

# 서울지역의 지표오존농도 예보를 위한 전이함수모델 개발

## Development of a Transfer Function Model to Forecast Ground-level Ozone Concentration in Seoul

김유근 · 손건태<sup>1)</sup> · 문윤섭 · 오인보  
부산대학교 대기과학과, <sup>1)</sup>부산대학교 통계학과  
(1999년 5월 12일 접수, 1999년 12월 6일 채택)

Yoo-Keun Kim, Keon-Tae Sohn<sup>1)</sup>, Yun-Seob Moon and In-Bo Oh  
*Department of Atmospheric Sciences, Pusan National University*  
<sup>1)</sup>*Department of Statistics, Pusan National University*  
(Received 12 May 1999; accepted 6 December 1999)

### Abstract

To support daily ground-level O<sub>3</sub> forecasting in Seoul, a transfer function model (TFM) has been developed by using surface meteorological data and pollutant data (previous-day [O<sub>3</sub>] and [NO<sub>2</sub>]) from 1 May to 31 August in 1997. The forecast performance of the TFM was evaluated by statistical comparison with O<sub>3</sub> concentration observed during September; it is shown that correlation coefficient (R), root mean squared error (RMSE), normalized mean squared error (NMSE) and mean relative error (MRE) were 0.73, 15.64, 0.006 and 0.101, respectively. The TFM appeared to have some difficulty forecasting very high O<sub>3</sub> concentrations. To compare with this model, multiple regression model (MRM) was developed for the same period. According to statistical comparison between the TFM and MRM, two models had similar predictive capability but TFM based on O<sub>3</sub> concentration higher than 60 ppb provided more accurate forecast than MRM. It was concluded that statistical model based on TFM can be useful for improving the accuracy of local O<sub>3</sub> forecast.

**Key words** Transfer function model (TFM), O<sub>3</sub> forecasting

### 1. 서 론

최근 유해성이 강한 지표오존의 고농도 현상이 중요한 대기환경의 문제로 부각되고 있다. 본 연구의 대상지역인 서울의 경우 자동차의 증가 등에 의해 오존 오염도가 심각한 수준까지 상승되었는데, 최근 3년(1995~1997) 동안 대기환경기준(100 ppb/h)을 초과하는 고농도 오존일이 63회(95년: 9회, 96년:

25회, 97년: 29회)나 이르고 있음이 조사되었다. 이에 따라 정확한 예보모델의 개발과 함께 오존 예보·경보시스템의 지속적인 개선이 요구되고 있다(김동술, 1998; 김영성, 1998; 김유근, 1998).

최근 지표오존의 농도예측에 관련된 국내연구의 경향을 살펴보면, 군집분석과 다중회귀분석과 같은 다변량 통계분석을 이용한 오존농도 예측(허정숙과 김동술, 1993), 신경망모델 및 광화학수치모델을 이용한 오존농도 예측(김용국, 1996; 김용국과 이종

범.1994), 현업운영 가능한 중회귀예보모델 개발에 관한 연구(김용준, 1997), 대기오염포텐셜 예측법을 이용한 고농도 오존일의 예측에 관한 연구(김유근, 1996), wavelet transform을 이용한 단기오존농도 예측에 관한 연구(김신도, 1998) 등 다양한 방법의 접근으로 지표 오존농도를 예측하고자 하였다.

국외의 경우를 보면, Ryan(1995)는 대도시지역인 Baltimore에서 오존 episode 예측에 관한 연구를 하였으며, Yi와 Prybutok(1996) 및 Acuna 등(1996)은 도시지역에서 신경망모델을 이용하여 일 최고 오존농도예측에 관한 연구를 수행한 바 있다. 최근에는 Hubbard 등(1998)이 미국 켄터키주의 Louisville을 대상으로 지표오존의 일 최고농도를 예측하기 위한 다중선형회귀 모델을 개발하였고, Torquera 등(1998)은 칠레 산티에고에서 일 최고 오존농도의 예측에 있어 새가지 모델(시계열모델, 신경망모델, fuzzy 모델)을 사용하여 비교연구를 하였다.

본 연구에서는 상기의 여러 연구를 바탕으로 현 오존예보시스템을 보완·개선할 수 있는 서울지역의 지표오존 단기 예보모델을 개발하고자 하였는데, 이를 위해 통계적 기법인 전이함수모형(TFM; transfer function model)을 이용하였다. 이 모형은 오존과 같이 대기오염물질과 기상요소 등 여러 개의 인자가 오존농도에 영향을 미칠 때 적합한 시계열 분석 방법으로, 모델입력자료의 부족으로 수치모델의 적용이 어려운 현 국내 오존예보 실정에 적합할 것으로 생각된다.

## 2. 연구 방법

본 연구에서 모델개발에 적용한 연구기간은 최근

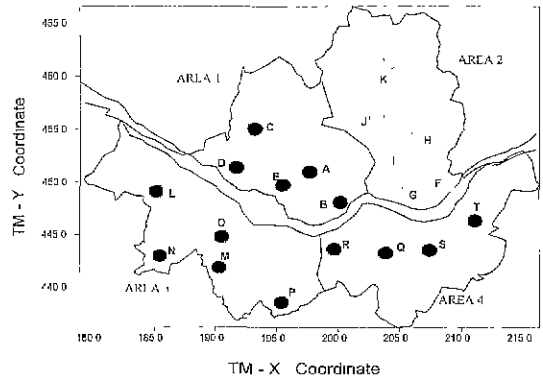


Fig. 1. Map of forecast area and air quality monitoring sites in Seoul. Forecast area and monitoring sites are marked as AREA 1~4 and A~T, A: City Hall, B: Hanam, C: Bulgkwang, D: Namgajwa, E: Mapo, F: Kuyui, G: Sungsoo, H: Myunmog, I: Shinseol, J: Kilm, K: Ssangmun, L: Hwangok, M: Kuro, N: Oryu, O: Munrae, P: Shinlim, Q: Taechi, R: Banpo, S: Chamsil, T: Bangi.

의 오존농도경향을 반영하기 위해 1997년을 택하였고, 현재의 예보기간인 5~9월로 설정하였다. 이러한 이유는 대부분의 유의농도(60 ppb 이상)가 5~9월 중에 발생하여 이 기간동안의 오존예보가 매우 중요하기 때문이다. 표 1은 서울지역의 현 오존예보 권역(그림 1)을 대상으로 3년(1995~1997년)동안 60 ppb 이상의 유의농도가 발생한 빈도를 월별로 조사한 것으로 90% 이상의 대부분의 유의농도가 5~9월 중에 발생했음을 나타내고 있다. 이렇게 본 연구에서 설정한 연구기간 중 전반부 122일은 예보모델을 구축하기 위한 모형화자료로, 후반부 30일은 구축된 예보모델의 검증자료로 나누어 사용하였다.

Table 1. Monthly distribution of number of days in which O<sub>3</sub> monitors in Seoul observed [O<sub>3</sub>] in excess of 60 ppb (one hour average) for the period 1995~1997.

Area	Month												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
I	-	-	-	16	29	44	34	35	20	4	4	-	186
II	-	5	26	54	41	63	42	51	38	16	4	-	340
III	4	1	6	16	21	47	26	33	23	7	1	-	185
IV	-	5	3	34	43	55	43	44	35	10	-	-	272
Total	4	11	35	120	134	209	145	163	116	37	9	-	983
Mean (Total/3Years×4Arcas)	0.3	0.9	2.9	10	11.2	17.4	12.1	13.6	9.7	3.1	0.8	-	81.9

예보모델의 입력자료를 살펴보면, 현재 서울지역 기상청 지상관측자료 중 8개의 기상인자(일최고온도, 풍속, 운량, 일사량, 상대습도, 강수량, 이슬점온도, 수증기압)와 현 예보권역의 대기오염자동측정망에서 관측되고 있는 오존(O<sub>3</sub>) 및 이산화질소(NO<sub>2</sub>)농도자료를 설정하였다. 표 2에서는 사용된 입력변수와 그 생성방법 및 적용시간을 나타내었는데, 여기서 적용된 평균 간격은 입력변수의 예측가능성을 고려하였고 오존농도와 상관성이 높은 시간대를 선택하였다. 이러한 입력변수는 오존의 생성·소멸에 크게 관여하고 있는 인자로서, 특히 일최고온도와 전일 오존농도의 경우 예측일의 오존농도와 가장 깊은 관련이 있으며 모델의 중요한 입력변수로 사용됨을 Robeson과 Steyn (1990), Hubbard와 Cobourn (1998) 등의 연구에서 알 수 있고 현 예보모델(중회귀모델)의 입력자료로도 사용되었다. 표 3은 1999년 5~9월을 대상으로 입력변수와 오존농도와의 상관관계를 분석한 결과로 일최고기온( $\bar{R}=0.573$ )과 전일 오존농도( $\bar{R}=0.538$ )가 상관성이 높은 것으로 나타나 그 중요성을 알 수 있다.

풍속의 경우 풍향과 관련해 오존전구물질의 이동과 오존의 형성에 영향을 미치는 기상인자이며 (Mckendry, 1993; 정용승, 1991), 오존과 약한 음의 상관( $\bar{R}=-0.240$ )을 보임을 표 3에서 알 수 있다. 또

한 일사량과 운량의 경우 대기의 일기상태를 반영하는 것으로 각각 오존과 비교적 높은 양의 상관( $\bar{R}=0.479$ )과 음의 상관( $\bar{R}=-0.374$ )을 보였고, 대기 중의 수분을 나타낼 수 있는 강수량, 상대습도, 이슬점온도 및 수증기압은 오존의 광화학메카니즘에 관여하는 중요한 인자로서 (Hubbard and Cobourn, 1998) 본 연구에서 예보모델 입력자료로 채택하였다. 이산화질소의 경우 오존생성 메카니즘에 있어 중요한 인자로서 여러 예측모델의 입력자료로 사용되고 있는데 (Jorquera *et al.*, 1998; 허정숙과 김동술 1993), 본 연구에서는 예측의 어려움으로 인해 전일 권역별 일최고 농도를 변수로 사용하였다.

이러한 입력변수들을 바탕으로 한 예보모델의 개발단계를 살펴보면, 첫 번째 단계로 서로 상관성이 높은 기상입력변수간의 독립성확보를 위해 8개의 기상인자를 대상으로 하여 주성분분석을 수행하였는데 이는 상관된 변수의 단순화와 서로간에 존재할 수 있는 다중공선성(multicollinearity)을 해결하기 위함이다 이렇게 얻어진 주성분과 대기오염물질 자료를 입력시계열로 하여 두 번째 단계인 TFM 분석을 수행하였고, 최종적으로 최적의 TFM을 예보모델로 결정하고 모델의 검진 및 검증과정을 실시하였다. 또한 본 연구에서는 개발된 모델의 적합성을 평가하기 위해 동일 입력자료와 기간을 이용, 중회귀모델(MRM; multiple regression model)을 구축하여 그 예측결과와 비교 검토하였다.

Table 2. List of input predictor variables.

Predictor	Content and timing	Unit
O <sub>3</sub>	Previous-day maximum 1 h [O <sub>3</sub> ] by areas	ppb
NO <sub>2</sub>	Previous-day maximum 1 h [NO <sub>2</sub> ] by areas	ppb
TMAX	Daily maximum 1 h surface temperature	°C
WS	Average wind speed (from 06h to 18h)	m/s
CL	Average cloud amount (from 06h to 18h)	-
RAD	Sum of irradiation (from 06h to 18h)	MJ/m <sup>2</sup>
RH	Average relative humidity (from 06h to 18h)	%
RAIN	Sum of precipitation (from 06h to 18h)	mm
TD	Dew point temperature at 15h	°C
VP	Water vapor pressure at 15h	hPa

### 3. 전이함수모형

전이함수모형은 하나의 시계열이 다른 여러 개의 시계열에 의하여 영향을 받는 경우에 사용하는 시계열분석 방법으로, 입력계열(x<sub>t</sub>, x<sub>t-1</sub>, x<sub>t-2</sub>, ...)의 값이 출력계열 y<sub>t</sub>에 전이된다는 의미를 가지고 있다. 일반적으로 분포시차모형에서 y<sub>t</sub>의 현재수준은 식 (1)과 같이 x<sub>t</sub>계열의 많은 과거의 값의 함수로써 나

Table 3. Correlation coefficients between O<sub>3</sub> concentration and main variables at each site of seoul.

Variables O <sub>3</sub> (Area)	previous-day								
	O <sub>3</sub>	NO <sub>2</sub>	TMAX	WS	RAD	RH	RAIN	TD	VP
I	0.541	0.319	0.640	-0.252	0.492	-0.296	-0.292	0.233	0.218
II	0.517	0.514	0.574	-0.194	0.446	-0.206	-0.216	0.241	0.261
III	0.553	0.409	0.509	-0.282	0.496	-0.293	-0.253	0.119	0.098
IV	0.542	0.143	0.567	-0.231	0.482	-0.180	-0.161	0.256	0.263

타내어진다.

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + N_t = v(B)x_t + N_t \quad (1)$$

여기서,  $x_t$ 와  $y_t$ 는 적절히 변환된 시계열로 둘 다 정상성을 만족하며,  $N_t$ 는 잡음과정 (noise process)으로 입력계열  $x_t$ 와 독립이고 일반 ARMA (p,q) (ARMA; autoregressive moving average)과정을 따른다. 여기서  $v_0, v_1, v_2 \dots$ 들은 전이함수가중치 (transfer function weight)로서,  $x_t$ 의 1단위 변화가  $y_t$ 에 주는 영향을 나타내는 계수이다. 그리고  $v(B)$ 는 전이함수이며, 이를  $B$  (backward operator)에 대한 다항식으로 표현하면 식 (2)와 같다.

$$v(B) = v_0 + v_1 B + v_2 B^2 + \dots = \sum_{i=0}^{\infty} v_i B^i \quad (2)$$

식 (1)은 유한개의 관측치를 가지고 있는 시계열  $y_t$ 에 무한개의  $v_i$ 를 적합시킨 모형으로  $v_i$ 에 대한 추정치가 불가능하다. 따라서 식 (3)과 같이  $v(B)$ 를 적은 수의 모수로 구성된 다항식 ( $\omega_s(B), \delta_r(B)$ )의 비로 표현하면 모형을 식별할 수 있다.

$$v(B) = \frac{\omega_0 - \omega_1 B - \dots - \omega_b B^b}{1 - \delta_1 B - \dots - \delta_r B^r} B^b = \frac{\omega_s(B)}{\delta_r(B)} B^b \quad (3)$$

여기서  $b$ 는 시차의 경과를 나타내는 지연모수이다. 결과적으로 백색잡음에 대한 적절한 모형인  $N_t$ 의 모형을  $\phi(B)N_t = \theta(B)a_t$ 라고 가정하면 전이함수 모형은 식 (4)와 같이 표현되고, 이를 도식적으로 간략히 나타내면 그림 2와 같다

$$y_t = \frac{\omega_s(B)}{\delta_r B} B^b x_t + \frac{\theta(B)}{\phi(B)} a_t \quad (4)$$

본 연구에서는 2개 이상의 입력계열을 가지고 있

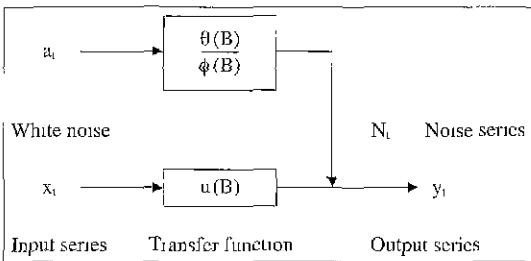


Fig. 2. Outline of Transfer function model (TFM).

는 다중입력 전이함수모형을 고려하였는데  $k$ 개의 입력계열이 있을 때 식 (1)은 식 (5)와 같이 표현할 수 있다.

$$y_t = v_1(B)x_{1t} + v_2(B)x_{2t} + \dots + v_k(B)x_{kt} + N_t = \frac{\omega_{s1}(B)}{\delta_{r1}(B)} B^{b1} x_{1t} + \frac{\omega_{s2}(B)}{\delta_{r2}(B)} B^{b2} x_{2t} + \dots + \frac{\omega_{sk}(B)}{\delta_{rk}(B)} B^{bk} x_{kt} + \frac{\theta(B)}{\phi(B)} a_t \quad (5)$$

여기서 각각의 입력계열은 서로간의 독립이고  $N_t$ 와도 독립이라고 가정한다. 이러한 다중입력 전이함수모형을 결정하기 위해 입력계열 ( $x_{kt}$ ) 들과 출력계열 ( $y_t$ )간에 각각의 교차상관함수 (CCF; cross correlation function),  $\hat{\rho}_{y, x_k}$ 를 계산하여 전이함수가중치  $\hat{v}_k$ 를 추정하게 되고,  $\hat{\rho}_{y, x_k}$ 와 비례관계에 있는  $\hat{v}_k$ 의 패턴과 이론적  $v_k$ 의 패턴을 비교 검토함으로써 지연모수인  $b_k$ 와  $\delta_{rk}(B), \omega_{sk}(B)$ 를 식별하게 된다. 그리고  $b_k, r_k, s_k$ 가 선택되면  $\delta_{rk}(B)v_k(B) = \omega_{sk}(B)B^{b_k}$ 의 관계식으로부터  $v_k(B) = \frac{\hat{\omega}_{sk}(B)}{\delta_{rk}(B)} B^{b_k}$ 를 얻게 되고, 이렇게 초

기전이함수  $\hat{v}_k(B)$ 이 얻어지면 잡음과정인  $N_t$ 의 모형식을 추정할 수 있다 이때  $N_t$ 에 대한 적절한 모형은 식 (6)으로 얻어진  $\hat{N}_t$ 에 대한 ACF와 PACF를 이용하여 식별할 수 있다.

$$\hat{N}_t = y_t - (v_1(B)x_{1t} + v_2(B)x_{2t} + v_3(B)x_{3t} + \dots + \hat{v}_k(B)x_{kt}) \quad (6)$$

마지막으로 백색잡음 모형을 포함시킨 최종모형을 구축하고 시계열모형의 모수추정방법을 사용하여 각각의  $\omega, \delta, \theta, \phi$ 를 추정한다. 그리고 추정결과를 이용하여 잔차에 대한 자기상관 유무를 통해 모형의 적합성 검정을 수행하게 된다

## 4. 연구 결과

### 4.1 주성분 분석

모형 개발의 첫 번째 단계로 변수들 사이의 독립성확보를 위해 8개 기상인자간의 상관행렬을 이용한 주성분 분석을 수행하였다. 그 결과 표 4(a)에서와 같이 전체분산을 대부분 설명할 수 있는 (누적비율이 0.85 이상) 3개의 주성분 (PRIN1 ~ PRIN3)이 선택되었다. 이러한 주성분들은 여러개의 인자들로

**Table 4. Principal component selected by principal component analysis (PCA) (a) and component loadings based on correlation matrix input (b).**

(a)			
Factor	Eigenvalue	Accounted-for "Variance"	Cumulative percentage of total "Variance"
PRIN1	3.50084	0.437605	0.43760
PRIN2	2.29330	0.28666	0.72427
PRIN3	1.03637	0.129546	0.85381
PRIN4	0.64163	0.080204	0.93402
PRIN5	0.28068	0.035085	0.96910
⋮	⋮	⋮	⋮

(b)			
	PRIN1	PRIN2	PRIN3
TMAX	0.042744	0.632228	-0.015980
WS	-0.041061	-0.091763	0.933659
CL	0.402469	-0.271993	0.131911
RAD	-0.406797	0.336828	0.006288
RH	0.493417	-0.08147	0.054286
RAIN	0.278480	-0.272233	-0.316581
TD	0.420526	0.398786	0.048734
VP	0.413725	0.404380	0.070988

구성되어 있는데, 주성분을 구성하고 있는 인자들의 기여도는 표 4(b)에서와 같이 각 인자의 고유벡터 값으로 알 수 있다. 본 연구에서는 이렇게 선택된 주성분과 권역별 전일 O<sub>3</sub>와 NO<sub>2</sub> 농도를 전이함수로

형의 입력시계열로 사용하였다

**4. 2 전이함수 모형분석**

본 연구에서 지표오존 예측모델의 권역별 구축을 위한 전이함수모형 결정단계를 살펴보면 우선 입력계열 x<sub>kt</sub> (PRIN1~PRIN3, O<sub>3</sub>, NO<sub>2</sub>)와 출력계열 y<sub>t</sub> 간에 각각의 표본교차상관함수  $\hat{\rho}_{xy}$ 를 계산하여 전이함수가중치  $\hat{v}_k$ 를 추정하였다. 그리고 교차상관함수를 통해 b<sub>k</sub>와 δ<sub>ik</sub>(B), ω<sub>sk</sub>(B)를 식별하여 초기전이함수인  $\hat{v}_k(B)$ 를 구하게 되고 식 (6)를 통해 백색잡음 모형인 N<sub>t</sub>를 추정하였다. 추정된 N<sub>t</sub>모형의 이동평균 부분(MA)의 차수는 없었고 자기회귀부분(AR)의 lag time은 I권역은 AR (12), II권역은 AR (12)와 AR (15), III권역은 AR (0), IV권역은 AR (3)으로 나타났다.

마지막으로 백색잡음 모형을 포함시킨 최종모형을 구축하고 모수추정방법의 하나인 조건부 최소제곱방법을 통해 권역별 각각의 모수 ω, δ, θ, φ와 오차분산을 추정하였다. 그리고 모형의 유의성 검정을 위해 잔차(a<sub>t</sub>)들의 자기상관함수값과 난수성(radomness)점정을 위한 퍼트먼트우 통계량 값을 분석하였는데, 그 결과 표 5에서와 같이 잔차들이 백색잡음과정을 따르고 있음을 알 수 있었다. 이러한 분석 결과를 바탕으로 최종적으로 표 6에서와 같이 서울

**Table 5. Autocorrelation check of residual (Area I).**

Lag	Square	DF	Prob						
6	10.10	5	<b>0.072</b>	-0.060	-0.024	0.133	-0.045	-0.028	0.235
12	13.00	11	<b>0.294</b>	0.071	-0.097	0.027	0.045	-0.071	-0.005
18	19.69	17	<b>0.290</b>	0.105	-0.082	-0.130	0.050	-0.104	-0.019
24	30.67	23	<b>0.131</b>	0.087	-0.077	-0.223	-0.044	-0.089	-0.039

**Table 6. Transfer function model (TFM) for Seoul.**

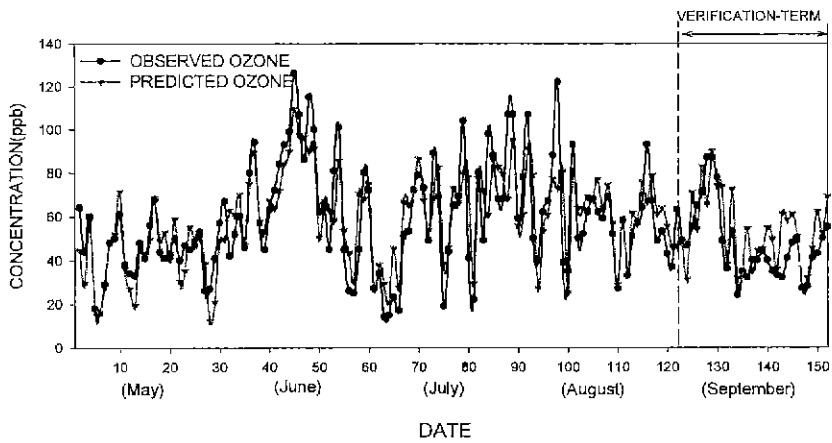
Area	Model
I	$\hat{y}_t = 29.543 + \frac{0.517}{(1 + 0.856B)} O_3 + \frac{0.162}{1 - 0.489B} NO_2 - \frac{4.456}{1 + 0.559B} PRIN1 + 5.957PRIN2 - \frac{5.464}{1 + 0.872B} PRIN3 + \frac{1}{(1 + 0.210 B^{12})} a_t$
II	$\hat{y}_t = 49.048 + \frac{0.206}{1 + 0.952B} O_3 + 0.328NO_2 + 8.719PRIN2 - \frac{3.069}{1 + 0.915B} PRIN3 + \frac{1}{(1 + 0.210 B^{12}) (1 + 0.294 B^{15})} a_t$
III	$\hat{y}_t = 32.638 + 0.356O_3 + 0.144 NO_2 - 3.515PRIN1 + \frac{6.473}{1 + 0.381B} PRIN2 - 4.756PRIN3 + a_t$
IV	$\hat{y}_t = 45.338 + 0.293O_3 - 3.344PRIN1 + 8.390PRIN2 - 5.530PRIN3 \frac{1}{1 - 0.349B} + a_t$

지역의 권역별 TFM 예보모델이 개발되었다.

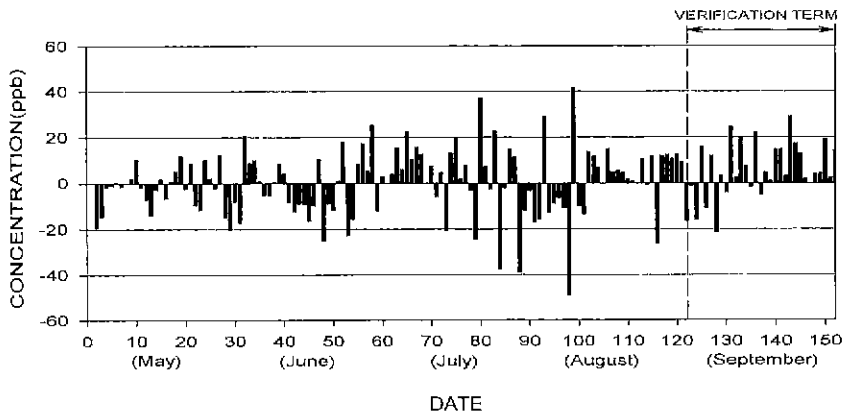
### 4. 3 예측결과 및 검증

본 연구에서 제시한 TFM 예보모델을 사용하여 익일의 지표오존의 농도를 권역별로 예측하여 보았다. 그림 3은 I권역을 대상으로 한 총 연주기간(152일) 동안의 예측치와 실측치간의 시계열 그림(a)과 예측치와 실측치간의 오차를 나타낸 그림(b)으로 검증기간(30일)을 포함하고 있다. I권역의 경우 예측치와 관측치가 매우 유사한 변동경향을 보이면서 비교적 좋은 예측결과를 보였고 나머지 세 권역 역

시 유사한 결과를 나타내었다 하지만 예측오차를 나타낸 그림에서 알 수 있듯이 고농도가 자주 발생한 7~8월 동안에는 예측치와 관측치간의 음의 차이가 커 고농도예측 부분에서는 다소 낮게 예측하는 경향이 있음을 알 수 있다. 이러한 결과는 II, III, IV권역에서도 나타남을 그림 4에서 알 수 있는데 그 원인을 두가지로 분석할 수 있다. 첫째 고농도 오존일의 자료수가 적어 모델의 설명력이 저하되었기 때문이고, 둘째 급격한 오존농도의 상승을 설명해 줄 수 있는 국지적 특성인자나 전구물질의 자료가 변수에 포함되지 못하였기 때문으로 분석될 수 있다.



(a)



(b)

Fig. 3. Time series of predicted ozone concentration using TFM model and observed ozone concentration at area I in Seoul (a) and plotting of error (predicted value-observed value) (b).

또한 검증기간동안의 예측농도치가 모델 구축기간에 비해 다소 실측치보다 높게 나타났는데, 이는 시계열모형의 특성상 검증기간동안의 자료가 모형의 입력자료로 사용되지 않아 앞부분의 고농도 경향이 뒤 부분으로 전이되어서 발생한 것으로 판단된다. 이러한 문제점은 변화된 최근 측정자료의 입력을 통한 지속된 모델의 보정으로 해결할 수 있을 것으로 생각된다.

그림 4와 5는 모델구축기간과 검증기간을 나누어 예측치와 실측치사이의 관계를 나타낸 산포도이다

모델구축기간동안의 검진결과 상관도(R)가 I권역이 0.82로 가장 좋게 나타났으며, II권역은 0.73, III권역은 0.74, IV권역은 0.73으로 나타나 비교적 우수한 모델이 구축되었음을 알 수 있었다 또한 검증결과 역시 R이 I권역은 0.72, II권역은 0.69, III권역은 0.73, IV권역은 0.82으로 나타나 비교적 좋은 결과를 보였다. 결과적으로 구축기간동안은 평균 0.75, 검증기간동안은 권역별로 차이가 다소 있지만 평균 0.73으로 모델의 신뢰성을 나타내고 있다. 여기서 각 권역마다 약간의 예측도의 차이가 있음을 알 수

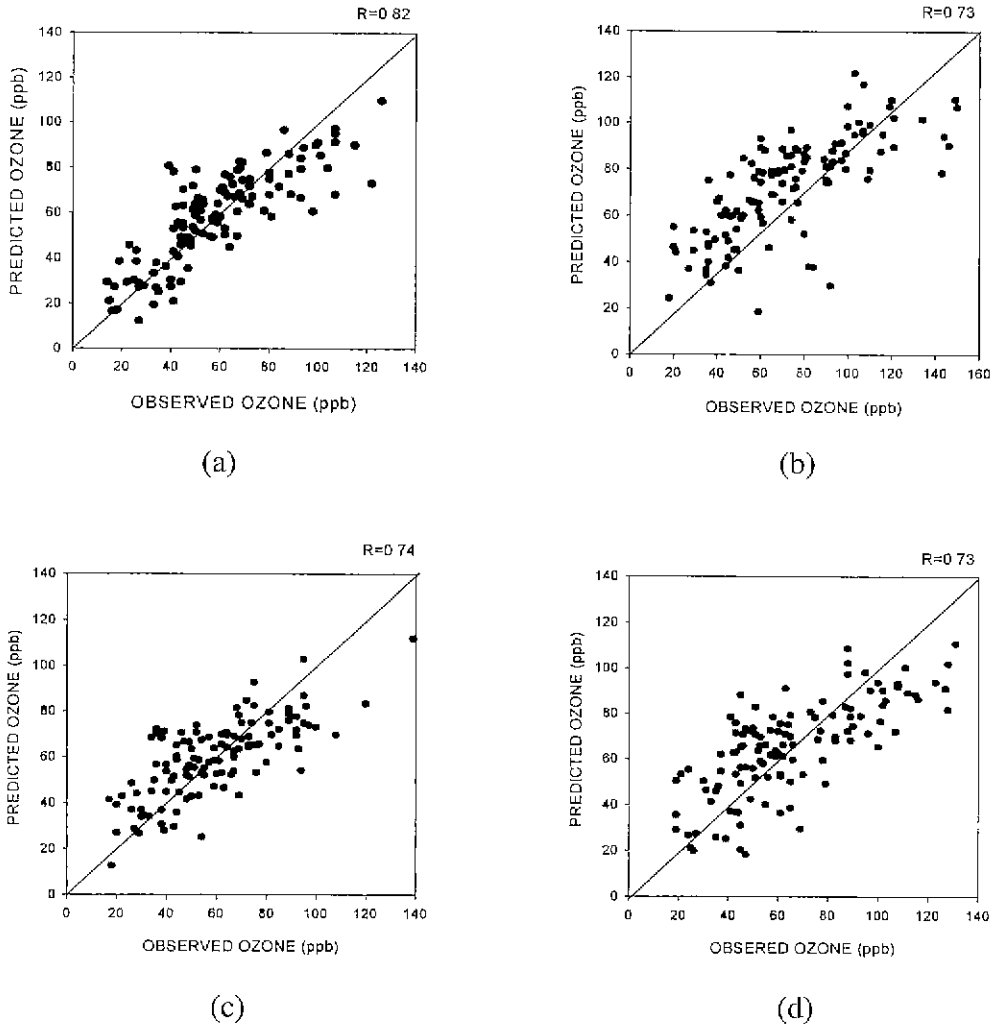


Fig. 4. Comparison between predicted ozone concentration using TFM and observed ozone concentration in Seoul for test period (122 days) (a: Area I, b: Area II, c: Area III, d: Area IV).

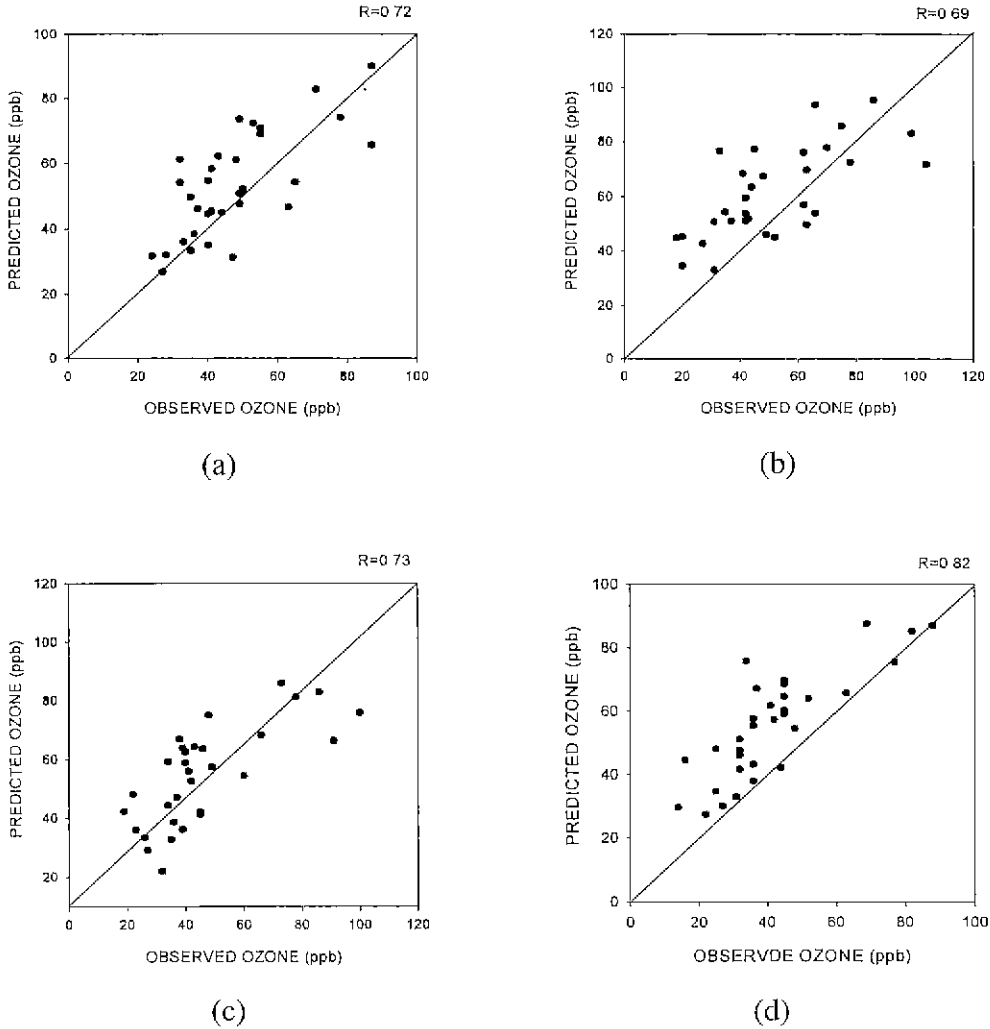


Fig. 5. Comparison between predicted ozone concentration using TFM model and observed ozone concentration in Seoul for verification period (30 days) (a: Area I, b: Area II, c: Area III, d: Area IV).

있는데, 이는 지역별 오존농도 상승의 기여인자가 다르고 오존전구물질의 이동 및 지형적 특성 때문으로 분석할 수 있다.

4. 4 모델의 평가

본 연구에서는 TFM 예보모델의 평가를 위해 동일자료와 기간을 바탕으로 중회귀모델(MRM)을 개발하여 비교하였다. MRM 개발을 위해 권역별 모델의 독립변수는 Stepwise 기법으로 선택하였고(유의

수준 0.05), 이를 바탕으로 권역별로 회귀방정식이 구축되었다(표 7). 그림 6은 I권역을 대상으로 한 MRM의 예측치와 실측치간의 시계열 그림(a)와 예측오차를 나타낸 그림(b)로 전반적인 변동패턴은 TFM 예보모델과 유사함을 알 수 있다. 하지만 TFM 과 비교하여 고농도발생 일이 많은 계절인 7~8월 동안에는 예측된 오염도의 변동폭이 관측값보다 작은 경향이 나타났고 예측오차 역시 음의 차이가 커 고농도 예측에 문제점을 들어냈다 따라서 상대적으로



Table 7. Multiple regression model (MRM) for Seoul

Area	Model
I	$Y = 42.621 + 0.234O_3 + 0.178NO_2 + 1.535TMAX - 6.526WS - 1.831CL - 0.317RH$
II	$Y = 0.307 + 0.189O_3 + 0.291NO_2 + 2.178TMAX - 2.564CL$
III	$Y = 30.851 + 0.412O_3 - 7.998WS + 1.577RAD$
IV	$Y = 4.237 + 0.315O_3 + 1.219TMAX - 6.655WS + 1.525RAD$

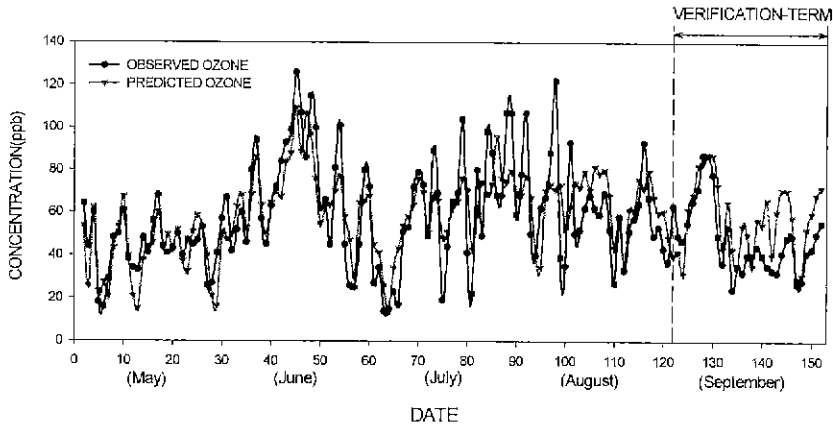
로 고농도 발생이 많지 않은 5월이나 검증기간인 9월의 예측결과가 좋게 나타났고, 이는 MRM의 경우

오염도의 변동폭이 적은 계절에는 예측결과가 비교적 잘 적응할 것으로 추정된다.

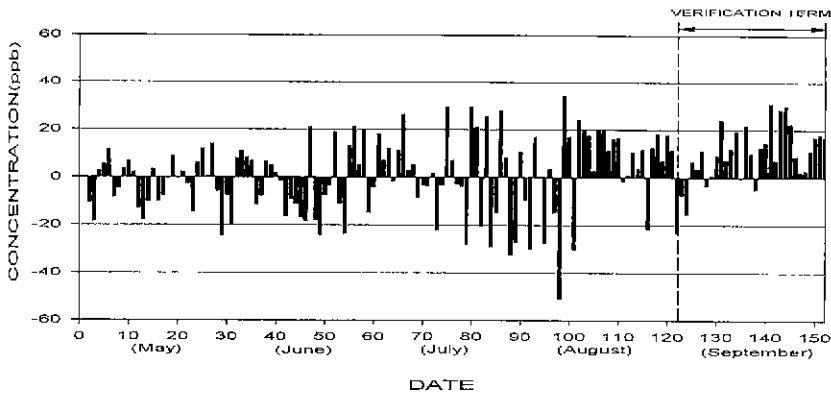
또한 모델간의 결과를 평가하기 위해 예측치와 실측치사이의 분산 정도를 나타내는 척도인 RMSE (root mean square error)와 NMSE (normalized mean square error), 예측치와 실측치사이의 1:1 대응관계를 나타내는 MRE (mean relative error)를 계산하여 분석하였다. 각각의 식은 다음과 같다.

$$RMSE = [\sum(P_i - O_i)^2/n]^{1/2} \tag{7}$$

$$NMSE = \frac{[\sum(P_i - O_i)^2/n]^{1/2}}{(\sum P_i/n)(\sum O_i/n)} \tag{8}$$



(a)



(b)

Fig. 6. Time series of predicted ozone concentration using multiple regression model (MRM) and observed ozone concentration at area I in Seoul (a) and plotting of error (predicted value-observed value) (b).

Table 8. Evaluation of TFM and MRM for Seoul Model (Area).

Model (Area)		Test period (122 days)				[O <sub>3</sub> ] > 60 ppb	Verification period (30 days)			
		R	RMSE	NMSE	RME	RMSE	R	RMSE	NMSE	RME
TFM	I	0.82	14.25	0.0041	0.0090	15.74	0.72	13.14	0.0052	0.0643
	II	0.72	21.15	0.0041	0.0236	22.31	0.69	16.53	0.0053	0.1012
	III	0.74	15.71	0.0046	0.0196	16.16	0.73	15.83	0.0060	0.0867
	IV	0.73	18.73	0.0045	0.0165	18.90	0.82	17.05	0.0072	0.1523
	Mean	0.75	17.46	0.0043	0.0172	18.27	0.73	15.64	0.0060	0.1011
MRM	I	0.77	15.90	0.0046	0.0150	18.21	0.77	15.24	0.0056	0.1022
	II	0.67	22.62	0.0044	0.0250	23.88	0.87	16.21	0.0049	0.1259
	III	0.72	16.29	0.0048	0.0140	17.36	0.71	14.34	0.0058	0.0538
	IV	0.69	19.79	0.0048	0.0180	21.09	0.82	14.38	0.0064	0.1254
	Mean	0.71	18.65	0.0047	0.0180	20.14	0.79	15.04	0.0057	0.1018

$$MRE = \sum \frac{(P_i - O_i)/(P_i + O_i)}{n} \quad (9)$$

식 (7), (8), (9)에서  $P_i$ 는 예측치,  $O_i$ 는 실측치를 그리고  $n$ 은 자료의 수를 나타낸다.

표 8에서는 TFM과 MRM 예보모델의 예측결과를 대상으로 모델 구축기간과 검증기간으로 나누어 R, RMSE, NMSE, MRE를 분석한 결과를 나타내고 있다. 모델 구축기간 동안에는 TFM의 R값(평균 0.75)이 MRM의 R(평균 0.71)보다 좋게 나타났고 RMSE, NMSE, RME 값 역시 MRM의 경우보다 대부분 낮게 나타나 TFM 예보모델이 더 우수함을 알 수 있다. 특히 유의농도 이상(60 ppb)의 고농도오존일을 대상으로한 평균 RMSE수치는 TFM이 18.27 ppb, MRM이 20.14 ppb로 각각 나타나 TFM 예보모델이 고농도오존의 예측에 있어 MRM보다 더욱 설명력이 좋을 수 있었다. 반면에 검증기간 동안은 두 모델의 R, RMSE, NMSE, MRE값이 유사하게 나타났는데, 이는 MRM의 검증결과가 구축된 모형의 설명력보다 높게 나타난 결과이다. 이러한 결과는 앞의 시계열분석결과와 마찬가지로 예측치의 변동폭이 적은 MRM의 특성으로 볼 수 있다. 그리고, 특징적으로 RME값을 보면 두 모델 모두 양의 값을 보여 전반적으로 실측치보다 다소 높게 예측하는 경향을 보였다.

결과적으로 표 8의 분석결과에서 구축된 모델의 전반적인 설명력은 MRM에 비해 TFM 예보모델이 우수한 것을 알 수 있었고, 또한 과거입력자료의 영향을 많이 받는 전이함수모형의 특성상 최근 측정

자료의 입력을 통한 지속된 모델의 보정이 이루어진다면 보다 나은 예측결과를 얻을 수 있을 것으로 분석된다.

## 5. 결 론

본 연구에서는 8개의 기상요소 및 전일의 O<sub>3</sub>와 NO<sub>2</sub>농도를 입력변수로 하여 서울지역의 일 최고 오존농도를 예측할 수 있는 권역별(4개) TFM 예보모델을 개발하였다. 구축된 4개의 TFM 예보모델은 평균상관도 R이 0.75, 평균 RMSE가 17.46 ppb로 나타나 비교적 좋은 예측결과를 보였다. 권역별 검증결과 역시 평균 R과 RMSE가 각각 0.73과 15.64 ppb로 나타나 모델의 적합성을 나타내었다. 또한 TFM 예보모델의 평가를 위해 MRM을 구축하여 비교한 결과( $\bar{R}=0.71$ ). 전반적인 예측률은 크게 차이가 없었지만 예측치와 관측치사이의 시계열 및 예측오차분석을 통해 TFM을 사용한 예측치가 실제 오존의 변동패턴을 비교적 잘 설명해 주고 있음을 알 수 있었다. 또한 유의농도(60 ppb) 이상의 고농도 오존일을 대상으로 구축된 TFM과 MRM의 예측오차를 분석한 결과 평균 RMSE가 18.27 ppb와 20.14 ppb로 나타나 고농도 오존일 예측에 있어 TFM 예보모델이 적합함을 알 수 있었다. 지표오존은 국지적 특성에 따라 그 오염도의 변화가 심하고, 생성·소멸에 있어 복잡한 광화학 반응을 가지고 있기 때문에 그 농도를 정확히 예측하기에는 사실상 불가능하다. 하지만 국지적 특성을 잘 고려할 수 있는 인자를 포함한 입력변수의 설정과 함께 예보모델의

개선이 지속적으로 이루어진다면 오존예측의 정확성은 더욱 향상될 것이다.

### 감사의 글

본 연구는 환경부/과학기술처 G-7 프로젝트 세부 과제인 “지구대기 조성변화 감시” 중 수도권 지역을 대상으로 하는 “국지성을 고려한 지표오존 예측 모델에 관한 연구”로서 환경부/과학기술처/기상연구소의 지원으로 이루어졌다.

### 참고 문헌

김동술 (1998), 오존예보모델 및 예보시스템의 개선, 오존 예보시스템에 관한 전문가토론회, 42-50.  
 김신도 (1998), 오존예보모델 및 예보시스템의 개선, 오존 예보시스템에 관한 전문가토론회, 16-23  
 김영성 (1998), 오존예보모델 및 예보시스템의 개선, 오존 예보시스템에 관한 전문가토론회, 51-58  
 김용국 (1996), 수도권 지역의 광화학 오존농도 예측에 관한 연구, 강원대학교 박사학위논문.  
 김용국, 이종남 (1994), 하계의 일 최고 오존농도 예측을 위한 신경망모델의 개발, 한국대기보전학회지, 10(4), 224-232.  
 김용준 (1997), 현업 운영 가능한 서울지역의 일 최고 대기오염도 예보모델 개발 연구 한국대기보전학회지, 13(1), 79-89.  
 김유근 (1998), 오존예보모델 및 예보시스템의 개선, 오존 예보시스템에 관한 전문가토론회, 26-38.  
 김유근, 이화운, 전병일, 홍정혜, 문윤섭 (1996), 부산연안역에서의 대기오염기상 예보시스템 개발에 관한 연구.

정용승, 정재섭 (1991) 서울수도권 지역의 광화학오존에 관한 연구, 한국대기보전학회지, 7(3), 169-179  
 허정숙, 김동술 (1993), 다변량 통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구, 한국대기보전학회지, 9(3), 207-215.  
 Acuna, G, H. Jorquera, and R. Perez (1996) Neural network model for maximum ozone concentration prediction, Lecture Notes in Computer Science No. 1112, 263-268, Springer, New York.  
 Feister, U and K. Balzer (1991). Surface ozone and meteorological predictors on a subregional scale, Atmospheric Environment, 25A(9), 1781-1790  
 Hubbard, M.C and W G Cobourn (1998) Development of a regression model to forecast ground-level ozone concentration in Louisville, KY, U.S.A., Atmospheric Environment, 32(14), 2637-2647.  
 McKendry, I G. (1993) Ground-level ozone in Montreal, Canada, Atmospheric Environment, 27B(1), 93-103.  
 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  
 Robeson, S M. and D.G. Steyn (1990) Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations Atmospheric Environment, 24B(2), 303-312.  
 Ryan, W.F (1995) Forecasting severe ozone episodes in the Baltimore metropolitan area. Atmospheric Environment, 29(17), 2387-2398.  
 Yi, J and V R. Piybutok (1996) A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialised urban area Environmental Pollution 92, 349-357.