

하계의 일 최고 오존농도 예측을 위한 신경망모델의 개발

Development of Neural Network Model for Prediction of Daily Maximum Ozone Concentration in Summer

김용국 · 이종범

강원대학교 환경학과
(원고접수: 1994. 6. 20.)

Yong-Goo Kim and Chong-Bum Lee

Department of Environmental Science, Kangwon National University,
Chunchon 200-701, Korea
(Received 20 June 1994)

Abstract

A new neural network model has been developed to predict short-term air pollution concentration. In addition, a multiple regression model widely used in statistical analysis was tested. These models were applied for prediction of daily maximum ozone concentration in Seoul during the summer season of 1991. The time periods between May and September 1989 and 1990 were utilized to train set of learning patterns in neural network model, and to estimate multiple regression model. To evaluate the results of the different models, several performance indices were used.

The results indicated that the multiple regression model tended to underpredict the daily maximum ozone concentration with small r^2 (0.38). Also, large errors were found in this model; 21.1ppb for RMSE, 0.324 for NMSE, and -0.164 for MRE. On the other hand, the results obtained from the neural network model were very promising. Thus, we can know that this model has a prominent efficiency in the adaptive control for the non-linear multi-variable systems such as photochemical oxidants. Also, when the recent new information was added in the neural network model, prediction accuracy was increased. From the new model, the values of RMSE, NMSE, MRE and r^2 were 13.2ppb, 0.089, 0.003 and 0.55 respectively.

1. 서 론

오존은 주로 자동차로부터 배출되는 질소산화물, 비메탄계 탄화수소 등의 오염물질의 반응에 자외선이 관여하여 생성되는 2차 오염물질로서 주간에 높은 농도를 유지하며 계절별로는 일사가 강한 여름철에 고농도를 나타내는 특징이 있다. 고농도의 오존이 발생하는 경우 광화학 스모그를 형성하여 시정 감소와 일사량의 감소를 초래하며 생태계 또는

인체의 건강에 피해를 주게 된다. 대류권 내에서 오존의 생성에 관한 반응 기작은 매우 복잡하며 최근 챔버실험 등을 통하여 많은 연구자들에 의해 연구되고 있다(Pitts and Pitts, Jr., 1993).

오존은 단기간에 고농도를 나타내는 특징이 있으므로 이러한 오존을 대상으로 대기질을 관리하는 경우 고농도 발생에 대비하여 신속히 대처할 수 있는 대기오염 예보제와 같은 정책적인 제도의 수립이 필요하다. 대기오염 예보제는 고농도의 대기오염

이 발생될 경우에 야기되는 피해를 줄이기 위한 제도로서(Stern et al., 1984) 여름철의 주 대상이 되는 물질은 오존이며, 이 제도의 수행을 위하여서는 최소한 수시간 전에 정확한 농도 예측이 가능해야 한다.

지금까지 이용되어 온 오존농도의 예측 방법에는 오염물질의 수송, 확산 과정 및 오염물질 상호간의 복잡한 반응 기작을 모사하는 광화학 확산모델을 이용하는 방법과 다변량 분석 등의 통계 기법을 적용한 통계모델을 이용하는 방법 등이 있다. 통계모델은 대기질 자료와 기타 기상 인자 등의 과거 자료에 의하므로 적절한 예측 모형이 선정될 경우 비교적 정확하게 대기질을 추정할 수 있는 장점이 있으나, 장래 배출량이 달라지면 모델 식을 다시 작성해야 하는 등의 단점이 있다. 대기오염 물질 중에서도 특히 오존은 복잡한 광화학 반응에 의하여 생성되므로 다른 오염물질과는 비선형적인 관계(홍낙기 등, 1993)가 있어 정확한 농도 추정에 어려움이 따른다.

최근 대기오염 농도를 예측하는 새로운 방법으로서 Boznar et al.(1993)은 신경망(neural network) 기법을 적용하여 복잡한 지형에서 측정된 기상 자료 및 대기질 자료를 이용하여 30분 후의 지역별 아황산가스 농도를 예측한 바 있다. 신경망은 인간의 두뇌 작용을 모방한 일종의 인공지능 시스템을 말하는 것으로 인간이 경험과 훈련에 의해 지식을 축적하듯이 주어진 자료들을 학습하는 능력을 지니고 있다. 또한 자료의 연속적인 학습을 통해 연결강도(connection strength)를 스스로 조절할 수 있으며, 새로운 정보의 유입시 이미 학습된 연결강도에 따라 출력을 재생하는 자기조직 시스템으로도 불리운다. 이러한 신경망의 학습은 병렬 분산 처리(parallel distributed processing)과정을 통하여 행해지므로 비선형적인 특징을 지닌 시스템의 재현에 유용한 장점(박민용과 최항식, 1991)을 가지고 있다. 신경망의 학습 방법에는 여러 가지 알고리즘이 제안되어 있으며 그 특성에 따라 영상인식, 음성인식 등의 분야에서 탁월한 기능을 보이므로 최근 관심이 고조되고 있는 분야이다.

본 연구에서는 신경망 기법 중에서 오존과 같은 대기오염 물질의 농도를 예측하기에 적절한 방법(Boznar et al., 1993)으로 알려져 있는 오차역전파(error back-propagation) 학습 알고리즘을 도입하여 서울 지역의 하계 일 최고 오존농도를 예측하는 신경망모델을 개발하였으며 이 모델을 다변량 통계 기법인 중회귀모델과 비교하였다.

2. 신경망모델

신경망은 인간 특유의 정보처리 시스템인 뇌를 모방하여 만든 계산 구조로서, 뉴런(또는 processing element)이라 불리는 기본 단위들의 집합으로 이루어져 있으며, 각 뉴런은 단층 또는 다층구조 사이에서 상호 연결되어 정보를 교환하도록 구성되어 있다(Rumelhart and McClelland, 1986). 다층의 병렬구조를 가진 신경망에 있어서 그 층수가 많으면 더 복잡한 자료까지도 처리가 가능하나 이 경우 컴퓨터의 계산 시간이 문제된다. 보통은 입력층과 중간층(또는 은닉층), 출력층으로 구성되는 3개의 층만으로도 적절한 예측을 기대할 수 있는 것(Niemann, 1989)으로 알려져 있다. 이러한 3층 네트워크의 기본 구조는 그림 1과 같으며 본 연구에서는 중간층의 수를 3개로 하였다. 오차역전파 학습 알고리즘은 이러한 다층 병렬구조를 기초로하고 있으며, 그 학습과정으로서 오차를 최소화하기 위한 정방향 또는 역방향의 반복과정을 수행한다.

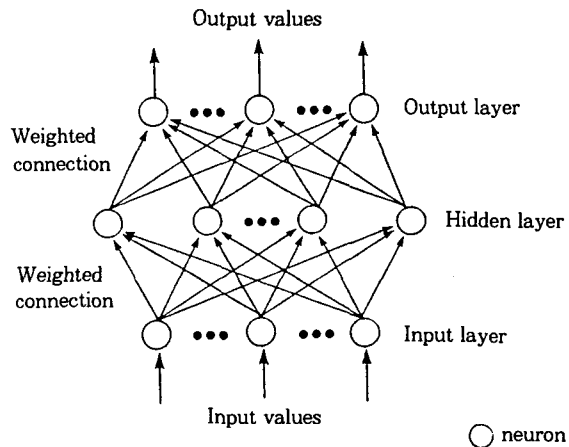


Fig. 1. The configuration of a three-layer perceptron neural network(after Lee, 1994).

학습 방법의 구체적인 과정(이상점, 1994)을 간략히 기술하면 다음과 같다. 식 (1)은 그 첫번째 과정으로서, U_i 는 입력층에 입력된 학습 패턴(learning pattern, I_i)과 입력층과 중간층의 가중값(weight, W_{ij}) 및 중간층의 오차항(θ_j)으로부터 계산된 값으로 입력층에서의 출력값이 된다.

$$U_i = (\sum_j W_{ij} \cdot I_i) + \theta_j \quad (1)$$

식 (1)로부터 계산된 U_i 는 중간층으로 입력되며,

중간층에서는 이 값을 제한된 영역 내로 조절시키기 위한 전달함수로서 연속적인 신호의 비선형 제어에 유용한 시그모이드 함수를 취한다. 시그모이드 전달함수는 식 (2)로 표현할 수 있으며 이를 통한 출력은 0에서 1 사이의 값으로 제한된다. 여기서 a 는 함수의 기울기를 결정하는 매개변수로 이 값이 크면 기울기가 완만하므로 본 연구에서는 0.3으로 하였다.

$$H_j = \frac{1}{1 + \exp(-U/a)} \quad (2)$$

위와 같은 과정은 출력층에서도 동일한 방법으로 진행된다. 식 (3)은 중간층의 출력값인 H_j 와 중간층과 출력층 사이의 가중값(W_{kj}), 출력층 뉴런의 오차함(θ_k)으로부터 출력층에 입력될 S_k 가 계산되는 과정을 나타낸 것이다. 이렇게 계산된 S_k 는 다시 식 (2)와 같은 형태의 시그모이드 함수를 거쳐 출력층의 출력값에 해당하는 O_k 로 변환된다(식 (4)). 위의 과정에서 첨자 i 및 k 는 입력층 및 출력층에서의 뉴런 수 즉, 독립변수 및 종속변수의 수에 해당하는 것으로 각각 14 및 1의 값을 입력하였으며, j 는 중간층의 수이다.

$$S_k = (\sum_j W_{kj} \cdot H_j) + \theta_k \quad (3)$$

$$O_k = \frac{1}{1 + \exp(-S_k/a)} \quad (4)$$

학습의 초기 단계에서 가중값(W_{ij} , W_{kj}) 및 오차함(θ_i , θ_k)은 난수화된 값으로 입력되나 이후에는 순차적으로 수정된 값으로 입력된다. 즉 각 뉴런은 계산된 오존농도와 실측 오존농도를 비교하여 가중값을 다시 조절하고, 오차가 최소화될 때까지 순방향 또는 역방향으로 반복적인 학습을 행하게 된다. 이러한 가중값의 조절 방법은 식 (5)로 간단히 표현할 수 있다.

$$\text{new weight} = (\text{learning rate}) \times (\text{error}) + (\text{momentum}) \times (\text{last weight}) \quad (5)$$

여기서 학습률(learning rate)과 모멘텀상수(momentum)는 매우 중요하며, 학습률은 연결강도의 변화 크기를 조절하는 것으로 이 값이 작으면 수렴 속도는 늦지만 보다 정확한 학습이 기대되므로 0.2의 값을 사용하였고, 모멘텀상수는 학습의 속도를 조절하는 변수로 본 연구에서는 0.9로 하였다. 이는 이전의 연결강도의 변화량을 반영한다. 위의 학습과정에서 사용된 시그모이드 함수의 기울기라든가 학

습률, 모멘텀상수 그리고 중간층의 수 등을 결정하는 데에는 많은 예비실험이 요구되나 본 연구에서는 이상점 등(1993)의 결과를 참고하였다.

지금까지의 식은 한개의 입력력 쌍에 대한 오차를 최소화시키는 과정을 나타낸 것으로 입력패턴 p 에 대한 오차(E_p)는 식 (6)으로 나타낼 수 있다. 여기서 T_k 는 실측값을 의미한다. 그러므로 p 개의 입력패턴에 대한 총 오차(E_{total})는 식 (7)과 같다.

$$E_p = 0.5 \sum_k (T_k - O_k)^2 \quad (6)$$

$$E_{\text{total}} = \sum_p E_p \quad (7)$$

식 (7)로부터 계산된 오차의 합이 허용치를 만족하게 되면 반복 학습을 끝내며, 다음으로는 실제 출력기 기대되는 새로운 독립변수들을 입력하여 예측값을 출력하게 된다. 즉 오존 및 이에 관한 설명 변수로서 과거의 자료들을 학습시킨 후 해당 일의 오존 발생에 관련된 오염물질 농도와 기상 자료들이 입력되면 기억된 학습 정보를 이용하여 최종 오존 농도를 예측하는 방법이다.

여러 분야에서 많이 응용되고 있는 오차역전파 신경망은 늦은 수렴속도, 실측치와 예측치의 오차를 줄여 나가는 과정 중에서 발생하는 local minima 등이 문제될 수 있으며, 이 외에도 이전에 학습되지 않은 새로운 패턴이 입력될 경우 예측하기 힘든 점이 해결해야 할 과제로 남아 있다. 이러한 문제는 알고리즘의 기술적인 개발과 함께 적절한 자료의 효율적인 이용 등 다각적인 연구를 통하여 개선될 수 있을 것으로 사료된다.

3. 신경망모델의 적용

3.1 대상 기간 및 자료

본 연구에서는 비교적 높은 오존농도가 발생하는 하계를 대상으로 1989년부터 1991년의 5월부터 9월까지 총 15개월을 분석하였다. 대기질 자료는 서울 및 주변 도시지역에 고루 분포되어 있는 35개 대기오염 자동측정망(그림 2)에서 측정된 오존, 일산화탄소, 이산화질소, 아황산가스, 총탄화수소 등 5개 항목의 매시간 자료를 이용하였다. 지상 기상 자료로는 서울기상대에서 매시간 측정된 풍속, 풍향, 강수량 및 일사량 자료와 3시간 간격으로 측정된 기온, 상대습도, 지표면 기압 및 운량 자료를 이용하였다. 이 자료 중 3시간 간격으로 측정된 기상 자료는 단순 내삽하여 매시간 자료를 구하여 사용

하였다. 상층 기상 자료는 오산에서 1일 2회 측정되는 고도별 기압, 기온 자료를 이용하였다.

한편 강수 현상이 있는 날은 오존농도가 낮으므로 아침 8시부터 16시까지의 총 강수량이 1mm 이상인 날과 같은 시간 동안의 평균 운량이 60% 이상인 날은 대상일에서 제외하였다. 위의 방법으로부터 선정된 날은 15개월의 459일 중 총 321일로 각각 1989년에 111일, 1990년에 97일, 1991년에 113일이었다.

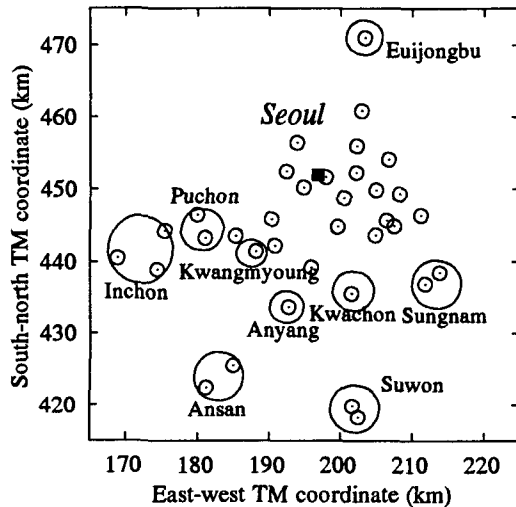


Fig. 2. Location of air pollution monitoring sites (○) and meteorological site (■).

Large circles represent groups of monitoring sites belong to each satellite city of Seoul.

3.2 입력 변수의 작성

서울시의 일 최고 오존농도를 예측하기 위하여 사용된 기상 입력 변수중 풍향은 산술적인 평균이 불가능하므로 풍향 및 풍속 자료를 이용하여 동서 성분 및 남북 성분의 풍속으로 계산하여 사용하였다. 또한 일사량이나 기온 이외에도 대기중의 수분은 대기오염 물질의 광화학 기작에 관여하는 중요한 요소이므로 기온 및 지표면 기압, 상대습도를 이용하여 절대습도를 계산하였다. 이러한 매시간 자료 중에서 기온과 절대습도는 14시에서 16시의 평균값을 이용하였으며 일사량은 16시까지의 적산량으로 하였고, 10시부터 16시까지의 동서 및 남북 성분의 풍속을 평균하여 이용하였다. 한편 혼합층고도는 오산의 상층 기상 자료와 서울의 지상 기상 자료를 이용하여 혼합층고도모델(이종범, 1991)로 계산하였으며 매일 16시의 혼합층고도를 일 최대 혼합층고도(MMD: maximum mixing depth)로 간주하여

분석에 이용하였다.

대기중으로 배출된 1차 오염물질들은 풍하측으로 수송, 확산되면서 광화학 반응을 일으켜 오존과 같은 광화학 오염물질을 생성(Altshuller, 1988) 또는 소멸(Leahey and Hansen, 1990)시키게 되므로 오존농도를 예측하기 위하여서는 오염물질 상호간의 반응 기작을 이해하는 것은 물론 풍상측의 대기질도 고려할 필요가 있다. 오존의 선구물질로서 질소산화물 및 탄화수소류는 중요한 요소로 작용하며, 우리나라의 경우에 있어서도 서울지역으로부터 배출된 선구물질이 풍하측으로 수송되어 약 80km 떨어진 춘천지역에서 고농도의 오존이 발생된 예(이태형, 1994)가 있다. 또한 이선기와 이종범(1991)은 서울지역의 오존농도 추정을 위하여 동풍계열과 서풍계열의 두 가지로 나누고 풍계별로 중회귀모델을 산정하여 개선된 결과를 제시한 바 있다. 이러한 결과는 서울에서 주간에 측정된 오존은 서울시 자체에서 배출된 오염물질에 의하여 생성될 수도 있으나, 주변의 도시지역에서 배출된 1차 오염물질이 서울지역의 상공으로 이동되어 생성되었기 때문이다.

본 연구에서는 이러한 점을 고려하여 서울지역에서 측정된 자료 이외에도 주변의 도시지역에서 오전에 측정된 1차 오염물질 자료를 입력 변수에 포함시켰다. 사용된 주변 지역의 농도 자료는 주간 동안의 평균 벡터풍향을 계산한 후 풍상측에 위치한 측정소의 농도 자료 중에서 중앙값을 취한 것이다. 서울지역의 매시간 평균 농도를 산출하는 방법으로는 해당 시간에 서울시의 20개 지점에서 측정된 자료들 가운데 중앙값을 취하여 이용하였다.

대기오염 농도를 예측하는데 있어서 대기오염 예보계 등 실용적인 면을 고려한다면 최소한 수시간 전에 농도가 예측되어야만 오염물질의 배출을 규제하기 위한 대상 지역을 설정한다든가, 자동차의 통행량을 조절한다든가 하는 규제를 취할 수 있다. 따라서 아황산가스, 총탄화수소, 일산화탄소 등 1차 오염물질 자료는 오전 7시부터 9시까지의 평균 농도를 이용하였다. 이산화질소는 초기에 배출된 일산화질소가 산화되어 약 1시간 후에 높은 농도를 나타내는 특징이 있으므로 8시부터 10시까지의 평균 농도를 이용하였다.

위의 방법으로 작성한 입력 변수는 서울지역의 오존농도 및 서울과 주변 지역 각각의 1차 오염물질 농도, 그리고 6개의 기상 자료로 전체 15개 항목이며 각 항목의 적용 시간대와 산출 방법을 표 1에 정리하였다.

Table 1. Input variables for neural network model.
* denote the concentrations measured in Seoul and it's surrounded area respectively.

Variable	Period (local standard time)	Method
Ozone(ppb)	14-16	Mean
Sulfur dioxide*(ppb)	07-09	Mean
Nitrogen dioxide*(ppb)	08-10	Mean
Total hydrocarbon*(ppm)	07-09	Mean
Carbon monoxide*(ppm)	07-09	Mean
Solar radiation(MJ/m ² /12hr)	05-16	Total
Air temperature(°C)	14-16	Mean
East-west wind component(m/sec)	10-16	Mean
South-north wind component(m/sec)	10-16	Mean
Absolute humidity(g/m ³)	14-16	Mean
Daily maximum mixing depth(km)	16	-

3.3 중회귀모델과 단단계 신경망모델의 비교

본 연구에서는 1989년과 1990년의 208일 자료를 사용하여 중회귀모델(multiple regression model, MRM)을 작성하였으며 또한 이와 동일한 자료를 학습 자료로 하여 신경망모델을 작성하였다. 이와 같이 일정한 기간의 자료를 학습시켜 얻어지는 신경망모델을 단단계 신경망모델(single step neural network model, SNM)이라 하며, 작성한 두 가지 모델의 정확도를 비교하기 위하여 1991년의 113일을 대상으로 서울의 전 지역 평균의 일 최고 오존 농도를 예측하고 실측 농도로 평가하였다. 중회귀모델 및 신경망모델에서 사용된 변수는 표 1과 같으며 이때 종속변수는 오존으로 하였고 독립변수는 나머지 14개 변수로 하였다. 중회귀모델의 작성은 SAS(Statistical Analysis System)의 변수증감법(stepwise procedure)을 이용하였다.

또한 모델간의 결과를 평가하기 위하여 회귀분석을 통한 결정계수(r^2) 및 회귀계수를 계산하였으며, 이 외에도 RMSE(root mean square error) 및 MRE(mean relative error)(Dennis et al., 1990), NMSE(normalized mean square error)(Cirillo and Poli, 1992)를 계산하여 각각의 결과를 비교하였다. RMSE와 NMSE는 식 (8)과 식 (9)로부터 계산되며 이들은 실측치(O_i)와 예측치(P_i) 사이의 분산 정도를 나타내는 척도로 그중 NMSE는 RMSE에 비하여 고농도 부근에서의 오차를 더 많이 반영하는 특징이 있다. MRE는 실측치와 예측치의 상대

적인 예측 정도를 평균한 값(식 (10))으로 0에 가까울수록 1:1 대응관계를 잘 나타낸다고 할 수 있다.

$$RMSE = \left[\frac{\sum (P_i - O_i)^2}{n} \right]^{1/2} \quad (8)$$

$$NMSE = \left[\frac{\sum (P_i - O_i)^2}{\sum P_i \cdot O_i} \right]^{1/2} \quad (9)$$

$$MRE = \left[\frac{\sum |P_i - O_i|}{\sum (P_i + O_i)} \right] \quad (10)$$

3.4 단단계 신경망모델의 평가

오차역전파 학습 알고리즘은 내삽(interpolation) 과정은 잘 수행하나 외삽(extrapolation)과정은 능률적으로 처리하지 못하는 단점(이상점 등, 1993)을 가지고 있다. 위에서 중회귀모델과 비교하기 위하여 작성한 단단계 신경망모델의 경우에는 1990년까지의 자료만으로 1991년의 오존농도를 예측하기 때문에 과거 학습 자료의 범위를 벗어나는 조건이 빈번히 발생하는 경우 예측 오차는 더욱 커질 수 있다. 그러므로 본 연구에서는 과거로부터 전일까지의 이용 가능한 모든 자료들을 학습시켜서 신경망모델을 작성하고 이 모델에 의해서 당일의 오존농도를 예측하였다. 즉 처음에는 1989년과 1990년에 측정된 208일간의 학습 자료로 모델을 작성하여 1991년의 113일 중 첫째 날을 예측하고, 다음에는 첫째 날을 포함한 209일간의 학습 자료로 모델을 다시 작성하여 둘째 날을 예측하는 방법이다. 이러한 과정을 반복적으로 진행시키면 마지막에는 320일간의 정보를 이용하여 최종일의 오존농도를 예측하게 된다. 이와 같이 매일 새롭게 작성되는 모델을 단단계 신경망모델(multi-step neural network model, MNM)이라 하며, 예측된 1991년의 113일에 대한 오존농도를 실측 농도로 평가하였다.

4. 결과 및 고찰

4.1 중회귀모델과 단단계 신경망모델의 비교

표 2는 표 1의 전체 14개 독립변수를 이용하여 수행한 중회귀 분석의 결과로서 변수증감법을 통해 선정된 최종 7개의 변수를 제시한 것이다. 이때 사용된 자료는 1989년과 1990년의 208일간의 자료이다.

기상 자료 중에서 선정된 변수는 일사량, 기온, 절대습도와 동서 성분의 바람으로 그 가운데서 일사량에 대한 부분결정계수(partial r^2)가 0.245로 가장 크게 나타났다. 또한 대기질 자료로는 서울의 주변 지역에서 측정된 총탄화수소 및 이산화질소가

선정되었고 여기에 서울지역의 일산화탄소가 포함되었다. 전체적인 중회귀분석 결과는 매우 유의하며

결정계수는 0.5 정도로 나타났다. 대기질 자료 중에서도 결정계수는 크지 않으나 서울지역 자체에서

Table 2. Result of stepwise procedure for dependent variable O₃(n=208, R²=0.5, overall F=28.68, P>F=0.001).

*; the concentrations measured in suburban area.

Variable	Parameter	Partial R ²	C(p)	P>F
Intercept	-31.89	-	-	0.0001
Solar radiation	0.01	0.245	96.39	0.0001
Total hydrocarbon*	5.35	0.129	47.11	0.0001
Air temperature	1.47	0.048	16.34	0.0001
Nitrogen dioxide*	0.52	0.040	33.36	0.0003
Carbon monoxide	-4.75	0.016	12.16	0.0152
Absolute humidity	-6.07	0.013	6.50	0.0229
East-west wind component	-1.01	0.011	9.72	0.0376

측정된 농도보다도 풍향을 고려하여 입력 변수로 추가시킨 주변 지역의 총탄화수소 및 이산화질소가 선정되었다는 것은 의미가 있다. 이것은 서울지역의 오존은 서울시 자체 뿐만 아니라 주변 도시지역으로부터 초기에 배출된 오염물질이 선구물질로 작용하여 바람을 따라 이동하면서 광화학 반응을 일으켜 생성된 것으로 판단할 수 있다. 따라서 위 결과는 서울시의 일 최고 오존농도를 예측하기 위해서는 서울지역의 자료 뿐만이 아니라 주변 지역의 자료도 포함하여 분석해야 할 필요성을 보여준 것으로 생각된다.

위에서 작성한 중회귀모델을 적용하여 1991년의 113일을 대상으로 예측한 오존농도와 실측 오존농도의 비교는 그림 3에 제시하였다. 예측된 오존농도는 실측 농도와 비교하여 약 40ppb 이상의 고농도에서 분산이 크고 기울기가 매우 완만하여 과소평가하는 결과를 나타냈다. 이러한 이유 중의 하나는 중회귀모델 산정시에 사용된 제반 조건에 비하여 모델 적용시의 오염물질 배출량의 변화 등 대기질 양상이 바뀔에 따라 나타난 결과로 사료된다. 따라서 이와 같은 중회귀모델은 일 최고 오존농도를 예측하는데 있어서 적절하지 못한 것으로 판단된다.

신경망모델의 실행에 앞서 학습횟수의 증가에 따라 식 (7)에 의하여 계산되는 총 오차(E_{total})의 감소 정도를 고찰하였다. 학습횟수에 따른 오차의 감소는 대상으로 하는 학습자료의 규모나 구조에 의하여 그 정도가 다를 수 있다. 또한 일반적으로 학습횟수가 많을수록 오차도 감소하겠지만 컴퓨터의 계산 부담을 고려하여 적절한 수준의 학습횟수

를 정할 필요가 있다. 그림 4는 신경망모델로부터 계산된 결과로서 반복횟수와 총 오차의 관계를 나타낸 것이다. 학습의 초기 단계에서 총 오차는 급격

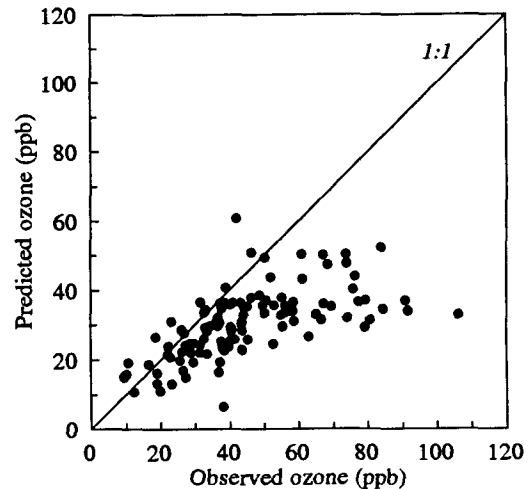


Fig. 3. Comparison between observed and predicted ozone concentration by estimated multiple regression model(MRM, n=113).

한 감소를 나타내나 약 5,000회 이상에서는 더 이상 감소하지 않고 대체로 일정한 상태를 유지하는 것으로 나타났다. 위의 결과로부터 본 연구에서는 신경망모델의 학습횟수를 5,000회로 적용하였다.

단단계 신경망모델을 이용하여 예측한 오존농도와 실측 오존농도를 비교한 결과는 그림 5와 같다. 특히 고농도 부근에서는 중회귀모델의 결과(그림

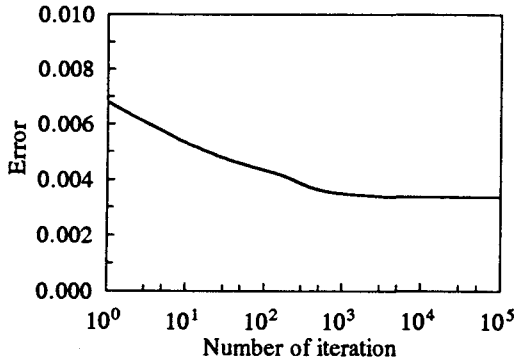


Fig. 4. Determination of the number of iteration for training set of learning patterns.

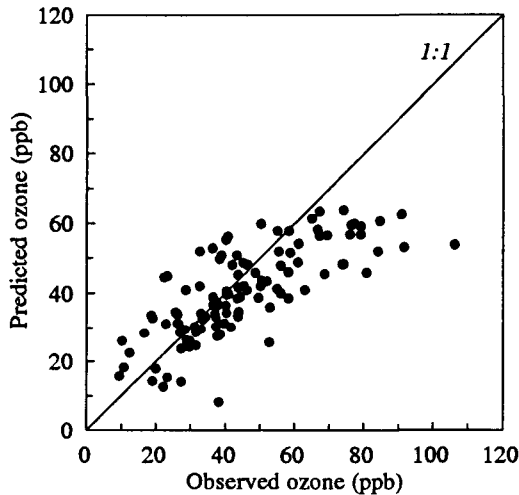


Fig. 5. Same as Fig. 3, except for single step neural network model(SNM).

3)에 비하여 훨씬 향상된 결과를 나타냈다. 그러나 50ppb 이상의 고농도 부근에서는 다소 낮게 예측하는 경향을 보이며, 이는 고농도의 오존이 발생한 경우 예측을 위한 입력 자료가 과거의 학습 자료의 범위에서 벗어나는 경우가 많았기 때문에 나타난 결과라고 생각된다. 이와 관련된 예를 그림 6에 제시하였다. 그림 6은 5월부터 9월까지 오존농도의 평균 일변화를 연도별로 나타낸 것으로 91년에 특히 높은 오존농도를 나타냈다. 이러한 이유는 오존 생성에 관련된 기상 조건이라든가 1차 오염물질의 배출 특성이 89년 또는 90년에 비하여 91년에는 크게 다르게 나타났기 때문으로 생각된다.

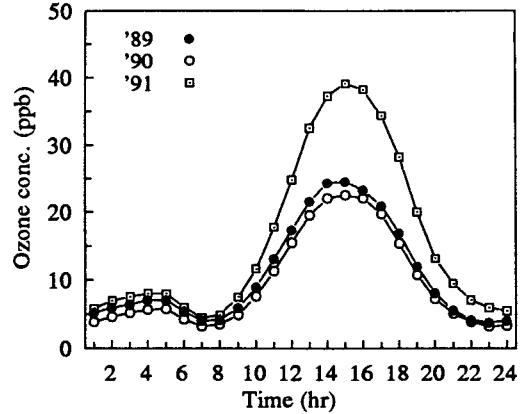


Fig. 6. Diurnal variations of ozone concentration in Seoul from May to September in 1989~1991.

4. 2 다단계 신경망모델의 평가

그림 7은 다단계 신경망모델을 이용하여 예측한 오존농도와 실측 오존농도를 비교한 것으로, 단단계 신경망모델의 결과(그림 5)에 비하여 약 40ppb 부근에서 분산이 다소 큰 경향이 있으나 그 이상의 고농도에서는 향상된 예측 결과를 나타냈다.

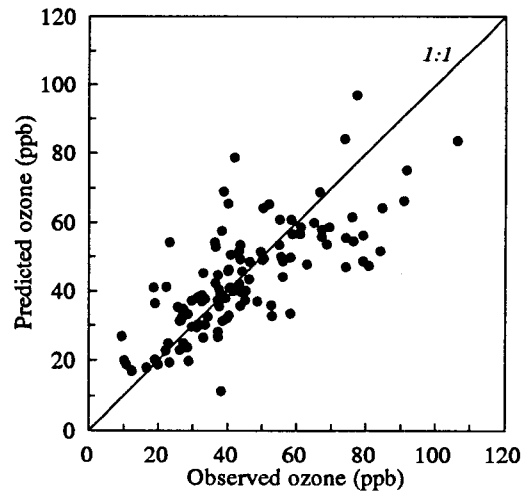


Fig. 7. Same as Fig. 3, except for multi-step neural network model(MNM).

중회귀모델 및 단단계 신경망모델, 다단계 신경망모델의 예측치와 실측치의 오차 분석 결과는 표 3과 같다. 실측치와 예측치의 분산 정도를 나타내는 RMSE는 중회귀모델에서 21.1ppb로 가장 컸으며, MRE 또한 -0.164로 큰 음의 값을 나타내 중회귀

모델은 오차가 가장 크고 과소 평가하는 결과를 나타냈다. 이에 반해 신경망모델의 오차는 전체적으로 작게 나타났다. 다단계 신경망모델의 경우 결정계수는 0.55로 단단계 모델에 비하여 0.01이 작으나 RMSE와 NMSE는 각각 13.2ppb와 0.089로 작은

오차를 나타냈으며, 특히 MRE는 0에 근접하는 결과를 나타내 실측치와 예측치의 1:1 대응관계를 잘 나타냈다. 이로부터 다단계 신경망모델은 일 최고 오존농도를 예측하는데 있어서 가장 적절한 모델로 평가되었다.

Table 3. Evaluation of the different model results(n=113).

Model	r ²	Regression coefficient	RMSE(ppb)	NMSE	MRE
MRM	0.38	0.31	21.1	0.324	-0.164
SNM	0.56	0.49	13.8	0.107	-0.037
MNM	0.55	0.60	13.2	0.089	0.003

이러한 신경망모델의 오차를 줄이기 위한 개선책의 하나는 학습 자료에 존재하지 않는 패턴을 수치적으로 재생할 수 있는 알고리즘의 기술적인 개발을 들 수 있다. 또한 모델 산출시 최근 자료의 영향이 크고 과거의 자료에 대해서는 영향이 작도록 가중값을 조절하는 방안과 학습시 자료의 범위를 크게 벗어나는 특이한 자료를 학습 자료에서 제외시키는 방법 등을 고려함으로써 예측 정확도를 향상시킬 수 있을 것으로 생각된다. 이 외에도 오존과 같은 대기오염 물질의 농도 예측은 저농도에 비하여 고농도시의 정확한 예측이 요구되므로 고농도인 경우의 학습 자료에 가중값을 크게 부여할 필요가 있다고 생각된다.

한편, 1차 오염물질 중에서 이산화질소와 일산화질소의 존재 비율, 비메탄계 탄화수소의 농도 등은 대기중의 광화학 기작을 이해하는데 매우 중요한 요소로 알려져 있으므로(Pitts and Pitts, Jr., 1993; Greenfield et al., 1992 등) 이들 농도 자료가 입력 변수로 추가된다면 모델의 예측력은 더욱 향상될 수 있을 것이다. 따라서, 우리 나라 환경처의 대기질 상시 측정망 항목에 1988년 이후 측정이 중단된 일산화질소와 비메탄계 탄화수소가 추가되는 것이 바람직하다고 생각된다.

본 연구에서 사용한 모델의 학습 자료 중 대기질 자료는 예측 대상 일의 오전에 측정된 자료를 이용하였으나 기상 자료는 16시까지의 자료를 사용하였다. 앞으로 다단계 신경망모델을 이용하여 대기오염 예보제 등에 실제로 활용할 경우 일 최고 오존농도가 발생하는 오후까지의 기상 조건을 오전에 추정하여 입력 자료로 사용하도록 하여야 하며, 오존농도의 예측 오차를 줄이기 위하여서는 기상 조건의 정확한 추정이 요구된다.

5. 결 론

중회귀분석의 변수선정 결과 서울지역에서 측정된 오존농도는 서울시 자체 뿐만이 아니라 서울의 주변에 위치한 도시지역의 이산화질소 및 총탄화수소에 의해서도 영향을 받는 것으로 나타났다. 따라서 비교적 대기오염 물질의 배출량이 많은 도시들과 인접되어 있는 서울의 경우에는 주변 지역에 의한 영향도 함께 고려하여 분석해야 할 것으로 생각된다.

단단계 신경망모델을 이용하여 1991년의 일 최고 오존농도를 예측한 결과 중회귀분석에 의한 결과보다 훨씬 유용한 방법으로 판정되었다. 그러나 고농도 부근에서는 실측 오존농도보다 다소 낮게 예측하는 경향을 나타냈다. 한편, 과거로부터 전달까지의 최근 자료를 포함하여 매일 새롭게 작성된 다단계 신경망모델을 평가한 결과 예측치와 실측치 사이의 분산 정도를 나타내는 RMSE는 다른 모델에 비하여 가장 작으며, NMSE 역시 작은 값을 나타냈다. 또한 MRE는 거의 0에 가까운 값을 나타내 비교적 정확한 예측이 가능하였으므로 대기오염 예보제 등의 실용적인 용도에도 충분히 활용될 수 있으리라 사료된다.

앞으로 본 연구에서 제시한 몇가지를 개선한다면 일 최고 오존농도를 예측하는 신경망모델의 정확도는 더욱 향상될 수 있을 것으로 생각된다.

감 사

본 연구의 신경망 기법의 적용에 관하여 조언하여 주신 인하대학교 전기공학과 이종호 교수와 연구에 필요한 대기질 자료를 제공해 주신 환경처 대기보전국의 여러분께 감사의 뜻을 표합니다.

참 고 문 헌

- 박민용, 최항식(1991), 뉴로 컴퓨터, 대영사, 301pp.
- 이상점, 김용국, 이종호(1993), 신경회로망을 이용한 오존농도의 예측에 관한 연구, 제3회 인공지능, 신경망 및 퍼지시스템 종합학술대회 논문집, 155~158.
- 이상점(1994), 신경회로망을 이용한 기상자료의 분석 및 예측에 관한 연구, 인하대학교 석사논문집, 46pp.
- 이선기, 이종범(1991), 서울지역의 광화학 오염농도 추정을 위한 중회귀모델 개발, 기상연구논문집, 8(1), 71~85.
- 이종범(1991), 중부지방 각지의 대기오염잠재력에 관한 연구. 한국대기보전학회지, 7(1), 41~47.
- 이태형(1994), 수도권 대기오염 물질이 풍하 지역의 오존농도에 미치는 영향예측을 위한 광화학 라그란지안 모델의 적용. 강원대학교 석사논문집, 57pp.
- 홍낙기, 이종범, 김용국(1993), 다변량해석을 이용한 서울시 하계 스모그의 형태분류. 한국대기보전학회지, 9(4), 278~287.
- Altshuller, A. P.(1988), Some characteristics of ozone formation in the urban plume of St. Louis, Mo. Atmospheric Environment, 20, 499~510.
- Boznar, M, M. Lesjak and P. Mlakar(1993), A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment, 27B(2), 221~230.
- Cirillo, M. C. and A. A. Poli(1992), An intercomparison of semiempirical diffusion models under low wind speed, stable conditions. Atmospheric Environment, 26A(5), 765~774.
- Dennis R. L. et al.(1990), National Acid Precipitation Assessment Program Report 5: Evaluation of regional acidic deposition models Part II. Clean Air Act Amendments.
- Greenfield, S. M., G. Z. Whitten, T. C. Myers, S. G. Douglas and J. F. Fieber(1992), Simulation of ozone reduction strategies for Mexico City utilizing the Urban Airshed Model. 85th annual A & WMA meeting, Kansas, 35pp.
- Leahey, D. M. and M. C. Hansen(1990), Observational evidence of ozone depletion by nitric oxide at 40km downwind of a medium size city. Atmospheric Environment, 24A(10), 2533~2540.
- Niemann, H.(1989), Pattern analysis and understanding(2nd ed.). Springer, New York, 199~205.
- Pitts, F. B. J. and J. N. Pitts, Jr.(1993), Atmospheric chemistry of tropospheric ozone formation: Scientific and regulatory implications. J. Air & Waste Manage. Assoc., 43, 1091~1100.
- Rumelhart, D. E. and J. L. McClelland(1986), Parellel distributed processing 1, 2. MIT Press, Cambridge, MA.
- Stern, A. C., R. W. Boubel, D. B. Turner and D. L. Fox(1984), Fundamentals of air pollution (2nd ed.). Academic Press, pp. 65~75.