

다변량 통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구

Prediction of High Level Ozone Concentration in Seoul by Using Multivariate Statistical Analyses.

허 정 숙 · 김 동 술

경희대학교 환경학과 대기오염연구실
(원고접수 : 1993. 7. 2)

Jeong-Suk Heo and Dong-Sool Kim

Air Pollution Lab.,
Dept. of Environmental Science,
Kyung Hee University-Suwon Campus
(Received 2 July 1993)

Abstract

In order to statistically predict O₃ levels in Seoul, the study used the TMS (telemetered air monitoring system) data from the Department of Environment, which have monitored at 20 sites in 1989 and 1990. Each data in each site was characterized by 6 major criteria pollutants (SO₂, TSP, CO, NO₂, THC, and O₃) and 2 meteorological parameters, such as wind speed and wind direction. To select proper variables and to determine each pollutant's behavior, univariate statistical analyses were extensively studied in the beginning, and then various applied statistical techniques like cluster analysis, regression analysis, and expert system have been intensively examined. For the initial study of high level O₃ prediction, the raw data set in each site was separated into 2 groups based on 60 ppb O₃ level. A hierarchical cluster analysis was applied to classify the group of data sets having over 60 ppb O₃ into small classes. Each class in each site has its own pattern. Next, multiple regression for each class was repeatedly applied to determine an O₃ prediction submodel and to determine outliers in each class based on a certain level of standardized residual. Thus, a prediction submodel for each homogeneous class could be obtained. The study was extended to model O₃ prediction for both on-time basis and 1-hr after basis. Finally, an expert system was used to build a unified classification rule based on examples of the homogeneous classes for all of sites. Thus, a concept of high level O₃ prediction model was developed for one of O₃ alert systems.

1. 서 론

공업지역과 도시에서 배출되는 주요 대기오염물질은 NO_x, SO_x, HC, TSP, CO 등이며, 이들 대기오염물질은 광화학적 산화물의 생성에 관여한다. 대표적 광화학적 산화물인 오존은 대도시에서 스모그

가 발생할 때 자주 관측된다. 대류권 내 오존은 전형적인 인위적 오염물질로서, 불안정하여 활성이 크며 분포가 광범위하여, 성층권 오존과는 달리 유해성이 크다. 일반적으로 대류권 내의 오존은 1) 성층권으로부터의 유입과 대류권 공기와의 혼합(Pou-lida, 1991; Reiter, 1991), 2) 고농도 오존과 광화

학적 전조물질 (precursor)이 오염된 지역에서 운송 (Burkhart, 1990; Edwards, 1991; McKendry, 1993), 3) 광화학적 생성, 4) 광화학적 파괴, 5) 습식 및 건식침적 (wet and dry deposition)등과 같은 생성·소멸과정의 복합산물이다. 최근, 이와 같은 오존의 생성과 소멸과정을 규명하여 제어전략을 세우는 연구들이 진행되고 있다.

오존제어전략과 더불어 기준치 이상의 오존노출을 예방하기 위한 방안으로써, 오존농도의 예측은 인간과 생물체에 미치는 영향을 최소화할 수 있다는 차원에서 중요하다. 이미, 미국 및 일본 등은 대기오염이 일정수준을 초과할 때, 조기경보를 발하는 대기오염 경보체계를 실시하고 있다. 오존농도 예측의 연구로, 박옥현 (1989)은 오존오염의 자기상관성과 기상특성을 고려하여, 고농도 오존을 예측하기 위해 판별분석 (discriminant analysis)을 적용한 바 있으며, Feister (1991)는 비선형 회귀분석을 이용하여 313개의 기상변수와 오존과의 관계를 정량화함으로써 오존농도를 예측한 바 있다. 또한, 시계열 분석을 이용한 오존농도 예측에 관한 연구도 있었다 (Kelly, 1986). 판별분석을 이용한 경우는 기준농도 이상의 오존발생 여부만이 판단되며, 시계열 분석에서는 변수선정에 제약점이 있었다.

따라서, 본 연구에서는 대규모 자료처리와 변수선정이 비교적 자유로운 다변량통계기법 중 군집분석 (cluster analysis)을 응용하고, 오염물질과 기상조건을 변수로 선정하여, 오존의 농도변화를 조사하였다. 또한, 군집분석 후 분류된 소집단을 대상으로 다중회귀분석을 응용하였다. 다중회귀분석은 종속변수인 오존을 가장 잘 설명할 수 있는 독립변수 (오염물질, 기상조건)를 확인할 수 있다. 따라서, 독립변수들의 함수로 표현된 모델회귀식을 구하여 오존농도를 예측할 수 있다. 본 연구의 마지막에는, 통계분석결과 분류된 소집단을 사례로 이용하여 전문가 시스템 (expert system)을 개발하여, 전체 대상 지역의 오존농도를 예측하는 종합적 오존예측모델을 개발하였다. 본 연구를 위해 1989년과 1990년 환경처 및 서울시의 대기오염자동측정망 중 서울시 20개소 자료를 대상으로 하였다. 오존과 관련하여 기타 대기오염물질과 기상조건을 단변수분석하였으며, 다각적 응용통계 기법을 검토하여 오존농도 예측모델을 개발하였다. 이는 오존조기 경보체계 개발을 위한 기초연구로써 유용하리라 사료된다.

2. 연구방법과 진행과정

2. 1 군집분석과정

군집분석은 자연과학 분야에서 폭넓게 응용되는

다변량 통계분석법으로 개체 (case)에서 유사도 (similarity)를 기준으로, 동질집단 (homogeneous class)을 유도 및 분류하는 방법이다. 군집분석은 집단간 분산 (between-group variance)을 최대화시키며, 집단내부의 분산 (within-group-variance)을 최소화시키는데 목적이 있다 (Dorling, 1992). 군집 분석을 응용하여 사전에 어떤 정보없이, 혹은 최소의 알고 있는 정보로 유용한 집단을 분류 및 확인할 수 있고, 대규모 자료를 처리하는데 유용하며, 많은 자료와 자료간의 관계를 설명할 수 있다. 기본 원리는 대상자료들의 유사도, 즉 특성통계치가 어느 정도 비슷한 값을 갖는지를 거리로서 환산하고, 이 거리를 기준으로 인접한 대상들은 동일한 집단으로 분류하며, 수상도 (dendrogram)라는 그림으로 표현한다. 군집분석에서 가장 큰 난제는 수상도 상에서 유사도를 주관적으로 결정하고 집단을 분류하는 과정이다. 본 연구는 군집분석법을 단지 변수확대의 도구로써 이용하였으며, 집단 상호간의 유의성을 기준으로 집단의 유의성을 향상시키는 직접적인 민감도 분석 (sensitivity analysis)를 채택하기 보다는, 다음 단계에 연결되는 회귀분석의 결과를 응용하여 반복적으로 이상치를 제거하고 각 소집단의 유사도를 향상시키는 간접적 방안을 시도하였다.

본 연구에서 군집분석법을 응용하기 위해, 우선, 입력된 자료를 정규화시켰으며, 표준편차는 개체 (case)간의 관련성을 계산하는 Q-기법을 이용하였다. 다음으로, 정규화된 변수값들의 유사도 측정은 두 개체간의 변수값의 거리제곱의 합에 제곱근을 취한 값 즉, Euclidian거리법을 이용하였다. 이 측정법은 다차원공간에서 최단 직선거리를 의미하는 것으로 각종 환경자료에 일반적으로 사용되는 거리법이다 (Hopke, 1985). 이 거리법으로 계산된 각 개체간의 평균거리 벡터를 이용하여 군집분석을 시도하였다. 분석방법은 위계분석법 중 응집위계분석법 (agglomerative hierarchical cluster analysis)을 이용하였으며, 군집화 방식은 합쳐진 집단간의 평균거리를 최소화하고, 상관관계를 최대화하는 평균결합법을 적용하였다. 이 결합법은 Lance & William (1967)에 의해, group average method라고 명명된 바 있다.

2. 2 연구과정

환경처에서 입수한 1989, 1990년 2년간의 서울시 20개소 자동측정망의 자료를 이용하였다. 본연구의 진행과정은 그림 1과 같다. 우선, 오존을 예측하기 위한 변수선정을 단변수분석을 통해 수행하였으며, 이를 위해 선정된 변수 (SO_2 , TSP, CO, NO_2 , O_3 , THC, 풍속, 풍향)가 모두 존재하는 자료 즉 개체

(case)만을 선택하였다. 이들 선택된 자료는 오존농도 60 ppb (한국 8시간 기준, 일본 1시간 기준)를 기준으로, 저농도 집단 (60 ppb 미만)과 고농도 집단 (60 ppb 이상)의 두 집단으로 분류하였다. 또한, 저농도 집단의 경우, 개인용 컴퓨터의 연산능력의 한계성과 과다한 개체수가 존재하기 때문에, 0~10 ppb, 10~30 ppb, 30~60 ppb의 세개의 소집단으로 나누어 군집분석을 수행하였으며, 고농도 집단은 개체의 수가 작아 집단분류는 시도하지 않았다.

이미, 오존의 특정농도를 기준으로 두개 집단으로 분류하였기 때문에, 군집분석의 변수로는 SO₂, TSP, CO, NO₂, THC와 풍속만을 선정하였다. 풍향은 정량화의 어려움이 있었으므로 변수에서 제외시켰다. 본 연구에서는 일단 군집분석을 통하여 한 측정지역에서 4~5개 소집단으로 인위적인 분류를 시도하였고, 각 소집단에 소속되지 않는 개체들은 이상치 (outlier) 집단으로 분류하였다. 저농도 집단은 군집분석 후 정성분석만을 하였고, 고농도 집단은 정성분석과 다중회귀분석을 병행하여 본 연구의 핵심으로 삼았다.

다중회귀분석을 이용한 오존농도의 추정과정은 다음과 같은 2단계로 수행되었다. 우선, 현시점의 오존농도를 오존을 제외한 주요오염물질과 풍속만을 변수로 이용하여 추정하였고, 다음, 1시간 후의 오존농도를 현시점의 모든 주요오염물질과 풍속자료를 이용하여 추정하였다. 이를 위해, 회귀분석과정에서는 첫번째, 대상지역에서 군집분석 전과 후의 통계적 의미를 조사하였으며, 두번째, 주요오염물질과 풍속이 오존과 단순 1차 함수관계가 아니라고 가정하고, 독립변수를 자료변환시킨 후 다중회귀분석을 하였다. 세번째, 회귀분석의 유의성을 높이고, 군집분석 후 유사도의 주관성이 내재되어 있는 소집단내에서 이상치를 제거하기 위해, 다중회귀식의 표준잔차 (standardized residual)를 이용하였다. 표준잔차란 다중회귀식내에서 개별잔차를 잔차의 표준편차 제곱으로 나눈 값이다. 일단, 표준잔차가 가장 큰 개체를 제거하고, R²이 0.85가 될 때까지 개체 제거작업을 반복하였다. 따라서, 유의성이 크게 향상된 최적모델이 개발되었다. 네번째, 군집분석 후 제거된 이상치와 회귀분석 과정에서 제거된 이상치를 제1 이상치 집단으로 분류하였고, 이상치 집단에서도 최종적으로 이상치로 판정된 자료는 회귀분석에 적용할 수 없는 자료로써, 제2 이상치 집단으로 분류하였다. 마지막으로, 총 20개의 측정장소 중 본 연구에 표본적으로 이용된 3개 측정장소에 대해, 이상치가 제거된 모든 동질집단과 이상치 집단들을 사례로 이용하여, 전문가 시스템을 개발하였다.

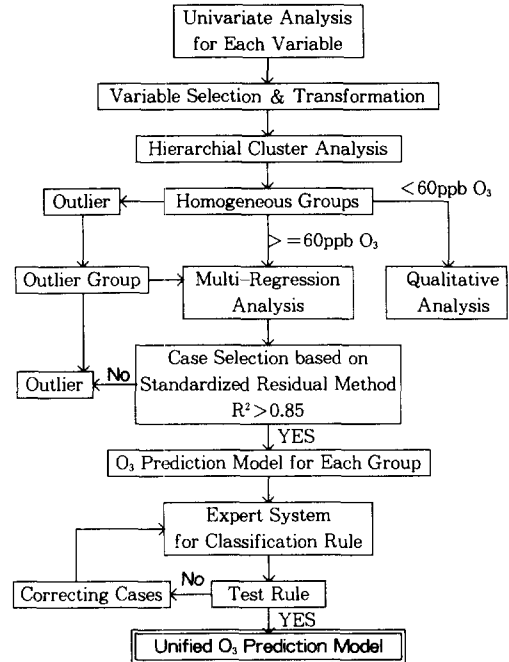


Fig. 1. Flow chart for developing O₃ prediction model.

3. 결과 및 논의

3. 1 주요 대기오염물질의 시간별, 계절별 특징

1989, 1990년 2년간의 SO₂, TSP, CO, NO₂, O₃, THC농도의 시간별 변화를 계절별로 조사하였다. 5가지 대기오염물질의 변화를 오존과 관련하여 분석하고, 오존생성에 영향을 미치는 오염물질의 거동을 파악하였다. 계절별 분석에서 봄은 3,4,5월 평균, 여름은 6,7,8월 평균, 가을은 9,10,11월 평균, 겨울은 12,1,2월 평균을 의미한다.

SO₂는 그림 2와 같이, 오전 9시까지 상승하다가, 오후 12시에서 6시 사이에 하향추세를 보인 후, 오후 6시 이후에 다시 상승하는 경향을 보였다. 계절별 변화는 화석연료를 많이 사용하는 겨울철에 높은 농도가, 여름철에 가장 낮은 농도가 나타났다. TSP는 그림 3과 같이, 오전 7~10시까지 상승하다가, 낮동안 감소한 후 오후 5시 이후 재상승하는 경향으로 조사되었다. 계절별로는 겨울철 농도가 가장 높았고, 습식침적이 많은 여름이 가장 낮았다. 고농도 TSP는 빛을 차단하여 일사량을 감소시키므로 오존생성을 억제할 것으로 사료된다. CO는 그림 4와 같이, 오후 4시에서 5시 사이에 최저농도를 보였으며, 이른 오전 중에 고

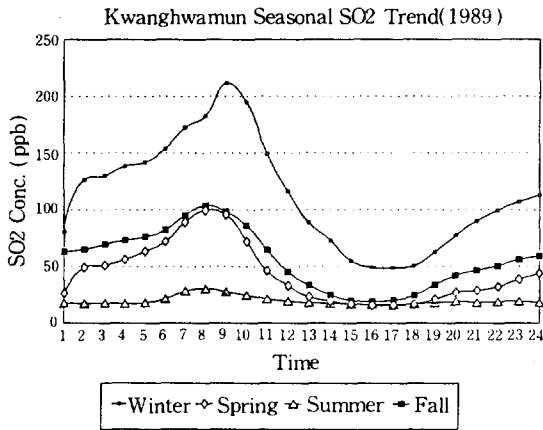


Fig. 2. Diurnal and seasonal variations of SO₂ at Seoul in 1989.

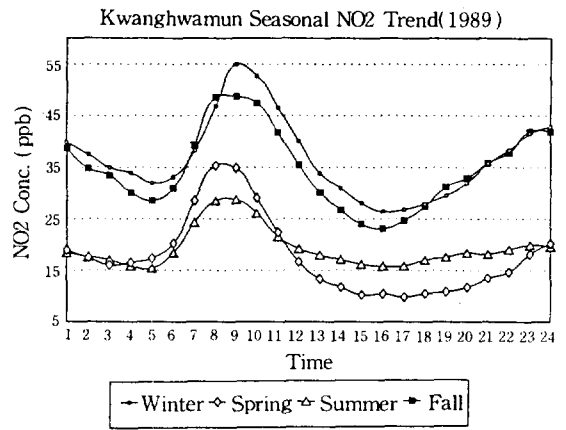


Fig. 5. Diurnal and seasonal variations of NO₂ at Seoul in 1989.

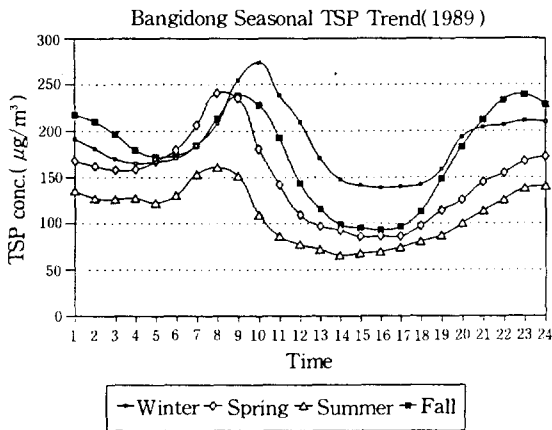


Fig. 3. Diurnal and seasonal variations of TSP at Seoul in 1989.

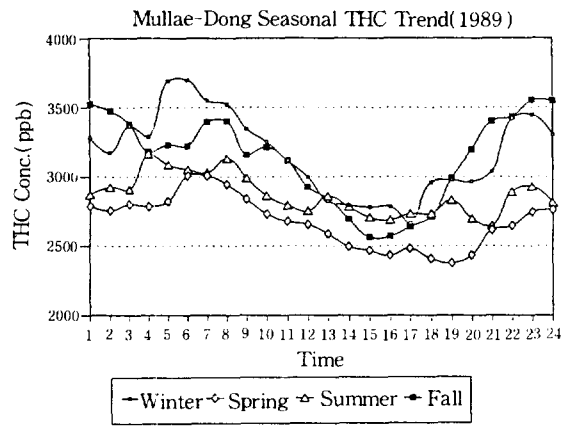


Fig. 6. Diurnal and seasonal variations of THC at Seoul in 1989.

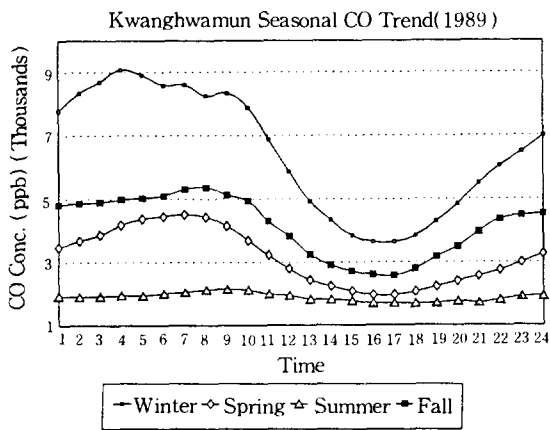


Fig. 4. Diurnal and seasonal variations of CO at Seoul in 1989.

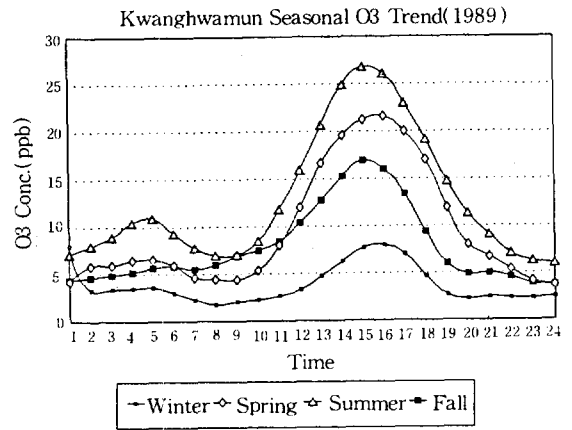


Fig. 7. Diurnal and seasonal variations of O₃ at Seoul in 1989.

농도가 나타났다. 계절별로는 겨울철이 가장 높았고, 가을, 봄, 여름 순서로 농도가 낮았다. CO는 OH기 (radical)에 의해 산화되어 상실되면서 NO 농도에 따라 O₃를 생성한다. 여름철은 OH기 생성에 유리한 조건 (높은 습도, 높은 오존농도, 일사량)이므로 (Poulida, 1991), CO농도는 여름에 낮고, 주간에 감소된 것으로 사료된다. NO₂와 THC는 오존생성에 가장 큰 영향을 미치는 오염물질로 알려져 있다. 그림 5와 같이, NO₂는 오전 8~9시 사이에 고농도가 나타났고, 낮동안에 감소하다가 오후 5시 이후 천천히 상승하다가 이른 새벽에 다시 감소하는 경향을 보였다. THC는 그림 6과 같이 가을과 겨울에는 오전 11시까지 상승하다가 오후 12~7시 사이에 하향한 후 8시 이후부터 재상승하는 경향을 띄었다. 여름과 봄철에는 시간에 따른 변화가 뚜렷하지 않았으나, 대개 오전에 농도가 높았다. 계절별 농도추이는 겨울이 가장 높았으며, 가을, 여름, 봄 순으로 농도가 높았다.

오존의 농도경향은 그림 7과 같이, 오후 3~4시 사이에 최대농도를 보였으며, 일몰 후에서 일출전까지는 저농도가 나타났다. 계절별로는 여름철에 가장 높은 농도가 나타났고, 겨울철에 가장 낮았는데, 이것은 겨울철의 적은 일사량과 한랭한 대기조건이 2차오염물질인 오존의 생성을 방해하기 때문이다. 다른 주요오염물질과 대조적인 양상으로 오존이 변화한다는 것은 높은 온도, 낮은 상대습도, 저풍속에서 오전 중에 오존전조물질의 축적이 있을 때 오존이 높게 나타난다는 보고와 일치한다 (Angle, 1991; Feister, 1991; Korsog, 1991; Poulida, 1991; Lefohn, 1992; Wolf, 1992).

3. 2 군집분석의 결과

군집분석 후 분류된 소집단의 특징을 조사하는 방법으로 우리나라 환경기준치 (1993년도 개정안)를 이용하였다. 즉, 각 소집단의 5가지 오염물질의 평균농도를 1시간 환경기준치 SO₂ 30 ppb, TSP 150 µg/m³, NO₂ 50 ppb, THC 3 ppm과 CO 8 시간 환경기준치 9 ppm으로 나눈 값으로 각 소집단별로 오염물질의 거동을 분석하였다. 오존농도가 60 ppb 미만인 저농도 분석의 경우, 단변수분석에서 오존농도가 가장 높은 방이동과 인근 지역인 성수동, 구의동, 신설동 지역을 집중적으로 연구하였다. 대체로 9~10가지의 특징을 갖는 소집단으로 분류되었는데, 각 소집단에서 높은 농도를 나타낸 오염물질은 SO₂, TSP, THC였다. 오존농도 0~10 ppb, 10~30 ppb의 경우, 가장 큰 농도변화를 나타낸 오염물질은 SO₂였고, 30~60 ppb에서는

TSP가 큰 기복을 나타냈으며, 가장 작은 기복을 보인 것은 CO였다. 이와 같이, 사전에 정보가 없었던 자료를 군집분석하므로써, 저농도 오존이 나타난 시점의 각 변수의 특성을 정성적으로 조사할 수 있었다.

오존농도가 60 ppb 이상인 고농도 분석의 경우, 구로동, 구의동, 남가좌동, 대치동, 방이동, 쌍문동, 오류동, 성수동을 대상으로 오존의 에피소드 출몰일수의 분석, 고농도 발생빈도, 주풍향의 분석 등 다각적인 통계분석을 시도하였다. 이 경우도 대기환경기준치를 중심으로 특징을 조사하여, 12가지 경향을 보이는 소집단으로 분류하였고, 저농도 분석때와는 달리, 각 오염물질이 큰 변화없는 평탄한 분포를 보였다. 이와 같이 각 오염물질의 특징을 조사하는 과정은 오존예측모델 개발시 고농도 오존에 영향을 미치는 인자 (factor) 파악의 준비작업으로 수행되었다. 정성분석의 결과는 문헌 (허정숙, 1993)을 참고하라.

3. 3 다중회귀분석의 결과

다중회귀분석은 구의동, 방이동 및 성수동을 대상으로 하였다. 현시점의 오존농도를 추정할 때, 대상지역에서 군집분석 이전과 군집분석 후 분류된 소집단을 회귀분석한 결과, 군집분석 후 분류된 소집단을 회귀분석한 모델이 높은 다중결정계수 R² 즉, 높은 유의성을 보였다. 군집분석 이전에는 회귀모델의 R²이 0.24~0.56이었지만 소집단의 경우 0.84~0.998로 매우 높게 나타났다. 이러한 결과는 분류된 소집단의 오존 추정식이 군집분석의 분류 이전의 회귀모델보다 더 큰 통계적 의미를 가짐을 뜻한다.

Wolff (1992)는 오존생성과정이 비선형적이라는 가정하에, VOC와 NO_x비율을 이용하여 오존소멸과정을 연구한 바 있으며, Korsog (1991)는 오존농도 변화에 대한 기상조건에 기여도를 평가하는데 대수선형 회귀식 모델을 개발한 바 있다. 본 연구에서도 대상지역의 오존농도를 예측하기 위해, 5가지 오염물질 (SO₂, TSP, CO, NO₂, THC)과 풍속 (WS) 등을 변수로 선정하고, 이들과 오존농도 간을 비선형적으로 자료를 변환한 후, SPSS를 이용하여 단계별 회귀분석 (stepwise regression)을 수행한 결과, 유의성이 개선된 추정식이 개발되었다. 즉, 구의동의 집단3은 0.998, 방이동의 집단3은 0.93, 성수동의 집단1은 0.92로 R²이 증가하였다. 또한, 각 소집단의 모델식에서 각 독립변수의 오존예측에 대한 결정계수 (기여도 : coefficient of determination)를 표 1에 제시하였다. 구

Table 1. Contribution of independent variables for the on-time O₃ prediction model.

Pollutants		SO ₂	TSP	CO	NO ₂	THC	W.S	Total%
Kuui-Dong	Class1	11.0	14.2	5.3	30.0	24.3	7.1	91.9
	Class2	6.7	7.5	3.7	62.7	—	15.1	95.6
	Class3	18.4	18.6	11.2	46.3	3.8	1.5	99.8
Bangi-Dong	Class1	4.2	21.1	13.5	16.7	13.9	17.9	87.3
	Class2	0.2	34.0	44.2	3.7	2.5	1.6	86.2
	Class3	6.7	7.8	17.0	6.5	27.0	28.3	93.3
Sungsu-Dong	Class1	0.1	0.9	8.6	8.0	0.7	74.2	92.4
	Class2	4.6	29.2	5.6	2.8	27.7	19.2	89.1
	Class3	12.6	30.6	12.0	1.6	—	28.4	85.2
	Class4	7.3	7.3	25.0	3.6	41.9	2.8	87.8
	Class5	0.2	16.8	54.3	8.7	2.5	1.7	84.0
	Class6	6.6	6.9	24.4	23.3	13.7	16.2	91.2

Table 2. Model equations for predicting O₃ levels by 1-hr after basis.

Site	Class	Prediction Equations	R ²
Kuui-Dong	Class1	$[O_3]_t = -73.5 - 0.048[SO_2]^2 + 0.021[TSP] - 2.68[CO] - 2.89[NO_2]^{1/2} + 1.6[O_3] - 1.5 \times 10^{-11} \exp[THC] + 46.6hr[W.S]$	0.91
	Class2	$[O_3]_t = 119.03 + 1.543 \times 10^{-5} \exp[SO_2] - 4.69[TSP]^{1/2} + 0.012[CO]^2 + 1.769 \times 10^{-2}[NO_2]^2 + 0.19[O_3] - 1188.5[THC]^{-1} - 0.057[W.S]^2$	0.99
Bangi-Dong	Class1	$[O_3]_t = -99.8 - 553[SO_2]^{-1} - 0.29[TSP] + 6.2 \times 10^{-6} \exp[CO] - 520[NO_2]^{-1} + 2.1[O_3] - 1.14[THC] + 21.3hr[W.S]$	0.91
	Class2	$[O_3]_t = 11.1 - 120.3[SO_2]^{-1} - 90.5[TSP]^{-1} - 0.054 \exp[CO] - 0.16[NO_2] + 1.2[O_3] + 0.1[THC] - 0.21[W.S]$	0.94
	Class3 (Outlier)	$[O_3]_t = -16.6 + 8.954 \times 10^{-4} [SO_2]^2 + 0.25[TSP] - 7.041 \times 10^{-6} \exp[CO] - 2.3[NO_2] + 8293[O_3]^{-1} + 1.503 \times 10^{-2} [THC]^2 - 199[W.S]^{-1}$	0.95
Sungsu-Dong	Class1	$[O_3]_t = 345.2 + 204.6[SO_2]^{-1} + 19841[TSP]^{-1} + 179.8 \ln[CO] - 9570[NO_2]^{-1} - 3843[O_3]^{-1} - 182.4 \ln[THC] - 3.95 \times 10^{-4} [W.S]$	0.95
	Class2	$[O_3]_t = -15.4 + 0.27[SO_2]^{-1} - 0.08[TSP] + 6.5 \times 10^{-7} \exp[CO] - 551[NO_2]^{-1} + 1[O_3] + 1[THC] - 0.9[W.S]$	0.93
	Class3	$[O_3]_t = -626.3 + 1079[SO_2] + 3392[TSP]^{-1} - 78.6[CO]^{1/2} + 13367[NO_2]^{-1} + 0.63[O_3] - 2.65 \times 10^{-8} \exp[THC]^2 + 30.3[W.S]^{1/2}$	0.96
	Class4	$[O_3]_t = -37.7 + 0.05[SO_2] + 9.5 \ln[TSP] - 1.982 \times 10^{-5} \exp[CO] - 0.93[NO_2] + 0.78[O_3] + 14.35[THC]^{1/2} - 17.1 \ln[W.S]$	0.87
	Class5 (Outlier)	$[O_3]_t = 122.1 - 0.56[SO_2] - 42.95 \ln[TSP] + 1.3[CO] - 1920.6[NO_2]^{-1} + 1.71[O_3] + 1946[THC]^{-1} + 96.5[W.S]^{-1}$	0.88

* [SO₂], [NO₂], [O₃], [O₃] - (ppb) ;
 [CO], [THC] - (ppb/100) ;
 [TSP] - (μg/m³)

Table 3. Contribution of independent variables of O₃ prediction model for 1-hr after basis.

Pollutants		SO ₂	TSP	CO	NO ₂	O ₃	THC	W.S	Total%
Kuui-Dong	Class1	8.4	8.1	3.2	0.3	57.0	10.6	2.9	90.5
	Class2	79.4	5.2	8.5	0.3	1.0	4.0	1.5	99.8
Bangi-Dong	Class1	6.3	3.7	2.7	0.7	61.1	2.7	13.4	90.5
	Class2	4.2	—	3.0	—	85.0	—	1.3	93.6
	Class3	1.5	1.6	47.4	9.6	10.7	18.8	5.7	95.3
Sungsu-Dong	Class1	14.3	14.0	4.5	23.9	22.7	12.6	2.9	94.9
	Class2	0.6	0.5	23.5	0.3	64.7	2.4	1.2	93.3
	Class3	16.9	0.2	22.5	12.0	20.2	8.0	16.1	95.8
	Class4	—	2.6	10.6	5.7	57.5	4.9	6.3	87.5
	Class5	2.1	1.6	1.2	2.2	63.0	11.8	6.6	88.3

의동에서는 NO₂가 오존생성에 가장 큰 기여를 하였으며, 방이동은 TSP, CO, 풍속 등의 변수가 모델식에서 커다란 기여를 하였고, 성수동의 집단1에서는 풍속이 오존생성에 가장 큰 기여를 하였다. 본 연구에서는 기여도가 작은 변수를 제거하고 모델식을 새로 개발하는 방식은 시도하지 않았다.

1시간 후의 오존농도 추정모델은 현시점의 오존을 포함한 모든 상기변수를 이용하여 1시간 후의 오존농도를 예측하는 것이다. 이 추정모델은 현시점의 오존농도 추정모델보다 유의성이 높았다. 또한, 분류된 소집단의 예측모델이 전체자료에 대한 예측모델보다 높은 유의성을 보였다. 구의동, 방이동, 성수동에서는 전체의 경우, R²이 0.44~0.50로 낮은 반면, 각 소집단 모델의 경우 0.87 이상의 유의성을 보였다. 1시간 후에 대한 오존예측도 변수를 자료변환하여 회귀분석한 결과, 유의성이 크게 향상된 회귀식이 유도되었다. 표 2는 SPSS의 단계별 회귀분석을 이용하여 대상지역의 오존농도 추정식을 개발한 것이며, 표 3은 표 1과 같이, 각 소집단의 모델에서 각 변수의 오존예측에의 기여도를 표시한 것이다. 구의동의 집단1에서는 현재 오존농도가 1시간 후 오존 예측에 가장 큰 기여를 했으며, 집단 2에서는 SO₂가 가장 큰 기여를 한 변수였다. 방이동, 성수동에서는 전반적으로 현재의 오존농도가 추정식에서 큰 기여를 했으며, CO, NO₂가 중요 변수로써 이용된 집단도 있었다.

4. 종합 오존농도 예측모델의 개발

4. 1 오존농도 예측모델의 실험.

각 대상지역의 오존농도 추정식으로 현시점의 오존농도를 예측한 결과, 높은 예측도가 나타났다.

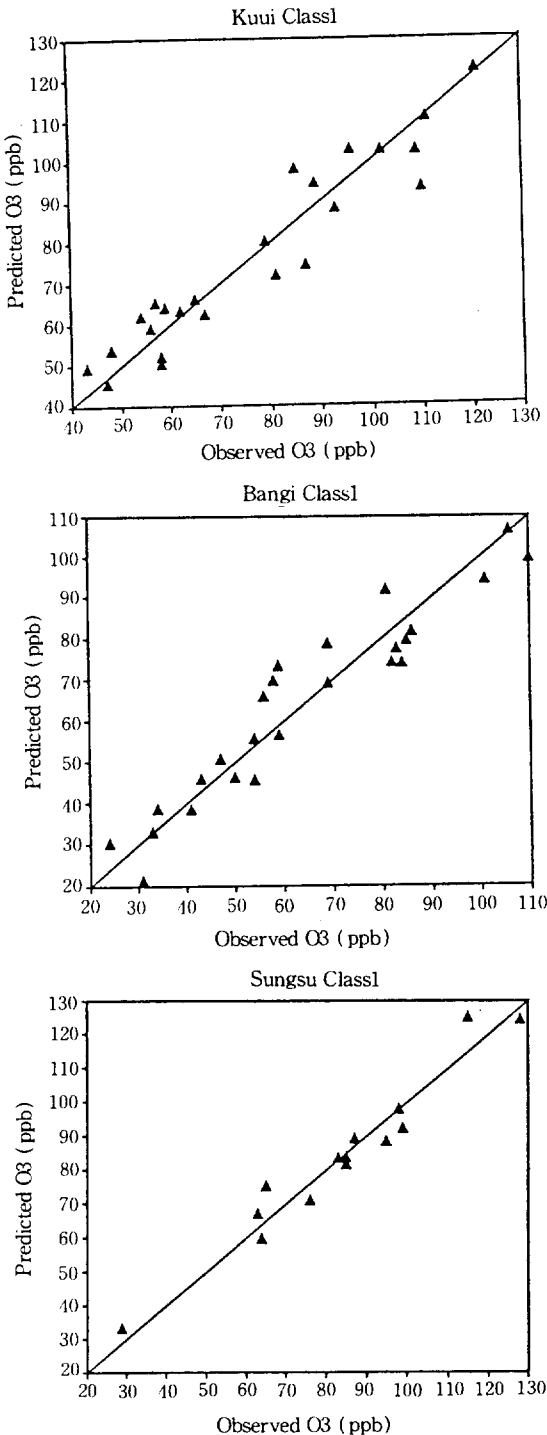


Fig. 8. Comparisons between the predicted O₃ and the observed O₃ with the O₃ prediction model for 1-hr after level.

구의동에서는 각 소집단의 예측오차가 평균 0.6~5.0%였으며, 방이동은 2.9~4.0%, 성수동은 2.5~7.2%로 나타났다.

1시간 후의 오존추정식은 현시점 오존예측모델보다 유의성은 높았으나 오차가 큰 경우들이 있다. 기본적인 이유는 자료의 연속성이 부족하였기 때문이다. 즉, 1시간, 3시간, 6시간 후의 예측모델을 개발하기 위해서는 대상지역에서 모델개발 전후 시간의 연속자료가 필수적이거나, 입수 원자료(raw data)에는 실종자료(missing data)가 대부분이었기 때문이다. 이에 따라, 성수동 집단5에서는 31.3%로 큰 오차를 보이기도 하였다. 하지만, 전반적으로 평균오차는 10.9%와 오차의 표준편차 7.8%로 비교적 높은 예측도를 나타냈다. 구의동에서는 각 소집단의 평균오차의 범위가 0.6~8.0%였으며, 방이동은 4.6~10.8%로 나타났다. 그림 8은 모델 추정식에 의한 추정치와 실제 오존농도의 측정치를 1:1로 비교한 것으로 추정모델의 우수함을 보여주고 있다.

4. 2 전문가 시스템의 개발과 검증.

오존을 제외한 임의의 주요오염물질의 농도를 입력하였을 때, 오존농도를 예측할 수 있는 종합적인 오존예측모델의 개발을 수행하였다. 종합모델 개발시 선행되어야 할 점은 임의의 자료를 경향(pattern) 별로 분류하는 작업이다. 즉, 임의의 자료가 입력되었을 때, 이 자료의 경향을 파악하고, 이미 군집분석으로 경향이 파악되어 모델화된 많은 소집단 추정식 중 하나에 적용될 수 있어야 한다. 이를 위해, 이미 모델화된 각 측정장소의 모든 소집단을 기지의 사례(case example)로 인정하고 전문가 시스템을 개발하였다. 전문가 시스템이란 인공지능의 한 부류로, 기지의 정보에 바탕을 두고, 시스템을 해석, 추정, 진단, 설계, 계획, 통제, 감시하는데 사용되고 있다. 이 분류법칙을 정하는 기준으로, 각 소집단 별로 오염물질 농도와 풍속에 대한 최대 및 최소의 범위를 이용하였으며, 이 범위가 각 소집단의 고유영역이 되도록 많은 반복과정을 거쳤다. 이러한 기준과 이미 개발된 오존추정식을 기반으로 FORTRAN 언어를 이용하여 전문가 시스템을 개발하였다. 개발된 프로그램은 종합적인 오존예측모델로써, 그림 9에는 이 모델의 흐름도를 표현하였다.

종합 오존예측모델은 구의동, 방이동, 성수동의 오염물질과 기상자료들이 입력되었을 때, 먼저, 대상지역의 경향을 진단한 다음, 진단된 지역의 사례를 조사하고, 해당 사례의 추정식으로 오존농도

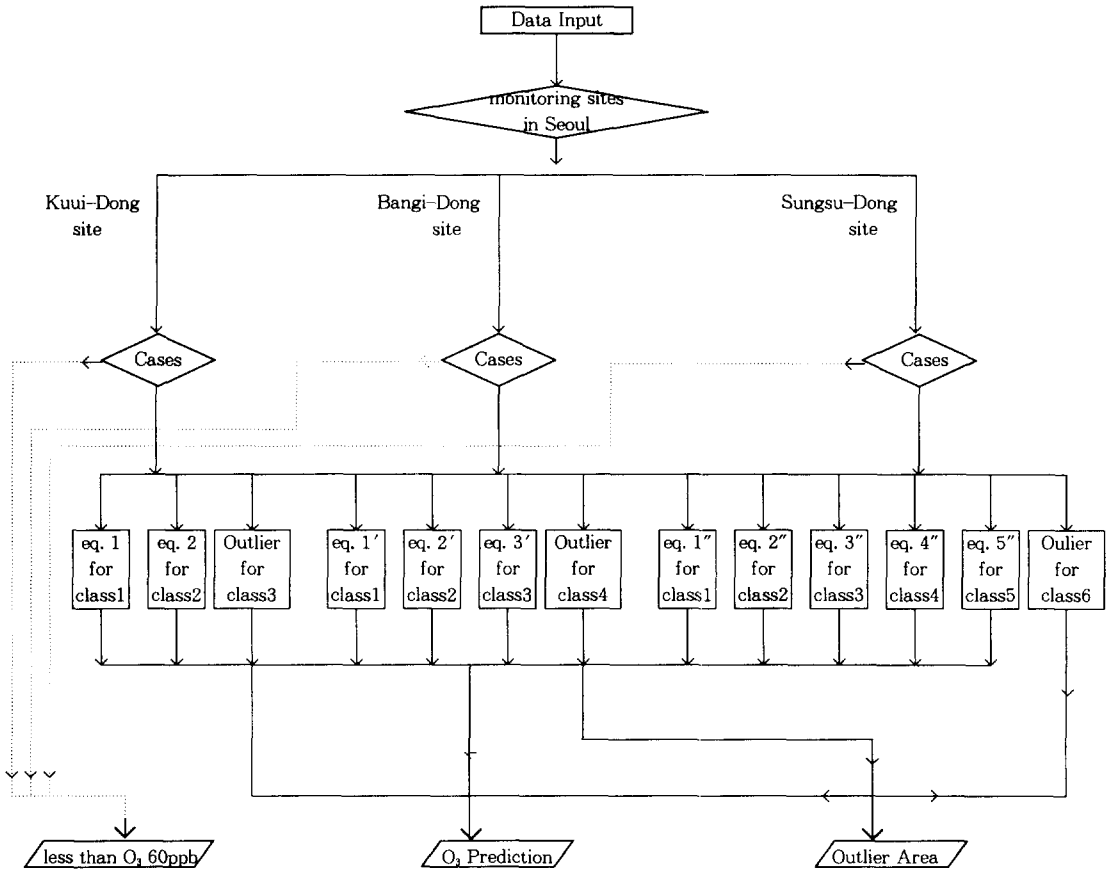


Fig. 9. Flow chart of O₃ prediction program.

를 예측하도록 하였다. 입력자료가 대상지역의 어느 범주에도 속하지 않을 경우, 오존농도를 60 ppb 미만으로 처리하였고, 이상치 자료범주의 경우에 속할 때는 이상치 집단으로 분류하였다. 1989년과 1990년도의 구의동, 방이동, 성수동에서 오존농도 60 ppb 미만의 자료를 종합 오존예측모델에 적용하여 검증한 결과, 현시점 오존을 예측하는데 있어서 구의동의 경우, 오존 60 ppb 미만을 진단하였을 때, 전체 3,636개의 자료에서 3,444개 (94.7%)가 올바르게 분류되었다. 방이동에서는 전체 2,261개의 자료 중에서 2,019개 (89.3%)가, 성수동에서는 전체 5,020개의 자료 중 4,678개 (93.2%)가 바르게 분류되어 높은 예측도를 보였다.

5. 결 론

본 연구는 1989년과 1990년의 서울시 대기오염

자동측정망자료를 이용하여 주요 오염물질 (SO₂, TSP, CO, NO₂, THC)과 기상조건을 오존생성에 영향을 미치는 변수로 선정하고, 다각적인 통계분석을 응용하여 오존농도를 추정하는 모델을 개발하였다. 연구결과 다음과 같은 결론을 얻었다.

- 1) 오존생성과 관련하여 주요오염물질 (SO₂, TSP, CO, NO₂, THC)의 시간별 변화를 계절별로 조사한 결과, 오전 중 저풍속시 주요오염물질의 축적이 오존생성에 영향을 미친다는 것이 확인되었다.
- 2) 군집분석의 결과, 저농도 오존이 나타날 때, SO₂와 TSP가 가장 큰 농도변화를 보였으며, 오존 60 ppb 이상인 고농도에서는 모든 주요오염물질이 유사한 경향을 보였다. 군집분석을 통해 고농도와 저농도의 오존이 생성될 때, 각 주요오염물질의 거동이 파악되었다.
- 3) 다중회귀분석의 결과, 군집분석 후 분류된

소집단의 통계적 유의성은 높았으며, 변수들의 비선형적 변환을 통해 유의성이 향상된 오존추정식을 얻을 수 있었다. 대상지역별로 현시점의 오존농도를 추정할 때, NO₂, CO, THC 등이 큰 기여를 했으며, 1시간 후 오존예측에서는 현재의 오존이 주요변수가 되었다. 현시점 오존예측모델의 예측 오차는 0.6~7.6%이었으며, 1시간 후 오존예측모델의 오차는 0.61~10.87% 정도로 매우 높은 예측도를 나타냈다.

4) 전문가 시스템을 이용하여 종합 오존예측모델을 개발하였다. 개발된 법칙을 1989년과 1990년 오존농도가 60 ppb 미만인 자료에 실행·검정한 결과 예측율은 매우 높았다.

이와 같은 결과로, 연속적인 원자료가 확보될 때, 즉, 실증자료가 작을 경우, 1시간 후의 오존농도의 예측뿐만 아니라, 3시간 혹은 6시간 후의 오존농도의 추정도 같은 방법으로 분석가능하며, 오존 조기경보체계 개발시 새로운 방법론으로 응용될 수 있다.

참 고 문 헌

- 박옥현 (1989) 대기오염 사전발생 예측에 관한 연구(Ⅲ), *대한환경공학회지*, 11(2), 9-16.
- 허정숙 (1993) 다변량 통계분석을 이용한 서울시 오존농도 예측에 관한 연구, 석사논문, 경희대학교.
- Angle, R. P. and H. S. Sandhu (1989) Urban and rural ozone concentrations in Alberta, Canada, *Atmos. Environ.*, 23(1), 215-221.
- Burkhart, R. P. (1990) A trajectory analysis of ozone exceedances in coastal Maine, AWMA for presentation at the 83rd Annual Meeting & Exhibition, Pittsburgh, Pennsylvania, June 24-29.
- Dorling, 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 applications, *Atmos. Environ.*, 26A(14), 2575-2581.
- Edwards, P. J., F. Wood, and J. N. Kochenderfer (1991), Characterization of ozone during consecutive drought and wet years at a rural West Virginia site, *J of AWMA*, 41(11), 1450-1453.
- Feister, U. and K. Balzer (1991) Surface ozone and meteorological predictors on a subregional scale, *Atmos. Environ.* 25A(9), 1781-1790.
- Hopke, P. K. (1985) Receptor Medeling in Environmental Chemistry, John Wiley & Sons, New York.
- Kelly, N. A., M. A. Ferman, and G. T. Wolf (1986) The chemical meteorological conditions associated with high and low ozone concentrations in Southeastern Michigan and nearby areas of Ontario", *JAPCA*, 36, 150-158.
- Korsog, P. E., and G. T. Wolff (1991) An examination of urban ozone trends in the Northeastern U. S.(1973-1983) using a robust statistical method, *Atmos. Environ.*, 25B(1), 47-57.
- Lance, G. N. and G. T. Williams (1967) A general theory of classificatory sorting strategies hierarchical system, *Computer Journal*, 9(4), 373-380.
- Lefohn, A. S., D. S. Shadwick, U. Feister, and V. A. Mohnen (1992) Surface-level ozone: Climate change and evidence for trends, *J. of AWMA*, 42(2), 136-144.
- McKendry, I. G. (1993) Ground-level ozone in Montreal, Canada, *Atmos. Environ.*, 27B(1), 93-103.
- Poulida, O., R. G. Wardell, and J.G. Watkins (1991) Trace gas concentrations and meteorology in rural Virginia -1. Ozone and carbon monoxide, *Journal of Geophysical Research*, 96D(12), 22461-22475.
- Reiter, R. (1991) On the mean daily and seasonal variations of the vertical ozone profiles in the lower troposphere, *Atmos. Environ.*, 25A(9), 1751-1757.
- Wolff, G. T. and P. E. Korsog (1992) Ozone control strategies based on the ratio of volatile organic compounds to nitrogen oxides, *J. of AWMA.*, 42(9), 1173-1177.