

자기회귀오차모형을 이용한 평택시 PM10 농도 분석

Analysis of PM10 Concentration using Auto-Regressive Error Model at Pyeongtaek City in Korea

이 훈 자*

평택대학교 디지털응용정보학과

(2010년 9월 6일 접수, 2011년 5월 6일 수정, 2011년 5월 30일 채택)

Hoonja Lee*

Department of Information Statistics, Pyeongtaek University

(Received 6 September 2010, revised 6 May 2011, accepted 30 May 2011)

Abstract

The purpose of this study was to analyze the monthly and seasonal PM10 data using the Autoregressive Error (ARE) model at the southern part of the Gyeonggi-Do, Pyeongtaek monitoring site in Korea. In the ARE model, six meteorological variables and four pollution variables are used as the explanatory variables. The six meteorological variables are daily maximum temperature, wind speed, amount of cloud, relative humidity, rainfall, and global radiation. The four air pollution variables are sulfur dioxide (SO₂), nitrogen dioxide (NO₂), carbon monoxide (CO), and ozone (O₃). The result shows that monthly ARE models explained about 17~49% of the PM10 concentration. However, the ARE model could be improved if we add the more explanatory variables in the model.

Key words : Autoregressive error (ARE) model, Explanatory variables, Meteorological variables, Pollution variables, PM10 concentration, Pyeongtaek City

1. 서 론

세계 많은 나라들이 급속한 산업성장 등으로 인하여 대기오염 문제를 갖고 있다. 특히, 대기 중의 높은 미세먼지, PM10 (Particulate Matter with a diameter less than 10 μm)은 인간의 건강에 심각한 피해를 주는 것으로 알려지고 있다. 우리나라 환경부에서는 2007년부터 PM10, 이산화질소(NO₂) 등의 대기환경 기준을

강화하였으며, PM10의 경우 연평균 70 μg/m³에서 50 μg/m³으로, 일평균 기준은 150 μg/m³에서 100 μg/m³으로 개정하여, PM10 저감도 정책을 강화하였다. 이러한 PM10 저감 정책이 효과적으로 진행되기 위해서는 PM10 농도변화 추이와 특성을 이해하고, 그 주요 원인을 파악하는 종합적인 연구가 뒷받침되어야 한다(Kim, 2006).

지금까지 국내외의 PM10 모형에 관한 연구는 주로 회귀모형, 신경망모형 등을 주로 많이 사용하였고, 박스모델이나 SVR (support vector regression) 등의 새로운 모형도 연구되었다.

*Corresponding author.

Tel : +82-(0)31-652-8710, E-mail : esther@ptu.ac.kr

PM10 농도의 모형에 관한 국내연구의 경향을 살펴보면, SVR (support vector regression) 모형을 이용하여 포항지역 미세먼지의 통계적 예보 모형을 연구하여, 회귀모형과 신경망모형을 비교하였고(Lee *et al.*, 2006), 회귀모형, 신경망모형, 의사결정모형을 이용하여 미세먼지 예보시스템을 개발하였고(Koo *et al.*, 2010), 박스모형을 이용하여 NO₂, SO₂, PM10 등의 대기오염의 추세(Kim and Joh, 2005) 등의 방법으로 PM10 농도모형에 관해 연구하였다.

국외의 대기오염에 관한 모형을 보면, 칠레 Santiago 지역의 PM10 농도 예측을 위해 신경망모형을 사용하였고(Perez and Reys, 2006), 이태리의 Palermo 지역의 PM10, SO₂, O₃ 등의 예측을 위해 신경망모형을 사용하였고(Brunelli *et al.*, 2007), 미국 Kentucky metropolitan 지역의 PM2.5의 예측을 위해 비선형 회귀분석을 이용하였다(Cobourn, 2010).

기상요인이 PM10 농도에 많은 영향을 준다는 사실이 국내외적으로 알려져 왔다. 국내에서는, PM10 농도를 지배하는 요인 중에는 배출원을 제외하고는 가장 영향을 주는 요소는 기상변수이며, 특히 풍향, 풍속, 기온, 습도 등의 상태에 따라 농도가 크게 좌우된다(Chung and Chung, 1991). 또한 미국 남서부지역의 1990~2003년 기간 동안 PM10 농도를 연구한 Wise and Comrie (2005)는 상대습도와 혼합고를 비롯한 일 최고온도, 이슬점온도, 지상 풍속을 모두 고려할 경우 전체 미세먼지 농도의 49%를 설명할 수 있다고 보고한 바 있다.

본 연구에서는, 경기 남부지역인 평택시 PM10 농도를 2003년 1월 1일부터 2009년의 12월 31일까지의 7년간 자료를 통계모형 중 시계열 모형에 적용하고자 한다.

지금까지 PM10 농도에 관한 모형으로 많이 사용된 방법은 회귀모형과 신경망모형이다. 그러나 PM10 농도가 시간에 따라 영향을 받기 때문에 회귀모형을 사용하면 오차항이 독립이어야 한다는 조건에 모순되며, 신경망 모형은 다른 모형에 비해 정확한 예측을 하는 특성이 있지만 결과를 해석하기 힘든 단점이 있다. 본 연구에서는 PM10 농도 같은 시계열자료에 적합한 새로운 통계모형인 자기회귀오차모형(ARE, Auto-Regressive Error Model)을 사용하였다. 자기회귀오차모형은 회귀분석과 같이 설명변수를 분석에 사용할 수 있는 모형이고, 어떤 설명변수들이 PM10

에 영향을 주는지 등의 분석 결과를 비교적 쉽게 해석하여 이로부터 유용한 정보를 얻을 수 있다(조신섭과 이정형, 1977 참고).

정확한 PM10 농도의 분석과 예측을 위해서는 PM10과 연관된 대기 및 기상자료를 포함시켜야 한다. 본 논문에서는 PM10과 연관이 있고, 기상청과 국립환경과학원에서 쉽게 자료를 구할 수 있는 6종류의 기상자료와 4종류의 대기자료를 설명변수로 포함한 다변량 분석인 ARE(Autoregressive Error, 자기회귀오차) 모형으로 분석하였다.

2. PM10 자료 및 설명변수 자료

PM10 자료는 국립환경과학원의 평택시 비전동 관측소의 시간별 자료(하루 24번)를 이용했고 분석을 위하여 일 최고 PM10 농도를 사용했다. 기간은 2003년부터 2009년까지의 1월 1일부터 12월 31일까지 7년간 자료 중 2003년부터 2007년까지의 5년간 자료를 모델링기간으로 사용했고, 2008년과 2009년의 2년간 자료를 검증기간으로 사용하여 월별로 분석하였다. 본 연구에서 1월에서 12월까지 모든 자료를 사용한 이유는 월별로 PM10 농도가 차이가 있기에 월별분석이 중요한 것으로 생각되었기 때문이다.

설명변수로 사용한 기상자료와 대기자료는 PM10과 연관이 있으면서, 쉽게 구할 수 있는 변수를 사용하였다. 기상자료는 시간별 자료로 평택에서 가장 가까운 곳에 위치한 수원시 기상청자료를 사용하였다.

기상자료도 PM10과 같이 2003~2009년 자료를 사용하였다. 6종류의 기상자료는 Lee *et al.* (2008)과 Wise and Comrie (2005)을 참고로 하여 일 최고온도(maximum temperature), 풍속(wind speed), 상대습도(relative humidity), 강수량(rainfall), 일사량(radiation), 운량(amount of cloud)을 사용했으며 표 1에 나타나 있다.

풍향(wind direction)은 PM10에 영향을 주지만 정량화의 어려움이 있어 설명변수에 포함하지 않았다. 설명변수에 6종류의 기상자료를 포함시킨 이유는, 서론부분에서 언급했듯이, Wise and Comrie (2005)의 연구에서 상대습도와 혼합고를 비롯한 일 최고온도, 이슬점온도, 지상 풍속을 모두 고려할 경우 전체 미세먼지 농도의 49%를 설명할 수 있다고 보고했고, 본

Table 1. List of pollution variables and meteorological variables.

Variable	Contents and Timing	Unit	Notation
PM10	Previous day maximum 1 hour PM10	$\mu\text{g}/\text{m}^3$	Y_{t-1}
SO ₂	Previous day maximum 1 hour SO ₂	ppb	S_{t-1}
NO ₂	Previous day maximum 1 hour NO ₂	ppb	N_{t-1}
CO	Previous day maximum 1 hour CO	0.1 ppm	I_{t-1}
O ₃	Previous day maximum 1 hour O ₃	ppb	O_{t-1}
Temperature	Daily maximum 1 hour surface temperature	°C	T_t
Wind	Average wind speed (from 9:00 to 15:00)	m/s	W_t
Cloud	Average cloud amount (from 9:00 to 15:00)	-	C_t
Humidity	Average relative humidity (from 9:00 to 13:00)	%	H_t
Rainfall	Sum of precipitation (from 1:00 to 24:00)	mm	A_t
Radiation	Sum of radiation (from 9:00 to 15:00)	MJ/M ²	R_t

평택 자료에서도 월별로 차이는 있지만 PM10과 상관관계가 높은 기상자료가 있기 때문이다. 예를 들면, 1월에는 PM10과 일 최고온도 상관관계 계수가 0.39이고, 10월에는 풍속과의 상관관계 계수가 -0.38로 나타났다.

4종류의 대기자료로는 이산화황(SO₂), 이산화질소(NO₂), 일산화탄소(CO), 오존(O₃)를 사용했고, PM10 자료와 같이 국립환경과학원의 평택시 관측소를 이용하여, 2003~2009년의 1월 1일~12월 31일의 시간별 자료를 이용하였다. 사용된 대기자료도 표 1에 나타나 있다. 설명변수에 4종류의 대기자료를 포함한 이유는 PM10과 상관관계가 높은 대기자료가 있기 때문이다. 예를 들면, 6월에는 PM10과 SO₂가 상관관계 계수가 0.43이고, 9월에는 O₃와 0.44이고, 특히 10월에는 NO₂와 0.63으로 매우 높게 나타났다.

6종류의 기상자료와 4종류의 대기자료의 시간별 선택과 차수 선택은 Hubbard (1998)를 참고로 하여 각 자료를 시간별 혹은 일별로 PM10 농도와의 상관관계가 가장 높은 시간대를 사용하였다. 예를 들면, 운량(cloud) 변수는 PM10 농도와 상관관계가 높은 일별 9~15시의 평균을 사용하였다. 각 설명변수의 구체적인 시간대는 표 1에 나타나 있다. 표 1의 Notation은 4장의 ARE모형 변수표시이다.

3. PM10 농도의 ARE 모형과 적합성

반응변수에 영향을 주는 설명변수가 있을 때는 설명변수를 분석에 이용하는 것이 효율적이다. ARE 모형은 설명변수를 분석에 사용할 수 있는 모형으로

시계열자료를 회귀모형에 적합시킬 때 적합한 모형이다. 반응변수인 PM10에 영향을 주는 NO₂농도나, 일 최고온도 같은 설명변수를 사용하여 자료를 분석한다.

ARE 모형은 설명변수를 분석에 사용할 수 있는 모형으로 시계열자료를 회귀모형에 적합시킬 때 적절한 모형이다.

일반적인 k차 ARE 모형은 다음과 같다.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp} + \varepsilon_t$$

$$t = 1, 2, \dots, n, \text{이며}$$

$$\varepsilon_t = e_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} \dots - \phi_k \varepsilon_{t-k} \quad (1)$$

이며, e_t 는 서로 독립이며, $N(0, \sigma^2)$ 인 정규분포를 갖는다고 가정한다.

식 (1)에서 오차항 ε_t 는 서로 상관관계가 있고 AR(k) 형태를 갖는다. 또한 식(1)에서 Y_t 는 반응변수인 PM10 농도를 나타내며, X_{tj} , $j=1, 2, \dots, p$ 는 Y_t 에 영향을 주는 NO₂, 일 최고온도 등의 설명변수들을 나타낸다. β_j 는 설명변수 X_{tj} 의 계수로 SAS 통계 패키지를 이용하여 구한다. 위의 ARE 모형과 회귀모형의 차이점은 ARE 모형은 식(1)에서 오차항 ε_t 가 서로 독립이 아니라 시차에 따른 자기 상관을 갖고 있으며 특히 AR (Autoregressive) (k) 모형의 형태를 갖는다는 점이다. 따라서 식(1)을 ARE (자기회귀오차) 모형이라 부르며 시계열자료에 회귀모형을 적합시킬 경우 많이 사용된다. 오차항의 계수 ϕ_i , $i=1, 2, \dots, k$ 도 SAS 통계패키지를 이용하여 구한다.

ARE 모형들의 적합성을 측정하는 데 있어 결정계수 R^2 와 잔차 분석의 RMSE를 사용하였다. 결정계수 R^2 은 PM10 농도 Y_t 가 설명변수에 의해 어느 정도 설

명되는지를 나타내는 지표로 값이 0이면 0% 설명되고 1이면 100% 설명됨을 의미한다. 따라서 값이 1에 가까울수록 Y_t 를 잘 설명하는 좋은 모형이다. 예측값과 실제값의 분산정도를 나타내는 RMSE (root mean square error)의 공식은 다음과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2}, \quad (2)$$

o_i 와 p_i 는 각각 i 일의 관측 PM10 농도와 추정 PM10 농도를 의미하며, n 은 통계량에 사용된 일수 (day)이다. RMSE는 값이 작을수록 좋은 모형이다.

4. 평택 PM10 농도의 ARE 모형

평택 지역의 일 최고 PM10 농도를 연도별 (그림 1a), 월별 (그림 1b)로 선 그래프를 통해 살펴보았다. PM10 고농도의 이상값들이 많아 평균값 대신 연도별 중앙값, 월별 중앙값을 사용해 그래프를 그렸다.

먼저, 연도별로 살펴보면 PM10 농도 중앙값이 2004년에는 $90 \mu\text{g}/\text{m}^3$ 로 나타났고, 2006년도가 $114 \mu\text{g}/\text{m}^3$ 로 매우 높게 나타났고 2007년에는 $106 \mu\text{g}/\text{m}^3$, 2008년에는 $104 \mu\text{g}/\text{m}^3$, 2009년에는 $95 \mu\text{g}/\text{m}^3$ 로 계속 감소 추세에 있다. 2004년에 PM10 농도가 비교적 적게 나타난 이유는 봄인 3월, 4월, 5월에 나타난 황사횟수가 5일로, 2006년의 11일, 2007년의 9일 등보다 비교적 적게 나타난 때문으로 생각된다. PM10 농도를 월별로 살펴보면, 봄철인 3월이 $124 \mu\text{g}/\text{m}^3$, 4월이 $122 \mu\text{g}/\text{m}^3$, 5월이 $123 \mu\text{g}/\text{m}^3$ 높게 나타났고, 7월이 $84 \mu\text{g}/\text{m}^3$, 8월이 $68 \mu\text{g}/\text{m}^3$, 9월이 $78 \mu\text{g}/\text{m}^3$ 로 비교적 낮게 나타났다. 봄에 PM10 농도가 높은 주요 원인은 중국에서 오는 황사 때문인 것으로 알려져 있다.

4.1 월별 PM10 농도 자료에 관한 ARE 모형

그림 1에서 월별로 PM10 농도가 매우 다르게 나타나 월별로 PM10 농도 모형을 분석할 필요를 느껴 분석하여 표 2에 나타냈다. 월별 분석은 2003~2007년의 5개월 자료를 모델링기간으로, 2008~2009년의 2개월 자료를 검증기간으로 SAS 통계 package를 이용하여 분석하였다. 모델링기간과 검증기간에 사용된 모든 설명변수들은 유의수준 10%에서 유의한 변수들만 선택하였다. ARE 모형의 변수 선택은 10종류의

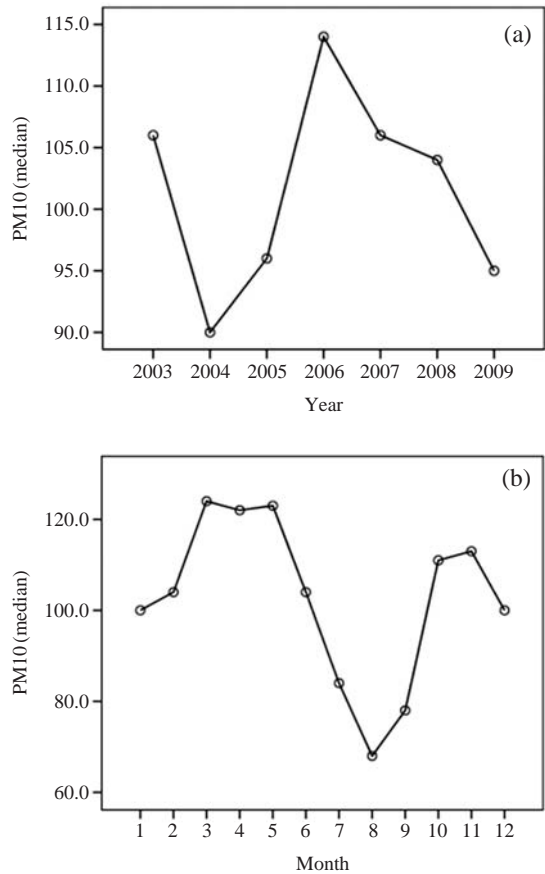


Fig. 1. The line plots of (yearly and monthly) PM10 concentrations in Pyeongtaek City of Gyeonggi-Do.

설명변수들 간의 다중공선성 (multi-collinearity)을 고려하면서 설명변수들 간의 다양한 combination 모형을 분석한 후 RMSE 값이 작은 모형을 선택하였다. ARE 모형의 설명변수 차수 선택은 표 1에 나타나 있다. 대기변수인 SO_2 , NO_2 , CO , O_3 은 일 최고농도를 사용하는데, 반응변수인 PM10의 일 최고농도 시간보다 더 늦게 나타날 수 있어, 예측을 고려하여 전일 최대 농도를 선택하여 전일자료($t-1$)를 사용하였다. 기상변수인 최고온도, 풍속, 상대습도 등은 PM10과 상관관계가 높은 시간대의 당일 자료(t)를 사용하였다.

ARE 모형의 모수 추정에는 YW (Yule-Walker), ULS (Unconditional Least Square), MLE (Maximum Likelihood Estimate) 등의 방법이 있지만, 본 연구에서는 PM10 모형과 같이 과거시점의 종속변수 (lagged

Table 2. The monthly ARE models for PM10 concentrations.

Month	Period	ARE Model	R ²	RMSE
Jan.	Modeling	$Y_t=52.35+0.49Y_{t-1}+223.92N_{t-1}+0.28T_t-6.82W_t+\epsilon_t,$ $\epsilon_t=0.21\epsilon_{t-19}+e_t$	0.43	21.81
	Verification	$Y_t=61.37+0.30Y_{t-1}+566.77N_{t-1}-8.68W_t+\epsilon_t,$ $\epsilon_t=0.28\epsilon_{t-6}+0.29\epsilon_{t-10}-0.25\epsilon_{t-16}+e_t$	0.44	21.54
Feb.	Modeling	$Y_t=64.84+0.34Y_{t-1}+1.06T_t-2.62W_t+0.03H_t+\epsilon_t,$ $\epsilon_t=0.26\epsilon_{t-10}+e_t$	0.23	24.54
	Verification	$Y_t=63.08+0.68H_t+\epsilon_t,$ $\epsilon_t=0.32\epsilon_{t-1}+e_t$	0.18	22.98
Mar.	Modeling	$Y_t=66.36+519.9O_{t-1}+2.24T_t-1.46A_t+\epsilon_t,$ $\epsilon_t=0.30\epsilon_{t-1}+e_t$	0.30	22.17
	Verification	$Y_t=68.29+555.8O_{t-1}+1.66T_t-1.03A_t+\epsilon_t,$ $\epsilon_t=0.25\epsilon_{t-1}+e_t$	0.17	25.88
April	Modeling	$Y_t=62.09+0.31Y_{t-1}+13.33I_{t-1}-0.42C_t+\epsilon_t,$ $\epsilon_t=-0.26\epsilon_{t-18}+e_t$	0.17	22.57
	Verification	$Y_t=71.07+83.07I_{t-1}-5.46C_t+\epsilon_t,$ $\epsilon_t=0.41\epsilon_{t-3}+e_t$	0.50	25.83
May	Modeling	$Y_t=62.69+0.48Y_{t-1}+826.2S_{t-1}+0.37T_t-0.37A_t-4.25W_t+\epsilon_t,$ $\epsilon_t=-0.26\epsilon_{t-6}+e_t$	0.42	21.69
	Verification	$Y_t=12.69+2007S_{t-1}+3.34T_t-0.77W_t+\epsilon_t,$ $\epsilon_t=0.49\epsilon_{t-1}+e_t$	0.28	21.75
June	Modeling	$Y_t=0.30Y_{t-1}+1164S_{t-1}+2.74T_t-9.64W_t-0.22H_t+\epsilon_t,$ $\epsilon_t=0.38\epsilon_{t-2}+e_t$	0.28	21.30
	Verification	$Y_t=165.55+0.26Y_{t-1}-22.15W_t-0.41H_t+\epsilon_t,$ $\epsilon_t=0.28\epsilon_{t-3}+e_t$	0.48	20.47
July	Modeling	$Y_t=46.49+0.44Y_{t-1}+1.41R_t-4.24W_t-0.10A_t+\epsilon_t,$ $\epsilon_t=0.18\epsilon_{t-13}+e_t$	0.39	22.81
	Verification	$Y_t=52.50+0.44Y_{t-1}-0.20A_t+\epsilon_t,$ $\epsilon_t=0.37\epsilon_{t-9}+e_t$	0.27	24.83
Aug.	Modeling	$Y_t=56.74+0.54Y_{t-1}-6.75W_t-0.83C_t+\epsilon_t,$ $\epsilon_t=0.18\epsilon_{t-4}+e_t$	0.45	19.57
	Verification	$Y_t=78.72+3189S_{t-1}-6.16W_t-2.35C_t+\epsilon_t,$ $\epsilon_t=0.38\epsilon_{t-1}+e_t$	0.25	22.81
Sep.	Modeling	$Y_t=37.01+0.35Y_{t-1}+477.35O_{t-1}+0.52R_t-5.34W_t+\epsilon_t,$ $\epsilon_t=0.19\epsilon_{t-18}+e_t$	0.41	19.95
	Verification	$Y_t=23.44+0.19Y_{t-1}+828.3O_{t-1}+0.95R_t-5.14W_t+\epsilon_t,$ $\epsilon_t=0.39\epsilon_{t-14}+e_t$	0.40	22.43
Oct.	Modeling	$Y_t=85.72+0.29Y_{t-1}+261.78N_{t-1}-9.72W_t-1.34A_t+\epsilon_t,$ $\epsilon_t=-0.24\epsilon_{t-14}-0.30\epsilon_{t-22}+e_t$	0.35	20.99
	Verification	$Y_t=27.71+0.52Y_{t-1}+557.4N_{t-1}-1.31W_t-0.21A_t+\epsilon_t,$ $\epsilon_t=-0.47\epsilon_{t-17}+e_t$	0.45	18.42
Nov.	Modeling	$Y_t=89.72+0.18Y_{t-1}+481.3O_{t-1}+1.09T_t-7.34W_t-2.67C_t+\epsilon_t,$ $\epsilon_t=-0.24\epsilon_{t-17}+e_t$	0.39	22.57
	Verification	$Y_t=62.05+1215O_{t-1}+2.67T_t-10.25W_t-1.62C_t+\epsilon_t,$ $\epsilon_t=0.48\epsilon_{t-2}+e_t$	0.42	27.19
Dec.	Modeling	$Y_t=37.70+0.38Y_{t-1}+389.2N_{t-1}+0.89T_t-6.41W_t+0.25H_t+\epsilon_t,$ $\epsilon_t=0.24\epsilon_{t-4}-0.22\epsilon_{t-5}+0.18\epsilon_{t-6}+e_t$	0.49	20.43
	Verification	$Y_t=61.28+731.8N_{t-1}+2.43T_t-5.37W_t+0.07H_t+\epsilon_t,$ $\epsilon_t=0.46\epsilon_{t-10}+e_t$	0.34	28.21

dependent variable)인 Y_{t-1} 가 설명변수로 사용되는 경우에 적합한 MLE 방법을 사용하였다.

ARE모형의 적합성 검정인 오차항의 자기상관의 존재여부를 검정하는 방법으로는, PM10 모형과 같이 과거시점의 종속변수가 설명변수로 사용되는 경우에 적합한 Durbin-h 검정을 사용해 검정하였고 또한 잔차의 독립성검정을 위해 ACF(Auto-Correlation Function)와 PACF(Partial Auto-Correlation Function)을 실시하였고 독립성임을 확인하였다(조신섭과 이정형, 1997 참고).

표 2의 월별 ARE 모형을 보면, PM10 농도를 설명하는 대기변수와 기상변수가 1월에서 12월까지 월마다 조금씩 다르게 나타났다. 1월, 2월, 5월, 6월, 11월, 12월의 모델링기간의 모형에 전일 PM10 농도(Y_{t-1}), 일 최고온도(T_t)와 풍속(W_t)이 포함되었다. 10%에서 유의한 변수들만 선택했기 때문에 모델링기간에 포함된 변수들이 검증기간에는 포함되지 않은 달도 있다. 모형에서 일 최고온도의 계수는 양수로 풍속은 음수로 나타나 온도가 오를수록 바람의 양이 적을수록 PM10 농도는 증가하는 것으로 나타났다. 7월, 9월의 모델링기간에 일사량(R_t)이 모형에 포함되었고, 5월, 7월, 9월에는 강수량(A_t)이 모형에 포함되었다.

대기변수는 1월, 10월, 11월에는 $NO_2(N_{t-1})$, 3월, 9월 11월에는 $O_3(O_{t-1})$, 4월에는 $CO(I_{t-1})$, 5월, 6월에는 $SO_2(S_{t-1})$ 가 모형에 포함돼 월별로 다르게 나타났고 대기변수의 계수는 모두 양수로 나타났다.

기상변수로는, 특히 일 최고온도와 풍속이 포함된 달이 많아 영향을 많이 주는 것으로 나타났다. 1~3월, 5~6월, 11~12월에 일 최고온도가 설명변수로 포함되었고, 1~2월, 5~12월에는 풍속이 설명변수로 포함되었다.

특이한 사실은 상대습도(H_t)는 6월의 모형에는 계수가 음수(-0.41)로 나타났지만, 2월, 12월에는 각각 0.03, 0.25인 양수로 나타났다. 즉 여름에는 상대습도가 높을수록 PM10 농도는 감소하지만, 겨울에는 상대습도가 낮을수록 PM10 농도가 감소하는 것으로 나타났다.

적합도 척도인 결정계수 R^2 는 모델링기간에 2월, 4월, 6월에 각각 0.23, 0.17, 0.28로 좀 낮게 나타났고, 그 외의 달에서는 0.30 이상으로 나타났다. 2~6월에 R^2 가 비교적 낮게 나타난 이유 중의 하나는 봄철에 중국에서 많이 불어오는 황사의 영향이 클 것으로 생

각되며 그러므로 본 연구에 사용된 대기변수나 기상변수가 PM10에 끼치는 영향이 상대적으로 적기 때문인 것으로 분석된다. 그러나 1월, 8월, 12월에는 R^2 가 각각 0.43, 0.45, 0.49로 높게 나타났다. 12월의 예를 들면, PM10 농도 Y_t 는 Y_{t-1} , $NO_2(N_{t-1})$, 일 최고온도(T_t), 풍속(W_t), 습도(H_t)에 의해 49% 설명되는 것으로 나타났다. 여기서 Y_{t-1} 과, N_{t-1} , T_t , H_t 는 양수로 나타났고, 반면에 W_t 는 음수로 나타나 이산화질소, 온도, 습도가 오를수록 PM10 농도는 높아지며, 풍속이 높을수록 PM10 농도는 낮아지는 것으로 나타났다. 또한 7월에는 PM10 농도 Y_t 는 Y_{t-1} , 일사량(R_t), 풍속(W_t), 강수량(A_t)에 의해서 30% 설명되는 것으로 나타나 날씨가 무덥고, 비가 많이 오는 7월에는 풍속과 함께 일사량, 강수량이 PM10 농도에 영향을 주는 것으로 분석되었다. 검증기간 동안의 R^2 도 0.17~0.50으로 모델링기간과 거의 비슷하게 나타났다.

적합도 척도로 실제값과 예측값의 차이를 나타내는 RMSE 값은 모델링기간에는 19.57~24.54 $\mu\text{g}/\text{m}^3$ 로 나타났고, 검증기간에는 18.42~28.21 $\mu\text{g}/\text{m}^3$ 로 나타났다. 본 연구에서 사용한 대기변수와 기상변수는 각각 국립환경과학원과 기상청에서 쉽게 구할 수 있는 자료이므로, 표 2를 이용하면, 월별 PM10 농도를 예측하는 데 도움이 될 것으로 생각된다.

4. 2 계절별 PM10 농도 자료에 관한 ARE 모형

PM10 농도를 봄, 여름, 가을, 겨울의 계절별 ARE 모형으로 분석한 결과를 표 3에 나타내었다. 계절 분석에서, 봄은 3~5월, 여름은 6~8월, 가을은 9~11월, 겨울은 12~2월의 자료를 사용하였다. 월별 분석과 같이 2003~2007년의 자료를 모델링기간으로, 2008~2009년의 자료를 검증기간으로 SAS 통계 package를 이용하여 분석하였다. 대체적으로 월별모형에 포함된 설명변수들이 계절별 모형에도 사용되었고 특히, 풍속 W_t 는 봄, 여름, 가을, 겨울 4계절 모두 포함되어 있어 영향을 제일 많이 주는 기상변수로 나타났다. 일 최고온도 T_t 는 여름을 제외한 봄, 가을, 겨울모형에 포함되어 있다. 봄, 여름모형에 일사량 R_t 도 포함되어 있다. 계절별 모델링기간의 ARE 모형의 결정계수 R^2 는 가을, 여름, 겨울은 각각 0.47, 0.42, 0.38로 비교적 높게 나타났고, 봄은 0.27로 낮게 나타났다. 가을 PM10 농도 Y_t 는 Y_{t-1} , N_{t-1} , T_t , W_t , C_t 에 의해 47% 정도 설

Table 3. The seasonal ARE models for PM10 concentrations.

Season	Period	ARE Model	R ²	RMSE
Spring	Modeling	$Y_t = 53.29 + 0.45Y_{t-1} + 0.43T_t - 2.98W_t + 0.62R_t + \epsilon_t$ $\epsilon_t = -0.13\epsilon_{t-22} + e_t$	0.27	23.36
	Verification	$Y_t = 93.68 + 1082S_{t-1} + 0.93T_t - 3.39W_t - 0.90C_t + \epsilon_t$ $\epsilon_t = 0.29\epsilon_{t-1} + 0.17\epsilon_{t-5} + e_t$	0.12	25.05
Summer	Modeling	$Y_t = 37.26 + 0.46Y_{t-1} + 165.9O_{t-1} - 6.98W_t + 1.02R_t + \epsilon_t$ $\epsilon_t = 0.13\epsilon_{t-4} + 0.16\epsilon_{t-7} + e_t$	0.42	21.10
	Verification	$Y_t = 40.51 + 0.40Y_{t-1} + 152.5O_{t-1} - 9.95W_t + \epsilon_t$ $\epsilon_t = 0.14\epsilon_{t-5} + e_t$	0.36	25.08
Fall	Modeling	$Y_t = 85.66 + 0.44Y_{t-1} + 145.2N_{t-1} - 0.37T_t - 9.05W_t - 1.88C_t + \epsilon_t$ $\epsilon_t = 0.10\epsilon_{t-6} + 0.13\epsilon_{t-23} + e_t$	0.47	21.36
	Verification	$Y_t = 56.14 + 0.41Y_{t-1} + 620.3N_{t-1} - 0.22T_t - 7.36W_t - 1.67C_t + \epsilon_t$ $\epsilon_t = 0.17\epsilon_{t-6} - 0.17\epsilon_{t-14} - 0.15\epsilon_{t-22} + e_t$	0.52	23.27
Winter	Modeling	$Y_t = 28.00 + 0.37Y_{t-1} + 330N_{t-1} + 0.95T_t - 4.10W_t + 0.27H_t + \epsilon_t$ $\epsilon_t = 0.11\epsilon_{t-6} + 0.12\epsilon_{t-7} + e_t$	0.38	22.31
	Verification	$Y_t = 30.95 + 0.28Y_{t-1} + 588N_{t-1} - 4.32W_t + 0.41H_t + \epsilon_t$ $\epsilon_t = 0.16\epsilon_{t-6} + 0.18\epsilon_{t-22} + e_t$	0.31	24.22

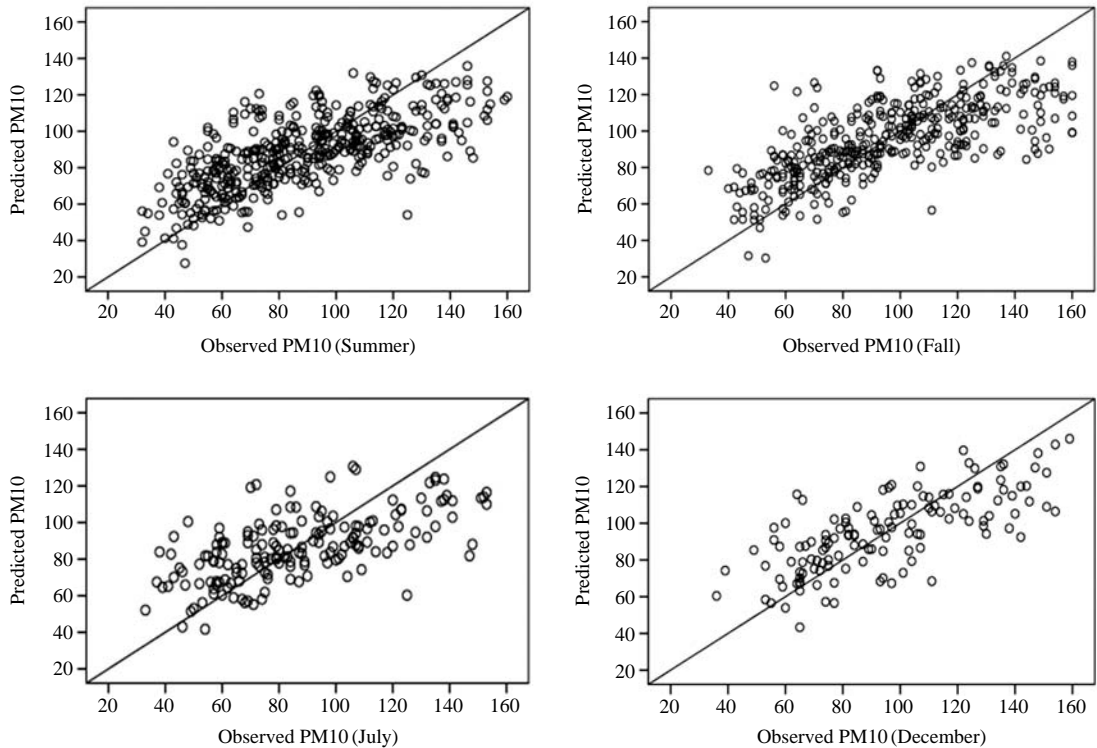


Fig. 2. The Scatter plots (monthly-July & December and yearly-Summer & Fall) between actual value and ARE estimated value of PM10 concentrations.

명되며, 여름 PM10 농도는 Y_{t-1} , O_{t-1} , W_t , R_t 에 의해 42% 설명된다. 검증기간의 R^2 는 가을모형은 0.52로 모델링모형의 0.47보다 향상됐으나, 다른 계절 검증 모형은 더 낮아졌다. RMSE 값은 모델링기간은 22.10 ~ 23.36 $\mu\text{g}/\text{m}^3$ 로 나타났고 검증기간은 23.27 ~ 25.05 $\mu\text{g}/\text{m}^3$ 로 모델링기간보다 조금 높게 나타났다.

4.3 ARE 모형의 실제값과 추정치의 산점도

PM10 농도 실제값과 ARE 모형 추정치의 산점도를 월별에서는 7월과 12월의 모델링모형, 계절별에서는 여름과 가을의 모델링모형을 그림 2에 나타내었다. 12월 모형의 산점도를 보면, 추정이 대체적으로 잘되었고, 7월 모형은 PM10 농도가 120 $\mu\text{g}/\text{m}^3$ 이하인 경우에는 추정이 잘 되었고, 120 $\mu\text{g}/\text{m}^3$ 이상인 경우에는 과소추정되는 경향이 있는 것으로 나타났다. 반면에 여름모형과 가을모형은 PM10 농도가 저농도인 60 $\mu\text{g}/\text{m}^3$ 이하인 경우 약간 과대추정된 경향이 있고, 60 ~ 120 $\mu\text{g}/\text{m}^3$ 사이는 추정이 잘되었고, 고농도인 125 $\mu\text{g}/\text{m}^3$ 이상인 경우에는 과소추정되는 경향이 있는 것으로 나타났다. ARE 모형은 PM10 농도에 영향을 주는 설명변수를 수식으로 표현하는 장점이 있으므로, 모형에 PM10 농도에 영향을 주는 기상변수와 대기변수를 좀 더 포함시켜 개발하면, 특히 PM10 농도가 120 $\mu\text{g}/\text{m}^3$ 이상인 고농도의 경우에, 실제값에 좀 더 가까이 접근하는 좋은 결과를 가질 수 있을 것으로 기대한다.

5. 결 론

경기 남부지역인 평택의 PM10 농도를 월별과 계절별로 시계열 모형인 ARE 방법으로 분석하였다. ARE 모형은 설명변수를 분석에 사용할 수 있는 모형으로 시계열자료를 적합시킬 때 적절한 모형이다. 본 연구에서는, PM10 농도에 영향을 주면서 쉽게 자료를 구할 수 있는 6종류의 기상자료와 4종류의 대기자료를 설명변수로 사용하였다. 6종류의 기상자료는 평택과 가장 가까운 수원시 기상청 자료로 일 최고온도, 풍속, 상대습도, 강수량, 일사량, 운량이며 4종류의 대기자료는 국립환경과학원의 평택시 SO_2 , NO_2 , CO , O_3 이다.

연구기간인 2003년부터 2009년의 1월 1일에서 12

월 31일까지의 7년간 자료 중 2003년부터 2007년까지의 5년간 자료를 모델링기간으로 사용했고, 2008년과 2009년 2년간 자료를 검증기간으로 사용하여 월별과 계절별 ARE 모형으로 분석하였다.

월별 PM10 농도를 분석한 결과, 기상변수는 일 최고온도와 풍속이 포함된 달이 많아 영향을 많이 주는 것으로 나타났다. 4월, 7 ~ 10월을 제외한 모든 달에 일 최고온도가 설명변수로 포함되었고, 3 ~ 4월을 제외한 모든 달에 풍속이 포함되었다. 특히하게, 상대습도는 더운 6월 모형에는 계수가 -0.41로 음수로 나타났고, 추운 2월, 12월 모형에는 각각 0.03, 0.25인 양수로 나타나 여름에는 상대습도가 높을수록 PM10 농도는 감소하지만, 건조한 겨울에는 상대습도가 낮을수록 PM10 농도가 감소한 것으로 나타났다. 대기변수 SO_2 , NO_2 , CO , O_3 는 월별로 각각 다르게 영향을 주는 것으로 나타났다. 적합도 척도인 결정계수 R^2 는 모델링기간에 2월, 4월, 6월에 좀 낮게 나타났으나 그 외의 달에는 0.30 이상으로 나타났으며, 특히 1월, 8월, 12월에는 각각 0.43, 0.45, 0.49로 높게 나타났다. 검증기간 동안의 R^2 도 0.17 ~ 0.50으로 모델링기간과 거의 비슷하게 나타났다. 적합도 척도로 실제값과 예측값의 차이를 나타내는 RMSE 값은 모델링기간에는 19.57에서 24.54 $\mu\text{g}/\text{m}^3$ 로 나타났고 검증기간에는 18.42 ~ 28.21 $\mu\text{g}/\text{m}^3$ 로 나타났다.

계절별 PM10 농도 분석 결과는 풍속이 봄, 여름, 가을, 겨울 4계절 모두 포함되어 있어, 영향을 제일 많이 주는 기상변수로 나타났다. 일 최고온도는 여름을 제외한 봄, 여름, 가을모형에 포함되어 있다. 봄, 여름모형에는 일사량도 포함되어 있다. 계절별 모델링기간의 ARE 모형의 결정계수 R^2 는 가을, 여름, 겨울은 각각 0.47, 0.42, 0.38로 비교적 높게 나타났고, 봄은 0.27로 낮게 나타났다.

ARE 모형은 표 2, 표 3과 같이 PM10 농도에 어떤 설명변수들이 영향을 주는지 등의 수식으로 간단히 표현할 수 있는 장점이 있으므로, 본 연구에서 사용하지 못한 설명변수들을 좀 더 첨가하여 분석하면 더 좋은 모형이 되리라 기대한다. 특히 PM10 농도가 120 $\mu\text{g}/\text{m}^3$ 이상인 고농도의 경우 ARE 모형이 과소추정되는 경향이 있으므로 좀 더 많은 기상자료와 대기자료를 첨가해야 한다.

앞으로, PM10 농도에 사용되는 설명변수에 관한 다양한 연구와 함께 다양한 모형에 관한 연구가 지

역별로 활발히 이루어져야 된다고 생각된다.

감사의 글

이 논문은 2010년도 평택대학교 학술연구비의 지원에 의하여 연구되었음.

참 고 문 헌

조신섭, 이정형 (1997) SAS/ETS를 이용한 경제시계열 분석, 자유아카데미.

Brunelli, U., V. Piazza, L. Pignato, F. Sorbello, and S. Vitabile (2007) Two-days ahead prediction of daily maximum concentrations of SO₂, O₃, PM10, NO₂, CO in the Urban area of Palermo, Italy, *Atmospheric Environment*, 41, 2967-2995.

Chung, Y.S. and J.S. Chung (1991) On surface ozone observed in the Seoul metropolitan area during 1989 and 1990, *J. KAPRA*, 7(3), 169-179.

Cobourn, W.G. (2010) An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations, *Atmospheric Environment*, 44, 3015-3023.

Hubbard, M. and W. Cobourn (1998) Development of a regression model to forecast ground-level ozone concentration in Louisville, KY, *Atmospheric Environment*, 32, 2637-2647.

Kim, S.C. and S.Y. Joh (2005) Derivation of new box model to analyze the air pollution trends in a metropolitan area, *J. Korean Soc. Atmos. Environ.*, 21(3), 377-397.

Kim, Y.P. (2006) Air pollution in Seoul caused by aerosols, *J. Korean Soc. Atmos. Environ.*, 22(5), 535-553.

Koo, Y.S., H.Y. Yun, H.Y. Kwon, and S.H. Yu (2010) A development of PM10 forecasting system, *J. Korean Soc. Atmos. Environ.*, 26(6), 666-682.

Lee, J.Y., B.J. Kong, J.S. Han and M.D. Lee (2008) Long term analysis of PM10 concentration in Seoul using KZ filter, *J. Korean Soc. Atmos. Environ.*, 24, 63-71.

Lee, Y.S., H.G. Kim, J.S. Park, and H.K. Kim (2006) A study on statistical forecasting models of PM10 in Pohang region by the variable transformation, *J. Korean Soc. Atmos. Environ.*, 22(5), 614-626.

Perez, P. and J. Reyes (2006) An integrated neural network model for PM10 forecasting, *Atmospheric Environment*, 40, 2845-2851.

Wise, E.K. and A.C. Comrie (2005): Meteorologically adjusted urban air quality trends in the Southwestern United States, *Atmospheric Environment*, 39, 2969-2980.