

선형 추계학적 모형을 이용한 PM₁₀ 농도 예측

환경조사팀

안미진 · 여인학 · 임귀철 · 김길식 · 고한성 · 신미영
이상훈 · 김두래 · 고현욱 · 허미숙 · 김홍기 · 이민환

Forecasting of PM₁₀ Concentration Using Linear Stochastic Model

Environmental Research Team

**Mi-jin Ahn, In-hak Yoe, Gui-cheul Lim, Gil-sik Kim, Han-sung Ko, Mi-young Shin,
Sang-hoon Lee, Du-rae Kim, Hyen-ug Ko, Mi-suk Heo, Hong-gi Kim and Min-hwan Lee**

Abstract

This study was carried out to analyze temporal characteristics of PM₁₀ measured at an official monitoring station neighboring traffic intersection in Sinsa-Dong, Seoul from Oct. 10, 1999 to Jan. 29, 2000 and to construct Auto Regressive Moving Average eXogenous Model. Meteorological variables were used to improve forecasting accuracy. The results were as followed.

1. Time series of PM₁₀ had the auto correlated structure including order less than 6 of auto regressive function and smaller order of moving average function.
2. The lag time between PM₁₀ and wind speed or relative humidity was 5 or 2 hour, thus it was estimated that wind speed and relative humidity would be good preceding indices.
3. It was possible to forecast hourly concentrations of PM₁₀ within 0.2 percent error in the case of 1 step ahead forecasting.
4. When wind speed was considered as input variable, forecasting accuracy was improved, but relative humidity didn't affect improvement of forecasting.

서론

방지시설 설치 강화, 청정연료의 보급 등의 대기질 개선 노력으로 인해 1990년대에 들어 서울시의 대기질은 아황산가스 등의 1차 대기오염물질은 전반적으로 저감되었으나, 미세먼지, 오존 등의 2차 대기오염물질은 오히려 증가추세에 있다. 오존은 여름철의 주 오염물질이

나, 미세먼지의 경우 전 계절에 걸쳐 고농도 현상이 빈번하다. 현재 서울시는 미세먼지에 대해 10 μ m이하의 공기역학적 직경을 가진 미세입자로 정의되는 PM₁₀을 상시 측정 감시하고 있다. PM₁₀은 호흡성 입자 크기로 공기 중에 부유 상태로 존재하면서 호흡에 의해 인체에 침입하여 기관지 및 폐에 부착되기 쉽기 때문에 호흡기 질환 등 인체에 악영향을 초래할 수 있으며 또한 빛을 흡

수, 산란시키기 때문에 시정 악화에 영향을 미친다.

최근 들어 대기오염 저감 대책과 더불어 장, 단기 대기오염 예측의 중요성이 인식됨에 따라 대기질 변화를 분석하고 예측하기 위한 연구가 많이 진행되어 왔다. 대기질 예측에 관한 국내 연구 사례에는 다변량 통계분석을 이용한 오존 예측(허, 1993)¹⁾, 오존 예측을 위한 신경망모델 개발(김, 1994)²⁾, 다원ARIMA 모형을 통한 오존 예측(박, 1997)³⁾ 등이 있는데 현재 이런 모델 대부분이 여름철 오존 예측에 대한 것이며, 다른 오염물(겨울철에 높은 농도를 가지는)에 대한 것은 아직 많지 않다. 외국의 경우 역시 오존 예측에 대한 것이 주류이나 신경망모형을 사용한 PM_{2.5} 예측(Patricio, 2000)⁴⁾이나 다변량 통계모형을 사용한 CO 예측(Prateek, 2000, 2001)^{5),6)} 등의 연구 사례가 있다.

현재 서울시에는 다량의 대기질 자료가 구축되어 있으므로 이를 바탕으로 통계모형을 구축하여 시간에 따른 대기질 변화를 분석하고 예측하는 것이 가능할 것이다. 따라서 본 연구에서는 과거의 PM₁₀ 및 기상 자료를 바탕으로 다변량 자기회귀 이동평균 모형(ARMAX)을 이용하여 PM₁₀의 시간적 거동 특성을 분석하였다.

연구 방법

1. 모형의 이론

연구에 사용된 모형은 Box와 Jenkins(1970)⁷⁾에 의해 정립된 다변량 자기회귀 이동평균 모형(Auto Regressive Moving Average eXogenous Model, ARMAX Model)으로서, 이 모형은 ARMA모형에 중회귀 모형의 특성을 결합시킨 선형 추계학적 예측방법으로써 선행지수인 투입변수와 종속변수간의 동적 관련성을 분석하여 모형의 예측 값을 개선시키고자 하는데 목적이 있다. 기본적인 입, 출력 시계열을 각각 $u(t)$, $y(t)$ 라 할 때, 모형의 일반적인 구조는 아래와 같다.

$$A(q)y(t) = B_1(q)u_1(t-nk_1) + \dots + B_n(q)u_n(t-nk_n) + C(q)e(t) \quad \text{식(1)}$$

여기서 $e(t)$ 는 분산 λ 인 백색잡음이며, q 는 이동(shift) 연산자이고, $A(q)$, $B(q)$, $C(q)$ 는 아래와 같이 지연 연산자(lag operator) q^{-1} 로 나타내어지는 다항식이다.

$$A(q) = 1 + a_1q^{-1} + \dots + a_nq^{-na}$$

$$\begin{aligned} B(q) &= b_1 + b_2q^{-1} + \dots + b_{nb}q^{-nb+1} \\ C(q) &= 1 + c_2q^{-1} + \dots + c_nq^{-nc} \end{aligned} \quad \text{식(2)}$$

여기서, na 와 nb , nc 는 각각 다항식의 차수를 나타내며, nk 는 입력에서 출력까지의 지연을 나타낸다.

2. 대상기간 및 자료

본 연구에서는 황사현상과 장마기간의 세정효과로 인해 외부 간섭효과가 큰 기간을 제외하고 비교적 먼지 오염도가 높은 가을~겨울을 중심으로 1999년 10월 10일 01시부터 2000년 1월 29일 24시까지의 자료를 분석하였다. 자료는 서울시 도로변 측정망 중 하나로 자동차 오염물질에 의한 대기오염 영향과역 및 자동차 관리대책의 기초자료로 활용하고자 신사 교차로에 설치, 운영되고 있는 신사지점의 PM₁₀(측정기기 : Mcsam II model-331 β , Kimoto, Japan, 절단경 10 μ m)농도의 시간 자료이다. 설명변수로서 위 기간 동안 동지점에서 측정된 풍속과 습도를 설명변수로 고려하여 분석하였다. 위의 대상기간으로부터 선정된 PM₁₀ 자료는 평균농도는 55 μ m/ m^3 , 최대농도 353 μ m/ m^3 , 표준편차 36 μ m/ m^3 였으며 자료의 개수는 2688points이고 이중 결측치는 단순내삽하여 보정하였다.

결과 및 고찰

1. PM₁₀ 자료 분석

PM₁₀의 시도표(그림.1)를 관찰한 결과 일정한 주기 패턴이 보여지는 면이 있으나, 기타 다른 영향 인자들로 인하여 시계열의 교란이 상당부분 포함되어 있는 것으로 보여졌다. 따라서 본 연구에서는 시계열의 거동이 주기

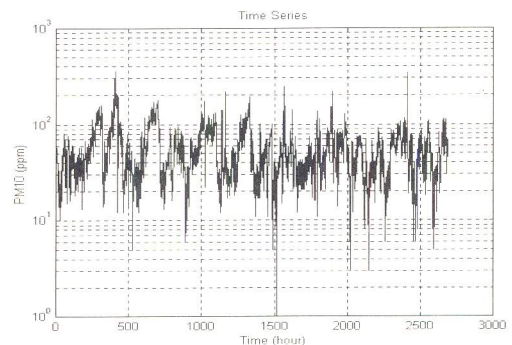


Fig. 1. Time series of PM₁₀ from Oct. 10, 1999 to Jan. 29, 2000

를 중심으로 진행한다기보다는 전체 평균을 기준으로 확률 보행하는 것으로 가정하고 이후 분석을 수행하였다.

자기상관함수(그림.2)에서 알 수 있듯이 PM₁₀의 경우 24시간 이상의 선행 자료들과 상당한 크기의 상관을 맺고 있는 것으로 보여지며, 지체 100시간 내에서는 특별한 주기성분을 찾아볼 수 없는 것으로 판단되었다. 편자기상관함수의 도시 결과는 지체 5시간 이내에서 절단되는 형태를 보여주고 있으며, 이는 비교적 작은 차수의 자기회귀모형이 적절하다는 것을 보여주는 것으로 판단되었다. 이러한 결과를 지수함수적으로 감소하고 있는 자기상관함수의 결과와 함께 비교하여 볼 때, 일정 차수의 자기회귀함은 보유하고 있으나, 이동평균함은 크게 나타나고 있지 않음을 알 수 있었다.

PM₁₀ 시간농도의 일일변동특성을 분석한 결과 오전 11시에 일 최고 농도를 기록하고 있으며 오후의 침투 농도는 특별하게 나타나지 않았다.(그림.3) 이는 출근시간 자동차의 집중에 의한 것으로 판단되었다. 자동차 배출

이 PM₁₀ 농도에 기여하므로 PM₁₀ 농도가 주말에 낮고 주중에 높을 것으로 판단되나 요일별로 그 차이가 크지 않아 모든 요일을 동일하게 처리하였다.

2. 기상 자료 분석

PM₁₀은 풍속과는 음의 상관을 맺고 있으며, 상대습도와는 양의 상관을 갖고 있다(그림4). 풍속 및 상대습도 모두 PM₁₀과의 최대상관정도가 0 지체에서 발생하는 것이 아니라 각각 5시간과 2시간 지체 후 발생하는 것으로 나타났으며, 이는 풍속 및 상대습도가 PM₁₀ 예측의 좋은 선행지수가 될 수 있음을 나타낸다.

풍속과 상대습도의 자기상관함수에서 볼 수 있듯이 풍속과 상대습도의 24시간 주기특성이 비교적 명확히 나타나고 있음에 따라 PM₁₀과는 달리 24시간을 주기로 진동하며 감소하는 특성을 보여주었다. 이에 따라 풍속이나 상대습도를 PM₁₀ 예측의 중요한 선행인자로 사용할 경우 자료의 주기특성을 고려하여야 할 것으로 판단되나, 본 논문에서는 보다 단순한 적용을 위하여 풍속이나 상대습도의 경우도 PM₁₀과 마찬가지로 주기성분을 고려하지 않고 이후 과정을 수행하였다. 편자기상관함수의 도시 결과는 PM₁₀과 마찬가지로 풍속, 상대 습도 자료 모두 지체 5시간 이내에서 절단되는 형태를 보여주었다.

3. 예측 모형의 적용

전술한 PM₁₀과 기상자료의 특성을 바탕으로 PM₁₀을 단일변수로 하는 단변량 자기회귀 이동평균 모형과 풍속과 습도를 각각 혹은 모두 설명변수로 추가한 다변량 자기회귀 이동평균 모형을 구축하였다. 구축된 모형에 대

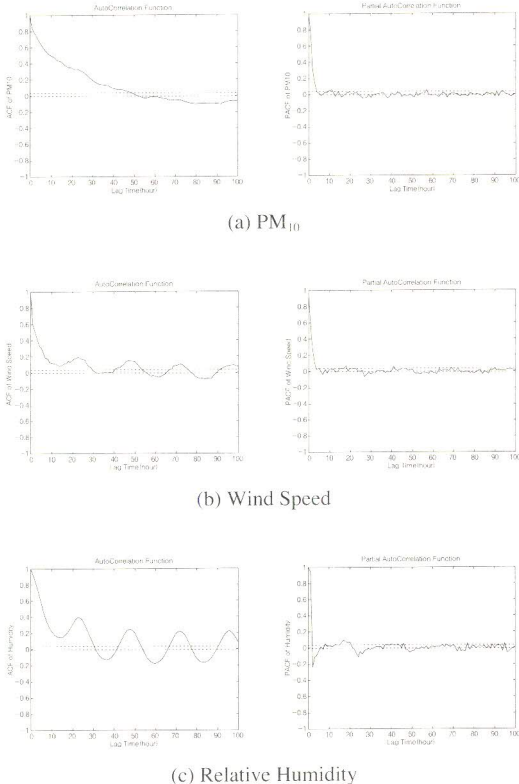


Fig. 2. Autocorrelation & Partial Autocorrelation

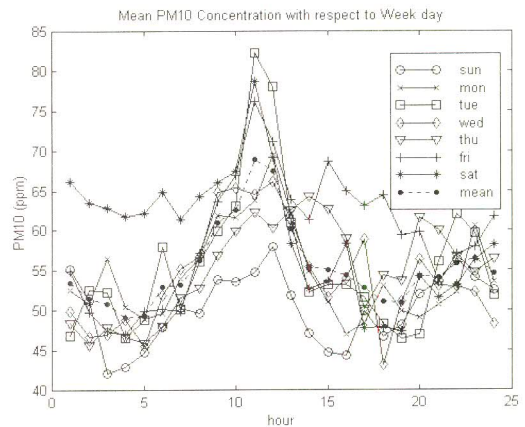
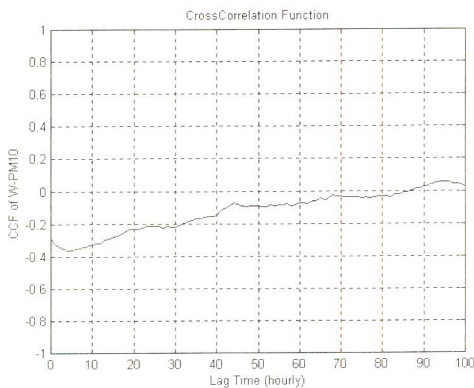


Fig. 3. Average hourly variation of PM₁₀ concentrations

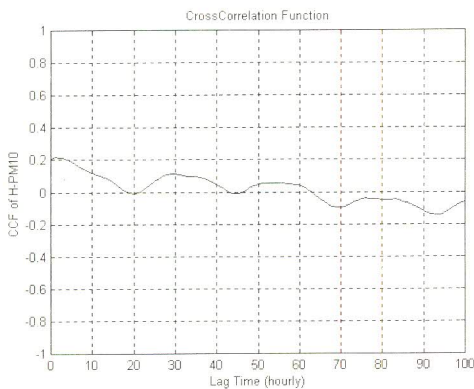
Table 1. Model Descriptions

Model Name	Model description	Estimated Parameters for 1 step ahead forecasting				
Model I	P Univariate	$a_1 = -0.9376$	$a_2 = 0.0420$	$a_3 = -0.0435$		
	ARMA(3,1) Model	$c_1 = -0.3465$				
Model II	W-P Multivariate	$a_1 = -0.9958$	$a_2 = 0.0876$	$a_3 = -0.0152$		
		$b_1 = -0.0136$	$b_2 = -0.0324$	$b_3 = -0.0199$		
	ARMAX(3,3,1) Model	$c_1 = -0.4494$				
Model III	H-P Multivariate	$a_1 = -0.8705$	$a_2 = 0.0028$	$a_3 = -0.0589$		
		$b_1 = 0.0954$	$b_2 = 0.0012$	$b_3 = -0.0936$	$b_4 = 0.0469$	$b_5 = -0.0273$
	ARMAX(3,5,1) Model	$c_1 = -0.2937$				
Model IV	W&H-P Multivariate	$a_1 = -1.0483$	$a_2 = 0.1127$	$a_3 = 0.0012$		
		$b_1 = -0.0291$	$b_2 = 0.0029$	$b_3 = -0.0223$		
	ARMAX(3,3,5,1) Model	$b'_1 = 0.1070$	$b'_2 = -0.1600$	$b'_3 = 0.0863$	$b'_4 = -0.0092$	$b'_5 = -0.0266$
		$c_1 = -0.4889$				

* Index : P = PM₁₀, W = Wind Speed, H = Relative Humidity



(a) Wind Speed - PM₁₀



(b) Relative Humidity - PM₁₀

Fig. 4. Cross Correlation

한 차수와 1단계 전방 예측 시 추정된 매개변수를 표1에 나타내었다. 여기에서 Model II의 차수는 (3,3,1)인데 이는 식(2)의 A, B, C의 차수를 의미한다.

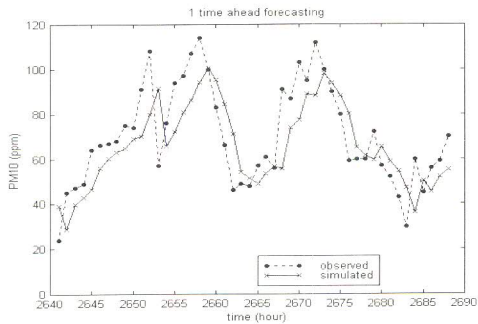
구축된 모형으로 1단계 전방 예측시의 결과를 그림 5에 나타내었다. 고농도 영역이 다소 과소평가 되는 경향이 있는 것으로 판단되었다.

모형의 예측정도를 평가하기 위해 다음과 같은 예측 오차를 사용하였다.

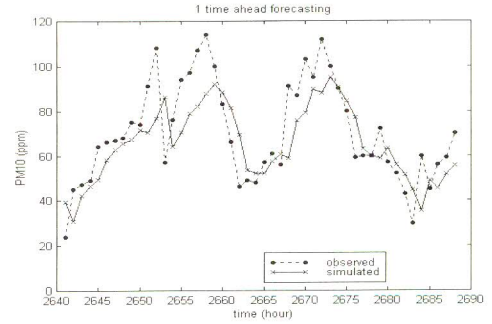
$$\text{Forecasting Percent Error} = \frac{1}{N} \sum_{t=0}^N \frac{|y(t) - y_p(y)|}{y(t)} \quad \text{식(3)}$$

여기서, N은 예측오차를 계산하는 데 사용된 자료의 개수이고, $y(t)$ 는 시간 t 에서의 관측값, y_p 는 시간 t 에서의 예측값이다.

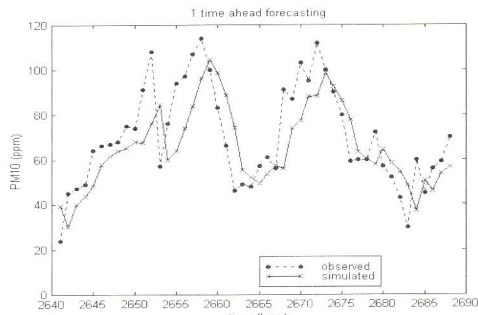
전방 예측시간 1시간에서 12시간까지의 예측오차에 대한 결과를 그림 6에 나타내었다. 그림에서 볼 수 있는 바와 같이 대체로 6시간 전방 예측 이후로는 전방 예측 시간이 증가함에 따라 예측오차가 점차적으로 수렴하는 양상을 나타냈으며, 이는 6시간 이상의 예측은 평균값 정도의 예측 값만을 주는 것으로 판단된다. 따라서 합리적인 전방예측시간은 5시간 이내인 것으로 판단된다. 1시간 전방 예측의 경우 모형 I의 예측오차는 0.20, 모형 II는 0.18, 모형 III은 0.20, 모형 IV는 0.18이며, 이로부터 기상자료로서 풍속을 고려하였을 때 모형의 예측결과가 향상되고 있음을 알 수 있다. 전체적인 예측오차는 0.18~0.35였다.



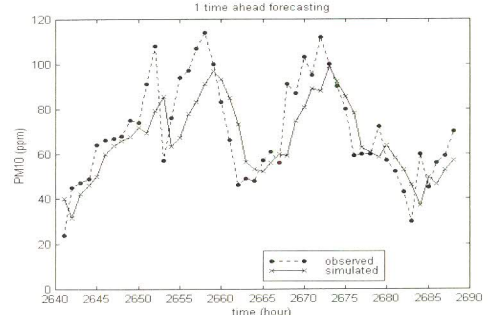
(a) Model I



(b) Model II

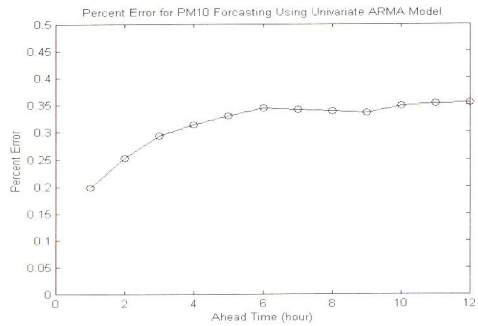


(c) Model III

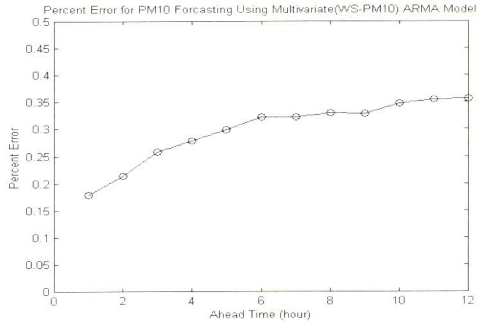


(d) Model IV

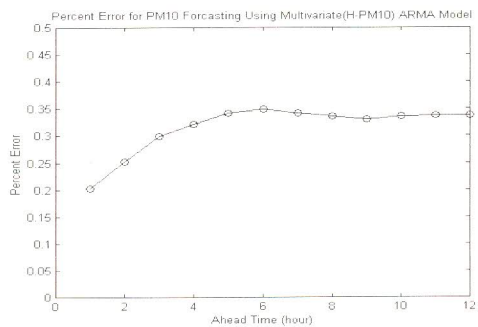
Fig. 5. One step ahead forecasting results



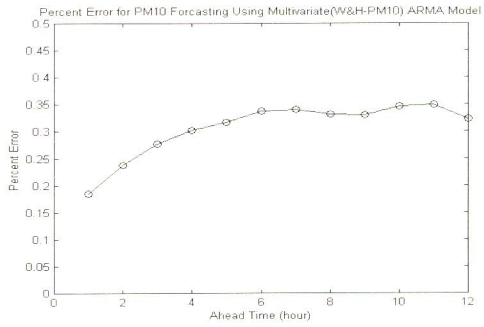
(a) Model I



(b) Model II



(c) Model III



(d) Model IV

Fig. 6. Forecasting Percent Error

결 론

본 연구에서는 1999년 10월 10일부터 2000년 1월 29일까지 서울시 신사동의 도로변 측정지점에서 관측된 PM₁₀자료의 시간적인 거동 특성을 분석하고자 적정 시계열 모형을 구축하였다. 또한 모형의 예측정도를 향상시키기 위하여 기상자료를 설명변수로 한 다변량 자기회귀 이동평균 모형(ARMAX)을 구축하여 PM₁₀의 시간농도예측시 기상자료의 사용가능성을 분석하였다. 본 연구 결과를 요약하면 다음과 같다.

1. PM₁₀ 시간농도의 경우 6차이내의 자기 회귀항을 갖는 것으로 보이며, 이동평균항은 크게 나타나지 않는 자기상관구조를 갖고 있었다.
2. PM₁₀과 풍속, PM₁₀과 상대습도는 각각 5시간과 2시간의 선행 시차를 가졌으며, 이로부터 풍속과 상대습도는 PM₁₀ 예측의 좋은 선행지수가 될 수 있을 것으로 판단되었다.
3. 자기회귀 이동평균 모형의 적용결과 1시간 선행 예측의 경우 0.2이내의 예측오차로 예측 가능함을 알 수 있었다.
4. 풍속을 설명변수로 고려하였을 때 예측오차는 0.18로 예측 정도 향상에 기여하였으나, 상대습도는 거의 역할을 하고 있지 않는 것으로 판단되었다.
향후 연구 시에는 본 연구에서 고려되지 않는 주기성분에 대한 고려를 모형구조에 포함하여 보다 더 일반적인 모형의 구축이 필요할 것으로 판단되며, 설명변수로서 기상자료 이외에 배출량자료가 추가된다면 보다 더 정도 높은 예측결과를 산출할 것으로 기대된다.

참 고 문 헌

1. 허정숙, 김동술: 다변량 통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구. 한국대기보전학회지, 9(3): 207-215. (1993)
2. 김용국, 이종범: 하계의 일 최고 오존농도 예측을 위한 신경망모델의 개발. 한국대기보전학회지, 10(4): 224-232. (1994)
3. 박진수, 김신도, 조용준: Wavelet transform을 이용한 오존특성 분석 및 다원 ARIMA모형을 통한 예측. 1997년도 한국대기보전학회 추계학술대회 요지집. 89-90 (1997)
4. Patricio, P., Alex, T., and Jorge, R.: Prediction of PM_{2.5} concentrations several hours in advance using neural networks in Santiago, Chile. Atmospheric Environment, 34: 1189-1196. (2000)
5. Prateek, S., and Mukesh, K.: Real time prediction of extreme ambient CO concentrations due to vehicular exhaust emissions using univariate linear stochastic models. Transportation Research Part D, 5: 59-69. (2000)
6. Prateek S., and Mukesh, K.: Short term, real-time prediction of extreme ambient CO concentrations due to vehicular exhaust emissions using transfer function-noise model. Transportation Research Part D, 6: 141-146. (2001)
7. Box, G. E. P. and Jenkins, G. M.: Time series analysis-forecasting and control. Holden Day. (1970)