

비선형 회귀 모형을 이용한 서울지역 오존의 고농도 현상의 모형화

정수연¹ · 최기현²

¹덕성여자대학교 정보통계학과, ²덕성여자대학교 정보통계학과

(2009년 4월 접수, 2009년 5월 채택)

요약

본 연구에서는 서울지역 오존의 기상상태와 추세경향에 따른 고농도 현상을 모수적 방법인 비선형회귀모형(nonlinear regression model)으로 모형화 하였다. 여기서는 1995년부터 1999년까지의 자료로부터 오존과 고농도 현상에 영향을 줄 수 있는 기상상태와 추세경향 등을 순차적으로 추가함으로써 고농도 현상을 예측하는 모형을 추정하였다.

주요용어: 비선형회귀모형, 오존, 장기추세, 계절추세.

1. 서론

성층권에 존재하는 오존은 태양의 해로운 자외선으로부터 동식물과 사람을 보호해주는 중요한 역할을 한다. 이와는 반대로 지상 근처 오염된 대기 중의 오존은 건강과 환경을 황폐하게 만든다. 자동차의 급격한 증가 및 산업 활동의 증가에 따라 대도시의 오존의 오염농도는 매년 악화되고 있는 실정이다. 지구표면의 오존(Ozone, O₃)은 눈과 목을 자극하고 식품에 해를 끼치는 광화학 스모그(Photochemical Smog)의 주된 성분이다. 그러나 대기 중 오존의 대부분(약 97%)은 산소원자와 산소분자가 결합하여 자연적으로 오존을 형성하는 성층권에 자리 잡고 있다. 이곳의 밀도는 부피로 따져 평균 0.002%도 못 된다. 그러나 이처럼 적은 양일지라도 태양의 해로운 자외선으로부터 동식물과 사람을 보호해 주기 때문에 중요한 역할을 한다. 이와는 반대로 지상 근처 오염된 대기 중의 오존(대류권 오존이라고도 함)은 대기로 직접 방출된 것이 아닌 2차 오염물질이다. 이것은 광화학적으로 휘발성 유기화합물과 질소 산화물 등의 전구물질에 의해 형성된다. 더욱이 오존 형성에는 기상 조건이 중요한 역할을 한다. 오염물질은 기상상태에 따라 다른 지역으로 수송되어 그 지역에서 고농도 대기오염 현상을 나타낼 수 있으며 배출원 부근의 대기 안정도가 안정한 경우에는 그 지역에서 오염물질의 고농도 현상이라는 문제점을 발생시킨다. 세계 여러 지역에서도 수송에 의한 오염물질의 고농도 현상에 대한 고찰이 있어왔다. Greater Athens Area(GAA)의 경우에도 우리나라와 마찬가지로 대기오염문제가 지난 20년 동안 문제시 되어왔다. 지난 20년 동안 GAA의 대기질에 관한 연구과제의 수행결과, 이산화황(SO₂)의 관측치가 겨울의 풍속과 최저기온 그리고 강우 등과 관련이 있으며 여름에는 상대습도와 최저기온, 풍속과 풍향과 관련이 있음이 밝혀졌다. 또한 전 날 오염물질의 최고농도와 오존, 질산(NO₂)농도의 상관관계를 분석하였고 상대습도, 기온, 바람과의 관계에 대해 연구되어왔다.

²교신저자: (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 정보통계학과, 교수.

E-mail: khchoi@duksung.ac.kr

우리나라에서는 저황유의 공급, 청정연료 사용 의무지역의 확대 등 정부의 적극적인 대기오염 개선 대책에 따라 90년대 들어와 황산화물, 먼지 등의 대기오염도는 계속 감소하고 있으나, 자동차의 급격한 증가 및 산업 활동의 증가에 따라 대도시의 오존의 오염농도는 매년 악화되고 있는 실정이다. 오존의 고농도 현상은 지상풍이 약하고 침체한 고기압역이 해당지역을 덮고 있는 맑고 더운 여름날에 최고조에 달한다. 이처럼 여러 변수가 작용하므로 연간 오존량의 추세는 일정하지 않다.

오존 농도와 기상상태간의 복잡한 관계를 선형적으로 모형화 하는 것은 무리가 있다. Cox와 Chu (1992)는 일반화 선형모형(generalized linear model)을 사용함으로써 오존의 고농도 현상을 예측하는 모델을 만들었다. 여기서는 오존의 고농도 현상을 설명하는데 있어서 기상상태를 조건으로 와이블 분포(weibull distribution)를 따른다고 가정하고 있으며, 기상상태간의 상호작용 효과는 선형조합으로 설명하였다. Bloomfield 등 (1996)은 시카고의 오존 농도에 대한 지표대기의 효과를 연구하였다. Davis 등 (1998)은 시카고 오존 농도에 대한 모수적 방법인 비선형회귀모형으로 모형화하였다.

본 연구에서는 서울지역 오존의 기상상태와 추세경향에 따른 고농도 현상을 모수적 방법인 비선형회귀모형으로 모형화 한다. 이 모형에는 1995년부터 1999년까지의 오존 농도와 고농도 현상에 영향을 줄 수 있는 기상자료, 추세경향 등을 순차적으로 추가하여 고농도 현상을 예측하는 모형을 추정하였다. 선형회귀분석에서는 변수들 간의 관계를 선형함수로 나타내어 설명을 하였으나, 우리 주변에서 나타나는 현상들은 변수들 간의 비선형관계의 결과로 설명되는 것이 더욱 자연스러운 경우가 많다. 물론 변수들 간의 관계가 알려져 있지 않은 경우에는 근사적 방법으로 선형모형을 사용할 수 있다. 그러나 컴퓨터의 발달로 비선형관계를 갖는 경우에 대해 활발한 연구가 진행되고 있다. 회귀모형에 주어진 회귀식이 모수들의 비선형함수로 나타내어지는 경우의 비선형회귀모형에 대한 통계적 분석은 회귀계수의 추정방법에서 선형회귀모형의 경우와 큰 차이가 있으며, 대부분의 결과는 근사적 방법으로 유도된다. 비선형회귀모형은

$$Y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, 2, \dots, n$$

으로 정의되며 $E(\epsilon) = 0$, $\text{var}(\epsilon) = \sigma^2$ 의 가정을 갖는다. 이때 $f(\mathbf{x}_i, \boldsymbol{\theta})$ 는 설명변수들 \mathbf{x} 와 미지의 p 차원 모수벡터 $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)'$ 의 함수로 주어지는 회귀식을 나타낸다. 이 회귀식이 모수 $\boldsymbol{\theta}$ 들의 비선형함수로 주어지면 모형은 비선형회귀모형이 된다.

2. 모수의 추정

2.1. 최소제곱법

비선형회귀분석에서도 모수 $\boldsymbol{\theta}$ 의 추정방법으로 최소제곱법을 사용한다. 즉, $\boldsymbol{\theta}$ 의 최소제곱추정량 $\hat{\boldsymbol{\theta}}$ 은 오차의 제곱합

$$\begin{aligned} S &= \sum_{i=1}^n \epsilon_i^2 \\ &= \sum_{i=1}^n \{Y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\}^2 \end{aligned}$$

을 최소화하는 $\boldsymbol{\theta}$ 의 값으로 주어진다. 선형회귀분석에서와 같이 제곱합에 대하여 편미분을 하면

$$\frac{\partial S}{\partial \theta_j} = (-2) \sum_{i=1}^n \{Y_i - f(\mathbf{x}_i, \boldsymbol{\theta})\} \frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_j}, \quad j = 1, 2, \dots, p$$

와 같이 주어지며 이들을 0으로 놓고 정리한 후 θ 에 추정량 $\hat{\theta}$ 을 대입하면 정규방정식을 얻을 수 있다. 그러나 정규방정식은 $\hat{\theta}$ 의 비선형함수이므로 정규방정식을 만족하는 $\hat{\theta}$ 의 값을 직접 구하지는 못하며 가우스-뉴턴(Gauss-Newton) 방법 등과 같은 반복추정방법을 사용하여야 한다.

2.2. 가우스-뉴턴 방법

비선형회귀분석에서 최소제곱추정량을 구하기 위하여 사용하는 반복추정방법들 중에서 가장 기본적인 것은 가우스-뉴턴 방법이다. 이 방법은 매 단계에서 비선형회귀식을 모수 θ 에 대한 선형함수로 근사화하여 선형회귀분석 방법으로 모수 θ 를 추정해 나가는 것이다. 즉, 모수 θ 에 대해 타당한 초기값 $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_p^0)'$ 이 주어진다고 가정하여 일차 테일러(Taylor) 전개를 하면 비선형회귀식을

$$f(\mathbf{x}_i, \boldsymbol{\theta}) \approx f(\mathbf{x}_i, \boldsymbol{\theta}^0) + \sum_{j=1}^p \frac{\partial f(X_i, \boldsymbol{\theta})}{\partial \theta_j} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} (\theta_j - \theta_j^0)$$

와 같이 θ 들의 선형함수로 근사화 시킬 수 있다.

2.3. 초기치의 계산

가우스-뉴턴 방법을 사용하기 위해서는 사전에 알려진 모수 θ 의 초기치가 필요하다. 이 초기치로는 과거의 분석에서 알려진 θ 의 값, 또는 해당 전문가의 경험이나 지식에 의해 주어지는 값을 사용할 수 있다. 가우스-뉴턴 방법에서의 반복추정량이 제대로 수렴을 하기 위해서는 정확한 초기치의 선택이 중요하다. 오차 제곱합의 공간은 매우 복잡한 형태를 가질 수 있으므로 시작점에 따라 값이 다른 국소 최적점(local optimum)에 수렴하는 경우가 있을 수 있다. 그러므로 정확한 초기치를 사용할수록 대역 최적점(global optimum)에 도달할 가능성이 높아진다. 경우에 따라서는 과거의 분석에서 알려진 θ 의 값이 없는 경우 현재의 관측치를 이용하여 θ 의 타당한 초기치를 구하는 방법을 이용하기도 한다. 여기서는 후자의 방법을 이용하여 비선형모형에 적합 시켜보았다.

3. 자료 및 분석결과

3.1. 자료

본 연구에서 사용된 자료는 1995년부터 1999년까지의 서울시의 오존 농도자료와 기상자료이며 모든 분석은 S-PLUS 6.1을 이용하였다.

3.1.1. 오존농도자료 본 논문에서 사용된 오존 농도는 통계청에서 입수한 1995년에서 1999년까지 일별 서울지역의 27곳의 각 측정소들의 자료를 사용하였다. 연구 기간 중 발생한 결측치에 대해서는 27곳의 평균을 이용한다. 오존 농도에 대한 그림 3.1, 3.2와 그림 3.3으로부터 매년 4월부터 10월에 걸쳐 오존 농도의 고농도 현상이 나타나고 있으며, 하루 중 오존 농도의 고농도 현상을 보이는 시간대는 오후 12:00부터 오후 18:00사이임을 알 수 있다.

3.1.2. 기상자료 기상상태의 지표자료는 서울시의 종로구 송월동 1번지에 위치하고 있는 기상청의 지상관측 점에서 측정된 자료로 최대기온(°C, Max.Temp), 최대풍속(m/s, Wind.Speed), 상대습도(%, Rel.Humidity), 일사량(0.01MJ/m², Radiation)에 대한 자료를 이용하였다. 여기서 선택된 설명변수들은 표 3.1과 같다.

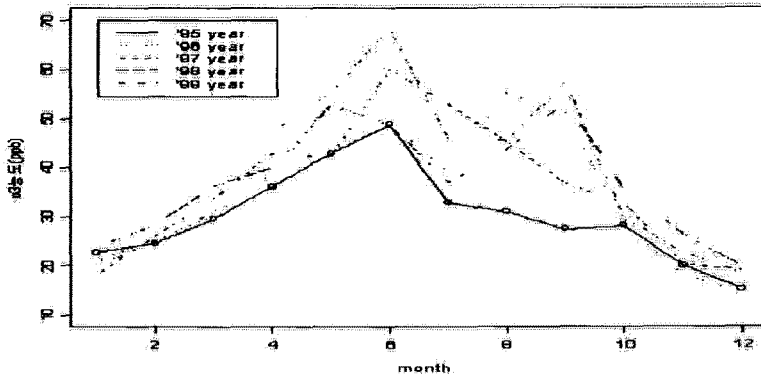


그림 3.1. 각 년도의 오존 농도의 월별 변화량

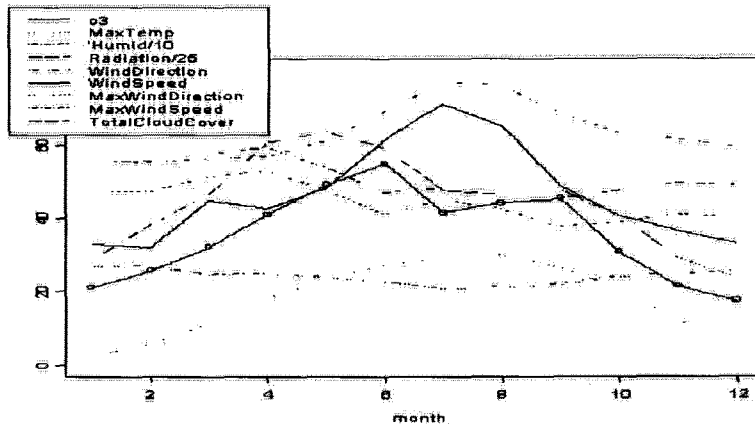


그림 3.2. 월별 오존 농도와 기상변수들의 변화량

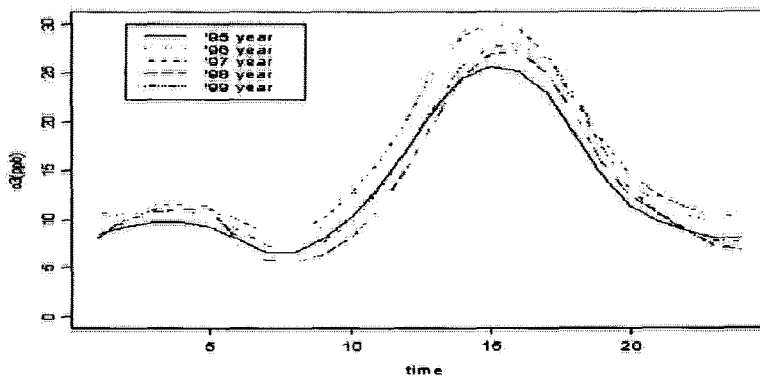


그림 3.3. 각 년도별 오존 농도의 일변화

표 3.1. 대기 변수

변수	단위	기호
최대기온	℃	mt
최대풍속	ms ⁻¹	mwspd
상대습도	%	rh
일사량	%	radi

표 3.2. 오존 농도와 대기 변수들의 기초통계량

변수	평균	편차	전체	백분률				
				최소	25%	50%	75%	최대
오존	35.37	18.35	1862	5.95	22.34	31.29	44.51	117.80
최대기온	17.52	10.28	1862	-7.90	8.30	19.05	26.50	36.10
최대풍속	88.75	34.55	1862	22.00	64.00	81.00	107.00	315.00
상대습도	63.36	14.50	1862	19.00	53.50	64.00	73.90	96.30
일사량	1108	589.39	1862	21	678	1056	1507	2668

표 3.3. 오존 농도와 대기 변수들 간의 상관관계

변수	상관관계				
	오존	최대기온	상대습도	최대풍속	일사량
오존	1	0.623	-0.081	0.040	0.599
최대기온		1	-0.177	0.412	0.426
최대풍속			1	-0.117	0.012
상대습도				1	-0.413
일사량					1

3.1.3. 자료의 특성 연구기간 동안 오존 농도와 기상 변수의 특징을 알아보기 위해 산점도(scatter plot)와 기초통계량, 상관관계를 살펴보았다. 다음 그림 3.4-3.7은 '95년-'99년의 자료로부터 산점도 (Tukey, 1977)와 LOESS 곡선 (Chambers와 Hastie, 1993)으로 오존 농도와 기상 요인들 간의 계절성을 확인할 수 있다. 기초통계량 결과 오존 농도의 최대값이 117.80ppb이었다. 120ppb를 넘을 때 발령하는 오존주의보가 1999년에 13번이나 발령한 것에 비해 여기서는 최대값이 120ppb보다 낮게 나타난 것은 한 시간 단위로 내려지는 오존 주의보와는 달리 본 연구에서 오존 농도의 하루 대표값은 27개 관측소의 평균값 중 최대값을 사용했기 때문이다. 기상 변수인 일일 최대기온, 상대습도, 풍속, 일사량과 오존 농도의 상관관계를 살펴보면 오존 농도는 최대기온과 일사량의 상관계수가 0.623과 0.599로 강한 양의 상관관계를 갖는다. 또한 기상 변수간의 상관관계를 살펴보면 일사량은 온도와 양의 관계를 보이거나 상대습도와는 음의 관계를 나타낸다.

3.2. 오존 농도의 모형화와 분석 결과

영향인자를 이용하여 오존 농도의 경향을 파악하기 위한 가장 기본적인 방법은 오존 농도와 영향인자간의 통계모형을 세우는 것이라 할 수 있겠다. 이때 자연 현상을 설명하는 많은 환경자료의 분석방법으로는 비선형모형이 더 적절하듯이 오존 농도와 영향인자간의 복잡한 관계를 선형모형으로 설명하는 것은 충분하지 못하다. 오존 농도는 기상변수에 의해 영향을 받는다. 오존 농도의 모형은 크게 설명변수의 유형별로 분류하여 모형을 고려하였다.

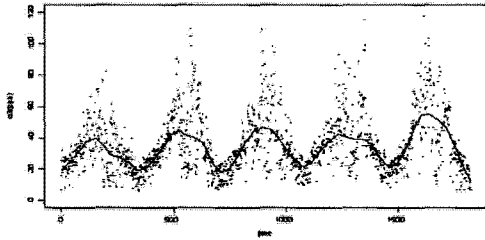


그림 3.4. 오존

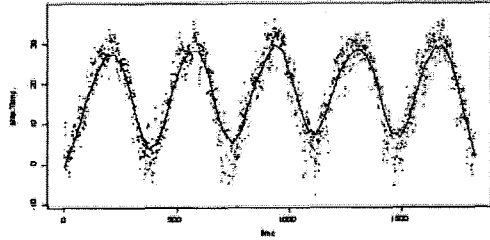


그림 3.5. 최대기온

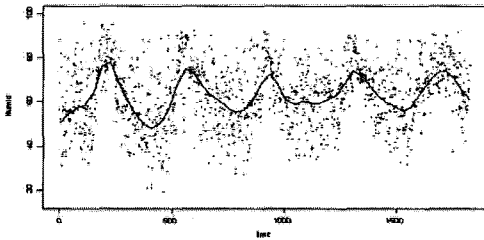


그림 3.6. 상대습도

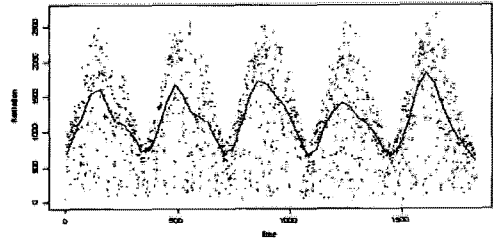


그림 3.7. 일사량

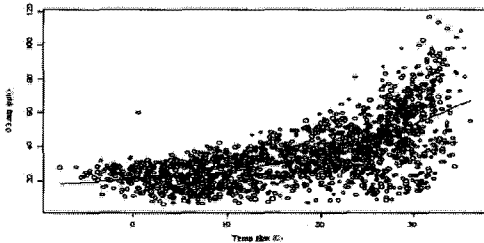


그림 3.8. 오존의 최대기온

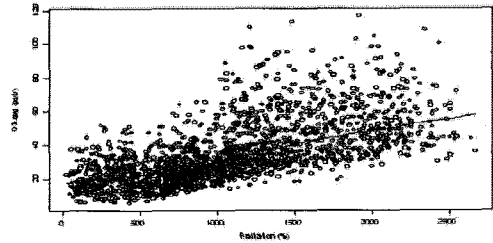


그림 3.9. 오존과 일사량

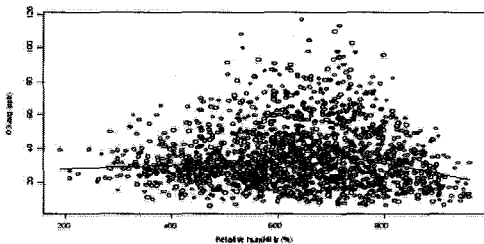


그림 3.10. 오존과 상대습도

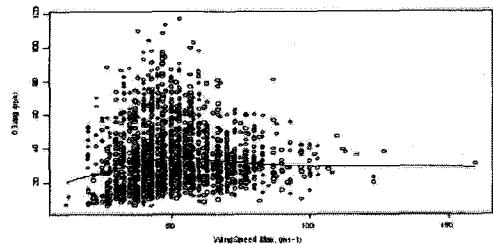


그림 3.11. 오존과 최대풍속

3.2.1. 모형화 단계 앞서 설명하였듯이 최대기온, 일사량, 상대습도, 풍속 등과 같은 기상변수들이 오존 농도의 고농도현상에 중요한 변수임이 확인되었다.

오존 농도의 고농도현상을 설명하는 모형은 다음 4단계에 걸쳐 설명한다.

Step 1. 자료에 대한 EDA(산점도와 비모수적 회귀방법)를 실시한다.

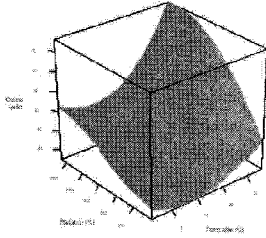


그림 3.12. 3차원 그래프

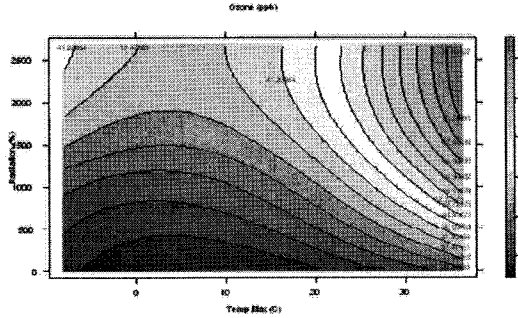


그림 3.13. 최대기온, 일사량 및 오존의 등고선 그래프

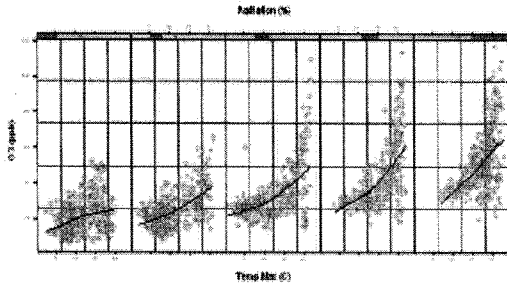


그림 3.14. 최대기온과 오존(조건: 일사량)

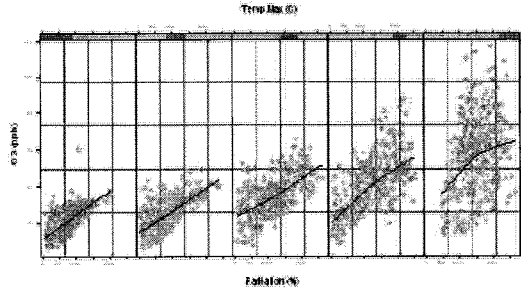


그림 3.15. 일사량과 오존(조건: 최대기온)

Step 2. 그 날의 기상자료와 오존 농도의 관계를 살펴본다. 그리고 상호작용 효과가 있는지 살펴본 후 변수들의 상호작용 효과의 구간을 모형에 추가한다.

Step 3. Step2에서 설정된 모형의 잔차들이 계절에 영향을 받는다면, 계절효과 구간을 모형에 추가한다.

Step 4. 마지막으로 계절추세경향(seasonal term trend)과 장기추세경향(long term trend)을 살펴본다.

각 단계의 모형은 R_p^2 값과 RSE(residual standard error)로 판단한다.

3.2.2. 지표 대기 오존 농도와 각 기상 변수들 간의 상관관계를 살펴 본 결과 오존 농도와 일사량, 최대기온들이 오존 농도와 가장 강한 관계를 갖고 있음을 알 수 있었다.

3.2.2.1. 오존 농도의 고농도 현상에 대한 최대기온과 일사량 그림 3.12는 오존 농도와 최대기온, 일사량의 관계를 3차원 공간에 나타낸 것이다. 3차원 그래프를 살펴보면 공간상에서 세 변수간의 관계는 일정한 형태의 평면이 아닌 최대기온이 증가함에 따라 오존 농도는 곡선의 형태로 증가하고 있음을 알 수 있다. 반면에, 오존 농도와 일사량의 관계는 일사량이 증가할 때, 오존 농도가 직선의 형태로 증가하고 있다. 그림 3.13은 오존 농도와 최대기온, 일사량의 관계를 등고선으로 나타낸 것이다. 오존 농도의 고농도 현상은 최대기온이 높고 일사량이 많을 때 나타난다. 그림 3.14와 3.15는 오존 농도에 대한 변수들의 관계를 좀 더 구체적으로 살펴보고자 각 요인을 일정 구간에서 겹치지 않도록 조건을 두고 좌표상에 나타낸 결과이다. 그림 3.14와 3.15는 각 구간에서 그래프의 모형이 약간씩 변화하는 것을 알 수 있다. 이는 오존 농도에 대한 최대기온과 일사량간의 상호작용이 존재하고 있음을 의미한다. 만일 상호작

표 3.4. 지표 대기를 고려한 오존 농도 모형화

I. 최대기온식					
	식	R^2	RSE	자유도 ₁	자유도 ₂
(i)	$a + t_1 \cdot (\text{최대기온} - 30)$	0.406	14.034	2	1813
(ii)	$(i) + t_2 \cdot (\text{최대기온} - 30)^2$	0.462	13.355	3	1812
(iii)	$(ii) + t_3 \cdot (\text{최대기온} - 30)^3$	0.464	13.338	4	1811
II. 최대기온과 일사량					
	식	R^2	RSE	자유도 ₁	자유도 ₂
(iv)	$(iii) \times (1 + r \cdot \text{일사량})$	0.591	11.645	4	1811
III. 최대기온, 일사량 및 상대습도					
	식	R^2	RSE	자유도 ₁	자유도 ₂
(v)	$(iv) \times (1 + h \cdot (\text{상대습도} - 60))$	0.591	11.648	5	1810
IV. 최대기온, 일사량 및 최대풍속					
	식	R^2	RSE	자유도 ₁	자유도 ₂
(vi)	$(iv)/(1 + \text{최대풍속}/w)$	0.596	11.573	5	1810

용 효과가 존재하지 않는다면 모든 구간에서 그래프의 형태는 동일한 결과를 갖게 된다. 따라서, 오존 농도에 대해 다음과 같은 모델을 제안한다.

$$\text{오존} = (\text{최대기온의 다중회귀식}) \times (\text{일사량의 선형식}) \tag{3.1a}$$

위 식 (3.1a)는 오존 농도와 최대기온의 관계는 다중회귀(Polynomial regression)식으로 설명하고, 일사량은 직선의 식으로 설명한다. 또한, 오존 농도에 대한 최대기온과 일사량간의 상호작용 효과를 고려해 주기 위해 곱의 형태로 표현한다. 최대기온의 다중회귀식의 차수는 표 3.4의 경우 차수증가에 따른 오존 농도의 R^2 값이 큰 차이를 보이지 않으므로 2차까지만 이용한다.

표 3.4의 II-(iv)에서 상수항 '1'을 더해 줌으로써 일사량이 '0'일 때, 오존 농도가 '0'이 되는 것을 막을 수 있다. 오존 농도와 최대기온, 그리고 일사량의 비선형회귀분석 결과 R^2 의 값은 0.591, 값은 11.645이다. 비모수적 방법인 LOESS (Hastie와 Tibshirani, 1990)의 R^2 값은 0.61이고, 모수의 수는 7.1이다. 표 3.4의 II-(iv)는 단지 4개의 모수만으로도 비모수적인 방법과 비슷한 결과를 얻는다. 여기서 선택된 모델은

$$\text{오존} = \{a + t_1 \cdot (\text{최대기온} - 30) + t_2 \cdot (\text{최대기온} - 30)^2\} \times (1 + r \cdot \text{일사량}) \tag{3.1b}$$

이다.

3.2.2.2. 오존 농도의 고농도 현상에 대한 최대기온과 최대풍속 그림 3.18을 보자. 이 그림에서 보듯이 최대풍속을 조건으로 한 오존 농도와 최대기온의 관계는 그림 3.15에서 살펴본 것처럼 구간에 따라 조금씩 다르게 증가하는 직선의 형태를 보인다. 그림 3.19의 경우는 최대기온을 조건으로 한 오존 농도와 최대풍속간의 관계를 나타낸다. 최대기온이 낮은 구간에서는 최대풍속이 증가하면, 오존 농도도 함께 증가한다. 반면, 최대기온이 높은 구간에서는 최대풍속이 증가함에 따라 오존 농도가 감소하는 것을 볼 수 있다. 따라서, 오존 농도와 최대기온, 최대풍속의 관계를 다음과 같은 모델로 제안한다.

$$\text{오존} = \text{상수항} + (\text{최대기온의 다중회귀식}) \times (\text{대풍속식}) \tag{3.2a}$$

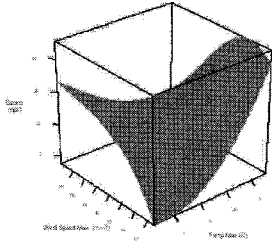


그림 3.16. 3차원 그래프

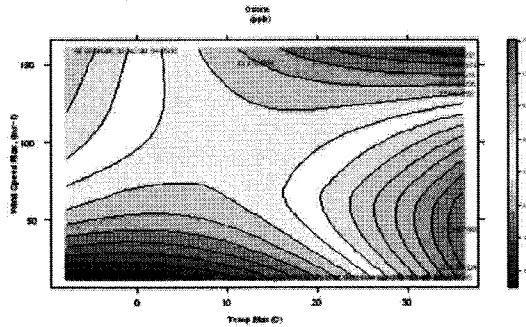


그림 3.17. 최대기온, 최대풍속, 오존의 등고선 그래프

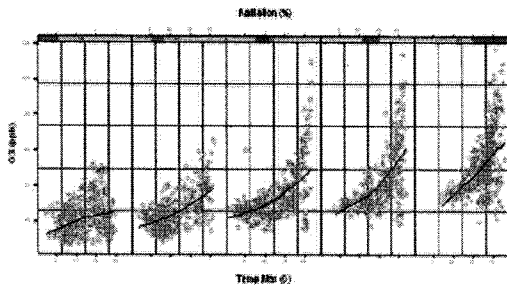


그림 3.18. 최대기온과 오존(조건: 최대풍속)

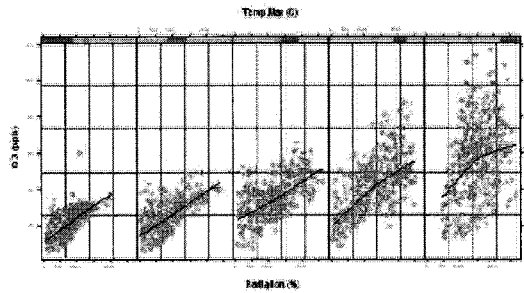


그림 3.19. 최대풍속과 오존(조건: 최대기온)

식 (3.2a)에서 오존 농도에 대한 최대풍속의 구간은 $1/[1 + \text{최대풍속}/w]$ 이다. 여기서 선택된 모델은

$$\text{오존} = \frac{a + t_1 \cdot (\text{최대기온} - 30) + t_2 \cdot (\text{최대기온} - 30)^2}{1 + \text{최대풍속}/w} \quad (3.2b)$$

이다. 식 (3.1a)와 (3.2a)를 결합하여

$$\text{오존} = \{ \text{상수항} + (\text{최대기온의 다중회귀식}) \times (\text{최대풍속식}) \} \times (\text{일사량의 선형식}) \quad (3.3a)$$

이다. 여기서 선택된 모형은

$$\text{오존} = \frac{\{ a + t_1 \cdot (\text{최대기온} - 30) + t_2 \cdot (\text{최대기온} - 30)^2 \}}{(1 + \text{최대풍속}/w)} \times (1 + r \cdot \text{일사량}) \quad (3.3b)$$

이다. 식 (3.3b)의 결과는 R^2 의 값이 0.596, RSE값은 11.573이다. 식 (3.3b)에서는 오존 농도에 대한 상대습도의 관계는 설명되지 않는다. 이는 표 3.4에서 살펴보듯이 상대습도를 추가해도 R^2 의 값이 크게 좋아지지 않으므로 여기서는 다루지 않았다.

3.2.3. 자연효과를 고려한 대기변수 자연효과를 생각 하지 않은 식 (3.3b)에 자연효과를 고려할 경우 하루 전의 최대기온이 오존 농도의 고농도 현상에 영향을 미치는 것으로 나타났다. 이때의 R^2 값은 0.600이고 RSE는 11.431이었다. 여기서 선택된 모형은

$$\begin{aligned} \text{오존} = & \frac{a + t_1 \cdot (\text{최대기온} - 30) + t_2 \cdot (\text{최대기온} - 30)^2 + t_{L1} \cdot (L1 \cdot \text{최대기온} - 30)}{1 + \text{최대풍속}/w} \\ & \times (1 + r \times \text{일사량}) \end{aligned} \quad (3.4)$$

표 3.5. 지표 대기의 자연효과를 고려한 오존 농도 모형화

I. 자연효과					
	식	R ²	RSE	자유도 ₁	자유도 ₂
(i)	표 3.4-(iii) + t _{L1} · (L1.최대기온 - 30)	0.600	11.431	6	1809
(ii)	(i) + t _{L2} · (L2.최대기온 - 30)	0.602	11.431	7	1808

이다.

3.2.4. 계절추세와 장기추세경향을 고려한 모형화 자연효과를 고려한 오존 농도 모형에 계절추세경향을 추가한다. 계절추세경향을 구하기 위해 푸리에 급수(Fourier series)를 사용한다. 푸리에 급수에 의한 계절추세의 효과는

$$\begin{aligned}
 & a_1 \left(\cos(2\pi year) - \overline{\cos(2\pi year)} \right) + b_1 \left(\sin(2\pi year) - \overline{\sin(2\pi year)} \right) \\
 & + a_2 \left(\cos(4\pi year) - \overline{\cos(4\pi year)} \right) + b_2 \left(\sin(4\pi year) - \overline{\sin(4\pi year)} \right) \\
 & = a_1c1 + b_1s1 + a_2c2 + b_2s2
 \end{aligned}$$

이다. 여기서 구한 계절추세경향을 추가한 결과 R² 값은 0.666이고 RSE는 10.541이었다. 여기서 선택된 모형은

$$\begin{aligned}
 \text{오존} = & \frac{\{a + t_1 \cdot (\text{최대기온} - 30) + t_2 \cdot (\text{최대기온} - 30)^2 + t_{L1} \cdot (L1.\text{최대기온} - 30)\}}{(1 + \text{최대풍속}/w)} \\
 & \times (1 + r \cdot \text{일사량}) + k + a_1c1 + b_1s1 + a_2c2 + b_2s2
 \end{aligned} \tag{3.5}$$

이다. 앞에서 살펴 본 기상변수와 계절적 요인이 아닌 시간이 흐르는 동안의 오존 농도 변화를 살펴보기 위해 장기추세구간을 추가한다. 장기추세구간은 natural cubic spline (Chambers와 Hastie, 1993)을 이용하여 구했다. 이 결과 R² 값은 0.673이고 RSE는 10.432였다. 그림 3.20은 1995년부터 1999년까지 서울지역 오존의 농도가 서서히 증가해 왔음을 보여준다.

4 단계에 걸쳐 실시한 결과 오존 농도의 고농도 현상을 예측하기 위한 최종모형은 다음과 같다.

$$\begin{aligned}
 \text{오존} = & \frac{\{a + t_1 \cdot (\text{최대기온} - 30) + t_2 \cdot (\text{최대기온} - 30)^2 + t_{L1} \cdot (L1.\text{최대기온} - 30)\}}{(1 + \text{최대풍속}/w)} \\
 & \times (1 + r \cdot \text{일사량}) \times (1 + tr \cdot yr) + k + a_1c1 + b_1s1 + a_2c2 + b_2s2.
 \end{aligned} \tag{3.6}$$

3.3. 결론

본 연구에서는 1995년부터 1999년까지의 서울 지역의 오존자료와 기상자료를 모수적 방법인 비선형회귀모형으로 설명하였다. 기존의 모형들은 비모수적 방법인 GAM을 사용하여 단순히 오존을 설명하는데 그쳤다면, 여기서는 모수적 방법을 사용함으로써 각 영향인자들로 인한 오존농도와의 관계를 살필 수 있었다. 회귀계수의 추정방법에서 선형회귀모형의 경우 큰 차이가 있으며, 대부분의 결과는 근사적 방법으로 유도된다. 이 때, 초기치가 필요하다. 초기치로는 과거의 분석에서 알려진 θ의 값이나, 또는 해당 전문가의 경험이나 지식에 의해 주어지는 값을 사용할 수 있다. 그러나 본 연구를 진행하면서 선택되었던 연구들이 부족하여 초기치를 설정하는데 있어 연구자의 주관에 의해 진행되다 보니 시작점에 따라 값이 다른 국소 최적점에 수렴 할 수 있음을 밝힌다. 그러므로 좀 더 시간을 갖고 정확한 초기치를

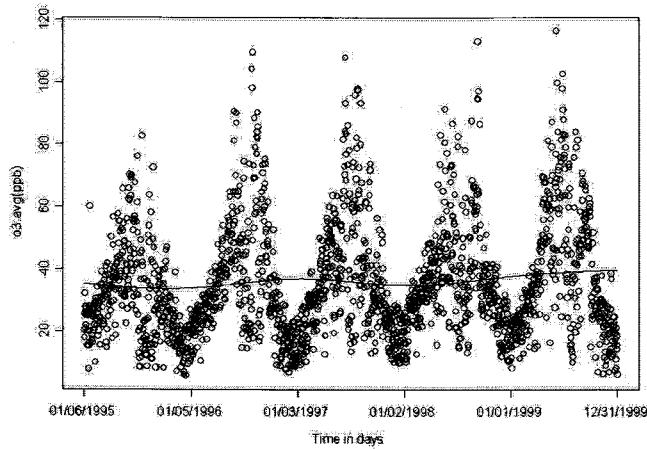


그림 3.20. 서울지역 오존 농도의 5년간의 변화

표 3.6. 최종 적합된 모형의 상관계수

계수	적합 값	표준오차	t 값
a	8.654832e+001	2.868798e+001	3.016885
t_1	3.867291e+000	3.195135e-001	12.103684
t_2	8.259761e-002	6.912581e-003	11.948882
t_{L1}	7.185113e-002	1.096154e-001	0.655484
r	9.084864e-005	3.560509e-005	2.551563
w	1.681402e+002	2.946521e+001	5.706397
$c1$	5.751300e+000	1.082068e+000	5.315102
$s1$	7.841348e+000	5.472791e-001	14.327878
$c2$	-3.883156e+000	4.687695e-001	-8.283722
$s2$	-4.023365e+000	4.280611e-001	-9.399044
k	-3.984804e+001	2.746946e+001	-1.450630
tr	1.484144e-002	5.839229e-003	2.541679

사용하여 오차를 줄여나가는 연구를 하고자 한다. 또한, 앞으로 정확한 오존과 영향 인자들 간의 관계를 살피기 위해 오존 생성의 전구물질인 NO_x 와 VOC 등을 설명변수로 첨가하고 상층대기(upper air meteorology)와 교호작용 등도 고려해 볼 수 있다. 또한 일반적으로 오존은 전구물질의 발생지역 뿐 아니라 바람으로 인한 수송으로 오염원이 없는 외곽지역에 영향을 주고 지역적 규모(regional scale) 이상의 수평적인 오존 수송을 나타낼 수도 있으므로 향후 오존의 통계적 모형에서 연구의 대상 지역 뿐 아니라 지리적으로 인접한 지역의 오존 및 전구물질의 유·출입을 이해 할 필요가 있겠다.

참고문헌

- Bloomfield, P., Royle, A., Steinberg, L. J. and Yang, Q. (1996). Accounting for meteorological effects in measuring urban ozone levels and trends, *Atmospheric Environment*, **30**, 3067-3078.
- Chambers, J. M. and Hastie, T. J. (1993). *Statistical Models in S*, Chapman & Hall, New York.

- Cox, W. M. and Chu, S. H. (1992). *Meteorologically Adjusted Ozone Trends in Urban Areas: A Probability Approach*, U.S. Environmental Protection Agency, Technical Support Division MD-14, Research Triangle Park, NC 27711.
- Davis, J. M., Brian, K. E. and Bloomfield, P. (1998). Modeling ozone in the Chicago urban area, *Case Studies in Environmental Statistics*, 5-26.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, Chapman & Hall, New York.
- Tukey, J. W. (1977). *Exploratory Data Analysis*, Addison Wesley, Reading, Massachusetts.

Modeling of High Density of Ozone in Seoul Area with Non-Linear Regression

Soo-Yeon Chung¹ · Kiheon Cho²

¹Department of Information and Statistics, Duksung Women's University;

²Department of Information and Statistics, Duksung Women's University

(Received April 2009; accepted May 2009)

Abstract

While characterized initially as an urban-scale pollutant, ozone has increasingly been recognized as a regional and even global-scale phenomenon. The complexity of environmental data dynamics often requires models covering non-linearity. This study deals with modeling ozone with meteorology in Seoul area. The relationships are used to construct a nonlinear regression model relating ozone to meteorology. The model can be used to estimate that part of the trend in ozone levels that cannot be accounted for by trends in meteorology.

Keywords: Long term trend, non-linear regression, ozone, seasonal trend.

²Corresponding author: Professor, Department of Information and Statistics, Duksung Women's University, 419 Ssangmun-Dong, Tobong-Ku, Seoul 132-714, Korea. E-mail: khchoi@duksung.ac.kr