

Pattern Recognition of Ozone Pollution for Predicting High Level Ozone

허정숙·김동술

경희대학교 환경학과 대기오염연구실 및 환경연구소

1. 서 론

오존농도의 증가원인은 인위적 배출물질이 증가하기 때문이며, 기온이 높고, 일사량이 풍부하며, 낮은 상대습도, 저풍속에서 오존전구물질(precursor)의 축적이 있을 때, 고농도 오존이 생성된다고 보고되고 있다 (Lazutin, 1996). 도심의 풍하쪽(down wind)에 위치한 교외지역에서 높은 오존농도가 관측된 것은 도심에서 고농도 오존의 이동과 오존전조물질이 이동하기 때문인 것으로 보고된 바 있다 (Ludwig, 1995). 오존은 전구물질 농도와 일사량, 풍향, 온도 등의 기상변수에 의해 생성되는 2차 오염물질로서 지역의 환경요인에 의하여 생성 정도가 다르다(Vukovich, 1995). 따라서, 장기간의 오존 및 1차오염물질 농도추이와 기상변수 등을 지역적 특징으로 파악하는 연구들이 진행되고 있다.

본 연구에서는 고농도 오존 추정을 위한 첫 번째 작업으로 오존생성과 관련한 여러 인자들을 대상으로 지역별로 단변수 분석을 수행한 바 있다(허정숙과 김동술, 1997). 다음 단계로 군집분석(cluster analysis)을 통해 오존오염자료를 유사한 특성을 지닌 군집별로 분류하고, 분산 주성분분석(disjoint principal component analysis)을 응용한 형태인식법(pattern recognition)을 이용하여 분류된 군집을 확률적으로 검증한 뒤, 최종적으로 산출된 순수군집(homogeneous cluster)의 특징 및 오존오염의 패턴을 분류하고자 하였다.

2. 연구방법

본 연구는 환경부에서 입수한 1989년부터 1996년까지 8년간 대기오염자동측정망의 시간별 자료(SO₂, TSP, CO, NO₂, O₃)와 기상청의 지상기상자료(풍향, 풍속, 일사량, 온도, 습도) 및 오산 상층기상자료(500hpa 고도면의 온도, 풍속, 풍향) 등 총 13개 변수를 이용하였다. 또한, 지상기상자료중 온도와 습도, 상층기상자료는 내삽하여 이용하였고, 대상자료는 13개 변수 모두가 갖추어진 자료만을 선택하였다. 대상지역은 광화문지역을 중심으로 하였다.

본 연구의 목적이 고농도 오존을 추정하는 것이므로, 첫 번째 단계로 오존농도 80 ppb (미국 EPA의 National Ambient Air Quality Standards 8시간 기준농도)와 100 ppb (우리나라의 1시간 기준농도)를 기준으로 두개 집단(오존농도 80~99 ppb 와 100 ppb 이상인 집단)으로 자료를 분류하였다. 두번째 단계로 상층풍향을 기준으로 자료를 분리하는 방법을 시도하였다. 풍향은 360° 범위로 다른 인근 각도에서 동일한 값으로 정량화되기 때문에, 180° 범위로 자료를 분리하였다. 풍향 고려시, 앞에서 분류된 두개 집단별로 고농도 오존이 나타난 날과 전일의 풍향 빈도수를 분석한 결과, 풍향 5°~175° 계열이 전체 5% 정도였고, 180°~360° 계열이 95% 정도였다. 따라서, 고농도 오존이 나타난 날과 전일의 상층풍향은 크게 서풍(180°~360°)과 동풍(5°~175°)계열로 분류할 수 있었다. 풍향은 식 (1)에 적용하여 정량화시켰으며, 풍향의 정량화된 값은 1~3사이 값으로 N=1, W(270°)=2, E=2, S=3이다. Ψ 는 각도별 radian값이다.

$$WD = 2 + \sin(\Psi - \pi/2) \quad (1)$$

세 번째 단계로 서풍과 동풍계열로 분류된 집단에 대해 군집분석과 분산 주성분분석법을 이용하여 고농도 오존오염의 패턴별 분류를 시도하였다. 군집분석은 자연과학분야에서 폭넓게 응용되는 다변량통계 분석법(multivariate statistical analysis)으로 object에서 유사도(similarity)를 기준으로, 동질집단을 유도 및 분류하는 방법이다. 군집분석법은 연구과정에서 자료의 형태 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. 일단 군집분석을 통해 유사한

특성을 가진 object들이 군집(class)으로 분류되면, 그 분류가 올바르게 되었는지를 정량적으로 분류해야 한다. 이를 위해 분산주성분분석을 응용한 형태인식법을 적용하였고, 이를 위해 SIMCA (Soft Independent Modeling of Class Analogy) package를 이용하였다. SIMCA 분석시 이용되는 자료는 traing set과 test set으로 구성되는데, object에 대한 기지의 정보 혹은 군집분석과 같은 통계방법으로 임의 분류될 수 있을 때, 이 군집(class)을 training set이라 하고, 소속이 불확실한 대상자료들의 모임을 test set이라 한다. SIMCA에 의한 자료 분석은 두 단계로 나누어 질 수 있는데, 첫단계는 training set 을 이용하여 주성분모델을 개발하는 것이고, 두 번째 단계는 test set 중의 각 대상자료를 이미 개발된 각 각의 군집별 주성분모델과 비교해 소속감을 부여하는 것이다. 모델링 이후 어느 군집에도 속하지 않는 object를 이상치(outlier)라고 한다.

통계분석시 자료의 변환은 중요하다. 만약, 한 변수의 변량분포가 대칭형이 아니고 다른 변수에 비해 한 쪽으로 치우쳐 있다면, 그 결과는 과장되거나 오류를 범할 수 있으므로, 적당한 변환을 수행하여 그 변수의 과잉 영향을 줄일 수 있다. 따라서, 본 연구에서는 군집분석에 앞서 총 13개 변수 각각에 대해 자료변환을 거치지 않은 자료, 로그(logarithmic)변환을 거친 자료와 제곱근(square root)변환을 거친 자료를 입력자료로 하였으며, 각 각의 입력자료는 z-score를 이용하여 표준화시킨 후 군집분석을 수행하였다. 군집분석시는 비교적 계산이 간편한 응집위계분석법을 적용하였으며, 비유사도 기준으로 유클리디안과 유클리디안 제곱거리 방법을 이용하였고, 군집화 기준으로 평균결합법(average linkage)과 Ward방법을 이용하였다. 군집분석을 수행한 후, 분류된 군집에 대해 분산 주성분분석을 통해 군집별 주성분의 수를 결정하였고, 유효한 주성분의 수가 결정된 분석 결과만을 선택하였다. 선택된 결과에 대해 Cooman 표준결정도를 이용하여 object들의 군집소속여부를 확률적으로 검증하였다. 이렇게 결정된 군집은 군집 분석시 포함되었던 주관적인 판단이 배제된 순수한 군집이며, 분류가 잘못된 object를 이상치로 정의할 수 있다. 이상치 개수가 최소로 결정될 때를 가장 이상적으로 분류가 잘 이루어진 것으로 판정할 수 있으며, 이 때 해당 변수의 자료변환이 가장 적합한 것으로 결정하고, 순차적으로 다음 단계 변수에 대해 위의 작업들을 반복적으로 수행하였다.

3. 결과 및 고찰

광화문 지역의 오존농도 100 ppb 이상 측정된 집단을 대상으로 13가지 변수 각각을 순차적으로 자료 변환하고, 군집분석과 주성분분석을 수행한 결과, SO₂와 NO₂는 제곱근 변환, TSP와 지상풍속은 로그변환, CO, O₃, 지상기상자료 (온도, 일사량, 풍향, 습도), 상층기상자료 (온도, 풍속, 풍향)는 자료변환을 시키지 않았을 때 가장 이상적으로 오존오염 형태 분류가 양호하였다.

표 1은 12가지 변수에 대해 순수한 군집을 결정하는 분류 작업을 수행하였고, 마지막 단계로 상층 풍향 변수를 고려하였을 때, 자료변환, 비유사도 기준, 군집화 방법에 따라 결정된 순수한 군집의 object수를 백분율로 표시한 것이다. 상층풍향 자료를 변환시키지않았을 때, 로그변환, 제곱근 변환을 시켰을 때, 유클리디안 거리와 평균결합법으로 군집화 시켰을 때, 이상치 개수가 가장 적은 것으로 분석되었다. 이 경우 군집분석시 임의적으로 분류된 군집 수는 4개였으며, 78개 object 중 5개 object는 어떤 군집에도 소속되지 않은 것으로 분류되었다. 군집분석 후 분류된 군집에 대해 SIMCA package를 이용하여 분산 주성분분석을 한 결과 주성분 수가 모든 군집에서 1개 이상으로 나타났으나, 군집 2는 유효 주성분이 결정되지 않았다. 그림 1과 2는 군집분석 결과 분류된 군집에 대해 Cooman 표준결정도를 이용하여 군집간 비교를 한 것이다. 그림에서 군집 1~4는 군집분석후 분류된 object들이며, 군집 99는 어떤 군집에도 소속되지않은 object이다. 그림에서 수직과 수평으로 그은 선은 95 % 확률의 임계거리선이고, 이 선에 의해 4개 영역으로 나누어진다. 이 영역 중 왼쪽 아래영역은 두 군집이 공존하는 영역이고, 오른쪽 위 영역은 두 군집에 속하지 않는 다른 군집의 영역이다. 따라서, 왼쪽 위, 오른쪽 아래 영역에 속할 때만이 95% 확률에서 해당하는 군집에 속하는 것이고, 그외 영역에 속할 경우, 군집에서 배제된다. 이 군집간 비교를 통해 각 군집 모델에 대한 새로운 군집이 형성되었고, 이 새로운 군집은 군집분석시 포함되었던 주관적인 판단이 배제된 순수한 군집이라 할 수 있다. 표 1에서 최적의 분류결과, 군집1은 85.7%(36/42), 군집2는 73.3%(11/15), 군집 3은 90.9%(10/11), 군집 4는 100%(5/5) 순수한 군집이 결정되

있으며, 분류가 잘못된 object는 전체 78개 중 16개 (20.5%) 이상치로 결정되었다.

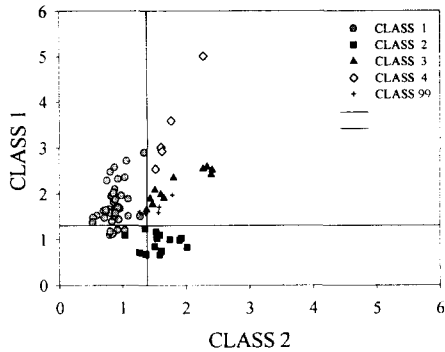


Fig. 1. Comparison between class 1 and 2.

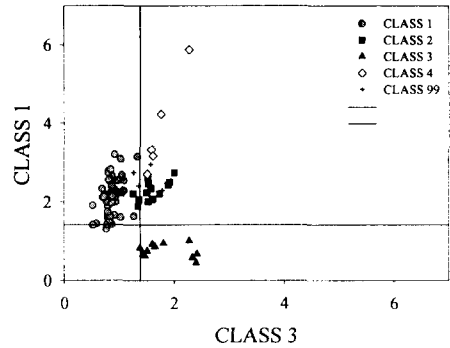


Fig. 2. Comparison between class 1 and 3.

Table 1. Fraction(%) and number of objects after deleting outliers in each class.

Dtata	Sqrt(SO ₂), Ln(TSP), CO, Sqrt(NO ₂), O ₃ , Ln(WS), TEMP, SOLA, WD, HUM, UWS, UTEM											
	No Transformation (UWD)				Log Transformation Ln(UWD)				Square Root Transformation Sqrt(UWD)			
	ED		SED		ED		SED		ED		SED	
Class	Ward	Avg.	Ward	Avg.	Ward	Avg.	Ward	Avg.	Ward	Avg.	Ward	Avg.
1	60.6 (20/33)	85.7 (36/42)	57.1 (24/42)	85.7 (36/42)	34.5 (10/29)	85.7 (36/42)	42.1 (16/38)	81.6 (40/49)	60.6 (20/33)	85.7 (36/42)	54.8 (23/42)	85.7 (36/42)
2	75.0 (6/8)	73.3 (11/15)	81.8 (9/11)	73.3 (11/15)	75.0 (6/8)	73.3 (11/15)	76.9 (10/13)	80.0 (12/15)	75.0 (6/8)	73.3 (11/15)	63.6 (7/11)	73.3 (11/15)
3	90.9 (10/11)	90.9 (10/11)	61.3 (8/13)	90.9 (10/11)	72.7 (8/11)	90.9 (10/11)	90.9 (10/11)	80.0 (4/5)	90.9 (10/11)	90.9 (10/11)	53.8 (7/13)	90.9 (10/11)
4	78.6 (11/14)	100 (5/5)	75 (9/12)	100 (5/5)	50.0 (7/14)	100 (5/5)	87.5 (14/16)		78.6 (11/14)	100 (5/5)	91.7 (11/12)	100 (5/5)
5	100 (12/12)				93.8 (15/16)				100 (12/12)			90.9 (10/11)
Outlier	24.4 (19/78)	20.5 (16/78)	35.9 (28/78)	20.5 (16/78)	41.0 (32/78)	20.5 (16/78)	35.9 (28/78)	28.2 (22/78)	24.4 (19/78)	20.5 (16/78)	38.5 (30/78)	20.5 (16/78)

* WS: Wind Speed, TEMP: Temperature, SOLA: Solarity, WD: Wind direction, HUM: Humidity, UWS: 500hpa Wind Speed, UTEM: 500hpa Temperature, UWD: 500hpa Wind Direction

참 고 문 헌

- 허정숙과 김동술 (1997) 고농도 오존의 추정을 위한 예비연구, 한국대기보전학회 추계학술대회, 동신대, 91-92pp.
- Lazutin, L., P.C. Bezerra, M.A. Fagani, H.S. Pinto, I.M. Martin, E.L.P. Da Silva, M.G. DaSilva Mello, A. Turtelli Jr, V. Zhavkov, and J. Zullo Jr (1996) Surface ozone study in Campinas, Sao Paub, Brazil, Atmos. Env., 30(15), 2729-2738.
- Ludwig, F.L., J.Y. Jiang, and T. Chen (1995) Classification of ozone and weather patterns associated with high ozone concentrations in the San Francisco and Monterey Bay Areas, Atmos. Env., 29(21), 2915-2928.
- Vukovich, F.M. (1995) Regional-scale boundary layer ozone variations in the Eastern United States and their association with meteorological variations, Atmospheric Environment, 29(17), 2259-2273.