_{ji}-X_{ki})^2 \quad (2)$$

$$\text{시티-블록 거리}: D_{jk}=\sum_{i=1}^{m}|X_{ji}-X_{ki}| \quad (3)$$

본 연구에서 군집분석을 응용하기 위해 응집위계분석법을 주사용 목적으로 하는 AGCLUS 프로그램

을 이용하였다. AGCLUS는 FORTRAN IV로 쓰여진 컴퓨터 프로그램으로 비유사도 측정을 위해 7가지 사양을 연구자에게 제공하고 있다(Oliver, 1973). 입력자료는 변량분포가 대칭형이 아닐 경우 결과가 과장되거나 오류를 범할 수 있으므로 원자료에 대한 제곱근(square root) 변환, 로그(logarithmic) 변환 또는 그 밖의 적당한 자료변환을 통해 변수의 분량분포가 좌우대칭형이 되도록 원자료를 정규화하여야 한다(Albano *et al.*, 1981). 군집분석법은 연구과정에서 자료의 형태나 윤곽은 쉽게 파악할 수 있으나, 유사도 측정방법의 선정, 군집수를 결정하는 유사도 준위의 결정 등 분석자의 주관적인 판단이 개입될 수 있으므로 연구의 최종 결과를 얻고자 할 때 많은 주의가 필요하다(김동술과 김형석, 1990).

2.3 분산주성분 분석과정

군집분석을 이용하여 유사한 개체들이 각각의 군집으로 분류되면, 각 군집에 분류된 개체들이 그 군집(class 또는 pattern)에 소속되어 있는지 여부를 정량적으로 규명하고, 분류된 군집을 확률적으로 검증한 뒤, 최종적으로 순수 군집을 도출하여야 한다. 이와 같은 연구를 수행하기 위해, Wold(1976)에 의해 개발된 분산주성분분석법을 응용한 형태 인식법(pattern recognition)을 적용하였으며, 이를 위해 SIMCA (Soft Independent Modeling of Class Analogy) 패키지를 사용하였다(SIMCA-3B manual, 1984). SIMCA 모델은 환경연구 뿐만 아니라 분석화학분야 등에 널리 응용되고 있으며, 1924년 Bronsted 등이 화학성분의 분석방법에서 처음으로 적용하였다(Albano *et al.*, 1981). 최근 연구로는 Branden과 Hubert(2005)가 숲에서 채취한 58개 토양시료의 양이온 분석값을 이용하여 시료 채취지점을 분류하였으며, Wang *et al.*(2005)은 근적외선분광기(near infrared spectroscopy: NIRS)에 의해 분석된 감초성분을 SIMCA를 이용하여 생산지역 및 생육조건 등을 분리하였다.

분산주성분분석법이란, 이미 알고 있거나 군집분석 등을 통해 생성된 군집 각각에 대해 독립적으로 주성분분석을 수행하는 방법론이다. 분산주성분분석을 통해 각 군집은 선형구조로 모델화 되고, 한 개의 개체가 고정된 확률값에서 이미 모델화된 특성 군집속에 소속할 수 있는지를 임계거리(critical distance)를 이용하여 결정할 수 있다. SIMCA를 수행 할 때의 입력 자료는 각 군집들이 미리 알고 있는 정보에 의해, 혹은 군집분석과 같은 방법으로 각 군집들이 임의로 분류되어 있는 군집인 트레이닝 세트(training set)와 어떤 군집으로도 분류되지 않은, 소속이 불확실한 개체들의 모임인 테스트 세트(test set)로 구성된다. 또한 SIMCA 수행 후 어떠한 군집에도 포함되지 않는 개체들을 이상치(outlier)라 부른다. SIMCA에 대한 자료 분석은 두 단계로 이루어지는데 첫 번째 단계로 트레이닝 세트를 이용하여 주인자 모델을 개발하는 것이고, 두 번째 단계는 테스트세트 중의 각 개체들을 이미 개발된 각 군집모델과 비교하여 소속감(membership)을 부여하는 것이다. 본 연구에서는 군집분석 시 모든 개체들이 각 군집에 소속되어 두 번째 단계의 분석이 요구되지 않았다. 또한 SIMCA를 응용할 경우 군집분석과 마찬가지로 자료의 변환은 매우 중요하다. 환경 자료는 한 쪽으로 치우쳐 있을 경우가 많으며, 이때 제곱근변환 및 로그변환 등 적합한 변환을 통해 변수의 과잉 영향을 줄일 수 있다(Box *et al.*, 1978).

SIMCA는 한 군집내의 유사 개체들의 측정값 X_{ik}는 경험적 모델에 의해 설명할 수 있다. SIMCA의 원리는 다중 Taylor 급수(multiple Taylor's expansion)를 기반으로 하는데, 이는 분산계수(discrete parameter)인 자료 X_{ik}는 식(4)와 같이 변수 i의 평균치 X_i, 변수관련항 β, 개체의 값 θ로 표현되며, 주성분 수 a개를 가진 주성분 모델에 맞게 조정(fitting)할 수 있다. 간단한 행렬식으로 나타내면 식(5)와 같다.

$$X_{ik}^{(q)} = X_i^{(q)} + \sum_{a=1}^{A} \beta_{ia}^{(q)} \theta_{ak}^{(q)} + \varepsilon_{ik}^{(q)} \quad (4)$$

$$X = A + B \cdot T + E \quad (5)$$

자료행렬 $X_{ik}^{(q)}$는 q번째 군집에서 a개의 주성분을 가진 i번째 변수와 k번째 개체의 측정치를 의미한다. 식(4)의 잔차 ε_{ik}는 X_{ik}의 임의의 부분(random part)을 설명하는 것으로 측정오차(measurement error)와 근사의 불완전성에 의한 모델오차(model error)를 포함한다. 잔차 ε_{ik}의 제곱의 합은 모델화된 군집과 그 군집에 속한 개체 사이의 거리를 나타낸다. 또한 잔차로부터 각 군집의 잔여표준편차(residual standard deviation) S_o가 계산되며 식(6)과 같이 나타낼 수 있다.

$$S_o^{(q)} = \left[\sum_{i=1}^{M} \sum_{k=1}^{M} \varepsilon_{ik}^{2(q)} / (M_q - A_q)(n_q - A_q - 1) \right]^{1/2} \quad (6)$$

여기서 n, A, M은 각각 군집의 개체 수, 주인자 수, 변수의 수를 의미한다. 잔여표준편차 S_o는 트레이닝 세트 중 잘못 분류된 군집, 즉 이상치(outlier)를 확인하기 위해 사용된다(Albano *et al.*, 1981). 각 개체의 잔여표준편차는 계산된 임계거리와 비교하여 군집에 소속여부를 결정할 수 있다. 하나의 개체는 하나의 군집에만 유일하게 소속되거나, 둘 또는 그 이상의 군집에 동시에 소속되거나, 또는 어떤 군집에도 속하지 않게 된다.

어떤 군집으로도 분류되지 않은 군집인 테스트 세트의 개체 p의 분류를 위하여 각 군집 q의 모델을 이용하여 식(7)과 같은 일반적인 다중선형회귀분석(multiple linear regression analysis)에 의해 각 군집모델에 맞게 개체를 조정하므로 계산할 수 있다.

$$X_{ip} - X_i^{(q)} = + \sum_{a=1}^{A} \beta_{ia}^{(q)} \theta_{ap}^{(q)} + \varepsilon_{ip}^{(q)} \quad (7)$$

테스트 세트의 분리 과정 및 SIMCA에 대한 더욱 자세한 원리 및 수식은 여러 문헌들을 참고할 수 있다(김동술과 김형석, 1990; Wold and Sjöström, 1977; Wold, 1976). 본 연구에서는 대상 개체들이 각 군집에 속해 있으므로 테스트 세트에 대한 분석은 필요하지 않았다. 한 개의 트레이닝 세트가 두 개 이상의 군집으로 구성되어 있을 때, 각 군집은 표준결정도(standard decision plot)를 이용하여 시각적으로 비교할 수 있다(Cooman *et al.*, 1981).

2.4 군집 분석법과 분산주성분 분석법에 의한 패턴 분류 과정

고등학교 교실 내 오염패턴 분류과정에 대한 순서도를 그림 3에 제시하였다. 본 연구에서는 다양한 군집분석 방법 중 비교적 알고리즘이 간단하고 계산이 간편하며, 보편적으로 가장 많이 사용되는 응집위계분석법을 적용하여 경기도에 위치한 고등학교 4곳에서 얻은 27개의 자료를 분류하였다. 군집분석을 위한 입력 자료로는 변환을 거치지 않은 원자료와 제곱근 변환과 로그 변환을 거친 자료들을 이용하였다. 로그변환 시 원자료의 농도 값이 "0"인 것이 상당수 있으므로, 자료 전체에 "1"을 더하여 로그변환을 실시하였다. 유사도 기준으로는 일반적으로 많이 사용하는 유클리디안 거리, 유클리디안 제곱 거리, 시티블록 거리 등 세 가지 방법을 이용하여 군집에 속한 개체의 해석이 용이한 최적의 방법을 선택하였다. 비유사도를 기하학적으로 볼 때, 군집분석은 공간상에서 두 개체 사이의 거리를 기준으로 비유사도를 측정하는 방법이므로, 변수간 편차가 큰 원자료를 그대로 군집분석에 이용할 경우 다른 변수들과 비교하여 상대적으로 큰 수치를 가지고 있는 변수가 거리계산시 가장 큰 영향을 미치게 된다(Hopke, 1985). 본 연구의 원자료 분석 결과 Ca의 최대값은 60.6 μg/m^3, 최소값은 0.79 μg/m^{3}이고, Mg의 최대값은 11.9 μg/m^3, 최소값은 0.06 μg/m^{3}으로 최소 몇 십 배에서 최대 200배 가까운 차이를 보이고 있다. 따라서 군집분석 이전의 변수들이 동일한 가중치를 가질 수 있도록 하기 위해 식(8)과 같이 z-score를 이용하여 표준화를 수행하였다.

$$Z_{ik} = \frac{(X_{ik} - \overline{X}_i)}{\overline{S}_i} \quad (8)$$

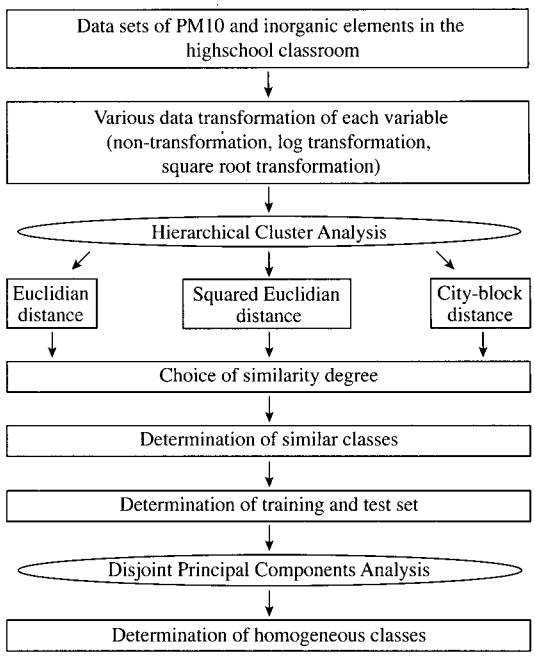

Fig. 3. A flow chart for the pattern recognition procedures of in this study.

여기서 Z_{ik}는 X_{ik}의 표준화 된 값이며, $\overline{X}_i$와 $\overline{S}_i$는 각각 i번째 변수의 평균과 표준편차이다. 각각의 비유사도 기준을 적용하여 분석한 결과 원자료에 로그변환을 거치고 두 개체간의 변수값의 거리제곱 값을 취한, 즉 유클리디안 제곱거리법을 적용한 경우 양호한 결과를 얻을 수 있었다. 분류된 그룹이나 케이스 사이의 거리를 표시하는 군집화 기준으로는 Ward's 방법을 이용하였다.

3. 결과 및 고찰

3.1 군집분석 결과

그림 4는 원자료를 로그변환한 후, 유클리디안 제곱 거리를 비유사도 기준으로 군집분석을 수행한 결과로 얻은 수상도이다. 그림에서 보는 바와 같이 밑으로 갈수록 적은 개체들이 높은 비유사도 준위(level)에서 산만하게 분산되어 있는 것을 볼 수 있다. 이 수상도에서 이론적 기준에 의하여 최적의 비유사도 준위를 결정할 수 있다면, 학교교실에서의 오염 패턴을 쉽게 파악할 수 있다. 하지만, 위계분집분석법의 수상도에서 최적의 군집수 결정방법은 통계학적인 견지에서 상당히 주관적이다. 따라서 본 연구에서는 군집분석에서 얻은 결과에 대한 객관적 판단

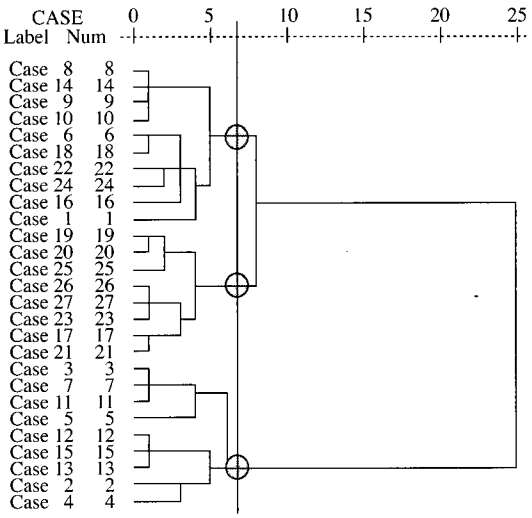

Fig. 4. Dendrogram using the squared Euclidian distance after logarithmic transformation.

을 위해 주인자분석법의 한 종류인 분산 주성분분석법을 통해 선택된 군집내의 개체들을 확률적으로 검증하였다. 본 연구에서는 비교적 낮은 비유사도 준위를 주관적으로 설정한 후, 잠정적 패턴 분류를 시도하였다. 즉, 주어진 비유사도에서, 군집 1은 10개의 개체(8, 14, 9, 10, 6, 18, 22, 24, 16, 1), 군집 2는 8개의 개체(19, 20, 25, 26, 27, 23, 17, 21), class 3은 9개의 개체(3, 7, 11, 5, 12, 15, 13, 2, 4)등으로 잠정적인 가분류를 시행 할 수 있었다. 어느 군집에도 포함되지 않은 특이치는 분류되지 않았다.

3.2 SIMCA의 분석 결과

위계군집분석법의 수상도를 이용하여 잠정적으로 3개의 군집을 만들었다. 본 연구의 경우 소속감을 부여할 테스트 세트가 존재하지 않으므로 SIMCA의 첫 번째 응용단계로서 트레이닝 세트를 이용한 주인자 모델을 개발하였다. 우선, 각 군집내의 개체가 군집 내에 확률적으로 존재할 수 있는지를 검사하였다. 이를 위해, 이미 잠정적으로 분류된 군집 1의 10개, 군집 2의 8개, 군집 3의 9개 개체들을 각각 트레이닝 세트로 하고, 분산 주성분분석을 시행하였다. 분산 주성분 분석의 첫 번째 단계는 횡유효도(cross validation) 검사를 수행하여 유효한 주성분 수를 결정하는 것이다. 각 군집의 개체 수가 3개 이하로 분류될 경우 유효한 주성분이 결정되기 어려우며, 이러한 결과는 군집 분석이 올바르게 수행되지 않았음을 의미한다. 본 연구의 각 군집의 object의 수는 3개 이상이었으며, 유효 주성분이 결정되었다. 그 다음 단계로 식(6)에 의해 잔여 표준편차를 구할 수 있었으며, 계산된 값을 이용하여 그림 5와 같이 Cooman의 표준결정도(standard decision plot)를 작성할 수 있었다. 그림에서 수직과 수평으로 그은 선은 95% 확률의 임계거리선(critical distance line)으로서 이 선에 의해 4개의 영역으로 나누어진다. 임계거리 선은 각 군집내의 개체가 해당 군집에 속하는 가를 평가할 수 있는 지시값이 된다. 이와 같이 각 평면위에 서로 다른 군집들을 각각의 임계거리에 준하여 상호비교하여 각 군집의 순수 개체와 이상치 개체(outlier object)를 분리 결정할 수 있다. 즉, 각 class의 횡유사도를 이용하여 이상치 개체를 제거한 후, 잠정적 군집을 순수 군집으로 만들 수 있었다.

그림 5(a)는 군집 1과 군집 2에 대한 표준 결정도

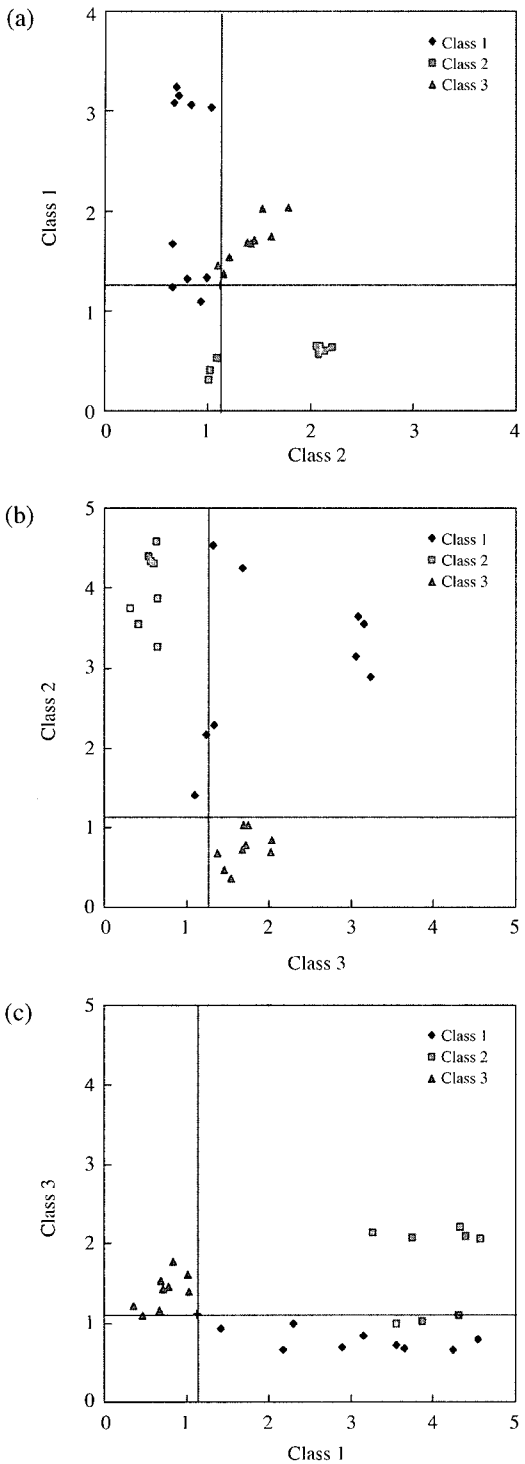

Fig. 5. Cooman's standard decision plot between specific classes.

Table 3. Classification objects for PM10 groups.

Class	Objects
Class 1	8, 14, 9, 10, 6, 22, 16, 1
Class 2	20, 26, 27, 23, 17
Class 3	7, 11, 5, 12, 15, 13, 2, 4
Class 99	18, 24, 25, 21, 19, 3

를 도식한 것으로 95% 확률의 임계거리 선에 의해 4개의 영역으로 나누어진다. 그림 왼쪽 상단 영역은 군집 1의 개체들만이 존재하는 영역이고, 오른쪽 하단의 영역은 군집 2의 개체들만이 존재하는 영역이다. 왼쪽 하단의 영역은 두 군집의 개체가 공존하는 영역이고, 오른쪽 상단 영역은 군집 1과 군집 2를 제외한 그 밖의 군집 개체가 존재하는 영역이다. 이 그림에서는 군집 3의 개체들이 오른쪽 상단에 위치한다. 이와 같은 표준결정도를 이용하여 군집분석 수행 시 주관적인 군집수의 결정, 즉 비유사도의 임의적 결정을 객관화하여 순수한 군집 (homogeneous cluster)을 도출하였다. 따라서 좌측 상단 및 우측 하단 영역에 개체가 존재하면 95% 확률로 해당 군집에 속하게 되는 것이고 그 이외의 영역에 개체가 존재하면 해당 군집에서 배제된다. 이와 같은 방법으로 그림 5의 b)와 c)을 작성할 수 있었다. 그림에 의하면, 잠정적 군집 1속의 10개의 개체 중 2개의 개체 (No. 18, 24), 군집 2의 8개 개체 중 3개 (No. 25, 21, 19), 군집 3의 9개 개체 중 1개 (No. 3)가 제거되어 나머지 개체로 구성된 순수 군집을 만들었다. 잠정적 군집에서 제거된 이상치 개체들은 표 3과 같이 상호 비교되어 기타 군집 99에 포함시켰다. 각 개체의 순수군집에의 소속율은 78%였다.

3. 3 군집의 특성 및 형태

군집분석에 의해 인위적으로 생성된 3개 군집을 SIMCA에 의해 이상치 개체들을 제거한 후 만든 순수 군집들만의 최종적으로 분류된 군집별 PM10 및 원소농도의 패턴 특성은 표 4와 같다. 군집별 각 원소평균농도는 그림 6에 제시하였다. 그림 7은 각 군집별 외기와 실내의 평균농도 비율을 보여 주고 있다. 외기 농도는 각 군집에 해당하는 개체와 동일한 날에 외기에서 분석된 자료를 비교하였다. 각 순수군집에서 도출된 특성 및 오염형태는 다음과 같다.

첫 번째 군집은 학생들의 출입이 잦아 준 개방 상

Table 4. verage concentrations of PM10 and inorganic components in each class after disjoint principal components analysis. (Unit : ng/m³)

Group	Class 1	Class 2	Class 3
PM10*	114.6	62.5	237.7
Al	403.6	0.0	3,484.4
Mn	65.1	20.8	221.4
Ti	0.0	0.0	85.9
V	7.8	0.0	7.8
Cr	5.2	0.0	161.5
Fe	888.0	154.2	5,085.9
Ni	31.3	0.0	67.7
Cu	46.9	0.0	296.9
Zn	190.1	0.0	708.3
As	78.1	133.3	122.4
Se	96.4	54.2	33.9
Cd	0.0	0.0	10.4
Ba	7.8	0.0	114.6
Ce	44.3	412.5	1,010.4
Pb	20.8	29.2	359.4
Na	203.1	695.8	3,932.3
Mg	614.6	75.0	4,036.5
K	510.4	633.3	2,489.6
Ca	4,515.6	3,241.7	19,447.9
Temp.	19.4	20.5	19.5
R.H.	57.1	66.4	72.4

*: Unit of PM10 is μg/m³

태로 샘플링한 개체들이 많았으며, 주로 지점 B와 C에 위치한 고등학교에서 측정된 개체가 주로 속하였다. 두 지점의 고등학교는 매우 발달된 상권으로 주변 교통량과 유동인구가 많은 지점이다. PM10의 농도는 114.6 μg/m³의 농도를 보였으며, 원소 성분 중에서 Ca (4,515.6 ng/m³), Fe (888.0 ng/m³), Mg (614.6 ng/m³), Al (403.6 ng/m³), K (510.4 ng/m³), Na (203.1 ng/m³) 등이 높게 나타났다. 이러한 물질은 주로 백묵과 토양기원 성분으로, 앞서 언급한 것과 같이 준개방상태의 환경이었으므로 토양 등 외부 유입물이 많았을 것이라 사료되며, 내부요인으로 백묵에 의한 영향으로 판단된다. 또한 군집 1에 속하는 개체의 75%가 3층에 위치하나 외부 영향인 토양성분의 경우 학생들의 외부활동에 의해 유입되는 경우로 2층과 3층간의 큰 차이는 없는 것으로 사료된다.

두 오염원에 기인한 입자로 평가하기 위한 비교자료로써 토양과 백묵의 조성에 대한 중량농도 조성비율 자료를 표 5에 제시하였다. 토양 조성은 타 연구에서 분석된 토양 성분분석 자료로 토양성분은 지각특성, 주변 대기오염도에 따라 차이를 보일 수 있다 (이학성 등, 2005). 토양 분석 자료에 의하면 Si, Al, Fe, K 순으로 높은 비율을 보였다. 분필성분 자료는

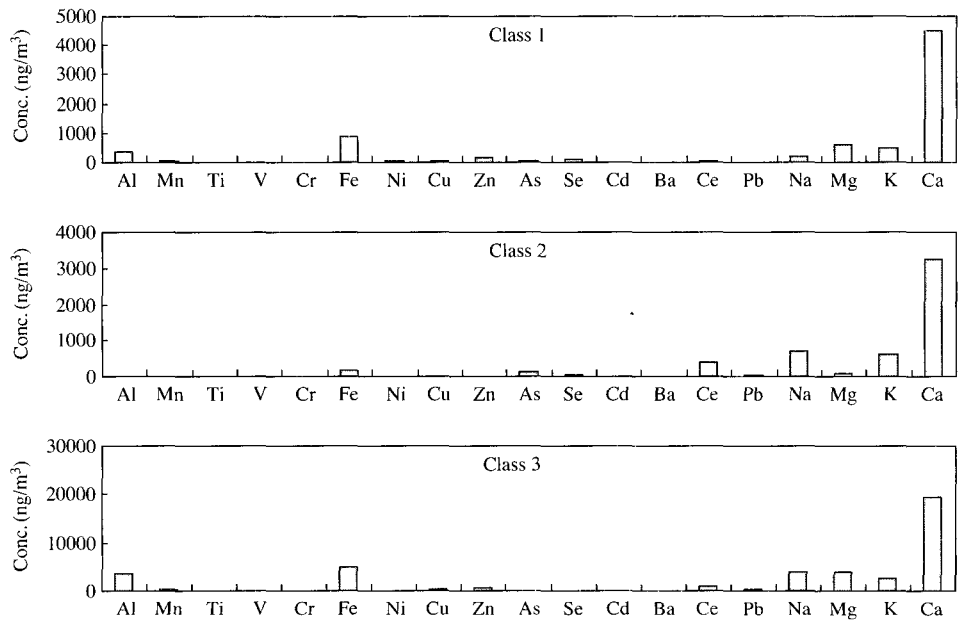

Fig. 6. Average concentrations of inorganic components for the three homogeneous PM10 classes.

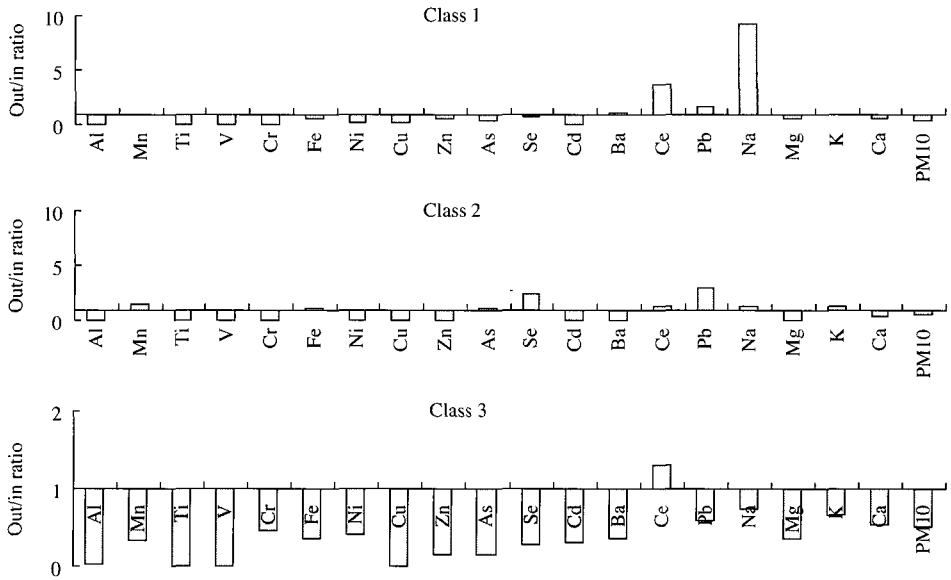

Fig. 7. Outdoor and indoor ratio of PM10 and inorganic components for each class.

Table 5. Composition of inorganic elements in the soil source and chalks. Each value is the weight percent of chemical species.

Chemical species	Soil[a)]	Chalk
Ca	0.34±0.04[b)]	42.41±0.08
Al	11.09±0.07	16.55±9.63
Ce	–	10.72±2.37
Na	0.17±0.04	8.37±1.41
Mg	0.06±0.01	7.31±3.51
Fe	8.44±0.03	4.76±3.31
K	2.47±0.06	3.62±1.23
Si	17.14±0.08	2.14±0.50
Cu	0.03±0.002	0.81±0.69
Mn	0.16±0.006	0.68±0.4
Ba	–	0.57±0.59
As	–	0.56±0.05
Se	–	0.41±0.03
Zn	0.04 ±0.004	0.40±0.20
Cr	–	0.17±0.02
Hg	–	0.17±0.10
Ti	0.66±0.022	0.12±0.05
Pb	0.01±0.002	0.10±0.01
Ni	–	0.06±0.01
V	–	0.06±0.01
Cd	–	0.01±0.00

[a)]Lee, H.S. *et al.* (2005): Each value is the weight percent of chemical species in the PM total mass
[b)]Standard error

일선 학교에서 가장 많이 사용되고 있는 국내 5개사의 흰색분필과 색분필(노랑, 빨강, 파랑)을 포함하여 총 40개의 분필을 분석한 자료이다. 각 분필은 분쇄한 후 앞장의 필터전처리 방법과 동일한 방법을 이용하여 전처리한 후 ICP-AES로 분석한 자료이다. 분필의 기본재료인 석회성분인 Ca이 총 42.4%로 비율이 가장 높았으며 그 뒤로 Al (16.6%), Ce (10.7%), Na (8.4%), Mg (7.3%), Fe (4.8%) 등의 순서로 분석되었다(장철순, 2005).

두 번째 군집은 교통량이 많지 않은 아파트 단지 내에 위치한 지점 D로 에어컨을 사용한 개체들로 주로 구성되었다. 군집 2로 분류된 각 개체들은 PM10 농도가 62.5 μg/m^{3}로 비교적 낮았으며, 대부분의 무기원소의 농도가 다른 두 군집에 비해 낮게 조사 되었다. 오염물질의 농도는 Ca (3,241.7 ng/m^3)이 가장 높았으며, Na (695.8 ng/m^3), K (633.3 ng/m^3) 등이 비교적 높은 농도를 보였다. 이러한 원소는 주로 백묵 성분으로 주로 실내에서 생성된 오염물질의 영향으로 판단된다.

세 번째 군집의 개체는 주로 비오는 날들이 많았으며, 준 밀폐 상태에서 샘플링을 한 군집으로 다른 두 군집에 비해 PM10 농도와 모든 무기원소의 농도가 월등하게 높게 나타났다. PM10 농도는 237.7

$\mu g/m^3$로 매우 높았으며, 무기원소는 백묵조성 및 토양기원 성분인 Ca (19,447.9 ng/m^3), Fe (5,085.9 ng/m^3), Al (3,484.4 ng/m^3) 순으로 높은 농도를 보였다. 또한 Mg (4,036.5 ng/m^3), Na (3,932.3 ng/m^3)과 같이 해염조성 물질이 다른 군집과 비교하여 매우 높은 농도를 보였다. 이러한 이유는 군집 3에 속하는 개체는 해안지역에 위치한 지점 A의 고등학교로 해염입자의 영향을 많이 받은 것으로 판단된다. 또한 Zn (708.3 ng/m^3), Pb (359.4 ng/m^3), Cu (296.9 ng/m^3), Mn (221.4 ng/m^3)등 기타 인위적 미량원소들이 매우 높은 농도를 기록하였다. 군집 3에서 높은 농도를 보인 미량원소의 특성을 고찰해보면, Zn의 주요 오염원은 석탄 및 기름연료 연소, 철 및 비철 관련 금속산업, 자동차 타이어 마모, 자동차 브레이크 라이닝 마모 등에 기인하는 성분이다 (Hopke, 1991). Pb은 페인트 안료, 도자기 유약, 포장지 등 산업 전반에 걸쳐 폭 넓게 사용되고 산업, 안료 등에 이용된다. 과거 유연휘발유 사용 시에는 자동차가 납의 주요 오염원이었으나 현재 유연휘발유 사용금지에 의해 자동차에 의한 납의 영향이 급속히 줄어들고 있다. Cu는 전기도금, 쓰레기 소각 및 다양한 금속 합금과정에서 발생하는 것으로 알려져 있다. Mn의 경우는 망간광석을 분쇄하는 작업, 망간강의 아크용접, 절단, 건전지 제조 등의 산업장에서 주로 발생한다. 따라서 군집 3의 개체의 경우 각종 산업공정 및 연료 연소 등에 의한 복합적인 인위적 배출원에 의한 영향을 받는 개체들로 분석되어지며, 이는 자동차관련 산업 및 다양한 업종을 포함하는 공단지역에 위치한 지점 A와 지점 B의 고등학교에 포함된 개체의 특성이 잘 반영된 것으로 사료된다. 또한 세 번째 군집에 해당하는 개체들에 포함된 PM10의 외기농도는 76.6 $\mu g/m^3$로 군집 1의 43.1 $\mu g/m^3$, 군집 2의 43.8 $\mu g/m^3$ 보다 높은 농도를 기록하였다. 이와 같은 특성은 실내 오염농도가 외기의 높은 영향을 받는 것으로 판단할 수 있다. 그러나 실내외 비율을 비교한 그림 7에 의하면, 군집 3의 모든 원소가 실내에서 매우 높은 비율을 보였으며, 이러한 특성은 외기에서의 유입된 오염물질의 영향뿐만 아니라, 비오는 날 분진농도 및 원소의 농도가 높게 나타난 또 다른 이유는 비와 우산에 묻은 수용성 및 불용성 분진이 건조 후 재비산된 것으로 사료된다. 한편, 세 번째 군집이 두 번째 군집과 비슷하게 밀폐상태를 유지시켰지만 많은 농도차이가 나는 것으로 보아 두 번째 군집의 경우 에어컨과 같은 기계적 환기가 실내공기질 개선에 많은 도움이 되는 것으로 판단된다.

4. 결 론

경기도 안산시와 수원시에 위치한 고등학교 4곳에서 포집한 실내 27개, 외기 11개의 PM10 자료를 이용하여 19개의 성분 및 환경조건 등을 분석한 후 형태 분류를 통한 오염 특성을 분석하였다. 이를 위해 군집분석법과 분산주성분분석법을 응용한 형태인식기법을 적용하였다. 먼저 군집분석 결과 3개의 군집을 분류하였으며, 분류된 군집은 분산 주성분분석을 적용하여 95% 확률로 검증하여 이상치를 제거한 순수한 군집을 도출하였다.

분류된 각 군집의 패턴은 뚜렷하게 분류되었으며, 첫 번째 군집은 준 개방 상태로 샘플링되었고, Ca, Fe, Mg, K, Al, Na 등 지각원소와 분필 성분이 높은 영향을 미치는 것으로 판단된다. 두 번째 군집은 주로 에어컨을 사용한 날로 PM10 및 대부분의 원소농도가 낮았으며, 주로 영향을 미친 원소는 Ca, K, Na, Ce 등 주로 분필조성과 유사한 성분이 높아 내부 오염원에 의한 영향으로 판단된다. 마지막 군집은 주로 비가 온 날들이 포함되었으며, PM10 농도 및 모든 원소농도가 매우 높은 것으로 분석되었다. 분필 및 지각원소 외에 Zn, Pb, Cu, Mn 등 기타 인위적 배출원에 의한 미량원소들이 높은 농도를 보였으며, 이는 각종 연소 및 산업공정에 의한 영향으로 판단된다.

본 연구의 결과는 많은 학생이 제한된 공간에서 생활하는 학교교실에서의 오염도 패턴을 분류하므로 학교 공기질 환경개선을 위한 기초자료 및 통계분석 기법을 제공할 수 있을 것이다. 또한 추후 채취장소 및 채취기간 등을 확대하고, 환기조건 및 창문의 개폐 여부 등 환경조건을 다양하게 고려하여 조사한다면 학교 공기질 개선에 도움을 줄 수 있는 새로운 방법론으로 응용될 수 있을 것으로 사료된다.

감사의 글

본 연구의 일부는 1999년 한국학술진흥재단 대학

부설연구소 지원과제 (과제번호 : KRF-2003-D00015)의 일환으로 수행되었으며, 이에 감사드립니다.

참 고 문 헌

교육인적자원부(2005) 학교보건법, 동법 시행규칙.

김기영, 전명식 (1994) 다변량 통계자료분석, 자유아카데미.

김동술, 김형석 (1990) Pattern recognition을 이용한 지하상가에서의 대기오염물질의 농도 분석에 관한 연구, 한국대기보전학회지, 6(1), 1-10.

남보현, 황인조, 김동술(2002) 분산주성분 분석을 이용한 실내환경 중 PM10 오염의 패턴 분류, 한국대기보전학회지, 18(1), 25-37.

신은상, 최민규, 선우 영, 정용삼(2002) 서울지역 PM10 중 미량원소의 특성 평가, 한국대기환경학회지, 18(5), 363-372.

이학성, 강충민, 강병욱, 이상권(2005) 수용모델을 이용한 서울지역 미세먼지에 영향을 미치는 배출원 특성에 관한 연구, 한국대기보전학회지, 21(3), 329-341.

임종한(2005) 학교 및 보육시설의 실내공기 오염과 어린이의 환경성 질환 실태, 아토피 스톱 프로젝트 심포지엄.

장철순(2005) 중등학교 내 PM10 농도경향 및 분필의 기여도 추정, 경희대학교 환경학과 석사학위 논문.

환경부(2003) 다중이용시설등의 실내공기질 관리법.

Albano, C., G. Blomqvist, D. Cooman, W.J. Dumm, U. Edlund, B. Eliasson, S. Helberg, E. Johansson, B. Norde'n, D. Johels, M. Sjöström, B. Söderström, H. Wold, and S. Wold (1981) Pattern recognition by means of disjoint principal component models (SIMCA): philosophy and methods, A plenary lecture given at the symposium on applied statistics, Copenhagen.

Box, G.E.P., W.G. Hunter, and J.S. Hunter (1978) Statistics for Experimenters, Wiley-Interscience, New York.

Branden, K.V. and M. Hubert (2005) Robust classification in high dimensions based on the SIMCA Method, Chemometrics and Intelligent Laboratory Systems, 79(28), 10-21.

Coomans, D., M. Jonckheer, I. Broeckaert, D.L. Massart, and S. Wold (1981) Pilot Study of the applicability of the SIMCA Pattern Recognition Method to clinical problems, using thyroid function tests as an example, Patroonherkenning Laboratoriu-monderzoeken, Hoofdstuk, 14.

Dolring, S.R., T.D. Davies, and C.E. Pierce (1992) Cluster Analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry-method and application, Atmos. Envrion., 26A(14), 2575-2581.

Hopke, P.K. (1976) Application of multivariate analysis to the interpretation of the chemical and physical analysis of lake sediments, J. Envi. Sci. Health, IIA, 367-383.

Hopke, P.K. (1985) Receptor Modeling in Environmental Chemistry, Wiley Interscience, New York, USA.

Hopke, P.K. (1991) Receptor Modeling for Air Quality Management, Elsevier Science Publishing Company Inc., New York, USA.

Massart, D.L. and L. Kaufman (1983) The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis, John Wiley & Sons, New York.

Olivar, D.C. (1973) Aggregative Hierarchical Clustering Program Write-up, National Bureau of Economic Research, Cambridge, Ma, USA.

SIMCA-3B Manual (1984) A Pattern Recognition Program for CPM and MS-DOS Based Microcomputers, Principal Data Components, 2505 Shepard Blvd, Columbia, Mo, USA.

Wang, L., F.S.C. Lee, and X. Wang (2005) Near-infrared spectroscopy for classification of licorice (Glycyrrhizia uralensis Fisch) and prediction of the glycyrrhizic acid (GA) content, Food Science and Technology, Article in Press.

Wold, S. (1976) Pattern recognition by means of disjoint principal components models, Pattern Recognition, 8, 127-139.

Wold, S. and M. Sjöström (1977) SIMCA: a Method for Analyzing Chemical Data in Terms of Similarity and Analogy, Chemometrics Theory and Application, edited by B.K. Kowalski, ACS Symposium Series 52, 243-282.