

경기도 수원시 미세먼지 농도의 시계열모형 연구

이혼자¹

¹평택대학교 디지털응용정보학과

접수 2010년 9월 11일, 수정 2010년 11월 12일, 게재확정 2010년 11월 23일

요약

미세먼지 농도는 국가의 중요한 환경 척도 중의 하나이다. 본 연구에서는 경기도 남부에 위치한 수원시 2003년-2009년 미세먼지 농도를 주위에서 쉽게 구할 수 있는 대기자료와 기상자료를 이용하여 자기회귀오차모형으로 월별로 분석하였다. 미세먼지 농도 분석을 위한 대기자료는 이산화황, 이산화질소, 일산화탄소, 오존 등을 사용했고, 기상자료로는 일 최고온도, 풍속, 상대습도, 강수량, 일사량, 운량을 사용하였다. 분석 결과, 자기회귀오차모형으로 월별 미세먼지 농도를 13%-49% 정도 설명할 수 있다.

주요용어: 기상변수, 대기변수, 미세먼지 농도, 설명변수, 자기회귀오차모형.

1. 서론

세계 많은 나라들이 급속한 산업성장 등으로 인하여 대기오염 문제를 갖고 있다. 특히, 대기 중의 높은 PM10 (미세먼지)은 인간의 건강에 심각한 피해를 주는 것으로 알려지고 있다. 우리나라 환경부에서는 2007년부터 PM10, 이산화질소 (NO_2) 등의 대기환경 기준을 강화하였으며, PM10의 경우 연 평균 $70 \mu g/m^3$ 에서 $50 \mu g/m^3$ 으로, 일평균 기준은 $150 \mu g/m^3$ 에서 $100 \mu g/m^3$ 으로 개정하여, 미세먼지 저감도 정책을 강화하였다. 이러한 미세먼지 저감 정책이 효과적으로 진행되기 위해서는 PM10 농도변화 추이와 특성을 이해하고, 그 주요원인을 파악하는 종합적인 연구가 뒷받침 되어야 한다 (김용표, 2006).

지금까지 다양한 방법으로 PM10 농도에 영향을 준 오염원에 관해 연구되어 왔다. PM10 농도의 모형에 관한 국내연구의 경향을 살펴보면, 전국 도시대기 측정망의 2000-2005년 PM10 농도 군집분석 (한지현 등, 2008), 대구지역 대기 중 미세먼지의 오염도 분석 및 기상인자에 따른 영향 평가 (황윤정 등, 2009), PMF (Positive Matrix Factorization)를 응용한 구미시 PM-10 오염원의 정량적 기여도 추정연구 (황인조 등, 2008), PMF모형을 이용한 용인·수원 경계지역에서 PM10 오염원의 확인과 상대적 기여도의 추정 (이형우, 2008), KZ (Kolmogorov-Zurbenko) 필터법을 이용한 서울지역 미세먼지 농도의 장기변화 분석 (이정영 등, 2008), 서울지역 분진 (PM)에 관한 장기추세 연구 (박혜련과 최기현, 2009) 등 다양한 방법으로 PM10 농도에 관해 연구하였다.

국외의 경우를 보면, 미국 북서부지역 PM2.5 분석을 위해 PMF모형을 사용했고 (Kim 등, 2003), 칠레 Santiago 지역의 PM10 과 오존 분석을 위해 군집분석을 사용하였고 (Gramsch 등, 2006), 오존과 PM10 분석을 위해 KZ 필터를 사용하였다 (Wise와 Comrie, 2005a).

기상요인이 미세먼지 농도에 많은 영향을 준다는 사실이 국내외적으로 알려져 왔다. 국내에서는, 미세먼지 농도를 지배하는 요인 중에는 배출원을 제외하고는 가장 영향을 주는 요소는 기상변수이며, 특히

¹ (450-701) 경기도 평택시 용이동 111, 평택대학교 디지털응용정보학과, 부교수. E-mail: esther@ptu.ac.kr

풍향, 풍속, 기온, 습도 등의 상태에 따라 농도가 크게 좌우된다 (정용승과 정재섭, 1991). 또한 미국 남서부지역의 1990-2003년 기간 동안 미세먼지 농도를 연구한 Wise와 Comrie (2005b)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 수원시 PM10농도를 2003년 1월1일부터 2009년의 12월 31일까지의 7년간 자료를 시계열 모형에 적합하고자 한다. PM10농도의 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 대기 자료를 설명변수로 포함한 다변량 분석인 ARE (Autoregressive Error, 자기회귀오차) 모형으로 분석하였다.

2. PM10자료 및 설명변수 자료

PM10 자료는 국립환경과학원의 수원시 영통동 관측소의 시간별 자료 (하루 24번)를 이용했고 분석을 위하여 일 최고 PM10 농도를 사용했다. 기간은 2003년부터 2009년까지의 1월부터 12월까지 7년간 자료 중 2003년부터 2007년까지의 각 월별자료를 모델링기간으로 2008년과 2009년을 검증기간으로 월별 분석하였다. 본 연구에서 1월에서 12월 까지 월별로 분석한 이유는 PM10 농도가 월별로 차이가 있기에 이 분석이 중요한 것으로 생각되었기 때문이다.

설명변수로 사용한 기상자료와 대기자료는 PM10과 연관이 있으면서, 쉽게 구할 수 있는 변수를 사용하였다. 기상 자료는 시간별 자료로 수원시 기상청자료를 사용하였다. 기상 자료도 PM10 과 같이 2003년-2009년 자료를 사용하였다. 6종류의 기상 자료는 이정영 등 (2008), Wise와 Comrie (2005b), 이훈자 (2009), 주용성 등 (2009), 김상용과 김진아 (2009)을 참고로 하여 일 최고온도 (maximum temperature), 풍속 (wind speed), 상대습도 (relative humidity), 강수량 (rainfall), 일사량 (radiation), 운량 (amount of cloud)을 사용했으며 표 2.1에 나타나 있다. 풍향 (wind direction)은 미세먼지에 영향을 많이 미치나 정량화의 어려움이 있어 설명변수에 포함하지 않았다. 4종류의 대기 자료로는 이산화황 (SO_2), 이산화질소 (NO_2), 일산화탄소 (CO), 오존(O_3)를 사용했고, PM10 자료와 같이 국립환경과학원의 수원시 관측소를 이용하여, 2003년-2009년의 1월1일-12월 31일의 시간별 자료를 이용하였다. 사용된 대기 자료도 표 2.1에 나타나 있다.

표 2.1 설명변수로 사용한 대기변수와 기상변수

변수	사용시간	단위	표시
PM10	전일 1시간 최고 농도	$\mu g/m^3$	Y_{t-1}
SO_2	전일 1시간 최고 농도	ppb	S_{t-1}
NO_2	전일 1시간 최고 농도	ppb	N_{t-1}
CO	전일 1시간 최고 농도	0.1ppm	I_{t-1}
O_3	전일 1시간 최고 농도	ppb	O_{t-1}
온도	일 최고 온도	$^{\circ}C$	T_t
풍속	일 9시-15시 평균	m/s	W_t
운량	일 9시-15시 평균	-	C_t
습도	일 9시-13시 평균	%	H_t
강수량	일 총강수량	mm	A_t
일사량	일 9시-15시 총 일사량	MJ/M^2	R_t

6종류의 기상 자료와 4종류의 대기 자료의 시간별 선택과 차수 선택은 Hubbard와 Cobourn (1998)를 참고로 하여 각 자료를 시간별 혹은 일별 시간대를 사용하였다. 예를 들면, 운량 (cloud) 변수는 일별 9시-15시의 평균을 사용하였다. 각 설명변수의 구체적인 시간대는 표 2.1에 나타나 있다. 표 2.1의 표

시는 4장의 모형식 변수표시이다.

3. PM10농도의 ARE 모형과 적합성

반응변수에 영향을 주는 설명변수가 있을 때는 설명변수를 분석에 이용하는 것이 효율적이다. ARE 모형은 설명변수를 분석에 사용 할 수 있는 모형으로 시계열자료를 회귀모형에 적합 시킬 때 적합한 모형이다. 반응변수인 PM10에 영향을 주는 SO_2 농도나, 일별 최고온도 같은 설명변수를 사용하여 자료를 분석한다. ARE 모형은 전이함수 모형의 특수한 경우로, 오차항 (ϵ_t)가 독립이 아니라 서로 상관관계를 갖게 되며, 특히 AR (Autoregressive) 형태를 갖는다. 일반적인 k 차 ARE 모형은 다음과 같다.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_p X_{tp} + \epsilon_t, \quad t = 1, 2, \dots, n,$$

$$\epsilon_t = e_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2} \dots - \phi_k \epsilon_{t-k}, \quad \text{이고 } e_t \sim i.i.d N(0, \sigma^2) \quad (3.1)$$

식 (3.1)에서 오차항 (ϵ_t)는 서로 상관관계가 있고 AR(k) 형태를 갖는다. 또한 식 (3.1)에서 Y_t 는 반응변수인 PM10 농도를 나타내며, X_{tj} ($j = 1, 2, \dots, p$)는 Y_t 에 영향을 주는 NO_2 , 일 최고온도 등의 설명변수들을 나타낸다. 위의 ARE 모형과 회귀모형의 차이점은 ARE 모형은 식 (3.1)에서 오차항이 서로 독립이 아니라 시차에 따른 자기 상관을 갖고 있으며 특히 AR (Autoregressive) (k) 모형의 형태를 갖는다는 점이다. 따라서 식 (3.1)을 ARE (자기회귀오차)모형이라 부르며 시계열자료에 회귀모형을 적합시킬 경우 많이 사용된다.

ARE 모형들의 적합성을 측정하는데 있어 결정계수 (R^2)와 잔차 분석의 RMSE를 사용하였다. 결정계수는 PM10 농도 Y_t 가 설명변수에 의해 어느 정도 설명되는지를 나타내는 지표로 값이 1에 가까울수록 설명력이 높은 좋은 모형이다. 예측값과 실제값의 분산정도를 나타내는 RMSE (Root Mean Square Error)의 공식은 다음과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2}, \quad (3.2)$$

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

4. 수원 PM10 농도의 ARE 모형

수원지역의 일 최고 PM10농도를 연도별, 월별로 선 그래프를 (그림 4.1) 통해 살펴보았다. PM10 고농도의 이상값 들이 많아 평균값 대신 중앙값으로 연도별과 월별로 그래프를 그렸다. 먼저, 연도별로 살펴보면 PM10농도 중앙값이 2003년에는 $66 \mu g/m^3$ 로 낮게 나타났고, 2006년도는 $101 \mu g/m^3$ 로 매우 높게 나타났고 2007년에는 $92 \mu g/m^3$, 2008년에는 $86 \mu g/m^3$, 2009년에는 $74 \mu g/m^3$ 로 계속 감소 추세에 있다.

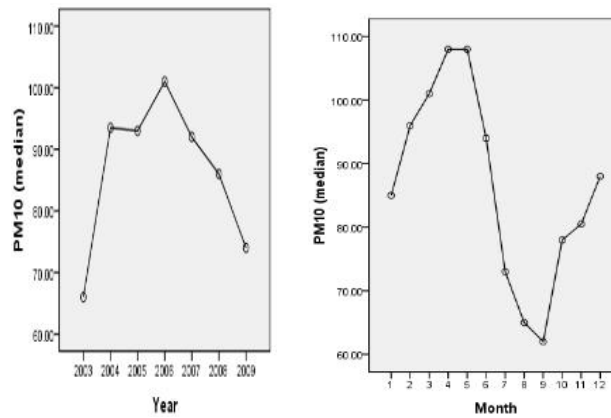


그림 4.1 수원시 연도별, 월별 PM10 농도의 선 그래프

PM10 농도를 월별로 살펴보면, 봄철인 3월이 $101 \mu\text{g}/\text{m}^3$, 4월이 $108 \mu\text{g}/\text{m}^3$, 5월이 $108 \mu\text{g}/\text{m}^3$ 로 높게 나타났고, 7월이 $73 \mu\text{g}/\text{m}^3$, 8월이 $65 \mu\text{g}/\text{m}^3$, 9월이 $62 \mu\text{g}/\text{m}^3$ 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 4.1에서 월별로 PM10 농도가 매우 다르게 나타나 월별로 PM10 농도 모형을 분석 할 필요를 느껴 분석하여 표 4.1에 나타났다. 월별 분석은 PM10 농도 연구기간인 2003년부터 2009년의 각 월별 자료 중, 2003년부터 2007년까지의 각 월별자료를 적합한 모형 구축을 위한 모델링 기간으로, 2008년과 2009년을 검증기간으로 사용하였다. 모델링기간과 검증기간에 사용된 모든 설명변수들은 유의수준 10%에서 유의한 변수들만 선택했다.

ARE 모형의 변수 선택은 10종류의 설명변수들 간의 다중공선성 (multi-collinearity)을 고려하면서 설명변수들 간의 다양한 모형을 분석한 후 RMSE 값이 작은 모형을 선택하였다. ARE 모형의 설명변수 차수 선택은 표 1에 나타나 있듯이 대기변수인 SO_2 , NO_2 , CO, O_3 는 전일 최대 농도를 선택하기 때문에 전일자료 ($t-1$)를 사용하였고, 기상변수인 최고온도, 풍속 등은 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 dependent variable)인 Y_{t-1} 가 설명변수로 사용되는 경우에 적합한 MLE 방법을 사용하였다.

ARE모형의 적합성 검정으로는 PM10 모형과 같이 과거시점의 종속변수가 설명변수로 사용되는 경우에 적합한 Durbin-h 검정 (조신섭과 이정형, 1997)을 사용해 검정하였고 또한 잔차의 ACF (Auto-Correlation Function)와 PACF (Partial Auto-Correlation Function)의 결과를 검정하였다 (조신섭과 이정형, 1997).

표 4.1의 월별 ARE 모형을 보면, PM10 농도를 설명하는 대기변수와 기상변수가 1월에서 12월까지 월 마다 조금씩 다르게 나타났다. 1월, 8월, 9월, 11월, 12월의 모델링기간에는 전일 PM10 농도 (Y_{t-1}), 일 최고온도 (T_t)와 풍속 (W_t)이 포함되었고 $\alpha=10\%$ 에서 유의한 변수들만 선택했기 때문에 검증기간에는 위의 변수들이 포함되지 않은 달도 있다. 6월, 7월에는 전일 SO_2 농도 (S_{t-1})와 일 최고온도 (T_t), 풍속 (W_t)이 포함되었다. 기상변수로는, 특히 일 최고온도와 풍속이 포함된 달이 많아

표 4.1 PM10 농도의 월별 ARE 모형

월	기간	ARE 모형	RMSE
1월	모델링	$Y_t = 65.03 + 0.30Y_{t-1} + 1.96T_t - 4.60W_t + \epsilon_t,$ $\epsilon_t = 0.21\epsilon_{t-4} - 0.20\epsilon_{t-16} + e_t$	0.33 24.49
	검증	$Y_t = 46.35 + 0.40Y_{t-1} + 0.72T_t + \epsilon_t,$ $\epsilon_t = 0.34\epsilon_{t-11} + e_t$	0.20 21.70
2월	모델링	$Y_t = 42.40 + 0.43Y_{t-1} + 1.58T_t + \epsilon_t,$ $\epsilon_t = -0.28\epsilon_{t-12} + e_t$	0.26 28.38
	검증	$Y_t = 74.47 + 3.33T_t + \epsilon_t,$ $\epsilon_t = -0.29\epsilon_{t-3} + e_t$	0.29 24.38
3월	모델링	$Y_t = 73.41 + 2.48T_t - 1.35A_t + \epsilon_t,$ $\epsilon_t = 0.32\epsilon_{t-1} + e_t$	0.19 24.88
	검증	$Y_t = 59.35 + 2.85T_t - 1.21A_t + \epsilon_t,$ $\epsilon_t = 0.32\epsilon_{t-1} - 0.31\epsilon_{t-3} - 0.27\epsilon_{t-6} + e_t$	0.29 23.29
4월	모델링	$Y_t = 84.56 + 0.14Y_{t-1} + 0.52T_t - 0.74A_t + \epsilon_t$ $\epsilon_t = -0.35\epsilon_{t-7} - 0.35\epsilon_{t-23} + e_t$	0.13 25.99
	검증	$Y_t = 55.58 + 2.44T_t - 1.16A_t + \epsilon_t,$ $\epsilon_t = -0.40\epsilon_{t-15} + e_t$	0.20 25.26
5월	모델링	$Y_t = 55.17 + 0.26Y_{t-1} + 2326S_{t-1} - 0.48A_t + \epsilon_t,$ $\epsilon_t = 0.36\epsilon_{t-14} + e_t$	0.29 26.85
	검증	$Y_t = 29.81 + 2.47T_t + \epsilon_t,$ $\epsilon_t = 0.53\epsilon_{t-1} + e_t$	0.11 22.93
6월	모델링	$Y_t = 12.14 + 0.20Y_{t-1} + 1569S_{t-1} + 2.49T_t - 4.54W_t + \epsilon_t,$ $\epsilon_t = 0.20\epsilon_{t-8} + e_t$	0.21 26.63
	검증	$Y_t = 93.70 + 1378S_{t-1} - 0.36A_t - 12.20W_t + \epsilon_t,$ $\epsilon_t = 0.49\epsilon_{t-1} - 0.39\epsilon_{t-12} + 0.57\epsilon_{t-13} + e_t$	0.47 19.20
7월	모델링	$Y_t = 0.27Y_{t-1} + 3169S_{t-1} + 3.64T_t - 0.17A_t - 5.72W_t + \epsilon_t,$ $\epsilon_t = 0.21\epsilon_{t-3} - 0.24\epsilon_{t-11} + e_t$	0.49 24.70
	검증	$Y_t = 46.63 + 0.56Y_{t-1} - 0.09A_t - 6.23W_t + \epsilon_t,$ $\epsilon_t = 0.51\epsilon_{t-13} + e_t$	0.42 18.91
8월	모델링	$Y_t = 0.32Y_{t-1} + 312.7O_{t-1} + 2.36T_t - 5.77W_t + \epsilon_t,$ $\epsilon_t = 0.20\epsilon_{t-20} + e_t$	0.42 24.54
	검증	$Y_t = 200.1O_{t-1} + 2.43T_t - 5.86W_t + \epsilon_t,$ $\epsilon_t = 0.38\epsilon_{t-1} + e_t$	0.15 22.53
9월	모델링	$Y_t = 0.32Y_{t-1} + 3.20T_t - 6.09W_t + \epsilon_t,$ $\epsilon_t = 0.22\epsilon_{t-4} + 0.20\epsilon_{t-10} + e_t$	0.48 19.45
	검증	$Y_t = 2.28T_t - 3.69W_t + \epsilon_t,$ $\epsilon_t = 0.40 - \epsilon_{t-1} + e_t$	0.12 19.74
10월	모델링	$Y_t = 37.43 + 0.46Y_{t-1} - 4.87W_t + 0.32H_t + \epsilon_t,$ $\epsilon_t = 0.20\epsilon_{t-18} - 0.25\epsilon_{t-23} + e_t$	0.31 26.09
	검증	$Y_t = 45.14 - 2.71W_t + 0.73H_t + \epsilon_t,$ $\epsilon_t = 0.71\epsilon_{t-1} + e_t$	0.31 17.94
11월	모델링	$Y_t = 36.30 + 0.41Y_{t-1} + 1.11T_t - 6.02W_t + \epsilon_t,$ $\epsilon_t = 0.23\epsilon_{t-8} + 0.23\epsilon_{t-9} + e_t$	0.37 25.76
	검증	$Y_t = 17.84 + 0.45Y_{t-1} + 1.68T_t + \epsilon_t,$ $\epsilon_t = -0.59\epsilon_{t-15} + e_t$	0.46 23.78
12월	모델링	$Y_t = 37.52 + 0.39Y_{t-1} + 1.25T_t + 0.36H_t - 3.48W_t + \epsilon_t,$ $\epsilon_t = 0.25\epsilon_{t-7} + 0.19\epsilon_{t-8} + e_t$	0.37 24.72
	검증	$Y_t = 63.02 + 0.21Y_{t-1} + 1.81T_t - 4.03W_t + \epsilon_t,$ $\epsilon_t = -0.41\epsilon_{t-10} + e_t$	0.21 28.84

영향을 많이 주는 것으로 나타났다. 대기변수는 8월에는 오존 (O_{t-1})이, 5월, 6월, 7월에는 이산화황 (S_{t-1}) 등 각 월별로 다르게 나타났다. 특이한 사실은 상대습도 (H_t)는 모델링기간 10월, 12월에 각각 양수인 0.32, 0.36로 나타났다. 보통 여름에는 상대습도가 높을수록 PM10 농도는 감소하는 것으로 알려져 있으나, 겨울에는 상대습도가 낮을수록 PM10 농도가 감소하는 것으로 나타났다.

적합도 척도인 결정계수 R^2 는 모델링 기간에는 0.13에서 0.49로 나타났고, 검증기간에는 0.11에서 0.47로 나타났다. 7월, 8월, 9월의 모델링기간에는 R^2 가 각각 0.49, 0.42, 0.48로 비교적 높게 나타났다. 7월 모델링기간의 예를 들면, PM10 농도 Y_t 는 Y_{t-1} , SO_2 (S_{t-1}), 일 최고온도 (T_t), 강수량 (A_t), 풍속

(W_t)에 의해 49% 설명되는 것으로 나타났다. 여기서 Y_{t-1} 과, S_{t-1} , T_t 는 양수로 나타났고, 반면에 A_t , W_t 는 음수로 나타나 이산화황, 온도가 오를수록 PM10 농도는 높아지며, 강수량, 풍속이 클수록 PM10 농도는 낮아지는 것으로 나타났고, 날씨가 무덥고 비가 많이 오는 7월에는 강수량도 PM10 농도에 영향을 주는 것으로 분석되었다.

적합도 척도로 실제값과 예측값의 차이를 나타내는 RMSE 값은 모델링기간에는 19.45에서 28.38 $\mu\text{g}/\text{m}^3$ 로 나타났고, 검증기간에는 17.94에서 28.84 $\mu\text{g}/\text{m}^3$ 로 나타났다. 본 연구에서 사용한 대기변수와 기상변수는 각각 국립환경과학원과 기상청에서 쉽게 구할 수 있는 자료이므로, 표 4.1을 이용하면, 월별 PM10 농도를 예측하는데 도움이 될 것으로 생각된다.

4.2. ARE 모형의 실제값과 추정치의 산점도

PM10 농도 실제값과 ARE 모형 추정치의 산점도를 10월의 모델링과 검증 모형을 그림 4.2에 나타내었다. RMSE가 17.94 $\mu\text{g}/\text{m}^3$ 으로 비교적 적게 나온 검증모형은 비교적 잘 적합되었다. 다만 PM10 농도가 고농도인 120 $\mu\text{g}/\text{m}^3$ 이상인 경우에는 약간 과소추정되는 경향이 있는 것으로 나타났다. 반면에, 모델링 모형의 산점도를 보면, PM10 농도가 100 $\mu\text{g}/\text{m}^3$ 이하인 경우에는 과대추정된 경향이 있고, 140 $\mu\text{g}/\text{m}^3$ 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, 실제값에 좀 더 가까이 접근하는 좋은 결과를 가질 수 있을 것으로 기대한다.

5. 결론

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

연구기간인 2003년부터 2009년의 각 월별자료 중, 2003년부터 2007년까지의 자료를 모형 구축을 위한

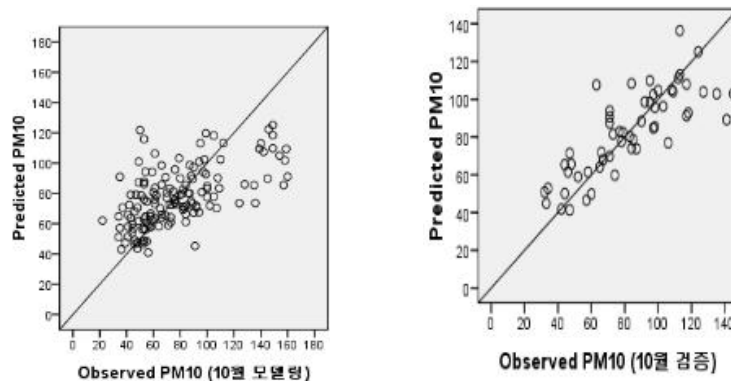


그림 4.2 PM10 농도 실제값과 ARE 모형 추정치간의 10월 모델링모형과 검증모형의 산점도

모델링 기간으로, 2008년과 2009년을 검증기간으로 ARE 모형으로 분석하였다. 모델링기간과 검증기간에 사용된 모든 설명변수들은 유의수준 10%에서 유의한 변수들만 선택했다.

월별 PM10 농도를 분석한 결과 기상변수로는 일 최고온도와 풍속이 ARE 모형에 포함된 달이 많아 PM10 농도에 영향을 많이 주는 것으로 나타났고, 대기변수는 SO_2 와 O_3 가 영향을 주는 것으로 나타났다. 적합도 척도인 결정계수 R^2 는 모델링 기간에는 0.13에서 0.49로 나타났고, 검증기간에는 0.11에서 0.47로 나타났으며 7월, 8월, 9월의 모델링기간에는 R^2 가 각각 0.49, 0.42, 0.48로 비교적 높게 나타났다. RMSE 값은 모델링기간에는 19.45에서 28.38 $\mu g/m^3$ 로 나타났고, 검증기간에는 17.94에서 28.84 $\mu g/m^3$ 로 나타났다.

ARE 모형은 표 4.1과 같이 PM10 농도에 어떤 설명변수들이 영향을 주는지 등의 수식으로 간단히 표현할 수 있는 장점이 있으므로, 본 연구에서 사용하지 못한 설명변수 들을 좀 더 첨가하여 분석하면 더 좋은 모형이 되리라 기대한다. 앞으로, PM10 농도에 사용되는 설명변수에 관한 다양한 연구와 함께 다양한 모형에 관한 연구가 지역별로 활발히 이루어져야 된다고 생각된다.

참고문헌

- 김상용, 김진아 (2009). 일반자기회귀 이분산모형을 이용한 시계열자료 분석. <한국데이터정보과학회지>, **20**, 475-483.
- 김용표 (2006). 서울의 미세먼지에 의한 대기오염. <한국대기환경학회지>, **22**, 535-553.
- 박혜련, 최기현 (2009). 서울지역에서 분전에 대한 장기추세 연구. <한국데이터정보과학회지>, **20**, 765-777.
- 이정영, 공부주, 한진석, 이민도 (2008). KZ 필터법을 이용한 서울지역 미세먼지 농도 의 장기변화 분석. <한국대기환경학회지>, **24**, 63-71.
- 이현미, 오세원 (2008). Dust monitor를 이용한 천안시 대기중 PM10, PM2.5 오염특성 조사. <한국대기환경학회지>, **24**, 367-375.
- 이형우, 이태정, 양성수, 김동술 (2008). PMF 모델을 이용한 용인·수원 경계지역에서 PM10 오염원의 확인과 상대적 기여도의 추정. <한국대기환경학회지>, **24**, 439-454.
- 이훈자 (2009). 경기도 파주시 오존농도의 통계모형 연구. <한국데이터정보과학회지>, **20**, 1085-1092.
- 정용승, 정재섭 (1991). 서울 수도권지역의 광화학 오존에 관한 연구. <한국대기보존학회지>, **7**, 169-179.
- 조신섭, 이정형 (1997). <SAS/ETS를 이용한 경제시계열 분석>, 자유아카데미.
- 주용성, 정현주, 김변준 (2009). 한국기상자료의 군집분석. <한국데이터정보과학회지>, **20**, 57-64.
- 한지현, 이미혜, 김영성 (2008). 전국 도시대기 측정망의 2000-2005년 PM10 농도 군집 분석. <한국대기환경학회지>, **24**, 300-309.
- 황윤정, 이순진, 도화석, 이윤기, 손태정, 권택규, 한정욱, 김동훈, 김종우 (2009). 대구 지역 대기 중 미세먼지 오염도 분석 및 기상인자에 따른 영향 평가. <한국대기환경학회지>, **25**, 459-471.
- 황인조, 조영혁, 최우건, 이혜문, 김태오 (2008). PMF를 응용한 구미시 PM-10 오염원 의 정량적 기여도 추정연구. <한국대기환경학회지>, **24**, 100-107.
- Chow, J. C., Watson, J. G. and Lowenlyhal, D. H. (1996). Source and chemistry of PM10 aerosol in Santa Barbara Country, CA. *Atmospheric Environment*, **30**, 1489-1499.
- Gramsch, E. F., Cereceda-Balic, Oyola, P. and Baer, D. von (2006). Examination of pollution trends in Santiago de Chile with cluster analysis of PM10 and ozone data. *Atmospheric Environment*, **40**, 5464-5475.
- Hubbard, M. and Cobourn, W. (1998). Development of a regression model to forecast ground-level ozone concentration in Louisville, KY. *Atmospheric Environment*, **32**, 2637-2647.
- Kim, E., Larson, T. V., Hopke, P. K., Slaughter, C., Sheppard, L. E. and Claiborn, C. (2003). Source identification of PM2.5 in Arid Northwest U.S. City by positive matrix factorization. *Atmospheric Research*, **66**, 291-305.
- Wise, E. K. and Comrie, A. C. (2005a). Extending the Kolmogorov-Zurbenko filter: Application to ozone, particulate matter, and meteorological Trends. *Journal of Air & Waste Management Association*, **55**, 1208-1216.
- Wise, E. K. and Comrie, A. C. (2005b). Meteorologically adjusted urban air quality trends in the Southwestern United States. *Atmospheric Environment*, **39**, 2969-2980.

Analysis of time series models for PM10 concentrations at the Suwon city in Korea

Hoonja Lee¹

¹Department of Information Statistics, Pyeongtaek University

Received 11 September 2010, revised 12 November 2010, accepted 23 November 2010

Abstract

The PM10 (Promethium 10) data is one of the important environmental data for measurement of the atmospheric condition of the country. In this article, the Autoregressive Error (ARE) model has been considered for analyzing the monthly PM10 data at the southern part of the Gyeonggi-Do, Suwon monitoring site in Korea. In the ARE model, six meteorological variables and four pollution variables are used as the explanatory variables for the PM10 data set. The six meteorological variables are daily maximum temperature, wind speed, relative humidity, rainfall, radiation, and amount of cloud. The four air pollution explanatory variables are sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3). The result showed that the monthly ARE models explained about 13-49% for describing the PM10 concentration.

Keywords: Autoregressive error (ARE) model, explanatory variable, meteorological variables, PM10 concentration, pollution variable.

¹ Associate professor, Department of Information Statistics, Pyeongtaek University, Pyeongtaek 450-701, Korea. E-mail: esther@ptu.ac.kr