

PA58) 경기도 남부지역 오존 농도의 자기회귀모형 연구

Analysis of Autoregressive Error Model for Ozone at the Southern Part of Gyeonggi-DO

이훈자 · 이현창¹⁾ · 김태진²⁾

평택대학교 디지털응용정보학과, ¹⁾(주)바이오텔, ²⁾수원대학교 화공생명공학과

1. 서 론

지표오존의 고농도 현상은 농작물이나 식물들의 피해뿐 아니라 인간의 건강에도 피해를 주는 것으로 알려져 있다. 최근 들어, 환경기준치 및 주의보 수준을 초과하는 오존농도가 수도권에 집중되어 빈도 높게 나타나고 있다(환경부, 2003). 지금까지 오존농도의 모형 적합에 관한 통계모형이 연구되어 왔다. 본 연구에서는 경기 남부지역 중 수원, 오산, 평택시의 2003년부터 2005년까지의 5월 1일부터 9월 30일 까지의 오존 농도 자료를 시계열 모형을 이용해 모델링하고자 한다. 본 논문에서는 오존에 영향을 주는 8 종류의 기상자료와 4 종류의 대기 자료를 설명변수로 포함한 다변량 분석인 자기회귀모형(Autoregressive Error) 방법으로 분석하고자 한다. 자기회귀오차모형은 설명변수를 분석에 사용할 수 있는 모형으로 오존과 같은 시계열 자료를 회귀모형에 적합할 때 적절한 모형이다.

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. 설명변수로 8 종류의 기상자료와 4 종류의 대기자료를 사용하였다. 8 종류의 기상자료는 일 최고온도, 일사량, 풍속, 상대습도, 강수량, 이슬점온도, 운량, 수증기압이다. 4 종류의 대기자료는 이산화황(SO₂), 이산화질소(NO₂), 이산화탄소(CO), 미세먼지(PM₁₀)를 사용했다.

3. 오존 농도의 ARE 모형 및 결과

경기도 남부인 평택, 오산, 수원지역 연도별(2003년, 2004년, 2005년) 오존농도 자료 분석을 위하여, 12 종류의 설명변수들 간의 다중공선성(multi-collinearity)을 분석하고, 오존 농도와 영향력이 제일 높은 각 설명변수들의 time-lag를 고려한 결과 표 1과 같은 설명변수를 포함한 ARE모형 등이 선택되었다.

표 1에서 Y_t 는 t 시점의 일 최고오존 농도, Y_{t-1} 는 t-1 시점의 일 최고오존 농도, P_t 는 일 최고 미세먼지 농도, N_t 는 일 최고 NO₂ 농도, S_t 는 일 최고 SO₂ 농도, A_t 는 1시-24시 강수량의 합, C_t 는 6시-18시 평균 운량, R_t 는 9시-15시 일사량의 합, W_t 는 6시-18시 평균 풍속, T_t 는 일 최고 온도이다.

모형의 적합성을 보기 위해 결정계수 R^2 과 RMSE(Root Mean Square Error)를 사용했다. 결정계수는 오존이 설명변수들에 의한 설명력으로 값이 1에 가까울수록 설명력이 높은 좋은 모형이다. 반면에 측정값과 예측값의 분산정도를 나타내는 RMSE는 값이 적을수록 좋은 모형이다. 각 지역별 및 연도별의 모형이 대부분 잘 적합되었고, 오산지역의 2004년 오존 모형식에서 보면 오존 Y_t 는 Y_{t-1} , P_t , R_t , C_t 에 의해 96% 설명되는 것으로 분석되었다. 또한, 본 연구에서는 평택, 오산, 수원 지역의 오존의 측정값과 ARE모형의 예측값의 비교와 산점도, 상관관계 분석을 하였고, 각 지역의 2005년 결과를 그림 1에 나타내었다.

Table 1. The ARE models for ozone concentration in Pyeongtaek, Osan, Suwon City.

Area	Year	ARE Model	R^2	RMSE
Pyeongtaek	2003	$Y_t = 0.256 Y_{t-1} + 0.400 P_t - 0.028 A_t + \epsilon_t,$ $\epsilon_t = -0.180 \epsilon_{t-12} - 0.203 \epsilon_{t-20} + e_t$	0.948	17.386
	2004	$Y_t = 0.316 Y_{t-1} + 0.340 P_t + 0.311 N_t - 1.140 C_t - 0.021 A_t + \epsilon_t,$ $\epsilon_t = 0.189 \epsilon_{t-5} + e_t$	0.856	19.394
	2005	$Y_t = 0.336 Y_{t-1} + 0.226 P_t + 0.012 R_t - 0.543 C_t + \epsilon_t,$ $\epsilon_t - 0.236 \epsilon_{t-13} + e_t$	0.967	13.063
Osan	2003	$Y_t = 0.151 Y_{t-1} + 0.381 N_t + 0.078 P_t + 0.017 R_t - 0.323 W_t + \epsilon_t,$ $\epsilon_t = 0.286 \epsilon_{t-1} + 0.337 \epsilon_{t-3} - 0.222 \epsilon_{t-9} - 0.178 \epsilon_{t-10} + 0.225 \epsilon_{t-11} + e_t$	0.890	13.373
	2004	$Y_t = 0.204 Y_{t-1} + 0.102 P_t + 0.014 R_t - 0.338 C_t + \epsilon_t,$ $\epsilon_t = -0.236 \epsilon_{t-11} - 0.266 \epsilon_{t-22} + e_t$	0.963	11.030
	2005	$Y_t = 0.277 Y_{t-1} + 0.215 N_t + 0.159 P_t + 0.014 R_t - 0.344 W_t + \epsilon_t,$ $\epsilon_t = -0.190 \epsilon_{t-15} + 0.281 \epsilon_{t-18} + e_t$	0.950	11.636
Suwon	2003	$Y_t = 0.237 Y_{t-1} + 0.678 S_t + 0.202 T_t - 2.462 C_t + \epsilon_t,$ $\epsilon_t = 0.332 \epsilon_{t-2} + 0.202 \epsilon_{t-4} + e_t$	0.792	15.103
	2004	$Y_t = 0.250 Y_{t-1} + 1.966 S_t + 0.146 T_t - 2.029 C_t + \epsilon_t,$ $\epsilon_t = 0.251 \epsilon_{t-5} + e_t$	0.901	13.986
	2005	$Y_t = 0.441 Y_{t-1} + 1.979 S_t + 0.132 T_t - 2.672 C_t + \epsilon_t,$ $\epsilon_t = 0.193 \epsilon_{t-8} - 0.225 \epsilon_{t-13} + 0.164 \epsilon_{t-18} - 0.173 \epsilon_{t-23} + e_t$	0.957	13.961

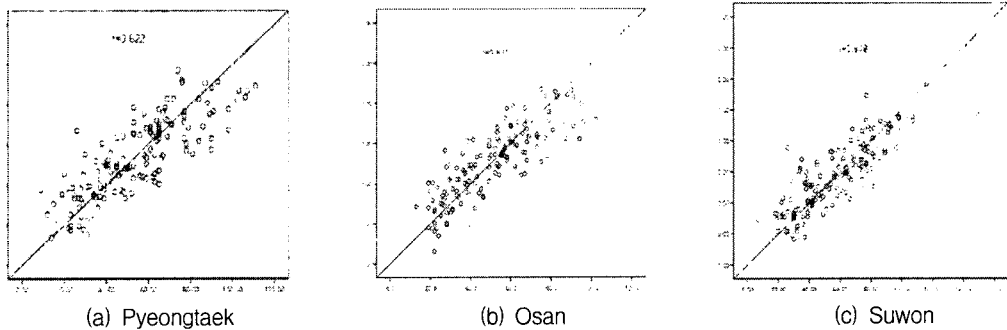


Fig. 1. The scatter plots & correlation between observed value and predicted value in 2005.

참고 문헌

- 김유근, 손건태, 문운섭, 오인보 (1999) 서울 지역의 지표오존농도 예보를 위한 전이함수 모델 개발, 한국 대기환경학회지, 15(6), 779-789.
- 환경부 (2003) 대기환경연보 2003, 65pp.
- 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.
- 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.