

## 인공신경망 모형을 이용한 울산공단지역 일 최고 SO<sub>2</sub> 농도 예측

이소영 · 김유근\* · 오인보\* · 김정규\*\*

국립기상연구소 예보연구팀, \*부산대학교 지구환경시스템학부, \*\*울산시 태화강관리단  
(2008년 3월 21일 접수; 2008년 9월 29일 수정; 2008년 11월 12일 채택)

### Prediction of Daily Maximum SO<sub>2</sub> Concentrations Using Artificial Neural Networks in the Urban-industrial Area of Ulsan

So-young Lee, Yoo-Keun Kim, In-bo Oh\* and Jung-Kyu Kim\*\*

*Forecast Research Laboratory, National Institute of Meteorological Research, KMA, Seoul 156-720, Korea*

*\*Division of Earth Environmental System, Pusan National University, Busan 609-735, Korea*

*\*\*Tae-hwa River Managing Agency, Ulsan metrocity, Ulsan 680-701, Korea*

(Manuscript received 21 March, 2008; revised 29 September, 2008; accepted 12 November, 2008)

#### Abstract

Development of an artificial neural network model was presented to predict the daily maximum SO<sub>2</sub> concentration in the urban-industrial area of Ulsan. The network model was trained during April through September for 2000-2005 using SO<sub>2</sub> potential parameters estimated from meteorological and air quality data which are closely related to daily maximum SO<sub>2</sub> concentrations. Meteorological data were obtained from regional modeling results, upper air soundings and surface field measurements and were then used to create the SO<sub>2</sub> potential parameters such as synoptic conditions, mixing heights, atmospheric stabilities, and surface conditions. In particular, two-stage clustering techniques were used to identify potential index representing major synoptic conditions associated with high SO<sub>2</sub> concentration. Two neural network models were developed and tested in different conditions for prediction: the first model was set up to predict daily maximum SO<sub>2</sub> at 5 PM on the previous day, and the second was 10 AM for a given forecast day using an additional potential factors related with urban emissions in the early morning. The results showed that the developed models can predict the daily maximum SO<sub>2</sub> concentrations with good simulation accuracy of 87% and 96% for the first and second model, respectively, but the limitation of predictive capability was found at a higher or lower concentrations. The increased accuracy for the second model demonstrates that improvements can be made by utilizing more recent air quality data for initialization of the model.

**Key Words :** SO<sub>2</sub>, Urban-industrial area, Potential parameters, Artificial neural network, Cluster analysis

#### 1. 서 론

우리나라 남동해안에 위치한 울산은 급격히 진행

된 도시화와 대규모 공단으로 인해 대기오염현상이 빈번하게 나타난다. 특히 다른 우리나라 대도시들과 비교해 석유화학공단에서 배출되는 SO<sub>2</sub>에 의한 대기질 악화가 특징적이다<sup>1)</sup>. 이에 최근 울산지역 SO<sub>2</sub> 고농도 현상을 사전에 방지하기 위한 연구노력이 있었고<sup>2)</sup> 이와 결부되어 신뢰성 있는 농도예측 연

Corresponding Author : Yoo-Keun Kim, Division of Earth Environmental System, Pusan National University, Busan 609-735, Korea  
Phone: +82-51-510-2282  
E-mail: kimyk@pusan.ac.kr

구의 필요성이 부각되고 있다.

1차 오염물질인 SO<sub>2</sub>는 근본적으로 배출량 삭감을 통해 농도수준이 크게 개선될 수 있다. 그러나 절대적 배출량삭감은 울산지역의 경제, 산업활동이 매우 중요함을 고려한다면 현실적인 한계가 있을 수밖에 없다. 이에 최근 총량규제를 통한 효율적 대기질 개선방안이 제시되고 있다<sup>3)</sup>. 오염물질의 확산이 잘 이루어지지 않는 기간 동안 배출량을 일시 규제함으로써 지역경제의 피해를 최소화하면서 대기질의 개선이 가능하다는 연구 결과가 Berlyand<sup>4)</sup>에 의해 소개되었다. 따라서 대기환경 능력의 사전평가와 함께 오염농도를 사전에 예측하는 일은 SO<sub>2</sub> 농도를 효율적으로 제어하는데 크게 기여할 것이다. 예로 아테네에서는 여러 기상 요소와의 관계를 사용해 회귀식으로 SO<sub>2</sub> 농도를 예측하였고<sup>5)</sup>. 최근 태국에서는 종관장을 군집분석 방법으로 분류하고 SO<sub>2</sub> 고농도 현상과 종관패턴과의 관련성을 연구한 바 있다<sup>6)</sup>. 우리나라에서는 기상요소와 대기오염 농도와의 단순 관계 비교를 통한 판정 작업을 바탕으로 한 연구<sup>7)</sup>나 수도권을 중심으로 한 포텐셜 예측 연구, 전국 대도시를 포괄적으로 평가하는 등의 연구가 몇몇 이루어져 왔다<sup>8~12)</sup>. 최근에는 김 등<sup>13)</sup>의 연구에서 회귀식을 이용하여 울산지역의 오존 포텐셜 예측 모형을 설계하여 종관패턴별로 오존농도를 예측하였다.

다른 대기오염물질과 마찬가지로 SO<sub>2</sub>의 3차원 농도예측에 있어서 수치모델의 사용은 궁극적으로 필요하다. 하지만 여전히 상세격자의 기상모델링 결과, 배출량자료의 불확실성이 크고, 무엇보다 현업의 활용측면에서 볼 때 신뢰성 있는 입력자료를 기반으로 하는 통계기법 등을 사용한 진단모델의 개발노력이 지속적으로 요구된다. 특히 통계모델은 다음날 농도수준의 빠른 예측을 바탕으로 전일 이후에 포괄적인 예보분석을 가능케 하는 장점을 가지고 있다.

본 연구에서는 울산공업단지지역의 일 최고 SO<sub>2</sub> 농도수준을 미리 예측하여 대기질 수치모델을 통한 예보방법을 보완할 수 있는 신경망모델 개발하였다. 입력자료는 지역고유의 환경특성을 반영하는 SO<sub>2</sub> 포텐셜 인자인 종관기상조건, 대기혼합 능력, 잔류 오염도 등을 과학적인 방법으로 수치화하여

생성시켰다. 아울러 개발된 모델은 배출 포텐셜을 고려하여 예측시간을 달리한 모델과 상호 예측도를 비교하였다.

## 2. 자료 및 연구방법

### 2.1. 자료의 구성

일 최고 SO<sub>2</sub> 농도 예측모델을 설계하기 위해서 우선 기상요소와 SO<sub>2</sub> 농도와의 관계를 파악해 농도의 증가와 감소에 직·간접적인 영향을 미치는 신뢰성 있는 포텐셜 인자들을 선정하였다. 이들을 추정하기 위해 사용된 기초자료는 기상청에서 제공하는 RDAPS(Regional Data Assimilation and Prediction System) 분석장 자료(이후 자세히 설명됨), 지상의 기상 및 SO<sub>2</sub> 농도자료, 포항에서 관측되는 고층기상 관측 자료, 울산 공장의 39개 굴뚝에서 측정된 대기질 원격관리 시스템(TeleMetering System, TMS) 자료이다. 연구 기간은 5년으로(2000~2005년), 단 TMS 자료는 1년(2003년), 연중 상대적으로 높은 SO<sub>2</sub> 농도가 나타나는 4~9월 동안<sup>14)</sup>을 대상으로 하였다.

지상의 기상 및 SO<sub>2</sub> 농도자료는 울산시에 위치한 울산 기상대와 연안에 위치한 공단무인기상관측소, 그리고 공단지역 3개의 대기질 측정망(부곡, 여천, 화산 지점)에서 각각 수집되었다. 관측지점의 위치는 김 등<sup>13)</sup>의 연구에 제시한 것과 동일하다. 분석된 기상요소는 울산기상대에서 관측된 기압, 기온, 풍향, 풍속, 습도, 강수량, 일조시간과 운량이며, 공단 무인기상관측소의 풍향, 풍속 자료이다. SO<sub>2</sub> 농도자료의 경우, 유효측정일(일 중 75% 이상의 시간별 자료가 있는 날)<sup>15)</sup>의 3개 측정망에서 측정된 1시간 평균값을 사용하였고, 선정된 날 중 두 지점 이상 울산 지역 1시간 환경 기준치인 120 ppb를 초과하는 날을 고농도일로 정의하였다.

### 2.2. 군집 분석

종관패턴 분류를 위해, 외부의 기준변량이 없는 경우 내부변량을 이용해 동질지역을 효과적으로 구분해 내는 방법인 군집분석을 사용했다<sup>16)</sup>. 군집분석은 개체들 간의 유사성(similarity) 또는 이와 반대 개념인 거리(distance)에 근거하여 개체들을 집단으로 묶는 통계적 방법을 뜻한다.

선정된 연구기간 1,073일(2000~2005년, 4~9월

중 결측자료 제외)을 대상으로, RDAPS 분석장 자료와 울산 기상대에서 관측된 노점편차와 평균 운량 자료를 이용하여 군집분석을 수행하였다. RDAPS 분석장 자료는 122° E, 39° N에 중심을 두고 동서방향으로 191개, 남북방향으로 171개의 격자점과 연직방향 33개 층으로 이루어져 있다. 격자간격은 30 km이며 3차원 최적내삽법으로 전국 예보장과 관측 자료를 혼합하여 생성하고, 12시간 간격으로 제공된다(0000, 1200 UTC). 본 연구에서는 전체 RDAPS 분석장 자료로부터 울산지역 중심의 중관기압패턴을 적절히 설명할 수 있는 4개 지점들의 1000 hPa 지오폠펌셜 고도 변화량을 사용하였다<sup>13)</sup>. 군집분석은 two-stage technique 방법<sup>17)</sup>으로 수행되었고 상세한 분석단계는 김 등<sup>13)</sup>의 연구에 제시되어 있다.

### 2.3. 인공신경망 모델

신경망은 대용량의 데이터로부터 일정한 규칙 또는 패턴을 찾아내는 데이터 마이닝 분류기법의 하나로서 복잡한 구조를 가진 자료에서의 예측문제를 해결하기 위해서 사용되는 비선형모델(nonlinear model)이다. 이 방법은 입력변수와 목표변수 사이의 관계를 설명하기는 어렵다는 단점이 있으나 예측력이 매우 높다는 장점을 가지고 있다. 신경망 모형은 은닉마디(hidden units)라고 불리는 구성요소를 포함하고 있으며 이는 인간의 신경세포를 모형화한 것으로서, 각 은닉마디는 입력변수들의 결합을 수신하여 목표변수에 전달한다. 이때 결합에 사용되는 계수들을 연결강도(synaptic weights)라고 부르며, 활성화함수(activation function)는 각각의 입력값을 변환하고 이를 다시 입력으로 사용하는 다른 마디로 출력하게 된다.

여러 가지 다양한 모델이 있으나 그 중에서도 자료 분석을 위해 널리 사용되는 모델은 MLP(Multi-layer Perceptron) 신경망이다<sup>18)</sup>. MLP는 입력층(input layer), 은닉마디로 구성된 은닉층(hidden layer), 출력층(output layer)으로 구성된 전방향(feed-doreward) 신경망이다. 입력층은 각각의 입력변수에 대응되는 마디들로 이루어져 있으며 은닉층은 여러 개의 은닉마디로 구성되어 있는데 각 은닉마디는 입력층으로부터 전달되는 변수 값들의 선형결합을 비선형함수로 처리하여 출력층 또는 다른 은닉층에 전달하는 역할을 한다. 출력층은 목표변수에 대응하는 마

디들로 이루어지며 구조는 Fig. 1과 같다. p개의 입력마디를 가진 입력층, N개의 은닉마디를 가진 하나의 은닉층, 하나의 출력층으로 이루어진 MLP 신경망 구조를 나타내며 이를 수식으로 표현하면 식 (1)과 같다.

$$\begin{aligned} H_1 &= f_1(w_{10} + w_{11}x_1 + \dots + w_{1p}x_p) \\ H_2 &= f_2(w_{20} + w_{21}x_1 + \dots + w_{2p}x_p) \\ &\vdots \\ H_N &= f_N(w_{N0} + w_{N1}x_1 + \dots + w_{Np}x_p) \\ EY &= g(w_0 + w_1H_1 + w_2H_2 + \dots + w_NH_N) \end{aligned} \quad (1)$$

결합함수(combination function)는 입력층 또는 은닉층의 마디들을 결합하는 함수를 말하며 위의 MLP 구조에서는 선형함수를 사용하였다. 활성화함수는 입력마디 또는 은닉마디의 결합을 변환하는 함수를 의미하며 여기서  $f_1, f_2, \dots, f_N$  은 은닉층에서의 활성화함수,  $g$ 는 출력층에서의 활성화함수를 나타내는데 MLP 신경망에서 은닉층에서의 활성화함수로는 하이퍼볼릭 탄젠트(hyperbolic tangent) 함수, 로지스틱(logistic) 함수 등과 같은 시그모이드(sigmoid) 활성화함수를 이용하며 출력층에서의 활성화함수로는 항등함수(identity function)등을 사용한다.

각 층 사이를 연결하는 네트워크에는 연결 강도들이 존재하며 이 가중치 값들의 변화에 따라 입력값에 따른 출력값이 결정되는데, 이러한 연결 강도는 임의의 초기값에서 시작하며 최적의 가중치를 구하기 위한 반복적인 학습과정이 이루어진다. 본 연구에서는 학습을 통한 매개변수 최적화를 위하여 최종 출력값과 실제 출력값 사이의 차이가 최소화 되도록 연결 가중치를 변화시키는 오류 역전파 알고리즘(Error Backpropagation Algorithms)을 사용하

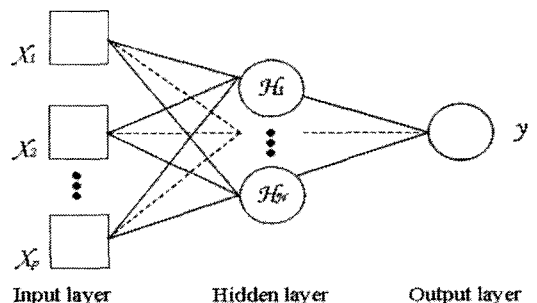


Fig. 1. MLP structure with on hidden layer, N is hidden units.

였다. 오류 역전파 알고리즘은 주어진 입력값과 출력값에 대한 적절한 연결 가중치를 구하기 위하여 비용함수(cost function)를 식 (2)의 형태로 정의하여 비용함수가 0에 가까워지도록 학습을 반복하는데, 이를 위한 수치해석적 방법으로 경사하강법(gradient decent method)을 이용하여 연결 강도의 증감분(식 (3))을 식 (4)와 같이 조정한다. 이 때 학습율(learning rate)  $\eta$ 의 크기에 대한 보편적인 법칙은 없으나<sup>19)</sup>, 일반적으로 0.1에서 0.9 사이의 값을 사용한다.

$$E(W) = \frac{1}{2} \sum (\zeta - o)^2 \quad (2)$$

ShouquanW)는 비용함수;  $\zeta$ 는 required output;  $o$ 는 network output)

$$W_{AB(\neq w)} = W_{AB(old)} - \Delta W_{AB} \quad (3)$$

$$= W_{AB(old)} - \eta \frac{\partial E(W)}{\partial W_{AB}} \quad (4)$$

(여기서,  $\eta$ 는 학습율)

이러한 일련의 과정을 거쳐서 구축되는 신경망 모델의 최적 가중치(weight)값들을 결정하는 학습과정과 새로운 입력 자료에 대한 추정값을 구하여 모델의 성능을 평가할 수 있는 검증과정으로 나누어진다. 본 연구에서는 SAS Enterprise Miner를 이용해 인공신경망 모델을 설계하였다.

### 3. SO<sub>2</sub> 포텐셜 인자의 선정

#### 3.1. 종관기상 조건

2000~2005년, 4~9월 동안 군집분석을 통해 종

관기상 조건을 객관적으로 분류하였고, 고농도 현상과 깊은 관련이 있는 패턴에 일정형태의 가중치를 주어 이를 일 최고 SO<sub>2</sub> 농도를 예측하는 포텐셜 인자로 사용하였다.

Table 1은 계층적 군집분석 결과 제시된 통계량으로 R<sup>2</sup>, Pseudo-F, Pseudo-t<sup>2</sup>의 군집수에 따른 변동을 보여주고 있다. 초기 군집의 개수는 각 연결단계 중에서 Pseudo-F 값이 전후 단계보다 높은 결합의 단계나 Pseudo-t<sup>2</sup> 값이 낮아지는 결합의 단계, R<sup>2</sup> 값이 급격히 증가하는 단계에서 결정하게 된다<sup>20~22)</sup>. 따라서 최적의 초기 군집의 개수는 8개로 선정하는 것이 통계적으로 유의함을 알 수 있다. 이렇게 선정된 초기 군집 수 K를 바탕으로 K-평균 군집분석을 수행하였고, 최종적으로 분류된 군집을 P1~P8이라고 하고 주요 종관패턴이라 정의하였다.

Table 2는 각 패턴별 빈도수와 운량의 평균, 패턴별 SO<sub>2</sub> 농도 평균을 나타낸 것이다. P2, P1, P6은 20%가 넘는 빈도를 나타내어 전체의 75%에 해당하

**Table 1.** Statistical tests for determining the number of clusters

Number of clusters	R <sup>2</sup>	Pseudo-F	Pseudo-t <sup>2</sup>
9	0.13	20.45	16.48
8	0.13	21.89	9.37
7	0.07	14.16	63.40
6	0.04	9.51	35.92
5	0.04	11.12	10.62
4	0.03	11.74	8.99
3	0.02	12.20	10.62
2	0.01	16.28	8.03
1	0.00	-	16.28

**Table 2.** Summary for statistics of SO<sub>2</sub> concentrations and cloud amounts for the classified seven synoptic patterns during April through September, 2000-2005

Pattern	No. of days <sup>a</sup>	Daily max. 1-h SO <sub>2</sub> con. (ppb)	No. of high SO <sub>2</sub> days (> 1-h/120 ppb) <sup>a</sup>	Daily mean cloud amount (10ths of sky cover)
P1	241 (22.5)	81.2	33 (13.7)	8.5
P2	340 (31.7)	89.4	58 (17.1)	5.3
P3	28 (2.6)	112.1	10 (35.7)	2.5
P4	62 (5.8)	103.0	17 (27.4)	5.7
P5	95 (8.9)	94.4	22 (23.2)	1.9
P6	225 (21.0)	93.4	55 (24.4)	3.9
P7	81 (7.5)	79.1	9 (11.1)	7.1
P8	1 (0.1)	-	-	3.1

<sup>a</sup> The values in parenthesis indicate the percentages for the individual pattern.

는 날들이 세 가지 패턴에 포함되었고 나머지 패턴은 모두 10% 미만의 빈도를 나타내었다. 특히 P8은 (군집분석에 포함된 1,073일 중 단 1일만 나타난 패턴) 통계적인 유의성이 결여되어 본 연구에서는 이를 제외한 P1~P7의 패턴에 대한 특성을 분석하였다.

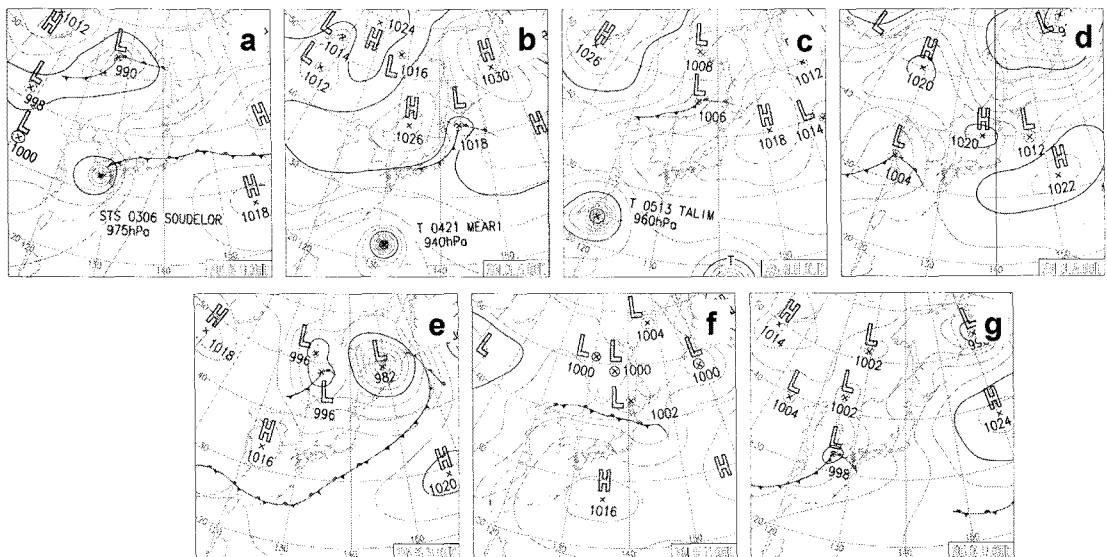
분류된 7개의 종관특성을 이해하기 위해 일본 기상청에서 발행한 0000 UTC의 지상 수치일기도를 중심으로 전반적인 기압배치 및 기상상태를 분석하였다. Fig. 2는 P1~P7의 패턴에 해당되는 대표적인 날의 지상 일기도이다. Table 2의 내용과 함께 각각의 패턴을 분석하면 다음과 같다.

P1(241일)은 저기압이나 장마전선의 영향으로 강수현상 또는 흐린 날씨를 보인 유형으로 비교적 낮은 SO<sub>2</sub> 농도수준을 보였다. P2(340일)는 한반도가 고기압의 후면에 위치하여 맑은 날씨와 함께, 울산 지역에 남동풍이 탁월하게 나타난다. 이는, 해풍의 방향과 동일하여 연안에 위치한 공단의 환기조건을 좋게 할 수 있지만, 내륙까지 영향반경이 확대되어 넓은 범위의 풍하측 농도증가가 가능해진다. 농도수준은 비교적 낮게 나타났고 전체의 17%가 고농도를 기록하였다.

P3(28일)은 한반도를 중심으로 기압경도력이 약

하게 작용하여, 뚜렷한 해륙풍 순환과 함께 국지적 대기정체, 혼합고 감소, 이로 인한 단거리 수송 등이 나타나 울산 전 지역에 심각한 고농도 현상이 가능한 패턴이다. 해당일의 수는 상대적으로 적지만 매우 높은 농도수준이 나타나(평균 일 최고농도: 112 ppb, 기준치초과 일수/비율: 10/약 36%) 울산지역의 심각한 SO<sub>2</sub> 고농도 발생과 가장 관련 깊은 패턴이라 할 수 있다. P4(62일)는 대부분 뚜렷한 기압패턴이 없고 일 중 맑은 후 흐려지는 날씨가 나타났다(약 6의 평균 운량). 고기압에 의해 안정되어 있던 대기에 약하게 구름이 끼면서 정체현상이 가중되어 P3에 이어 두 번째로 높은 농도수준을 기록한 패턴이었다. P5(95일)는 약한 기압골의 영향을 받는 형태로 울산지역은 북서 종관풍이 탁월하며 해풍의 발달이 억제되고 이로 인해 연안 오염물질의 국지적 고농도현상이 빈번하게 나타난다.

P6(225일)은 북태평양 고기압이 우리나라 남쪽에, 북쪽에는 저기압이 자리하고 있는 형태로 여름철에 주로 나타나며 P1, P2와 함께 많은 고농도일이 포함되었다. 비교적 맑은 날씨가 나타나고 고기압의 영향으로 안정적인 대기상태를 유지하게 되어 P5와 비슷한 농도수준이 기록되었다. P7(81일)은 우



**Fig. 2.** Surface weather maps at 0000 UTC representing each pattern: (a) P1 (19 June 2003), (b) P2 (25 September 2000), (c) P3 (1 September 2005), (d) P4 (26 April 2004), (e) P5 (20 May 2005), (f) P6 (20 July 2004), and (g) P7 (12 June 2003).

리나라 남북으로 저기압이 자리잡고 있어 이 패턴에 해당하는 대부분의 날들이 강수현상이 나타났고, 이로 인해 7개의 패턴 중 가장 낮은 평균 일 최고농도(79 ppb)와 고농도 발생 비율(11%)을 보였다.

이러한 패턴분류결과가 SO<sub>2</sub>의 일 최고 농도 예측에 활용되기 위해 고농도 발생의 가능성을 잘 표현할 수 있는 각 패턴의 고농도 발생일수 비율(%)을 중관기상장의 특성을 반영하는 포텐셜 인자로 선정하였다. 그러나 이 비율은 같은 중관패턴에 포함되는 모든 날들이 동일한 수치를 가지므로 각 날짜의 특징 있는 중관기상 조건을 반영하기 어렵다. 따라서 인근 포항에서 관측된 고층기상 관측자료를 활용하여 각 해당일의 중관풍의 영향을 고려하였다. 즉 0000 UTC의 850 hPa 바람의 u와 v성분을 중관조건 특성을 나타내는 또 다른 포텐셜 인자로 사용하였다.

### 3.2. 대기 혼합 능력

대기 혼합 능력은 배출된 오염 물질을 넓은 곳으로 빠르게 확산시켜 한 곳에 정체되는 것을 방지하는 능력을 나타내는 것으로 대기오염 농도를 예측하는데 필수적인 요소라 할 수 있다<sup>23,24</sup>. 대기혼합 능력을 결정하는 것은 주간과 야간에 따라 혼합고와 기온역전 강도로 나누어 생각할 수 있는데, 각각은 연직적인 관측값을 필요로 한다. 그러나 현실적으로 연속적인 연직 관측값을 확보하기란 어려운 실정이므로 본 연구에서는 연구대상기간 동안 중규모 기상모델인 MM5(NCAR/PennState Fifth Generation Mesoscale Model)<sup>25</sup>를 수행하여 연직적인 값을 산출해 사용하였다. 본 연구에 사용된 모델의 세 개 도메인은 Fig. 3에 나타내었으며, 각각 18 km(62×62), 6 km(82×82), 2 km(40×40)의 해상도를 가진다. 연직층은 sigma 좌표계의 23개 층이며 지표근처의 간격을 조밀하게 설정하였다.

MM5의 초기 기상장 입력자료는 NCEP(National Centers for Environmental Prediction)에서 제공하는 6시간 간격의 연직 17층, 수평 2.5° 전지구 분석자료(Climatic Data Assimilation System, CDAS)를 사용하였다. 대기경계층내의 물리과정은 대기경계층 기상을 보다 상세히 모사할 수 있는 MRF-PBL 방법을 선택하였다. 모델 수행기간은 2000~2005년 4~9월, 총 36개월이며, 혼합고와 기온 역전 강도를 산출한 지

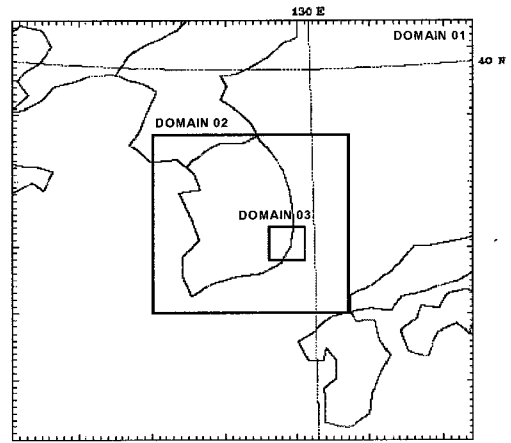


Fig. 3. Three horizontal domains for MM5 simulations.

점은 울산 기상대 지점이다.

주간의 연직 혼합능력을 알아보기 위해 MM5 결과에서 혼합고(PBLH)를 산출하여 일 최고 혼합고를 구하고, SO<sub>2</sub> 농도와의 관계를 분석하였다. Fig. 4는 일반적인 자유대기 고도(약 1500 m)의 약 70% 수준인 1000 m를 기준으로 SO<sub>2</sub>의 평균 시간별 농도를 나타낸 것이다. 1000 m이하의 일 최고 혼합고를 나타내는 경우의 주간 농도가 반대의 경우보다 최고 3 ppb가량 높은 농도가 주간에 지속됨을 볼 수 있다. 이는 일최고 혼합고가 낮은 경우 배출된 SO<sub>2</sub>가 확산되지 못하고 정체되어 주간의 농도를 높이는 원인이 되는 것을 의미한다.

야간의 경우 역전강도로 연직 혼합능력을 판단할

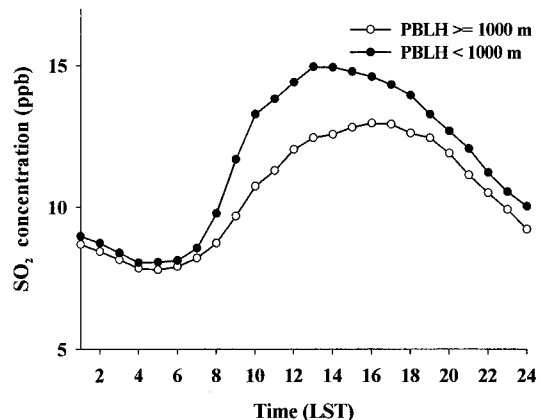


Fig. 4. Diurnal variations of SO<sub>2</sub> concentrations for max. PBLH ≥ 1,000 m and max. PBLH < 1,000 m.

수 있다. 이는 김<sup>14)</sup>의 연구에서 사용했던 지표부근의 두 연직 층 간 기온차를 이용하는 방법으로 추정할 수 있다. 김<sup>14)</sup>의 연구에서는 울산 BNG STEEL 내에 설치된 10 m 기상관측 타워의 2 m와 10 m 두 층의 기온구배를 이용하였으나, 본 연구에는 MM5에서 계산된 지표와 시그마 레벨 0.99(약 70 m)의 기온차를 0시부터 6시까지 합산(이하 'TSUM'이라 칭함)하여 야간 역전강도를 평가하는 지표로 사용하였다. 즉, TSUM의 값이 클수록 역전 강도가 강하며 오전의 대기정체를 유발할 가능성이 높은 것으로 판단할 수 있다.

Fig. 5는 TSUM이 2 이상인 비교적 역전강도가 강한 집단(Group I)과 TSUM 2 이하로 상대적으로 역전강도가 약한 집단(Group N)에 해당되는 SO<sub>2</sub> 농도의 일변화를 비교한 것이다. Group I의 경우 주간의 SO<sub>2</sub> 농도가 Group N보다 약 8 ppb 가량 높게 나타났다. 이는 역전 강도가 강할수록 SO<sub>2</sub>의 고농도 가능성이 높아짐을 보여주는 것으로서 TSUM이 SO<sub>2</sub>의 포텐셜 인자로 고려될 수 있다.

일 최고 혼합고와 야간 역전 강도가 대기의 연직 혼합능력을 보여주는 인자라면, 수평 바람은 대기의 수평 혼합능력을 평가할 수 있다. Fig. 6은 울산 기상대에서 관측된 SO<sub>2</sub> 고농도일(> 120 ppb)과 비교농도일(< 60 ppb)의 시간별 풍향빈도와 평균풍속을 나타낸 것이다. 풍향의 경우(Fig. 6a와 b) 고농도일은 해풍유입을 의미하는 남동풍이 주간에 탁월하게 나타나 SO<sub>2</sub>의 고농도 현상과 해풍이 상당한 관련이

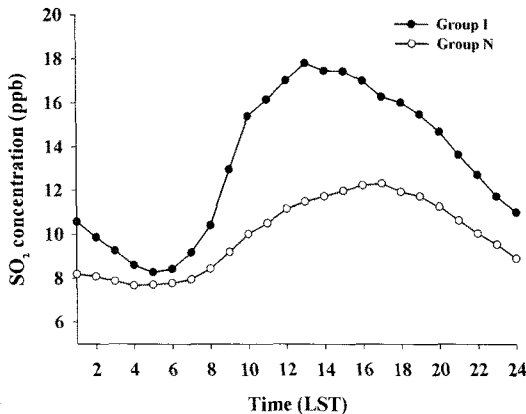


Fig. 5. Diurnal variations of SO<sub>2</sub> concentrations for Group I (TSUM ≥ 2) and Group N (TSUM < 2).

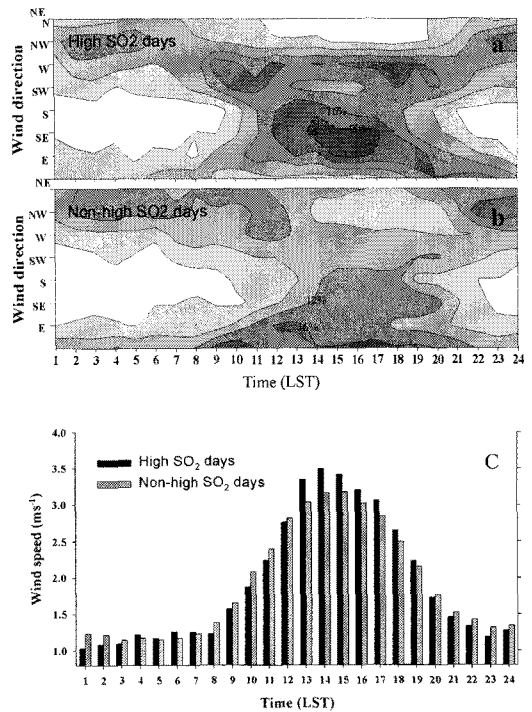


Fig. 6. Diurnal variations of frequency of wind direction for (a) the high SO<sub>2</sub> days and (b) the non-high SO<sub>2</sub> days, and (c) the averaged wind speed for the high and non-high SO<sub>2</sub> days.

있음을 보여준다. 반면 비교농도일의 경우 상대적으로 뚜렷한 풍향특징을 볼 수 가 없다. 고농도일과 비교농도일의 시간별 평균풍속을 비교한 Fig. 6c에서는, 고농도일의 경우 오전에 상대적으로 약한 바람이, 오후에는 강한 바람이 부는 것을 알 수 있다. 즉, 오전의 대기 정체 현상과 오후의 해풍의 영향을 반영하는 것이다. 이는 해풍유입과 관련해서 일반적으로 나타나는 오전의 대기정체와 대기안정도의 증가 등의 현상이 SO<sub>2</sub>의 고농도 현상과 깊은 관련이 있음을 말해준다.

이상과 같이 SO<sub>2</sub>의 농도는 주·야간 연직·수평 대기혼합능력과 상관이 있음을 알 수 있었다. 따라서 본 연구에서는 연직 대기 혼합 능력을 대변하는 인자로 일최고 혼합고와 TSUM을 사용하였고, 수평 대기혼합능력을 대변하는 인자로 울산 기상대 지점 1300 LST u, v와 0500 LST u, v를 선정하여 사용하였다. 또한 해풍이 발달한 시간인 1000~1600 LST의 평균풍속 자료와 오전의 대기정체를 나타낼 수

있는 시간인 0100~0500 LST의 평균 풍속값을 사용하였다. 아울러 대기혼합능력과 SO<sub>2</sub>의 농도변화에 간접적으로 기여할 수 있는 일 최고기온, 일 평균운량, 일조율자료도 고려하였다.

### 3.3. 잔류 오염도

일반적으로 하루 전 대기경계층 내의 SO<sub>2</sub> 농도, 즉 잔류 오염도는 다음날 지표 농도변화에 기여할 수 있다. 정확한 잔류 오염도 평가를 위해서는 대기경계층내의 연직적 농도분포를 고려해야 하나 본 연구에서는 연직 관측자료의 활용이 불가능하여 지표농도를 잔류오염도로 간주 하였다.

잔류오염도의 영향을 조사하기 위해 전일의 일 최고 SO<sub>2</sub> 농도를 3개의 구간(120 ppb~, 60~120 ppb, ~60 ppb)으로 나누어 각각의 구간에 해당되는 당일의 최고 농도와의 관계를 분석하였다. 그 결과 각 구간에 해당되는 농도가 평균 104.2 ppb, 91.3 ppb, 72.5 ppb로 나타나 전일의 농도가 당일 농도 수준에 영향을 미침을 알 수 있다. 각 구간별 당일의 기준치 초과 비율 역시 35.1%, 18.2%, 8.5%로 나타나 연속된 두 날의 농도수준이 서로 깊은 상관성이 있음을 간접적으로 보여준다. 결과적으로 본 연구에서는 전일의 SO<sub>2</sub> 최고 농도를 대기 중 잔류오염도로 가정하여 일 최고 SO<sub>2</sub> 농도를 예측하는 포텐셜 인자로 고려하였다.

## 4. 일 최고 SO<sub>2</sub> 농도 예측 모델의 설계 및 평가

### 4.1. 모델의 설계

앞서 SO<sub>2</sub> 농도와 관계 깊은 여러 기상인자와 대기질 자료로부터 포텐셜 인자들을 선정할 수 있었다. Table 3은 신경망모델에 수치화되어 입력자료로 사용되는 포텐셜 인자들을 나타낸 것으로, 앞서 설명한 큰 항목을 기준으로 세부 인자를 구하여 일 최고 SO<sub>2</sub> 농도 예측모델을 설계하였다. 모델에서 고려하는 기상인자들은 기상모델의 예측장에서 산출될 수 있는 인자들로 기상모델링의 결과를 바탕으로 현장에서 적용 가능하다.

예측은 대기오염 가능성을 시민들에게 알리고 고농도에 대한 피해저감대책을 실시하기 위해 최대한 일찍 이루어 져야 한다. 본 연구에서는 예측시간의 차이를 두어 두 개의 모델을 설계하였다. 첫 번째는 (모델 A) 기상청의 정기 일기예보 시각(16:00), 오염

**Table 3.** Potential factors associated with the episodes of high SO<sub>2</sub>

The probability of SO <sub>2</sub> exceedances at each synoptic pattern (%)
u and v component for upper winds on 850 hPa (m s <sup>-1</sup> )
Daily maximum PBL(planetary boundary layer) height (m)
Vertical temperature gradient (°C)
Mean surface wind speed (1000 ~ 1600 LST and 0100 ~ 0600 LST) (m s <sup>-1</sup> )
u and v component of surface winds at 0500 and 1300 LST (m s <sup>-1</sup> )
Daily maximum temperature (°C)
Rate of sunshine (fraction)
Cloud amount (10ths of skycover)
The maximum SO <sub>2</sub> concentration on previous day (ppb)

예보 담당자의 근무시간(9:00~18:00), 언론매체를 이용한 보도와 시민의 대비 편의성 등을 고려하여<sup>8)</sup> 매일 17:00에 예측하는 것으로 하였다. 두 번째는 (모델 B), 대상물질인 SO<sub>2</sub>는 배출량의 영향을 직접적으로 받는 1차 오염물질이므로 야간 배출량의 영향을 고려하여 당일 오전 10시에 예측하는 조건으로 모델을 설계하였다.

모델 B는 오전 시간(0500~0700 LST)에 증가한 배출량 영향을 감안하기 위한 것이므로 오전 시간의 3개의 대기질 측정망 SO<sub>2</sub> 농도(1시간 간격)와 대단위 공장의 굴뚝 39개에 설치된 대기질 원격 관리 시스템(TeleMetering System, TMS) 농도자료(30분 간격)를 사용하여 추가적으로 포텐셜 인자를 고려하였다. TMS 자료는 제한된 공개로 인해 2003년 자료만을 분석할 수 있었고 결과적으로 모델 B의 평가기간은 2003년으로 하였다. 포텐셜 인자 선정은 0500~0700 LST 각 시간대별 기준치(1시간 120 ppb)를 초과하는 측정망의 비율(%)를 계산하였고 이 초과비율의 평균을 포텐셜 인자로 사용하였다. 아울러 모델 B는 전일 일 최고 농도 대신 당일 오전 시간대의 최고농도를 사용하였다.

### 4.2. 예측모델의 평가

설계된 두 모델의 전반적인 평가를 위해 전체 날짜에 대한 예측정확도(SO<sub>2</sub> Simulation Accuracy)를 분석하였다. 예측정확도는 고농도 기준인 120 ppb를 중심으로 나누어진 분할표(Table 4)를 바탕으로 계산되어 진다. 예측값과 관측값이 동일한 범주(A 또는 D)에 해당되는 날짜수의 합에 총 날수로 나누



**Table 4.** Contingency table used for verification of the SO<sub>2</sub> prediction accuracy

	Predicted SO <sub>2</sub> con. < 120 ppb	Predicted SO <sub>2</sub> con. ≥ 120 ppb
Observed SO <sub>2</sub> con. < 120 ppb	A	B
Observed SO <sub>2</sub> con. ≥ 120 ppb	C	D

어 퍼센트 단위로 표현되며(식 5 참조) 값이 높을수록 정확도가 높다. 또한, 고농도일(1시간 농도 120 ppb 이상인 날)과 비고농도일의 편의(Bias)와 평균 제곱근오차(root mean square error, RMSE)값을 계산하였다. Bias는 Table 4의 분할표와 식 6과 7을 통해 계산되어지며, 모델 결과값의 치우침 정도를 나타내는 것으로 1에 가까울수록 정확도가 높다. RMSE는 식 8을 통해 얻어지며 관측값과 결과값 사이의 평균적인 오차 정도를 의미한다.

$$SO_2 \text{ Simulation Accuracy} = (A+D)/N \times 100 \quad (5)$$

$$\text{Bias(H)} = (B+D)/(C+D) \quad (6)$$

$$\text{Bias(NH)} = (A+C)/(A+B) \quad (7)$$

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

여기서, N(자료개수)=A+B+C+D, H는 고농도일의 경우, NH는 비고농도일의 경우를 의미한다. P<sub>i</sub>는 모델의 결과값, O<sub>i</sub>는 관측값, n은 예측 날짜 수이다.

두 예측모델의 평가결과, 모델 A의 경우, 87.3%의 예측정확도를 나타내어 대략적으로 농도예측에 대한 분별력이 있음을 보여준다. 하지만 고농도일의 RMSE가 46.2로 크게 나타나 모델 예측값이 관측값과 정량적으로 상당한 차이가 있었다(Table 5). 이는 통계모델의 한계로, 반복성이 없이 단시간에 급격히 증가하는 관측치의 경향을 모델이 제대로 학습하지 못했기 때문으로 해석된다. 아울러 SO<sub>2</sub>는 1차 오염물질로서 국지배출원 변화에 매우 민감할 수 있기 때문에 이를 정확히 반영할 수 없는 모델 A의 한계가 오차를 크게 하는 주요 원인으로 볼 수 있다. 비고농도일의 경우 예측 농도수준이 대체로

관측 값과 비슷하게 모의되어 RMSE 값이 27.6으로 줄어들었다. Bias 역시 고농도일과 비고농도일 각각 0.3과 1.2로 나타나 고농도일은 과소 모의를, 비고농도일은 약간의 과대 모의를 하는 결과가 나타나 고농도 값의 예측은 정확성이 떨어지는 것을 알 수 있다. 모델 B의 경우 오전시간에 증가한 배출량과 갱신된 일 최고 SO<sub>2</sub> 농도의 영향을 고려한 것으로 모델 평가기간의 차이는 있으나 모델 A 보다 약 10% 가량 높아진 96.1%의 예측정확도를 나타냈다. 또한 RMSE와 Bias 값의 경우 모델 A와 비교해 상당히 향상된 결과를 Table 5에서 볼 수 있다. 특히 고농도일의 RMSE가 6.7로 나타나 오전의 국지배출과 관련된 포텐셜인자(TMS 측정망 기준치 초과 비율과 오전의 최고농도)를 고려한 것이 모델 정확도 향상에 매우 중요함을 알 수 있었다. Bias 역시 고농도일에 0.8로 나타나 모델 A와 비교해 관측값과의 차이가 크게 줄어들었다.

## 5. 결론

본 연구에서는 울산지역의 공단지역을 대상으로 SO<sub>2</sub> 농도분포와 관계하는 중요한 포텐셜 인자들을 선정하고 일 최고 SO<sub>2</sub> 농도를 예측하는 신경망모델을 설계하였다. 선정된 포텐셜 인자는 종관기상장, 대기혼합능력 및 국지기상요소, 잔류오염도를 반영한 것으로, 적절히 수치화하여 모델 입력자료로 사용하였다.

예측모델의 설계는 전일 오후 5시에 예측하는 방법(모델 A)과 당일 오전의 TMS 기준치 초과 비율과 최고농도를 반영하여 당일 오전 10시에 예측하는

**Table 5.** Statistics for the evaluation of different two models

	Model A			Model B		
	N <sup>a</sup>	RMSE	Bias	N <sup>a</sup>	RMSE	Bias
High SO <sub>2</sub> day	126	46.2	0.3	31	6.7	0.8
Non-high SO <sub>2</sub> day	540	27.6	1.2	91	13.2	1.1

<sup>a</sup> N indicates the number of days for each group.

방법(모델 B)으로 나누어 이루어졌다. 모델 A의 경우 120 ppb를 평가기준으로 약 87%의 예측 정확도를 보였으나 실측치와 비교적 큰 편의와 오차범위를 나타내었다. 하지만 모델 B의 경우, 약 96%의 예측 정확도와 신뢰할 만한 수준의(Bias = 0.8, RMSE = 6.7) 검증결과를 나타내어 개발된 신경망모델이 일 최고 SO<sub>2</sub> 농도를 실측값과 상당히 근접하게 예측할 수 있음을 보여 주었다. 이는 예측시간과 가까운 대기질 정보가 신경망 모델에 학습됨으로써 일 최고 SO<sub>2</sub> 농도의 예측 정확성이 향상됨을 의미하는 결과이다.

본 연구에서는 군집분석 등과 같은 다양한 방법으로 울산지역 SO<sub>2</sub> 농도변화과 깊이 관련하는 지역의 기상 및 대기질 포텐셜 인자를 정량화하는 방법을 제시하였고, 정량화된 포텐셜 인자를 기반으로 한 신경망 모델의 일 최고 SO<sub>2</sub> 농도예측이 실제 현장에서 활용 가능함을 보여주었다. 대기질 예측에 있어 수치모델의 적용이 궁극적이며 지향해야 할 방법이지만 신뢰성있는 통계모델을 통한 사전예측은 긴 시간이 요구되는 수치모델링 예보에 앞서 개괄적 예보분석 정보를 제공해 줄 수 있는 역할을 할 수 있을 것이다.

## 감사의 글

본 연구는 환경부 지정 울산지역 환경기술개발센터의 2006년 연구비로 수행되었으며 이에 감사드립니다.

## 참고 문헌

- 이화운, 원경미, 정우식, 오은주, 김민선, 도우곤, 2002, 해륙풍을 고려한 대기오염물질농도의 수치모의, 한국환경과학회지, 11(7), 933-943.
- 구윤서, 2005, 지역대기환경용량평가 및 배출허용기준의 효율적인 적용방안, 울산지역환경기술개발센터, 225pp.
- 이영준, 2004, 고정오염원에서 발생하는 SO<sub>2</sub> 배출량 저감을 위한 효율적인 환경정책수단의 연구, 한국환경과학회지, 13(4), 339-347.
- Berlyand M. E., 1991, Prediction and regulation of air pollution, Kluwer Academic Pub., 312pp.
- Ioannis C. Z., Dimitrios M., Christos S. Z., Alkiviadis F. B., 1995, Forecasting peak pollutant levels from meteorological variables., Atmos. Environ., 29(24), 3703-3711.
- Oanh N. T., Chutimon P., Ekbordin W., Supat W., 2005, Meteorological pattern classification and application for forecasting air pollution episode potential in a mountain-valley area, Atmos. Environ., 39(7), 1211-1225.
- 김유근, 이화운, 전병일, 홍정혜, 1996, 부산연안역의 오존 농도에 미치는 해풍의 영향, 한국환경과학회지, 5(3), 265-275.
- 김용준, 1997, 현업운영 가능한 서울지역의 일 최고 대기오염도 예보모델 개발 연구, 한국대기보전학회지, 13(1), 79-89.
- 이보람, 박순웅, 1997, 종관 기상 상태를 고려한 한반도 대기 오염 퍼텐셜 예측법, 한국기상학회지, 33(4), 641-656.
- 김용국, 이종범, 1994, 하계의 일 최고 오존농도 예측을 위한 신경망모델의 개발, 한국대기보전학회지, 10(4), 224-232.
- 허정숙, 김동술, 1993, 다변량 통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구, 한국대기보전학회지, 9(3), 207-215.
- 이종범, 1991, 중부지방 각지의 대기오염 잠재력에 관한 연구, 한국대기보전학회지, 7(1), 41-47.
- 김유근, 이소영, 임윤규, 송상근, 2007, 중회귀 모형을 이용한 울산지역 오존 포텐셜 모형의 설계 및 평가, 한국대기환경학회지, 23(1), 14-28.
- 김유근, 2003, 울산의 대기오염 기상조건과 기상여건에 따른 대기오염실태조사, 울산지역환경기술개발센터, 244pp.
- 환경부, 2004, 대기환경월보 2004년 6월, 60pp.
- Brain S. E., 1993, Cluster analysis, 3rd ed., Halsted Press, 170pp.
- Davis R. E., Kalkstein L. S., 1990, Development of an automated spatial synoptic climatological classification, Int. J. Climatol., 10(8), 769-794.
- Potts W. J. E., 2000, Neural network modeling course notes, SAS Institutes Inc., 240-259.
- 손건태, 2001, 전산통계개론(개정판), 자유아카데미, 330pp.
- Cheng S., Lam K. C., 2000, Synoptic typing and its application to the assessment of climatic impact on concentrations of sulfur dioxide and nitrogen oxides in Hong Kong, Atmos. Environ., 34(4), 585-594.
- 박정균, 이동규, 1998, 군집분석에 의한 아시아 동안에서 급격히 발달하는 저기압의 분류와 그 발달기구, 한국기상학회지, 34(4), 523-537.
- Eder B. K., Davis J. M., Bloomfield P., 1994, An automated classification scheme designed to better elucidate the dependence of ozone on meteorology, J. Appl. Meteo., 33(10), 1182-1199.
- 전병일, 김유근, 이화운, 황수진, 1996, 해풍효과에 의한 저층대기구조 변화의 측정, 한국환경과학회지, 5(4), 441-451.
- 전병일, 김유근, 1998, 부산연안에서 관측된 저층대

기의 특성에 관한 연구, 한국환경과학회지, 7(2), 195-201.  
25) Dudhia J., 1993, A nonhydrostatic version of the penn

state/NCAR mesoscale model : variation tests an simulation of an Atlantic cyclone and cold front, Mon. Wea. Rev., 121(5), 1493-1513.