

## 중회귀식을 이용한 원주시 SO<sub>2</sub> 오염도 예보기법 개발에 관한 연구

송 동 웅  
상지대학교 환경공학과  
(1998년 6월 10일 접수)

### On the Development of the Statistical SO<sub>2</sub> Forecasting Technique by the Multiple Regression Analysis in Wonju City

Dong-Woong Song  
Dept. of Environmental Engineering, Sangji University, Wonju, Korea.  
(Manuscript received 10 June 1998)

Statistical SO<sub>2</sub> forecasting technique by multiple regression analysis was designed and developed to predict SO<sub>2</sub> concentration in Wonju City.

SO<sub>2</sub> concentration data measured from air pollution monitoring system and meteorological factors data such as : wind speed, atmospheric stability, surface temperature, relative humidity and precipitation were used in Wonju City during the 1996~1997.

As the results, correlation model for forecasting was well fitted with some parameters including minimum temperature, wind speed and the SO<sub>2</sub> concentration of the previous day.

Key words : forecasting technique, multiple regression analysis, SO<sub>2</sub> concentration, meteorological factors.

#### 1. 서 론

오늘날 효율적인 대기오염의 관리를 위하여 오염도의 현황파악 및 관리기준을 설정하고 있으며 이를 위하여 대기오염농도를 예측하는 방법이 개발되고 있다.

대기오염농도 예측모델에는 크게 수학적, 물리학적, 통계학적 모델로 나눌 수 있으며 이들 대부분은 기상학적 요소들과의 상관성을 토대로 분석되고 있다. 국내에서는 기상요소와 오염도의 통계분석에 관한 연구로 신찬기 등(1992)이 4가지 오염물질(SO<sub>2</sub>, CO, TSP, O<sub>3</sub>)과 5가지 기상인자(기온, 상대습도, 풍향, 풍속, 일사량)에 대한 상관과 중회귀분석을 실시하였으며 홍낙기 등(1993)은 6가지 오염물질(O<sub>3</sub>, CO, NO<sub>2</sub>, SO<sub>2</sub>, THC, TSP)과 8가지 기상인자(풍속, 풍향, 강수량, 일사량, 시정, 기온, 상대습도, 운량)를 대상으로 중회귀분석 등의 통계분석을 실시한 바 있다. 한편 기상요소와 오염도의 관계에서 장래 오염도를 예측하는 연구로는 허정숙과 김동술(1993)이 5가지 오염물질(SO<sub>2</sub>, TSP, CO, NO<sub>2</sub>, THC)과 풍속을 변수로 선정, 단계별 회귀분석을 실시하여 고농도 오존 예측식을 제시한 바 있으며 김용국과 이종범(1994)은 신경회로망을 이용한 오존농도 예측을 실시하였고, 김용준(1997)은 오염도와 기상자료를 이용한 오염도 예보모델개발을 위한 연구를 수행한 바 있다. 국외의 경우 Turner(1961), Simth와Jeffrey(1972), Joukoff와Malet(1982), Allender와 Dejardin(1982) 등이 SO<sub>2</sub> 예측식을 제시하였으며, 이들 대부분은 안정도,

풍향, 풍속, 온도, 혼합고 등의 기상인자와 전일 SO<sub>2</sub>농도 등 지역별 보정계수를 적용하여 오염도를 예측하고 있으나 각각 제시된 식이 지역특성에 따라 차이가 있어 이를 일반적으로 적용하기는 불가능하다. 이외에도 현재 주로 이용되는 예측모델은 배출량과 기상인자를 고려한 확산식에 의하여 계산하는 방식이다. 그러나 배출원이 다양화해짐에 따라 모든 plume의 확산농도를 계산하기란 불가능하고 복잡한 광화학반응과 여러 가정이 현실과 맞지 않아 많은 문제점을 가지고 있다. 따라서 본 연구에서는 여러 가지 기상인자들에 대한 적용 가능성을 검토하고 기상요소와 대기오염농도와의 관계에서 주요인자를 선별하고 분석하여 이들 인자와 대기오염도와의 일정한 경향을 중회귀모형을 이용하여 통계적으로 수식화하여 오염농도를 예측하는 예보식을 제시하였다.

#### 2. 연구 방법

본 연구에서 사용된 SO<sub>2</sub> 오염도 자료는 96-97년 2년간의 원주시 학성동에 위치한 대기오염 자동측정망(127° 57', 37° 21') 자료를 이용하였고, 기상자료는 96-97년 원주시 기상대의 시간별 자료(풍속, 기온, 습도, 강수량)를 이용하였다. 자료의 선정은 총 관측시간의 2/3이상 관측된 값을 유효관측치로 선정하였다.

본 연구에서 고려된 SO<sub>2</sub> 오염도와의 관련인자는 Table 1과 같다.

분석방법은 기본적으로 자연과학 분야에서 폭넓게 응

Table 1. Factors related to SO<sub>2</sub> concentrations

1) Air temperature - 24-h mean temperature, maximum temperature, minimum temperature
2) Stability - 24-h mean stability
3) Wind speed - 24-h mean wind speed, number of hours of mean wind less than 1.5m/sec during the 24-h
4) Humidity - 24-h mean humidity
5) Precipitation
6) Previous day's 24-h average SO <sub>2</sub> concentration

용되는 다변량통계분석법(multivariate statistical analysis)으로 독립변수의 인자가 대상(object)에 대한 기여도를 중심으로 전체의 대상에 대한 일정한 경향을 파악하여 간단한 식으로 표현하였으며, SPSS (Statistical Package for the Social Sciences)통계팩키지(SPSS version 7.5)를 사용하였다.

기본원리는 대상자료의 기여도, 즉 특성통계치가 어느 정도 영향을 미치고 있는지를 구별하여 각각의 인자별 요인분석(factor analysis)과 분산(variance), 상관관계(correlation)를 조사하여 다변량통계분석의 기본자료로 확보하여 stepwise selection 방식에 의하여 중회귀 분석을 실시하였다.

본 연구에서는 예보기법개발을 위해 종속변수의 변화를 2개 이상의 독립변수로 설명하는 중회귀분석을 실시하였으며 일반적인 중회귀식은

$$X_i = b_0 + b_1X_{i1} + b_1X_{i1} + b_2X_{i2} + \dots + b_pX_{ip} + \epsilon_i \quad (i= 1, \dots, n) \quad (1)$$

로 표시되며, 여기에서

- $b_0, \dots, b_p$ : 회귀계수
- $X_{i1}, \dots, X_{ip}$ : 독립변수
- $X_i$ : 종속변수
- $\epsilon_1, \dots, \epsilon_n$ : 오차항

이다.

회귀모델의 적합성 판단에는 독립변수로서 설명할 수 있는 종속변수의 분산비율(variance rate)인 다중결정계수(coefficient of multiple determination) R<sup>2</sup>을 이용하였다. 이산치의 판정기준은 표준잔차(standardized residual, E) 분석법을 이용하였으며 표준잔차(E)란 다중회귀식내에서 개별잔차(individual residual, e)를 잔차의 표준편차( $\hat{\sigma}_e$ ) 제곱으로 나눈 값으로 다음 식과 같이 표현할 수 있다.

$$E = e / \hat{\sigma}_e^2$$

$$\hat{\sigma}_e^2 = \left[ \sum_{i=1}^n (e - E(e))^2 \right]^{1/2}$$

### 3. 결과 및 고찰

#### 3.1 SO<sub>2</sub>농도 특성

SO<sub>2</sub>농도의 연중 변화를 살펴보면 Fig. 1과 같다. 전형

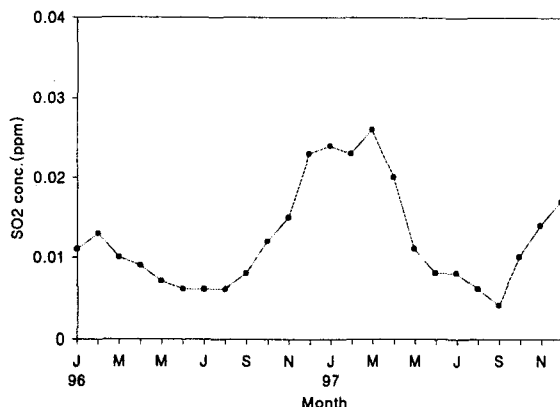


Fig. 1. Monthly average SO<sub>2</sub> concentration in Wonju City from January 1996 to December 1997.

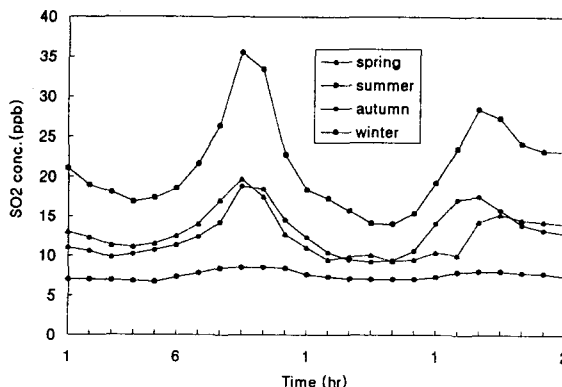


Fig. 2. Diurnal variations of the seasonal average SO<sub>2</sub> concentration in Wonju City from January 1, 1996 to December 31, 1997.

적인 분포 현황으로 동절기에 높은 농도가 그리고 하절기에는 매우 낮은 농도가 관측되었다. 월별로는 하절기인 6, 7, 8월에 0.01ppm 이하의 낮은 농도를 보였으나 동절기인 12, 1, 2월에는 대기환경기준 0.03ppm에 근접한 수준을 나타냈다.

계절별 SO<sub>2</sub> 농도의 일변화를 Fig. 2에 도시하였다. 4계절 모두 오전 8시-10시대에 최고 농도가 나타나고 일몰후 다시 2차 peak가 나타나고 있다. 특히 고농도가 나타나는 동절기에는 이러한 double peak가 뚜렷하게 나타나며 하절기에서만 2차 peak가 약화된 양상을 보이고 있다. 이는 출퇴근시 자동차에 의한 배출가스 증가와 난방연료의 사용에 기인한 것으로 판단된다.

#### 3.2 SO<sub>2</sub>농도와 관련 인자와의 관계

대기 오염 농도에 영향을 미칠 수 있는 인자 중 본 연구에서 고려한 관련인자는 안정도, 풍속, 기온, 습도, 강수량, 전날의 농도이며 각 인자들의 상관 계수와 이들이 SO<sub>2</sub>농도에 미치는 상관 계수를 Table 2에 정리하였다.

SO<sub>2</sub>농도와 높은 상관을 보이는 기상인자는 평균온도가 -0.8070으로 가장 높았고, 그 다음이 안정도, 강수량

Table 2. Correlation coefficient between SO<sub>2</sub> concentration and meteorological parameters

	SO <sub>2</sub>	Stability	Ave. W.S.	No. of times (<1.5m/s)	Ave. temp.	Ave. hum.	Prec.	Previous day's conc.
SO <sub>2</sub>	1.0000							
Stability	.4499**	1.0000						
Ave. W.S.	-.1258*	.0525	1.0000					
No. of times (<1.5m/s)	.2270**	.5481**	-.6628**	1.0000				
Ave. temp.	-.8070**	.2193**	-.0342	.1168	1.0000			
Ave. hum.	-.0252	.5951**	-.1530*	.6156**	.3672**	1.0000		
Prec.	-.2688**	-.6971**	-.0113	-.4268**	.0414	-.4084**	1.0000	
Previous day's conc.	.8525**	.1109	-.0949	-.0503	.8290**	-.3086**	-.0028	1.0000

\*\*\* significance at P≤0.05 or 0.01, respectively 순이었다.

안정도의 경우 예상보다 낮은 상관성이 나타났는데 이는 안정도가 고농도에서 강한 상관을 보이는데 여름철의 경우 대체적으로 오염도가 낮았고 또한 시간별 안정도의 평균치를 일평균치로 가정하였기 때문에 이로 인하여 상관계수가 비교적 낮게 나타난 것으로 사료된다. 이를 확인하기 위해 동절기(12, 1, 2월)와 하절기(6, 7, 8월)에 대한 SO<sub>2</sub>농도와 안정도와의 상관계수를 구한 결과 예상대로 동절기에는 0.7657, 하절기에는 -0.0828로 동절기에 SO<sub>2</sub>농도와 안정도가 높은 상관을 보였다. 김정필(1989)의 실제 고농도에서의 기상의 변화를 조사한 논문에서도 이러한 결과를 입증하고 있다.

기상인자 이외에 전일SO<sub>2</sub>농도가 0.8525의 높은 상관계수를 나타냈으며, 앞으로 대기 오염 예보에 있어서 중요하게 취급되어야 할 예보인자라고 사료된다.

3.3 중회귀분석

중회귀모델을 이용하여 SO<sub>2</sub>농도와 관련 9개인자(안정도, 평균풍속, 풍속 1.5m/sec이하의 시간수, 평균온도, 최고온도, 최저온도, 평균습도, 강수량, 전일SO<sub>2</sub>농도)와의 중회귀분석결과 다음의 회귀식이 구해졌다.

$$X = 0.61X_1 - 0.25X_2 + 0.60X_3 - 2.1X_4 + 1.14X_5 + 0.17X_6 - 1.21 \quad (2)$$

여기서 X=SO<sub>2</sub>예측농도, X<sub>1</sub>=전일SO<sub>2</sub>농도, X<sub>2</sub>=최저온도, X<sub>3</sub>=1.5m/sec 이하의 시간수, X<sub>4</sub>=평균온도, X<sub>5</sub>=최고온도, X<sub>6</sub>=평균습도이다.

식(2)에서 6개 인자만이 선정되었는데 이는 다중회귀모형의 인자분석에서 적합도가 떨어지는 3개 인자가 제외된 것이다. 식(2)의 R<sup>2</sup>은 0.944로 매우 높게 나타났다. 그러나 이 결과는 온도가 최고, 평균, 최저의 3가지 형식으로 중복사용되는 등 지나치게 적합도의 향상에만 치중한 경향이 있다. 따라서 중회귀분석시 적합도가 가장 높은 최저온도만을 취하고 최고온도와 평균온도를 제외하여 회귀분석 한 결과 식(3)과 같이 전일 SO<sub>2</sub>농도, 평균풍속, 최저온도만이 선정된 회귀식이 구해졌다.

$$X = 0.62X_1 - 1.22X_2 - 7.19X_3 + 35.15 \quad (3)$$

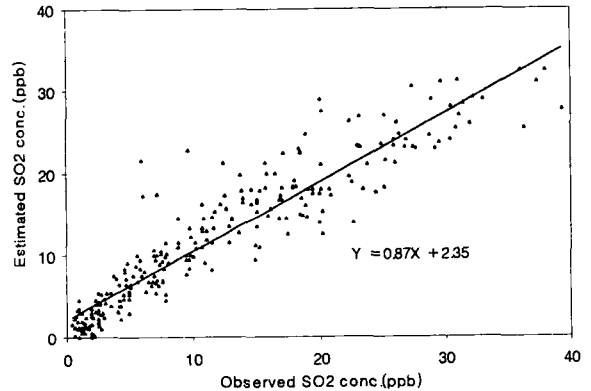


Fig. 3. Scatter diagram of observed and estimated SO<sub>2</sub> concentration by multiple regression model.

여기서 X=SO<sub>2</sub> 예측농도, X<sub>1</sub>=전일 SO<sub>2</sub>농도, X<sub>2</sub>=평균 풍속, X<sub>3</sub>=최저온도이다.

이때 식(3)의 R<sup>2</sup>은 0.938로 식(2)에 비해서는 적합도가 약간 떨어지나 식이 간략하고 기상 인자의 중복이 없으므로 식(2)보다는 바람직한 예보식이라 판단된다.

Table 3은 식(3)이 계산되는 과정에서 인자들이 식에 기여한 정도에 따라 인자선택기준과 계산에 참여한 인자들의 표준오차값을 나타내고 있다. 여기서 기여도(T)란 식에 대한 그 인자의 기여율을 나타내는 것으로 절댓값이 높을수록 그 인자의 식에 대한 기여율은 큰 것이며 유의도는 인자의 식에 대한 기여율의 유의도를 나타내는 것으로 일반적으로 유의도 0.05이하이면 그 인자를 채택하게 되는데 강우량과 안정도는 기여도가 낮게 분석되었고 유의수준도 벗어났다. 여기서의 1번 인자인 전일농도의 경우 기여도는 18.843이고 유의도는 0.0001이하로 식(3)에 대해 대단히 높은 기여를 하고 있으며 평균풍속과 최저온도도 비교적 기여도가 높은 것으로 나타났다.

여기서 식(3)의 결과에 대하여 고찰하면 기온과 풍속

Table 3. Variables in the equation of T-test

	STD ERROR	T	SIGNIFICANCE
Previous day's conc.	0.03274	18.843	<0.0001
Ave. W.S.	0.11557	-10.518	<0.0001
Min. temp.	1.03960	-6.920	<0.0001
Prec.	*	-0.801	0.4238
Stability	*	-0.282	0.7783

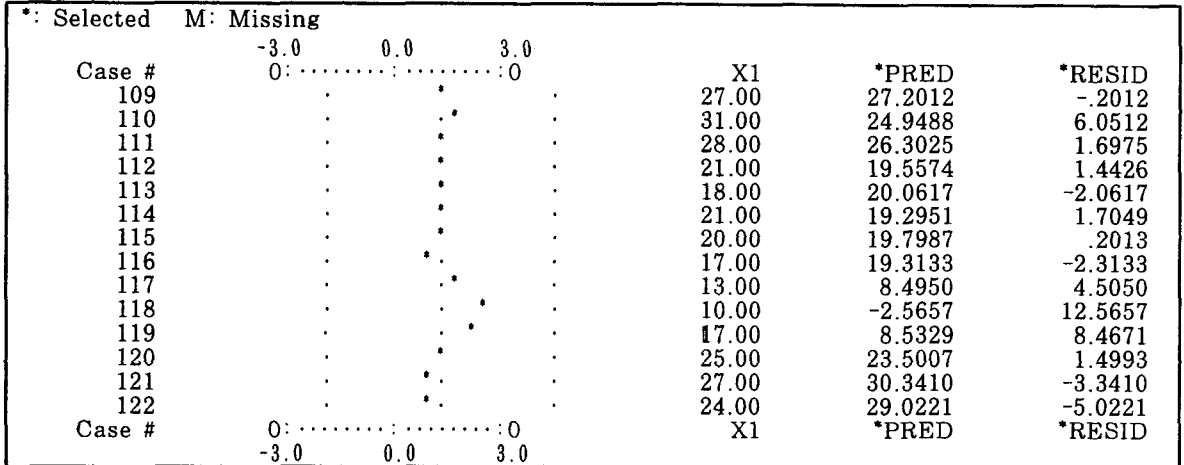


Fig. 4. Casewise plot of standadized residual in summer season.

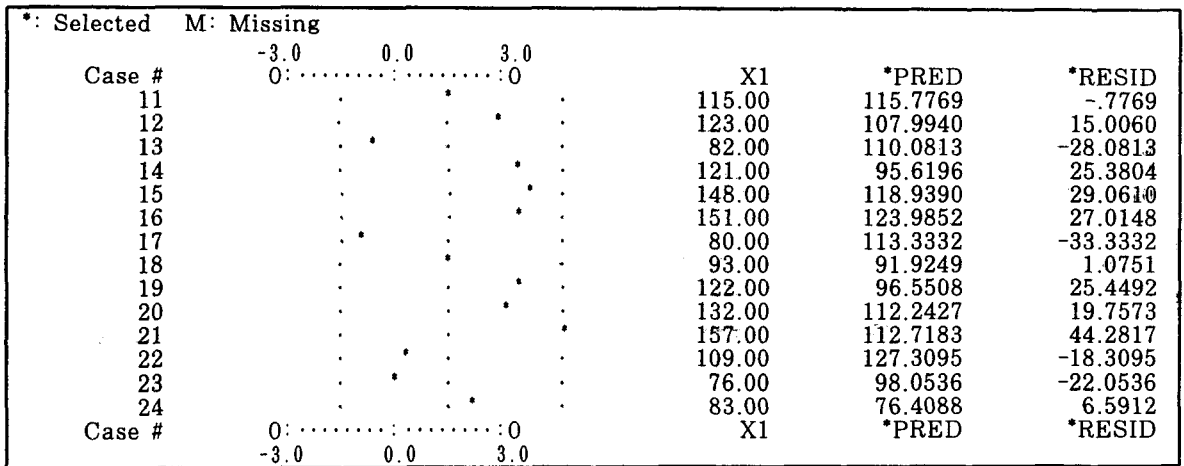


Fig. 5. Casewise plot of standadized residual in winter season.

은 SO<sub>2</sub>농도와 역상관으로써 기온의 증가는 SO<sub>2</sub> 농도의 감소를 나타내고 있으며 풍속의 증가는 바람에 의한 확산효과로 SO<sub>2</sub>농도를 감소시키고 있음을 의미하고 전날의 농도는 양의 상관으로 전날의 농도가 높을수록 높은 농도를 나타내는 양상을 보이고 있다. 중회귀분석에 의한 SO<sub>2</sub> 예측치와 실측치의 분포는 Fig. 3과 같다.

이 식을 실제로 적용하여 계산해본 결과 비교적 변동의 폭이 크지 않은 하절기의 경우에는 적합도가 매우 우수하며, 예측값과 실측값의 차이로 인한 분산의 결과가 매우 적었으나 고농도의 동절기의 경우에는 계산 값의

적중도와 집중성이 나쁘게 나타났다. 이러한 예 중에서 가장 대조를 잘 이루고 있는 사례를 Fig. 4와 5에 제시하였다.

그림에서 X<sub>1</sub>은 실측된 SO<sub>2</sub>의 농도이고 \*PRED는 식을 적용하였을 때의 예측치이고 \*RESID는 예측치와 실측치의 차이를 나타내고 있으며 case #(no.)의 숫자는 2년간 전체 자료 중 날짜 일련번호를 의미하는 것이다.

중회귀 분석 결과 제시된 최적 예보식 (3)의 적합도를 분석하기 위해 다중결정계수(coefficient of multiple determination : R<sup>2</sup>)와 RMSE(root mean square er-

Table 4. Seasonal R<sup>2</sup> and RMSE(ppb) of multiple regression equation-(3)

	R <sup>2</sup>	RMSE
Spring	0.94	5.72
Summer	0.97	3.84
Fall	0.93	4.11
Winter	0.82	10.26

ror)를 계절별로 분석하여 표 4에 제시하였다.

표 4에서와 같이 여름, 봄, 가을철의 경우 예보식 (3)의 적합도가 매우 우수한 것으로 나타났으나 겨울철의 경우 적중도와 집중성이 떨어지는 것으로 분석되었다. 또한 RMSE의 경우도 오염도가 낮은 여름에 3.84로 가장 작았고 오염도가 큰 겨울에 10.26으로 크게 나타났다. 따라서 겨울철의 경우 적합도를 높이기 위해 별도의 중회귀 분석을 실시한 결과 나타난 회귀식은 식 (4)와 같다.

$$X = 0.51X_1 + 12.37X_2 - 23.42 \quad (4)$$

여기서 X=SO<sub>2</sub> 예측농도, X<sub>1</sub>=전일 SO<sub>2</sub> 농도, X<sub>2</sub>=안정도이다.

겨울철에 적용된 식 (4)의 경우 R<sup>2</sup>은 0.89이며, 기상요소 중 안정도는 1번 인자로 선정하고 있음을 알 수 있다. 이는 SO<sub>2</sub> 고농도일 경우에는 SO<sub>2</sub> 농도와 대기안정도의 상관관이 높다는 기존의 연구와도 일치하며 본 연구에서 동절기와 하절기의 SO<sub>2</sub> 농도와 안정도의 상관분석 결과도 동절기에는 0.7657로 매우 높았으나 하절기에는 -0.0828로 거의 상관성이 없는 것으로 나타나 SO<sub>2</sub> 오염도가 높은 동절기에는 식(4)의 적용이 적합할 것으로 판단된다.

#### 4. 결 론

본 연구는 1996년 1월부터 1997년 12월까지 2년간 원주시 대기오염 자동측정망의 SO<sub>2</sub> 관측자료와 원주시 상대의 기상관측자료를 이용하여 SO<sub>2</sub> 오염도와 SO<sub>2</sub> 오염도에 영향을 주는 기상인자 (대기안정도, 풍속, 기온, 습도, 강우량)와의 관련성을 각종 통계분석을 통해 검토하였고 이를 토대로 원주시역의 SO<sub>2</sub> 오염도 예보를 위한 최적식을 제시하였으며 다음과 같은 결론을 얻었다.

1) 원주시역의 SO<sub>2</sub> 오염도에 영향을 많이 미치는 기상인자는 기온(평균온도), 대기안정도 순으로 상관계수가 각각 -0.8070, 0.4499로 분석되었으며 기상인자의 예는 전날의 SO<sub>2</sub> 농도가 0.8525로 높은 상관성을 보였다.

2) 중회귀분석결과 원주시역의 SO<sub>2</sub> 오염도 예보는 식 (3)  $X=0.62X_1-1.22X_2-7.19X_3+35.15(R^2=0.94)$ 이 가장 적합한 것으로 분석되었으며 이 식을 이용해 예측한 값과 실측치와의 회귀분석 결과 상관계수 0.97의 높은 상

관을 보였다.

3) 식(3)의 경우는 Table 4와 같이 봄, 여름, 가을에는 우수한 적합도를 보였으나 동절기에 적합도가 떨어지는 경향을 보였으며 동절기에는 안정도가 포함된 식(4)  $X=0.51X_1+12.37X_2-23.42$ 의 적용이 적합할 것으로 판단된다.

이러한 통계분석을 통한 대기오염 예측 및 예보방법이 아직까지는 널리 이용되고 있지는 않지만 배출원 중심의 확산모델에 비해 식이 비교적 간단하며 반응성 물질에 대해서도 폭넓게 적용할 수 있는 등 많은 장점을 갖고 있어 이에 대한 지속적인 연구와 활용이 기대된다.

#### 참 고 문 헌

기상청, 1996-1997 원주시 기상자료  
 기상청, 1996-1997 기상연보  
 김용국,이종범, 1994, 하계의 일 최고 오존농도 예측을 위한 신경망 모델의 개발, 한국대기보전학회지, 10(4), 224-232  
 김용준, 1997, 현업 운영 가능한 서울지역의 일 최고 대기오염도 예보모델 개발 연구, 한국대기보전학회지, 13(1), 79-89  
 김정필, 1989, 복잡한 지형에서의 대기오염확산계수 개선에 관한 연구, 인천대 석사학위논문  
 신찬기,한진석,김윤신, 1992, 대기오염농도와 기상인자의 관련성 연구, 한국대기보전학회지, 8(4), 213-220  
 허정숙,김동술, 1993, 다변량통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구, 한국대기보전학회지, 9(3), 207-215  
 홍낙기,이종범,김용국, 1993, 다변량해석을 이용한 서울시 하계 스모그의 형태 분류, 한국대기보전학회지, 9(4), 278-287  
 환경처, 1996-1997 대기오염 측정자료  
 Joukoff and Malet, 1982, Daily Forecasting of Air Pollution Potential, Atmospheric Pollution 1982, 97-102  
 P.Allender and J.M. Dejardin, 1982, The Forecasting Method of Air Pollution Peaks Developed and Used in the Nord-Pas-De-Calais Area, Atmospheric Pollution 1982, 103-105  
 Smith, F.B., and Jeffrey, G.H., 1972, Aspects Air Quality Data, Chapel Hill, N.C., 9 -10  
 Turner, D. B., 1961, Tennessee. J. Air Poll. Control Assoc., 11  
 Turner, D.B., 1964, A Diffusion Model for on Urban Area. Journal of Applied Meteorology, Vol. 3, 83-91