

## PA6) 경기도 광명시 오존 농도의 모형 연구

### Analysis of Model for Ozone Concentrations at Gwang Myeong City of Gyeonggi-Do in Korea

이 훈 자

평택대학교 디지털응용정보학과

#### 1. 서 론

지금까지 오존농도의 모형 적합에 관한 통계모형이 연구되어 왔다. 본 연구에서는 경기 중부지역 중 광명시의 2003년부터 2007년까지 5년간의 5월 1일부터 9월 30일까지의 오존 농도 자료를 시계열 모형인 자기회귀오차모형(ARE, Autoregressive Error Model)과 데이터마이닝 모형인 신경망모형(NN, Neural Network Model)을 이용해 모델링하고자 한다. 본 연구에서는 8종류의 기상자료와 4종류의 대기자료를 오존에 영향을 주는 설명변수로 사용하였다. 자기회귀오차모형과 신경망모형은 설명변수를 분석에 사용할 수 있는 모형으로 오존과 같은 시계열 자료를 적합할 때 적절한 모형이다.

#### 2. 오존자료 및 설명변수 자료

오존자료는 국립환경과학원의 광명시 관측소의 시간별 자료(하루 24번)를 이용했고, 분석을 위하여 일 최고 1시간 오존농도를 사용했다. 기간은 2003년부터 2007년까지의 5월 1일부터 9월 30일까지 자료를 사용했다. 설명변수 중 대기자료는 고양시의 이산화황(SO<sub>2</sub>), 이산화질소(NO<sub>2</sub>), 일산화탄소(CO), 미세먼지(PM<sub>10</sub>)를 사용하였고 기상자료는 광명시에서 가까운 곳에 위치한 서울시의 일 최고온도, 일사량, 풍속, 상대습도, 강수량, 이슬점온도, 운량, 수증기압을 사용하였다.

#### 3. 오존 농도의 모형 및 결과

광명시 지역의 오존농도를 연도별, 월별로 상자그림(box plot) 그림 1을 통해 살펴보았다. 먼저, 연도별로 살펴보면, 2005년도에는 분포가 넓게 퍼져있고 중앙값이 비교적 높게 나타났고 2004년에는 오존 농도가 100ppb보다 높은 고농도 요일이 5일 정도 나타났다. 월별로 살펴보면, 오존 중앙값이 7, 8, 9월에 비해 5, 6월이 높게 나타났다. 자기회귀오차모형은 어떤 설명변수들이 오존에 영향을 주는지 등의 분석 결과를 비교적 쉽게 해석하여 이로부터 유용한 정보를 얻을 수 있다. 반면에 신경망 모형은, 다른 모형에 비해 정확한 예측을 하는 특성이 있지만 결과를 해석하기 힘든 단점이 있다.

자기회귀오차모형으로 월별 오존농도 자료 분석을 위하여, 12종류의 설명변수들 간의 다중공선성(multi-collinearity)을 고려하고 분석한 결과 표 1과 같은 ARE 모형 들이 선택되었다. 표 1에서  $Y_t$ 는 t 시점의 일 최고오존 농도,  $Y_{t-1}$ 는 t-1 시점의 일 최고오존 농도,  $P_{t-1}$ 는 t-1 시점의 PM<sub>10</sub> 농도,  $T_t$ 는 일 최고 온도,  $C_t$ 는 9시-15시 평균 운량,  $A_t$ 는 일 강수량 등 이다. 월별로 모형식은 조금씩 차이가 있지만, 각 월에서 오존을 가장 잘 설명하는 변수로 일 최고온도와 운량이 포함되었다. 모형의 적합성을 보기 위해 결정계수( $R^2$ )와 RMSE(root mean square error)를 사용했다. 결정계수는 설명변수들에 의한 오존 설명력으로 값이 1에 가까울수록 설명력이 높은 좋은 모형이다. 반면에 측정값과 추정값의 분산정도를 나타내는 RMSE는 값이 적을수록 좋은 모형이다. 8월의  $R^2$ 는 0.66으로 설명력이 비교적 낮게 나타났으나, 그 외의  $R^2$ 는 0.80 이상으로 설명력이 높게 나타나 월별 모형이 대부분 잘 적합 되었다. 예를 들어 9월에는 오존  $Y_t$ 는  $Y_{t-1}$ ,  $T_t$ ,  $C_t$ 에 의해 82% 설명되며 RMSE도 12.58ppb로 모형이 잘 적합된 것으로 분석되었다. 그림 2는 2007년 오존농도를 자기회귀오차모형과 신경망모형으로 분석한 후 실측치와 예측치를 산점도로 나타낸 것이다. 산점도에서 신경망모형이 더 좋은 예측결과를 나타냈고 또한 RMSE 값도 자기회귀오차모형이 17.85ppb이고 신경망모형은 12.60ppb로 나타나 신경망모형이 더 좋은 모형으로 나타났다.

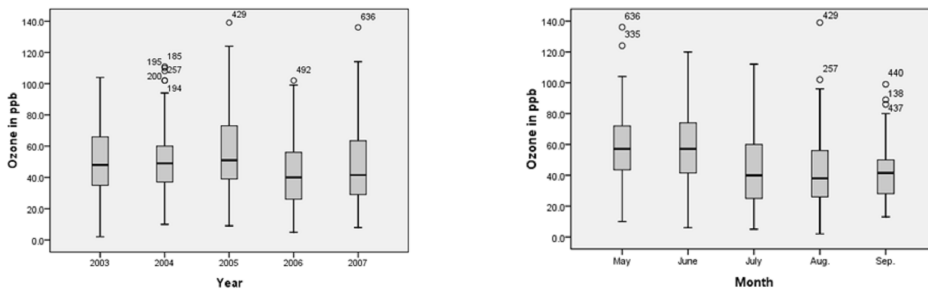
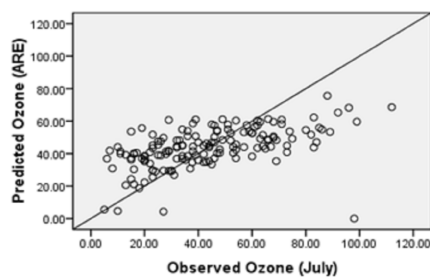


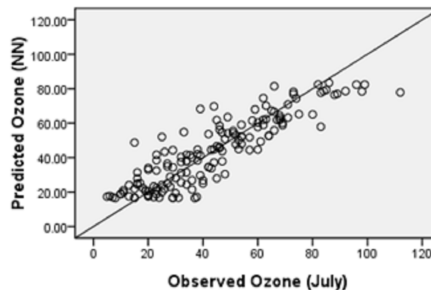
Fig. 1. The box plots of (yearly and monthly) ozone concentrations in Gwang Myeong City. The circle indicates outlier of the ozone concentration.

Table 1. The monthly ARE models for ozone concentration at Gwang Myeong City.

Month	ARE Model	R <sup>2</sup>	RMSE
May	$Y_t = 0.150 Y_{t-1} + 0.054 P_{t-1} + 2.314 T_t - 2.024 C_t + \epsilon_t,$ $\epsilon_t = 0.215 \epsilon_{t-2} + e_t$	0.92	14.05
June	$Y_t = 0.133 Y_{t-1} + 2.422 T_t - 2.127 C_t + \epsilon_t,$ $\epsilon_t = 0.178 \epsilon_{t-8} - 0.183 \epsilon_{t-13} + e_t$	0.92	17.63
July	$Y_t = 0.273 Y_{t-1} + 1.262 T_t - 0.142 A_t + \epsilon_t,$ $\epsilon_t = 0.215 \epsilon_{t-16} + e_t$	0.82	17.85
August	$Y_t = 2.008 T_t - 2.487 C_t + \epsilon_t,$ $\epsilon_t = 0.501 \epsilon_{t-1} + e_t$	0.66	17.04
September	$Y_t = 0.199 Y_{t-1} + 1.765 T_t - 1.662 C_t + \epsilon_t,$ $\epsilon_t = 0.197 \epsilon_{t-3} + 0.219 \epsilon_{t-4} + e_t$	0.82	12.58



(a) Autoregressive Error model



(b) Neural Network model

Fig. 2. The scatter plots between observed ozone and predicted ozone concentration in July.

### 참 고 문 헌

- 강현철, 한상태, 최종후, 이성건, 김은석, 엄익현, 김미경 (2006) 고객관계관리(CRM)을 위한 데이터마이닝 방법론, 자유아카데미, 165-193.
- 김용국, 이종범 (1994) 하계의 일 최고 오존농도 예측을 위한 신경망 모델의 개발, 한국대기보전학회지, 10(4), 224-232.
- 천성표, 김성신, 이종범 (2007) 고농도 오존 예측을 위한 향상된 변화기법과 예측성능 평가, 퍼지 및 지능 시스템학회 논문지, 17(4), 435-442.