

경기도 안양시 오존농도의 시계열모형 연구

Analysis of Time Series Models for Ozone Concentration at Anyang City of Gyeonggi-Do in Korea

이 훈 자*

평택대학교 디지털응용정보학과

(2008년 9월 18일 접수, 2008년 10월 20일 채택)

Hoonja Lee*

Department of Information Statistics, Pyeongtaek University

(Received 18 September 2008, accepted 20 October 2008)

Abstract

The ozone concentration is one of the important environmental issue for measurement of the atmospheric condition of the country. This study focuses on applying the Autoregressive Error (ARE) model for analyzing the ozone data at middle part of the Gyeonggi-Do, Anyang monitoring site in Korea. In the ARE model, eight meteorological variables and four pollution variables are used as the explanatory variables. The eight meteorological variables are daily maximum temperature, wind speed, amount of cloud, global radiation, relative humidity, rainfall, dew point temperature, and water vapor pressure. The four air pollution variables are sulfur dioxide (SO₂), nitrogen dioxide (NO₂), carbon monoxide (CO), and particulate matter 10 (PM10). The result shows that ARE models both overall and monthly data are suited for describing the ozone concentration. In the ARE model for overall ozone data, ozone concentration can be explained about 71% by the PM10, global radiation and wind speed. Also the four types of ARE models for high level of ozone data (over 80 ppb) have been analyzed. In the best ARE model for high level of ozone data, ozone can be explained about 96% by the PM10, daily maximum temperature, and cloud amount.

Key words : Ozone, High level of ozone, Autoregressive error (ARE) model, Anyang City, Meteorological variables, Pollution variables, Explanatory variables

1. 서 론

지표오존(O₃)의 고농도는 농작물이나 식물들의 피해뿐 아니라 인간의 건강에도 피해를 주기 때문에

이와 관련된 연구는 매우 중요한 것으로 알려지고 있다.

지금까지 오존농도의 예측 모형에 관해 연구되어 왔다. 오존농도 모형 설정은 크게 다음의 4가지 범주로 나뉘질 수 있다. Box-Jenkins methods, transfer function method, regression techniques, neural networks이다.

*Corresponding author.
Tel : +82-(0)31-659-8353, E-mail : esther@ptu.ac.kr

지표오존농도 모형에 관련된 국내연구의 경향을 살펴보면, wavelet transform을 이용한 단기오존농도 예측에 관한 연구(김신도, 1998), 전이함수모델 개발에 관한 연구(김유근 등, 1999), 기상요인들의 교차상관 관계를 이용해 서울시의 오존농도 모형을 연구(전의찬과 우정현, 1999), 울산지역에 관해 군집분석을 이용한 중회귀모형 설계 및 평가(김유근 등, 2007), 경기도 남부지역 지표오존농도의 시계열모형 연구(이훈자, 2007) 등 다양한 방법으로 지표오존농도 모형에 관해 연구하였다.

국외의 경우를 보면, 칠레 Santiago 지역의 오존 일최고농도 예측을 위해 세 가지 모형인 시계열모형, 신경망모형, fuzzy모형을 사용해 비교 분석하였고(Jorquera *et al.*, 1998), 오스트리아 Vienna지역의 오존농도의 단기예측을 위해 시계열모형인 ARX (Autoregressive Exogenous)모형을 사용하였고(Bauer *et al.*, 2001), 미국 동부와 캐나다 남부지역의 오존농도 예측을 위해 동적회귀모형(dynamic regression model)을 사용하였고(Pagowski *et al.*, 2006), 미국 Kentucky 지역의 오존 예측을 위해 비선형 회귀분석과 fuzzy모형을 사용하였다(Lin and Cobourn, 2006).

본 연구에서는, 경기 중부지역 안양시의 2003년부터 2007년까지의 5월 1일부터 9월 30일까지 오존농도 자료를 이용하여 시계열 모형인 ARE (Autoregressive Error)모형을 구축하고자 한다. 정확한 오존농도의 분석과 예측을 위해서는 오존과 연관된 대기 및 기상자료를 포함시켜야 한다. 본 논문에서는 오존농도와 연관이 있는 8종류의 기상자료와 4종류의 대기 자료를 설명변수로 포함한 다변량 분석인 ARE 방법으로 분석하고자 한다.

2. 오존자료 및 설명변수 자료

오존자료는 국립환경과학원의 안양시 호계동의 시간별 자료(하루 24번)를 이용했고 분석을 위하여 일최고 1시간 오존농도를 사용했다.

기간은 2003년부터 2007년까지의 5월 1일부터 9월 30일까지 자료를 사용했다. 본 연구에서 5월에서 9월까지 자료를 사용한 이유는 1년에 걸쳐 이 기간의 오존농도가 높기 때문이고 그래서 이 기간이 중요하기 때문에 분석하였다.

Table 1. List of air pollution variables and meteorological variables.

Variable	Contents and timing	Unit	Notation
O ₃₋₁	Previous day's maximum 1 hour O ₃	ppb	Y _{t-1}
SO ₂	Previous day's maximum 1 hour SO ₂	ppb	S _{t-1}
NO ₂	Previous day's maximum 1 hour NO ₂	ppb	N _{t-1}
CO	Previous day's maximum 1 hour CO	0.1 ppm	I _{t-1}
PM10	Previous day's maximum 1 hour PM10	µg/m ³	P _{t-1}
Temperature	Daily maximum 1 hour surface temperature	°C	T _t
Wind	Average wind speed (from 09 hour to 15 hour)	m/s	W _t
Cloud	Average cloud amount (from 09 hour to 15 hour)	-	C _t
Radiation	Sum of radiation (from 09 hour to 15 hour)	MJ/M ²	R _t
Humidity	Average relative humidity (from 09 hour to 13 hour)	%	H _t
Rainfall	Sum of precipitation (from 01 hour to 24 hour)	mm	A _t
Dew	Average dew point temperature (from 09 hour to 15 hour)	°C	D _t
Waterp	Average water vapor pressure (from 09 hour to 15 hour)	hPa	V _t

설명변수 중 기상 자료는 시간별 자료로 안양에서 가장 가까운 곳에 위치한 기상청 수원지점의 기상 자료를 사용하였다. 기상자료도 오존과 같이 2003년 ~ 2007년 5월부터 9월까지의 자료이다. 8종류의 기상 자료는 김유근 등(1999)을 참고로 하여 일최고온도(maximum temperature), 일사량(global radiation), 풍속(wind speed), 상대습도(relative humidity), 강수량(rainfall), 이슬점온도(dew point temperature), 운량(amount of cloud), 수증기압(water vapor pressure)을 사용했으며 표 1에 나타나 있다. 풍향(wind direction)은 정량화의 어려움이 있어 설명변수에 포함하지 않았다.

설명변수인 4종류의 대기 자료로는 이산화황(SO₂), 이산화질소(NO₂), 일산화탄소(CO), PM10을 사용했고, 오존자료와 같이 국립환경과학원의 안양지역 관측소를 이용하여, 2003년~2007년의 5월 1일~9월 30일의 시간별 자료를 사용하였다. 사용된 대기 자료

도 표 1에 나타나 있다. 8종류의 기상 자료와 4종류의 대기 자료의 시간별 선택은 각 자료를 시간별 혹은 일별로 비교해 오존과의 상관관계가 가장 높은 시간대를 사용하였다. 예를 들면, 일사량(R_i) 변수는 오존과 제일 상관관계가 높은 일별 9시~15시의 합을 사용하였다. 각 설명변수의 구체적인 시간대는 표 1에 나타나 있다. 표 1의 Notation은 4장과 5장의 ARE모형식의 변수 표기이다.

3. 오존농도의 ARE모형과 적합성

ARE모형은 transfer function model의 특수한 경우로 설명변수를 분석에 사용할 수 있는 모형으로 시계열자료를 회귀모형에 적합시킬 때 적절한 모형이다.

일반적인 k차 ARE모형은 다음과 같다.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp} + \varepsilon_t, \quad t=1, 2, \dots, n, \text{ 이며}$$

$$\varepsilon_t = \varepsilon_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_k \varepsilon_{t-k} \quad (1)$$

이때 ε_t 는 서로 독립이며 $N(0, \sigma^2)$ 인 정규분포 형태를 갖는다고 가정한다.

식(1)에서 Y_t 는 반응변수인 일 최고 오존농도를 나타내며, $X_{tj}, j=1, 2, \dots, p$ 는 Y_t 에 영향을 주는 설명변수인 $NO_2, PM10$ 와 같은 대기변수와 일사량, 강수량 등의 기상변수들을 나타낸다. β_j 는 설명변수 X_{tj} 의 계수로 SAS 통계 패키지를 이용하여 구한다. ARE모형과 일반 회귀모형과의 차이점은 오차항 ε_t 가 서로 독립이 아니라 시차에 따른 자기 상관관계를 갖고 있으며 특히 AR (Autoregressive) (k) 모형의 형태를 따른다. 따라서 식(1)의 모형을 ARE (자기회귀오차) 모형이라 부르며 시계열자료에 회귀모형을 적합시킬 경우 많이 사용된다. 오차항의 계수 $\phi_i, i=1, 2, \dots, k$ 도 SAS 통계패키지를 이용하여 구한다.

오존농도의 ARE모형에 관한 적합성을 측정하는데 있어 결정계수와 잔차 분석의 RMSE를 이용하였다. 결정계수 R^2 은 오존 농도 Y_t 의 설명변수들에 의한 설명력으로, 값이 0과 1 사이에 있으며 1에 가까울수록 설명력이 높은 좋은 모형이다. 예측치와 실측치의 분산정도를 나타내는 RMSE (root mean square error)의 공식은 식(2)와 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2}, \quad (2)$$

o_i 와 p_i 는 각각 i 일의 관측 오존량과 추정 오존량을 의미하며, n 은 통계량에 사용된 일수(day)를 의미한다. RMSE는 값이 작을수록 좋은 모형이다.

4. 안양 오존농도의 ARE모형

안양 지역의 오존농도를 연도별, 월별로 상자그림(box plot) 그림 1을 통해 살펴보았다. 먼저, 연도별로 살펴보면 중앙값(상자그림의 가운데 직선)은 2006년이 제일 낮고, 2003년이 제일 높은 것으로 나타났으며, 2007년과 2003년에는 사분위 범위(IQR)가 넓게 퍼져 있다. 참고로, 상자그림에서 상자는 첫째 사분위수(Q_1)에서 시작하여 셋째 사분위수(Q_3)까지 나타내

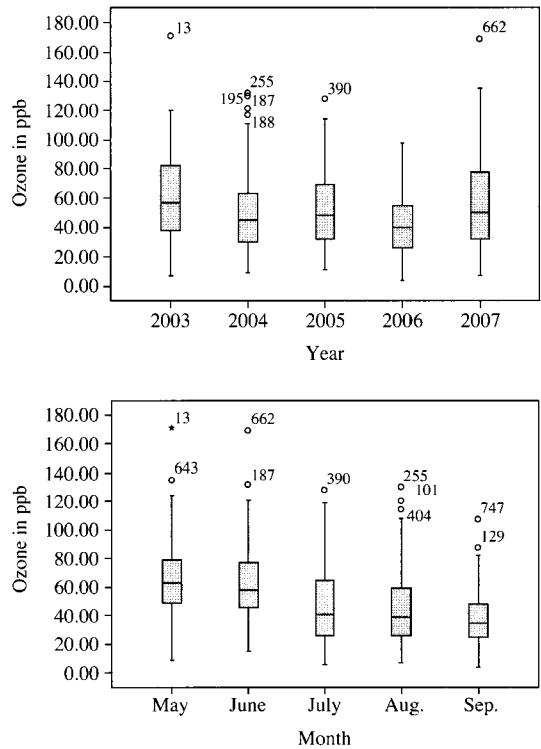


Fig. 1. The box plots of (yearly and monthly) ozone concentrations in Anyang City of Gyeonggi-Do. The circle indicates outliers of the ozone concentration.

Table 2. The ARE models for ozone concentrations.

Period	ARE model	R ²	RMSE
Modeling (n=612 days)	$Y_t=0.291Y_{t-1}+0.084P_{t-1}+2.631R_t-2.468W_t+\varepsilon_t$, $\varepsilon_t=0.113\varepsilon_{t-5}+0.114\varepsilon_{t-8}+0.132\varepsilon_{t-12}+0.127\varepsilon_{t-18}+0.128\varepsilon_{t-19}+\varepsilon_t$	0.714	17.428
Verification (n=153 days)	$Y_t=0.317Y_{t-1}+0.091P_{t-1}+3.345R_t-4.467W_t+\varepsilon_t$, $\varepsilon_t=0.268\varepsilon_{t-19}+\varepsilon_t$	0.870	18.745

며 상자의 폭이 사분위 범위이다. 상자그림에서 상단과 하단의 직선은 양쪽 안올타리(inner face, 허명희와 문승호, 2006) 값의 안쪽에 있으면서 경계에 가장 인접한 자료를 찾아 상자 양쪽을 연결한 것이다.

2006년을 제외한 각 년도마다 오존 농도가 120 ppb보다 넘는 이상치(outlier)가 1일~4일 정도씩 나타났다. 이상치란 상자그림에서 작은 circle, “○”로 표시되며 정상범위에서 많이 떨어진 관측값을 의미한다. 오존농도를 월별로 살펴보면 5, 6월이 7, 8, 9월에 비해 오존농도가 높게 나타났다. 월별로도 9월을 제외하고는 120 ppb보다 넘는 이상치가 1일~3일 씩 나타났다.

4.1 오존 전체농도자료에 관한 ARE모형

오존농도 연구기간인 2003년부터 2007년의 5월부터 9월까지의 자료 중, 2003년 5월부터 2006년 9월까지 612일 자료를 적합한 모델 구축을 위한 모형 구축기간으로, 2007년 5월부터 9월까지 153일을 검증기간으로 사용하였다. 오존농도 자료 분석을 위하여, 12종류의 설명변수들 간의 다중공선성(multi-collinearity)을 분석하고, 오존농도와 영향력이 제일 높은 각 설명변수들의 시차(time-lag)를 고려한 결과 표 2와 같은 설명변수를 포함한 ARE모형이 선택되었다. 표 2에서 Y_t 는 t일(day) 최고 오존농도, Y_{t-1} 는 t-1일 최고 오존농도, P_{t-1} 는 t-1일 최고 PM10 농도, R_t 는 t일 9시~15시까지의 일사량의 합, W_t 는 t일의 9시~15시의 평균 풍속 등이며 기타 설명변수에 관한 해석은 표 1을 참고하기 바란다. 모형구축기간의 결정계수 R^2 는 0.714로 오존농도 Y_t 는 P_{t-1} (PM10), R_t (일사량)와 W_t (풍속)에 의해 71.4% 설명되고 RMSE는 17.428 ppb로 나타났다. 설명변수 중 P_{t-1} 와 R_t 의 계수는 양수로 W_t 의 계수는 음수로 나타나 PM10과 일사량이 증가하면 오존농도는 증가하는 반면, 풍속이 증가하면 오존 농도는 감소하는 것으로 나타났다. 검증기간의 R^2 는 0.870으로 오존 농도 Y_t 의 설명력

이 매우 높게 나타난 반면 RMSE는 18.745 ppb로 모형구축기간보다 조금 높게 나타났다.

4.2 월별 오존농도 자료에 관한 ARE모형

그림 1의 상자그림에서 월별로 오존농도가 다르게 나타나, 월별로 오존농도 모형을 따로 연구할 필요성을 느껴 분석한 결과를 표 3에 나타냈다. 월별 오존 분석은 2003년~2005년의 3개월 자료를 모형구축기간으로, 2006년~2007년의 2개월 자료를 검증기간으로 사용하였다. 표 3을 보면, 오존을 설명하는 변수는 5월, 8월, 9월에는 T_t (일 최고온도)와 C_t (운량)으로 나타났고, 6월에는 N_{t-1} (이산화질소)와 R_t (일사량), 7월에는 P_{t-1} (PM10)와 R_t (일사량)로 월별로 조금씩 다르게 나타났다. 결정계수 R^2 는 5월 검증기간에는 0.305로 매우 낮게 나왔지만, 그 외의 기간에는 0.70 이상으로 오존이 설명변수에 의해 잘 설명되었다. RMSE 값은 6월의 검증기간에 20.339 ppb로 조금 높게 나타났고 그 외의 기간은 20 ppb 이하로 나타나 모형은 적절한 것으로 분석되었다.

5. 오존농도 80 ppb 이상 자료에 관한 ARE모형

오존에 관한 주요 관심사 중의 하나는 고농도에 관한 분석이다. 최근 들어, 환경기준치 및 주의보 수준을 초과하는 오존농도가 수도권에 집중되어 빈도 높게 나타나고 있다(환경부, 2003). 이는 수도권 지역의 오존농도에 관한 실태 및 오존농도에 영향을 끼치는 요인들에 관한 연구가 이루어져야 함을 시사한다.

오존의 고농도에 대한 정의는 학자마다 차이가 있다. Lyons *et al.* (1991)과 Ludwig *et al.* (1995)에 의하면 NAAQS (National Ambient Air Quality Standard) 기준으로 1시간 동안 120 ppb를 초과하면 오존 고농도

Table 3. The monthly ARE models for ozone concentrations.

Month	Period	ARE model	R ²	RMSE
May	Modeling (n=93 days)	$Y_t = 3.333T_{t-1} - 2.015C_t + \epsilon_t$ $\epsilon_t = 0.281\epsilon_{t-1} + e_t$	0.882	18.800
	Verification (n=62 days)	$Y_t = 1.572T_{t-1} - 1.228C_t + \epsilon_t$ $\epsilon_t = 0.368\epsilon_{t-1} - 0.604\epsilon_{t-4} + e_t$	0.305	15.176
June	Modeling (n=90 days)	$Y_t = 0.389N_{t-1} + 3.301R_t + \epsilon_t$ $\epsilon_t = 0.401\epsilon_{t-1} + e_t$	0.861	17.264
	Verification (n=60 days)	$Y_t = 0.510N_{t-1} + 2.133R_t + \epsilon_t$ $\epsilon_t = 0.378\epsilon_{t-8} + e_t$	0.819	20.339
July	Modeling (n=93 days)	$Y_t = 0.227P_{t-1} + 2.943R_t + \epsilon_t$ $\epsilon_t = 0.252\epsilon_{t-1} - 0.300\epsilon_{t-2} + e_t$	0.909	18.223
	Verification (n=62 days)	$Y_t = 0.074P_{t-1} + 3.546R_t + \epsilon_t$ $\epsilon_t = 0.433\epsilon_{t-1} + e_t$	0.741	17.430
August	Modeling (n=93 days)	$Y_t = 0.295T_t - 2.780C_t + \epsilon_t$ $\epsilon_t = 0.421\epsilon_{t-1} + e_t$	0.708	19.598
	Verification (n=62 days)	$Y_t = 2.174T_t - 3.845C_t + \epsilon_t$ $\epsilon_t = 0.395\epsilon_{t-1} + e_t$	0.733	17.622
September	Modeling (n=90 days)	$Y_t = 2.218T_t - 2.661C_t + \epsilon_t$ $\epsilon_t = 0.280\epsilon_{t-2} + e_t$	0.816	15.277
	Verification (n=60 days)	$Y_t = 1.661T_t - 0.891C_t + \epsilon_t$ $\epsilon_t = 0.273\epsilon_{t-1} - 0.272\epsilon_{t-8} - 0.305\epsilon_{t-16} + e_t$	0.905	15.132

Table 4. Yearly & monthly days and average of high level of ozone concentrations (over 80 ppb).

	2003	2004	2005	2006	2007	Monthly total
May	17 (102.74)	2 (98.50)	6 (91.83)	0	12 (97.50)	37 (99.04)
June	11 (101.64)	7 (108.57)	6 (91.50)	1 (82.00)	10 (101.40)	35 (100.66)
July	8 (105.13)	2 (89.50)	4 (102.75)	0	7 (97.43)	21 (100.62)
Aug.	2 (114.00)	3 (105.67)	4 (99.00)	2 (89.00)	4 (89.75)	15 (98.53)
Sep.	3 (83.67)	0	1 (82.00)	0	2 (94.50)	6 (87.00)
Yearly total	41 (102.06)	14 (103.79)	21 (94.71)	3 (86.67)	35 (97.54)	114 (99.13)

로 정의하고 있다. Bower *et al.* (1994)은 영국에서 두 세 지역에서 일 평균 오존농도가 60 ppb를 넘으면, 고농도 오존일로 정의하고 있다. McKendry (1993)은 캐나다의 두 세 지역에서 시간당 평균 82 ppb를 초과했을 때를 고농도일로 보고 있다.

Poulida *et al.* (1991)은 고농도일을 시간당 농도가 80 ppb 이상으로 3시간 이상 지속된 경우로 정의하고 있다. 우리나라는 한 지점 이상 1시간 환경기준치 100 ppb를 초과하는 날을 오존오염일 (환경부, 2004)로 정의하였다.

고농도 오존 모형에 관련된 국내연구는 광화학모델인 CIT를 이용한 서울 수도권 고농도 오존 모델링

(김진영과 김영성, 2001), 군집분석과 회귀분석을 이용한 서울시 고농도 오존 예측 (허정숙과 김동술, 1993) 등이 있으며 좀 더 다양한 모형을 이용한 연구가 필요하다. 국외의 경우를 보면 비선형회귀모형을 이용하여 Louisville 지역의 고농도 오존 모형을 연구하였고 (Cobourn and Hubbard, 1999), MM5/SMOKE/CMAQ air quality 모형을 이용해 Ontario 지역 고농도 오존을 연구했고 (Brulfer *et al.*, 2007), CMAQ모형을 이용해 Houston 지역의 오존 고농도를 연구하였다 (Byun *et al.*, 2006).

표 4는 안양지역의 2003년 5월에서 2007년 9월까지 일반적인 고농도로 볼 수 있는 80 ppb 이상되는

Table 5. Results of ANOVA and Duncan's multiple range test of the explanatory variables for ozone of the concentrations.

Variable (unit)	ANOVA p-value	Duncan's multiple range test		Variable (unit)	ANOVA p-value	Duncan's multiple range test	
		Grouping	Mean			Grouping	Mean
NO ₂ (ppb)	0.000	A ¹	49.12	CO (0.1 ppm)	0.000	A ¹	8.30
		B ²	66.29			B ²	10.70
		C ²	69.06			C ²	11.72
SO ₂ (ppb)	0.000	A ¹	7.47	PM10 (µg/m ³)	0.000	A ¹	89.70
		B ²	11.88			B ²	133.04
		C ³	14.21			C ³	158.98
Temperature (°C)	0.000	A ¹	26.32	Rainfall (mm)	0.000	C ¹	0.52
		B ²	28.05			B ¹	0.75
		C ³	29.35			A ²	8.33
Wind (m/s)	0.000	B ¹	2.13	Humidity (%)	0.000	C ¹	52.78
		C ¹	2.22			B ¹	54.89
		A ²	2.69			A ²	66.01
Dew (°C)	0.003	B ¹	14.24	Waterp (hPa)	0.004	B ¹	17.41
		C ^{1,2}	15.07			C ^{1,2}	18.04
		A ²	16.39			A ²	19.69
Radiation (MJ/M ²)	0.000	A ¹	10.75	Cloud	0.000	C ¹	3.23
		B ²	16.03			B ¹	4.14
		C ²	17.06			A ²	7.00

Note: The A, B, C letters means ozone concentrations of [0, 79], [80, 99], [100, ∞] respectively. In 1, 2, 3 letters, the same letters are not significantly different at 0.05 level.

날짜와 평균값을 월별, 연도별로 분석한 표이다. 월별, 연도별 고농도 일수와 고농도 평균값이 조금씩 차이를 보이고 있다. 고농도 일수는 2003년에 41일, 2007년에 35일, 2005년에 21일, 2004년에 14일, 2006년에 3일 순으로 나타났다. 월별로 살펴보면, 5년간 5월에는 37일, 6월에는 35일, 7월에는 21일, 8월에는 15일로 나타나 5월, 6월이 고농도 일수가 많이 나타났고 9월에는 6일로 고농도 일수가 작게 나타났다.

본 연구에서는 일반적 고농도 정의를 고려하여, 안양시의 2003년 5월에서 2007년 9월까지의 오존 농도를 79 ppb 이하, 80 ppb에서 99 ppb 사이, 100 ppb 이상의 3그룹으로 나누었고 이에 따른 일수는 각각 651일, 67일, 47일로 나타났다. 보편적 고농도로 볼 수 있는 80 ppb 이상도 114일로 나타났다.

그룹별 오존 농도에 따라 12종류의 설명변수들의 평균값들이 차이가 있는지를 분산분석(ANOVA)과 Duncan 평균비교 검정 분석을 한 결과를 표 5에 나타냈다. 12개의 모든 설명변수는 유의 수준 1%에서 유의한 차이가 있는 것으로 분석되었다. 특히 설명변

수 중 SO₂, PM10, 최고온도는 오존농도의 3그룹별 모두 유의한 차이가 있는 것으로 나타났고, SO₂, PM10, 최고온도(Temperature)가 각각 증가함에 따라 오존농도도 증가하는 것으로 분석되었다.

강수량(Rainfall), 풍속(Wind), 습도(Humidity), 운량(Cloud), 일사량(Radiation)은 2그룹 [80, 99], [100, ∞]은 유의한 차이가 없고, [0, 79]그룹과는 유의한 차이가 있는 것으로 분석되었고 강수량, 풍속, 습도, 운량이 증가함에 따라 오존농도가 감소하는 것으로 분석된 반면, 일사량이 증가하면, 오존농도도 함께 증가하는 것으로 분석되었다.

일반적 고농도인 일 최고 농도가 80 ppb 이상 되는 자료에 관한 ARE모형식을 표 6의 General ARE, HL ARE, PM10 add ARE, Temperature add ARE, Radiation add ARE의 5가지 방법으로 분석하였다. 첫 번째 방법인 80 ppb 이상되는 자료만 분석한 모형을 General ARE모형이라 명하였다. 이 방법으로 오존농도 Y_t는 P, T, C_t에 의해 96.3% 설명되어 설명력이 매우 높다.

Table 6. The five ARE models for high level ozone concentrations (over 80 ppb).

Method	Model	R ²	RMSE
General ARE (n=114)	$Y_t=0.106P_t+3.036T_t-0.934C_t+\varepsilon_t$ $\varepsilon_t=0.198\varepsilon_{t-1}+e_t$	0.963	15.891
HL ARE (n=227)	$Y_t=0.160P_t+4.536R_t-4.676W_t+\varepsilon_t$ $\varepsilon_t=0.235\varepsilon_{t-1}+0.144\varepsilon_{t-7}+0.144\varepsilon_{t-11}+e_t$	0.806	20.727
PM10 add ARE (n=221)	$Y_t=0.405N_t+3.602R_t-1.489W_t+\varepsilon_t$ $\varepsilon_t=0.214\varepsilon_{t-1}+0.152\varepsilon_{t-2}+0.185\varepsilon_{t-4}+0.173\varepsilon_{t-7}-0.135\varepsilon_{t-13}+e_t$	0.761	20.693
Temperature add ARE (n=331)	$Y_t=3.081I_t+3.697R_t-7.568W_t+\varepsilon_t$ $\varepsilon_t=0.362\varepsilon_{t-1}+0.142\varepsilon_{t-2}+0.136\varepsilon_{t-15}+e_t$	0.693	19.452
Radiation add ARE (n=247)	$Y_t=0.215P_t+2.648T_t-9.817W_t+\varepsilon_t$ $\varepsilon_t=0.226\varepsilon_{t-1}+0.230\varepsilon_{t-2}+0.173\varepsilon_{t-8}+0.134\varepsilon_{t-14}+e_t$	0.680	17.781

고농도 자료에 관한 2번째 모형으로, Hubbard and Cobourn (1998)이 제안한 HL (High-Low)모형을 생각할 수 있다. Hubbard 등은 미국 Kentucky주의 Louisville 지역의 고농도 오존 모형을 오존농도 100 ppb 이상과 50 ppb 이하의 자료를 회귀분석으로 분석하였다. HL 모형은 오존과 설명변수들의 분산이 매우 큰 단점이 있는 반면에 추정계수들이 오존농도 분포의 중간값으로 가려는 성향을 감소시킬 수 있는 장점과 오존 고농도와 저농도 간의 대비가 뚜렷해 오존 고농도에 영향을 주는 설명변수들이 모형에 포함될 가능성이 높은 특성이 있다. 본 논문에서는 HL 모형에서 고농도와 저농도의 자료수를 동일하게 사용하여 ARE 방법으로 분석하였다.

80 ppb 이상 되는 고농도가 114일로 나타나 대칭되는 저농도도 24 ppb 이하를 포함시켜 모두 227개의 자료로 분석한 결과 P_t , R_t , W_t 가 오존의 영향을 주는 설명변수로 분석되었다.

고농도 자료에 관한 3번째에서 5번째 방법으로 표 5의 설명변수들의 ANOVA 및 Duncan 비교검정 결과를 이용하여 설명변수 중 특히 오존에 영향을 많이 주는 PM10, 최고온도, 일사량을 고려한 add ARE 모형을 분석하였다.

표 5의 Duncan 비교검정 결과, PM10은 3그룹 간에 유의한 것으로 나타났고 PM10 농도가 $130 \mu\text{g}/\text{m}^3$ 이상이면 오존농도가 80 ppb 이상인 고농도가 될 가능성이 있다고 사려되어 오존농도가 80 ppb 이상이거나 PM10 농도가 $130 \mu\text{g}/\text{m}^3$ 이상인 자료를 함께 분석하고 이를 PM10 add ARE모형이라 명하였다. 표 6의 결과를 보면, 오존농도는 N_t , R_t , W_t 에 의해 76.1%

설명되는 것으로 나타났다.

최고온도는 28° 이상인 자료를 포함시켰고, 이를 Temperature add ARE모형이라 명하였다. 오존농도는 I_t , R_t , W_t 에 의해 69.3% 설명되는 것으로 나타났다.

일사량은 $16 \text{ MJ}/\text{M}^3$ 이상인 자료를 포함시켜 이를 Radiation add ARE모형이라 명하였다. 오존농도는 P_t , T_t , W_t 에 의해 68% 설명되는 것으로 나타났다.

오존 고농도에 영향을 주는 설명변수는 5가지 ARE모형에 대해 조금씩 다르게 나타났고, 적합성 검정인 결정계수 R^2 와 RMSE를 비교했을 때, General ARE모형이 제일 적합한 것으로 나타났다. 고농도 오존의 특성을 명확히 보기 위해 저농도 오존을 포함한 HL ARE모형도 R^2 는 0.806, RMSE는 20.727 ppb로 비교적 적절한 모형으로 나타났다. 또한 오존 고농도에 영향을 많이 주는 설명변수인 PM10 add 모형도 R^2 는 0.761, RMSE는 20.693 ppb로 비교적 적절하게 나타났으나, Temperature add ARE와 Radiation add ARE모형은 RMSE는 적절하나, R^2 가 조금 낮게 나왔다.

6. 결 론

본 연구에서는 경기 중부지역인 안양지역의 오존농도를 시계열모형인 ARE모형으로 분석하였다. ARE 모형은 설명변수를 분석에 사용할 수 있는 모형으로 시계열자료를 회귀모형에 적합시킬 때 적절한 모형이다.

본 연구의 ARE모형에서는, 오존농도에 영향을 주

는 8종류의 기상 자료와 4종류의 대기 자료를 설명 변수로 사용하였다. 8종류의 기상 자료는 일 최고온도, 일사량, 풍속, 상대습도, 강수량, 이슬점온도, 운량, 수증기압이며 4종류의 대기 자료는 SO₂, NO₂, CO, PM10이다. 연구기간인 2003년부터 2007년의 5월부터 9월까지의 자료 중, 2003년에서 2006년의 자료 612일을 모형구축기간으로 2006년 자료 153일을 검증기간으로 사용하였다. ARE모형에서 일 최고 오존농도는 PM10농도, 일사량, 풍속에 의해 영향을 받는 것으로 분석되었다.

월별로 오존농도를 분석한 결과 월별로 영향을 주는 설명변수가 조금씩 다르게 나타났다. 5월, 8월, 9월에는 일 최고온도와 운량이 영향을 주는 것으로 나타났고, 6월에는 이산화질소와 일사량이, 7월에는 PM10과 일사량이 영향을 주는 설명변수로 나타났다.

그리고, 오존 연구의 주요 관심사인 고농도에 관한 자료를 분석하였다. 일반적 고농도로 볼 수 있는 일 최고 오존농도가 80 ppb 이상 되는 요일에 대해 5가지 방법의 ARE모형을 이용하여 분석하였다. 분석 결과 80 ppb 이상의 자료만 사용한 General ARE모형이 제일 적합한 것으로 나타났고, 고농도 오존의 특성을 명확히 보기 위해 저농도 오존을 포함한 HL ARE모형도 적합한 것으로 나타났다. 또한 오존 고농도에 영향을 주는 고농도 PM10를 포함한 PM10 add모형도 비교적 적절한 것으로 분석되었다.

ARE모형이 자료를 적용하기가 비교적 간단하기 때문에 오존농도 분석에 ARE모형을 사용하는 것이 좋을 것으로 사려된다. 앞으로, 오존농도에 사용되는 설명변수에 관한 다양한 연구와 함께 고농도 오존에 관한 연구가 지역별로 좀 더 활발히 이루어져야 된다고 생각된다.

참 고 문 헌

김신도(1998) 오존예보모델 및 예보시스템의 개선, 오존예보시스템에 관한 전문가토론회, 16-23.
 김진영, 김영성(2001) 상세한 기상관측 자료를 이용한 1997년 서울 수도권 고농도 오존 사례의 모델링, 한국 대기환경학회지, 17(1), 1-17.
 김유근, 손건태, 문윤섭, 오인보(1999) 서울 지역의 지표오존농도 예보를 위한 전이함수 모델 개발, 한국대기

환경학회지, 15(6), 779-789.
 김유근, 이소영, 임윤구, 송상근(2007) 중회귀모형을 이용한 울산지역 오존 포텐셜 모형의 설계 및 평가, 한국 대기환경학회지, 23(1), 14-28.
 이훈자(2007) 경기도 남부지역 지표오존농도의 시계열모형 연구, 한국대기환경학회지, 23(3), 364-372.
 전의찬, 우정현(1999) 오존 농도에 영향을 미치는 주 기상요소의 도출 및 예측모형 수립, 한국대기환경학회지, 15(3), 257-266.
 환경부(2003) 대기환경연보 2003, 65pp.
 환경부(2004) 대기환경연보 2004년 6월, 60pp.
 허명희, 문승호(2006) 탐색적 자료분석, 자유아카데미.
 허정숙, 김동술(1993) 다변량 통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구, 한국대기보전학회지, 9(3), 207-215.
 Bauer, G., M. Deistler, and W. Scherrer (2001) Time series models for short term forecasting of ozone in the eastern part of Austria, *Environmetrics*, 12, 117-130.
 Bower, J.S., K.J. Stevenson, G.F.J. Broughyon, and J.E. Lampert (1994) Assessing recent surface ozone concentrations in the U.K., *Atmospheric Environment*, 28D, 115-128.
 Brulfert, G., O. Galvez, F. Yang, and J.J. Sloan (2007) A regional modeling study of the high ozone episode of June 2001 in southern Ontario, *Atmospheric Environment*, 41, 3777-3788.
 Byun, D.W., S. Kim, and S. Kim (2006) Evaluation of air quality models for the simulation of a high ozone episode in the Houston metropolitan area, *Atmospheric Environment*, 41, 837-853.
 Cobourn, W. and M.C. Hubbard (1999) An enhanced ozone forecasting model using air mass trajectory analysis, *Atmospheric Environment*, 33, 4663-4674.
 Hubbard, M. and W. Cobourn (1998) Development of a regression model to forecast ground-level ozone concentration in Louisville, KY, *Atmospheric Environment*, 32(14/15), 2637-2647.
 Jorquera, H., R. Perez, A. Cipriano, A. Espejo, M.V. Letelier, and G. Acuna (1998) Forecasting ozone daily maximum levels at Santiago, Chile, *Atmospheric Environment*, 32(20), 3415-3424.
 Lin, Y. and W.G. Cobourn (2006) Fuzzy system models combined with nonlinear regression for daily ground-level ozone predictions, *Atmospheric Environment*, 41(2007), 3502-3513.
 Ludwig, F.L, I.J. Jiang, and J. Chen (1995) Classification of ozone and weather pattern associated with high

- ozone concentrations in the San Francisco and Monterey Bay Areas, *Atmospheric Environment*, 29, 2915-2828.
- Lyons, W.A., J.L. Eastman, R.A. Pielke, C.T. Tremback, P.A. Moon, and K.R. Lincoln (1991) The meteorological of ozone episodes in the lower lake Michigan air quality region, air & waste management association, For presentation at the 83th Annual Meeting & Exhibition, Vancouver, British Columbia, June 16-21.
- McKendry, I.G. (1993) Ground-level ozone in Montreal, Canada, *Atmospheric Environment*, 27B, 93-103.
- Pagowski, M., G.A. Grell, D. Devenyi, S.E. Peckham, S.A. McKeen, W. Gong, L. Delle Monache, J.N. McHenrt, J. McQueen, and P. Lee (2006) Application of dynamic linear Regression to improve the skill of ensemble-based deterministic ozone forecasts, *Atmospheric Environment*, 40(2006), 3240-3250.
- Poulida, O., R.P. Dickerson, B.G. Doddridge, J.Z. Holland, R.G. Wardell, and J.G. Wartjins (1991) Trace gas concentrations and meteorology in rural Virginia: ozone and carbon monoxide, *Journal of Geo-physical Research*, 96, 22461-22475.