韓國大氣保全學會誌 第 9 卷 第 4 號
J. KAPRA Vol. 9, No. 4, (1993) pp. 278-287

다변량해석을 이용한 서울시 하계 스모그의 형태 분류

Multivariate Analysis for Classification of Smog Type during the Summer Season in Seoul, Korea

홍낙기 · 이종범 · 김용국
*강원도 보건환경연구원
강원대학교 환경학과
(원고접수 : 1993. 9. 24)

Nak-Gi Hong*, Chong-Bum Lee and Yong-Goog Kim
*Kangwon-Do Institute of Health and Environment
Dept. of Environmental Science, Kangwon National University
(Received 24 September 1993)

Abstract

In order to classify smog type during the summer season in Seoul, air quality and meteorological data were analyzed by multivariate analysis. Among 15 variables relating to visibility, 10 variables were selected by multiple regression analysis for clustering of smog types; total suspended particle, sulfur dioxide, ozone, nitrogen dioxide, total hydrocarbon, south-north wind component, relative humidity, precipitable water, mixing height and air temperature. Smog types were grouped into three clusters using cubic clustering criterion and the numbers of days in each cluster were contained 74, 28 and 16 days. Each cluster was seperated clearly by sulfur dioxide, precipitable water and air temperature. The first cluster was representative of high ozone concentration and prevailing meteorological conditions for ozone formation. Therefore, visibility in the first cluster was considered to be affected by photochemical smog. The third cluster showed characteristics of sulphurous smog type due to the higher concentration of primary pollutant, based on the dry condition than that in another cluster. On the other hand, the characteristic of the second cluster was not relatively clear, but considered to be in an intermediate characteristic between photochemical smog and sulphurous smog type.

1. 서 론

스모그는 매연, 분진과 아황산가스 등의 1차 오염물질에 의한 매연성스모그와 주로 자동차에서 배출되는 질소산화물과 비메탄계 탄화수소가 태양의 자외선에 의하여 복잡한 광화학반응을 일으켜 생성되는 옥시단트에 의한 광화학스모그(Oke, 1978)로 크게 구분된다. 스모그에 기인되어 나타나는 대표적인 현상은 시정(visibility)악화로서, 시정은 수평하늘에 대해서 어떤 사물을 식별할 수 있는 최대 가시거리를 말하는 것으로(William and Lewis, 1980) 가스상, 입자상 물질에 의한 가시광선의 흡수나 산란에 의하여 감소하며 대기오염지표 중의 하나로 사용되어 왔다.

시정과 관련된 인자들에 대한 연구에서 Leaderer et al.(1979)과 Trijonis(1979) 등은 이산화질소, 아황산가스, 부유분진, 상대습도 등과 관련된 감쇄계수와의 선형적인 관계식을 제시하였다. Barone et al.(1978)은 다변량 통계분석을 이용하여, 각 지역마다 시정을 감소시키는 어떤 특정변수가 있으며, 황산염을 포함한 분진의 구성원소와 입자의 크기는 시정과 밀접한 관계가 있음을 밝혔다. 또한 Philip et al.(1976)과 Philip(1980)은 분진의 발생원을 규명하기 위하여 분진 구성원소의 자료를 이용하여 인자분석을 행하고, 계보적 군집분석에 의하여 각 인자를 해석하였으며, William과 Clark(1981)은 인자분석을 이용하여 미국 동부지역의 오존농도 패턴을 분석한 바 있다. 이와같이 인자분석 등의 다변량 해석은 분진의 발생원을 규명하거나 오염농도의 패턴을 파악하는데 폭넓게 이용되어 왔다.

우리나라의 경우 일반적으로 겨울철에는 난방에 의한 연료의 사용이 많고 부유분진의 농도가 높아 매연성스모그의 발생가능성이 크며, 여름철에는 난방연료의 사용이 적은 반면 차량 등에 의하여 배출된 오염물질이 주 요인이 되어 이에 의한 광화학스모그의 발생 가능성이 증가하고 있는 추세이다(정용승 등, 1992).

하계의 시정악화는 매연성스모그 또는 연무에 의한 영향뿐만 아니라 광화학스모그 등 여러 요인에 의해서도 복합적으로 나타날 수 있다. 따라서 하계의 시정악화를 초래하는 스모그를 주요인에 따라 분류할 수 있는 방법을 정립한다면 광화학스모그의 발생빈도를 파악하는데 유용할 뿐만 아니라 대기오염 관리와 대책면에서도 매우 유용할 것으로 사료된다.

본 연구에서는 서울시에서 하계동안에 발생하는 스모그의 형태를 분류하기 위하여 시정이 특히 악화된 날들을 선택하여 중회귀분석 및 군집분석을 통하여 특성있는 몇 개의 군집으로 구분하고, 주성분분석을 통하여 각 군집의 특성과 시정악화 요인을 파악하였다.

2. 대상기간 및 자료

본 연구에서는 하계에 해당되는 1989년부터 1991년까지의 4월부터 9월까지(이하 하계라 함) 18개월을 대상기간으로 하였다. 분석에 사용된 대기질자료는 서울시 전역에 고루 분포되어 있는 20개의 대기오염 자동측정망에서 매시간 측정된 오존(O_3), 일산화탄소(CO), 이산화질소(NO_2), 아황산가스(SO_2), 총탄화수소(Total HydroCarbon: THC)와 총부유분진(Total Suspended Particle : TSP) 등의 6개 항목을 이용하였으며, 이들 자료 중에서 각 월별로 70% 이상 비교적 충실하게 측정된 지역들의 농도자료를 매시간 평균하여 사용하였다.

지상기상자료로는 서울기상대에서 매시간 측정한 풍속, 풍향, 강수량 및 일사량 자료와 3시간 간격으로 측정한 시정, 기온, 상대습도 및 운량 자료를 이용하였다. 이들 중 3 시간 간격으로 측정된 기상자료는 단순 내삽하여 매시간 자료로 환산하였으며, 풍향은 산술적인 평균이 불가능하므로 풍향 및 풍속자료를 이용하여 동서성분 및 남북성분의 풍속으로 환산한 후 분석하였다. 한편 혼합층고도 및 가강수량은 오산에서 측정한 상층기상자료를 이용하여 계산하였다.

가강수량(precipitable water; Wp)은 단위면적위의 수직기주내에 포함되어 있는 수증기가 모두 응결하여 강수로 되었다고 가정한 경우의 강수량을 말한다. 즉, 이 양은 어느 지역위에 있는 수증기 총량을 나타내는 것으로 다음의 식 (1)과 같은 Solot (1939)의 근사식을 이용하였다.

$$Wp = \Sigma\, 0.01 \cdot \bar{q} \cdot \triangle p \qquad (1)$$

여기서 $\bar{q}$는 두 등압면 사이의 평균 비습(g/kg)이고, $\triangle p$는 두 등압면사이의 기압차(mb)로서, 본 연구에서는 오전 9시에 측정된 상층기상자료중 기압이 850mb가 되는 고도까지 각 층별로 기온과 이슬점 온도를 이용하여 비습을 구하고, 이로부터 오전 9시의 가강수량을 계산하였다. 한편 혼합층고도는 오산의 상층기상자료와 서울의 지상기상자료를 이용하여 혼합층고도 모델(이종범, 1991)로 계산하였으며 매일 16시의 혼합층고도를 일최대 혼합층고도(MMD: maximum mixing depth)로 간주하여 분석에 이용하였다.

3. 연구방법

3.1 자료의 선정

기상자료 및 대기질자료는 시간에 따른 변화가 특히 크므로 통계분석을 위한 자료의 선정을 위하여서는 무엇보다도 먼저 시간에 따른 변화특징을 충분히 고려해야 한다. 본 연구에서는 위의 대상기간 동안 전체 자료에 대하여 평균 일변화 특징(그림 1)을 분석하고 이를 기초로하여 각 변수별로 시간범위를 선정하였다. 각 변수들의 적용시간대와 산출방법 등은 표 1에 제시하였다.

O_3농도는 일중 최고를 나타내는 14시에서 16시까지의 평균치를 이용하였으며, TSP를 포함한 SO_2, NO_2 등 1차 오염물질들은 대체로 오전중에 농도가

높으나 주간에 나타나는 시정악화를 고려하기 위하여 14시에서 16시까지의 평균농도로 하였다. 또한 기상자료중 기온, 상대습도는 14시에서 16시까지의 평균값을, 동서 및 남북방향의 바람성분과 운량은 12시에서 16시까지의 평균값을 사용하였고 일사량은 일출 후부터 16시까지의 적산량으로 하였다.

한편, 연구 목적중의 하나는 스모그에 의한 시정악화의 특징을 파악하고 형태를 분류하는 것이므로 가급적 시정이 현저히 악화된 경우만을 대상으로 하는 것이 바람직 하겠으나 통계처리를 위하여서는 적정량의 자료가 필요하므로 본 연구에서는 전체 대상일 중 14시부터 16시까지의 평균 시정이 12km 이하인 날을 분석대상으로 하였다. 또한 대상일 중 강수현상이 있는 날은 오염물질의 농도가 낮으므로 스모그의 발생 가능성이 적다고 사료되어, 5시부터 16시까지의 총강수량이 0.1mm 이상인 날을 대상에서 제외하였으며, TSP의 농도가 300$\mu g/m^3$ 이상으로 특히 높은 날은 황사에 의한 영향이 있을 수 있으므로 대상에서 제외하였고, 단순한 연무에 의하여 시정이 악화된 경우를 제외하기 위하여 운량이 80 % 이상인 날을 제외하였다. 하계동안의 총 549일 중 위의 조건이 고려되어 선정된 날은 89년에 31일, 90년에 38일, 91년에 49일로서 모두 118일이었으며 본 연구에서는 선정된 118일의 자료를 이용하여 분석하였다.

3.2 중회귀분석 및 군집분석

위의 자료선정 방법을 통하여 선정된 날을 대상으로 시정을 악화시키는 요인이 비슷한 날들로 구분하기 위하여 군집분석을 행하였다. 그러나 군집분석을 행하는데 있어서 변수의 수가 많은 경우에는 해석하기가 곤란하므로 자료를 충분히 설명할 수 있는 몇개의 변수만으로 군집분석을 행할 필요가 있다. 이를 위하여 중회귀분석을 수행하였으며 이때 변수선정 방법으로는 모든 가능한 회귀방법(all possible regressions)을 이용하였다. 이 방법에서 변수

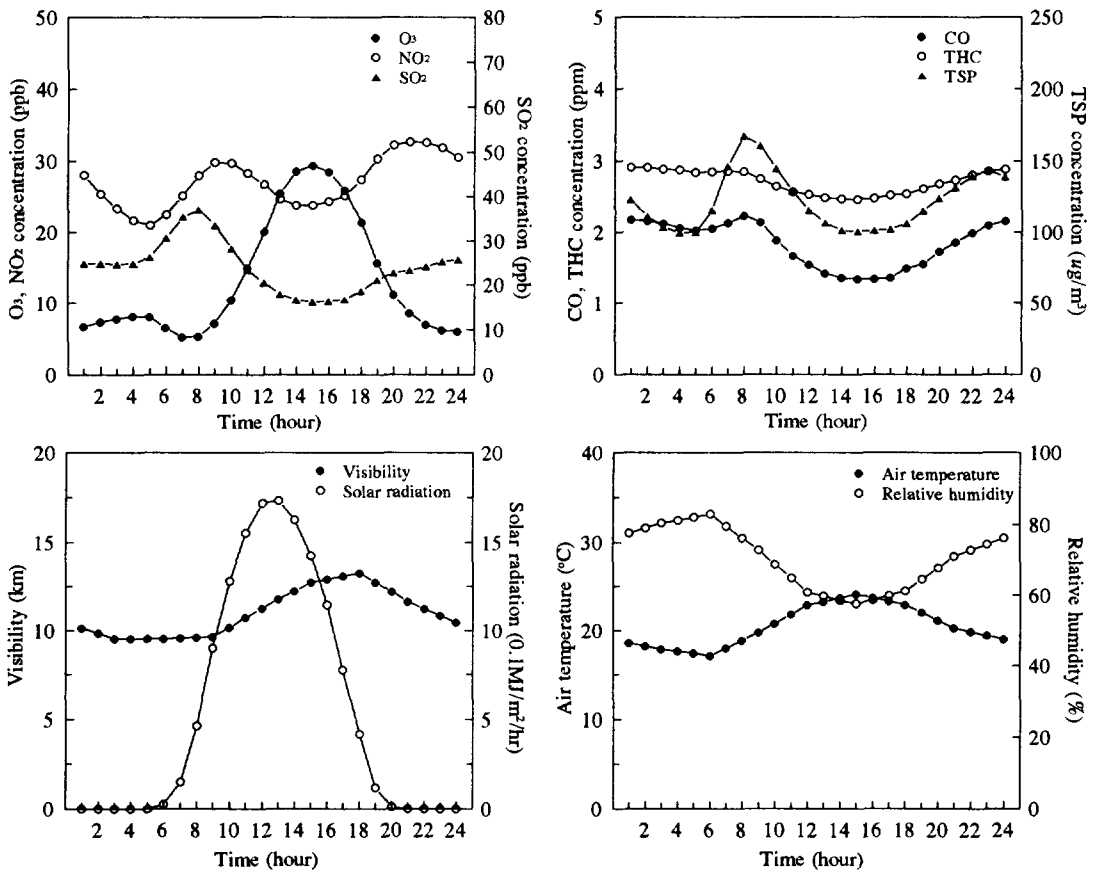

Fig. 1. Diurnal variations of air pollutants and meteorological parameters in Seoul from April to September in 1989-1991.

의 선정은 결정계수(R^2)의 결점을 보완한 수정결정계수($\overline{R}^2$) 값이 가장 큰 축소모형을 최적의 회귀모형으로 판정하였으며, 수정결정계수의 값이 같을 경우에는 Mallows Cp값이 작은 모형을 선정하였다(허명회, 1991). 수정결정계수는 다음과 같이 구하며 여기에서 n은 자료의 수, p는 선택된 변수의 수를 말한다.

$$\overline{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2) \qquad (2)$$

중회귀분석의 변수선정 과정에서 시정악화에 큰 영향을 미칠 것으로 판단되는 상대습도와 가강수량, 매연성스모그의 주원인물질인 TSP, SO_2 및 광화학반응의 생성물인 O_3는 항상 포함되어 수행되도록 하였다.

이와같은 방법에 의하여 선정된 변수를 이용하여 군집의 수를 결정한 후 군집분석을 행하였으며, 군집의 수를 결정하는 방법으로는 CCC(Cubic Clustering Criterion) 판정기준을 이용하였다. 군집분석은 대상일이 여러 군집중 어느 하나에만 속하게 하기 위하여 유클리드거리를 기초로 반복적 알고리즘을 통해 상호배반적(disjoint)인 군집을 형성시키는 최적분리방법을 이용하였다. 이때 변수들의 측정단위가 서로 다를 경우에는 변수들간의 직접비교가 곤란하므로 각 변수의 평균을 0, 분산을 1로 표준화하였다. 또한 군집의 seed값을 다시 계산하기 위하여 필요한 최대 반복회수는 20회로 하였으며, seed값의 수렴기준은 0.002로 하여 seed값 변화의 최대거리가 초기 seed값들 사이의 최소거리에 0.002를 곱한 값 이하가 되면 반복이 중단되도록 하였다.

Table 1. Variable descriptions and selected times.

Variables	Time	Method
VIS ; visibility(≤12km)	14–16	Mean
SR ; solar radiation	05-16	Sum
Ta ; air temperature	14–16	Mean
CL ; cloud amount(≤80%)	12-16	Mean
U_x ; east-west wind component	12-16	Mean
U_y ; south-north wind component	12-16	Mean
RH ; relative humidity	14-16	Mean
Wp ; precipitable water	09	-
MH ; mixing height	16	-
O_3 ; ozone	14-16	Mean
SO_2 ; sulfur dioxide	14-16	Mean
NO_2 ; nitrogen dioxide	14-16	Mean
CO ; carbon monoxide	14-16	Mean
THC ; total hydrocarbon	14-16	Mean
TSP ; total suspended particle ($\leq 300\mu g/m^3$)	14-16	Mean

3.3 주성분분석

본 연구에서는 군집수의 결정과정을 거쳐 최적분리 군집방법에 의하여 구분된 군집별로 표 1에 제시한 변수를 대상으로 주성분분석을 행하였다. 주성분분석은 각각의 분산이 1이 되도록 모든 변수를 표준화하여 이로부터 얻어진 상관행렬에 기초하여 유도되었으며, 표준화된 고유벡터를 구한 다음 이를

Table 2. Adjusted coefficient of determination($\overline{R}^2$) and Mallows Cp using all possible regressions for dependent variable, visibility(n=118). The first 5 variables in each model are included variables.

No. of Variables	R^2	C_p	$\overline{R}^2$	Variables in model										
.....														
8	0.544	12.498	0.511	RH	Wp	TSP	SO_2	O_3	U_x	U_y	NO_2			
.....														
9	0.569	8.436	0.533	RH	Wp	TSP	SO_2	O_3	MH	U_y	Ta	NO_2		
9	0.568	8.685	0.532	RH	Wp	TSP	SO_2	O_3	MH	U_y	NO_2	THC		
.....														
10	0.577	8.388	0.537	RH	Wp	TSP	SO_2	O_3	MH	U_y	Ta	NO_2	THC	***
10	0.573	9.220	0.534	RH	Wp	TSP	SO_2	O_3	MH	U_y	Ta	NO_2	CL	
.....														
11	0.579	9.758	0.537	RH	WP	TSP	SO_2	O_3	MH	U_y	Ta	NO_2	THC	CL
.....														

*** indicates selected regression model.

각 주성분에 대해 척도화된 고유벡터를 구하여 사용하였다.

본 연구에서 모든 통계처리는 SAS(Statistical Analysis System)를 이용하였다.

4. 결과 및 고찰

4.1 중회귀분석

자료선정 방법을 통해 선택된 118일에 대한 군집분석을 위하여 수행된 중회귀분석의 변수선정 결과는 표 2에 제시하였다. 최적의 회귀모형을 선정하는데 있어서 수정결정계수의 값이 가장 큰 회귀모형은 항상 포함되어 수행되도록 한 상대습도, 가강수량, TSP, O_3, SO_2 등 5개의 변수외에 혼합고, 남북바람성분, 기온, NO_2 및 THC 등 5개의 변수가 추가된 경우와, 여기에 운량이 추가된 2가지 경우였으나, 운량이 추가되기 전의 회귀모형에서 Mallows Cp값이 더 작았으므로 이를 최적의 모형으로 선정하였다. 군집분석은 이와같이 선정된 10개의 변수로 수행하였다.

4.2 최적분리를 이용한 군집분석

군집의 수를 결정하는 문제는 분석하고자 하는 자료의 특성에 따라 여러 방법이 사용되나, 최적분리방법과 같이 어떤 판정기준의 최적화를 따르는 군집방법은 판정기준의 값이 급격히 증가된 곳에서 대응되는 수를 결정하는 방법이 통상적으로 많이 사용된다(김기영과 전명식, 1991). 본 연구에서는

Table 3. Determination of number of clusters based on the cubic clustering criterion.

Number of Clusters	R-squared	Cubic clustering criterion	Pseudo-F
12	0.861	-0.872	59.6
11	0.846	-1.587	58.6
10	0.830	-2.119	58.7
9	0.815	-2.445	59.9
8	0.799	-2.571	62.5
7	0.779	-2.790	65.3
6	0.754	-3.112	68.5
5	0.722	-3.330	73.5
4	0.687	-2.600	83.3
3	0.647	-1.659	105.5
2	0.466	-3.139	101.4
1	0.000	0.000	—

Table 4. Result of cluster analysis.

Cluster	Number of days	RMS STD*	Maximum Distance**	Nearest cluster	Centroid distance
1	74	0.77	3.69	2	3.02
2	28	0.80	5.45	1	3.02
3	16	0.91	4.68	2	3.72

* indicates root mean square standard deviation.

** indicates maximum distance from seed to observation.

CCC 판정기준을 이용하여 군집의 수를 결정하였는바 결과는 표 3에 제시하였다. 표 3의 결과에 의하면 군집의 수가 3개인 곳에서 CCC 값은 국부적 최고(local peak)를 나타냈으며, F-test의 결과 역시 이 지점에서 최고치를 나타냈다. 또한 결정계수의 경우에 있어서도 3개 이상의 군집에서 단조증가하는 경향을 나타냈으므로 군집의 수는 3개가 가장 적당한 것으로 판단된다.

군집의 수를 3개로 정하여 최적분리방법에 의해 군집분석을 실행한 결과 표 4와 같이 각 군집별로 각각 74일, 28일 그리고 16일이 할당되었다. 3개의 군집중 군집 1과 군집 3은 군집 2와 인접해 있으며, 군집내 거리평방근은 군집 1에서 가장 작았다.

사용된 각 변수들의 통계량은 표 5에 제시하였다. 전체의 군집으로부터 변수를 예측하는 정도를 나타내는 결정계수 및 군집간 분산에 대한 군집내 분산의 비율을 나타내는 분산비(variance ratio, $R^2/(1-R^2)$)는 SO_2가 가장 큰 값을 보였으며, 기온 및 가강수량 또한 큰 값을 나타내 각 군집은 이들 세 변수에 의하여 잘 구분됨을 나타냈다.

한편 군집분석을 위하여 사용된 변수들 중 SO_2, O_3, 가강수량, 기온 등 4개의 변수에 대한 군집별

Table 5. Simple statistics of variables by clustering procedures.

Variable	Total STD	Within STD	R^2	$R^2/(1-R^2)$
Ta	1.00	0.63	0.61	1.56
U_v	1.00	0.88	0.24	0.32
RH	1.00	0.77	0.42	0.70
Wp	1.00	0.63	0.61	1.57
MH	1.00	0.91	0.18	0.23
O_3	1.00	0.88	0.24	0.32
SO_2	1.00	0.60	0.65	1.87
NO_2	1.00	0.77	0.42	0.71
THC	1.00	0.96	0.10	0.11
TSP	1.00	0.85	0.30	0.41

Table 6. Means and standard deviations(STD) of variables for three clusters.

Variables	Cluster 1		Cluster 2		Cluster 3	
	Mean	STD	Mean	STD	Mean	STD
VIS(km)	8.20	2.54	9.14	1.77	6.87	1.65
SR(W/m^2)	334.52	77.34	389.42	63.96	368.83	55.18
Ta(℃)	28.53	2.40	21.28	3.49	21.44	3.34
CL(Tenth)	5.12	2.05	3.76	2.69	2.54	2.91
U_x(m/sec)	−2.00	1.39	−3.05	1.32	−2.25	1.18
U_y(m/sec)	−0.29	0.81	−1.39	1.04	−0.38	0.29
RH(%)	56.56	8.99	47.23	10.04	34.21	10.47
Wp(mm)	16.87	4.09	8.70	2.27	6.51	2.34
MH(km)	0.83	0.20	1.02	0.27	0.71	0.10
O_3(ppb)	50.48	14.74	35.35	9.48	36.33	7.73
SO_2(ppb)	15.14	5.25	21.81	6.68	43.50	12.81
NO_2(ppb)	25.53	6.49	26.80	6.85	44.44	12.61
CO(ppm)	1.24	0.28	1.40	0.33	1.93	0.42
THC(ppm)	2.66	0.39	2.40	0.25	2.85	0.78
TSP($\mu g/m^3$)	119.38	31.27	125.61	33.57	177.90	23.02

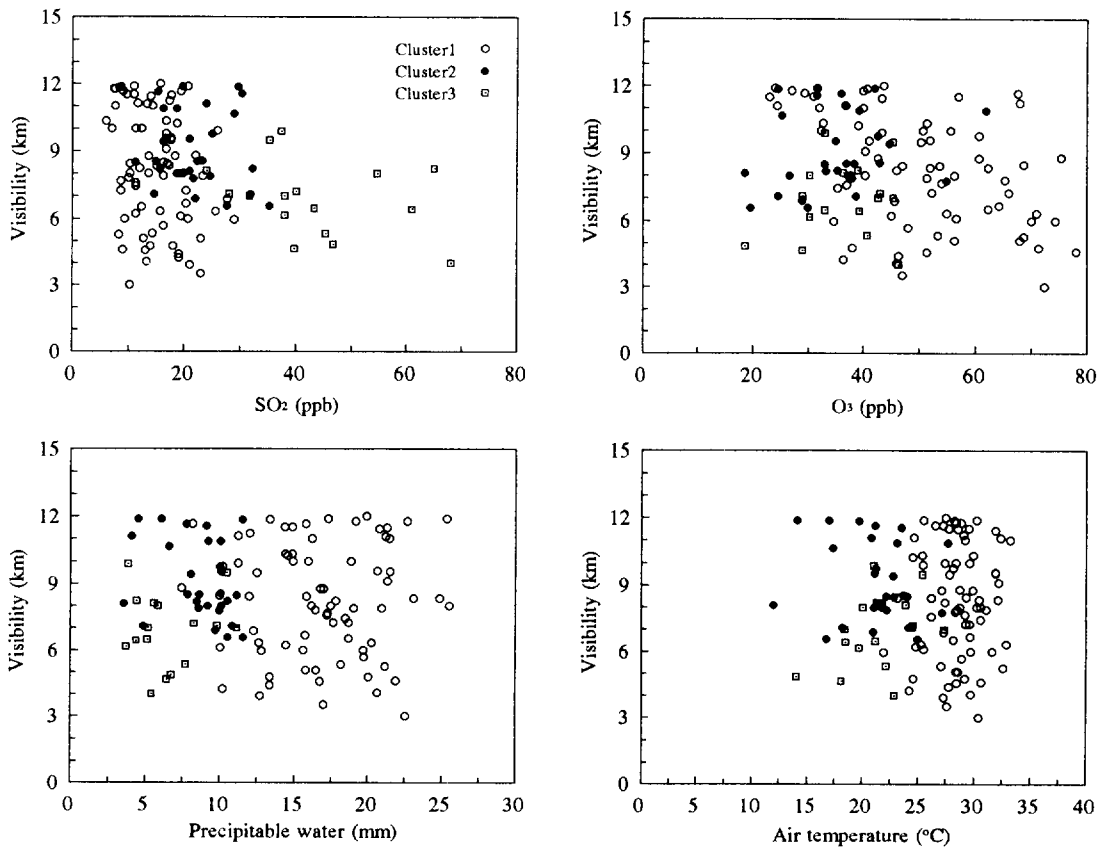

Fig. 2. Scatter diagrams of visibility vs. four variables.

산포도를 시정과 관련하여 그림 2에 제시하였다. SO_2의 경우 군집 3은 다른 군집과 비교하여 뚜렷하게 구분되었으나, O_3 및 가강수량, 기온의 경우에 군집 2와 3은 비슷한 값으로 분포한 반면 군집 1의 구분이 뚜렷한 특징을 나타냈다. 시정은 군집 2에서 비교적 높은 값을 나타냈으나 군집 3에서는 작은 값으로 분포하는 특징을 나타냈다.

각 군집별로 계산한 변수들의 평균과 표준편차는 표 6에 나타냈으며 시정, 가강수량, 혼합층고도, 기온, O_3, SO_2, NO_2, TSP 등 8개의 변수에 대한 군집별 중앙값과 25, 75 percentile 값을 그림 3에 나타내었다. 각 군집의 평균시정은 6.9-9.1km로 비슷한 값을 보이나 시정에 영향을 미치는 주요 인자는 다르게 나타났다.

군집 1은 기온과 상대습도가 높고 가강수량이 많은 기상상태와 함께 TSP 및 1차 오염물질의 농도는 낮으나 O_3의 농도가 현저히 높은 특징을 나타냈다. 따라서 군집 1은 광화학스모그에 의한 시정악화가 고려되는 군집으로 판단된다. 군집 1에서 기온은 높으나 일사량 및 혼합층고도가 다소 낮은 이유는 이미 형성된 스모그에 의해 일사량이 감쇄되고 따라서 혼합층고도 또한 낮아진 것으로 사료된다. 군집 3은 시정이 평균 6.9km로 가장 악시정을 보인 군집으로 대부분의 변수가 군집 1과는 반대의 경향을 나타냈다. 즉, 대기중 가강수량이 가장 작고, 상대습도 또한 34%로 아주 건조하며, 기온은 낮은 상태를 보이고 있다. 그러나 O_3를 제외한 TSP 등 모든 1차 오염물질농도는 가장 높은 분포를 보이고 있어 군집 3은 매연성스모그의 경향이 짙은 날들로 판단된다. 군집 2의 특징을 보면 기상요소들의 경우에는 군집 1 및 군집 3과 비교하여 뚜렷한 특징을 나타내지 않았으며, O_3는 군집 3에 가까운 농도를 보였으나 다른 오염물질은 군집 1에 가까운 농도를 나타냈다. 따라서 군집 2는 위의 군집 1 또는 군집 3의 중간적인 특징을 나타내고 있는 것으로 사료된다.

전체 118일에 대한 월별 각 군집의 발생빈도를 표 7에 제시하였다. 군집 1의 경우에는 대부분 6월부터 8월에 걸쳐 발생되었으며, 군집 2와 3은 4월과 5월에 주로 발생된 것으로 나타나 각 군집은 계절적인 특성이 다소 반영된 것으로 분석되었다. 한

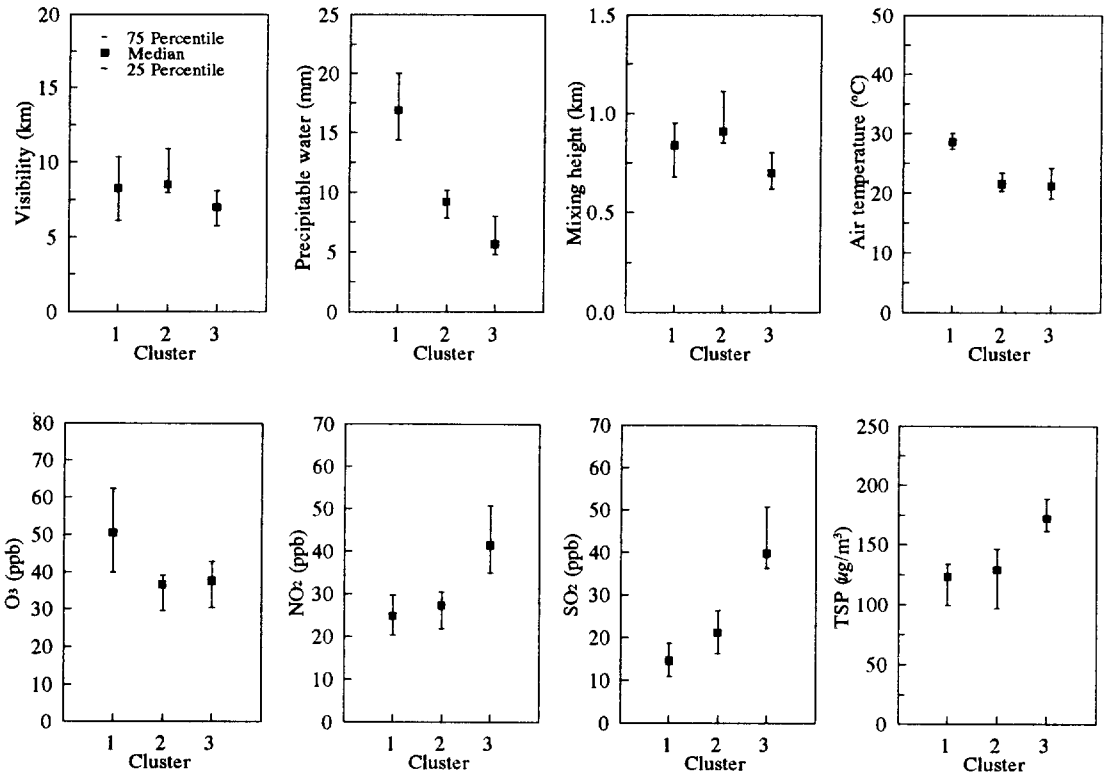

Fig. 3. Median, 25 and 75 percentile values for each clusters.

Table 7. The monthly frequencies of smog days for each cluster.

Cluster	Year	Month						Total
		Apr.	May	Jun.	Jul.	Aug.	Sep.	
1	1989	1	2	4	6	2	1	16
	1990	-	1	2	6	10	4	23
	1991	3	4	10	8	5	5	35
2	1989	2	2	-	2	2	1	9
	1990	3	3	1	-	1	1	9
	1991	2	3	1	2	1	1	10
3	1989	2	2	1	-	-	1	6
	1990	1	1	1	-	1	2	6
	1991	2	-	-	1	-	1	4

편, 광화학스모그의 경향을 나타내는 군집 1은 89년에 16일, 90년에 23일, 91년에 35일로서 이 기간중 매년 약 50% 정도씩 발생빈도가 증가하는 것으로 나타났으며, 군집 2는 89년과 90년에 각각 9일, 91년에 10일로 나타났고, 매연성 스모그의 경향을 나타내는 군집 3의 경우에는 89년과 90년에 각각 6일, 91년에 4일이 발생된 것으로 나타났다. 또한 군집 1은 91년의 4월과 5월 중에도 7일이 발생 되었으며, 군집 2 또는 매연성스모그의 경향을 나타내는 군집 3은 6월부터 9월중에도 발생되는 것으로 나타나 각 군집은 계절적인 특성 뿐만이 아니라 날에 따른 기상조건의 차이가 원인이 된 것으로 분석되었다.

4. 3 주성분분석을 이용한 각 군집의 특성

각 군집별로 시정을 제외한 14개의 변수에 대하여 상관행렬에 기초한 주성분을 유도하였으며 고유값(eigen value)이 1 이상이 되는 주성분까지의 누적기여율과 고유벡터(eigen vector)를 표 8에 나타냈다.

군집 1은 첫번째 주성분에서 일사량 및 남북방향의 바람성분을 제외한 기상변수들이 부의 적재를 보인 반면 오염물질은 모두 양의 적재를 나타냈으며 오염물질 중에서도 SO_2 및 TSP의 적재치가 크게 나타났다. 두번째 주성분에서는 수분과 관련된 상대습도 및 가강수량의 적재치가 큰 반면 일사량은 부의 적재를 나타냈다. 또한 네번째 주성분에서는 O_3가 0.61로 높은 적재를 나타냈으며 다섯번째 주성분까지의 누적기여율은 78%를 나타냈다. 군집 2에서 첫번째 주성분은 CO, NO_2, THC 등 1차 오염물질의 적재가 크며, 두번째 주성분에서는 O_3 및 이의 생성과 관련된 기온, 일사량에 높은 적재치를 나타냈다. 군집 3의 첫번째 주성분에서는 TSP 및 SO_2, CO의 적재가 큰 반면 상대습도와 가강수량은 부의 큰 적재를 나타냈다. 또한 두번째 주성분은 O_3와 기온의 적재가 크며, 세번째 주성분에서는 대기중의 수분과 TSP 및 SO_2의 상관이 큰 것으로 나타났고 다섯번째 주성분까지의 누적기여율은 79%로 나타났다.

5. 결 론

본 연구에서는 서울시에서 1989년부터 1991년 하계의 주간에 발생한 스모그의 형태를 분류하기 위하여 14시부터 16시까지 평균 12km 이하의 악시정을 나타냈던 날중 총 118일을 선정하여 분석하였다. 선정된 118일을 대상으로 시정을 악화시키는 요인이 비슷한 날들로 구분하고자 군집분석을 행하였으며, 군집분석에는 중회귀분석의 변수선정 방법을 통해 선정된 TSP, O_3, SO_2, NO_2 및 THC 등의 대기오염 농도자료와 기온, 상대습도, 가강수량, 혼합고, 남북바람성분 등 모두 10개의 변수가 사용되었다. 또한 구분된 각 군집에 대하여 주성분분석을 통해 그 특성을 파악하였다.

군집분석 과정에서 CCC(Cubic Clustering Criterion) 판정기준을 이용한 결과 군집의 수는 3개가 가장 적당한 것으로 분석되었으며, 최적분리 군집분석에 의하여 분류된 군집에는 각각 74일, 28일, 16일이 해당되었다. 또한 3개의 군집은 SO_2, 기온 그리고 가강수량에 의하여 뚜렷하게 분류된 특징을 나타냈다. 각 군집에서 시정에 영향을 주는 요인은 다소 복잡하게 작용하나 군집의 특성을 반영하는 주요인은 구분할 수 있었다. 군집 1은 O_3농도가 가장 높으며 기온이 높은 등 O_3의 생성에 유리한 기상조건을 나타내 광화학반응에 의해 생성된 2차 오염물질에 의한 시정악화가 고려되는 군집으로 생각된다. 군집 3은 3개의 군집중 가장 건조하면서, TSP와 SO_2 등 1차 오염물질의 농도가 가장 높은 분포를 보이는 반면 O_3는 가장 낮은 농도를 나타냈다. 따라서 군집 3은 건조하면서 매연성스모그의 경향이 짙은 날들로 판단된다. 군집 2의 경우에는 O_3를 제외한 다른 오염물질은 대체로 군집 1과 비슷한 농도를 나타냈으나, 기상요소들의 경우에는 군집 1 또는 군집 3과 비슷한 특징을 나타내, 군집 2는 광화학스모그 또는 매연성스모그의 중간적인 특징을 나타낸 날들로 사료된다.

따라서 12km 이하의 악시정을 기록한 전체 118

Table 8. Eigen values and Eigen vectors for each cluster. PC denotes principal component.

Var.	Eigen vector													
	Cluster 1					Cluster 2				Cluster 3				
	PC1	PC2	PC3	PC4	PC5	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4	PC5
SR	0.14	-0.48	0.22	-0.12	0.07	-0.22	0.42	0.11	0.05	0.29	0.15	0.38	-0.18	-0.21
Ta	-0.14	0.13	0.60	0.17	0.07	0.24	0.49	-0.10	-0.06	-0.18	0.46	0.31	0.18	-0.06
CL	-0.24	0.18	-0.37	0.05	0.38	0.23	-0.13	-0.43	-0.15	-0.16	-0.17	-0.18	0.11	0.48
U_x	-0.30	0.02	-0.08	0.31	-0.43	0.41	0.02	0.15	-0.13	-0.30	0.11	0.15	-0.28	0.39
U_y	-0.17	0.14	0.34	-0.20	0.17	0.08	0.29	0.07	0.68	-0.14	-0.23	0.28	0.35	0.13
RH	-0.13	0.43	-0.15	-0.21	0.28	0.10	-0.36	-0.37	0.28	-9.32	-0.25	-0.24	0.33	-0.21
Wp	-0.21	0.42	0.31	-0.11	0.21	0.31	0.13	-0.37	0.23	-0.41	0.17	0.08	0.39	-0.06
MH	-0.08	-0.39	0.16	-0.36	0.36	-0.37	0.01	0.04	0.37	0.25	-0.34	0.33	0.01	-0.28
O_3	0.20	-0.04	0.24	0.61	0.27	0.06	0.47	-0.06	-0.36	0.03	0.42	0.34	0.16	0.26
SO_2	0.44	-0.01	-0.12	0.14	-0.10	0.19	-0.41	0.40	-0.09	0.38	0.04	-0.21	0.29	0.26
NO_2	0.35	0.03	-0.24	0.28	0.39	0.35	0.11	0.22	0.14	0.08	0.35	-0.42	-0.30	0.10
CO	0.28	0.26	0.05	-0.38	-0.24	0.39	-0.05	0.14	0.01	0.35	-0.06	-0.02	0.32	0.20
THC	0.28	0.29	0.22	0.09	-0.27	0.33	-0.04	0.11	0.25	-0.11	0.31	-0.30	0.12	-0.49
TSP	0.45	0.16	-0.10	0.06	0.09	0.02	-0.03	0.50	0.05	0.34	0.26	-0.16	0.38	-0.02
Eigen value	3.71	2.78	1.90	1.51	1.03	4.14	2.69	2.21	1.13	3.68	2.51	2.16	1.43	1.32
Cumul. prob.(%)	26.5	46.4	60.0	70.7	78.1	29.6	48.8	64.6	72.7	26.3	44.2	59.6	69.9	79.3

일중 광화학스모그의 경향을 나타냈던 날은 74일로서 약 62.7%를 차지했으며, 매연성스모그의 경향을 나타냈던 날은 16일로 약 13.6%, 위의 광화학스모그 및 매연성스모그의 중간적인 경향을 나타냈던 날은 28일로서 약 23.7%가 발생된 것으로 분석되었다. 이러한 각 군집의 발생일은 계절적인 특성에 의하여 주로 구분되었으나 기온, 습도 등 날에 따른 기상조건의 차이 또한 다소 반영되어 구분된 것으로 판단된다. 한편, 광화학스모그를 나타낸 74일 중에서도 각 연도별 발생일수는 89년에 16일, 90년에 23일, 91년에 35일로서 이 기간중에 매년 약 50%씩 증가한 것으로 나타났다. 비록 본 연구에서는 대상기간을 3년으로 하였으므로 경년변화를 파악하기에는 다소 짧은 기간이라고 생각되나, 이 기간중에 광화학스모그의 발생빈도가 많고 또한 매년 현저히 증가하는 특징을 나타내었으므로 이에 대한 대책이 절실히 요구되는 것으로 사료된다.

참 고 문 헌

김기영, 전명식 (1991) SAS 군집분석. 자유아카데미, 68pp.

이종범 (1991) 중부지방 각지의 대기오염잠재력에 관한 연구. 한국대기보전학회지, 7(1), 41-47.

정용승, 김태군, 정재섭 (1992) 도시의 대기오염과 시정감소에 관하여. 한국대기보전학회지, 8(1), 1-6.

허명회 (1991) SAS 회귀분석. 자유아카데미, 91pp.

Barone, J.B., T.A. Cahill, R.A. Eldred, R.G. Flocchini, D.J. Shadoan, and T.M. Dietz (1978) A Multivariate statistical analysis of visibility degradation at four California cities. Atmospheric Environment, 12, 2213-2221.

Leaderer, B.P., T.R. Holford, and J. Stolwijk (1979) Relationships between air pollutant and visibility. J. Air Pol. Control Assoc., 29, 154-157.

Oke, T.R. (1978) Boundary Layer Climates. Methuen & Co., 281-284.

Philip, K.H. (1980) Source identification and resolution through application of factor and

cluster analysis. NYAS, 103-115.

Philip, K.H., E.S. Gladney, and A.G. Jones (1976) The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. Atmospheric Environment, 10, 1015-1025.

Solot, S.B. (1939) Computation of depth of precipitable water in a column of air. Monthly Weather Review, 67(4), 100-103.

Trijonis, J.C. (1979) Visibility in the southwest an exploration of the historical data base. Atmospheric Environment, 13, 833-843.

William, H. and J. Lewis (1980) Protection against visibility impairment under the Clean Air Act. J. Air Pol. Control Assoc., 30(2), 118-120.

William, M.C. and J. Clark (1981) Ambient ozone concentration patterns among eastern U.S. urban areas using factor analysis. J. Air Pol. Control Assoc., 31(7), 762-766.