

# ARIMA 모델에 의한 서울시 일부지역 SO<sub>2</sub> 오염도의 월별 변화에 대한 시계열 분석

환경조사과

김 광 진

## A time series analysis for the monthly variation of SO<sub>2</sub> in the certain areas of Seoul using ARIMA model

*Environmental research division*

**Kwang-Jin Kim**

### =Abstract=

The typical ARIMA model which was developed by Box and Jenkins, was applied to the monthly SO<sub>2</sub> data collected at Seoungsoo and Oryudong in metropolitan area over five years, 1982 to 1986. To find out the changing pattern of SO<sub>2</sub> concentration, autocorrelation and partial autocorrelation analysis were undertaken. The three steps of time series model building were followed and the residual series was found to be a random white noise.

The results of this study is summarized as follows.

- 1) The monthly SO<sub>2</sub> series was found to be a non-stationary series which has a periodicity of 12 months. After eliminating the periodicity by differencing, the monthly SO<sub>2</sub> series became a stationary series.
- 2) The ARIMA seasonal model of the SO<sub>2</sub> was determined to be ARIMA (1, 0, 0)(0, 1, 0)<sub>12</sub> model.
- 3) The model equations based on the prediction were:

for Seoungsoodong:

$$Y_t = 0.5214 Y_{t-1} + Y_{t-12} - 0.5214 Y_{t-13} + a_t$$

for Oryudong:

$$Y_t = 0.8549 Y_{t-1} + Y_{t-12} - 0.8549 Y_{t-13} + a_t$$

- 4) The validity of the model identified was checked by comparing the measured SO<sub>2</sub> values and one-month-ahead predicted values. The result of correlation and regression analysis is as follows.

Seoungsoodong:

$$Y = 0.8710X + 0.0062 \quad r = 0.8768$$

Oryudong:

$$Y = 0.8758X + 0.0073 \quad r = 0.9512$$

### I. 서 론

일반적인 통계기법은 관측치의 정상분포와 독립성에

기초를 둔데 반하여 시계열분석은 관측치의 발생순서를 보전하는 시계열을 중심으로 관측치간의 자기상관성 (autocorrelation)에 기초를 두고 있다. 특정변수의 시간에 따른 변화양상을 파악하기 위한 시계열분석은

각 관측치간의 상호관계를 밝히고 이를 바탕으로 한 모형을 확정하여 미래를 예측하고자 하는 체계적인 방법이다. 시계열분석은 과거의 시계열 형태가 그 특성을 잃지 않고 미래에도 반복된다고 가정하여, 일일이 그 변화 원인을 규명할 수 없는 연속적 자료의 단기에측에 유용한 추계학적 통계분석기법이다(Box 등, 1970). 시계열분석은 최초에 공학분야에서 개발되었으나 그 후, 경제학, 사회과학, 그리고 환경과학분야에서 주로 단기에측을 위한 모델로서 많이 응용되고 있다. 환경오염물질 변화에 대한 추계학적분석은 기상분야와 하천수문분야에서 응용되고 있으며 세계적으로는 1970년대 후반부터 많은 연구결과가 보고 되고 있다. (이상훈, 1987), Box와 Jenkins(1970)가 ARIMA model을 제시한 이후 Mehta와 Ahlert(1975) 등은 유량, 수온, BOD 및 DO의 일별 변화자료를 가지고 ARIMA모형을 이용하여 추계학적 분석을 시도하였다. 대기오염에 대한 추계학적인 분석으로 Box와 Tiao(1975) 등에 의하여 환경자료에 대하여 시계열분석방법의 일종인 개입분석법을 도입하였다. 또한 이들은 Los Angeles 광화학 smog를 시계열분석으로 그 변동을 연구하였다. 국내에서 연구하는 이흥근(1977)이 한강의 수질변동에 대하여 Spectral density를 이용한 용존산소(DO)의 시간적 변동분석을 보고한 바 있으며 또한 1982년에는 노량진과 죽도 유역에서 DO와 탁도에 대한 수질변화의 주기성을 발견하고, 수질의 시간적변동 상태를 규명하여 시간별 및 일별에 대한 ARIMA 모형을 추론하였다. 이종남 등(1984)도 섬진강 월유출량을 대상으로 ARIMA 모형에 의한 시계열분석을 시도하였으며, 안령미(1984)은 죽도, 구의, 구미, 왜관에서 하천의 일별 수온차에 대한 용존산소(DO)변화의 시계열적 특성을 분석하였다. 그러나 현재까지 대기분야에 대한 시계열분석이 응용된 연구는 아직까지 국내에 보고된 바 없다.

우리나라는 4계절이 뚜렷하여 특히 겨울동안에 난방으로 인한 연료사용량이 증가하여 도시에서는 대기오염도가 높은 특성을 나타내고 있으므로 시계열적 오염도의 변화가 예상된다. 따라서 본 연구는 한도시의 대기오염의 시계열적 특성을 분석하고 대기오염농도의 단기에측의 가능성을 탐색하여 대기오염 방지대책의 기초자료를 제공하기 위하여 수행되었다.

## II. 연구방법

### 1. 대상 및 자료

서울 시내 성수, 오류동에 설치되어 있는 대기오염

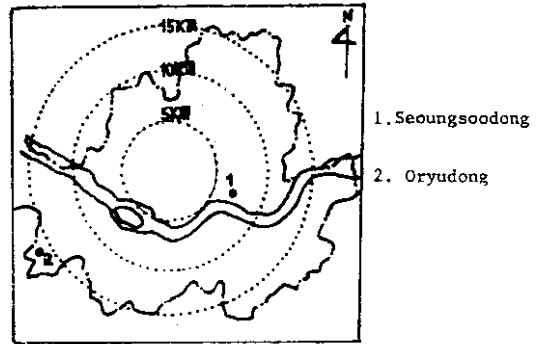


Fig. 1. Sampling sites in Seoul area.

자동 측정망에서 측정되는 아황산가스를 대상으로 1982년 3월부터 1986년 12월까지 5년간의 오염도자료를 수집하였다. 자동측정소의 아황산가스의 측정은 KIMOTO McSAM-1 model에 의한 용해전도율법으로 매시간 연속으로 측정된 자료로 부터 월평균치를 산출하였다. 서울특별시에서 아황산가스는 대기오염물질(critical pollutant)로서 보고되고 있다(정용, 1986).

### 2. 조사지점

Fig. 1과 같이 서울시 성수 및 오류동을 택하였다.

가. 성수동 : 측정소를 중심으로 반경 2km 이내 서편으로는 준공업지역으로 산업장이 산재하고 있으며 북으로는 중랑천과 종말처리장이, 남동편으로는 어린이대공원이, 남으로는 한강이 흐르는 준공업지역으로서 아파트와 주택, 상가 등이 혼재하고 있다. 대체로 2km 반경이내의 지역인구는 약 20만 정도이다. 성수동 대기오염 자동측정소는 1982년 3월부터 상시 측정하고 있으며 측정기기의 위치는 지상 12m에서 시료를 흡입하였다.

나. 오류동 : 측정소를 중심으로하여 동쪽으로는 구로동 공업단지가, 남쪽으로는 농경지와 철도 및 공장이 산재하고 있으며 서쪽으로는 은수동 기계공업단지가 주거지와 혼재하고 있으며 북쪽으로는 아산이 있다. 이 지역의 인구는 대체로 10만인으로 추정된다. 오류동 대기오염 자동측정소는 1982년 3월부터 상시 측정하고 있으며 측정기기의 위치는 지상 5m에서 시료를 흡입하였다.

### 3. 시계열분석 방법

일별 대기오염도 자료는 Box와 Jenkins의 ARIMA model을 이용하여 분석하였다. 시계열분석에는 PC형 시계열분석 Package인 Autobox(Box-Jenkins package)를 이용하였다.

시계열분석은 다음 Fig. 2와 같이 모형의 확인(identification), 변수의 추정(estimation), 적합성 검토

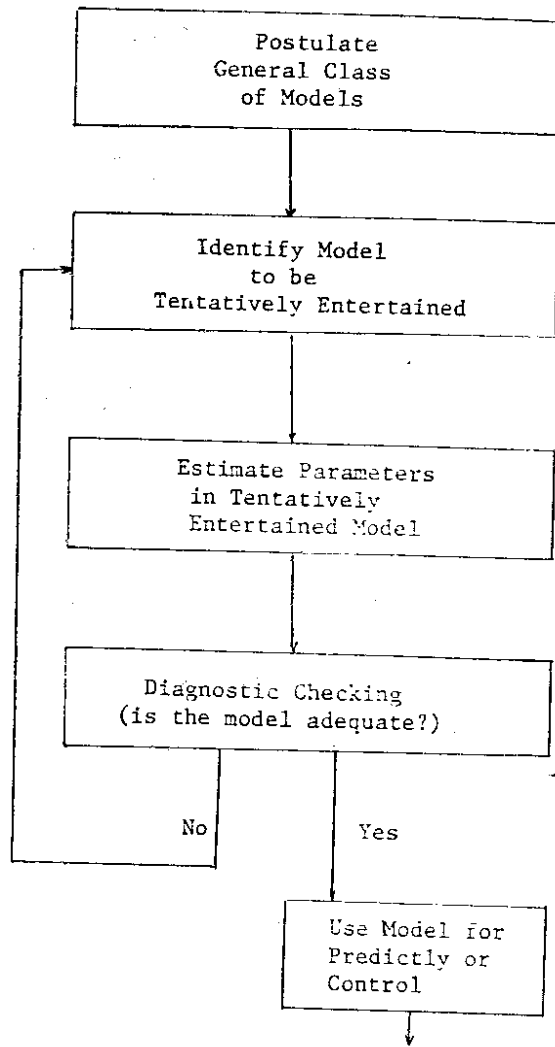


Fig. 2. Stages in the iterative approach to model building.

(diagnostic checking)의 3단계 절차를 통하여 수행하였다. 적합성 검토에서 잠정적으로 선택한 모형이 적합하지 않다고 판단되면 제 1 단계인 확인단계로 되돌아가며 이러한 과정은 더 이상 모형을 개선할 수 있는 정보를 얻을 수 없을 때까지 거듭한다.

가. 모형의 확인(model identification)

잠정적인 모형을 정하는 것으로 원시계열  $Y_t$ (original time series)의 자기상관함수(ACF: autocorrelation function)를 조사하여 ARIMA(autoregressive integrated moving average) model을 적용하여 주기성을 검토하였다. 자기상관함수(ACF)에 의하여 주기성이 존재하는 경우에는 원시계열에서 차이계산(differencing,  $d$ )을 취하여 주기성분을 제거하였다. 원시계열  $Y_t$ 가

안정시계열이면  $d=0$ 이며  $Z_t = Y_t - \bar{Y}_t$ 이다.

여기서  $Z_t$ 는 안정시계열

$Y_t$ 는 원시계열

$\bar{Y}_t$ 는 원시계열의 평균값

안정시계열  $Z_t$ 가 얻어지면 자기상관함수(ACF)와 편 자기상관함수(PACF: partial autocorrelation function)의 상관도(correlogram)를 분석하여 모형에 필요한 자기회귀모형(AR)과 이동평균모형(MA)의 변수(parameter)의 수를 결정하였다. 자기회귀모형(AR: autoregressive model)은 시계열에서 현재 관측치를 설명하여 줄수 있는 시계열의 이전 값들( $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ )과 설명하여 주지 못하는 부분  $a_t$ 의 선형 결합으로써 표시되므로 다음 식 (1)과 같이 표시할 수 있다.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t \quad (1)$$

여기서 방정식(1)은 차수가  $p$ 인 자동회귀과정 AR( $p$ )이라 불리우며  $\phi_p$ 는  $p$ 번째 전의 기간에서의 가중치를 나타낸다. 예를 들어 기간  $t$ 의 값이 단지 이전 기간  $t-1$ 의 관측치에 의존할 때 이를 1차 회귀모형 AR(1)이라고 하며, 기간  $t$ 에서의 관측치가 이전의 2개의 관측치에 모두 의존할때 이를 2차의 자동회귀모형 AR(2)라 한다. 즉 AR(1)이면 자동상관함수는 Lag  $K$ (시간차  $k$ )의 증가에 따라 AR(1)의 경우보다 다소 완만하게 감소되며 상하로 움직인다. 한편 이동평균모형(MA: moving average model)은 시계열의 현재값은 이전의 예측오차( $Y_t - \hat{Y}_t = a_t$ )의 선형 결합으로 표시할 수 있다는 가정에 기반을 둔 것으로 다음 식 (2)와 같이 나타낼 수 있다.

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2)$$

MA(1)은 Lag  $K$ (시간차  $K$ )  $K=1$ 에서 spike를 나타내며 MA(2)이면 Lag  $K$ (시간차  $K$ )  $K=1, 2$ 에서 spike를 나타낸다.

관측치 상호간의 연관성을 나타내는 자동상관함수(ACF)는 측정된 자료를 다음 식 (3)에 대입하여 분석을 행하였으며 시간차에 의한 두 관측치의 상관관계가 다른 시간차의 관계를 고정시킬때 얼마나 밀접한가를 나타내는 척도인 편자기상관함수(PACF)는 다음 식 (4)에 관측치를 대입하여 분석하였다.

$$ACF(k) = \frac{\sum_{i=1}^{N-k} (Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N-k} (Y_i - \bar{Y})^2} \cdot \frac{N}{N-k} \quad (3)$$

여기서

$N$ 은 관측치의 수

$Y_t$ 는 시계열에서 관측한 값

$\bar{Y}$ 는  $Y_t$ 의 평균값

$$r_1; k=1$$

$$r_k = \frac{\sum_{j=1}^{k-1} \phi_{k-1,j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j}; k=2, 3, \dots, k \quad (4)$$

여기서

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j} \quad (5)$$

실질적으로 자동상관함수의 유용한 추정치를 얻기 위해 적어도 50개의 관측치가 필요하며 추정된 자동상관계수  $r_k (k=0, 1, 2, \dots, k)$ 의 수효는 관측치의 갯수의 1/4 (즉  $n/4$ )보다 크지 않아야 한다(Box, 1970).

나. 변수의 추정(parameter estimation)

본 연구의 대상인 아황산가스 오염도의 시계열분석으로 시간에 따른 변수인 아황산가스 농도의 변화를 ARIMA model에 적용하여 구하여 보았다.

$$(1 - \phi_1 B)(1 - B^{12}) Y_t = a_t \quad (6)$$

$B$ : 역변환인자(즉  $BY_t = Y_{t-1}$ )

$$(1 - B^{12} - \phi_1 B + \phi_1 B^{13}) Y_t = a_t$$

$$Y_t - Y_{t-12} - \phi_1 Y_{t-1} - \phi_1 Y_{t-13} = a_t$$

$$Y_t = \phi_1 Y_{t-1} + Y_{t-12} - \phi_1 Y_{t-13} + a_t \quad (7)$$

실측시계열인  $Y_t$ 와 모형으로 부터 구해진 예측시계열  $\hat{Y}_t$ 의 residual의 자승합을 최소화 시키도록 하였다. Residual 자승합( $S$ )을 식으로 표시하면 다음과 같다.

$$S(\hat{\phi}, \hat{\theta}) = \sum_{t=1}^N (Y_t - \hat{Y}_t)^2 \quad (8)$$

여기서  $Y_t - \hat{Y}_t$ 은 실제 관측과 추정된 관측치간의 차이인 residual error를 나타낸다. 또한  $\hat{\phi}$ 과  $\hat{\theta}$ 은 각기 자동회귀모형과 이동평균모형의 계수를 나타낸다.

이와 같이 residual 합이 최소화되는 모형을 구하였다.

다. 모형의 적합성 검토(Diagnostic checking)

모형의 적합성은 model residual의 자동상관함수를 평가하는 것으로 이루어진다. 만약 residual가 배경농도(white noise)와 다르지 않다면 residual의 모든 Lag  $K$ (시간차)에서의 자동상관함수는 0으로 기대될 것이다. 따라서 residual 총 통계적 유의를 알기 위하여 다음  $x^2$ 변이의 총수인  $Q$ 통계치를 이용하였다.

$$Q = n \sum_{k=1}^k r_k^2(\hat{a}_t) \quad (9)$$

$Q$ : Chi-square 변이의 총수

$n$ : 잔차 시계열의 관측치수

$K$ : 계산한 자동상관관계의 수

$r_k(\hat{a})$ : 시간차  $K$ 에서  $\hat{a}_t$ 시계열의 자동상관계수

$Q$ 은 자유도가  $K - p - q$ 인  $x^2$ 표준값과 비교되는데  $p$ 와  $q$ 는 추정하려는 자동회귀모형과 이동평균모형의 항의

수를 나타낸다.  $Q$ 값이  $x^2$ 보다 클때 이는 현재모형이 부적당하며 모형을 다시 세워야 함을 말한다(Box, 1970; McDowall, 1981; 이홍근, 1982).

또한 시간차  $K$ 일때 자동상관 계수의 분산은 다음과 같은 식으로 구하여 보았다.

$$\text{Var}(r_k) = \frac{1}{N} \sum_{j=-\infty}^{\infty} (r_j^2 + r_{j+k} \cdot r_{j-k} - 4r_k \cdot r_j \cdot r_{j-k} + 2r_j^2 \cdot r_k^2)$$

여기서  $r_j$ : 시간차  $j$ 에서 자동상관계수

$r_k$ : 시간차  $K$ 에 있어 자동상관 관계

$j > p$ 일때  $r_j$ 가 0인 과정에서 시간차  $K$ 인 자동상관 관계의 표준오차의 근사치는 다음과 같다.

$$Sr_k \approx \sqrt{\frac{1}{n} (1 + 2 \sum_{j=1}^{k-1} r_j^2)} \quad (11)$$

$Sr_k$ : 시간차  $K$ 에 있어 자동상관 관계 표준오차

$n$ : 시계열 자료수  $N$ 에서 차이계산  $d$ 를 제한

$(N-d)$

$r_j$ : 시간차  $i$ 에서 자동상관계수

또한 예측치의 정확도를 나타내는 지수인 RMSE (root mean square error)는 다음과 같이 계산하였다.

$$\text{RMSE} = \sqrt{\frac{(\sum Y_{act} - Y_{pre})^2}{n}} \quad (12)$$

$Y_{act}$ : 실측값

$Y_{pre}$ : 추계값

### III. 결과 및 논의

#### 1. 실측오염도 분석

Fig. 3은 성수, 오류동의 대기오염 자동측정소에서 1982년 3월부터 1986년 12월까지 측정된 아황산가스에 대한 월 평균농도를 나타낸 것이다. 5개년 동안의 월 변화를 보면 두지점 모두 겨울철중 가장 기온이 하강하는 1월을 정점으로 서서히 하향곡선을 그리다가 6,7월에 최저농도를 나타내고 있으며 9월부터 서서히 증가하기 시작한다. 아황산가스의 오염도는 계절에 따라 뚜렷한 차이를 보이고 있어 겨울철에는 높은 오염도를 여름철에는 낮은 오염도를 나타내고 있다. 권 등(1985)이 조사한 대기오염도에 영향을 미치는 기후요소 분석에 의하면 아황산가스 농도는 기온과 가장 높은 역 상관관계를 나타내고 있다. 반면 오염물질 배출량과는 높은 순 상관관계를 나타냈다고 보고하였다. 아황산가스가 기온과 높은 역 상관관계를 나타내는 것은 기온이 높은 여름철에는 연료의 사용량이 낮기 때문에 아황산가스의 배출량이 낮아 오염도가 낮으며 기온이 낮은 겨울철에는 연료의 사용량이 증가하여 아황산가스

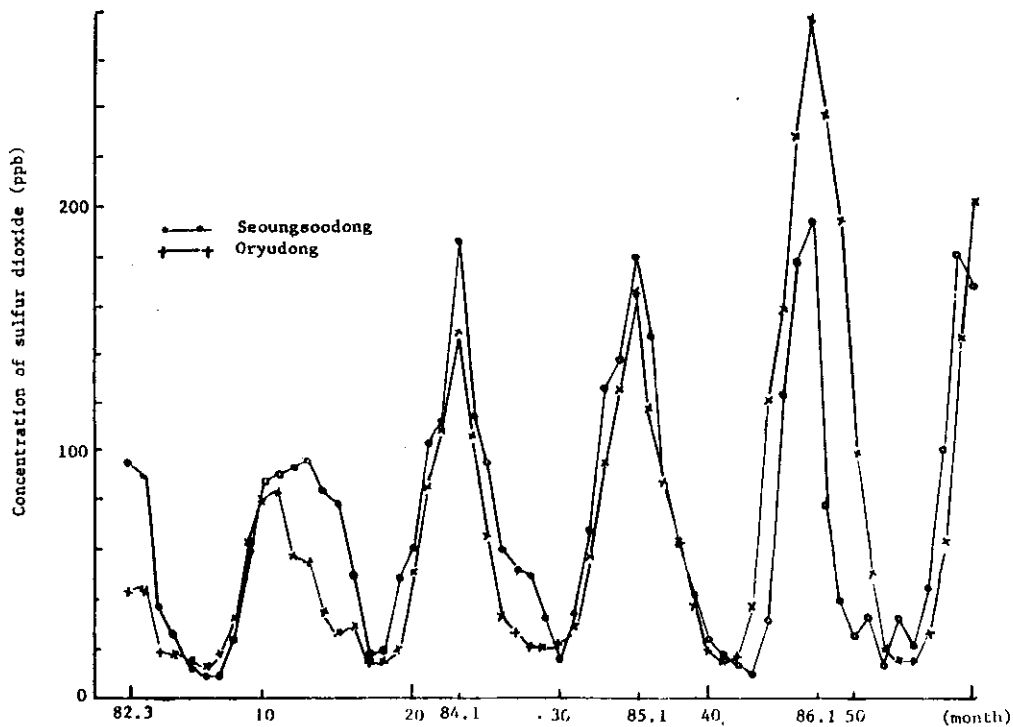


Fig. 3. Monthly variation of sulfur dioxide in Seung Soo and Oryudong area.

의 배출량이 증가하여 농도가 높아지기 때문에 기온이 연중 아황산가스 농도와 높은 역 상관관계를 나타내고 있다.

## 2. 모형의 확인

Fig. 4은 성수동에서 원시계열에 대한 자동상관함수를 나타낸 것으로 시간차  $K$ (Lag  $K$ )의 증가에 따라 자

동상관계수( $r_k$ )가 sine곡선을 그리며 증감을 나타내므로 이는 주기성을 가진 불안정 시계열임을 알 수 있었다. 즉 시간차 12인  $r_{12}$ 값이 크게 나타나고 있어 시계열  $Y_t$ 는 12개월의 주기를 가지는 불안정 시계열임을 알 수 있다. Fig. 5은 오류동에 대한 자기상관함수의 변화로서 성수동과 같은 양상을 보여주고 있다.

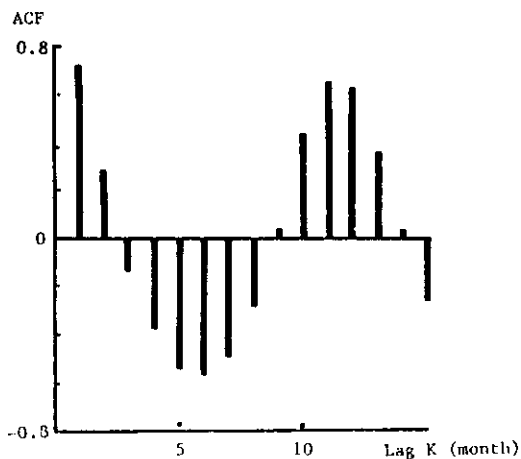


Fig. 4. The autocorrelation of the original series in Seungsoodong.

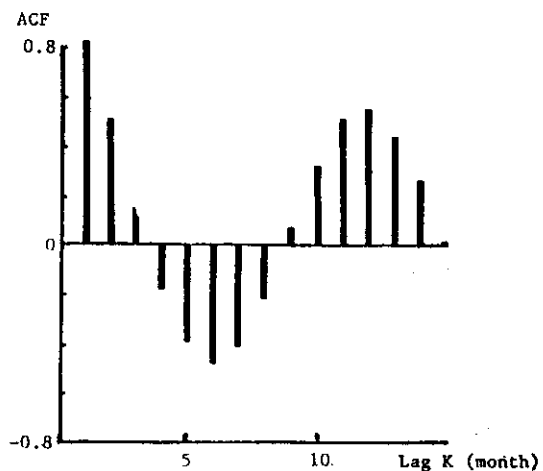


Fig. 5. Autocorrelation of the original series in Oryudong.

시계열  $Y_t$ 가 불안정성을 나타낼 때는 차이계산을 행하여  $Y_t$ 를 안정시계열  $Z_t$ 가 되도록 하였다. 안정시계열  $Z_t$ 의 자기상관함수(ACF)와 편 자기상관함수(PACF)의 상관도(correlogram)를 분석하여 본결과는 Fig. 6, 7과 같다.

Fig. 6은 성수동의 경우로서 대체로 안정시계열  $Z_t$ 의 자동상관함수에서는 그 차이들이 적은 안정시계열이 되었다. Fig. 7은 오류동의 경우로서 차이계산한 자동상관함수로 시간차(Lag  $K$ ) 1, 2에서만 의의가 있고 나머지 자동상관계수( $r_k$ )는 모두 의의가 없는 것으로 나타났다. 또한 12개월의 주기성이 제거된 안정시계열이 됨을 알 수 있다. 한계 Fig. 8은 성수동의 차이계산

한 시계열에 대한 편자기 상관함수로 시간차  $K=1$ 에서  $r_1=0.496$ 이 크게 나타나고 있으며  $r_{14}$ 는  $-0.349$ 로 파상형으로 기록이 심하게 나타나고 있다.

본 연구에서는 원시계열에서 차이계산( $d$ : differencing)을 행하고 주기성분은 제거하여 보편 안정시계열  $Z_t$ 인 ARIMA model( $p, d, q$ )로 분석되었다. 계절성 주기를 제거한( $D=1$ ) 확률성분  $Z_t$ 의 ( $p, d, q$ ) 모형을 판별하기 위하여  $d=0, 1$ 에 대한 자기상관함수와 편 자기상관함수를 분석하여 보면, Fig. 4에서  $d=0$ 인 경우 자기상관함수는 시간차(Lag  $k$ )  $k=1$ 에서  $r_1=0.724$ 으로 부터 시작하여 완만하게 감소하고 있으나 Fig. 6에서는 차이계산을 위한 결과 자동상관계수( $r_k$ )가 급격

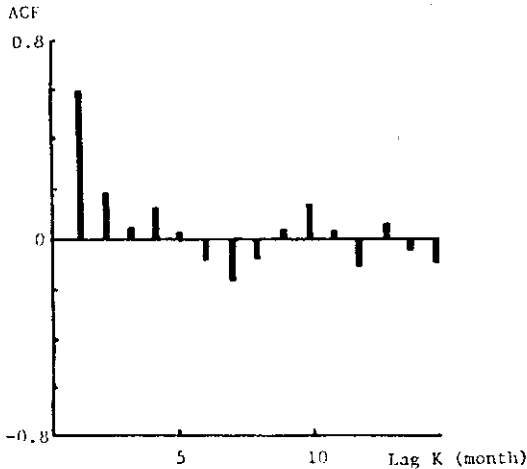


Fig. 6. The autocorrelation of the differenced series in Seungsoodong.

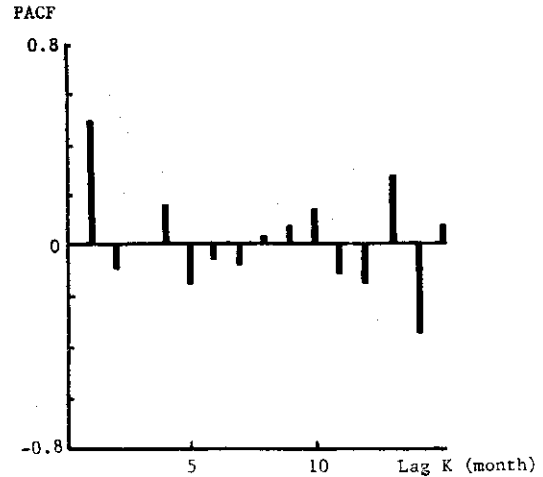


Fig. 8. The partial autocorrelation of the differenced series in Seungsoodong.

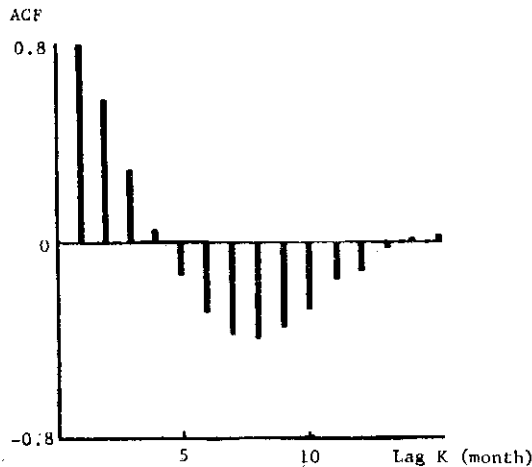


Fig. 7. The autocorrelation of the differenced series in Oryudong.

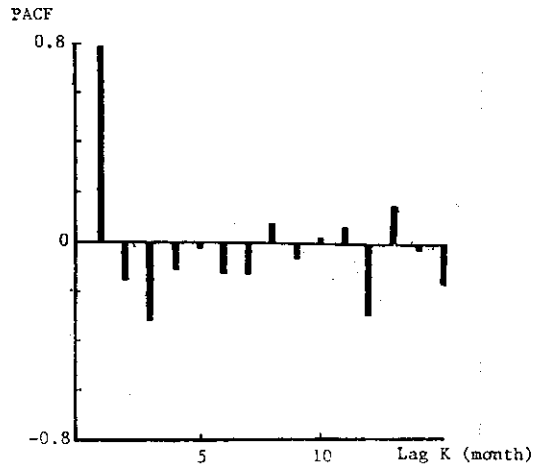


Fig. 9. The partial autocorrelation of the differenced series in Oryudong.

히 감소한다. 시간차  $K=2$ 에서 부터는 자동상관계수 ( $\gamma_k$ )의 표준오차( $Sr_k$ )에 대한 95% 유의수준  $\pm 1.96 \times Sr_k$ 를 조사한 바 전부 신뢰구간 이내에 들어와 소멸되어 의의가 있음을 알 수 있었다. Fig. 8은 차이계산한 편 자기 상관함수로  $K=1$ 에서만 돌출되었으며 나머지 모두 편 자기상관함수의 (PACF)의 표준오차( $\phi_{kk}$ ):

$$\phi_{kk} = \frac{1}{\sqrt{k}} \text{에 대한 95\% 유의수준 } \pm 1.96 \times \phi_{kk} \text{ 이내에}$$

있었다. 이처럼 자기상관함수와 편 자기상관함수를 분석한 결과 자기회귀모형 (AR: autoregressive model)은 AR(1)임을 알 수 있었다. 또한 Fig. 9는 오류동에 대한 것으로 성수동과 같은 양상을 나타내고 있다.

이상에서 자기상관함수와 편 자기상관함수를 검토한 바 ARIMA( $p, d, q$ )( $P, D, Q$ )형태서 (1, 0, 0)(0, 1, 0)<sub>12</sub>

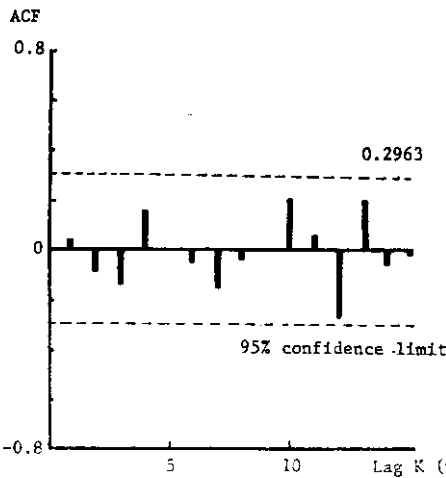


Fig. 10. The autocorrelation of the residual series in Seungsoodong.

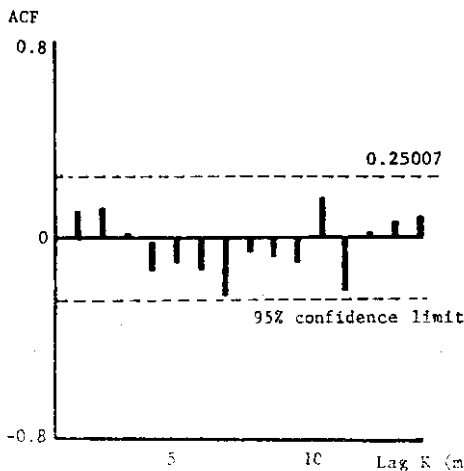


Fig. 11. The autocorrelation of the residual series in Oryudong.

Table 1. Lag coefficient of ARIMA model for Seung Soo and Oryudong.

	Differencing (D)	Order	Lag Coefficient ( $\phi_1$ )
Seung Soo	1	12	0.5214
Oryudong	1	12	0.8550

의 모형을 제시해 주고 있다.

여기서  $P$ 는 자동회귀모형의 차수를  $D$ 는 차이계산을,  $Q$ 는 이동평균모형의 차수를,  $S$ 는 계절주기를 표시한 것이다. 이 모형에 대한 식은 (6)과 같다

Fig. 10과 11은 성수, 오류동에 대한 잔차시계열 자동상관관계를 나타낸 것으로 점선은 95% 신뢰구간을 나타낸 것으로 모두 적합함을 알 수 있었다.

### 3. 추계 모형

실측치로부터 연속적 자동전자 계산에 의하여 주기성을 갖은 ARIMA model에서의 지체상관계수( $\phi$ : lag coefficient)를 계산한 결과는 Table 1과 같이, 성수동에서는  $\phi=0.5214$ , 오류동에서는  $\phi=0.8550$ 이었다.

따라서 각 연구대상 지역의 시계열 추계모형은 각기 다음 식과 같이 표현된다.

$$\begin{aligned} \text{성수: } Y_t &= 0.5214 Y_{t-1} + Y_{t-12} \\ &\quad - 0.5214 Y_{t-13} + a_t \end{aligned} \quad (13)$$

$$\begin{aligned} \text{오류: } Y_t &= 0.8550 Y_{t-1} + Y_{t-12} \\ &\quad - 0.8550 Y_{t-13} + a_t \end{aligned} \quad (14)$$

$Y_t$ :  $t$ 시간에 있어  $SO_2$ 농도

$Y_{t-1}$ :  $t-1$ 시간에 있어  $SO_2$ 농도

$Y_{t-12}$ :  $t-12$ 시간에 있어  $SO_2$ 농도

$Y_{t-13}$ :  $t-13$ 시간에 있어  $SO_2$ 농도

### 4. 모형의 적합도 검증

위 식 (13), (14)에서 얻은 추계모형에 의하여 각기 추계된 결과를 실측치가 비교하면 다음 Fig. 12, 13과 같다.

Fig. 12는 성수동에 대한 실측치와 추계치를 나타낸 것으로 최대차이는 0.086ppm이며, 최소차이는 0.002 ppm였다. Residual series  $a_t$ 에 대한 전체합은 0.000194로서 배경농도(white noise)의 이론적 값인 0에 접근하는 정규분포를 나타내고 있으며, Fig. 13은 오류동에 대한 실측치와 추계치를 나타낸 것으로 최대차이는 0.059ppm이며 최소차이는 0.001ppm였다. Residual series  $a_t$ 가 0.000199로서 0에 접근하는 정규분포를 하고 있다.

통계적 유의성을 검증하기 위하여 식 (9)에 대입하여  $Q$ 값을 구하면 성수는  $Q=10.5255(K=15)$ , 오류

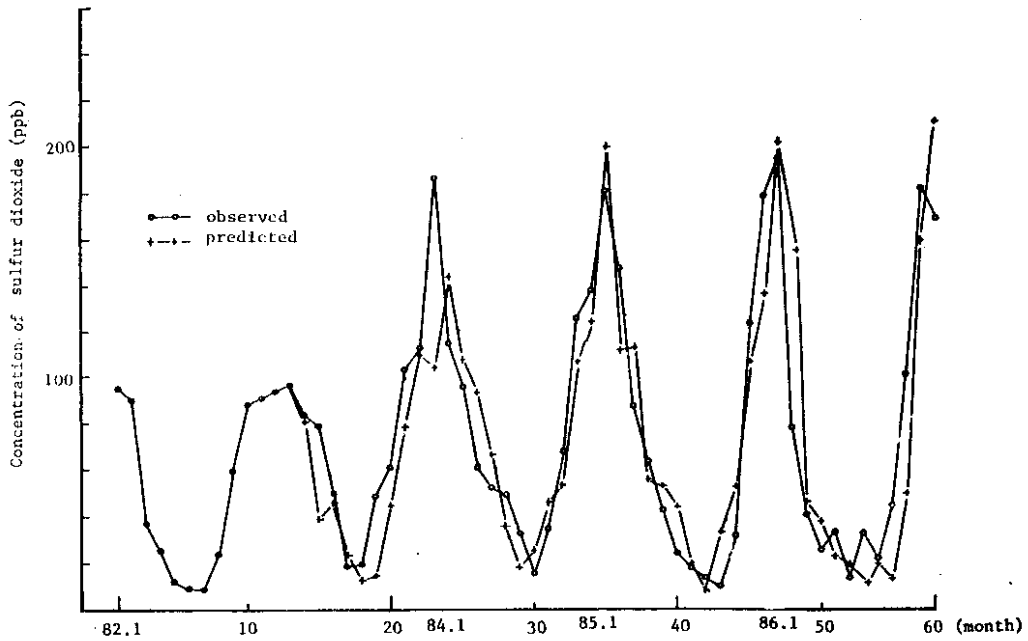


Fig. 12. Predicted and observed value of sulfur dioxide in Seungsoodong.

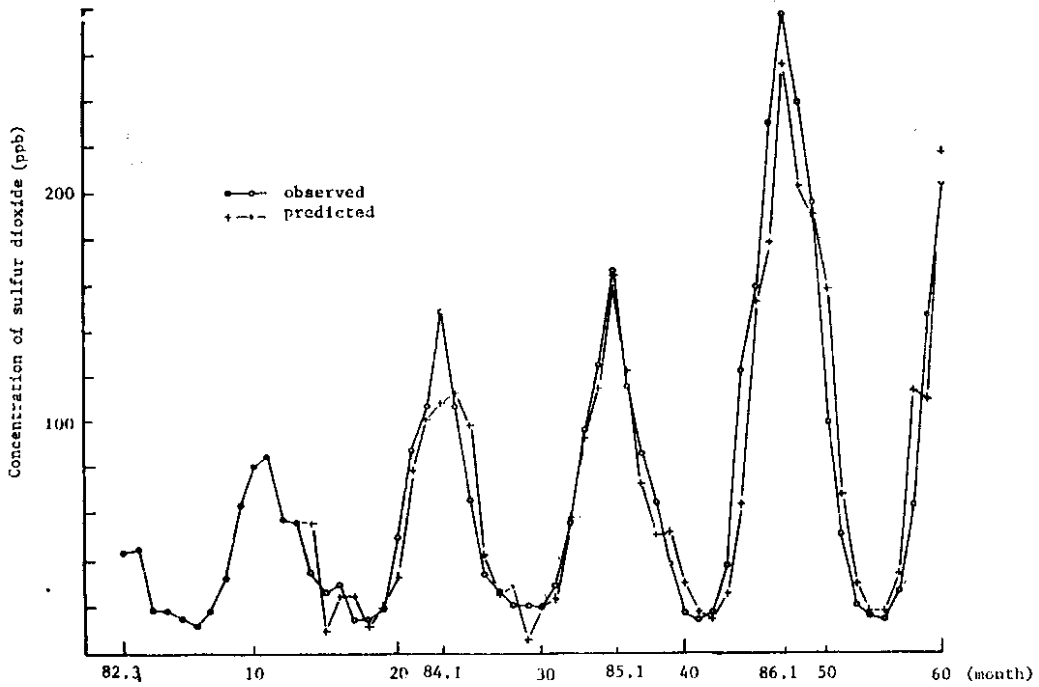


Fig. 13. Predicted and observed value of sulfur dioxide in Oryudong.

$Q=9.8397(K=15)$ 이며  $P<0.05$ 일때  $x^2=23.685$ 이므로 두 지점 residual time series도 모두 배경농도(white noise)이다. 따라서 확인된 모형과 추정된 계수는 모

두 적합함을 알 수 있었다.

추계치와 실측치와의 상관에 의한 적합도를 구하여 본 결과 Fig. 14, 15와 같이 실측치(X)와 추계치(Y)



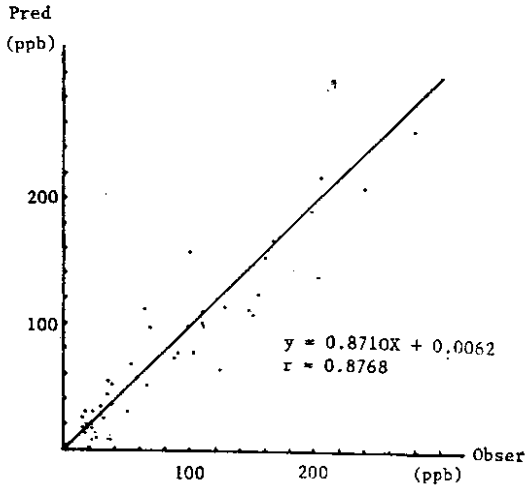


Fig. 14. Correlation between the observed and predicted value of sulfur dioxide concentration in Seoungsoodong.

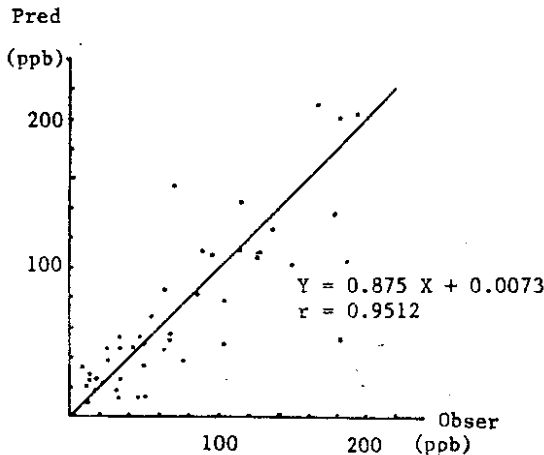


Fig. 15. Correlation between the observed and predicted value of sulfur dioxide concentration in Oryudong.

의 상관관계와 회귀방정식은 다음과 같다.

$$\text{성수} : Y = 0.8771X + 0.0062 \quad r = 0.8786$$

$$\text{오류} : Y = 0.8758X + 0.0073 \quad r = 0.9512$$

두 지점 모두 높은 상관성을 나타내고 있었다.

#### 5. 연구의 제한점

대기오염 관리 측면에서 현재의 대기질의 파악과 더불어 미래의 대기오염물질 배출원이 대기질에 어떠한 영향을 미칠 것인가를 알아야 할 필요가 있다. 이러한 확산현상을 규명하기 위하여 여러가지 방법이 있으나 가장 많이 이용되고 있는 방법으로는 simulation modeling 등이 사용되고 있으며 그 방법도 수십종에 달하

고 있다. 대기오염 확산모델을 이용한 추계를 위해서는 오염원에 대한 자료, 기후자료, 대기질의 자료가 정확해야 한다. 그러나 이들에 관한 상세하고도 정확한 자료를 확보한다는 것은 매우 어렵다. 따라서 modeling 시에 많은 단순화된 가정 및 추정을 하게 된다(정용, 1986).

시계열분석법의 기본 가정은 과거 시계열의 형태가 그 특성을 잃지 않고 미래에도 반복된다는 것이며 인과관계를 상세히 밝히기 어려운 자료의 단기예측에 가장 유용한 통계적 방법이다. 그러나 시계열분석법을 모든 종류의 자료에 대해서 사용할 수 없다. 시계열분석에 있어서 ARIMA모형이 갖는 제한점으로는

첫째, ARIMA모형을 통한 예측은 과거의 관행을 미래로 연장하여 하는 것으로 어떠한 현상이 발생하게 되는 인과관계에 대한 고려가 없기 때문에 예측결과를 해석하거나 이해하는데 어려움이 뒤 따른다.

둘째, 예측치의 시간단위가 짧을 수록 좋으나 결측치(missing value)가 많을시 이를 보완 수정하여 시계열자료로 이용할 수는 있지만 예측에 대한 신뢰도가 낮아지기 때문에 자료수집에 어려움이 있으며

셋째, 예측시 1단계(한단계 앞)까지는 추계방정식은 실측자료를 이용하므로 비교적 정확한 예측은 할 수 있으나 2단계 또는 그 이상의 추계는 추계치를 가지고 다시 추계를 하므로 신뢰도가 낮아지는 등 문제점을 내포하고 있어 단기추계에는 적합하나 장기추계에는 부적당하다. 장기추계는 대기오염 확산모델 등을 상호 보완적으로 사용하는 것이 바람직 하다고 사료된다.

#### IV. 결 론

시계열분석방법인 Box-Jenkins의 ARIMA model을 적용 서울특별시 일부지역(성수 및 오류동)의 대기오염도를 분석하였다. 1982년부터 1986년까지 5년간 측정된 아황산가스의 월 평균농도에 대하여 그 주기성과 변동상태에 대해서 자기상관관계(auto correlation function)와 편자기상관관계(partial autocorrelation function) 등을 분석하여 모형을 확인하고 추정된 후 적합성을 검토한 결과 다음과 같은 결론을 얻었다.

1. 조사대상 지역의 아황산가스에 의한 대기오염도는 월별로 주기성을 나타내었다. 기온이 낮은 겨울동안에 오염도가 높았고 여름에는 가장 낮았으며 연료사용시기와 그 오염도는 대체로 일치하였다.

2. 월별 대기질의 시계열을 분석한 바 12개월의 주기성을 가진 불안정 시계열임을 확인했으며 주기성을 제거하면 안정 시계열이 됨을 알았다.

3. ARIMA model( $p, d, q$ )( $P, D, Q$ )<sub>s</sub>의 모형에서(1, 0, 0)(1, 0, 1)<sub>12</sub>형태를 얻었다.

4. 모형방정식은 다음과 같이 얻었다.

$$\text{성수} : Y_t = 0.5214Y_{t-1} + Y_{t-12} - 0.5214Y_{t-13} + a_t$$

$$\text{오류} : Y_t = 0.8549Y_{t-1} + Y_{t-12} - 0.8549Y_{t-13} + a_t$$

5. 추계방정식의 적합도를 조사하기 위하여 실측치( $X$ )와 예측치( $Y$ )간의 상관관계 및 회귀방정식을 조사한 바

$$\text{성수} : Y = 0.8710X + 0.0062 \quad r = 0.8768$$

$$\text{오류} : Y = 0.8710X + 0.0073 \quad r = 0.9512$$

와 같이 두지점 모두 높은 상관성을 나타내고 있다.

이상의 결과로 보아 ARIMA model은 도시지역의 대기오염도의 변화를 모형화하는데 좋은 방법이며 오염도의 단기예측에도 적용이 가능하다. 그러나 장기예측은 확산모델 등을 상호보완적으로 사용하는 것이 바람직하다고 생각된다.

### 참 고 문 헌

- 권속표, 정용, 임동구. Hanna-Gifford Model에 의한 서울시 대기오염도 추정, 환경보전 협회지, 1980;1(1):21-38.
- 권속표, 환경대책과 자연보호. 연세대학교, 1985.
- 김광진 외 9명. 자동측정기에 의한 서울시대 SO<sub>2</sub> 오염도 조사. 서울특별시 보건환경연구소보, 1984; 20:294-313.
- 김미숙. 낙동강 수질변동에 대한 추계학적 연구. 한국과학기술연구소, 1983;9(1):15-30.
- 김양균 외 5명. 대기오염 확산모델 개발에 관한 연구. 대기보전학회지, 1985;1(1):53-70.
- 김정년. 통계학. 경문사, 1984.
- 김재영. 시계열 모형에 의한 건설활동 예측. 국토연구, 1987;7:85-110.
- 신응배 외 8명. 전국 지역별 대기관리에 관한 연구. 한국과학기술연구소, 1980.
- 안령미, 하철의 일별 수온차에 대한 추계학적 특성, 서울대학교, 1984.
- 이민희 외 3명. 대기오염 측정망 관리에 대하여. 서울특별시 보건환경연구소, 1978;14:85-97.
- 이상훈. 수자원 연구에 응용되는 시계열분석. 건설기술정보, 1987;42:1-4.
- 이종남, 이홍근. 섬진강 월유출량의 추계학적 모형. 한국수문학회, 1984;17(4):281-291.
- 이홍근, 한강 수질오염의 추계학적인 변동에 대한 연구. 한국보건협회지, 1977;3(11):43-51.
- 이홍근. 한강 하류부의 수질변동에 대한 추계학적 특성. 서울대학교, 1982.
- 정용, 강제연. 대기오염 모델의 적합도에 대한 연구. 한국대기보전학회지, 1986;2(1):81-90.
- Box GEP, Time series analysis forecasting and control. Holden-day 1970.
- Box GEP, Tiao GC. Intervention analysis with application to economic and environmental problems. Journal of the American Statistical Association, 1975;70(349):70-79.
- Franklin SL, Maidment DR. An evaluation of weakly and monthly time series forecasts of municipal water use. Water Resources Bulletin, 1986;22(4):611-621.
- Glass GV. Estimating the effects of intervention into a nonstationary time series. American Educational Research Journal, 1972;9(3):463-477.
- Gottman JM. Time series analysis. Cambridge University Press, 1981.
- Kotteyada NT. Stochastic water resources technology. The Macmillan Press, 1980.
- Medowall D, McCleary R, Medidinger EC. Interrupted time series analysis. A sage University Papers, 1981.
- McLeod AI, Hipel KW, Lennox WC. Advance in box-jenkins Modeling Z. Application. Water Resource Res. 1977;13(3):577-586.
- Mehta BM, Ahlert RC. Stochastic variation of water quality of the Passaic River. Water Resour. Res., 1975;11(2):300-318.
- Panodtsh HA, Brier GW. Some application of statistics to meteorology. The Pennsylvania state University, 1968.
- Stern AC. Air pollution 3rd ed, Vol. II. Academic Press, 1977.
- Sim CH, A mixed gamma ARMA(1.1) model for river flow time series. Water Resources Research, 1987;23(1):32-36.
- Tiao GC, Box GEP. Analysis of Los Angeles photo Chemical smog date. Journal of the air pollution Contral Association, 1975;25(3):260-268.
- U.S. EPA DAQPS. Guideline series guideline on air quality model EPA-450/22-78-027, 1978.