

경기도 평택지역과 서울 정동지역 지표오존농도의 시계열모형 연구

이 훈 자¹⁾

요약

최근 유해성이 강한 지표오존농도가 대기환경의 주요한 문제로 부각되고 있다. 본 연구에서는 경기도 평택과 서울 정동지역의 오존농도를 설명변수를 사용할 수 있는 다변량 시계열 모형인 ARE(자기회귀오차) 모형으로 분석 하였다. ARE 모형에서는 오존 전체자료를 사용한 전체모형과 오존농도가 41ppb 이상 되는 자료를 사용한 부분모형 두 가지 모형을 비교하였다. ARE의 오존농도 설명변수로는 오존농도와 연관 있는 8종류의 기상자료와 4종류의 대기오염자료를 고려하였다. 기상자료의 8가지 설명변수로 일 최고온도, 일사량, 풍속, 상대습도, 강수량, 이슬점온도, 수증기압, 운량 자료를 사용하였다. 대기오염자료의 4가지 설명변수로는 아황산가스(SO₂), 이산화질소(NO₂), 코발트(CO)와 프롬테움 10(PM₁₀)를 사용하였다.

1. 서론

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 등(2000)에 따르면 오존농도 모형설정은 크게 다음의 3가지 범주로 나뉘질 수 있다. 첫째, regression-based modeling(Feister and Balzer, 1991), 둘째, extreme value approach(Smith and Huang, 1993), 셋째, spatio-temporal(Carroll, 1997 : Guttorp, 1994) 이다. 본 연구에서는, 평택시와 서울 정동의 2003년부터 2005년까지의 5월 1일부터 9월 30일까지 오존농도 자료를 시계열 모형에 적합하고자 한다.

정확한 오존농도의 분석과 예측을 위해서는 오존과 연관된 대기 및 기상자료를 포함시켜야 한다. 본 논문에서는 오존과 연관이 있는 8개의 기상자료와 4개의 대기 자료를 설명변수로 포함한 다변량 분석인 ARE 방법으로 분석하고자 한다.

1) 경기도 평택시 평택대학교 디지털응용정보학과 부교수, 450-701
E-mail: esther@ptu.ac.kr

고전적인 방법인 ARIMA(Autoregressive Integrated Moving Average)모형은 시계열분석에서 자주 사용된다. ARIMA모형은 다양한 종류의 일변량 시계열자료를 분석할 수 있고 미래의 값을 쉽게 예측할 수 있다. 반면에 분석하고자 하는 자료와 연관된 설명변수를 포함 시킬 수 없다. 오존자료 분석에도 오존과 연관성이 많은 강수량이나 이산화질소 같은 설명변수를 포함 시킬 수 없다.

본 연구의 ARE(Autoregressive Error)모형 분석에서는 오존과 연관이 있는 8종류의 기상 자료와 4종류의 대기 자료를 설명변수로 사용하였다. 8 종류의 기상 자료로는 일 최고 온도(temperature), 일사량, 풍속, 상대습도, 강수량, 이슬점온도, 운량, 수증기압 이다. 4 종류의 대기 자료로는 SO₂(Sulfur dioxide), NO₂(Nitrogen dioxide), CO(Cobalt), PM₁₀(Promethium 10)을 사용했다.

2. 오존자료 및 설명변수 자료

오존자료는 국립환경과학원의 평택 비전동 monitoring site와 서울 정동 monitoring site의 시간별 자료(하루 24번)를 이용했고 분석을 위하여 일 최고오존농도를 사용했다. 기간은 2003년부터 2005년까지의 5월 1일부터 9월 30일 까지 자료를 사용했다. 본 연구에서 5월에서 9월 까지 자료를 사용한 이유는 1년에 걸쳐 이 기간 동안이 오존 농도가 높기 때문이고 그래서 이 기간동안의 분석도 중요할 것으로 생각되었다.

설명변수인 기상 자료는 시간별 자료로 평택자료는 평택에서 가장 가까운 곳에 위치한 기상청의 수원 지점 자료를 사용하였다. 정동자료는 기상청 서울 자료를 사용하였다. 기상 자료도 오존과 같이 2003년-2005년 5월부터 9월까지의 자료이다. 8 종류의 기상 자료는 [표 1]에 나타나 있으며, 일 최고온도(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) 이다.

설명변수인 4 종류의 대기 자료로는 SO₂(Sulfur dioxide), NO₂(Nitrogen dioxide), CO(Cobalt), PM₁₀(Promethium 10)을 사용했고, 오존자료와 같이 국립환경과학원의 평택, 정동지역 monitoring sites를 이용하여, 2003년-2005년의 5월1일-9월 30일의 시간별 자료를 이용하였다. 사용된 대기 자료도 [표 1]에 나타나 있다.

8종류의 기상 자료와 4종류의 대기 자료의 시간별 선택은 각 자료를 시간별 혹은 일별로 오존과의 상관관계가 가장 높은 시간대를 사용하였다. 예를 들면, 일사량변수는 오존과 제일 상관관계가 높은 일별 5시-20시의 합을 사용하였다. 각 설명변수의 구체적인 시간대는 [표 1]에 있다.

Variable	Contents and Timing	Unit
O3_1	Previous day maximum 1 hour O3	ppb
SO2	Current & previous days maximum 1 hour SO2	ppb
NO2	Current & previous days maximum 1 hour NO2	ppb
CO	Current & previous days maximum 1 hour CO	0.1ppm
PM10	Current & previous days maximum 1 hour PM10	Ug/m3
Temperature	Daily maximum 1 hour surface temperature	° C
Wind	Average wind speed (from 06 hour to 18 hour)	m/s
Cloud	Average cloud amount (from 06 hour to 18 hour)	-
Radiation	Sum of radiation (from 05 hour to 20 hour)	MJ/M ²
Humidity	Average relative humidity (from 09 hour to 15 hour)	%
Rainfall	Sum of precipitation (from 01 hour to 24 hour)	mm
Dew	Dew point temperature at 15 hour	° C
Waterp	Water vapor pressure at 15 hour	hPa

[표 1] 설명변수로 사용된 대기 자료와 기상 자료

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 적합한 모형이다. 반응변수인 오존 농도에 영향을 주는 이

산화질소나 프로메튬 같은 대기 자료와 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 k 차 ARE 모형은 다음과 같다.

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

$$\epsilon_t = e_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2} - \dots - \phi_k \epsilon_{t-k}, \quad \text{이고} \quad e_t \sim \text{i.i.d. } N(0, \sigma^2). \quad (1)$$

식 (1)에서 오차항 ϵ_t 는 서로 상관관계가 있고 AR(k) 형태를 갖는다. 또한 식 (1)에서 Y_t 는 반응변수인 오존 농도를 나타내며, $X_{t,j}, j = 1, 2, \dots, p$ 는 Y_t 에 영향을 주는 설명변수들을 나타낸다.

4. 평택지역 오존농도의 ARE 모형

4.1 평택 전체 오존농도자료에 관한 ARE 모형

평택 전체 오존농도자료 분석을 위하여, 12종류의 설명변수들 간의 multi-collinearity를 분석하고, 각 설명변수들이 오존 농도와 영향력이 제일 높은 time-lag를 고려한 결과 다음과 같은 설명변수를 포함한 ARE모형이 선택되었다.

$$Y_t = 0.290 Y_{t-1} + 0.834 S_t + 0.210 D_t - 0.215 W_t + 0.014 R_t + 0.064 T_t + \epsilon_t \quad \text{이며}$$

$$\epsilon_t = 0.155 \epsilon_{t-2} + 0.113 \epsilon_{t-12} + 0.118 \epsilon_{t-19} + e_t, \quad \text{이고} \quad (2)$$

$Y_t = t$ 시점의 ozone 농도, $Y_{t-1} = t-1$ 시점의 오존농도, $S_t = t$ 시점의 일 최고SO2 농도, $D_t = t$ 시점 15시의 이슬점온도, $W_t = t$ 시점 15시의 수증기압, $R_t = t$ 시점의 5시-20시 사이의 일사량의 합, $T_t = t$ 시점의 최고온도 이다.

ARE 모형의 결정계수는 $R^2 = 0.8595$ 로, 오존농도 Y_t 는 설명변수들 $Y_{t-1}, S_t, D_t, W_t, R_t, T_t$ 에 의해 86% 설명된다. 설명변수들 중 수증기압(W_t)의 계수가 음이므로 수증기압이 오존 농도를 감소시키는 것으로 분석되었다. ARE 모형의 총결정계수인 total $R^2 = 0.9339$ 로 Y_t 는 독립변수와 식(2)의 오차항 ϵ_t 에 의해 93% 설명되어 진다.

4.2 평택 오존농도 41ppb 이상 자료에 관한 ARE 모형

2003년 5월에서 2005년 9월까지의 평택지역에서 보편적인 고농도로 볼 수 있는 80ppb 이상이 62일 이나 나타났다. 평택 오존농도가 41ppb 이상되는 자료 분석을 한 결과 다음과 같은 ARE모형이 선택되었다.

$$Y_t = 0.203 Y_{t-1} + 0.198 P_t + 0.010 R_t + 0.051 T_t + \epsilon_t, \text{ 이고} \quad (3)$$

$$\epsilon_t = 0.204 \epsilon_{t-20} + e_t, \text{ 이며}$$

$Y_t = t$ 시점의 ozone 농도, $Y_{t-1} = t-1$ 시점의 오존농도, $P_t = t$ 시점의 일 최고 pm10 농도, $P_t = t$ 시점의 5시-20시까지의 일사량의 합, $T_t = t$ 시점의 최고온도 이다.

오존 농도 41ppb 이상 되는 ARE모형 (3)의 결정계수는 $R^2=0.946$ 로 매우 높게 나타났고, 식 (2)의 전체자료의 결정계수 $R^2=0.860$ 보다 높게 나타났다.

5. 서울 정동지역 오존농도의 ARE 모형

5.1 정동 오존농도 전체자료에 관한 ARE 모형

서울 정동지역의 ARE 모형을 위해, 평택의 ARE 모형과 같이, 12종류의 설명변수들 간의 multi-collinearity를 분석하고, 오존 농도와 영향력이 제일 높은 time-lag를 고려한 결과 다음과 같은 ARE모형이 선택되었다.

$$Y_t = 0.254 Y_{t-1} + 0.099 P_t + 0.094 D_t + 0.012 R_t - 1.409 C_t + \epsilon_t \text{ 이며}$$

$$\epsilon_t = 0.110 \epsilon_{t-8} + 0.103 \epsilon_{t-12} + e_t \text{ 이고} \quad (4)$$

$Y_t = t$ 시점의 ozone 농도, $Y_{t-1} = t-1$ 시점의 오존농도, $P_t = t$ 시점의 일 최고 PM10 농도, $D_t = t$ 시점 15시의 이슬점온도, $P_t = t$ 시점의 5시-20시까지의 일사량의 합, $C_t = t$ 시점의 6시-18시까지의 운량의 합이다.

ARE 모형의 결정계수는 $R^2=0.852$ 이므로, 오존농도 Y_t 는 설명변수들 Y_{t-1} , P_t , D_t , R_t , C_t 에 의해 85% 설명된다. 설명변수들 중 운량(C_t)의 계수가 음이므로 구름양이 오존 농도를 감소시키는 것으로 분석되었다. ARE 모형의 총결정계수 total R^2 는 0.8962이

다, 즉 Y_t 는 독립변수와 식 (4)의 오차항 ϵ_t 에 의해 90% 설명되어 진다.

5.2 정동 오존농도 41ppb 이상 자료에 관한 ARE 모형

2003년 5월에서 2005년 9월까지의 정동지역에서 보편적인 고농도로 볼 수 있는 80ppb 이상이 31일 이상이나 나타났다. 정동오존농도가 41ppb 이상되는 자료 분석을 한 결과 다음과 같은 ARE모형이 선택되었다.

$$Y_t = 0.148Y_{t-1} + 0.095P_t + 0.004R_t + 1.214T_t + \epsilon_t \text{ 이며,}$$

$$\epsilon_t = 0.173\epsilon_{t-7} + 0.141\epsilon_{t-9} - 0.136\epsilon_{t-14} + e_t \text{ 이고} \quad (5)$$

$Y_t = t$ 시점의 ozone 농도, $Y_{t-1} = t-1$ 시점의 오존농도, $P_t = t$ 시점의 일 최고 PM10 농도, $R_t = t$ 시점의 5시-20시까지의 일사량의 합, $T_t = t$ 시점의 최고온도이다.

오존 농도 41ppb 이상 되는 ARE모형 (5)의 결정계수는 $R^2=0.914$ 로 나타나 전체자료의 결정계수 $R^2=0.852$ 보다 높게 나타났다.

5. 오존농도에 관한 ARE모형의 적합성

오존농도에 관한 두 가지 ARE 모형, 전체 자료 ARE모형과 부분자료 ARE모형(오존농도 ≥ 41)에 관한 적합성을 측정하는데 있어 잔차 분석을 이용하였다. 잔차 분석에 사용된 두 가지 통계량은 RMSE(root mean square error) 와 IA(index of agreement) 이며 공식은 다음과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2}, \quad IA = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (|o_i'| + |p_i'|)^2} \text{ 이며,} \quad (6)$$

o_i 와 p_i 는 각각 i 일의 관측 오존량과 추정 오존량을 의미하며, n 은 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 일치 하지 않음을 의미하며, 1은 완벽하게 일치함을 의미한다. 즉 IA은 1에 가까울수록 좋은 모형임을 나타낸다. 반면에 RMSE은 값이 작을수록 좋은 모형이다.

Statistics	ARE Model (all ozone data)	ARE Model (ozone density \geq 41)
	Modeling terms n=days	Modeling terms n=309 days
RMSE	15.2861	13.3209
IA	0.8521	0.7970

[표 2] 평택 전체자료 ARE모형과 부분자료 ARE모형 간의 적합도 비교

Statistics	ARE Model (all ozone data)	ARE Model (ozone density \geq 41)
	Modeling terms n= 454days	Modeling terms n=246 days
RMSE	16.5315	15.9957
IA	0.8080	0.6525

[표 3] 서울 정동 전체자료 ARE모형과 부분자료 ARE모형 간의 적합도 비교

[표 2]와 [표 3]에서 보듯이, 전체자료 ARE 모형은 IA 통계량이 좀 높게 나왔고, 부분자료 ARE 모형에서는 RMSE 이 좀 작게 나왔다. 오존 농도에 적합한 모형을 찾는 큰 목적은, 미래의 오존농도에 관한 예측이다. 특히 고농도 오존일 예측에 관해 관심이 많으므로 비교적 높은 오존농도에 관한 부분자료 ARE모형을 사용하고 좀 더 다양한 독립변수를 이용하는 것이 좋을 것으로 판단된다.

6. 결론

평택시와 정동지역 오존농도를 시계열 모형인 ARE 방법으로 분석하였다. ARE 방법에서는, 오존 농도 전체 자료를 사용한 전체 ARE모형과 오존 농도가 41ppb 보다 높은 자료를 사용한 부분 ARE 모형을 비교 분석하였다. 부분 ARE 모형이 설명력이 높게 나타났고, 잔차 분석에서는 두 모형 간에 통계량에 차이가 조금 있었다.

앞으로, 미래의 오존농도 중 특히 고농도 예측을 고려하면 비교적 높은 오존 자료를 사용한 다양한 부분자료 ARE 모형에 관한 연구가 이루어져야겠다고 사려 된다. 또한 ARE 모형에 사용되는 독립변수에 관한 연구가 활발히 이루어져야 된다고 생각된다.

참고문헌

- [1] 환경부 (2003): 대기환경연보 2003, 65pp.
- [2] G. Bauer, 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.
- [3] L. S. Smith 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.
- [4] M. L. Thompson, J. Rynolds, 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.
- [5] P. Guttorp, W. Meiring, and DP Samson (1994): *A space-time analysis of ground-level ozone Data*, Environmetrics 5, 241-254.
- [6] S. M. Roberson and D. G. Steyn (1990): *Evaluation and Comparison of statistical forecast models for daily maximum ozone concentrations*, Atmospheric Environment 24B, 303-312.
- [7] U. Feister and K. Balzer (1991): *Surface Ozone and Meteorological Predictors on a Subregional Scale*, Atmospheric Environment 25, 1781-1790.