

경기도 남부지역 지표오존농도의 시계열모형 연구

Analysis of Time Series Models for Ozone at the Southern Part of Gyeonggi-Do in Korea

이 훈 자*

평택대학교 디지털응용정보학과

(2007년 4월 17일 접수, 2007년 6월 19일 채택)

Hoonja Lee*

Department of Information Statistics, Pyeongtaek University

(Received 17 April 2007, accepted 19 June 2007)

Abstract

The ozone concentration is one of the important environmental issue for measurement of the atmospheric condition of the country. In this article, two time series ARE models, the direct ARE model and applied ARE model have been considered for analyzing the ozone data at southern part of the Gyeonggi-Do, Pyeongtaek, Osan and Suwon monitoring sites in Korea. The result shows that the direct ARE model is better suited for describing the ozone concentration in all three sites. In both of the ARE models, eight meteorological variables and four pollution variables are used as the explanatory variables. Also the high level of ozone data (over 80 ppb) have been analyzed at the Pyeongtaek, Osan and Suwon monitoring sites.

Key words : Autoregressive error (ARE) model, Direct ARE model, Applied ARE model

1. 서 론

지표오존의 고농도 현상은 농작물이나 식물들의 피해뿐 아니라 인간의 건강에도 피해를 주는 것으로 알려지고 있다. 최근 들어, 환경기준치 및 주의보 수준을 초과하는 오존농도가 수도권에 집중되어 빈도 높게 나타나고 있다(환경부, 2003). 지금까지 오존농도의 모형 적합에 관해 연구되어 왔다. Thompson *et al.* (2000)에 따르면 오존농도 모형 설정은 크게 다음의 3가지 범주로 나뉘질 수 있다. 첫째, regres-

sion-based modeling (Feister and Balzer, 1991), 둘째, extreme value approach (Smith and Huang, 1993), 셋째, spatio-temporal (Guttorp *et al.*, 1994)이다.

지표오존농도 모형에 관련된 국내연구의 경향을 살펴보면, 중회귀 예보모델 개발에 관한 연구(김용준, 1997), wavelet transform을 이용한 단기오존농도 예측에 관한 연구(김신도, 1998), 전이함수모델 개발에 관한 연구(김유근 등, 1999), 군집분석을 이용한 중회귀모형 설계 및 평가(김유근 등, 2007) 등 다양한 방법으로 지표오존농도 모형에 관해 연구하였다.

국외의 경우를 보면, 미국 켄터키주의 Louisville 지역 오존 일 최고농도 예측을 위해 비선형 다중회귀분석을 사용했고(Hubbard *et al.*, 1998), 칠레 San-

*Corresponding author.

Tel : +82-(0)11-296-8710, E-mail : esther@ptu.ac.kr

tiago 지역의 오존 일 최고농도 예측을 위해 세 가지 모형인 시계열모형, 신경망모형, fuzzy 모형을 사용해 비교 분석하였고(Jorquera *et al.*, 1998), 오스트리아 Vienna지역의 오존농도의 단기예측을 위해 시계열모형인 ARX (Autoregressive Exogenous) 모형을 사용하였다(Bauer *et al.*, 2001).

본 연구에서는, 경기 남부지역의 평택시, 오산시, 수원시의 2003년부터 2005년까지의 5월 1일부터 9월 30일까지 오존농도 자료를 이용하여 시계열 모형인 ARE 모형을 구축하고자 한다.

ARIMA (Autoregressive Integrated Moving Average)모형은 시계열분석에서 자주 사용된다. ARIMA 모형은 다양한 종류의 일변량 시계열자료를 분석할 수 있고 미래의 값을 쉽게 예측할 수 있다. 반면에 분석하고자 하는 자료와 연관된 설명변수를 포함 시킬 수 없는 단점이 있다. 예를 들면, 오존자료 분석에 오존과 연관성이 많은 온도나 이산화질소 같은 설명변수를 포함 시킬 수 없다. 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 (Autoregressive Error) 방법으로 분석하고자 한다.

2. 오존자료 및 설명변수 자료

오존자료는 국립환경과학원의 평택시 비전동, 오산시 오산동, 수원시 영통동 관측소의 시간별 자료(하루 24번)를 이용했고 분석을 위하여 일 최고 1시간 오존농도를 사용했다.

기간은 2003년부터 2005년까지의 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.

설명변수 중 기상자료는 시간별 자료로 평택, 오산, 수원지역 모두 기상청의 수원 지점 자료를 사용하였다. 수원지점은 평택과 오산에서 가장 가까운 곳에 위치한 기상청 지점이다. 기상자료도 오존과 같이 2003~2005년 5월부터 9월까지의 자료이다. 8종류의 기상자료는 김유근 등(1999)을 참고로 하여 일 최고

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	Daily maximum 1 hour SO ₂	ppb	S _t
NO ₂	Daily maximum 1 hour NO ₂	ppb	N _t
CO	Daily maximum 1 hour CO	0.1 ppm	O _t
PM10	Daily maximum 1 hour PM10	Ug/m ³	P _t
Temperature	Daily maximum 1 hour surface temperature	°C	T _t
Wind	Average wind speed (from 06 hour to 18 hour)	m/s	W _t
Cloud	Average cloud amount (from 06 hour to 18 hour)	-	C _t
Radiation	Sum of radiation (from 05 hour to 20 hour)	MJ/M ²	R _t
Humidity	Average relative humidity (from 09 hour to 15 hour)	%	H _t
Rainfall	Sum of precipitation (from 01 hour to 24 hour)	mm	A _t
Dew	Dew point temperature at 15 hour	°C	D _t
Waterp	Water vapor pressure at 15 hour	hPa	V _t

온도 (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.

설명변수인 4종류의 대기 자료로는 이산화황 (SO₂), 이산화질소 (NO₂), 일산화탄소 (CO), PM10를 사용했고, 오존자료와 같이 국립환경과학원의 평택, 오산, 수원지역 관측소를 이용하여, 2003~2005년의 5월 1일~9월 30일의 시간별 자료를 사용하였다. 사용된 대기 자료도 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_t) 변수는 오존과 제일 상관관계가 높은 일별 5~20시의 합을 사용하였다. 각 설명변수의 구체적인 시간대는 표 1에 나타나 있다. 표 1의 Notation은 4장과 6장의 ARE 모형식의 변수 표기이다.

3. 오존농도의 ARE 모형

반응변수에 영향을 주는 설명변수가 있을 때는 설명변수를 분석에 이용하는 것이 효율적이다. ARE 모형은 설명변수를 분석에 사용 할 수 있는 모형으로 시계열자료를 회귀모형에 적합 시킬 때 적절한 모형이다. 반응변수인 오존 농도에 영향을 주는 이산화질소(NO₂)나 PM10 같은 대기자료와 일별 최고온도나 강수량과 같은 기상자료를 사용하여 자료를 분석한다.

ARE 모형은 transfer function model의 특수한 경우로, 오차항 ϵ_t 가 독립이 아니라 서로 상관관계를 갖게 되며, 특히 AR(Autoregressive) 형태를 갖는다. 일반적인 k차 ARE 모형은 다음과 같다.

$$\begin{aligned}
 Y_t &= \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp} + \epsilon_t, \\
 t &= 1, 2, \dots, n \text{ 이며} \\
 \epsilon_t &= c_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2} - \dots - \phi_k \epsilon_{t-k} \text{ 이고} \\
 \epsilon_t &\sim \text{i.i.d } N(0, \sigma^2).
 \end{aligned}
 \tag{1}$$

위의 모형과 일반 회귀모형과의 차이점은 식(1)에서 오차항 ϵ_t 가 서로 독립이 아니라 시차에 따른 자기상관관계를 갖고 있으며 특히 AR(k) 모형의 형태를 따른다는 점이다. 따라서 식(1)의 모형을 ARE(자기회귀오차)모형이라 부르며 시계열자료에 회귀모형을 적합 시킬 경우 많이 사용된다. 또한 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 대기변수와 설명변수들을 나타낸다.

4. 경기 남부 오존농도의 ARE 모형

경기 남부지역인 평택, 오산, 수원 지역의 오존농도를 연도별, 월별로 상자그림(box plot) 그림 1을 통해 살펴보았다. 먼저, 연도별로 살펴보면 평택지역은 2003년에서 2005년까지의 중앙값은 거의 비슷하고 2004년에 농도 오존이 120ppb보다 넘는 이상치(outlier)가 며칠 나타났다. 이상치란 정상범위에서 아주 많이 떨어진 관측값을 의미한다. 오산지역은 오존농도 중앙값이 2004년도가 좀 낮게 나타났고 작게 퍼져 있는 반면, 2003년에는 사분위수범위(IQR)가

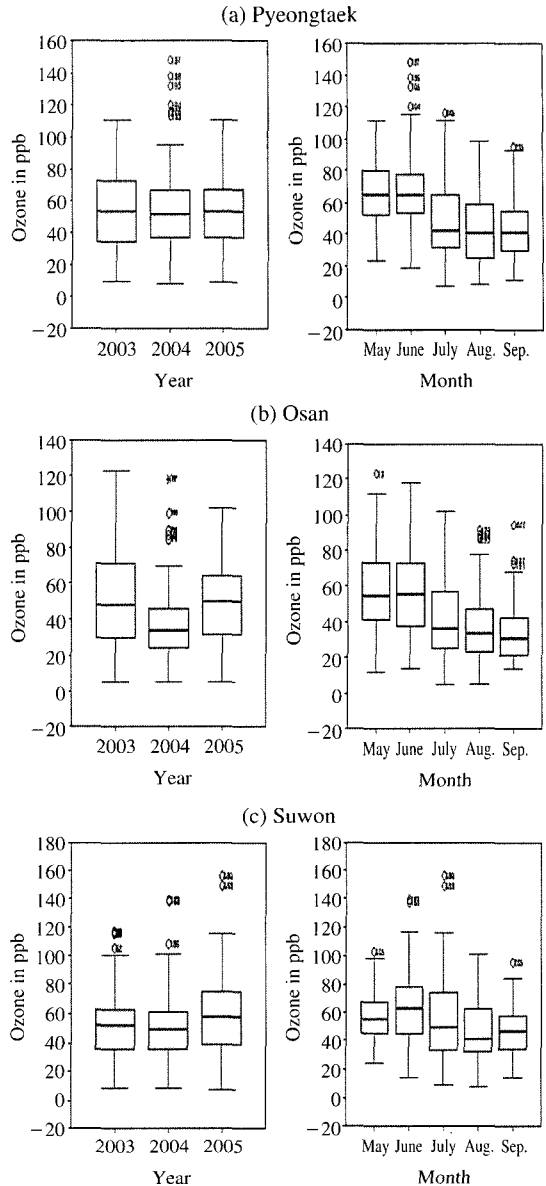


Fig. 1. The box plots of (yearly and monthly) ozone concentrations at southern part of Gyeonggi-do. The circle indicates outliers of the ozone concentration.

넓게 퍼져 있다. 수원지역은 2005년도 중앙값이 조금 높게 나타났고 각 년도 마다 100ppb가 넘는 고농도 오존일이 2, 3일씩 나타났다. 오존농도를 지역별 월별로 살펴보면, 평택과 오산, 수원지역 모두 5, 6

Table 2. The direct ARE model for ozone concentration.

Area	Direct model
Pyeongtaek	$Y_t = 0.148Y_{t-1} + 0.230P_t + 0.017R_t - 0.146W_t + \epsilon_t$ $\epsilon_t = 0.124\epsilon_{t-5} + e_t$
Osan	$Y_t = 0.111Y_{t-1} + 0.411N_t + 0.070P_t + 0.013R_t$ $- 0.383C_t + \epsilon_t$ $\epsilon_t = 0.186\epsilon_{t-2} + 0.275\epsilon_{t-5} + 0.188\epsilon_{t-14} + e_t$
Suwon	$Y_t = 0.290Y_{t-1} + 1.472S_t + 0.138T_t - 0.388W_t + \epsilon_t$ $\epsilon_t = 0.147\epsilon_{t-4} + 0.172\epsilon_{t-5} + 0.116\epsilon_{t-22} + e_t$

월이 7, 8, 9월에 비해 오존농도가 높게 나타났다. 최고온도가 5, 6월이 7, 8, 9월 보다 낮음에도 불구하고, 오존 농도가 높게 나타난 이유는 강수량을 포함한 설명변수에서 분석될 것으로 생각된다.

4.1 경기 남부 오존농도자료에 관한 direct ARE 모형

경기 남부 지역인 평택, 오산, 수원지역 각각의 오존농도자료 분석을 위하여, 12종류의 설명변수들 간의 다중공선성 (multi-collinearity)를 분석하고, 오존농도와 영향력이 제일 높은 각 설명변수들의 시차 (time-lag)를 고려한 결과 표 2와 같은 설명변수를 포함한 direct ARE모형이 선택되었다. 표 2에서 Y_t 는 t일 (day) 최고 오존농도, Y_{t-1} 는 t-1일 최고 오존농도, P_t 는 t일 최고 PM10 농도, R_t 는 t일 5~20시까지의 일사량의 합, W_t 는 t일의 6시~18시의 평균 풍속 등이며 기타 설명변수에 관한 해석은 표 1을 참고하기 바란다. 지역별로 영향을 주는 대기변수와 기상변수가 조금씩 다르게 나타났다. 평택지역의 오존농도 Y_t 는 P_t (PM10), R_t (일사량) 와 W_t (풍속)에 의해 영향을 받고, 오산지역의 오존농도 Y_t 는 N_t (이산화질소), P_t (PM10), R_t (일사량)와 C_t (운량)에 의해 영향을 받고, 수원지역 오존농도 Y_t 는 S_t (이산화황), T_t (온도)와 W_t (풍속)에 의해 영향을 받는 것으로 분석되었다. 특히 기상변수 중 W_t (풍속)과 C_t (운량)의 계수가 음수로 나타나 W_t 와 C_t 가 증가하면 오존농도는 감소하는 것으로 나타났다.

4.2 경기 남부 오존농도자료에 관한 applied ARE 모형

설명변수 중 최고온도, 일사량, 강수량 등의 기상변수들은 서로 연관성이 높은 변수가 있을 수 있으며

Table 3. Factor loading of meteorological variables.

Variable	Factor 1	Factor 2	Factor 3
Maximum temperature	-.504	.803	-9.416E-02
Rainfall	.615	4.878E-02	.391
Wind speed	6.609E-02	-7.589E-02	.961
Humid	.860	.376	3.224E-02
Dew point temperature	.467	.866	-1.688E-02
Water vapor point	.448	.880	-2.858E-02
Radiation	-.951	-4.060E-02	-4.778E-02
Cloud amount	.881	.161	3.455E-02

로, 이 절에서는 먼저 기상변수들 간의 요인분석을 한 후 그 결과를 ARE 모형에 적용하는 방법인 applied ARE 모형을 적용하였다.

요인분석 (Factor Analysis)은 오존자료 분석과 같은 다변량 자료 분석에 많이 쓰이는 기법이다. 요인 분석의 기본적인 틀은 설명변수가 많이 있는 경우, 변수들 간의 상호연관성을 사용하여 변수 속에 내재된 요인 (factor)라고 부르는 작은 수의 공통적인 새로운 변수를 찾아내고, 찾아낸 변수를 이용하여 전체 자료의 특성을 설명하려는 자료 분석기법이다.

요인분석의 가장 큰 장점은 설명변수들의 정보가 되도록 많이 유지하면서 간단히 하는 분석이다. 산출된 인자적재량 (factor loading)을 해석이 용이하도록 직교회전인 Varimax 회전방법을 이용하여 요인분석한 결과 3개의 요인 (Factor 1~Factor 3)이 선택되었다 (표 3 참고). 8종류의 기상자료가 3개의 요인으로 되었을 때의 설명력은 86%로 나타났다. Factor 1에서는 rainfall, humid, radiation, cloud amount 변수가 포함되었고, Factor 2에서는 maximum temperature, dew point temperature, water vapor point 변수가 포함되었으며, Factor 3에서는 wind speed 변수만이 포함되었다.

오존농도를 요인분석 결과로 나온 3개의 요인과 대기변수인 SO₂, NO₂, CO, PM10을 포함한 7개의 설명변수를 이용한 applied ARE 방법으로 한 결과가 표 4에 나타나 있다. 설명변수인 요인 1 (Fact 1)은 평택, 오산, 송탄지역의 오존농도를 설명하는데 모두 포함하고 있다.

Table 4. The applied ARE model for ozone concentration.

Area	Applied model
Pyeongtaek	$Y_t = 0.221Y_{t-1} + 0.389P_t - 10.132\text{Fact } 1_t + \epsilon_t,$ $\epsilon_t = \epsilon_t + 0.186\epsilon_{t-19}$
Osan	$Y_t = 0.141Y_{t-1} + 0.622N_t + 0.102P_t - 8.676\text{Fact } 1_t$ $+ 3.054\text{Fact } 2_t + \epsilon_t,$ $\epsilon_t = 0.176\epsilon_{t-4} + 0.200\epsilon_{t-5} + 0.181\epsilon_{t-14} + \epsilon_t$
Suwon	$Y_t = 0.422Y_{t-1} + 0.619N_t - 7.017\text{Fact } 1_t$ $+ 5.809\text{Fact } 2_t + \epsilon_t,$ $\epsilon_t = 0.141\epsilon_{t-11} + \epsilon_t$

5. 두 가지 ARE 모형의 적합성

오존농도에 관한 두 가지 ARE 모형, direct ARE 모형과 applied ARE 모형에 관한 적합성을 측정하는데 있어 결정계수와 잔차 분석을 이용하였다. 결정계수 R²은 오존 농도 Y_t가 독립변수들에 의해 설명되는 설명력으로 값이 1에 가까울수록 설명력이 높은 좋은 모형이다. 잔차 분석에 사용된 두 가지 통계량은 RMSE (root mean square error)와 IA (index of agreement) 이며 공식은 식 (2)와 같다.

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

$$IA = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (|o_i| + |p_i|)^2} \quad \text{이며,} \quad (2)$$

o_i와 p_i는 각각 i일의 관측 오존량과 추정 오존량을 의미하며, n은 통계량에 사용된 일수 (day)이며, p_i' = p_i - o_m, o_i' = o_i - o_m, o_m = 관측 오존 평균값을 의미한다.

통계량 IA는 [0, 1] 사이의 값을 가지며 0은 전혀 일치하지 않음을 의미하며, 1은 완벽하게 일치함을 의미한다. 즉 IA은 1에 가까울수록 좋은 모형임을 나타낸다.

반면에 예측치와 실측치의 분산정도를 나타내는 RMSE (root mean square error)는 값이 작을수록 좋은 모형이다. 연구기간인 2003년부터 2005년의 5월부터 9월까지의 자료 중, 2003년 5월부터 2004년 9월까지 306일 자료를 적합한 모델 구축을 위한 모형

구축기간으로 2005년 5월부터 9월까지 153일을 검증기간으로 사용하였다.

두 모형을 지역별로 살펴보면, 평택지역과 오산지역은 모형 구축기간과 검증기간 모두 두 모형이 다 적절하나, direct ARE 모형이 좀 더 적합한 것으로 나타났다. 특히 평택지역의 direct ARE모형의 결정계수 R²이 모형 구축기간과 검증기간 각각에 0.922와 0.969로 높게 나타나 오존농도 Y_t는 독립변수들 Y_{t-1}, P_t, R_t, W_t에 의해 잘 설명된다(표 2 참고). RMSE는 오산지역이 제일 적게 나타났다. 수원지역은 모형 구축기간에는 두 모형이 모두 적합하나, applied ARE 모형이 좀 더 적합한 것으로 나타났으나, 검증기간에는 applied ARE의 결정계수 R²이 0.453으로 설명력이 매우 낮게 나타났고 direct ARE의 R²은 0.936으로 높게 나타났다. 결과적으로 오존농도 분석에 세 지역 모두 direct ARE 모형이 좀 더 적합한 것으로 나타났다.

더욱이, direct ARE 모형은 자료 적용하기가 비교적 간단하기 때문에 오존농도 분석으로 direct ARE 모형을 사용하는 것이 적합할 것으로 생각된다.

6. 경기 남부지역의 오존농도 80 ppb 이상 자료에 관한 direct ARE 모형

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 1시간동안 120 ppb를 초과하면 오존 고농도로 정의하고 있다. Bower *et al.* (1994)는 영국에서 두 세 지역에서 일 평균 오존농도가 60 ppb를 넘으면, 고농도 오존일로 정의하고 있다. McKendry (1993)은 캐나다의 두 세 지역에서 시간당 평균이 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

Table 5. The goodness-of-fit test of direct and applied ARE models.

Method	Area	Modeling period (n=306 days)			Verification period (n=153 days)		
		R ²	RMSE	IA	R ²	RMSE	IA
Direct ARE	Pyeongtaek	0.922	14.913	0.873	0.969	12.869	0.901
	Osan	0.769	12.387	0.903	0.951	11.967	0.897
	Suwon	0.878	16.419	0.764	0.936	16.771	0.859
Applied ARE	Pyeongtaek	0.889	17.003	0.896	0.958	14.466	0.889
	Osan	0.792	13.276	0.896	0.906	13.082	0.884
	Suwon	0.906	15.323	0.856	0.453	15.073	0.904

100 ppb을 초과하는 날을 오존오염일 (환경부, 2004)로 정의하였다. 표 6은 평택, 오산, 수원지역의 2003년 5월에서 2005년 9월까지 일반적인 고농도로 볼 수 있는 80 ppb 이상 되는 날자와 평균값을 월별, 년도별로 분석한 표이다. 지역마다 고농도 일수와 고농도 평균값이 조금씩 차이를 보이고 있다. 고농도 일수는 2003년에는 평택과 오산지역이, 각각 25일, 24일로 나타났고, 2004년에는 평택지역이 19일, 2005년에는 수원지역이 30일로 높게 나타났다. 월별로 살펴보면, 전체적으로 5, 6월이 고농도 일수가 많이 나타났고 2005년도에는 5, 6월 뿐만 아니라 7, 8월에도 고농도 일수가 지역별로 2~6일 까지 나타났다.

본 연구에서는 일반적 고농도 정의를 고려하여, 경기 남부지역 중 평택시의 2003년 5월에서 2005년 9월까지의 오존농도를 59 ppb 이하, 60 ppb에서 79 ppb 사이, 80 ppb에서 99 ppb 사이, 100 ppb 이상의 4 그룹으로 나누었고 이에 따른 일수는 각각 276일, 117일, 45일, 17일로 나타났다. 보편적인 고농도로 볼 수 있는 80 ppb 이상도 62일로 나타났다.

그룹별 오존농도에 따라 12종류의 설명변수들의 평균값들이 차이가 있는지를 분산분석(ANOVA)과 Duncan 평균비교 검정 분석을 한 결과를 표 7에 나타냈다. 일산화탄소(CO)만 제외 한 다른 11개의 모든 독립변수는 유의 수준 1%에서 유의한 차이가 있는 것으로 분석되었다. 특히 독립변수 중 PM10은 오존농도의 4 그룹별 모두 유의한 차이가 있는 것으로 나타났고, 오존농도가 증가함에 따라 PM10도 증가하는 것으로 나타났다. 강수량(Rainfall) 이슬점 온도(Dew), 수증기압(Waterp)은 3그룹 [60, 79], [80, 99], [100, ∞]은 유의한 차이가 없고, [0, 59]그룹과는 유의한 차이가 있는 것으로 분석되었고 오존농도가 증가함에 따라 강수량, 이슬점 온도, 수증기압은 감소

Table 6. Monthly days and average of high level of ozone concentrations (over 80 ppb) at the southern part of Gyeonggi-do.

Year	Month	Pyeongtaek days (mean)	Osan days (mean)	Suwon days (mean)
2003	May	11 (90.73)	12 (94.33)	0
	June	8 (87.88)	8 (91.50)	6 (98.50)
	July	4 (89.50)	3 (87.33)	5 (101.60)
	Aug.	1 (83.00)	1 (88.00)	2 (86.00)
	Sep.	1 (92.00)	0	2 (82.50)
	Total	25 (89.36)	24 (92.25)	15 (95.73)
2004	May	6 (84.83)	1 (84.00)	1 (84.00)
	June	9 (113.22)	3 (102.00)	6 (108.67)
	July	2 (102.00)	0	2 (89.50)
	Aug.	1 (87.00)	1 (86.00)	4 (89.25)
	Sep.	1 (95.00)	0	1 (95.00)
	Total	19 (100.74)	5 (95.20)	14 (97.64)
2005	May	8 (93.50)	6 (89.67)	12 (90.08)
	June	5 (90.80)	2 (82.50)	8 (91.38)
	July	2 (106.50)	2 (96.00)	6 (116.17)
	Aug.	4 (90.75)	2 (89.00)	4 (89.50)
	Sep.	0	1 (94.00)	0
	Total	19 (93.58)	13 (89.77)	30 (95.57)

하는 것으로 분석되었다.

오존에 관한 주요 관심사 중의 하나는 고농도에 관한 분석이다. 일반적 고농도인 일 최고 농도가 80 ppb 이상 되는 요일에 관해 평택, 오산 송탄 지역의 direct ARE 모형식을 분석한 결과가 표 8에 나타났다. 표 8의 설명변수는 모든 오존농도가 포함된 표 2의 설명변수와는 좀 다르게 나타났다. 특히 표 2의 모형에서 계수가 음수로 나타났던 수증기압(W_i), 운량(C_i)은 표 8에서는 나타나지 않았다. 그 이유는 표 7의 Duncan 비교 검정 결과에서 보듯이 오존농도가 [80, 99], [100, ∞]에서는 수증기압과 운량은 유의한

Table 7. Results of ANOVA and Duncan's multiple range test of the explanatory variables for ozone concentrations in Pyeongtaek.

Variable (unit)	ANOVA p-value	Duncan's multiple range test		Variable (unit)	ANOVA p-value	Duncan's multiple test	
		Grouping	Mean			Grouping	Mean
NO ₂ (ppb)	0.000	A ¹	35.01	CO (0.1 ppm)	0.301	D ¹	6.12
		B ^{1,2}	39.98			B ¹	6.31
		C ²	44.27			A ¹	6.48
		D ²	45.47			C ¹	7.09
SO ₂ (ppb)	0.000	A ¹	7.04	PM10 (Ug/m ³)	0.000	A ¹	81.11
		B ²	8.65			B ²	103.90
		C ³	10.89			C ³	129.11
		D ³	11.94			D ⁴	155.18
Temperature (°C)	0.000	A ¹	25.77	Rainfull (mm)	0.000	D ¹	0.00
		B ¹	26.85			C ¹	0.28
		C ¹	27.26			B ¹	24.05
		D ²	29.47			A ²	107.66
Wind (m/s)	0.000	D ¹	16.01	Humidity (%)	0.000	D ¹	44.03
		C ^{1,2}	17.54			C ^{1,2}	49.34
		B ^{2,3}	19.93			B ²	55.08
		A ³	22.29			A ³	67.15
Dew (°C)	0.000	D ¹	118.43	Waterp (hPa)	0.000	D ¹	149.61
		C ¹	138.18			C ¹	168.47
		B ¹	141.81			B ¹	174.08
		A ²	170.40			A ²	203.57
Radiation (MJ/M ³)	0.000	A ¹	1248.3	Cloud	0.000	D ¹	3.22
		B ²	1969.6			C ¹	3.81
		C ^{2,3}	2228.8			B ²	5.14
		D ³	2344.9			A ³	7.54

Note: The A, B, C, D letters means ozone concentrations of [0, 59], [60, 79], [80, 99], [100, ∞] respectively. In ^{1,2,3,4} letter, the same letters are not significantly different at 0.05 level.

Table 8. Direct ARE model for high level of ozone concentrations (over 80 ppb).

Area	Model	R ²	RMSE
Pyeongtaek	$Y_t = 0.181P_t + 0.252T_t + \epsilon_t$ $\epsilon_t = 0.378\epsilon_{t-4} - 0.316\epsilon_{t-19} + \epsilon_t$	0.984	10.777
Osan	$Y_t = 0.550N_t + 0.156D_t + 0.020R_t + \epsilon_t$ $\epsilon_t = 0.362\epsilon_{t-2} + 0.501\epsilon_{t-3} - 0.458\epsilon_{t-6} + \epsilon_t$	0.984	7.512
Suwon	$Y_t = 3.286S_t + 0.186T_t + 0.078D_t + \epsilon_t$ $\epsilon_t = -0.403\epsilon_{t-5} + \epsilon_t$	0.985	12.389

차이가 없기 때문인 것으로 생각된다.

평택, 오산 수원 의 세 지역 결정계수는 각각 0.984, 0.984, 0.985로 표 5의 결정계수보다 높게 나타났다.

또한, 세 지역 모두 RMSE가 표 5의 RMSE보다 적게 나타나 오존농도가 80 ppb 이상인 경우 표 8의 direct ARE 모형식이 오존농도를 잘 설명해 주며 적절한 것으로 나타났다.

7. 결 론

본 연구에서는, 경기 남부 지역인 평택, 오산, 수원의 오존농도를 시계열 모형인 두 가지 형태의 ARE 모형으로 분석하였다. ARE 모형은 설명변수를 분석에 사용할 수 있는 모형으로 시계열자료를 회귀모형에 적합시킬 때 적절한 모형이다.

본 연구의 ARE 모형에서는, 오존농도에 영향을 주는 8종류의 기상자료와 4종류의 대기자료를 설명변

수로 사용하였다.

8종류의 기상자료는 일 최고온도, 일사량, 풍속, 상대습도, 강수량, 이슬점온도, 운량, 수증기압이며 4종류의 대기자료는 SO₂, NO₂, CO, PM10 이다. 연구기간인 2003년부터 2005년의 5월부터 9월까지의 자료 중, 2003년과 2004년의 자료 306일을 모형 구축기간으로 2005년 자료 153일을 검증기간으로 사용하였다.

ARE 모형에서, 12종류의 모든 설명변수들을 직접 분석한 direct ARE 모형과 설명변수 중 8종류의 기상 자료들을 먼저 요인분석한 후, 그 결과인 3개의 요인과 4종류의 대기 자료를 설명변수로 이용한 applied ARE 모형을 비교 분석하였다.

평택과 오산지역에는 모형 구축기간과 검증기간 모두 두 모형이 다 적절하나, direct ARE 모형이 좀 더 적합한 것으로 나타났다. 수원지역은 모형 구축기간에는 두 모형이 다 적절하나, applied ARE 모형이 좀 더 적합한 것으로 나타났으나, 검증기간에는 applied ARE 모형은 적합하지 않은 것으로 나타났다. 결과적으로 세 지역 모두 direct ARE 모형이 더 적합한 것으로 분석되었다. 더욱이 direct ARE 모형이 자료를 적용하기가 비교적 간단하기 때문에 오존농도 분석에 direct ARE 모형을 사용하는 것이 좋을 것으로 생각된다.

그리고, 오존 연구의 주요 관심사인 고농도에 관한 자료를 평택, 오산, 수원지역에서 분석하였다. 일반적 고농도로 볼 수 있는 일 최고 오존농도가 80 ppb 이상 되는 요일에 대해 direct ARE 모형을 이용하여 분석하였다. 평택, 오산, 수원의 세 지역의 고농도 direct ARE 모형이 매우 적합한 것으로 나타났다. 특히 세 지역 모두 결정계도 높고 RMSE가 적게 나타나 고농도 direct ARE 모형이 매우 적합한 것으로 나타났다.

앞으로, 오존농도에 사용되는 설명변수에 관한 다양한 연구와 함께 고농도 오존에 관한 연구가 지역 별로 좀 더 활발히 이루어져야 된다고 생각된다.

감사의 글

이 논문은 2005년도 평택대학교 학술연구비의 지원에 의하여 연구되었음.

참 고 문 헌

- 김신도(1998) 오존예보모델 및 예보시스템의 개선, 오존예보시스템에 관한 전문가토론회, 16-23.
- 김용준(1997) 현업운영 가능한 서울지역의 일 최고 대기오염도 예보모델 개발 연구, 한국대기보전학회지, 13(1), 79-89.
- 김유근, 손진태, 문윤섭, 오인보(1999) 서울지역의 지표오존농도 예보를 위한 전이함수모델 개발, 한국대기환경학회지, 15(6), 779-789.
- 김유근, 이소영, 임윤구, 송상근(2007) 중회귀 모형을 이용한 울산지역 오존 포텐셜 모형의 설계 및 평가, 한국대기환경학회지, 23(1), 14-28.
- 환경부(2003) 대기환경연보 2003, 65pp.
- 환경부(2004) 대기환경연보 2004년 6월, 60pp.
- 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. Broughyoun, and J.E. Lampert (1994) Assessing recent surface ozone concentrations in the U.K., *Atmospheric Environment*, 28D, 115-128.
- Feister, U. and K. Balzer (1991) Surface ozone and meteorological predictors on a subregional scale, *Atmospheric Environment*, 25, 1781-1790.
- Guttorp, P., W. Meiring, and D.P. Samson (1994) A space-time analysis of ground-level ozone Data, *Environmetrics*, 5, 241-254.
- 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.
- 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. Limcoln (1991) The meteorological of ozone episodes in the lower lake Michigan air quality region, air & waste management associ-

- ation, 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.
- 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 Geophysical Research*, 96, 22461-22475.
- Smith, L.S. and L.-S. Huang (1993) Modeling the threshold exceedence of urban ozone, Technical Report No. 6, National Institute for Statistical Science, Research Triangle Park, NC.
- Thompson, M.L., J. Rwynolds, L.H. Cox, P. Guttorp, and P.D. Sampson (2000) A review of statistical methods for the meteorological adjustment of tropospheric ozone, Technical Report No. 26, National Research Center for Statistics and Environment, Seattle.