

서울 지역에서 분진에 대한 장기 추세 연구

박혜련¹ · 최기현²

¹²덕성여자대학교 정보통계학과

접수 2009년 2월 12일, 수정 2009년 9월 12일, 게재확정 2009년 9월 21일

요약

본 연구에서는 대기오염물질 중에서 주목을 받고 있는 분진 직경이 10마이크로미터 (micrometer) 이내의 것을 대상으로 실제로 증가하는지 감소하는지를 알아보기 위하여 교락 요인을 제외한 분진을 대상으로 순수한 장기 추세를 연구하였다. 자료는 1996년에서 2000년까지 서울시의 기상 변수들 (최대기온, 평균습도, 최대풍속, 일사량)과 27지점에서 얻은 분진 직경이 10마이크로미터이내의 것을 이용한다. 이 자료를 이용하여 분진과 비선형 관계를 보이는 기상 변수들의 회귀 스플라인을 이용하여 계절성을 통계한 일반화 부가모형을 세웠다. 그 결과 증가가 아닌 감소하는 순수 장기 추세를 얻을 수 있었다.

주요용어: 분진, 일반화 부가모형, 회귀 스플라인.

1. 서론

대기오염물질이란 대기의 물질 중에서 인공적 또는 천연적으로 발생한 것으로 생물이나 물질에 악영향을 끼치는 미량 물질을 말한다. 대기오염물질 중 분진은 흔히 총부유분진 (total suspended particulates: TSP)에 비하여 분진 직경이 10 μ m이내의 것인 PM10 (particular matters: PM)을 말하며 이 중에서 특히 직경이 2.5 μ m이내의 분진을 미세분진 (PM2.5)이라고 한다. 게다가 아황산가스나 이산화질소가 묻쳐 형성되는 가스상 물질도 전체 분진의 40%가량을 차지한다. 분진은 오존과 더불어 증가추세에 있고 다양한 독성물질로 구성되기 때문에 건강 유해와 관련, 많은 관심을 받고 있다. 세계적 연구 결과 (Annetand와 Jiri, 2000; Bloomfield 등, 1996; Davis 등, 1998; Schwartz, 1993)를 보면 PM10의 하루 농도가 10 μ g/m³ 증가하면 1-5일 안에 사망률이 1% 정도 늘어난다. 또 PM2.5는 10 μ g/m³ 증가하면 장기적으로 심혈관계 질환으로 인한 사망은 5%, 폐암으로 인한 사망은 8% 증가한다. 네덜란드에서는 10 μ g/m³ 증가시 25-30세 인구의 잔여 수명이 51.32년에서 50.21년으로 단축된다고 발표했다. 또한 서울, 인천, 경기 등 수도권에서만 연간 1만 1000여명이 조기사망하고 있다는 연구 결과도 나오고 있는 실정이다. PM10으로 인한 초과 사망자는 교통사고 사망자의 3배꼴인 연간 약 2천명으로 추정되고 있다. 환경정책기본법 시행령 제 3조 (환경관계법규, 2008)에 의거하여 환경기준항목으로 설정되어 있는 PM10의 이런 유해성 때문에 우리나라에서는 총부유분진과 더불어 연평균 기준치는 70 μ g/m³ 이하이고 일평균 기준치는 150 μ g/m³ 이하로 설정하여 1995년 1월 1일부터 시행하고 있다. 환경부가 발표한 '2004 환경백서'에 따르면 서울시는 PM10의 오염도에서 69 μ g/m³을 기록하여 타 도시에 비해 월등히 높다. 반면 대구시는 1995년 (81 μ g/m³)부터 1999년 (66 μ g/m³)까지 연속 1위를 기록한 뒤 오염도가 현재

¹ (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 대학원 통계학과, 석사졸업.

² 교신저자: (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 정보통계학과, 교수.

E-mail: khchoi@duksung.ac.kr

까지 ($59\mu\text{g}/\text{m}^3$) 계속 낮아지고 있다. 따라서 기상관련 변수 등 알려진 다양한 교락 요인의 영향을 배제한 순수한 도시 미세먼지 수준의 추세를 추정하는 것은 중요한 이슈가 되고 있다. 이에 본 연구에서는 서울시 27지점 측정소별로 얻은 기상 자료들 중에서 교락 요인으로 판단되는 것을 선별한 후 적은 가정으로 복잡한 비선형 관계를 파악할 수 있는 일반화 부가모형 (general additive model: GAM)을 통하여 PM10의 순수한 장기 추세를 알아보려고 한다.

김호 (2000)는 대기 중 총분진과 사망자수에 관한 분석에서 일반화 가법모형을 이용하여 1996년-1997년의 서울 자료를 통해 대기 중 총분진 $100\mu\text{g}/\text{m}^3$ 증가할 때마다 초과사망자는 3%로 추정되는 결과를 얻었다. 또한 1997년-2001년의 자료로 PM10의 하루 중 변동량이 사망률에 미치는 영향에 대한 연구를 통해 PM10의 농도가 높을수록 사망률이 높다는 발표도 하였다. 조완근과 권기동 (2004)은 대구 지역 PM10 오염 관리를 위한 시간적 및 공간적 오염 특성 평가에서 PM10의 유해성을 인식, 황사의 영향으로 봄에 특히 PM10의 농도가 높은 점을 착안하여 오염의 특성을 파악해 PM10을 예방하는 연구를 하였다. 또한 김성천 등 (2000)은 군산 지역에서 PM10의 농도 및 성분 특성에 관한 연구를 하였고, 백성욱 등 (1995)는 대구지역 대기 중 PM10에 함유된 유해물질의 농도를 조사하였다. 그리고 양수미 등 (2003)는 제주지역 PM10의 오염원 규명에 관한 연구를 하였다.

본 연구에서는 1996년-2000년에 해당하는 5년간의 자료를 이용하여 영향력 있는 변수들에 대한 적절한 통계를 통해 PM10의 장기 순수 추세를 보고자 한 것으로 스플라인을 이용한 모수적 일반화 부가모형을 사용한다.

2. 자료 및 특성

대기오염자료는 서울시 자동 측정소 27지점에서 시간당 4회 정도에 걸쳐 24시간 연속적으로 측정되고 있다. 1996년 1월 1일부터 2000년 12월 31일까지 측정된 PM10을 시간별 자료 중에서 최대값을 취한 후 27지점으로 나눈 평균치를 그 날의 자료로 사용하였다. 표 2.1을 보면 후반으로 갈수록 황사의 발생일이 잦아지면서 PM10의 양이 증가하는 것을 볼 수 있다. 그러나 본 연구에서는 서울시라는 지역적 제한을 두었으므로 외부적인 요인 즉 중국으로부터 불어오는 황사에 의한 영향은 배제하기로 한다. 따라서 $300\mu\text{g}/\text{m}^3$ 을 상한으로 정하여 그 이상의 PM10 수치는 PM10의 5개년의 평균치로 대체하였다.

표 2.1 서울 지역의 황사발생 연도별 일수 및 해당일

1996년	1997년	1995년	1999년	2000년
1일	1일	13일	6일	10일
5월 5일	3월 3일	3월 28일-30일	1월 25일-28일	3월 7일
		4월 14일-22일	2월 27일	3월 23일-24일
		4월 28일	4월 5일	4월 7일-8일
				4월 26일-27일

그림 2.1의 주기적인 포물선을 통해 PM10이 계절적인 영향을 받고 있음을 알 수 있다.

기상자료 역시 같은 기간인 1996년 1월부터 2000년 12월까지 서울시 종로구에 위치하는 기상청에서 측정된 것으로 최대기온과 평균습도, 최대풍속, 일사량에 대한 자료를 이용하였다. 기상자료를 이용하는 것은 장기적인 추세변동과 계절변동을 통제했을 때 PM10의 추세에 영향을 미칠 수 있는 요인으로 생각되기 때문이다. 그림 2.2부터 그림 2.5까지는 1996년-2000년까지의 년 수와 일 년의 4계절을 고려하여 자유도를 20으로 스플라인 (spline)곡선을 그린 그래프이다.

PM10과 최대기온, 평균습도, 일사량은 음의 관계를 보이고 최대풍속과는 양의 관계를 보인다. 특히 평균습도와 최대기온의 양의 상관관계가 두드러진다. 표 2.2를 통해 온도가 낮고 일사량이 낮으면서 풍속의 세기가 약할 때 PM10의 양은 증가한다고 볼 수 있다.

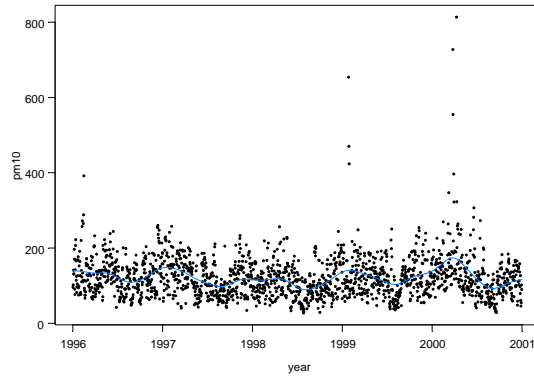


그림 2.1 1996년-2000년 일별 PM10의 분포

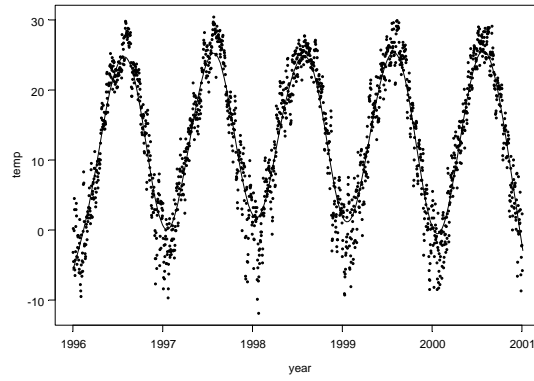


그림 2.2 1996년-2000년 일별 최대기온의 분포

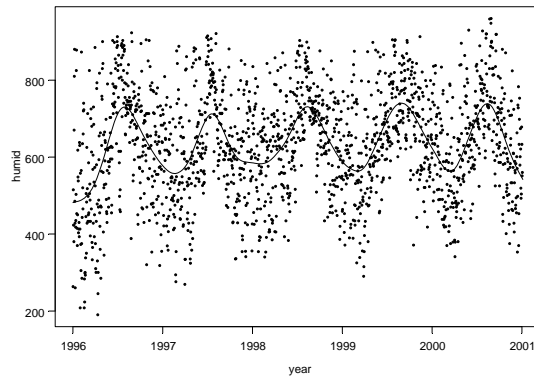


그림 2.3 1996년-2000년 일별 평균습도의 분포

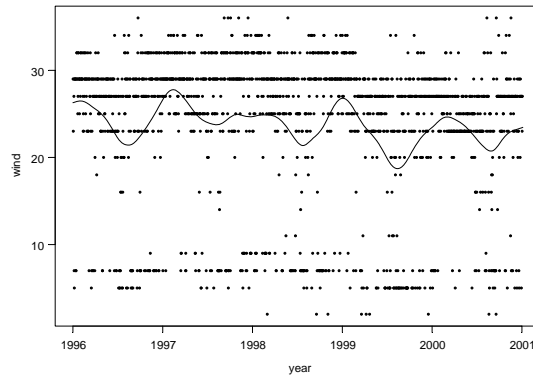


그림 2.4 1996년-2000년 일별 최대풍속의 분포

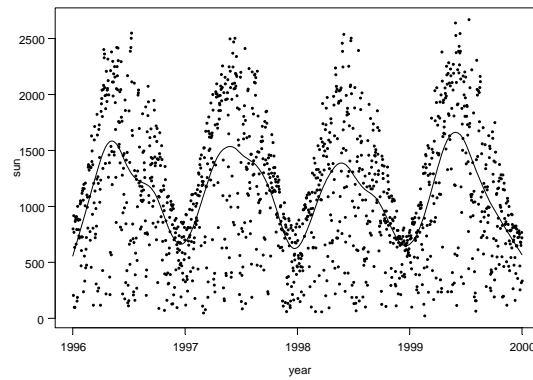


그림 2.5 1996년-2000년 일별 일사량의 분포

표 2.2 데이터의 상관관계

변수	PM 10	최대기온	평균습도	최대풍속	일사량
PM 10	1.000				
최대기온	-0.128	1.000			
평균습도	-0.043	0.465	1.000		
최대풍속	0.170	-0.219	-0.166	1.000	
일사량	-0.034	0.367	-0.413	0.121	1.000

3. 일반화 부가모형

일반화 부가모형은 다음 모형과 같이 표현할 수 있다.

$$g(\mu) = \alpha + \sum_{j=1}^p f_j(X_j)$$

여기서 α 와 f_j^1 에 대한 추정치는 아래와 같은 로컬 스코어링 (local scoring)과정으로 구해진다.

알고리즘 : 로컬 스코어링 과정

1. 초기화
2. 반복 수정된 반응변수를 아래와 같이 정의한다.

$$z_i = \eta_i^0 + (y_i - \mu_i^0) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)_0$$

여기서 $\eta_i^0 = \alpha^0 + \sum_{j=1}^p f_j^0(x_{ij})$, $\mu_i^0 = g^{-1}(\eta_i^0)$, 가중치 $w_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^{2_0} (V_i^0)^{-1}$ 이다. 추정함수 f_j^1 , 가법 예측치 η^1 , 적합값 μ_i^1 을 구하기 위해 z_i 에 가중치를 주어 적합을 시킨다. 수렴구간까지 아래의 식을 계산한다.

$$\Delta(\eta^1, \eta^0) = \frac{\sum_{j=1}^p \|f_j^1 - f_j^0\|}{\sum_{j=1}^p \|f_j^0\|}, \quad \| \cdot \| = \text{Euclidean norm}$$

3. $\Delta(\eta^1, \eta^0)$ 가 근소한 차이가 날 때까지 2.를 반복한다.

그런데 자료에 대한 $g(X_i)$ 의 적합도의 측정방법은 주로 아래와 같은 잔차 제곱합 (residual sum of squares: RSS)이다.

$$\Sigma(Y_i - g(X_i))^2$$

차이를 줄이기 위한 $g(\cdot)$ 의 함수를 정의 내리기는 쉬우나 그 함수의 형태가 유일하지 않고 회귀분석의 구조를 해석하기도 쉽지 않으므로 스플라인을 사용한다. 자료와 적합 값들의 차이가 큰 값에는 가중치를 주는 변수를 넣는다. 변동 값을 측정하는 여러 방법이 있지만 가장 손쉬운 방법으로 $s_\lambda(g) = \Sigma(Y_i - g(X_i))^2 + \lambda \|g''\|^2$ 통한 스플라인 평활법 (smoothing)은 양수인 모수 λ 를 갖는 다음 식을 최소화하는 함수로 정의된다. 여기서는 장기적인 추세변동 및 계절변동을 통제하기 위한 적절한 평활법으로 B-스플라인을 이용하였다. 시간이 0에서 T 까지인 추세를 다음처럼 정의한다.

$$f(t) = \alpha_0 + \sum_{k=1}^{K-1} \alpha_k \delta_k(t), \quad 0 \leq t \leq T$$

여기서 $\delta_1(\cdot), \dots, \delta_K(\cdot)$ 는 다음과 같다.

$$\delta_k(t) = B \frac{K}{T} (t - \tau_k) - \frac{1}{K}$$

$B(\cdot)$ 는 B-스플라인 기초함수이다. K 는 전체 나트 (knot)의 수이고 k 번째 점은 $\tau_k = T_k/K$ 로 주어진다. K 는 $\delta_k(t)$ 에 의해서 정해진다. $1/K$ 를 빼는 이유는 $f(t)$ 값의 평균을 0 근처로 만들기 위해서이다. 여기서 $k = K$ 보다는 $k = K - 1$ 까지 계산하는 이유는 적합이 좋지 않더라도 끝의 점에 연결시키기 위한 것이다. 프로그래밍인 S-plus에서 제공하는 비에스 (bs)함수 (B-스플라인, Chambers와 Hastie, 1997; Venables와 Ripley, 1999)를 사용하여 분석을 하였다.

국소적 가중 평활함수 (locally-weighted smoother)의 계산 방법은 창 (window)을 구분하여 하나의 창 내에 존재하는 자료를 이용하여 회귀분석을 한다. 모수의 조절은 창의 크기와 자료의 범위를 고려하여 조절한다. k 근처 값들을 사용한 국소적 가중 평활함수는 다음의 과정으로 구해진다.

1. x_0 의 가장 근접한 k 개를 정하고 $N(x_0)$ 으로 표기한다.

2. x_0 로부터 가장 가까운 값들의 거리를 구한다.

$$\Delta(x_0) = \max_{N(x_0)} |x_0 - x_i|$$

3. 3차 입방 가중치 함수를 사용하여 가중치 w_i 를 구한다.

$$W\left(\frac{|x_0 - x_i|}{\Delta(x_0)}\right), \quad W(u) = \begin{cases} (1 - u^3)^3, & 0 \leq u \leq 1 \\ 0, & \text{그 이외} \end{cases}$$

4. 가중치 부여된 y 의 최소제곱법으로부터 $N(x_0)$ 를 구하고 이를 이용하여 3.에서 구한 w_i 를 통해 x_0 의 추정치인 $s(x_0)$ 를 얻는다.

또한, 편차는 모형에서 잔차 제곱합의 역할을 함으로서 모수적, 비모수적인 경우 모두에서 적합도 추정이나 모형을 비교할 때 사용된다. 적합된 모형에 대한 편차는

$$D(y; \hat{\mu}) = 2l(\mu_{\max}; y) - l(\hat{\mu}; y)$$

이다. 여기서 μ_{\max} 는 포화 모형에서 $l(\mu; y)$ 를 최대화시키는 추정량이고, $\hat{\mu}$ 은 현재 모형에서 추정량이다. η_1 과 η_2 가 선형 모형이라고 가정할 경우, 적당한 정규조건을 만족하고 η_1 이 옳다는 가정 하에서

$$D(\hat{\eta}_2; \hat{\eta}_1) = D(y; \hat{\eta}_1) - D(y; \hat{\eta}_2)$$

위의 편차는 자유도가 두 모형의 차원의 차이인 χ^2 을 근사적으로 따른다. 편차 값이 작을수록 현재 모형의 설명력이 좋다는 것을 의미한다. 마지막으로 모형 적합성을 측정하는 한 가지 측도는 결정계수이다.

$$R^2 = \frac{SSR}{TSS}$$

즉, 결정계수는 총변동 (total sum of square: TSS) 중 모형으로 설명 가능 한 처리변동 (sum of square due to regression: SSR)의 비율이다. 이 비율이 1에 가까우면 가정된 모형이 데이터를 매우 잘 설명하고 있는 것이다. 반면에 이 비율이 0에 가까우면 총변동의 대부분이 오차에 기인하는 변동이 되어 모형 설정의 의미가 없어진다. 결정계수는 새로운 변수가 추가되면 그 값이 무조건 증가한다. 최소모형일 때는 그 값이 0이고 최대모형일 때는 최대값을 갖기 때문이다. 따라서 결정계수를 이용하여 최적의 모형을 찾기란 쉽지 않다. 단, 그 증가하는 폭이 작을 때 우리는 그 모형을 최적의 모형이라고 판단한다. 이에 같이 판단할 수 있는 다른 기준으로 잔차의 표준편차 $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ 를 고려한다. 즉 이것을 잔차 표준오차라고 부른다. 이 값은 항상 떨어지지 않기 때문에 어느 정도 작아질 때까지는 추정된 회귀직선이 독립변수와 종속변수와의 통계적 관계를 적절하게 설명한다고 판단할 수 있다.

4. 모형 설정

여러 기상 변수들 중에서 스플라인을 이용하여 PM10에 영향을 미칠 것이라고 판단되는 변수를 선별한 후 계절 변수와 시간 변수를 첨가하여 모수적 일반화 부가모형을 설정, PM10의 장기 순수 추세를 보고자 한다. 통상적으로 사용되는 비모수적 일반화 부가모형이 아닌 모수적 방법을 적용하는 것은 추세 도출이 용이하고 PM10과 기상 변수들 간의 비선형적 특성을 반영하기가 쉽기 때문이다. 자료 탐색을 위해 기상 변수들과 PM10의 국소적 가중 평활함수를 이용한 플롯을 한다. 이 함수를 이용하기 위한 간격 (span)을 정하기 위해 모형에서 모든 변수의 간격을 0.10부터 0.50까지 살펴본 결과 0.5부터 0.3까지

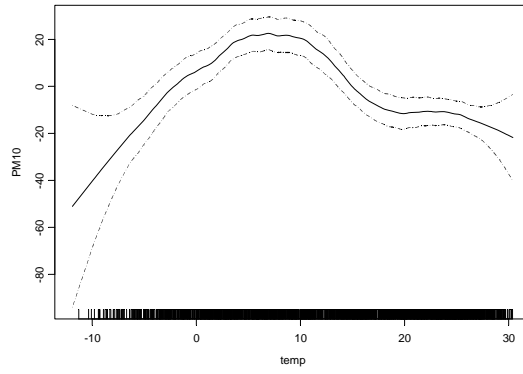


그림 4.1 PM10과 최대기온의 플롯과 95% 신뢰구간

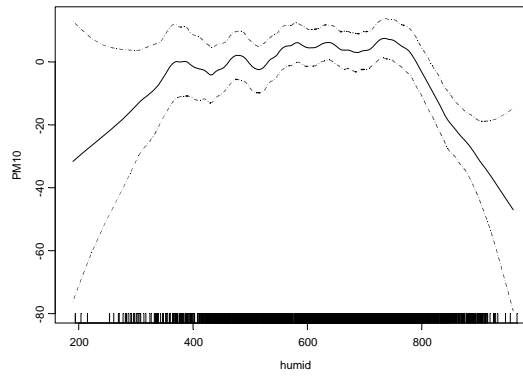


그림 4.2 PM10과 평균습도의플롯과 95% 신뢰구간

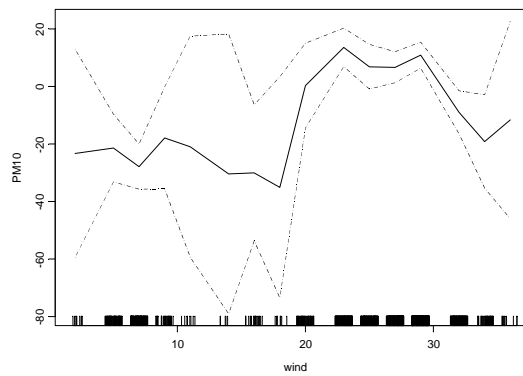


그림 4.3 PM10과 최대풍속의플롯과 95% 신뢰구간

는 감소폭이 작아지다가 0.25부터 감소폭이 크게 변화했다. 그리하여 자료의 개수, 평활 정도를 함께 고려하여 간격을 0.2로 정하였다 (그림 4.1-그림 4.4).

전반적으로 기상 변수들과 PM10과의 분명한 관계성이 있음을 알 수 있다. 통계적으로도 이 모든 변수들이 유의한 관계가 있는지 확인하기 위해 전체에서 하나씩 제거하는 방법인 후진 선택법을 사용하였다. T검증 결과 모든 변수들을 넣은 모형이 하나씩 제거한 모형들보다 편차값이 더 작으므로 모든 변수들이 유의하다고 할 수 있다. F검증도 해 본 결과 표 4.1과 같이 모든 기상 변수들이 영향력이 있다고 보인다. 여기에 기상 변수들의 특징상 이 변수들이 미치는 영향은 바로 그 날의 하루치 값이기 보다는 며칠 전 값들의 누적 효과로 나타날 수 있으므로 변수들의 지연 효과 (lag effect)를 고려해 보았다. 아래의 표 4.2를 보면 최대기온에 대한 3일 전의 지연 효과는 유의수준 10%에서 0.071로 유의하다고 판단되나 다른 변수들은 p 값이 크므로 유의하지 못하므로 영향이 거의 없다고 생각되어 제거한다.

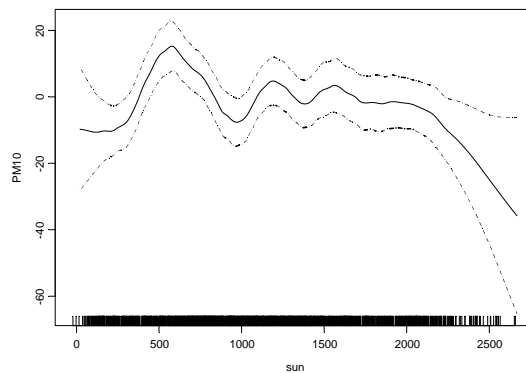


그림 4.4 PM10과 일사량의플롯과 95% 신뢰구간

표 4.1 일반화 부가모형의 기상 변수선택

변수	F 값	자유도	Pr(F)
최대기온	11.45	7.90	0.000
평균습도	7.85	8.50	0.000
최대풍속	5.16	14.00	0.000
일사량	2.54	8.20	0.009

따라서 PM10에 대한 기본 모형은 다음과 같다.

$$PM10 = \alpha_0 + \beta_1 \cdot (\text{최대기온}) + \beta_2 \cdot (\text{평균습도}) + \beta_3 \cdot (\text{최대풍속}) \\ + \beta_4 \cdot (\text{일사량}) + \beta_5 \cdot (\text{3일전 최대기온})$$

위의 그림 4.5를 보면 PM10에서 뚜렷한 계절성이 보이므로 1년의 4계절을 고려, 5년 동안의 자료이므로 자유도를 20으로 해서 계절 변수를 첨가함으로써 계절성을 조정해준다. 계절 변수를 추가한 후의 모형은 아래와 같다.

$$PM10 = \alpha_0 + \beta_1 \cdot (\text{최대기온}) + \beta_2 \cdot (\text{평균습도}) + \beta_3 \cdot (\text{최대풍속}) \\ + \beta_4 \cdot (\text{일사량}) + \beta_5 \cdot (\text{3일전 최대기온}) + \beta_6 \cdot (\text{계절성})$$

표 4.2 지연효과를 고려한 일반화 부가모형의 기상 변수선택

변수	F	df	Pr(F)
최대기온	3.42	7.90	0.001
평균습도	4.00	8.50	0.000
최대풍속	3.91	14.00	0.000
일사량	2.04	8.20	0.037
1일 전 최대기온	0.75	7.90	0.646
2일 전 최대기온	0.79	7.90	0.606
3일 전 최대기온	1.81	7.90	0.071
1일 전 평균습도	0.89	8.50	0.531
2일 전 평균습도	1.15	8.50	0.327
3일 전 평균습도	0.67	8.50	0.711
1일 전 최대풍속	1.16	14.00	0.298
2일 전 최대풍속	1.07	14.00	0.377
3일 전 최대풍속	1.28	14.00	0.214
1일 전 일사량	1.21	8.20	0.287
2일 전 일사량	0.88	8.20	0.532
3일 전 일사량	0.78	8.20	0.623

표 4.3 최종 선별된 기상 변수들

변수	F 값	자유도	Pr(F)
최대기온	6.27	7.90	0.000
평균습도	4.69	8.50	0.000
최대풍속	5.69	14.00	0.000
일사량	2.61	8.20	0.007
3일전 최대기온	1.94	7.90	0.050

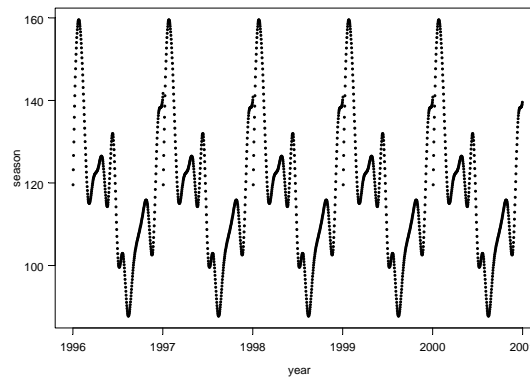


그림 4.5 1996년-2000년 PM10에 대한 계절성

시간 변수의 생성은 올리우스 즉, 일상생활에서 YYYY년 MM월 DD일과 같은 형태로 날짜를 세는 것과는 달리 어느 시점으로부터 며칠째 되는 날과 같이 날짜를 세는 방식을 이용하여 만들었다. 최종적으로 PM10에 대한 추세 변수, 계절 변수, 기상 변수들 등이 포함된 최종 모형은 다음과 같다.

$$PM10 = \alpha_0 + \beta_1 \cdot (\text{최대기온}) + \beta_2 \cdot (\text{평균습도}) + \beta_3 \cdot (\text{최대풍속}) + \beta_4 \cdot (\text{일사량}) + \beta_5 \cdot (\text{3일전 최대기온}) + \beta_6 \cdot (\text{계절성}) + \beta_7 \cdot (\text{시간})$$

표 4.4 계절성이 고려된 모형

변수	계수	표준오차	T 값	$Pr(> t)$
constant	-10.61	49.82	-0.21	0.831
bs.season1	55.29	20.77	2.66	0.008
bs.season2	13.68	14.06	0.97	0.330
bs.season3	40.61	15.92	2.55	0.011
bs.season4	-3.52	14.17	-0.25	0.804
bs.season5	-8.06	15.17	-0.53	0.595
bs.season6	-27.50	16.03	-1.72	0.086
bs.season7	-40.70	17.26	-2.36	0.018
bs.season8	-67.30	17.10	-3.93	0.000
bs.season9	-83.90	18.41	-4.56	0.000
bs.season10	-145.00	18.11	-8.01	0.000
bs.season11	-144.09	19.00	-7.58	0.000
bs.season12	-196.49	19.25	-10.21	0.000
bs.season13	-145.28	18.68	-7.75	0.000
bs.season14	-97.30	17.74	-5.48	0.000
bs.season15	-65.08	16.82	-3.89	0.000
bs.season16	-30.54	16.04	-1.90	0.057
bs.season17	-29.32	15.05	-1.95	0.052
bs.season18	31.51	16.38	1.92	0.055
bs.season19	8.69	17.68	0.49	0.623
bs.season20	27.87	17.67	1.58	0.115

표 4.5 추세 변수가 고려된 모형

변수	계수	표준오차	T 값	$Pr(> t)$
constant	21.38	48.20	0.44	0.657
bs.trend1	20.77	9.96	2.09	0.037
bs.trend2	-29.73	6.65	-4.47	0.000
bs.trend3	-25.45	9.15	-2.78	0.006
bs.trend4	22.25	7.27	3.06	0.002
bs.trend5	-36.80	8.11	-4.54	0.000
bs.season1	53.48	20.02	2.67	0.008
bs.season2	12.08	13.55	0.89	0.370
bs.season3	42.26	15.35	2.75	0.006
bs.season4	-11.95	13.69	-0.87	0.383
bs.season5	-15.03	14.64	-1.03	0.305
bs.season6	-40.49	15.51	-2.61	0.009
bs.season7	-64.48	16.80	-3.84	0.000
bs.season8	-90.11	16.64	-5.41	0.000
bs.season9	-110.35	17.96	-6.15	0.000
bs.season10	-168.48	17.692	-9.56	0.000
bs.season11	-169.82	18.52	-9.16	0.000
bs.season12	-217.40	18.67	-11.65	0.000
bs.season13	-171.23	18.24	-9.39	0.000
bs.season14	-119.45	17.27	-6.91	0.000
bs.season15	-83.93	16.27	-5.16	0.000
bs.season16	-40.66	15.55	-2.61	0.009
bs.season17	-31.67	14.58	-2.17	0.030
bs.season18	36.12	15.84	2.28	0.023
bs.season19	12.97	17.13	0.76	0.449
bs.season20	28.42	17.06	1.67	0.096

아래 그림 4.6은 원시자료에 대한 PM10의 플롯 위에 영향을 미칠 것이라 예상되는 변수들을 통제한 후의 플롯을 오버랩(overlap)시킨 것이다. PM10의 두드러진 증가나 감소를 보이지는 않으나, 미약하게나마 줄어드는 추세라고 짐작할 수 있다. 그림 4.6에서 기상 변수와 계절성을 제거한 순수 장기 추세 반응변수 y 의 값들인 y_1, \dots, y_n 을 가지는 다음의 모형을 생각하자.

여기서 f 는 미지의 함수이고, ϵ_i 는 독립이며 평균 0과 공분산 σ^2 을 가지는 정규 확률 변수들이다. 이 모형에 대하여 예측변수 t 가 반응변수 y 에 대하여 효과가 없음(no effect)을 가지는가를 검정하고자 한다. 즉 이 검정은 $H_0 : f = 0$ 인가를 검정하는 것과 동일하다. 이러한 효과가 없는 모형에 대한 검정은 고전적 적합도 검정법이지만 다른 모형들을 검정함에 있어서 기초가 되는 중요한 검정법이다. 5년

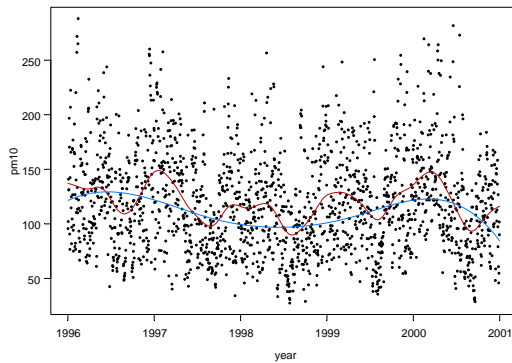


그림 4.6 기상 변수와 계절성을 제거한 순수 장기 추세 $y_i = f(t_i) + \epsilon_i, i = 1, \dots, n$.

간의 PM10에 대한 추세의 존재 여부를 근사 F 검정을 사용하여 분산 분석법 (analysis of variance: ANOVA)을 통해 확인해 본 결과 p 값이 영 (zero)임을 통해 확실한 추세가 존재함을 확인하였다. 추세에 대한 확대 그림으로 아래 그림 4.7를 보면 향후 감소할 추세라고 짐작할 수 있다.

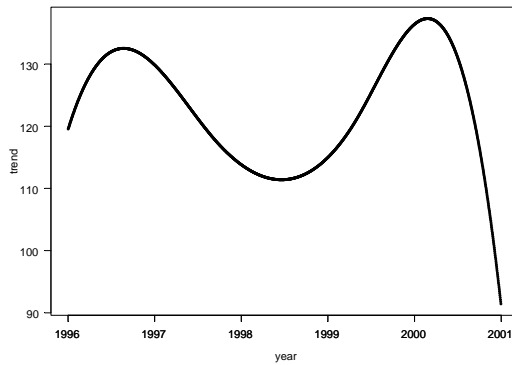


그림 4.7 확대된 추세 곡선

5. 결론

입자의 크기가 $10\mu\text{m}$ 보다 작은 입자상 물질 PM10은 도시지역에서 오존과 함께 최우선적으로 고려되는 오염물질로 알려져 있다. 왜냐하면 입자가 미세하고 가벼워서 좀처럼 침강하기 어려워 대기 중 체류시간이 길 뿐만 아니라 인체 내로 흡인되기 쉬우며 다른 유해물질을 함유할 가능성이 크기 때문이다. PM10은 자연적으로 화산폭발, 자연적인 화재, 바람 등으로 인해 발생하고, 인위적으로 화석연료를 사용하는 연소시설, 유리 및 도자기 제조시설, 금속의 용융, 제련 및 열처리 시설 등과 자동차 배출가스 및 타이어 마모에서 배출된다.

대기 위해성이 중요한 관심사가 됨에 따라 우리나라 PM10 환경기준은 연간 평균치 $70\mu\text{g}/\text{m}^3$ 이하이고 싱가포르의 $50\mu\text{g}/\text{m}^3$ 이하, 미국 캘리포니아 주가 가장 강한 연간 $30\mu\text{g}/\text{m}^3$ 로 지정되어 있다. 이처럼

PM10에 대한 정부의 적극적인 방침에도 불구하고 이에 관련된 기존 연구는 주로 오존에 따른 영향과 사망률, PM10과 PM2.5의 비교, PM10과 사망률에 대한 것이고 PM10에 영향을 줄 수 있는, 즉 교락 요인을 배제한 상태의 PM10의 순수 증가량에 대한 구체적인 연구는 미비한 실정이다. 이에 본 연구에서는 PM10이 다른 요인을 원인으로 증감하는지 아니면 실제로 증감하는지를 알아보기 위하여 중요한 혼란 변수로 생각되는 기상 변수들, 기상 변수들의 당일부터 3일 전까지의 지연 효과 변수, 계절 변수를 이용하여 연구하였다. 기상 변수들로 채택한 최대기온, 평균습도, 최대풍속, 일사량은 모두 유의한 값들이었고, 지연 효과로는 3일전 최대기온만이 유의하였다. 기상 변수는 비모수 평활방법인 국소적 가중 평활을 통해 통제된 후 예측변수에 대한 비선형적 관계를 반영시키는 모수적 평활 함수를 이용한 일반화 부가모형으로 분석하였다. 그 결과 PM10은 미약하게나마 감소하고 있는 추세를 보여주었다.

참고문헌

- 김성천, 임성호, 강달선, 송재중 (2000). 군산 지역에서 PM10의 농도 및 성분 특성에 관한 연구. <한국환경위생학회지>, **26**, 18-25.
- 김호 (2000). 대기 중 총분진과 사망자수에 관한 분석. <응용통계연구>, **13**, 265-274.
- 백성욱, 김기남, 최진수 (1995). HPLC와 형광검지기를 이용한 대기분진중 다환방향족탄화수소의 농도측정. <대한환경공학회>, **17**, 85-99.
- 양수미, 이기호, 허철구 (2003). 제주지역 미세먼지의 오염원 규명에 관한 연구. <한국환경과학회지>, **12**, 217-226.
- 조완근, 권기동 (2004). 대구지역 PM10 오염 관리를 위한 시간적 및 공간적 오염 특성 평가. <한국환경과학회지>, **13**, 27-36.
- 홍문관 편집부 (2008). 환경관계법규 IV, 홍문관.
- Annetand, P. and Jiri, S. (2000). Associations between mortality and air pollution in central Europe. *Environmental Health Perspectives*, **108**, 283-287.
- Bloomfield, P., Royle, A., Steinberg, L. J. and Yang, Q. (1996). Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmospheric Environment*, **30**, 3067-3078.
- Chambers, J. M. and Hastie, T. J. (1997). *Statistical models in S*, Chapman and Hall, London.
- Choi, K. (2008). Two-stage procedure with lower bound variance based on bootstrap. *Journal of Korean Data & Information Science Society*, **19**, 657-666.
- Davis, J. M., Brian, K. E. and Bloomfield, P. (1998). *Case Studies in environmental statistics*, Springer, New York.
- Kook, S., Han, H., Kim, G., and Choi, K. (2008). The anti-hepatotoxic effect of ginseng in rats: Meta-analysis. *Journal of Korean Data & Information Science Society*, **19**, 937-949.
- Schwarz, J. (1993). Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology*, **137**, 1136-1147.
- Venables, W. N. and Ripley, B. D. (1999). *Modern applied statistics with S*, Springer, New York.

Long term trend for particular matters in Seoul

Hyeryun Park¹ · Kiheon Choi²

Department of Information and Statistics, Duksung Women's University

Received 12 February 2009, revised 12 September 2009, accepted 21 September 2009

Abstract

Our study aimed to illustrate long term trend in 10 micrometer particular matters excluding confounding effect. Daily 10 micrometer particular matters data were measured in 27 places and meteorological data (maximum temperature, humidity and maximum wind speed, solar radiation) were obtained from the national institute of environmental research for the period from January, 1996 to December 2000. To estimate the increasing and decreasing long term trend in a set of observed data, set up the model. The model included regression spline smooth function on the time and meteorological factors to capture the seasonal time trend and any possible nonlinear relationship. The result was estimated to decrease slightly after adjusting for meteorological factors and seasonal time trend.

Keywords: Generalized additive model, particular matters, regression spline.

¹ Graduate Student, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.

² Corresponding author: Professor, Department of Information and Statistics, Duksung Women's University, Seoul 132-714, Korea. E-mail: khchoi@duksung.ac.kr