

군집분석법과 분산주성분분석법을 이용한 대기분진시료의 분류 Classification of Ambient Particulate Samples Using Cluster Analysis and Disjoint Principal Component Analysis

유 상 준 · 김 동 술

경희대학교 환경학과 대기오염연구실 및 자연과학종합연구소
(1996년 10월 2일 접수, 1997년 1월 24일 채택)

Sang-Joon Yoo, Dong-Sool Kim

*Air Pollution Laboratory, Department of Environmental Science, Kyung Hee University
and Institute of Natural Science*

(Received 2 October 1996; accepted 24 January 1997)

Abstract

Total suspended particulate matters in the ambient air were analyzed for eight chemical elements (Ca, Co, Cu, Fe, Mn, Pb, Si, and Zn) using an x-ray fluorescence spectrometry (XRF) at the Kyung Hee University - Suwon Campus during 1989 to 1994. To use these data as basis for source identification study, membership of each sample was selected to represent one of the well defined sample groups. The data sets consisting of 83 objects and 8 variables were initially separated into two groups, fine ($d_p < 3.3 \mu\text{m}$) and coarse particle groups ($d_p > 3.3 \mu\text{m}$). A hierarchical clustering method was examined to obtain possible member of homogeneous sample classes for each of the two groups by transforming raw data and by applying various distances. A disjoint principal component analysis was then used to define homogeneous sample classes after deleting outliers. Each of five homogeneous sample classes was determined for the fine and the coarse particle group, respectively. The data were properly classified via an application of logarithmic transformation and Euclidean distance concept. After determining homogeneous classes, correlation coefficients among eight chemical variables within all the homogeneous classes for calculated and meteorological variables (temperature, relative humidity, wind speed, wind direction, and precipitation) were examined as well to intensively interpret environmental factors influencing the characteristics of each class for each group. According to our analysis, we found that each class had its own distinct seasonal pattern that was affected most sensitively by wind direction.

Key words : aerosol classification, cluster analysis, principal component analysis, SIMCA, Suwon

1. 서 론

서울 및 수도권 지역에는 다양한 오염원이 산재하고 있으며 각종 오염물질이 배출되고 있다. 이중 입자의 크기, 화학적 조성 등 물리화학적으로 다양한 특성을 가진

분진은 주요 오염물질의 하나로, 수많은 오염원으로부터 배출되고 있다. 부유분진은 보통 2.5~5 μm 를 중심으로 미세입자와 거대입자로 분류된다. 거대입자는 토양과 해염 등 자연적 발생원에서 방출되나, 미세입자군은 주로 인간의 산업활동에 의해 배출된다. 입자상 물질의 크기, 화학적 조성, 농도분포, 물에 대한 친화력 등의 변

수는 인체에 미치는 영향과 밀접한 관계가 있다. 특히 직경 10 μm 이상의 입자들은 대부분 구강에서 제거되지만, 10 μm 이하의 입자들은 10~60% 정도 기도를 통과해 폐포에 침착되기 때문에 인체에 악영향을 줄 수 있다(Chow, 1995).

대기 중 부유분진의 농도에 직접적 영향을 미치는 각종 오염원의 정성적 및 정량적 분석은 분진오염의 관리 측면에서 매우 중요하다. 이를 위한 일반적인 접근 방법으로는 오염발생원(source)을 중심으로 하는 분산모델과 수용체(receptor)를 중심으로 한 수용모델이 있다. 후자는 수용체에서 오염물질의 특성을 분석한 후, 오염원의 기여도를 정량적으로 추정하는 방법론으로, 현재 우리나라에서와 같이 오염원 관련자료의 확보가 어려운 실정에는 적합한 방법이다(김동술과 이태정, 1993). 또한, 최근들어 분진에 함유된 미량성분에 대한 시료채취장치 및 분석기법의 발달, 컴퓨터의 보급 확산, 그리고 다양한 수리통계학적 모델의 개발 등으로 수용모델의 응용영역은 급속히 확대되고 있다. 특히, 다변량 통계분석법(multivariate statistical analysis)은 화학질량수지법(CMB: chemical mass balance analysis)과 함께 최근 들어 가장 많이 사용하는 수용모델이다.

본 연구는 수원근교에서 분진오염의 현황과 오염의 형태(pattern)를 파악하기 위하여, 우선 군집분석법(cluster analysis)을 이용하여 대기중 부유분진을 유사특성의 특정군집으로 분류하고, 분산 주성분 분석법(disjoint principal component analysis)을 이용하여 분류된 군집을 확률적으로 검증하였다. 또한, 최종적으로 도출된 순수군집(homogeneous cluster)을 해석하여, 연구대상지역에서의 분진오염의 특성을 구명하고자 하였다. 이 방법론은 이미 김동술과 김형석(1990)이 서울 시내 지하상가 및 지하통로에서 실내공기질의 관리 차원에서 응용한 바 있다.

2. 이론적 배경

2.1 군집분석

군집분석법은 자연과학 분야에서 널리 사용하는 응용 통계 분석법으로서, 자료집단 내에서 유사한 성질을 갖는 특정 그룹을 분류하는데 이용되고 있다. 군집분석은 군집의 개수, 내용, 구조 등이 완전히 알려지지 않은 상태에서, 개체 사이의 거리 또는 비유사도에 근거하여 군집의 특성과 관계를 분석하는 것이 주요 목적이다(김기영과 전명식, 1994). 이 분석법의 기본 원리는 두 object 사이의 거리를 기준으로 공간에서 비유사도를 측

정함으로써, 동일한 pattern을 가진 object를 분류하는 것이며, 거리 또는 비유사도가 작은 두 object는 동일한 군집에 소속된다(Hopke, 1991). 일반적으로 군집분석은 크게 위계 군집분석법(hierarchical cluster analysis)과 비위계 군집분석법(non-hierarchical cluster analysis)으로 구분한다. 비위계분석법은 군집의 수를 이미 알고 있거나 추정할 수 있을 때, object들을 정해진 특정 군집으로 최적 분배하는 방법이다. 반면, 위계 분석법은 한 군집이 다른 군집에 포함되지만, 군집간 중복이 허용되지 않고 군집의 수가 미지일 때, 계보형식의 수상도(dendrogram)를 이용하는 분석법이다. 위계 분석법은 비위계분석법에 비해 algorithm이 간단하고 계산시간이 상대적으로 짧으며, 전체 군집들간의 구조적인 관계를 수상도라는 2차원 상의 공간에 간단히 표현할 수 있다는 장점이 있다(Massart and Kaufman, 1983; 김기영과 전명식, 1994).

Object간 비유사도를 측정하는 방법에는 여러 가지 방법이 있으나, 본 연구에서는 일반적으로 많이 사용되는 유클리디안 거리(Euclidean distance), 유클리디안 제곱거리(squared Euclidean distance), 시티-블럭 거리(city-block distance 또는 Manhattan distance) 등 세 가지 방법을 이용하였으며, 그 수식은 다음과 같다;

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

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

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

여기서, D_{jk} 는 object X_j 와 X_k 사이의 거리를 의미하며, m 은 총 변수의 숫자이다.

군집분석법은 자료의 형태 윤곽을 쉽게 파악할 수 있으나, 거리산출 방법의 선정 및 유사도 준위의 결정 등 분석자의 주관적인 판단이 개입될 수 있으므로 연구의 최종 결과를 얻고자 할 때 세심한 주의가 필요하다.

2.2 주성분분석법

일단 군집분석을 통해 유사한 특성을 가진 object들이 군집으로 분류되면, 그 분류가 과연 올바르게 수행되었

는지의 여부를 정량적으로 규명해야 한다. 따라서 본 연구는 분산주성분분석법에 입각한 형태인식법 (pattern recognition)을 응용하였고, 이를 위해 SIMCA package (SIMCA-3B manual, 1984)를 이용하였다. 분산주성분분석법이란, 이미 알고 있거나 군집분석 등을 통해 생성된 군집 각각에 대해 독립적으로 주성분분석을 수행하는 방법론이다. SIMCA는 Wold (1976)에 의해 개발된 형태인식 기법으로, 분산주성분분석 모델을 기초로 한 통계 package이다. 국내에서는 김동술과 김형석 (1990)이 서울 시내 지하상가 및 지하통로의 공기오염도 분석에 적용한 바 있으며, 최근 들어서는 Sjögren *et al.* (1996)이 디젤연료의 배출가스 분석에 응용한 바 있다.

SIMCA의 분산주성분 분석에 의해 각 군집은 선형구조로 모델화되고, 한 개의 object가 고정된 확률 값에서 이미 모델화된 특정군집에 소속될 수 있는지의 여부를 임계거리 (critical distance)에 입각하여 결정한다. 일반적으로 입력자료는 training set과 test set을 포함하고 있다. 이전의 정보 또는 군집분석과 같은 통계적 방법을 통해 이미 알고 있는 군집을 training set이라 하며, 이 기지의 정보를 주성분 모델의 개발에 사용한다. Test set은 군집으로 분류되어야 하는, 즉 소속이 불확실한 object들의 집단을 말하며, 때로는 training set으로부터 확보된 군집에 대한 정보의 정확성 여부를 규명하는데 사용되기도 한다. 그리고, 모델 개발 후 어떤 군집에도 소속되지 않는 object를 특이점 (outlier)이라 한다.

SIMCA의 원리는 다중 Taylor 급수 (multiple Taylor's expansion)에서 부터 출발하며, 자료 y_{ik} 는 식 (4)와 같이 주성분수 k 개를 가진 주성분 모델로 fitting할 수 있다.

$$y_{ij}^{(q)} = y_i^{(q)} + \sum_{k=1}^A b_{ik}^{(q)} t_{kj}^{(q)} + e_{ij}^{(q)} \quad (4)$$

자료행렬 y_{ij} 는 q 번째 군집에서 k 개의 주성분을 가진 i 번째 변수와 j 번째 object의 측정치를 의미한다. 여기서 우변의 y 는 변수 i 의 평균치, b 는 변수 관련항, t 는 object의 값을 나타내고, 잔차 e_{ij} 는 y_{ij} 의 확률부분 (random part)을 설명하며, 측정오차 (measurement error)와 모델오차를 포함하고 있다. 측정오차가 모델오차보다 크다면, 모델은 자료를 잘 해석하고 있다고 볼 수 있다 (Wold, 1976). 잔차 e_{ij} 의 제곱 합은 모델화 된 군집과 그 군집에 속해 있는 특정 object 간의

거리를 나타낸다. 또한, 잔차로부터 각 모델에 대한 잔여표준편차 (residual standard deviation)가 계산되며, 이것은 측정공간에서 주성분 모델 주위에 신뢰구간을 형성하게 된다. 따라서, 잔여표준편차는 training set에서 outlier를 감지하는데 사용된다 (Albano *et al.*, 1981).

Test set의 object들은 군집모델들에 대한 fitting의 정도를 서로 비교하여 분류할 수 있다. 이것은 다중선형 회귀분석에 의해 각 군집모델에 object를 fitting하므로써 계산할 수 있다. 만약 소속이 불확실한 object p 를 분류하기 위해 군집 q 의 모델을 이용할 경우, object p 의 잔여표준편차가 군집 q 의 잔여표준편차보다 작으면 object p 는 군집 q 로 분류된다. 이와 같이 소속이 확실하지 않은 object는 모델화된 각 군집의 F-test를 통해 가장 가까운 군집에 속하게 되며, F의 임계값을 이용해 임계거리를 계산할 수 있다.

계산된 임계거리와 object의 잔여표준편차를 비교하여 군집에의 소속여부를 결정할 수 있다. 하나의 object는 하나의 군집에만 유일하게 소속되거나, 둘 또는 그 이상의 군집에 동시에 소속되거나, 또는 어떤 군집에도 속하지 않게 된다. 일반적으로 주성분 모델은 주성분의 수가 증가할수록 잔차는 작아지며, 따라서 모델은 자료에 더욱 근접하여 fitting이 가능하게 된다. 그러나, 결과에 대한 모델의 유효성 (validity)은 제약을 받게 될 것이다 (Wold *et al.*, 1984). 따라서 주성분의 수를 결정할 때는 이 두가지 측면을 동시에 고려해야 하며, SIMCA에서는 횡유효도 검사 (cross validation)를 통해 주어진 군집에 대한 통계적으로 유효한 주성분의 수를 결정한다. 더욱 자세한 SIMCA의 원리 및 수식은 다음 문헌을 참고할 수 있다 (Wold and Sjöström, 1977; Wold, 1976).

3. 실험방법

분진의 포집은 89년 11월부터 현재까지 경희대학교 수원캠퍼스에서 진행 중에 있다. 이들 자료 중 94년 9월까지의 무기원소 분석자료를 본 연구의 원자료 (raw data)로 이용하였다. 분진의 포집장소는 경부고속도로에 인접한 경기도 용인시 신갈지점에서 서측으로 2 km, 수원시에서 동측으로 약 10 km에 위치해 있으며, 주변에는 면적 231 ha의 신갈호수 및 크고 작은 공장들이 위치하고 있다. 이 기간동안 포집위치 주변에서 크고 작은 공사가 이루어졌는데, 중앙도서관 신축공사가 1992년 6월 17일에 착공되어 1995년 12월에 완공되었

으며, 정문 앞 도로공사가 94년 5월에 착공되어 94년 하반기에 완공되었다. 포집장치로는 입경별 분진포집기인 9단 cascade impactor (일본 Dylac co., Model KA-200)를 이용하여 대기 부유분진을 입경별로 포집하였다. 유량은 28.3 L/min으로 고정하여 7일에서 30일 간격으로 연속 측정하였으며, 포집 부하량의 증가로 인해 유량이 25.1 L/min 이하로 급속히 떨어질 때 교체하였다. 여지로는 자체오염이 적고, 미량원소의 측정에 적절한 직경 80 mm, pore size 0.45 μm 인 membrane filter (Gelman Science사: model GN-6)를 사용하였으며, 이 기간동안 총 83개의 set가 채취되었다. 각 set는 9장의 여지로 이루어져 있다. 분진을 포집한 여지는 테시케이터에 보관한 후 원소분석을 하였다. 포집된 시료는 일정한 크기로 절취하여 비파괴분석법인 X-선 형광분광기를 이용하여 각 단별로 농도를 측정하였다. 본 연구에 사용된 X-선 형광분광기는 일본 Rigaku사의 X-ray fluorescence spectrometer (model 3063)이다. 대기분진 시료의 원소분석은 backup filter ($<0.43 \mu\text{m}$)를 제외한 제 1단에서 8단까지의 membrane 여지시료만을 분석하였다. Membrane 여지는 전처리 과정없이 직경 3cm 크기로 절취하여 holder에 넣고, 포집된 분진의 비산을 막기위해 투명한 mylar막으로 덮은 후, 진공상태에서 무기원소의 농도를 측정하였다. 표준시료로 미국 NIST사의 표준물질 SRM 1832 (일련번호 195)와 SRM 1833 (일련번호 1207)을 사용하였다.

참고로 본 연구에 사용된 filter의 바탕시험 및 포집공기량을 감안한 XRF의 실제검출한계 (working detection limit)는 Si 1.7 ng/m³, Ca 7.8 ng/m³, Mn 16.3 ng/m³, Fe 9.1 ng/m³, Cu 2.5 ng/m³, Zn 1.9 ng/m³, Pb 6.4 ng/m³, Co 1.2 ng/m³, V 2.2 ng/m³, K 6.8 ng/m³이다.

4. 결과 및 고찰

4.1 자료의 구조 및 변환

입경별로 분석된 Ca, Co, Cu, Fe, Mn, Pb, Si, Zn 등 8개 무기원소의 농도자료를 원자료로 이용하였다. 본 연구에서는 포집기 1단에서 8단까지의 무기원소 농도자료 중, 입자직경 3.3 μm 를 기준으로 전체 입자를 양분하였다. 즉, 포집기 1단에서 4단까지의 농도자료를 합하여 미세입자 그룹으로 ($d_p < 3.3 \mu\text{m}$), 3.3 μm 이상인 포집기 5단에서 8단까지의 자료를 합하여 거대입자 그룹으로 분류하였다 ($d_p > 3.3 \mu\text{m}$). 포집기간은 89년

11월부터 94년 9월까지이며, 이 기간동안 포집된 시료의 수, 즉 object의 수는 총 83개이다. 각 object는 8개 무기원소에 대한 농도자료, 즉 변수(variable)를 가지고 있다.

본 연구에서는 군집분석에 앞서 원자료의 변량분포가 좌우대칭형이 되도록, 로그변환과 제곱근변환 등 두가지 방법으로 자료를 변환하였다. 또한, z-score를 이용한 표준화를 통해 변수들이 평균 0, 분산 1의 동일한 가중치를 갖도록 하였다.

4.2 군집분석 결과

다양한 군집분석 방법 중 응집위계분석법 (hierarchical cluster analysis)을 적용하였으며, 비유사도 기준으로는 일반적으로 많이 사용되는 유클리디안, 유클리디안 제곱, 맨하탄 거리 등 세가지 방법을 각각 적용하였다. 군집분석시 원자료에 대해 자료변환을 거치지 않은 자료, 로그변환을 거친 자료, 그리고 제곱근변환을 거친 자료 등 3개의 자료를 입력자료로 사용하였다. Z-score를 이용한 자료의 표준화는 3개의 자료 모두에 적용하였다. 이 3개의 자료군에 대하여 상기 3가지의 비유사도 기준을 각기 적용하고 군집분석을 수행하였다. 따라서, 미세입자와 거대입자 그룹에 대해 각각 9개씩의 수상도(dendrogram)를 얻을 수 있었다. 그림 1은 미세입자의 원자료를 로그변환한 후, 유클리디안 거리를 비유사도 기준으로 군집분석을 수행한 결과 얻은 수상도이다. 수상도에서 임의의 수직선 B를 그렸을 때 좌측에 위치한 object 집단들이 각기 군집을 형성한다. 그림에서 어느 군집에도 포함되지 않은 object (*표)를 outlier로 분류하였다. 그리고 이 outlier들을 묶어 또 하나의 군집(class 99)을 만들었다.

수상도에서 비유사도의 준위를 선정하고 군집의 수를 결정하는 과정은 군집분석에서 가장 중요한 과제이지만, 분석자의 주관적인 판단에 전적으로 의존하기 때문에 많은 주의가 필요하다. 본 연구에서는 상호 결과비교가 용이하도록 가급적 군집의 수를 일치시켜 주었으며, 연구 초기에는 outlier 집단을 포함하여 5~8개 사이에서 군집의 수를 결정하였다.

4.3 SIMCA 응용 결과

군집분석을 수행하여 생성된 군집 각각을 training set로 하고, SIMCA package를 이용하여 분산주성분 분석을 수행하였다. 주성분분석시 가장 중요한 과정 중 하나는 통계적으로 유효한 주성분의 수를 결정하는 것이며, 이를 위하여 횡유효도(cross validation)검사를 수

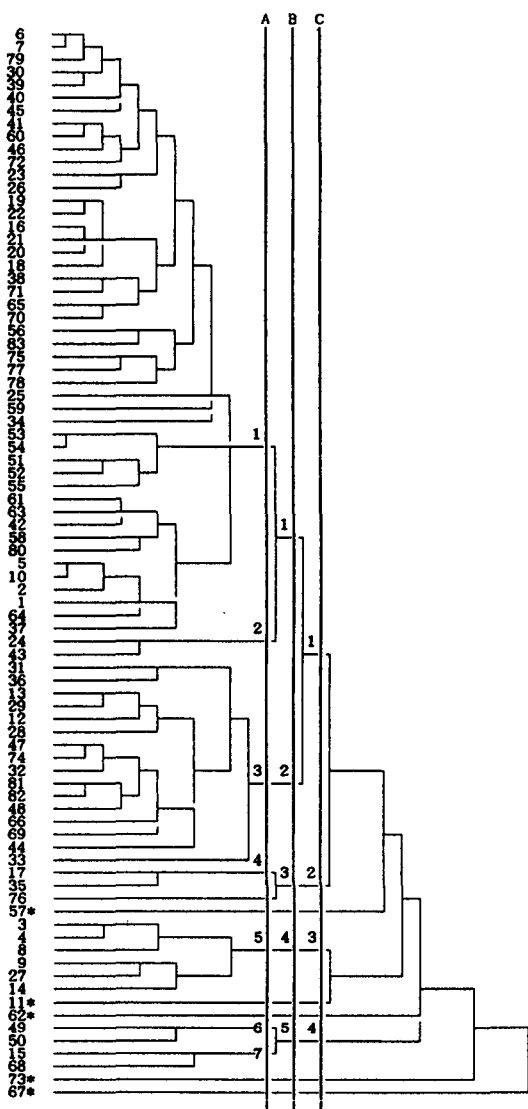


Fig. 1. A dendrogram for fine particle group based on the Euclidean distance after logarithmic transformation shows various levels of dissimilarities as example.

행하였다. 표 1은 미세입자와 거대입자 그룹 각각에 대하여 분산주성분분석을 수행한 결과 산출된 군집별 주성분의 수이다. 또한 괄호 안의 숫자는 각 군집의 object 수이다. 하나의 군집에 소속된 object의 수가 3개 이하로 매우 작을 경우, 유효한 주성분의 수가 결정되지 않았다. 이것은 군집분석시 자료에 대한 분류가 올바르게 수행되지 않았음을 의미한다.

따라서 이를 배제하고 유효한 주성분의 수가 결정되는 분석결과만을 선택하였다. 그 결과, 미세입자의 경우, 로그 변환과 제곱근 변환을 거친 자료 각각에 대하여 유클리디안 거리와 유클리디안 제곱거리를 비유사도 기준으로 군집분석을 수행한 결과가 가장 양호하였다. 그리고, 거대입자의 경우, 로그 변환을 거친 자료 중 유클리디안 거리를 적용한 경우와, 제곱근 변환을 거친 자료중 유클리디안 제곱거리를 적용한 경우만이 양호하였다.

일단 모든 군집에 대해 주성분의 수가 결정되면 군집과 object에 대한 잔여표준편차를 구할 수 있으며, 이를 이용하여 Coomans의 표준결정도(Coomans *et al.*, 1981)를 작성할 수 있었다. Coomans의 표준결정도를 이용하여 object들의 군집 소속여부를 확률적으로 검증할 수 있었다. 그림 2는 미세입자에 대하여, 로그 변환을 거치고 유클리디안 거리를 이용하여 군집분석을 수행한 결과 얻은 군집에 대해, 표준결정도를 이용하여 군집간을 비교한 예이다. 그림에서 수직과 수평으로 그은 선은 90% 확률의 임계거리선이고, 이 선에 의해 4개의 영역으로 나누어진다. 이 영역 중 좌측하단은 두 군집이 공존하는 영역이고, 우측상단은 두 군집의 어디에도 속하지 않는 영역이다. 따라서 좌측상단 및 우측하단 영역에 object가 소속될 때 90% 확률로 해당군집에 속하게 되고, 그 외 영역에 소속될 때 object는 해당군집에서 배제된다. 이와 같이 표준결정도에 의거하여 군집분석 수행시 내재된 주관적인 결정, 즉 비유사도의 임의적 결정을 객관화하고 순수군집을 도출하였다. 표 2는 순수군집과 군집에 속한 object를 보여주고 있다. 또한 미세입자군과 거대입자 그룹을 모두 분석한 결과, 로그 변환을 수행하고 유클리디안 거리를 비유사도 기준으로 분석하였을 때가 기타 거리를 기준으로 분석하였을 때보다 outlier의 수가 적게 생성되었다. 따라서, 본 연구에 사용된 자료의 해석에 있어서, 자료의 전처리는 로그 변환을, 군집분석시에는 비유사도 기준으로 유클리디안 거리를 이용할 때, 가장 이상적인 분류가 가능하였다.

앞의 그림 1의 수상도는 미세입자에 대하여 이미 결정되었던 비유사도 준위(line B)를 하부(line A)와 상부(line C)로 한 단계씩 이동조절하여 표시하고 있다. 여기서, 미세입자와 거대입자 각각에 대하여 outlier 군집을 포함한 군집의 수가 5개와 8개로 조절되었으며, 새롭게 만들어진 군집들 각각에 대하여 다시 주성분분석을 수행하였다. 이중 군집 수가 6개일 때 군집간 중복과 outlier의 수가 비교적 적었고, 따라서 군집의 수가 6개인 경우(outlier 군집포함)를 선택하여 자료해석에 이용하였다. 즉, 미세입자에서 line A를 비유사도 준위로

Table 1. Number of significant principal components and of the assorted objects (in parenthesis) for fine and coarse particle groups based on a cross validation test.

Group	Class	Without Transformation			Log Transformation			Square Root Transformation		
		ED	SED	CBD	ED	SED	CBD	ED	SED	CBD
Fine Particle Group	Class 1	0 (54)	0 (54)	1 (31)	0 (49)	0 (50)	0 (25)	1 (33)	1 (33)	1 (57)
	Class 2	0 (17)	0 (17)	0 (7)	0 (16)	0 (15)	0 (20)	0 (12)	0 (12)	0 (9)
	Class 3	0 (4)	0 (4)	0 (7)	0 (3)	0 (3)	0 (24)	0 (21)	0 (21)	- (2)
	Class 4	- (2)	- (2)	- (3)	0 (6)	0 (6)	0 (6)	0 (6)	0 (6)	1 (7)
	Class 5			0 (14)	1 (4)	1 (4)	- (2)	1 (4)	1 (4)	1 (4)
	Class 6			1 (7)			- (2)			
	Class 7			0 (6)						
	Class 99	0 (6)	0 (6)	0 (8)	0 (5)	0 (5)	0 (4)	0 (7)	0 (7)	0 (4)
Coarse Particle Group	Class 1	2 (49)	1 (38)	2 (49)	0 (11)	0 (11)	1 (52)	0 (15)	2 (49)	2 (53)
	Class 2	0 (4)	1 (15)	0 (14)	1 (44)	1 (44)	0 (9)	1 (20)	0 (9)	1 (8)
	Class 3	0 (14)	0 (10)	0 (8)	0 (9)	0 (4)	1 (9)	1 (32)	0 (4)	0 (8)
	Class 4	0 (7)	0 (4)	0 (4)	0 (4)	0 (9)	1 (8)	0 (7)	2 (7)	1 (8)
	Class 5	- (2)	1 (6)	- (2)	1 (8)	1 (8)	- (2)	- (3)	1 (7)	- (3)
	Class 6		- (2)			- (2)				
	Class 99	0 (7)	0 (8)	0 (6)	0 (7)	2 (5)	0 (3)	0 (6)	0 (7)	- (3)

ED : Euclidean distance
 SED : squared Euclidean distance
 CBD : city-block distance

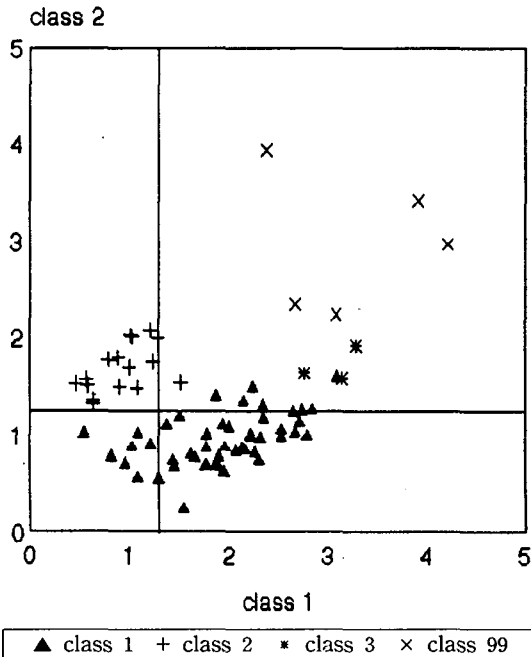


Fig. 2. Coomans' standard decision plot for fine particles between classes 1 and 2.

결정하였을 때, 8개의 입자군을 얻을 수 있었으나 object의 수가 3개 미만인 군소군집이 4개가 생성되어

유효 주성분을 결정할 수 없었다. 또한, line C를 비유사도 준위로 결정하였을 때, 5개의 군집을 얻을 수 있었으나, 황유효도 검사에 의한 각 object의 순수군집에의 소속율은 66%이었다. 반면, line B일 때의 소속율은 90%이었다. 따라서 본 연구에서는 line B를 비유사도 준위로 선정하고 군집의 수를 6개로 결정하였다. 이와 같은 방법을 거대입자에도 적용하였다.

4.4 군집의 특성 및 형태

군집분석과 분산주인자분석을 통하여 군집의 수를 결정할 수 있었고, outlier를 배제한 순수군집을 얻을 수 있었다. 미세입자 및 거대입자 그룹 각각에 대하여 순수군집의 수는 6개 (Class 99포함)로 결정되었으며, 이들 군집은 연구대상지역의 오염형태를 대변할 수 있다. 본 연구는 생성된 군집의 특성을 파악하고 물리적으로 해석하기 위해 기상자료를 활용하였으며 제 2차 통계분석을 시도하였다.

표 3은 최종 생성된 순수군집들에 대한 원소별 평균농도를 보여주고 있다. 또한, 도표는 각 object의 시료채취기간대의 온도, 습도, 풍속, 강수량 등 기상자료(기상청, 1989~1994)를 취합하고 군집별로 평균한 자료를 보여주고 있다. 그림 4는 표 3의 미세입자 그룹에 대하여, 5개 군집의 경향을 재구성한 것이다. 각 군집에 소속된 object들은 통계적으로 매우 유사한 특성을 지니

Table 2. Classified objects for fine and coarse particle groups.

Group	Class	Objects
Fine Particle Group	Class 1	1, 2, 5, 10, 16, 19, 20, 21, 22, 23, 25, 37, 40, 41, 42, 52, 53, 54, 56, 58, 59, 60, 61, 63, 64, 65, 70, 71, 72, 77, 78, 80, 83
	Class 2	13, 28, 29, 31, 32, 36, 47, 48, 66, 69, 74, 81, 82
	Class 3	17, 35, 76
	Class 4	3, 4, 8, 9, 14, 27
	Class 5	15, 49, 50, 68
	Class 99	6, 7, 11, 12, 18, 24, 26, 30, 33, 34, 38, 39, 43, 44, 45, 46, 51, 55, 57, 62, 67, 73, 75, 79
Coarse Particle Group	Class 1	48, 49, 50, 69, 82, 83
	Class 2	6, 7, 13, 25, 26, 27, 29, 30, 32, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 51, 52, 53, 54, 55, 56, 58, 59, 60, 61, 65, 66, 67, 70, 72, 75, 77, 78, 79, 80, 81
	Class 3	16, 17, 18
	Class 4	5, 10, 62, 63
	Class 5	1, 2, 3, 4, 8, 9, 12, 28
	Class 99	11, 14, 15, 19, 20, 21, 22, 23, 24, 31, 33, 34, 44, 47, 57, 64, 68, 71, 73, 74, 76

Table 3. The mean concentrations of each component within each homogeneous class after disjoining principal component analysis.

Group	Class	Ca	Co	Cu	Fe	Mn	Pb	Si	Zn	Temp	RH	WS	PREP
Fine Group	Class 1	248.0	1.4	20.0	336.1	16.9	236.1	507.0	147.0	9.0	67.4	1.6	6.1
	Class 2	74.1	3.3	15.2	164.6	16.9	163.6	268.8	97.0	19.4	77.7	1.4	9.7
	Class 3	633.5	3.9	44.8	515.3	57.0	478.7	473.4	212.2	0.6	64.9	1.3	2.3
	Class 4	158.3	N.D.	1.1	166.4	N.D.	174.9	378.7	110.7	9.5	75.8	1.4	12.0
	Class 5	235.6	0.5	15.7	124.5	7.9	63.8	155.6	34.9	22.3	78.6	1.3	18.7
Coarse Group	Class 1	465.6	1.9	9.2	351.2	23.8	45.3	969.2	24.3	24.4	79.8	1.4	18.9
	Class 2	998.3	2.1	5.7	760.5	9.3	59.8	2302.8	39.8	12.0	71.1	1.5	7.0
	Class 3	1046.1	1.6	39.0	959.3	90.6	183.4	2781.3	118.3	-0.2	68.1	1.1	4.0
	Class 4	1463.9	1.5	0.3	1296.9	1.0	47.3	3916.8	49.6	6.6	75.0	1.9	6.0
	Class 5	736.3	0.3	0.1	517.2	N.D.	15.2	1760.5	33.9	7.6	69.4	1.5	6.9

Unit : chemical element = ng/m³
 temperature (TEMP) = °C
 relative humidity (RH) = %
 wind speed (WS) = m/s
 precipitation (PREP) = mm
 N.D. = Non Detectable

고 있었다. 또한, 표 4는 미세 및 거대입자 그룹에서 각각 도출된 5개의 순수군집에 대한 변수간의 상관성을 보여주고 있다. 각 순수군집에서 도출된 특성 및 오염형태는 다음과 같았다.

미세입자 그룹에서, 군집 1은 Si의 농도(507.0 ng/m³)가 높고, 풍속도 1.6 m/s로 높았다. 군집 1에 속한 object들의 구성을 살펴보면, 계절별로는 봄철 자료가 15개 (45.5%)로 가장 많았고, 연도별로는 포집장소 주변에서 토목공사가 활발히 진행되었던 92년 이후 자료가 13개 (39.4%)로 가장 많았다. 군집 1내의 원소간의 상관성은 매우 낮았다. 비록 이 군집이 Si 농도가 높았으

나, 이 군집이 미세입자군이며 Si, Fe, Ca 사이의 상관성도 낮아 토양오염원 (soil source)의 영향을 받았다고 해석하기에도 어려움이 있었다. 또한, 군집내 변수간의 음의 상관성을 현상적으로 해석하는 것은 불가능하였다.

군집 2는 Ca의 농도(74.1 ng/m³)가 매우 낮았으며, Co (3.3 ng/m³), Cu (15.2 ng/m³), Fe (164.6 ng/m³), Pb (163.6 ng/m³), Si (268.8 ng/m³), Zn (97.0 ng/m³)도 전체자료의 평균보다 낮은 수치를 보이고 있었다. 군집 2의 object들을 연도별로 분류하여 보면, 90년 (13)이 7.7%, 91년 (28, 29, 31, 32, 36)이 38.4%, 92년 (47, 48)이 15.4%, 93년 (66, 69, 74)

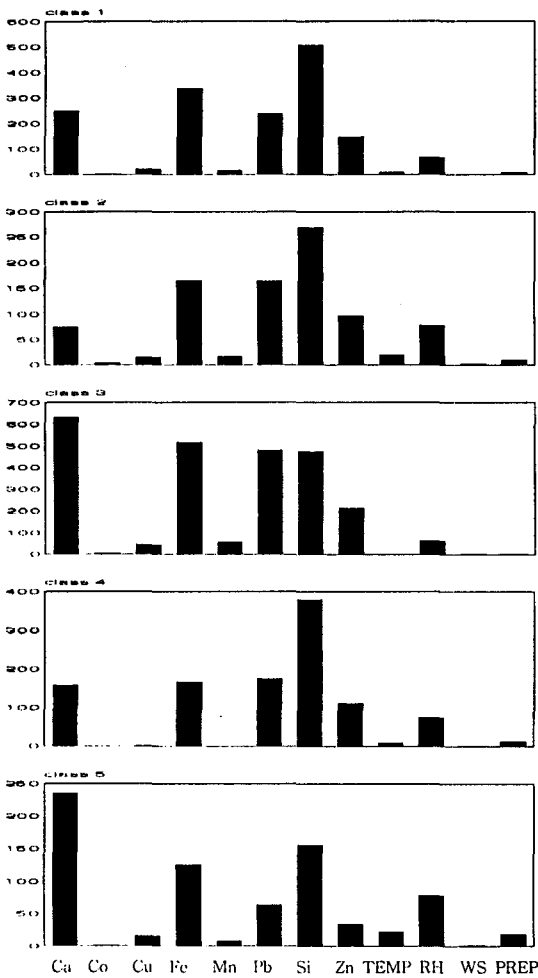


Fig. 3. The means for major chemical components / and environmental variables for the five ho- / mogeneous fine particle classes.

이 23.1%, 그리고, 94년(81, 82)이 15.4%로 91년 이후 자료가 군집 전체의 92.3%를 차지하고 있었다.

군집 3은 Ca (633.5 ng/m³), Co (3.9 ng/m³), Cu (44.8 ng/m³), Fe (515.3 ng/m³), Mn (57.0 ng/m³), Pb (478.7 ng/m³), Zn (212.2 ng/m³) 등 대부분의 원소의 농도가 높았으며, 온도(0.6°C), 상대습도(64.9%), 풍속(1.3 m/s), 강수량(2.3 mm)이 모두 낮았다. 또한, 군집 3은 겨울철 기간동안의 objects (91년 1월 29~2월 6일, 91년 11월 27일~12월 12일, 94년 2월 14일~2월 28일)로만 구성되어 있었다. 특히 이군집은 Fe와 Co, Zn와 Co, Fe와 Zn, Mn과

Pb, Mn과 Cu, Pb와 Cu 등이 매우 큰 상관성을 가지고 있었다. 즉, 각종 인위적 금속관련 오염원이 농축되어 있어 추후 오염원의 marker 추정연구에 도움이 될 것으로 사료된다.

군집 4는 Co, Cu, Mn의 농도가 특히 낮아 검출한계 이하의 수치를 보였다. 군집 4의 object를 연도별로 분류해 보면, 89년(3, 4)이 2개, 90년(8, 9, 14)이 3개, 91년(27)이 1개를 차지하며, 92년 이후 자료는 하나도 포함되지 않았다. 군집 4는 91년 이전 자료가 83.3%를 차지하고 있어, 91년 이후 자료가 92.3%인 군집 2와 대조되는 것을 알 수 있었다. 이 군집에서는 Si와 Ca가 상관성을 보이고 있었다.

군집 5는 Fe (124.5 ng/m³), Pb (63.8 ng/m³), Si (155.6 ng/m³), Zn (34.9 ng/m³)의 농도가 낮았으며, 온도(22.3°C), 상대습도(78.6%), 강수량(18.7 mm)은 높고, 풍속(1.3 m/s)은 낮았다. 또한, 군집 5는 대부분 여름철 기간동안의 objects (75%, 90년 9월 22일~10월 21일, 92년 8월 3일~8월 17일, 92년 8월 18일~8월 17일, 93년 7월 21일~8월 9일)로 구성되어 있어 여름철의 기후조건, 예를 들어 강우에 의한 세정효과 등의 영향을 받은 것으로 사료된다. 이 군집은 Ca와 Co, Cu와 Co, Mn과 Co, Cu와 Mn이 강한 상관성을 보이고 있었다.

거대입자 그룹에서, 군집 1은 objects 중 83.3%가 여름철 기간(92년 7월 16일~7월 31일, 92년 8월 3일~8월 17일, 92년 8월 18일~8월 31일, 93년 8월 11일~8월 28일, 94년 8월 16일~8월 30일, 94년 8월 31일~9월 16일)동안의 자료이며, Ca (465.6 ng/m³), Fe (351.2 ng/m³), Si (969.2 ng/m³), Zn (24.3 ng/m³)의 농도가 낮고, 온도(24.4°C), 습도(79.8%), 강수량(18.9 mm)은 비교적 높았다. 또한, 이 군집은 Si와 Fe이 강한 상관성을 보여 토양 오염원의 영향을 일부 받은 군집으로 사료되었다.

군집 2는 Co (2.1 ng/m³)가 높았으며 object들의 구성비를 연도별로 보면, 90년(6, 7, 13)이 7.3%, 91년(25, 26, 27, 29, 30, 32, 35, 36)이 19.5%, 92년(37, 38, 39, 40, 41, 42, 43, 45, 46, 51, 52, 53, 54, 55, 56)이 36.6%, 93년(58, 59, 60, 61, 65, 66, 67, 70, 72)이 22.0%, 94년(75, 77, 78, 79, 80, 81)이 14.6%로 91년 이후 자료가 전체의 92.7%를 차지하고 있었다. 이 군집의 변수사이의 상관성은 뚜렷하지 않았다.

군집 3은 Cu (39.0 ng/m³), Mn (90.6 ng/m³), Pb (183.4 ng/m³), Zn (118.3 ng/m³)의 농도가 높았

Table 4. Correlation coefficients among 8 chemical variables within each homogeneous class for the fine and coarse particle groups.

Group	Class	Variable	Ca	Co	Cu	Fe	Mn	Pb	Si	Zn
Fine Group	Class 1	Ca	1.000							
		Co	-0.011	1.000						
		Cu	-0.384	0.090	1.000					
		Fe	-0.039	-0.168	-0.077	1.000				
		Mn	-0.207	0.299	0.380	-0.040	1.000			
		Pb	-0.362	-0.195	0.249	-0.172	0.186	1.000		
		Si	0.039	0.256	-0.019	0.254	-0.070	-0.377	1.000	
		Zn	-0.293	0.030	0.102	0.002	0.177	0.101	-0.142	1.000
	Class 2	Ca	1.000							
		Co	-0.205	1.000						
		Cu	-0.535	0.293	1.000					
		Fe	-0.791	0.389	0.307	1.000				
		Mn	-0.211	0.022	0.137	0.402	1.000			
		Pb	-0.430	0.218	0.506	0.123	-0.338	1.000		
		Si	-0.093	0.640	0.145	0.336	0.247	0.157	1.000	
		Zn	-0.102	0.386	0.159	0.153	0.278	0.440	0.682	1.000
	Class 3	Ca	1.000							
		Co	-0.727	1.000						
		Cu	-0.844	0.246	1.000					
		Fe	-0.973	0.867	0.696	1.000				
		Mn	-0.508	-0.222	0.891	0.294	1.000			
		Pb	-0.508	-0.222	0.890	0.293	0.999	1.000		
		Si	0.309	-0.878	0.248	-0.522	0.662	0.662	1.000	
		Zn	-0.699	0.999	0.207	0.846	-0.261	-0.262	-0.896	1.000
	Class 4	Ca	1.000							
		Co	-	1.000						
		Cu	0.325	-	1.000					
		Fe	0.167	-	0.260	1.000				
Mn		-	-	-	-	1.000				
Pb		0.187	-	-0.737	-0.355	-	1.000			
Si		0.722	-	0.065	0.194	-	0.032	1.000		
Zn		-0.448	-	-0.123	0.214	-	-0.008	-0.152	1.000	
Class 5	Ca	1.000								
	Co	0.925	1.000							
	Cu	0.706	0.919	1.000						
	Fe	0.011	0.078	0.021	1.000					
	Mn	0.698	0.909	0.957	0.309	1.000				
	Pb	0.498	0.668	0.808	-0.562	0.602	1.000			
	Si	-0.935	-0.897	-0.681	-0.364	-0.775	-0.280	1.000		
	Zn	0.950	0.915	0.772	-0.266	0.676	0.727	-0.794	1.000	

으며, 온도(-0.2°C), 상대습도(68.1%), 풍속(1.1 m/s), 강수량(4.0 mm)이 낮았다. 특히, 군집 3는 겨울철 기간동안의 objects로 (91년 1월 16일~91년 2월 17일)만 구성되어 있었다. 이 군집에서는 Cu와 Fe, Cu와 Mn, Fe와 Si, Fe와 Zn, Pb와 Si, Pb와 Zn, Si와 Zn 등 많은 변수사이에 강력한 상관성이 발견되어 연구대상지역 인근에 존재하는 특정 오염원의 영향을 집

중적으로 받은 군집으로 사료되었다.

군집 4는 Ca (1,463.9 ng/m³), Fe (1,296.9 ng/m³), Si (3,916.8 ng/m³)의 농도가 높았으며, 풍속(1.9 m/s)도 빨랐다. 또한, 군집 4는 대부분 봄철 기간동안의 object들(90년 1월 7일~1월 20일, 90년 3월 30일~4월 27일, 93년 3월 22일~4월 6일, 93년 4월 7일~4월 23일)로 구성되어 있었다. 이 군집은 특

Table 4. Continued.

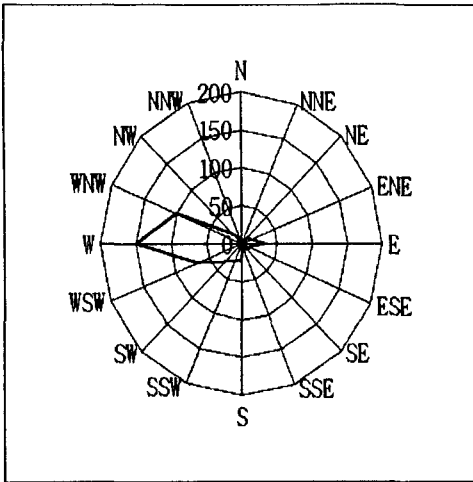
Group	Class	Variable	Ca	Co	Cu	Fe	Mn	Pb	Si	Zn
Coarse Group	Class 1	Ca	1.000							
		Co	-0.453	1.000						
		Cu	-0.414	-0.523	1.000					
		Fe	0.460	-0.531	0.363	1.000				
		Mn	-0.275	0.281	-0.276	-0.420	1.000			
		Pb	0.089	0.838	-0.855	-0.307	0.155	1.000		
		Si	0.281	-0.296	0.310	0.956	0.310	-0.140	1.000	
	Zn	-0.349	0.267	0.227	0.377	-0.397	0.163	0.559	1.000	
	Class 2	Ca	1.000							
		Co	0.274	1.000						
		Cu	-0.142	0.007	1.000					
		Fe	0.276	0.170	-0.071	1.000				
		Mn	-0.004	0.435	0.330	-0.091	1.000			
		Pb	0.145	0.399	0.129	0.063	0.366	1.000		
		Si	0.545	0.223	-0.017	0.616	-0.004	0.262	1.000	
	Zn	-0.049	-0.005	0.308	0.262	-0.059	0.080	0.102	1.000	
	Class 3	Ca	1.000							
		Co	-0.817	1.000						
		Cu	-0.426	-0.174	1.000					
		Fe	0.018	-0.591	0.897	1.000				
		Mn	-0.615	0.047	0.976	0.778	1.000			
		Pb	0.555	-0.933	0.516	0.842	0.315	1.000		
		Si	0.312	-0.803	0.726	0.955	0.558	0.964	1.000	
	Zn	0.298	-0.794	0.737	0.960	0.570	0.960	0.999	1.000	
	Class 4	Ca	1.000							
		Co	0.184	1.000						
		Cu	-0.056	-0.971	1.000					
		Fe	0.991	0.046	0.080	1.000				
Mn		-0.417	-0.970	0.909	-0.288	1.000				
Pb		0.437	-0.298	0.187	0.484	0.179	1.000			
Si		0.537	-0.731	0.796	0.648	0.543	0.553	1.000		
Zn	-0.173	-0.915	0.970	-0.047	0.883	-0.057	0.668	1.000		
Class 5	Ca	1.000								
	Co	-0.951	1.000							
	Cu	-0.624	0.698	1.000						
	Fe	0.262	-0.041	0.081	1.000					
	Mn	-	-	-	-	1.000				
	Pb	0.396	-0.501	-0.255	0.001	-	1.000			
	Si	0.521	-0.491	-0.164	0.658	-	0.646	1.000		
Zn	0.673	-0.496	-0.204	0.665	-	0.274	0.496	1.000		

히 Fe와 Ca, Cu와 Mn, Cu와 Zn, Mn과 Zn 사이에 강한 상관성이 조사되었다.

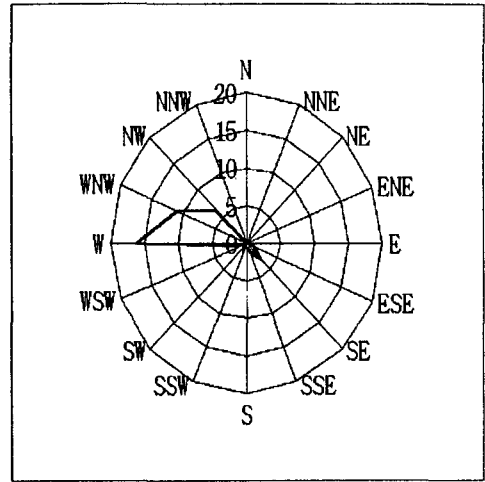
군집 5에서 Co, Cu, Mn의 농도는 검출한계 이하였으며, Pb의 농도는 15.2ng/m³로 낮았다. 연도별로 보면, 89년(1, 2, 3, 4)이 50%로 가장 많았고, 90년(8, 9, 12)이 37.5%, 그리고 91년(28)이 12.5%로 91년 이전 자료가 군집 전체자료 중 87.5%를 차지해, 91년 이후 자료가 92.7%를 차지하는 군집 2와는 상반

되는 면을 볼 수 있었다. 이 군집에서 Ca와 Co 사이에 강한 음의 상관성이 조사되었으나 해석할 수는 없었다.

이 중 계절 및 기상요인과 관련이 있는 군집에 대해 풍향자료(기상청, 1989~1994)를 결합하여 비교해 보았다. 미세입자 그룹에서 풍속이 높은 수치를 기록한 군집 1(그림 4a)과 겨울철 자료로 구성된 군집 3(그림 4b)의 경우, 서풍의 빈도가 가장 높았다. 거대입자에서 겨울철 자료로 구성된 군집 3(그림 5a)과 봄철 자료로

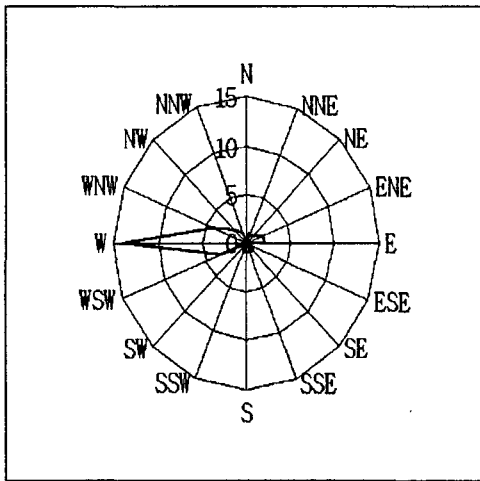


(a) class 1

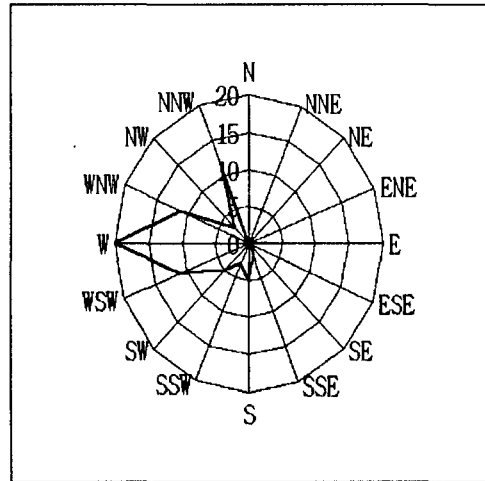


(b) class 3

Fig. 4. Wind direction for classes 1 and 3 in case of the fine particle group.



(a) class 3



(b) class 4

Fig. 5. Wind direction for classes 3 and 4 in case of the coarse particle group.

구성된 군집 4(그림 5b)의 경우, 역시 서풍의 빈도가 가장 높았다. 반면, 여름철 자료로 구성된 미세입자의 군집 5와 거대입자의 군집 1의 경우에는 풍향에 의한 뚜렷한 특징은 확인되지 않았다. 연구대상지역의 서쪽에는 각종 공장들이 위치하고 있고, 또한 크고 작은 건축 및 토목공사가 빈번히 수행된 바 있으며, 현재 영통 주택단지 개발이 진행되고 있다. 따라서 이들 오염원에서 배출된 각종 분진의 영향을 많이 받은 것으로 사료된다.

5. 결 론

경희대학교 수원캠퍼스에서 포집된 입경별 분진자료를 이용하여, 미세 및 거대분진의 오염 특성 및 형태분류를 시도하였다. 이를 위해 군집분석법과 주성분분석을 동시에 응용한 형태인식기법을 적용하였으며, 이 결과 미세입자와 거대입자에 대하여 각각 6개의 pattern을 분류할 수 있었다. 분류된 군집은 90% 확률로 검증하여

outlier를 제거한 후, 통계적으로 양호한 pattern을 도출하였다. 또한, 통계분석 과정에서 다양한 자료변환과 비유사도 기준을 적용하여 자료의 해석에 가장 유용한 방법을 결정하였다. 그 결과, 수원인근의 연구대상지역 대기 중 분진의 입경별 자료를 분류할 경우, 자료의 전 처리로는 로그 변환을, 군집분석시에는 유클리디안 거리를 비유사도 기준으로 이용할 때 가장 이상적으로 분류할 수 있었다.

미세분진과 거대분진 각 경우, 각 군집의 pattern은 비교적 뚜렷하게 분류될 수 있었고 물리적 해석이 가능하였다. 각 군집들은 계절적인 특성을 가지고 있었다. 미세입자의 군집 3과 5는 각각 겨울과 여름철 자료로 구성되어 있었고, 거대입자의 군집 1, 3, 4는 각각 여름, 겨울, 봄철 자료로 구성되어 있었다. 따라서, 수원지역의 오염특성은 계절적인 영향이 극심함을 알 수 있었다. 또한 미세입자와 거대입자 모두 겨울철 자료로 구성된 군집은 서풍의 빈도가 압도적으로 높았으며, 봄철 자료로 구성된 거대입자의 군집 4와 미세입자의 군집 1는 서풍, 서북서풍, 서남서풍의 빈도가 높았다.

감 사

본 연구의 일부는 삼성전자(주)의 SAMP (Suwon Aerosol Monitoring Program)의 일환으로 수행되었으며, 이에 도움을 주신 임직원께 감사드린다.

참 고 문 헌

- 기상청 (1989~1994) 기상월보.
- 김기영, 전명식 (1994) 다변량 통계자료분석, 자유아카데미, 129-152, 253-274pp.
- 김동술, 김형석 (1990) Pattern Recognition을 이용한 지하상가에서의 대기오염물질의 농도 분석에 관한 연구, 한국대기보전학회, 6(1), 1-10.
- 김동술, 이태정 (1993) 입경분류에 입각하여 목표변환 인자분석법을 이용한 수원지역 분진오염원의 정량적 기여도 추정, 한국대기보전학회, 9(1), 44-50.
- 백성옥, 황승만 (1992) 주성분/중회귀분석을 이용한 대구지역 대기 중 부유분진의 발생원별 특성 평가, 한국대기보전학회, 8(3), 179-190.
- 신훈중 (1995) 대기부유분진 중 입경에 의한 미량 금속 원소의 농도분포에 관한 연구, 경희대학교 대학원 환경학과 석사학위논문.
- 이태정 (1992) X-선 형광분광법을 이용한 대기 부유분진 중 중금속의 농도 분포에 관한 연구, 경희대학교 대학원 환경학과 석사학위논문.
- Albano, C., G. Blomqvist, D. Cooman, W.J. Dunn III, U. Edlund, B. Eliasson, S. Hellberg, E. Johansson, B. Nordén, D. Johnels, 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.
- Chow, J.C. (1995) Measurement methods to determine compliance with ambient air quality standards for suspended particles, *A&WMA*, 45(5), 320-382.
- Cohen, B.S. (1986) *Cascade Impacter, Sampling and Data Analysis*, edited by J.P. Lodge, Jr. and T.L. Chan, American Industrial Hygiene Association, 1-21pp.
- 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.
- Duewer, D.L., B.R. Kowalski, and T.F. Schatzki (1975) Source identification of oil spills by pattern recognition analysis of natural elemental composition, *Analytical Chemistry*, 47(9), 1573-1583.
- Dunn III, W.J., S.L. Emery, and W.G. Glen (1989) Preprocessing, variable selection, and classification rules in the application of SIMCA pattern recognition to mass spectral data, *Environ. Sci. & Technol.*, 23(12), 1499-1505.
- Hopke, P.K. (1985) *Receptor Modeling in Environmental Chemistry*, John Wiley & Sons, New York.
- Hopke, P.K. (1991) *Receptor Modeling for Air Quality Management*, Elsevier Science Publishing Company Inc., New York.
- Kim, D.S. (1987) *Particle Class Balance for Apportioning Aerosol Mass*, Ph.D. Thesis, University of Illinois at Urbana-Champaign.

- ampaign.
- Kim, D.S., P.K. Hopke, D.L. Massart, L. Kaufman, and G. Casuccio (1985) Multivariate analysis of CCSEM data, JAPCA, 85-21.6., Presented at the 78th Annual Meeting of APCA, Detroit, Michigan.
- Massart, D.L. and L. Kaufman (1983) *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, John Wiley & Sons, New York.
- Scott, D.R. (1986) Determination of chemical classes from mass spectra of toxic organic compounds by SIMCA pattern recognition and information theory, *Analytical Chemistry*, 58(4), 891-890.
- SIMCA-3B manual (1984) A pattern recognition program for CPM and MS-DOS based microcomputers, principal components models, 2505 Shepard Blvd, Columbia, MO.
- Simenov, V., R. Tsitouridou, C. Samara, M. Karadjov, and A. Andreev (1994) Chemometrical approach to study rain composition from the region of Thessaloniki, Greece, *Toxicol. Environ. Chem.*, 46, 97-106.
- Sjögren, M., H. Li, U. Ranning, and R. Westerholm (1996) Multivariate analysis of exhaust emissions from heavy-duty diesel fuels, *Environ. Sci. & Technol.*, 30(1), 38-49.
- Stratis, J.A., R.D. Tsitouridou, and V.D. Simenonov (1995) Chemometrical classification of aerosol analytical data, *Toxicol. Environ. Chem.*, 47, 191-196.
- Whitby, K.T., R.B. Husar; and B.Y.H. Liu (1972) The aerosol size distribution of Los Angeles smog, *J. Col. & Inter. Sci.*, 39, 203.
- Wold, S., (1976) Pattern recognition by means of disjoint principal components models, *Pattern Recognition*, 8, 127-139.
- Wold, S., C. Albano, W.J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hollberg, E. Johansson, W. Lindberg and M. Sjöström (1984) *Multivariate Data Analysis in Chemistry, Chemometrics: Mathematics and Statistics in Chemistry*, edited by B.R. Kowalski, NATO ASI series, 138, 17-95pp.
- Wold, S. and M. Sjöström (1977) SIMCA: a method for analyzing chemical data in terms of similarity and analogy, in *Chemometrics: Theory and Application*, edited by B.R. Kowalski, ACS symposium series 52, 243-282pp.