

## DR19) 중회귀 모형을 이용한 일최고 오존 농도 예측성 검토에 관한 연구

### Prediction of Daily Maximum Ozone Concentration using Multi-Regression

김영은 · 조석연

인하대학교 환경공학과

#### 1. 서 론

대기질의 통계예측모형은 주로 오존 농도 예측에 사용된다. 통계예측 방법은 중회귀 모형, 신경망 모형, Fuzzy 논리 모형 등이 있다. 중회귀 모형은 종래 통계분석 방법으로 예전부터 많이 사용되고 있는 방법인 반면에 신경망 모형과 Fuzzy 논리 모형은 최근에 개발되어 적용가능성을 검토 중인 방법이다. 국내의 연구결과에 의하면 각 방법에 의한 고농도 오존 예측성은 크게 다르지 않았다. 국내에서는 중회귀 모형<sup>1)</sup>과 신경망 모형<sup>2)</sup>이 적용되었는데, 상관계수는 0.6-0.7 정도로 보고되었다. 국내 연구에서 보고된 상관계수는 외국연구의 상관계수보다 낮게 보고되고 있는데, 이는 국내 지형이 복잡하여 기상관측소 주변 기상과 대기 측정소 주변 기상이 다른 경우가 많고, 대기 측정소 주변에 주요 오염원이 위치하여 대기 측정소의 대표성이 낮은 경우가 많으며, 대기오염물질 배출량의 증가속도가 커서 통계분석에 부적절한 것 등이 주요 이유이다.

대기 측정소와 기상 관측소의 자료를 실시간으로 관측하는 종합대기관리 시스템에 익일 오존 최고농도를 예측하도록 하는 것이 본 연구의 목적이다. 한국은 산업발전이 급속도로 이루어져서 대기질의 변화가 선진국에 비하여 심하므로, 오존을 정확히 예보하기 위해서는 관련자료가 계속해서 갱신된 최근자료를 이용하여 통계분석이 이루어지도록 하여야 한다. 본 연구에서 오존 예측 모형으로는 관련 계수의 계산이 효율적이고도 계산과정이 쉬운 중회귀 모형을 채택하였다.

#### 2. 중회귀 모형의 검토

중회귀 모형은 다음 산술식으로 표현된다.

$$y_0 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} \cdots \beta_p x_{1p} \quad i=1, \dots, n$$

여기서  $y_0$ 는 종속변수인데 본 모형에서는 일 최고 오존 농도이며,  $x_{11}, \dots, x_{1p}$ 는 독립변수인데 본 모형에서는 기상인자와 전일 대기오염물질 농도이다.  $\beta_0, \dots, \beta_p$ 는 회귀계수로써 과거의 자료를 회귀 분석하여 산출한다. 상용 통계분석 프로그램을 이용하면 회귀계수를 구할 수 있으나, 본 시스템의 독립성을 높이기 위하여 본 연구에서는 회귀계수를 구하는 프로그램을 독자 개발하였다. 위 식에서  $x_{11}, \dots, x_{1p}$ 가 대상 변수의 1차 함수면 선형 중회귀 모형이며, 대상변수의 2차 함수 이상이면 비선형 중회귀 모형이다.

국내에 적합한 회귀분석모형을 개발하기 위하여 본 연구에서는 표 1에 나열한 6개의 선형 및 비선형 중회귀 모형을 개발하여 국내 오존 농도를 예측하였다.

#### 3. 중회귀 모형 예측 결과

표 1에 나열된 중회귀 모형을 1996년에 적용하여 회귀계수를 구하고 이를 1997년에 적용하여 모형의 예측성을 평가하였다. 대상 측정소는 면목동, 불광동, 마포, 신림동과 잠실로 선정하였다. 모든 경우에서 1년 전체에 대하여 적용시킨 경우보다 계절별로 적용시킨 경우가 예측성이 좋았다. 표 2는 모형별로 각 측정소의 일 최고 오존 농도를 예측한 뒤에 예측성을 나타내는 인자를 평균한 것이다.

표 2는 관측변수와 예측변수만을 사용한 모형 3, 5, 7과 8의 예측성이 관측변수와 예측변수를 모두 사용한 경우 보다 낮다는 것을 보여 주고 있다. 또한 모형 4와 모형 6의 비교는 전날 오존 관측치를 첨가하면 예측성이 증가하는 것으로 나타났는데, 이는 오존의 자기회귀성(auto-correlation)이 높다는 것을 보여 준다. 모든 모형의 상관계수, RMSE와 오존 농도 구간별 출현회수의 정확성은 높지 않은 것으로 평가된다. 모형간에 예측성은 큰 차이는 없으나 비선형 모형인 모형 2의 예측성이 가장 높은 것으로 나타났다.

다. 향후 연구에서는 기상유형별로 분류하여 모형의 예측성이 높아지는 가를 검토할 예정으로 있다. 본 시스템에는 3차원 이동/확산/평화확/침착 모형인 STEM-ENG을 포함시켜서 미래 대기질을 모사하는 기능을 갖추려 한다. 이를 위하여 격자 간격이 6km이며 격자구간이 남한 전체와 북한 일부를 포함한 모사구간을 선정하였다. 또한 기상측정자료로부터 3차원 바람장, 소용돌이 확산계수, 건조 침착 속도 등을 산출하는 프로그램을 개발하였다. 대기오염물질 배출량 자료의 산출이 완료되면 STEM-ENG 모형에 의한 대기질 예측을 행할 예정이다.

Table 1. Multi-regression models compared in the present study

| 번호 | 모형명                | 독립변수   |
|----|--------------------|--|
| 1  | 주요변수 선형 중회귀 모형     | 관측변수: 일 최고 온도, 최고 오존 농도<br>예측변수: 일 최고/최저 온도, 아침풍속, 오후 풍속, 습도, 운량, 일사량  |
| 2  | 주요변수 비선형 중회귀 모형    | 모형 1과 동일하되 lowess 방법으로 독립변수를 비선형 함수로 나타냄 <sup>3)</sup>   |
| 3  | 주요변수 비선형 중회귀 모형-예측 | 모형 1과 동일하되 관측변수를 제외함   |
| 4  | 다중 변수 선형 중회귀 모형    | 관측변수: 일 평균 풍속, U, V 풍속, 일평균온도, 운량, 습도, 최고/최저온도, 일사량, 강수량, 표면기압경사<br>예측변수: 일 평균 풍속, U, V 풍속, 운량, 최고/최저온도, 강수량 |
| 5  | 다중 변수 선형 중회귀 모형-관측 | 모형 4와 동일하되 예측변수를 제외함   |
| 6  | 다중 변수 선형 중회귀 모형-오존 | 모형 4와 동일하되 전일 오존 최고농도를 추가함   |
| 7  | 다중 변수 선형 중회귀 모형-관측 | 모형 6과 동일하되 예측변수를 제외함   |
| 8  | 다중 변수 선형 중회귀 모형-예측 | 모형 4와 동일하되 관측변수를 제외함.  |

Table2. Performance analysis of various multi-regression models in terms of correlation coefficients, RMSE, NMSE

|      | 상관계수 | RMSE  | NMSE  |      | 상관계수 | RMSE  | NMSE  |
|------|------|-------|-------|------|------|-------|-------|
| 모형 1 | 0.77 | 13.42 | 0.013 | 모형 5 | 0.60 | 19.80 | 0.017 |
| 모형 2 | 0.72 | 14.77 | 0.014 | 모형 6 | 0.72 | 15.73 | 0.014 |
| 모형 3 | 0.66 | 16.34 | 0.015 | 모형 7 | 0.66 | 18.03 | 0.015 |
| 모형 4 | 0.67 | 17.42 | 0.015 | 모형 8 | 0.69 | 15.52 | 0.014 |

( RMSE는 root mean square error, NMSE는 normalized mean square error 임.)

Table3. Performance analysis of various multi-regression models in predicting high ozone concentrations

| 모형치<br>측정치 | 모형 1     |            |             |           | 모형 2     |            |             |           | 모형 4     |            |             |           | 모형 6     |            |             |           |
|------------|----------|------------|-------------|-----------|----------|------------|-------------|-----------|----------|------------|-------------|-----------|----------|------------|-------------|-----------|
|            | 80<br>이하 | 80-<br>100 | 100-<br>120 | 120<br>이상 | 80<br>이하 | 80-<br>100 | 100-<br>120 | 120<br>이상 | 80<br>이하 | 80-<br>100 | 100-<br>120 | 120<br>이상 | 80<br>이하 | 80-<br>100 | 100-<br>120 | 120<br>이상 |
| 80이하       | 1743     | 13         | 3           | 0         | 1734     | 18         | 6           | 1         | 1716     | 41         | 2           | 0         | 1718     | 39         | 2           | 0         |
| 80-100     | 41       | 10         | 0           | 0         | 40       | 9          | 2           | 0         | 38       | 9          | 4           | 0         | 38       | 11         | 2           | 0         |
| 100-120    | 4        | 2          | 2           | 0         | 3        | 3          | 1           | 1         | 5        | 1          | 2           | 0         | 4        | 2          | 2           | 0         |
| 120이상      | 1        | 0          | 1           | 0         | 0        | 1          | 1           | 0         | 1        | 1          | 0           | 0         | 0        | 1          | 1           | 0         |

### 참 고 문 헌

1. 김 용준 (1997) 현업 운영 가능한 서울 지역의 일 최고 대기오염도 예보모델 개발 연구, 한국대기보전학회지, 13(1), 79~89
2. 김 용국 · 이 중범 (1994) 하계의 일 최고 오존농도 예측을 위한 신경망 모델의 개발, 한국대기보전학회지, 10(4), 224~232
3. Milton C. Hubbard and W. Geoffrey Cobourn (1998) Development of a Regression Model to forecast Ground-level ozone concentration in Louisville, KY, *Atmospheric Environment*, 32, 2637~2647