

시계열 분석을 활용한 한·중 월경성 미세먼지 연구*

김기환** · 김오석***

Analyzing Transboundary Particulate Matters of Korea and China Using Time-series Analyses*

Keewhan Kim** · Oh Seok Kim***

요약 : 본 연구의 주된 목적은 중국 수도권의 주요 도시인 베이징과 톈진의 미세먼지(PM_{10}) 농도가 서울의 PM_{10} 농도에 영향을 미치는지 미치지 않는지를 통계적으로 파악하는 것이며, 더 나아가 만약에 영향을 미친다면 중국 수도권의 PM_{10} 이 어떠한 패턴으로 서울에 영향을 미치는지 알아보는 것이다. 교차상관계수와 자기상관회귀모형을 사용하여 2014년 1월 1일부터 2017년 12월 31일까지의 서울, 베이징, 톈진의 PM_{10} 일별 농도와 서울의 평균기온, 일강수량, 평균풍속, 최다풍향, 해면기압을 분석하였다. 교차상관계수 결과에 따르면, 베이징과 톈진의 PM_{10} 은 지속적으로 서울에 영향을 미치는 것으로 나타났다. 자기상관회귀모형 결과에 따르면, 서울의 PM_{10} 은 서울 자체에서 발생한 PM_{10} 과 베이징과 톈진의 하루 전날 PM_{10} 에 영향을 받는 것으로 나타났다. 또한, 동풍이 부는 날은 PM_{10} 농도가 낮아지고 대기정체가 지속되는 날은 PM_{10} 농도가 높아지는 것을 확인했다.

주요어 : 시계열 분석, 교차상관계수, 자기상관회귀모형, 대기 오염, 중국

Abstract : The primary purpose of this research is to analyze the statistical relationship between particulate matter 10 μm or less in diameter (PM_{10}) of Korea and that of China. Specifically, the pattern of transboundary influence from China to Korea is analyzed, if the relationship turns out statistically significant. Using cross-correlation analyses and autoregressive models, data on PM_{10} of Seoul, Korea, and Beijing and Tianjin, China are analyzed in accordance with climatic data of Seoul, namely mean temperature, precipitation, average wind speed, prevailing wind directions, and average sea-level pressure. According to the results of cross-correlation analysis, PM_{10} of Beijing and Tianjin have affected the air quality of Seoul in a consistent manner; according to the result of autoregressive models, the air quality of Seoul has been affected by its domestic emissions and transboundary PM_{10} from Beijing and Tianjin that were emitted the day before. Seoul's PM_{10} tends to get lower when the wind blows from East to West but to get higher when atmospheric circulation decreases.

Key Words : Time-series analysis, Cross-correlation, Autoregressive model, Air pollution, China

I. 서론

1. 연구배경

미세먼지가 인체에 해로운 영향을 미친다는 것은 다

양한 역학조사를 통해 이미 밝혀진 과학적 사실이고, 입자의 크기가 작을수록 그 해로운 영향 정도가 큰 것으로 알려져 있다. 우리나라는 2018년 3월 초미세먼지 예보를 기존의 기준에서 선진국 수준으로 강화했고, 미세먼지 배출 규제 역시 강화하고 있으나, 배출 자체를 저감할

*본 연구는 KEI에서 수행한 “미세먼지 통합관리 전략 수립 연구”와 “중국의 미세먼지 농도추이 및 한중 간의 오염도 상관관계 분석”의 지원을 받았음.

**고려대학교 경제통계학부 국가통계전공 교수(Professor, Department of National Statistics, Korea University, korpen@korea.ac.kr)

***한국환경정책·평가연구원 부연구위원(Research Fellow, Korea Environment Institute, oskim@kei.re.kr)

수 있는 뚜렷한 정책은 제시하지 못하고 있다. 우리나라의 미세먼지는 중국의 영향이 클 것이라는 생각이 지배적이지만, 최근 중국 환경보호부의 '환경상황공보'에 의하면 중국 주요 도시의 초미세먼지는 지속적으로 감소한 것으로 나타났다. 또한, 2018년 1월 우리나라 수도권의 미세먼지 고농도 에피소드는 국외보다는 국내의 영향이 컸다는 국립환경과학원의 분석결과도 있다. 이는 미세먼지가 중국의 영향이라는 기존의 생각과는 배치되는 결과이다. 하지만, 반대로 중국의 영향이 크다는 연구 결과도 다수 있는바, 최종적으로는 한·중 월경성 미세먼지 연구 시 국내외 영향을 함께 고려하여 분석할 필요가 있다.

2. 선행연구

중국으로부터 날아 들어오는 월경성 미세먼지와 관련하여 서울과 텐진 지역의 오염물질이 서울의 미세먼지 또는 초미세먼지 농도변화에 대한 추정연구가 대표적이다(Jeong *et al.*, 2011). 이 연구에서는 PACF(Potential Source Contribution Function)와 CWT(Concentration Weighted Trajectory) 수용모형을 이용하여 중국의 영향력이 있음을 정량적으로 추정하였다. 또한 서울에 고농도 미세먼지 에피소드가 발생하였을 때로 제한되었기는 하지만 중국 오염물질의 영향력에 대한 연구도 비슷한 맥락의 결과를 나타낸다(Lee *et al.*, 2013). 도시지역이 아닌 고산 배경지역에서의 미세먼지 화학분석과 공기 궤의 이동 경로 분석을 통해 중국 대기오염물질의 장거리 이동시 발생하는 미세먼지의 물리화학적 특성 분석과 관련된 연구도 다수 수행된 바 있다(Lee *et al.*, 2007; Lim *et al.*, 2012; Han *et al.*, 2017; Kang *et al.*, 2018; Shang *et al.*, 2018). 국내 미세먼지 또는 초미세먼지에 대한 통계적 방법론을 적용한 연구는 먼저 초미세먼지에 대한 공간패널 모형을 이용한 연구(Kim, 2018), 시계열 모형을 이용한 서울의 미세먼지 농도 변화 자체를 예측한 연구(Oh *et al.*, 2018), 미세먼지로 인해 우리나라 지역별 유병률의 변화를 분석한 연구(Cho and An, 2018)가 있다.

앞에서 정리한 것과 같이 미세먼지의 중국 영향에 관한 기존의 연구들은 우리나라의 미세먼지 성분분석을 통한 중국의 영향력 존재 여부, 일시적 시점에서의 대기 모형에 의한 영향력 예측에 관한 것이었으므로 일상적으로 체감하는 미세먼지에 대한 중국과의 영향력 관계

를 밝히기에는 어려움이 있다.

한·중 간 월경성 대기오염물질 이동 연구는 대기 중의 미세먼지 농도를 직접 측정한 자료를 사용하는 것보다는 배출원 자료에 기반을 둔 대기모델링 방법론을 적용하는 것이 일반적이다. 하지만, 한국환경정책·평가연구원(Korea Environment Institute, 이하 KEI)은 최근에 들어 미세먼지 농도 자료에 기반을 둔 월경성 대기오염 연구의 중요성을 깨닫고 이에 대한 연구를 수행해왔다. 미세먼지의 배출원은 화력발전소, 산업단지, 경유차량 이동경로 등 그 종류가 다양하고, 배출원 자료에 근거한 연구는 배출원 별로 얼마만큼의 미세먼지가 발생할 것이라는 예상치를 모델링을 통해 도출하는 것이 주요 골자이다. 하지만, 이러한 예측치는 대기 중의 미세먼지 농도를 실제로 측정한 관측치와는 근본적으로 다른 것이다. 미세먼지의 배출량을 저감하는 차원에서는 배출원 별 기여도를 파악하는 것이 중요하지만, 인간에게 미치는 부정적인 건강영향을 가늠할 수 있는 지표로는 미세먼지 측정 농도가 더 유효할 수 있다.

주현수 등(2019a)은 기초 통계분석을 통해 한·중·일의 미세먼지와 초미세먼지 농도 변화의 추이를 비교하였다. 국가 별 농도의 통계적 관계를 분석하기 위해 피어슨 상관관계수(Pearson's correlation coefficient)를 활용하였고, 해당 통계량 특성 상 중국에서 한국으로 대기오염물질이 이동하는 거리 또는 시간은 별도로 고려하지 않았다. 한편, 그 후속 연구에서는 대기오염물질이 이동하는 거리 또는 지연된 시간을 고려할 수 있는 통계량인 교차상관관계수(cross-correlation coefficient)를 활용하여 중국의 여러 도시와 서울 간의 미세먼지 농도의 통계적 관계를 분석했다(주현수 등, 2019b).

피어슨 상관관계수는 양방향으로 해석이 가능하지만, 교차상관관계수는 시차를 단방향으로만 고려하므로 피어슨 상관관계수처럼 양방향 해석이 어렵다. 예를 들어, 베이징과 서울의 미세먼지 농도 자료에 피어슨 상관관계수를 적용하여 높은 상관관계수가 도출되면, 이는 베이징의 미세먼지가 서울에 영향을 미친다는 주장에 대한 근거가 될 수도 있으나, 이와는 반대로 서울의 미세먼지가 베이징에 영향을 미친다는 주장에 대한 근거도 될 수 있다. 하지만, 이와는 달리 교차상관관계수를 이용하여 베이징 미세먼지 농도가 서울에 큰 영향을 미치는 것으로 나타나면, 이 결과는 역으로 서울 미세먼지 농도가 베이징에 영향을 미친다는 주장에 대한 근거가 될 수 없으며,

역을 입증하기 위해서는 변수의 순서를 바꿔서 새로이 교차상관계수를 계산해야 한다.

3. 연구목적

본 연구는 중국 수도권의 주요 도시인 베이징과 톈진의 미세먼지(이하, PM_{10}) 농도가 서울의 PM_{10} 농도에 영향을 미치는지 미치지 않는지를 통계적으로 파악하는 것이 주된 목적이다. 더 나아가, 중국 수도권의 PM_{10} 이 어떠한 패턴으로 한국 수도권에 영향을 미치는지 알아보고자 한다.

기존의 한·중 월경성 PM_{10} 농도 비교 연구는 중국 도시 하나의 PM_{10} 농도와 서울의 PM_{10} 농도를 하나의 식을 통해 두 도시 간의 통계적 관계를 분석하고 이를 중국 7개의 도시에 반복·적용하였다면, 본 연구에서는 서울과 베이징, 톈진의 PM_{10} 농도를 하나의 식으로 적합하여 분석하였다. 즉, 주현수 등(2019b)은 베이징과 톈진 간 PM_{10} 농도의 통계적 관계를 고려하지 않았지만, 본 연구에서는 이들 도시 PM_{10} 농도 간의 관계까지 확장하여 고려했다는 점에서 차별성이 있다. 베이징과 톈진은 인접해 있으므로 이들의 PM_{10} 농도가 서울에 영향을 미치는지 미치지 않는지를 분석하기 위해서 이 둘을 연계해서 고려하는 방안이 더 합리적이다.

II. 연구방법

1. 자료

본 연구에서 사용한 PM_{10} 일별 농도($\mu g/m^3$) 자료는 2014년 1월 1일부터 2017년 12월 31일에 걸쳐 총 1461개의 관측치로 구성되어 있다. 이처럼 시간적 범위가 제한적인 이유는 2014년 이전의 중국 PM_{10} 자료 확보가 어려웠기 때문이다. 같은 기간 동안의 서울의 평균기온($^{\circ}C$), 일강수량(mm), 평균풍속(m/s), 최대풍향(degree), 해면기압(hPa) 자료도 함께 확보하여 본 연구의 시계열 분석이 기상요인을 고려할 수 있도록 설계하였다. 평균풍속은 임의의 10분 동안 평균적으로 가장 세게 불었던 풍속을 가리키며, 최대풍향은 시간 별 관측치 중 풍향 별 관측 횟수가 가장 많은 풍향을 의미한다. 자료 출처는 한국의 국립환경과학원과 서울특별시 대기환경정보, 그리

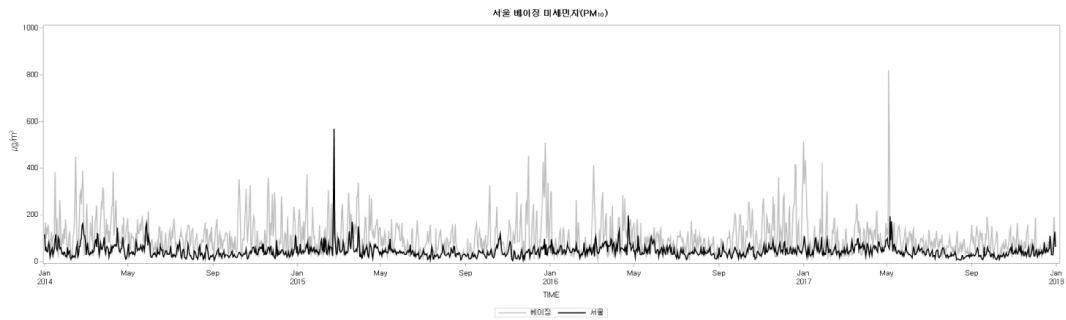
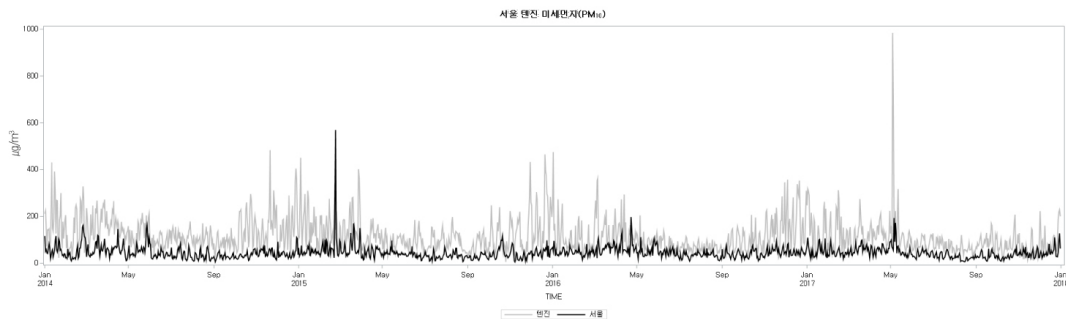
고 중국의 환경보호부와 중국환경모니터링센터 등이며, 한국환경정책·평가연구원으로부터 제공받았다(주현수 등, 2019a; 2019b).

서울과 톈진의 PM_{10} 농도 자료는 결측치가 없었으나, 베이징의 자료는 42개의 결측치(2015년 20개, 2016년 16개, 2017년 6개)가 있었다. 결측치는 과거 동일시점 관측치의 평균값으로 대체(impute)하였다. 한편, 일강수량의 원자료에서 0.0mm는 실제로 강수가 있으나 강수량이 미미한 경우를 뜻하고, null값은 당일 강수가 아예 없는 경우를 의미한다. 통계 프로그램이 null값을 결측치로 인식하는 것을 방지하기 위해 미미한 강수량은 0.01mm로, 강수가 전무한 경우는 0.0mm으로 전처리한 후, 시계열 분석을 수행하였다.

본격적인 시계열 모형 적합에 앞서, 세 도시의 시도표(time plot)를 작성하면 그림 1과 같다. 서울의 PM_{10} 농도가 중요하므로 서울을 기준으로 해서 베이징과 톈진의 PM_{10} 농도 변화를 살펴보고, 세 도시 간의 원활한 비교를 위해 시도표의 축을 통일하였다. 세로축은 PM_{10} 농도, 가로축은 시간을 가리킨다. 전반적으로 베이징과 톈진의 PM_{10} 농도가 서울 농도보다 월등히 높은 것을 확인할 수 있다. 베이징과 톈진의 PM_{10} 농도 변화는 추세를 비슷하지만 진폭은 다소 차이가 있다(그림 1).

세 도시의 PM_{10} 농도의 기초통계를 도출하면 표 1과 같다. 평균을 기준으로 해서 비교해보면, 2014년 서울의 PM_{10} 농도는 베이징의 39.29%, 톈진의 33.87% 수준, 2017년 서울의 PM_{10} 농도는 베이징의 48.40%, 톈진의 43.15% 수준임을 확인할 수 있다. 평균 이외에 최소·최대값과 표준편차를 통해서도 확인할 수 있듯이, 베이징과 톈진의 PM_{10} 농도는 서울의 농도보다 더 심각하다. 하지만, 서울의 PM_{10} 농도는 4년 동안 큰 차이가 없는 반면, 베이징과 톈진의 PM_{10} 농도는 큰 폭으로 개선되었다(표 1).

세 도시의 연월별 비교를 위해 월별 자료에 대한 수염상자를 그리면 그림 2와 같다. 좌우 수평으로 표시된 수염부분의 끝은 각각 최소값과 최대값을 나타내며, 상자부분은 4분위수를 의미한다. 다른 계절에 비해 기온이 낮은 겨울과 봄에 PM_{10} 농도가 높아지는 경향은 세 도시에서 공통적으로 발견되는 특징이다. 각 도시를 비교해보면 앞서 보았던 내용과 마찬가지로 서울의 PM_{10} 농도 변화가 베이징이나 톈진에 비해 상당히 낮은 수준임을 알 수 있고, 서울의 PM_{10} 농도 수준은 4년 동안 큰 변화가

(a) 서울(흑색), 베이징(회색) PM₁₀ 농도(b) 서울(흑색), 텐진(회색) PM₁₀ 농도**그림 1.** 서울과 베이징, 서울과 텐진의 PM₁₀ 농도 시도표(2014~2017년)**표 1.** 서울, 베이징, 텐진의 PM₁₀ 농도 연도별 기초통계(단위: $\mu\text{g}/\text{m}^3$)

도시 별 PM ₁₀ 농도		2014	2015	2016	2017
서울	최소값	6.00	5.00	9.00	7.00
	평균	45.83	45.19	47.84	43.81
	최대값	164.00	569.00	198.00	194.00
	표준편차	26.89	37.5	21.66	23.35
베이징	최소값	9.00	9.00	7.00	10.00
	평균	116.64	106.12	100.84	90.51
	최대값	449.00	510.00	415.00	820.00
	표준편차	74.56	79.13	72.99	74.86
텐진	최소값	21.00	6.00	24.00	11.00
	평균	135.33	120.13	107.67	101.54
	최대값	483.00	465.00	475.00	984.00
	표준편차	70.78	74.84	67.22	79.17

없는 반면, 베이징과 텐진의 PM₁₀ 농도는 2017년 들어 눈에 띄게 감소한 것을 확인할 수 있다(그림 2).

서울의 기상요인은 서울 PM₁₀ 농도에 큰 영향을 미치는 것으로 알려져 있다. 대기순환이 원활한 날은 그렇지

못한 날에 비해 서울 내에 적체되는 PM₁₀이 적기 때문이다. 이러한 기상학적 특징을 반영하기 위해서는 서울의 평균기온(°C), 일강수량(mm), 평균풍속(m/s), 최대풍향(degree), 해면기압(hPa)를 고려한 시계열 모형이 필요

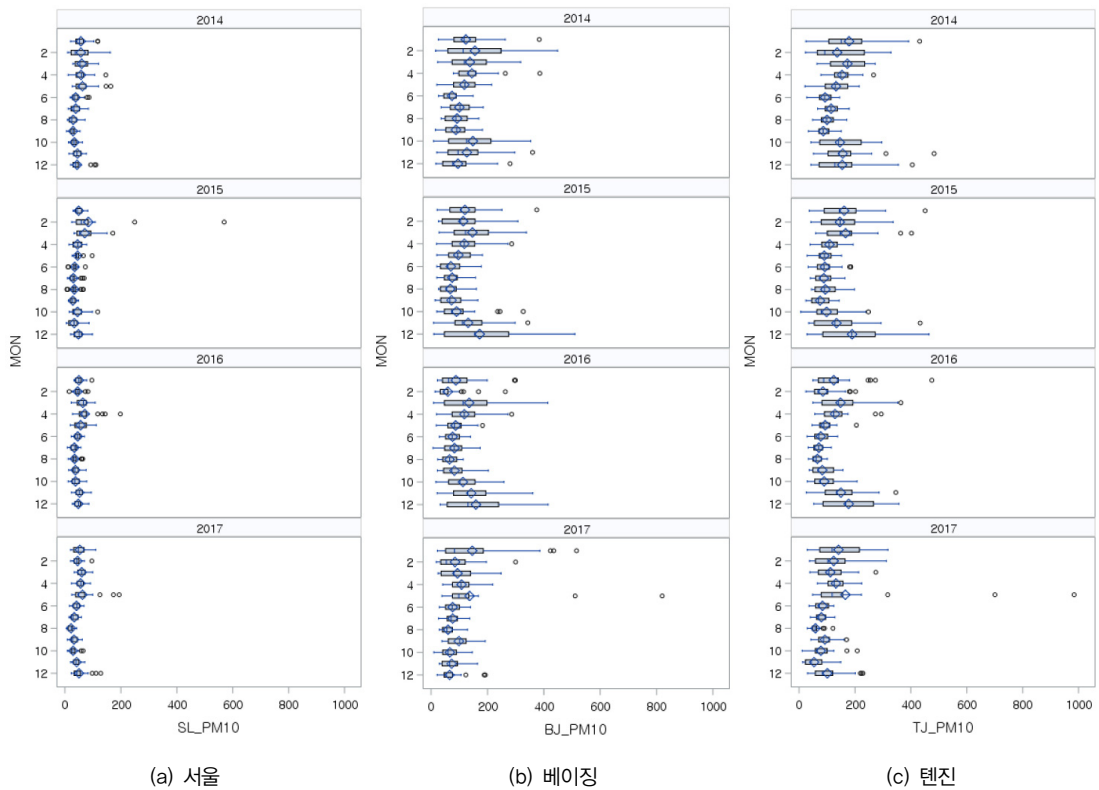


그림 2. 서울과 베이징, 텐진의 PM₁₀ 농도 수염상자(2014~2017년)

하다. 서울의 PM₁₀ 농도와 5개의 주요 기상요인을 시각화하면 그림 3과 같다. 각 그림에서 서울의 일별 기상요인의 변화와 PM₁₀ 농도 변화를 비교할 수 있도록 기상요인은 흑색(왼쪽 축)으로 표시하였으며, 서울의 PM₁₀ 농도는 회색(오른쪽 축)으로 표시하였다. 평균온도(그림 3의 (a))와 해면기압(그림 3의 (e))은 뚜렷한 주기를 보여주고 있으나, PM₁₀ 농도 변화와는 눈에 띄는 연관성을 발견하기가 어렵다. 일강수량(그림 3의 (b))과 최대풍향(그림 3의 (d)) 역시 PM₁₀ 농도 변화와는 당장에 크게 관련이 있어 보이지는 않는다. 하지만, 평균풍속(그림 3의 (c))는 전반적인 변동 추세가 PM₁₀ 농도 변화와 유사한 경향을 나타낸다.

2. 방법론

1) 교차상관계수

베이징과 텐진에서 발생한 PM₁₀이 서울에 영향을 미

치는 것이 사실이라면, 세 도시 간의 PM₁₀ 농도 변화는 연동된 움직임을 보여야하고 이는 통계 모형을 통해 설명할 수 있다. 다만, 중국 대륙에서 한반도로 PM₁₀이 이동하는데 시간이 걸리므로 이를 고려한 통계 분석이 필요하다. 즉, 시차(time lag)의 개념이 없는 일반적인 피어슨의 상관계수로는 중국과 한국 간의 PM₁₀ 영향의 연관성을 제대로 설명할 수 없다. 교차상관계(cross correlation)는 시차를 고려한 시계열적 상관관계를 의미한다. 따라서 세 도시 간의 교차상관계수를 도출하면 이들 간의 통계적 연관성을 파악할 수 있다.

교차상관계수의 정의는 다음과 같다. 두 시계열 x_t, y_t 에 대하여 시차 k 에서 교차공분산(cross covariance)는 다음과 같이 식 (1)로 정의된다.

$$\hat{\gamma}_{x,y}(k) = \frac{1}{T} \sum_{t=k+1}^T (x_t - \bar{x})(y_{t-k} - \bar{y}) \quad (1)$$

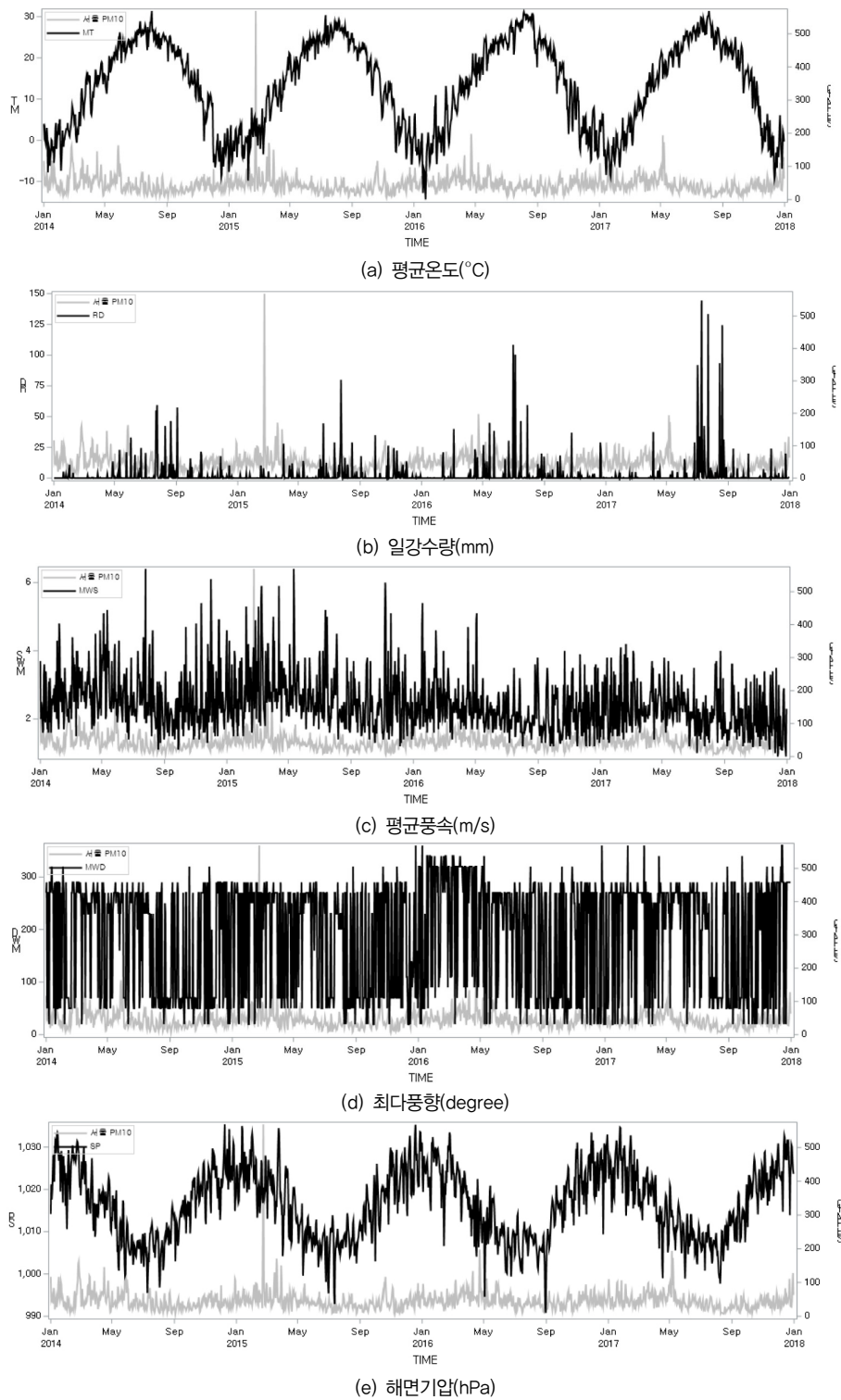


그림 3. 서울의 PM₁₀ 농도와 기상요인 변화 추이(2014~2017년)

식 (1)을 이용하여 시차 k 에서 교차상관계수는 식 (2)로 계산된다.

$$\hat{\rho}_{x,y}(k) = \frac{\hat{\gamma}_{x,y}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}, \quad -1 \leq \hat{\rho}_{x,y}(k) \leq 1 \quad (2)$$

2) 자기상관회귀모형

기존 한·중 월경성 미세먼지 농도 자료를 분석한 연구와 비교했을 때, 본 연구에서 사용하는 자기상관회귀 모형(autoregressive model)의 특징은 크게 두 가지로 정리할 수 있다. 첫째는 경험적인 방법을 통해 서울의 기상요인을 가변수화하는 부분이고, 둘째는 하나의 식을 통해서 세 도시의 미세먼지 농도 변화를 한꺼번에 통계적으로 모델링하는 것이다.

(1) 기상요인의 가변수화

기초적인 모형 적합을 통해 서울의 기상요인을 가변수화하는 것이 필요함을 파악하였다. 먼저, 베이징과 텐진의 PM₁₀ 농도만을 설명변수로 사용한 모형과 여기에 서울의 기상요인을 추가한 모형을 각각 적합해 이들의 적합도(goodness-of-fit)를 비교하면, 전자는 $R^2 = 0.417$, 후자는 $R^2 = 0.451$ 로 나타나, 두 모형의 설명력(R^2)은 큰 차이를 보이지 않음을 확인하였다. 하지만, 서울의 PM₁₀ 농도가 서울의 대기상태와 밀접한 연관이 있음은 기존 선행연구를 통해 이미 밝혀진 과학적 사실이므로, 기상요인 원자료를 그대로 모형에 사용하는 것이 적절하지 않다는 것을 알 수 있다.

일반적으로 대기정체가 심할수록 PM₁₀ 농도가 높아지는 것으로 알려져 있다. 이러한 경험을 활용하는 것이 효율적이라 판단되어 아래와 같은 과정을 통해 서울의 기상요인을 가변수화하였다.

$$D_{1,t} = \begin{cases} 1, & 5 \leq MT \leq 25 \\ 0, & \text{else} \end{cases}, \quad D_{2,t} = \begin{cases} 1, & RD = 0 \\ 0, & \text{else} \end{cases},$$

$$D_{3,t} = \begin{cases} 1, & MWS \leq 2 \\ 0, & \text{else} \end{cases}, \quad D_{4,t} = \begin{cases} 1, & 260 \leq MWD \leq 280 \\ 0, & \text{else} \end{cases},$$

$$D_{5,t} = \begin{cases} 1, & SP \geq 1020 \\ 0, & \text{else} \end{cases} \quad (3)$$

$$SD_{i,t} = \sum_{k=0}^2 D_{i,t-k}, \quad i = 1, 2, \dots, 5, \quad CD_t = \prod_{i=1}^5 SD_{i,t}$$

가변수 $D_{j,t}$, $j = 1, 2, 3, 4, 5$ 에서, j 는 순서대로 평균온도(MT), 일강수량(RD), 평균풍속(MWS), 최대풍향(MWD), 해면기압(SP)을 나타내고, t 는 시간(일)을 나타낸다. 각 가변수 $D_{j,t}$ 는 식 (3)에 따라 1 또는 0의 값을 갖는다. 4년간 서울의 PM₁₀ 농도 일평균이 $100\mu\text{g}/\text{m}^3$ 이상인 날을 추출하여 기상요인 측정치의 분포를 파악한 후, 이를 근거로 하여 가변수 $D_{j,t}$ 가 1이 되는 조건을 식 (3)과 같이 정하였다. $D_{2,t}$ 의 경우를 예로 들면, 일강수량이 없는 날은 1의 값을, 일강수량이 있는 날은 0의 값을 갖는다.

특정 기상조건이 3일간 지속되는 경우를 반영하기 위하여 $SD_{i,t}$, $i = 1, 2, 3, 4, 5$ 를 고안하였다. 이는 시차 $k = 0, 1, 2$ 에서의 $D_{j,t-k}$ 값을 더한 것이다. 예를 들어, $D_{1,t}$ (평균기온)이 3일 연속 1의 값을 갖는다면, $SD_{1,t} = 3$ 이 되며, 3일 중 2일이 1의 값을 갖는다면, $SD_{1,t} = 2$ 가 된다. 3일간 기상조건이 지속되는 경우를 고려하는 근거는 교차상관관계 분석 결과에 있다.

마지막으로 CD_t 는 $SD_{i,t}$ 의 곱으로 만들어지며 조건에 따라 $0 \leq CD_t \leq 243$ 의 범위를 갖는다. CD_t 는 서울의 PM₁₀ 농도가 극단적으로 높게 측정되는 경우도 모형에서 설명될 수 있도록 하는 역할을 한다. 그림 4는 CD_t 값의 변화(검은색, 왼쪽 축)와 서울의 PM₁₀ 농도 변화(회색, 오른쪽 축)를 함께 표시한 것이다. 모든 경우가 일치

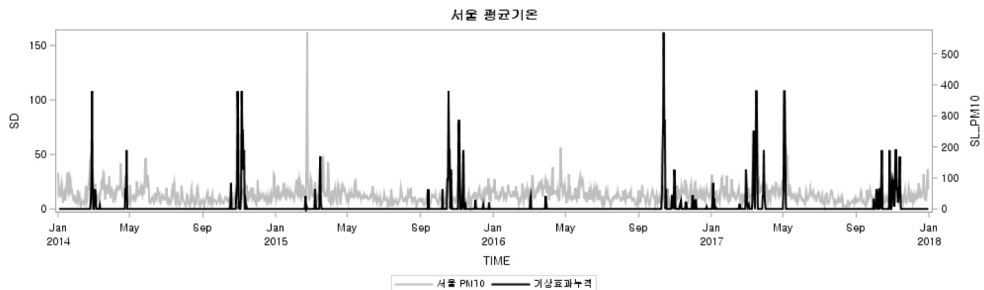


그림 4. CD_t 와 서울의 PM₁₀ 농도 변화

하는 것은 아니지만 서울 PM₁₀ 농도가 갑자기 증가하는 지점과 겹치는 것을 확인할 수 있다.

앞서 기상요인 자료의 기초통계 분석을 통해 살펴본 것들이 최다풍속의 자료 패턴은 서울의 PM₁₀ 농도 변화의 추세와 유사하므로 이를 가변수화하여 모형에 반영할 필요가 있다. 식 (3)에서도 일부 최다풍속을 반영하기는 하지만, 중국 대륙에서 불어오는 제한된 각도의 서풍만 고려하는 반면, 가변수 $W_{1,t}$, $W_{2,t}$, $W_{3,t}$ 는 북풍, 동풍, 남풍, 서풍에 따라 0과 1이 각각 적절하게 할당하여 모든 방향의 바람을 고려하였으며, 자세한 내용은 식 (4)에서 확인할 수 있다.

$$W_{i,t} = \begin{cases} W_{1,t} = 1, W_{2,t} = 0, W_{3,t} = 0, & MWD > 315 \text{ or } MWD \leq 45 \\ W_{1,t} = 0, W_{2,t} = 1, W_{3,t} = 0, & 45 < MWD \leq 135 \\ W_{1,t} = 0, W_{2,t} = 0, W_{3,t} = 1, & 135 < MWD \leq 225 \\ W_{1,t} = 0, W_{2,t} = 0, W_{3,t} = 0, & 225 < MWD \leq 315 \end{cases} \quad (4)$$

, $i = 1, 2, 3$.

(2) 최종모형

서울의 PM₁₀ 농도 변화를 설명하기 위한 최종적인 자기관화귀모형은 식 (5)와 같다. 이를 먼저 4년 동안의 전체 일별 자료에 먼저 적합하였고, 이후 연월별 자료에 적합하여 최종모형의 설명력을 확인하였다.

$$S_t = \alpha_0 + \alpha_1 S_{t-1} + \sum_{i=0}^4 \beta_i B_{t-i} + \sum_{i=0}^4 \gamma_i T_{t-i} + \delta_1 W_{1,t} + \delta_2 W_{2,t} + \delta_3 W_{3,t} + \delta_4 CD_t + \epsilon_t$$

$$\epsilon_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \theta_4 \epsilon_{t-4} + \nu_t \quad (5)$$

종속변수 S_t 는 서울에서 측정된 시점 t 의 PM₁₀ 농도를 의미하며, 독립변수 중 S_{t-1} 은 하루 전날의 PM₁₀ 농도를 뜻한다. S_{t-1} 을 독립변수군에 포함한 이유는 서울 내에서 자체적으로 발생하는 PM₁₀ 농도를 제어하기 위함이지만, S_{t-1} 에 영향을 미치는 베이징과 텐진의 PM₁₀ 농도의 효과를 완전히 배제할 수 없는 한계가 있다. B_t 와 T_t 는 각각 베이징과 텐진에서 측정된 시점 t 의 PM₁₀ 농도이다. 그림 5의 교차상관계수 결과를 반영하여 두 지역 모두 0과 4 사이의 시차를 적용하였다. $W_{i,t}$, $i = 1, 2, 3$ 은 최다풍향, CD_t 는 기상요인을 반영하기 위한 가변수이다. 오차항 ϵ_t 는 시계열적 자기상관(serial autocorrelation)을 제어하여 최종모형 적합 시 오차항의 독립성 가정을 충족시키는 역할을 하며 AR(4) 모형을 적합하였다.

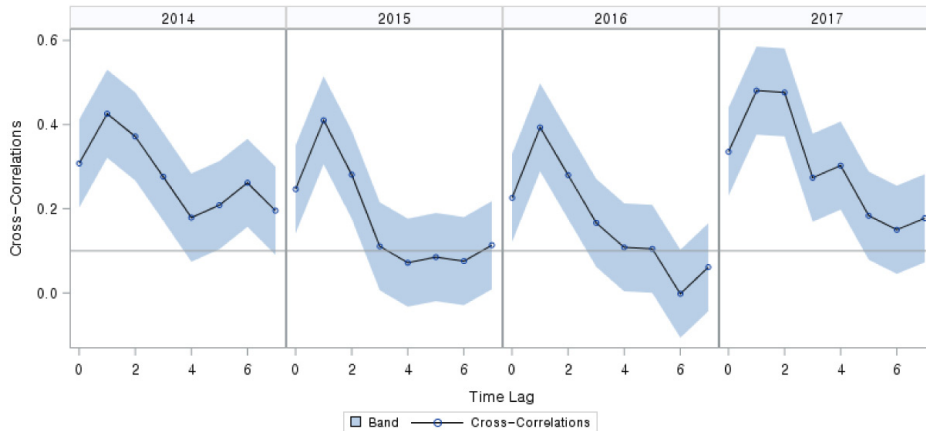
최종적으로 식 (5)의 오차항의 가정을 만족하는 최종 오차항은 ν_t 가 된다.

III. 연구결과

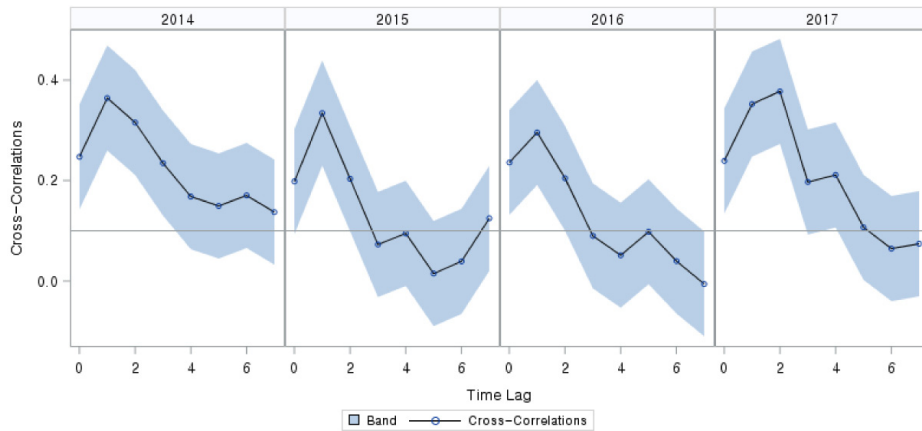
1. 교차상관계수

서울과 베이징 간의 교차상관계수는 그림 5의 (a)와, 서울과 텐진 간의 교차상관계수는 그림 5의 (b)와 같다. 식 (1)에서 x_t 는 서울의 일별 PM₁₀ 농도이며, y_{t-k} 가 베이징이나 텐진의 PM₁₀ 농도이다. 그림 5에서 4개의 칸은 각 연도(2014~2017년)를 의미하고 y 축은 교차상관계수를, x 축은 시차(0~7일)를 가리킨다. 각 시차에서 점으로 표시된 것이 식 (2)에 의해 도출된 교차상관계수이며 이를 선으로 연결하여 변화 양상을 볼 수 있도록 그림을 설계하였다. 즉 첫 번째 칸은 2014년 한 해에 걸쳐 일별 PM₁₀ 농도와 통계적 관계를 분석했을 때, 서울 PM₁₀ 농도가 베이징으로부터 어느 정도 영향을 받는지를 시차를 달리하여 나타낸 것이다. 즉, 시차가 0인 것은 베이징 PM₁₀ 농도가 당일에 직접 서울 PM₁₀ 농도에 영향을 주는 것을 의미하고, 시차가 1인 것은 베이징 PM₁₀ 농도가 그 다음 날에 서울 PM₁₀ 농도에 영향을 주는 것을 의미한다.

모든 결과에서 공통적으로 시차가 1일 때 가장 높은 교차상관계수를 보인 후, 시차가 커질수록 그 영향이 점차 줄어드는 것을 확인할 수 있으며, 시차가 0일 때도 유의한 교차상관계가 있음을 확인할 수 있다. 이는 서울의 PM₁₀ 농도가 베이징 또는 텐진의 PM₁₀ 농도와 밀접한 관계가 있음을 의미한다. 음영으로 표시된 부분은 시차별 교차상관계수의 95% 신뢰구간을 나타내며, 가로선은 교차상관계수가 0.1을 의미한다. 본 분석에서는 교차상관계수가 0.1이하인 경우, 신뢰하한이 0보다 작아지는 것으로 나타났다. 이 모두를 고려하여 결론을 내리자면, 연도별로 도시별로 다소 차이는 존재하지만, 시차 0에서부터 4까지를 고려한 교차상관계수만 유의하다고 보는 것이 적절하며, 시차 1~4에서의 결과는 시점 0과는 달리 인과관계의 의미부여가 가능하다. 즉, 당일 또는 1~4일 전 시점의 베이징이나 텐진의 PM₁₀ 농도가 현재 서울의 PM₁₀ 농도에 유의한 영향을 준다고 볼 수 있다. 한편, 이는 서울특별시 보건환경연구원에서 수행한 연구 결과와도 부합되는 내용이다. 해당 연구에 따르



(a) 서울과 베이징 간의 PM₁₀ 농도 일별 교차상관계수



(b) 서울과 톈진 간의 PM₁₀ 농도 일별 교차상관계수

그림 5. 서울과 베이징, 서울과 톈진의 PM₁₀ 농도 일별 교차상관계수

면, 베이징과 선양에서 발생한 PM₁₀이 서울에 영향을 주는데 12~30시간이 걸리는 것으로 나타났다(매일경제, 2019년 3월 6일자).

마지막으로 본 분석 결과는 일별 PM₁₀ 농도 측정치를 1년 단위로 모두 분석한 것이므로 PM₁₀ 농도가 높은 날과 낮은 날 모두를 포함한 것임을 유의해야 한다. 달리 말하면, PM₁₀ 고농도 에피소드에 국한하지 않고 PM₁₀ 농도가 낮은 날도 상시적으로 서울은 베이징과 톈진의 영향을 받고 있다는 의미이다.

2. 최종모형

SAS/ETS V9.4의 PDLREG 프로시저를 사용하여 식 (5)를 서울, 베이징, 톈진의 PM₁₀ 측정 농도 시계열 자료와

서울의 기상요인을 가변수화한 자료에 적합하였다(표 2). 모든 독립변수의 VIF(variance influence factor)가 다중공선성(multicollinearity) 무무의 기준값인 10보다 작게 나타나, 모형 적합 시 다중공선성을 염려할 필요가 없음을 확인하였다.

표 2의 첫째 열은 추정대상이 되는 모수이며, 둘째 열은 4년 전체의 일별 PM₁₀ 농도 자료를 모두 사용해 모형을 적합한 결과로 첫 행에 '전체'로 표시하였다. 이후 열은 각 연도별로 일별 PM₁₀ 농도 자료를 적합한 결과이며 첫 행에 각각의 연도로 표시하였다.

먼저 전체의 결과를 확인해 보면, 모형의 적합도인 $R^2 = 0.676$ 으로 서울의 2014년 1월 1일부터 2017년 12월 31일까지의 PM₁₀ 농도의 일별 변화 중 약 68%를 설명하고 있다. 4년간의 일별 자료를 적합하는 것임을 고려하

표 2. 최종 모형 적합 결과

Par.	전체		2014		2015		2016		2017	
	Est.	$p > t $	Est.	$p > t $	Est.	$p > t $	Est.	$p > t $	Est.	$p > t $
α_0	8.7106	<.0001	6.5497	0.0126	9.2085	0.0047	10.3809	0.0004	7.6933	0.0014
α_1	0.7024	<.0001	0.6646	<.0001	0.6864	<.0001	0.718	<.0001	0.6433	<.0001
β_0	0.0049	0.6668	0.0256	0.1804	-0.0330	0.2664	0.0274	0.1345	-0.0232	0.3265
β_1	0.0228	0.0820	0.0208	0.3287	0.0771	0.0205	-0.0008	0.9697	0.0034	0.8984
β_2	-0.0214	0.1067	-0.0198	0.3560	-0.0803	0.0180	-0.0250	0.2343	0.0401	0.1253
β_3	-0.0132	0.3186	0.0158	0.4632	-0.0098	0.7811	0.0063	0.7638	-0.0200	0.4410
β_4	0.0413	0.0004	0.0422	0.0392	0.0701	0.0242	0.0204	0.2623	0.0160	0.4569
γ_0	0.0017	0.8863	0.0211	0.3179	0.0020	0.9526	-0.0174	0.4056	0.0368	0.0888
γ_1	0.0936	<.0001	0.0673	0.0025	0.1837	<.0001	0.1089	<.0001	0.0320	0.2023
γ_2	-0.0095	0.4942	-0.0140	0.5321	-0.0683	0.0997	-0.0278	0.2371	0.0440	0.0793
γ_3	-0.0371	0.0062	-0.0090	0.6865	-0.0557	0.1545	-0.0143	0.4971	-0.0820	0.0009
γ_4	-0.0163	0.1702	-0.0417	0.0532	-0.0116	0.7287	-0.0329	0.0883	0.0382	0.0600
δ_1	1.2089	0.4084	-1.1835	0.7844	-6.6033	0.2856	2.0901	0.2925	5.8823	0.0287
δ_2	-8.3841	<.0001	-11.2270	<.0001	-8.9667	0.0027	-6.2425	0.0012	-7.1551	0.0004
δ_3	-0.2268	0.9320	-7.2989	0.2163	-13.3997	0.1490	-4.2221	0.2507	12.9989	0.0006
δ_4	0.0579	0.0826	0.0745	0.2466	0.2415	0.0333	0.0269	0.6257	0.0276	0.5834
θ_1	0.1820	<.0001	0.0483	0.5840	0.1625	0.0554	0.2003	0.0280	0.1453	0.1033
θ_2	0.2477	<.0001	0.2211	0.0029	0.2801	0.0002	0.2679	0.0019	0.1988	0.0110
θ_3	0.1217	0.0006	0.0615	0.3840	0.1905	0.0073	0.0335	0.6824	0.0525	0.4738
θ_4	0.0722	0.0226	-0.0256	0.6975	0.0896	0.1825	0.0004	0.9953	0.0703	0.2639
R^2	0.6757		0.7139		0.6379		0.6596		0.7376	

면 상당히 높은 적합도를 보인다고 할 수 있으며, 기상요인을 원자료 그대로 사용한 모형에 비하면 약 20% 이상의 설명력이 상승하였다.

유의수준 10%를 기준으로 봤을 때, 서울의 하루 전날 PM₁₀ 농도의 영향을 나타내는 α_1 의 추정값은 통계적으로 유의하며 부호가 양인 것으로 나타났다. 이는 서울 내에서 자체적으로 발생하는 PM₁₀의 영향이 통계적으로 유의함을 의미한다. 베이징과 텐진의 하루 전날 PM₁₀ 농도의 영향인 β_1 와 γ_1 의 추정값이 모두 통계적으로 유의하고 부호가 양인 것으로 나타났다. 이를 통해, 서울의 PM₁₀은 하루 전날의 베이징과 텐진의 PM₁₀ 농도에 영향을 받는 것으로 해석할 수 있다. 기상요인을 가리키는 $\delta_1 \sim \delta_4$ 의 경우, δ_2 와 δ_4 의 추정값만 통계적으로 유의하였다. δ_2 의 추정값은 최다풍향 중 동풍의 효과를 나타내는 것으로 부호가 음으로 나타났다. 이는 동풍이 불 때, 서울의 PM₁₀ 농도가 낮아지는 효과가 있음을 의미한다. δ_4 은 식 (3)에서 설정한 것처럼 4년간 서울의 PM₁₀ 농도 일평균이 $100\mu\text{g}/\text{m}^3$ 이상일 때의 기상요인을 의미한다.

δ_4 의 추정값의 부호가 양으로 나타났으므로 이 값이 커질 때 서울 PM₁₀ 농도가 증가함을 의미한다. 또한, 이 결과는 고농도 에피소드를 발생하게끔 하는 특정 기상요인이 통계적으로도 유의하게 감지가 된다는 뜻이기도 하며, 동시에 이러한 기상조건이 갖춰졌을 때, 서울의 PM₁₀ 농도가 증가함을 의미한다.

베이징과 텐진 PM₁₀ 농도의 영향을 나타내는 $\beta_2 \sim \beta_4$ 와 $\gamma_2 \sim \gamma_4$ 는 시차를 고려하여 모형에 넣었으나 효과가 결합되면서 교차상관관계 분석 결과에서 나타났던 것처럼 명확한 영향을 보여주지는 못하고 있다. 그러나 전반적으로 베이징과 텐진의 PM₁₀ 농도 변화가 시차별로 서울의 PM₁₀ 농도 변화에 양의 영향을 주고 있음을 보여주고 있으며, 부호가 음인 경우는 대부분 통계적으로 유의하지 않았다. 오차항에 적합한 AR(4) 모형의 계수 $\theta_1 \sim \theta_4$ 의 추정값은 모두 통계적으로 유의하지만 오차항의 가정과 관련된 함이므로 추정계수의 의미를 해석하지는 않는다.

연도별 적합 결과 역시 동일한 방법으로 해석할 수 있다. 전체 결과에 비하여 연도별 결과는 모수 추정값에

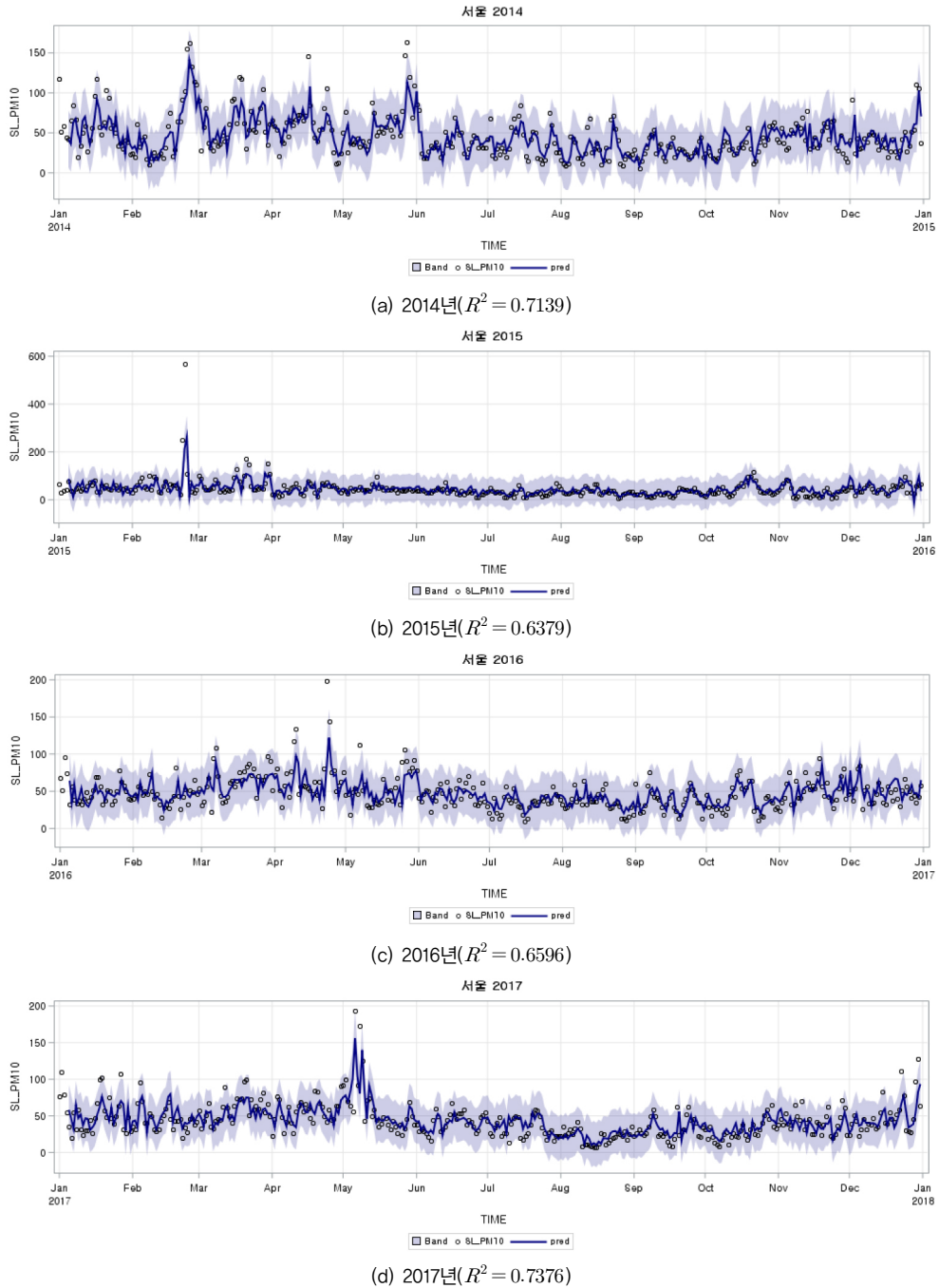


그림 6. 최종 모형을 이용한 연도별 서울 PM₁₀ 농도 적합 결과

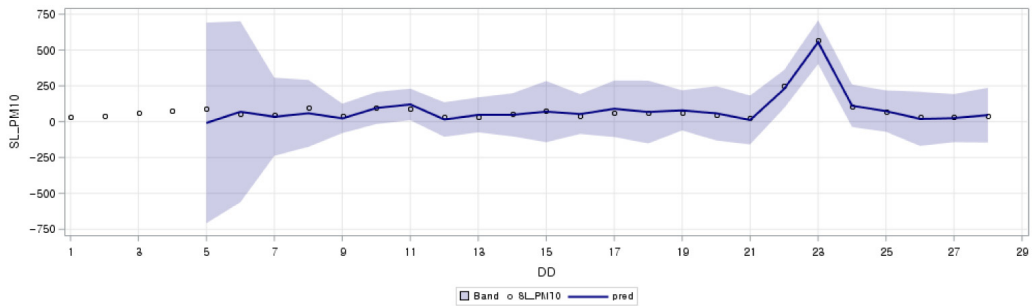
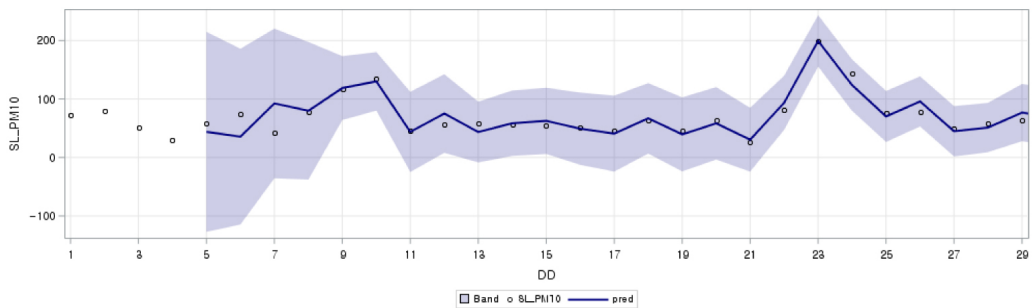
대한 통계적 유의성이 변하는 것을 확인할 수 있다. 모형의 적합도는 연도별로 다소 차이는 있으나 60% 중반에서 70% 중반 정도의 값을 보이며 이는 만족스러운 결

과로 판단된다.

그림 6은 최종 모형의 추정결과를 이용하여 연도별 자료에 적합한 결과를 시도표로 표현한 것이다. 그림 6의

표 3. 최종 모형을 이용한 연월별 서울 PM₁₀ 농도 적합 결과 적합도

연월	2014	2015	2016	2017
1	0.996273	0.994511	0.987788	0.997687
2	0.999959	0.996967	0.988831	0.999841
3	0.928069	0.950812	0.997785	0.990055
4	0.990436	0.977347	0.993819	0.965815
5	0.972327	0.925273	0.995776	0.999702
6	0.989997	0.999503	0.946864	0.993438
7	0.970352	0.971015	0.999695	0.864131
8	0.991210	0.995574	0.998657	0.994783
9	0.994086	0.998532	0.999862	0.999324
10	0.978324	0.997448	0.999175	0.999255
11	0.977497	0.966937	0.990385	0.989568
12	0.949094	0.986231	0.933304	0.983504

(a) 2015년 2월($R^2 = 0.997$)(b) 2016년 4월($R^2 = 0.994$)그림 7. 최종 모형을 이용한 서울 PM₁₀ 농도 연월별 적합 결과 시도표

(a)를 예로 들어 설명하면, 점으로 표현한 것은 서울의 2014년 일별 PM₁₀ 농도이며, 선으로 표현한 것은 모형에 의한 추정값을 나타낸다. 음영으로 표현된 것은 추정값의 95% 신뢰구간이다. 연도별 적합 시도표를 보면, 서울의 PM₁₀ 농도가 극단적으로 높아지는 경우, 적합값이 잘

따라가지 못하는 경우들이 보이지만, 대부분은 PM₁₀ 농도의 변화를 적합값이 잘 따라가고 있는 것을 확인할 수 있다. 이는 2015~2017년 자료에서도 동일하게 확인된다(그림 6의 (b)~(d)).

표 3는 최종 모형의 추정결과를 이용하여 연월별로 총

48개 모형을 적합한 결과 중 적합도(R^2)를 정리한 것이다. 2017년 8월($R^2=0.864$)를 제외하고는 모두 90% 이상의 설명력을 보여주고 있어 매우 만족스러운 결과임을 알 수 있다.

연월별 48개 모형의 추정결과를 제시하는 대신, 표 3의 적합 결과 중에서 2015년 2월과 2016년 4월만 선별하여 적합 그래프를 작성하면 그림 7과 같다. 두 시점을 선택한 이유는 그림 6에서 서울의 PM_{10} 농도가 급격히 증가하여 연도별 모형이 실제 자료의 변화를 잘 따라가지 못하는 모습을 보였기 때문이다. 그림 7의 (a)와 그림 7의 (b)에서 각각 점, 선, 음영은 연도별 모형의 적합과는 달리 적합 결과가 자료의 변동을 잘 따라가고 있는 것을 볼 수 있다.

본 분석에서처럼 장기간 많은 시점의 관측치를 포함하는 시계열 자료를 이용하여 모형을 적합하는 경우 두 가지 특징을 갖는다. 첫 번째 모형의 적합도를 나타내는 R^2 가 낮아지는 경향이 있다. 이는 자료가 많아짐으로써 너무나 다양한 변동을 포함하게 되고 이를 하나의 모형으로 설명하기 어려워지기 때문이다. 두 번째는 관측치 개수 증가로 모수 추정값의 통계적 유의성이 높아지는 경향이 생긴다. 이는 통계적 유의성을 계산하는 식의 특성에 의한 것이다. 따라서 4년 전체 자료를 이용하여 적합한 모형을 연월별 자료에 적합하는 경우, R^2 는 낮아지고 모수 추정값의 통계적 유의성은 떨어지게 된다.

IV. 결론

본 연구에서는 중국 수도권의 주요 도시인 베이징과 톈진의 PM_{10} 이 서울에 영향을 미치는지 미치지 않는지를 통계적으로 파악하였고, 중국 수도권의 PM_{10} 이 어떠한 패턴으로 서울에 영향을 미치는지 시차를 고려한 시계열 분석을 통해 알아보았다. 교차상관계수 결과에 따르면, 베이징과 톈진의 PM_{10} 은 지속적으로 서울에 영향을 미치는 것으로 나타났다. 자기상관회귀모형 결과에 따르면, 서울의 PM_{10} 은 서울 자체에서 발생한 PM_{10} 과 베이징과 톈진의 하루 전날 PM_{10} 에 영향을 받는 것으로 나타났다. 또한, 동풍이 부는 날은 PM_{10} 농도가 낮아지고 대기정체가 지속되는 날은 PM_{10} 농도가 높아지는 것을 확인했다.

서울의 PM_{10} 은 국내의 영향을 모두 받으므로 서울의

대기오염 문제가 한국이나 중국 어느 한 국가의 책임이라고만 보기에는 어렵다. 하지만, 중국으로부터의 영향은 고농도 에피소드가 발생하던 하지 않던 상시적으로 존재한다고 보는 것이 합리적이다.

참고문헌

- 이종현·김영민·김용구, 2017, “공간패널모형을 이용한 국내 초미세먼지 농도에 대한 분석” 한국데이터정보과학회지, 28(3), 473-481.
- 주현수·김소연·최민욱, 2019a, “한·중·일 3국의 미세먼지 농도 분석” KEI 포커스, 6(10), 3-23.
- 주현수·김소연·최민욱, 2019b, 「중국의 미세먼지 농도추이 및 한중 간의 오염도 상관관계 분석」, 한국환경정책·평가연구원(출간 준비 중).
- Cho, Y.E. and An, H., 2018, Analysis of the relationship between regional prevalence and the average concentration of particulate matter using beta regression, *Journal of the Korean Data Analysis Society*, 20(4), 1791-1800.
- Han, J., Lee, M., Shang, X., Lee, G., and Emmons, L.K., 2017, Decoupling peroxyacetyl nitrate from ozone in Chinese outflows observed at Gosan Climate Observatory, *Atmospheric Chemistry and Physics*, 17, 10619-10631.
- Jeong, U., Kim, J., Lee, H., Jung, J., Kim, Y.J., Song, C.H., and Koo, H., 2011, Estimation of the contributions of long range transported aerosol in East Asia to carbonaceous aerosol and PM concentrations in Seoul, Korea using highly time resolved measurements: A PSCF model approach, *Journal of Environmental Monitoring*, 13(7), 1905-1918.
- Kang, E., Lee, M., Brune, W.H., Lee, T., Park, T., Ahn, J., and Shang, X., 2018, Photochemical aging of aerosol particles in different air masses arriving at Baengnyeong Island, Korea, *Atmospheric Chemistry and Physics*, 18(9), 6661-6677.
- Kim, H., 2018, Estimating changes of causative factors' influences: Focusing on diesel, *Journal of the Korean Data Analysis Society*, 20(2), 747-757.

- Lee, M., Song, M., Moon, K.J., Han, J.S., Lee, G., and Kim, K., 2007, Origins and chemical characteristics of fine aerosols during the northeastern Asia regional experiment (Atmospheric brown cloud East Asia regional experiment 2005), *Journal of Geophysical Research*, 112, D22S29 (doi:10.1029/2006JD008210).
- Lee, S., Ho, C., Lee, Y.G., Choi, H., and Song, C., 2008, Influence of transboundary air pollutants from China on the high-PM₁₀ episode in Seoul, Korea for the period October 16-20, *Atmospheric Environment*, 77, 430-439.
- Lee, S.M., Ho, C., Lee, Y.G., Choi, H., and Song, C., 2013, Influence of transboundary air pollutants from China on the high-PM₁₀ episode in Seoul, Korea for the period October 16-20, 2008, *Atmospheric Environment*, 77, 430-439.
- Lim, S., Lee, M., Lee, G., Kim, S., Yoon, S., and Kang, K., 2012, Ionic and carbonaceous compositions of PM₁₀, PM_{2.5} and PM_{1.0} at Gosan ABC Superstation and their ratios as source signature, *Atmospheric Chemistry and Physics*, 12(4), 2007-2024.
- Moon, K., Cheo, H., Jeon, K., Yang, X., Fan, M., Kim, D., Park, H., and Kim, J., 2018, Review on the current status and policy on PM_{2.5} in China, *Journal of Korean Society for Atmospheric Environment*, 34(3), 373-392.
- Oh, J., Shin, H., Shin, Y., and Jeong, H.C., 2018, Forecasting the particulate matter in Seoul using a univariate time series approach, *Journal of the Korean Data Analysis Society*, 19(5), 2457-2468.
- Shang, X., Lee, M., Han, J., Kang, E., Kim, S.W., Gustafsson, Ö., and Chang, L., 2018, Identification and chemical characteristics of distinctive Chinese outflow plumes associated with enhanced submicron aerosols at the Gosan Climate Observatory, *Aerosol and Air Quality Research*, 18(2), 330-342.
- 매일경제, 2019년 3월 6일자, “올초 미세먼지, 중국요인이 70% 이상”
- 교신 : 김오석, 30147, 세종특별자치시 시청대로 370 세종국책연구단지 B동 1003호, 한국환경정책·평가연구원 (이메일: oskim@kei.re.kr)
- Correspondence : Oh Seok Kim, 30147, 1003 Bldg. B, 370 Sicheong-daero, Sejong Special Self-Governing City, Korea, Korea Environment Institute (Email: oskim@kei.re.kr)
- 투 고 일: 2019년 3월 29일
심사완료일: 2019년 4월 9일
투고확정일: 2019년 4월 18일