

## Cluster Analysis with Air Pollutants and Meteorological Factors in Seoul<sup>1)</sup>

Jaehee Kim<sup>2)</sup> . Jiwon Lim<sup>3)</sup>

### Abstract

Principal component analysis, factor analysis and cluster analysis have been performed to analyze the relationship between air pollutants and meteorological variables measured in 1999 in Seoul.

In principal analysis, the first principal has been shown the contrast effect between  $O_3$  and the other pollutants, the second principal has been shown the contrast effect between  $CO$ ,  $SO_2$ ,  $NO_2$  and  $O_3$ ,  $PM_{10}$ ,  $TSP$ . In factor analysis, the first factor has been found as  $PM_{10}$ ,  $TSP$ ,  $NO_2$  concentrations which are related with suspended particulates. As a result of cluster analysis, three clusters respectively have represented different air pollution levels, seasonal characteristics of air pollutants and meteorological situations.

**주제어** : 대기오염물질, 기상변수, 주성분분석, 인자분석, 군집분석

## 1. 서론

### 1.1 연구배경 및 목적

우리나라는 60~70년대 이후 약 30년 간 산업이 급속히 발달함에 따라 대기오염의 문제 또한 심각하게 발생하기 시작했다. 그리고 전 세계적으로 연간 약 300만 명이 대기오염으로 사망하는 것으로 추정된다. 그리스 아테네에서는 매년 매연으로 인한

---

1) This work was supported by the research fund of Duksung Women's University in 2003

2) Associate Professor, Dept. of Statistics, Duksung Women's University, 419 Ssangmun-Dong Tobong-Ku Seoul, Korea.  
E-mail : jaehee@duksung.ac.kr

3) Graduate Student in Master Program, Dept. of Statistics, Duksung Women's University, 419 Ssangmun-Dong Tobong-Ku Seoul, Korea.  
E-mail : jiwon9711@hanmail.net

사망자가 5%씩 증가하고, 연100만 명 정도를 대기오염에 의한 사망으로 보고있다. 그리고 이종현(2002)에 의하면 WHO는 1999년 보고서에서 천식의 30~40%, 모든 호흡기 질환의 20~30%가 대기오염과 관련된 것으로 추정하고 있다. 우리나라에서도 김윤신(1998) 등에 의하면 대기오염물질 중  $O_3$ 의 농도가 100ppb 증가함에 따라 사망자수가 울산지역은 127%, 여천지역은 116% 증가함을 보여줬고, Lee(2000)등은 우리나라 7대 광역시에 대한 오염물질과 사망률과의 유의한 관계를 제시하였다. 대기오염의 심각성은 이미 전문가들뿐 아니라 일반 시민들의 인식에서도 분명히 확인되고 있다. 수도권 남녀 1,400명을 대상으로 한 최근의 한 설문 조사에서 건강에 가장 해를 끼치는 오염으로 대기오염이 61.6%로 1위를, 수질오염과 쓰레기오염이 각각 15.8%, 9.4%로 2, 3위를 차지했다. 따라서 대기오염현상 파악과 더불어 대기오염현상에 대한 효율적인 규제 및 방지를 위하여 체계적이고 종합적인 대기오염물질의 관리가 필요하다. 이전에 오광중(1998) 등은 인자분석과 중회귀분석 결과모형을 제시하여 부산지역의 대기오염물질과 기상변수간의 관계를 파악하였고, 어수미(1997) 등은 기상인자가 대기오염물질에 미치는 영향을 연구하였다. 그리고 Bruno and Cocchi(2002), Kassomenos et al.(1999), Prabhat and Tyagi(1999) 등은 기존의 오염지수를 보완한 새로운 대기오염지수를 제안했다.

본 연구에서는 우리나라에서 가장 경제활동이 활발하고 인구가 많은 서울의 27개 지역에서 1년간 측정된 대기오염물질 자료를 가지고 대기오염들의 특성과 그것들간의 관계를 알아봄으로서 대기오염 방지대책을 위한 기초자료를 제공한다. 특히, 군집분석의 결과는 대기오염 예보체제에서 환경기준의 판단자료로도 제공이 가능할 것으로 보인다.

## 1.2 대기오염자료와 기상자료

본 연구에서 사용하는 대기오염물질 자료는 1999년 1월1일부터 12월31일까지 서울 27개 지역에서 매일 매시간 측정된  $CO$ ,  $SO_2$ ,  $NO_2$ ,  $O_3$ ,  $PM_{10}$ ,  $TSP$  자료를 이용하였다. 우선 매시간 24개의 지역산출평균을 구하여 3시, 9시, 15시, 21시 자료들 중에서 최대값을 취해 하루의 오염 대표값으로 사용하였다. 유효자료의 숫자는 365개이며 자료는 환경청으로부터 받았다. 기상자료는 기온(temperature), 강우량(precipitation), 습도(humidity), 풍속(wind speed), 풍향(wind direction)을 기상청으로부터 받아 사용하였다. <표1.1>은 본 연구에서 사용한 대기오염물질의 측정단위 및 국내환경기준을 보여준다.

<표1.1> 대기오염물질 측정단위 및 국내환경기준

기간	대기오염물질	측정단위	국내환경기준
1999년 1월1일~ 12월31일	CO(일산화탄소)	ppm	8시간평균치9ppm이하/1시간평균치25ppm이하
	SO <sub>2</sub> (이산화황)	ppb	연간평균치20ppb이하/24시간평균치50ppb이하/1시간 평균치150ppb이하
	NO <sub>2</sub> (이산화질소)	ppb	연간평균치50ppb이하/24시간평균치80ppb이하/1시간 평균치150ppb이하
	O <sub>3</sub> (오존)	ppb	8시간평균치60ppb이하/1시간평균치100ppb이하
	PM10(미세먼지)	μg/m <sup>3</sup>	연간평균치70μg/m <sup>3</sup> 이하 /24시간평균치150μg/m <sup>3</sup> 이하
	TSP(총분진)	μg/m <sup>3</sup>	연간평균치150μg/m <sup>3</sup> 이하 /24시간평균치300μg/m <sup>3</sup> 이하

다음에서는 환경오염의 주요원인이 되는 대기오염물질에 대한 이해를 돕기 위해 다음과 같이 간단하게 설명하고자 한다.

(1) CO

일산화탄소라고도 하며 석탄·석유등을 대량으로 소비하는 공장지대에서는 상당한 양(5ppm 정도)에 달하는 수도 있다. 또, 가정에 공급되고 있는 도시 가스의 주성분이다. 물에 잘 녹지 않으며, 공기보다 약간 가볍고, 공기 중에서 점화하면 청색 불꽃을 내며 타서 이산화탄소가 된다. 환원성이 있으며, 촉매 존재 하에서 탄소와 이산화탄소로 분해한다. 한편, 일산화탄소를 흡입하면 혈액 속에 있는 헤모글로빈과 결합하는데 그 친화력은 산소의 200배이며, 보다 안정된 화합물을 만든다. 공기 중에 소량이 존재해도 호흡을 저해한다.

(2) SO<sub>2</sub>

아황산가스·아황산무수물이라고도 하며 자극적인 냄새가 나는 무색 기체로 유독하다. 물에 잘 녹으며, 수용액은 아황산을 생성하며 산성을 띤다. 황산 제조의 원료로서 중요할 뿐 아니라, 표백제·환원제로도 사용되며, 액체는 붉은 인·요오드·황 등의 용매로도 사용된다. 짙은 기체를 흡입하면 콧물·담·기침이 나며 목구멍이나 가슴이 아프고, 호흡이 곤란해진다. 기관지염·폐수종·폐렴 등이 되는 수도 있다. 석유의 정제시나 중유가 연소할 때 원유에 함유되어 있는 황이 산화되어 공중에 방출되는데, 최근 에너지원이 석유로 전환됨에 따라 아황산가스의 대기 중 농도가 증가하여, 대기오염물질 중에서 큰 비중을 차지하게 되었다.

(3) NO<sub>2</sub>

석유나 석탄의 연소에 의하여 발생하며, 발생원은 배기가스를 내뿜는 자동차가 대부분을 차지하며, 그 다음으로 광공업, 발전소 순이다. 가정의 난방에 석유난로 등을 사용하면 실내 질소산화물이 외부의 공기보다 몇 배나 많아지는 경우가 있다. 연소에 의하여 발생하는 것은 주로 일산화질소인데, 이것이 대기 중에 방출되면 산화되어 이산화질소가 된다. 이산화질소는 인체에 유해하며 고농도 하에서는 폐기종, 기관지염 등 호흡기 질환의 원인이 된다.

(4)  $O_3$ 

특이한 냄새가 나며, 공기 속에 0.0002부피 %만 존재해도 냄새를 감지할 수 있다. 기체는 물에 잘 녹지 않는다. 물에 녹은 오존은 서서히 분해한다. 오존은 직접 공기로 배출되는 것이 아니라 공기중의 질소산화물과 휘발성유기 화합물이 태양광선을 받아 화학반응을 일으킨 결과로써 생성된다. 상온에서 자발적으로 분해되어 산소가 되고, 이산화망간 · 백금 분말 등은 분해를 촉진시킨다. 인체가 오존에 반복적으로 노출되면 폐에 비가역적 손상이 초래될 수 있다. 오존이 비록 낮은 농도일지라도 흡입하게 되면 가슴통증, 기침, 메스꺼움, 인후자극, 충혈과 같은 다양한 건강문제를 야기한다. 그리고 기관지염, 심장병, 폐 기종, 천식을 악화시킬 수 있고, 폐 용량을 감소시킬 수 있다. 또한 건강한 사람도 오존에 노출되면 호흡의 곤란을 경험한다.

(5)  $PM_{10}$ (Particulate Matter Less than  $10\mu m$ )

매우 작은 먼지 입자로 지름  $10\mu m$  이하이다. 연소작용에 의해 발생되므로 황산염, 질산염, 암모니아 등의 이온 성분과 금속화합물, 탄소화합물 등 유해물질로 이루어져 있다. 대도시의 미세먼지는 70% 이상이 자동차 배기가스에서 나오는데, 일반 먼지보다 더욱 엄격하게 규제한다. 사람의 폐포까지 깊숙하게 침투해 각종 호흡기 질환의 직접적인 원인이 되며 우리 몸의 면역 기능을 떨어뜨린다.

(6)  $TSP$ (Total Suspended Particulate)

대기중에 떠다니거나 흩날려 내려오는 입자상 물질의 하나로 일명 분진이라 한다. 보통  $0.1\sim 500\mu m$ 의 입경범위를 가지며, 입자의 크기에 따라 무거워서 침강하기 쉬운 것을 강하분진이라 하고, 입자가 미세하고 가벼워서 좀처럼 침강하기 어려워 장기간 대기중에 떠다니는 것을 부유분진이라 한다. 1995년부터 TSP 측정기가 미세먼지 ( $PM_{10}$ ) 측정기로 단계적으로 전환되었고, TSP농도는 2000년 대기환경기준 항목에서 제외되었다.

### 1.3 분석방법

기본적으로 각 대기오염물질들의 기술통계량과 상관관계를 구한다. 주성분분석을 통하여 다변량 자료를 요약하는 몇 개의 주성분을 찾아내어 원래 자료들을 식별하고 해석한다. 인자분석을 실시하여 원래 자료들에 내재된 잠재변수인 공통인자를 찾아낸다. 마지막으로 대기오염물질들을 이용한 군집분석과 주성분점수를 이용한 군집분석을 실행하여 관측치들을 몇 개의 군집으로 분류하고, 각 군집들의 특징을 알아보며 비교하고자 한다. 여기서, 다변량분석기법을 실행하기에 앞서 측정단위가 서로 다르기 때문에 모든 대기오염자료들을 평균이 0, 분산이 1인 자료로 표준화하였다. 통계분석에는 통계패키지 SAS 8.1을 이용하였다.

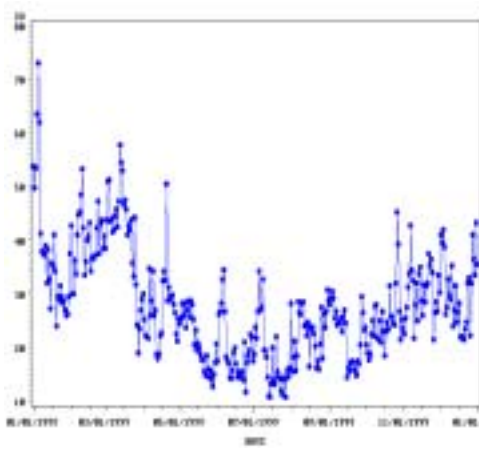
## 2. 대기오염 자료에 대한 통계분석

### 2.1 기술통계량 및 상관관계

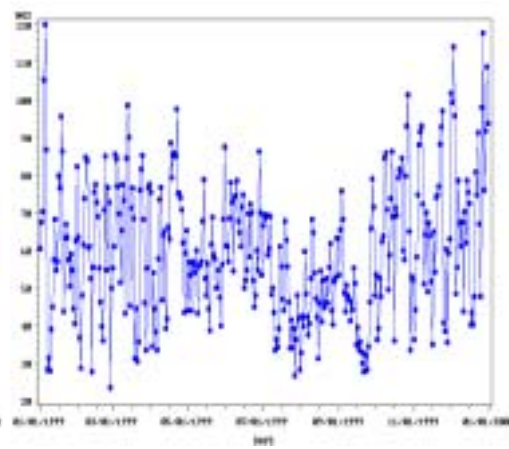
대기오염물질들의 기술통계량 <표2.1>과 <그림2.1> ~ <그림2.6>를 보면  $CO$ ,  $SO_2$ ,

$NO_2$ 의 농도는 여름에 낮고, 난방연료 사용량이 늘고 자동차 이용량이 많은 겨울일수록 높아지는 경향이 있는데 반해  $O_3$ 는 여름에 높아졌다가 기온이 떨어질수록 낮아지는 경향을 보였는데 이는 여름에 일사량이 강하기 때문인 것으로 보인다.  $PM_{10}$ ,  $TSP$ 는 황사현상의 영향을 받는 3월에서 6월까지 높은 농도를 나타내다가 한여름인 8월에는 가장 낮아지는 경향을 보이고 있다. <표2.2>의 상관관계는  $CO$ ,  $SO_2$ ,  $NO_2$  간에, 그리고  $PM_{10}$ ,  $TSP$ 간에 높은 것으로 나타났다.

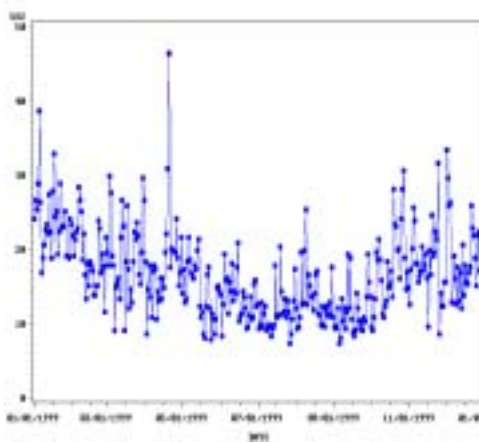
<그림2.1>  $CO$ 의 일일 변동



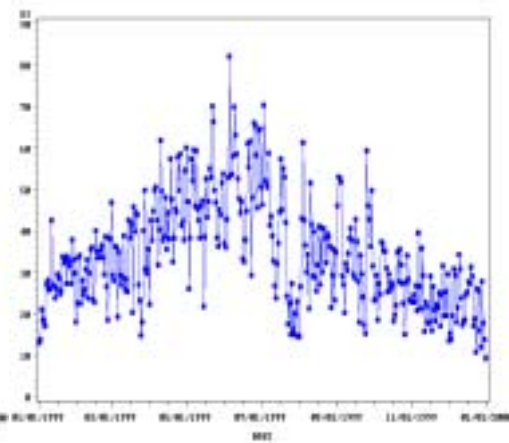
<그림2.2>  $NO_2$ 의 일일 변동

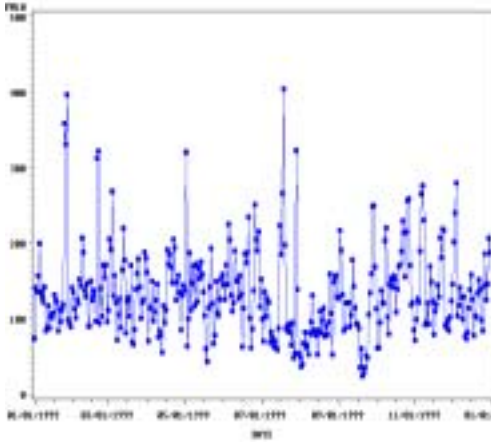
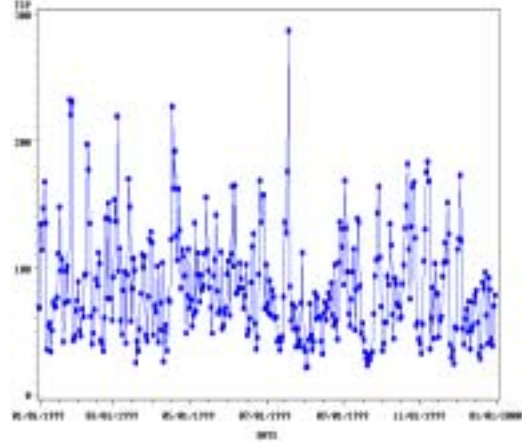


<그림2.3>  $SO_2$ 의 일일변동



<그림2.4>  $O_3$ 의 일일변동



<그림2.5>  $PM_{10}$ 의 일일변동<그림2.6>  $TSP$ 의 일일변동

&lt;표2.1&gt; 1999년 대기오염물질들의 월별 평균과 표준편차

	$CO$	$NO_2$	$O_3$	$PM_{10}$	$SO_2$	$TSP$
	평균 ± 표준편차					
1999년	27.96 ± 10.20	59.98 ± 18.25	34.45 ± 13.07	132.84 ± 58.35	16.53 ± 5.77	85.42 ± 42.81
1월	37.79 ± 12.14	62.79 ± 21.67	26.85 ± 6.77	139.72 ± 78.57	23.83 ± 4.50	98.41 ± 55.65
2월	39.86 ± 5.43	58.70 ± 17.99	30.78 ± 5.99	145.68 ± 56.60	19.05 ± 4.21	87.15 ± 43.23
3월	41.22 ± 9.34	65.16 ± 18.98	32.83 ± 9.16	138.63 ± 48.08	19.16 ± 5.57	92.67 ± 44.32
4월	27.33 ± 6.48	64.01 ± 17.56	44.08 ± 8.78	129.65 ± 39.63	18.13 ± 6.85	95.27 ± 47.43
5월	21.05 ± 5.12	55.66 ± 10.62	47.62 ± 10.39	134.16 ± 51.73	14.26 ± 3.85	86.66 ± 27.07
6월	19.49 ± 5.69	63.88 ± 10.45	52.18 ± 12.53	153.45 ± 48.59	13.46 ± 3.01	96.44 ± 38.51
7월	18.38 ± 6.56	48.93 ± 13.47	37.02 ± 15.52	121.89 ± 83.60	11.44 ± 2.86	77.93 ± 50.53
8월	22.73 ± 4.52	49.00 ± 9.43	34.37 ± 9.32	90.03 ± 34.32	13.29 ± 3.61	66.10 ± 29.19
9월	22.36 ± 5.17	47.05 ± 14.50	35.42 ± 11.39	115.10 ± 60.60	11.52 ± 3.21	80.75 ± 43.22
10월	25.01 ± 5.76	65.52 ± 17.73	27.33 ± 5.84	150.37 ± 53.31	17.39 ± 5.51	91.17 ± 40.63
11월	30.48 ± 4.91	64.35 ± 18.57	23.73 ± 5.61	140.30 ± 57.24	17.85 ± 4.78	82.67 ± 46.35
12월	30.58 ± 6.74	74.54 ± 22.52	21.41 ± 6.89	136.61 ± 48.39	19.06 ± 5.22	70.39 ± 32.03

<표2.2> 대기오염물질들간의 상관관계

오염물질	CO	NO <sub>2</sub>	O <sub>3</sub>	PM10	SO <sub>2</sub>	TSP
CO	1.00	0.49**	-0.31**	0.25**	0.61**	0.31**
NO <sub>2</sub>		1.00	-0.11*	0.50**	0.62**	0.60**
O <sub>3</sub>			1.00	0.097	-0.29**	0.11*
PM10				1.00	0.38**	0.77**
SO <sub>2</sub>					1.00	0.45**
TSP						1.00

\*\* : p<0.01

\* : p<0.05

## 2.2 대기오염자료에 대한 주성분분석

대기오염물질들의 주요 변동을 설명할 수 있는 주성분을 찾고자 주성분분석을 실시하여 <표2.3>과 같이 고유값, 누적설명비율 그리고 주성분계수를 얻었다.

<표2.3> 주성분계수 및 고유값

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6
고유값	3.053	1.408	0.583	0.424	0.324	0.205
누적분산비율	0.508	0.743	0.840	0.911	0.965	1.000
CO	0.394	-0.379	0.458	0.637	-0.290	0.0001
NO <sub>2</sub>	0.480	0.002	0.161	-0.654	-0.532	0.178
O <sub>3</sub>	-0.105	0.691	0.699	0.026	0.115	0.080
PM10	0.421	0.403	-0.428	0.317	0.066	0.608
SO <sub>2</sub>	0.461	-0.266	0.207	-0.237	0.783	0.041
TSP	0.459	0.378	-0.217	0.089	-0.026	-0.767

주성분분석을 시행한 결과는 최대 고유값은 3.053이고 전체 표본 변동량 중 약

50%를 설명하며, 처음 두 개 주성분에 의해 전체 변동량 중 총 75%를 설명할 수 있어 2개의 주성분으로 차원을 줄여 자료를 설명할 수 있다.

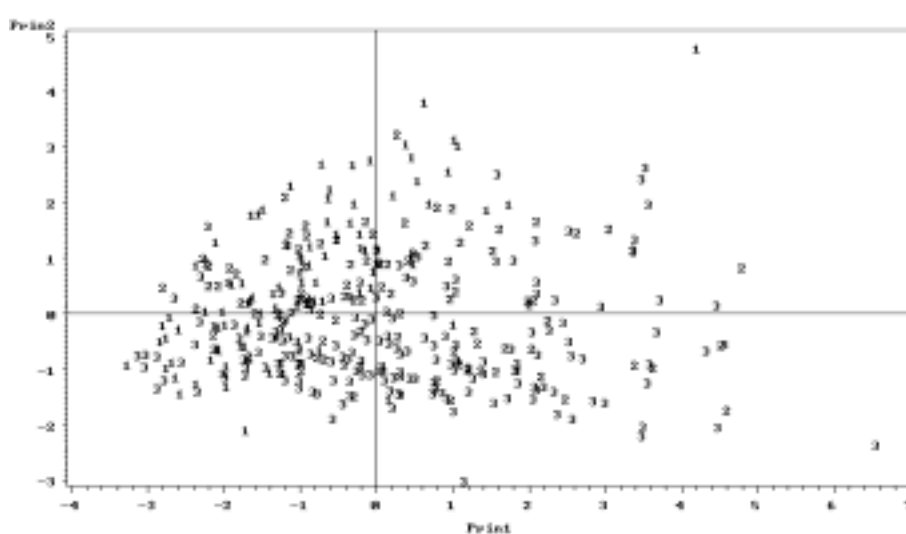
제1주성분은  $O_3$ 과 나머지 오염물질들간의 대비효과를 나타내며  $O_3$ 를 제외한 나머지 오염물질들의 농도가 높아지면 주성분점수가 커진다. 즉, 제1주성분은 대기오염도를 나타내는 지표로  $NO_2$ 가 가장 큰 영향을 미치고, 그 뒤로  $SO_2$ ,  $TSP$ ,  $PM10$ ,  $CO$ ,  $O_3$  순으로 영향을 미치는 것으로 나타났다. 제2주성분은  $CO$ ,  $SO_2$ ,  $NO_2$ 와  $O_3$ ,  $PM10$ ,  $TSP$ 간의 대비효과를 나타내고  $O_3$ 의 농도가 가장 큰 영향을 미치며,  $O_3$ 과 분진오염도를 나타낸다. 이 주성분들은 다음과 같은 식으로 표현된다:

$$Y_1 = 0.394CO + 0.480NO_2 - 0.105O_3 + 0.421PM10 + 0.461SO_2 + 0.459TSP$$

$$Y_2 = -0.379CO + 0.002NO_2 + 0.691O_3 + 0.403PM10 - 0.266SO_2 + 0.378TSP.$$

다음은 주성분점수들간의 산점도를 그려 주성분들의 계절에 따른 특징을 찾아보았다. <그림2.7>은 오염물질들에 대한 제1주성분점수와 제2주성분점수간의 산점도이다. 기온이 높은 여름을 나타내는 관측치들의 제1주성분점수가 커질수록 제2주성분점수도 커지는 모습을 보인다. 반대로 기온이 낮은 겨울을 나타내는 관측치들의 제1주성분과 제2주성분은 반비례 경향을 보인다. 봄과 가을 관측치들은 여름과 겨울자료들 사이에 분포하며 눈에 띄는 경향을 보이지는 않는다.

<그림2.7> 주성분점수들간의 산점도



여름(6월~8월):1, 봄(3월~5월)-가을(9월~11월):2, 겨울(12월~2월):3

### 2.3 대기오염자료에 대한 인자분석

대기오염물질들간에 내재하는 공통인자를 찾아내기 위하여 직교인자모형을 이용한



인자분석을 실시하였다. 6개 대기오염변수에 대한 직교인자모형은

$$X_i = l_{i1}F_1 + l_{i2}F_2 + \dots + l_{ij}F_j + \dots + l_{im}F_m + \epsilon_i, \quad i = 1, 2, \dots, 6, \quad m < 6$$

여기서,  $X_i$ 는 측정된  $i$ 번째 변수이고,  $l_{ij}$ 는  $i$ 번째 변수의  $j$ 번째 인자의 적재값(factor loading)이다.  $F_j$ 는  $j$ 번째 공통인자(common factor)이며, 오차항  $\epsilon_i$ 는  $X_i$ 와 관련한 특수인자(specific factor)로 공통인자에 의해 설명할 수 없는 부분을 말한다. 이것을 다시 행렬로 표현하면

$$X = LF + \epsilon$$

이고,  $L$ 은 인자적재행렬,  $F$ 는 인자행렬을 나타낸다. 모형에 대한 가정은

$$E(F) = 0, \quad Cov(F) = I, \quad E(\epsilon) = 0, \quad Cov(\epsilon) = \Psi, \quad Cov(\epsilon, F) = 0$$

인자분석의 결과 상관행렬의 고유값이 1보다 큰 공통인자의 수는 2개가 나왔으며, 이들에 의해 설명되는 변동량은 약 75%이므로 인자수의 결정은 적절하다고 판단된다.

인자적재값을 해석의 편리를 위해 VARIMAX 회전을 실시하였고, 결과는 다음 <표 2.4>와 같다. 인자1은 분진을 나타내는  $PM_{10}$ ,  $TSP$ 와 분진 생성에 영향을 많이 주는  $NO_2$ 가 인자에 대해 비중이 높게 의존하는 것으로 나타났고 높은 양의 관계를 보였다. 인자2는 주로 자동차나 난방연료의 연소로부터 발생하는  $CO$ ,  $SO_2$ 가 인자에 대해 높은 양의 상관관계를 가졌으며  $O_3$ 는 높은 음의 관계를 나타냈다.

대기오염물질들이 공통인자에 의해 설명되는 정도를 나타내는 것이 공통성(communality)인데 제일 낮은 공통성을 갖는  $CO$ 는 추출된 2개의 인자에 의해 약 68%가 설명되며 나머지 오염물질들의 공통성은 70% 이상으로 공통인자에 의해 설명되는 비율이 대체적으로 높다.

<표 2.4> 대기오염물질들의 인자분석 적재값도표

	인자1( $F_1$ )	인자2( $F_2$ )	공통성
고유값	3.053	1.408	·
누적분산비율	0.508	0.743	
$CO$	0.324	0.756	0.677
$NO_2$	0.700	0.462	0.704
$O_3$	0.300	-0.785	0.708
$PM_{10}$	0.878	0.009	0.772
$SO_2$	0.496	0.710	0.750
$TSP$	0.918	0.070	0.847

인자분석 결과 각 오염 물질에 대해 인자1  $F_1$ 과 인자2  $F_2$ 를 가진 다음의 식으로 표현된다:

$$CO = 0.324F_1 + 0.756F_2$$

$$NO_2 = 0.700F_1 + 0.462F_2$$

$$O_3 = 0.300 F_1 - 0.785 F_2$$

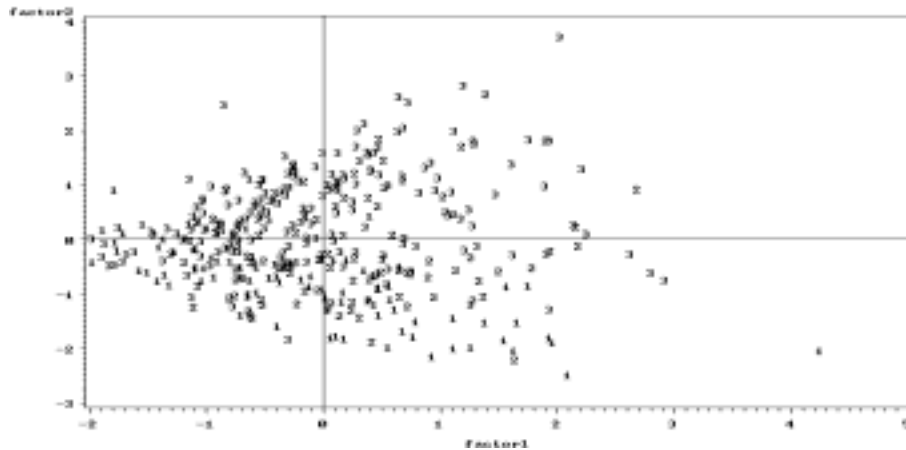
$$PM10 = 0.878 F_1 + 0.009 F_2$$

$$SO_2 = 0.496 F_1 + 0.710 F_2$$

$$TSP = 0.918 F_1 + 0.070 F_2$$

<그림2.8>은 주성분법에 의한 인자추정으로 구해진 두 인자의 인자점수를 산점도로 나타낸 것이다. 하절기를 나타내는 관측치들은 인자1점수가 커질수록 인자2점수는 작아지는 반비례 경향을 보인다. 반대로 동절기를 나타내는 관측치들은 인자1점수가 작을수록 인자2점수는 커지는 정비례 경향을 나타낸다. 봄과 가을 관측치들은 인자2점수에 대해서는 여름과 겨울자료의 중간에 나타나며 인자점수1을 기준으로 보면 골고루 퍼져있는 것으로 보인다.

<그림2.8> 인자1점수와 인자2점수와의 산점도



여름(6월~8월):1, 봄(3월~5월)·가을(9월~11월):2, 겨울(12월~2월):3

### 3. 군집분석

대기오염정도가 서로 유사한 날들이 자발적으로 묶이도록 군집분석을 실시하였다. 우선 정보의 손실이 최소가 되도록 하는 Ward방법을 이용한 위계적 군집분석의 결과를 바탕으로 하여 군집의 수를 3개로 하는 K-Means 군집분석을 실행하였다. <표 2.6>는 각 군집의 오염물질들은 물론이고, 대기오염정도에 따라 생성된 군집의 대기오염과 관련있는 기상변수들의 기술통계량 및 분산분석(ANOVA) 결과를 보여준다.

군집1은 259개, 군집2는 63개, 군집3은 43개로 대기오염물질 자료들에 의한 관측백터들이 분류되었다. 여기서 대기오염원과 기온, 풍속은 유의수준 1%에서 분산분석 실시 결과 군집들간에 차이가 있었으며, 모든 군집에서 평균적으로 서풍 및 서남서풍 경향을 보여주었다.

군집1의 특징은  $O_3$ 를 제외한 모든 오염물질들의 농도가 군집들 중에서 가장 낮았고, 기온이  $14.1^\circ\text{C}$ 로 1999년도 평균기온( $13.25^\circ\text{C}$ )정도이며, 풍속이 2.01%로 다른 군집의 자료들에 비해 센 편이었다. 풍속이 높은 경우 2차오염원인 오존의 생성과정을 촉진하는 것으로 보인다. 군집1의 관측치들은 전체적으로 일년 중 오염정도가 비교적 낮은 날들이 포함된다.

군집2는  $CO$ ,  $SO_2$ ,  $NO_2$ 의 농도가 군집들 중에서 가장 높았으며 특히,  $CO$ 는 41.37ppm로 국내환경기준치(1시간평균치25ppm이하)보다 훨씬 높은 수치였다.  $O_3$ 의 농도는 가장 낮았다.  $PM_{10}$ ,  $TSP$ 의 상대적으로 세 군집중 두 번째였으나,  $PM_{10}$ 은  $165.39\mu\text{g}/\text{m}^3$ 으로 국내환경기준(24시간평균치  $150\mu\text{g}/\text{m}^3$ 이하)에 비교해봤을 때 매우 높은 수치이다. 평균기온이  $5.66^\circ\text{C}$ 로 매우 낮고, 풍속이 1.43%로 비교적 낮았으며, 관측치들은 대부분 11월에서 4월 사이에 분포하는 동절기 자료들이었다.

군집3은  $PM_{10}$ ,  $TSP$ 의 농도가 다른 군집들에 비해 가장 높았고,  $PM_{10}$ 은  $223.13\mu\text{g}/\text{m}^3$ 으로 국내환경기준을 훨씬 뛰어넘었다.  $CO$ ,  $SO_2$ ,  $NO_2$ 는 군집중에서 두 번째였으나,  $CO$ 는 국내환경기준에 거의 넘어서는 수치였다.  $O_3$ 의 농도는 상대적으로 가장 높았고, 평균기온도  $19.27^\circ\text{C}$ 로 높은 편이며, 관측치들이 주로 5월과 7월 사이에 분포하는 것으로 나타났다. 이는  $O_3$ 가 광화학반응에 의해 생성되므로 일사량이 많은 하절기 자료들로 분포하는 것은 당연한 결과로 보인다. 월별 자료를 보면 6월에  $PM_{10}$ ,  $TSP$ ,  $O_3$ 의 농도가 가장 높았으며 이를 반영하는 군집으로 여겨진다.

<표2.5> 주성분점수를 이용한 군집분석의 주성분점수와 인자점수의 기술통계량

군집	관측치 수	주성분	평균 ± 표준편차	인자	평균 ± 표준편차
1	254	제1주성분*	$-0.89 \pm 1.01$	인자1*	$-0.48 \pm 0.63$
		제2주성분*	$-0.11 \pm 0.91$	인자2*	$-0.20 \pm 0.72$
2	63	제1주성분	$2.50 \pm 1.16$	인자1	$0.75 \pm 0.66$
		제2주성분	$-0.93 \pm 0.73$	인자2	$1.45 \pm 0.62$
3	48	제1주성분	$1.44 \pm 1.18$	인자1	$1.55 \pm 0.68$
		제2주성분	$1.85 \pm 0.93$	인자2	$-0.84 \pm 0.78$

\*: 유의수준 0.01에서 3개의 군집간 분산분석시 유의한 변수

또한 2.1절에서 구한 주성분점수를 이용하여 군집분석을 실시해 보았다. 주성분점수에 의해 분류된 각 군집의 주성분점수와 인자점수의 기술통계량을 나타낸 것이 <표 2.5>이다. 제1주성분은  $O_3$ 의 농도가 높아지면 주성분점수의 값이 작아지고, 나머지 오염물질들의 농도가 높아지면 값이 커지므로 하절기 관측치들의 주성분점수들은 다른 군집에 비해 작은 값으로 나타나고, 동절기 관측치들의 주성분점수는 비교적 높은 값으로 나타난다. 그러나 제2주성분은  $O_3$ ,  $PM_{10}$ ,  $TSP$ 의 농도가 높아지면 주성분점수가 커지므로, 하절기 관측치들의 주성분점수는 다른 군집에 비해 높은 값으로 나

타나고, 동절기 관측치들의 주성분점수는 비교적 작은 값으로 나타난다. 따라서 하절기 관측치가 주로 분포하는 군집3은 제2주성분점수가 군집 중에서 가장 높고, 동절기 관측치들이 대부분인 군집2는 제1주성분점수가 군집 중에서 가장 높은 것이다. 즉, 군집2가 가장 오염도가 심함을 알 수 있다.

앞의 2.3절에서 하절기 관측치들은 대체적으로 인자2 점수가 낮았고, 동절기 관측치들은 높았다. 그런데 주성분점수를 이용한 군집분석 결과에서 하절기 관측치들이 많이 분포하는 군집3의 인자2점수가 가장 작고, 동절기 관측치들이 많이 분포하는 군집2의 인자점수가 가장 크므로, 인자분석의 결과와 일관된 결과를 보여준다.

주성분점수로 분류된 각 군집의 오염물질들과 기상변수에 대한 기술통계량과 분산분석 결과를 나타낸 것이 <표2.7>이다. <표2.7>과 대기오염물질 원자료를 이용한 군집분석 결과 <표2.6>를 비교해보면 거의 차이가 없어 주성분이 대기오염상태를 잘 반영하고, 주성분과 인자가 계절적요인과 관련된 군집의 특성을 나타냄을 알 수 있다.

#### 4. 결론

1999년 서울지역 대기오염물질 자료와 기상변수 자료를 가지고 대기오염의 특징과 그것들간의 관련성을 파악하기 위해 여러 가지 기본적 통계적 분석 및 군집분석을 실시하였다. 주성분분석을 실시한 결과 제1주성분은 대기오염성분으로 구성되었으므로 그날의 오염도의 대표값으로 사용할 수도 있겠다. 제2주성분은  $CO$ ,  $SO_2$ ,  $NO_2$ 와  $O_3$ ,  $PM_{10}$ ,  $TSP$ 간의 대비효과를 나타냈다. 인자분석에서는 분진과 관련이 높은 인자와 자동차나 난방연료의 연소로 발생하는  $CO$ ,  $SO_2$ 와  $O_3$  이 관련이 많은 인자가 나타났다. 대기오염물질 자료와 주성분점수를 가지고 각각 군집을 나누어보고 군집의 특성을 파악한 결과 두가지 방법에 의한 군집의 차이가 거의 없는 것을 알 수 있어 2개의 주성분이 6개의 대기오염변수를 반영하고 있음을 알 수 있었다. 3개의 각 군집에서 대기오염물질들과 기상변수들의 계절적인 특징이 나타나 대기오염이 계절과 관계가 있음을 알 수 있었다.

이외에도 기상관계를 고려한 다각적인 대기오염현상 파악을 위해 여러 통계적 기법을 이용할 수 있으며 대기오염자료에 대한 새로운 통계적 방법을 기대한다.

#### 참고문헌

1. 김기영, 전명식(1997). 다변량 통계자료분석, 자유아카데미, 서울.
2. 김윤신, 최원목, 김무채(1998). 공단지역 대기오염과 일별사망수와의 연관성에 관한 연구, 한국보건통계학회지, 23, 124-136.
3. 어수미, 김광진, 이규남(1997). 기상인자가 대기오염에 미치는 영향, 한국환경위생학회지, 23, 104-114.
4. 오광중, 광진, 정덕영, 손건태(1998). 부산지역의 대기오염물질농도와 기상인자간의 통계분석(광안리 지역을 중심으로), 대한환경공학회지, 20, 1235-1245.

5. 이강청, 김우성(2002). SAS 통계분석, 자유아카데미, 서울.
6. 이종현(2002). 서울시 대기오염도 실태와 대기질 개선을 위한 관리방안, 시민환경연구소
7. 최용석, 정광모(2001). SAS를 활용한 응용다변량 자료분석, 교우사, 서울.
8. Bruno, F. and Cocchi, D. (2002). A unified strategy for building simple air quality indices, *Environmetrics*, 13, 243-261.
9. Lee Jong-Tae, Kim Ho, Hong Yun-Chul, Kwon Ho-Jang, Schwartz J., Christiani, D.C. (2000). Air pollutants and daily mortality in seven major cities, *Environmental Research*, 84, 247-254.
10. Kassomenos, P., Skouloudis, A.N., Lykoudis, S., Flocas, H.A.(1999). Air-quality indicators for uniform indexing of atmospheric pollution over large metropolitan areas, *Atmospheric Environment*, 33, 1861-1879.
11. Prabhat K. S., Tyagi, A.(1999). Formation of air pollution index, *Air & Waste Management Association*, 49, 88-91.

[ 2003년 6월 접수, 2003년 9월 채택 ]

&lt;표2.6&gt; 대기오염물질자료를 이용한 군집분석에 따른 오염물질, 기상변수 기술통계량

군집	관측치(일)수	오염물질, 기상변수	평균 ± 표준편차	최소값	최대값
1	259	<i>CO</i> *	24.97 ± 7.94	10.75	53.00
		<i>NO<sub>2</sub></i> *	52.55 ± 13.68	23.50	91.75
		<i>O<sub>3</sub></i> *	35.11 ± 11.92	14.50	70.25
		<i>PM10</i> *	109.94 ± 40.12	24.25	322.25
		<i>SO<sub>2</sub></i> *	14.34 ± 3.96	7.25	25.25
		<i>TSP</i> *	65.61 ± 24.32	21.50	136.25
		<i>temperature</i> *	14.10 ± 10.70	-9.30	29.90
		<i>precipitation</i>	16.01 ± 35.40	0	261.60
		<i>humidity</i>	65.24 ± 13.71	29.00	91.30
		<i>wind wpeed</i> *	2.01 ± 1.08	0.20	5.30
		<i>wind direction</i>	20.79 ± 10.26	1.00	35.00
2	63	<i>CO</i>	41.37 ± 9.69	23.50	73.00
		<i>NO<sub>2</sub></i>	83.60 ± 13.53	60.50	120.00
		<i>O<sub>3</sub></i>	23.08 ± 7.08	9.25	41.50
		<i>PM10</i>	165.39 ± 43.35	73.75	279.25
		<i>SO<sub>2</sub></i>	24.80 ± 5.11	16.00	46.25
		<i>TSP</i>	119.53 ± 31.87	68.50	219.50
		<i>temperature</i>	5.66 ± 5.21	-7.20	18.40
		<i>precipitation</i>	6.24 ± 12.39	0	42.20
		<i>humidity</i>	64.08 ± 10.83	37.60	81.40
		<i>wind wpeed</i>	1.43 ± 0.77	0.10	4.00
		<i>wind direction</i>	21.42 ± 11.83	1.00	35.00
3	43	<i>CO</i>	26.36 ± 6.98	13.25	43.75
		<i>NO<sub>2</sub></i>	70.09 ± 13.64	44.00	97.50
		<i>O<sub>3</sub></i>	47.13 ± 13.18	21.25	82.00
		<i>PM10</i>	223.13 ± 60.60	143.75	403.75
		<i>SO<sub>2</sub></i>	17.60 ± 4.53	9.25	30.75
		<i>TSP</i>	154.75 ± 40.85	74.75	287.00
		<i>temperature</i>	19.27 ± 6.39	3.80	27.60
		<i>precipitation</i>	5.23 ± 12.89	0	39.50
		<i>humidity</i>	66.67 ± 10.21	46.50	81.90
		<i>wind wpeed</i>	1.70 ± 0.90	0.40	4.30
		<i>wind direction</i>	23.90 ± 6.63	4.00	30.00

\*: 유의수준 0.01에서 3개의 군집간 분산분석시 유의한 변수

<표2.7> 주성분점수를 이용한 군집분석에 따른 오염물질, 기상변수 기술통계량

군집	관측치(일)수	오염물질, 기상변수	평균 ± 표준편차	최소값	최대값
1	254	CO*	25.02 ± 7.93	10.75	53.00
		NO <sub>2</sub> *	52.17 ± 13.43	23.50	91.75
		O <sub>3</sub> *	34.87 ± 11.88	14.50	70.25
		PM10*	108.55 ± 39.49	24.25	322.25
		SO <sub>2</sub> *	14.34 ± 4.07	7.25	28.75
		TSP*	64.79 ± 23.77	21.50	136.25
		temperature*	13.92 ± 10.76	-9.30	29.90
		precipitation	16.31 ± 35.69	0	261.60
		humidity	65.35 ± 13.73	29.00	91.30
		wind wpeed*	2.02 ± 1.09	0.20	5.30
		wind direction	20.79 ± 10.32	1.00	35.00
2	63	CO	41.60 ± 9.56	23.50	73.00
		NO <sub>2</sub>	83.50 ± 13.67	60.50	120.00
		O <sub>3</sub>	23.02 ± 7.01	9.25	41.50
		PM10	167.12 ± 42.71	73.75	279.25
		SO <sub>2</sub>	24.63 ± 5.17	16.00	46.25
		TSP	120.43 ± 32.03	68.50	219.50
		temperature	5.68 ± 5.21	-7.20	18.40
		precipitation	6.24 ± 12.39	0	42.20
		humidity	63.62 ± 10.95	37.60	81.40
		wind wpeed	1.43 ± 0.76	0.10	4.00
		wind direction	21.44 ± 11.83	1.00	35.00
3	48	CO	25.64 ± 6.38	13.25	42.75
		NO <sub>2</sub>	70.43 ± 13.40	44.00	97.50
		O <sub>3</sub>	47.26 ± 12.46	21.25	82.00
		PM10	216.38 ± 60.64	143.75	403.75
		SO <sub>2</sub>	17.51 ± 4.35	9.25	30.75
		TSP	148.62 ± 42.20	74.75	287.00
		temperature	19.66 ± 5.96	3.80	29.50
		precipitation	4.50 ± 11.65	0	39.50
		humidity	66.53 ± 10.21	45.90	84.30
		wind wpeed	1.70 ± 0.40	0.40	4.30
		wind direction	23.60 ± 4.00	4.00	30.00

\*: 유의수준 0.01에서 3개의 군집간 분산분석시 유의한 변수