

# 미세먼지(PM<sub>10</sub>)의 지역적 대푯값 산정 방법에 관한 연구 - 서울특별시를 대상으로\*

송인상<sup>1</sup> · 김선영<sup>2\*</sup>

## Estimation of Representative Area-Level Concentrations of Particulate Matter(PM<sub>10</sub>) in Seoul, Korea\*

In-Sang SONG<sup>1</sup> · Sun-Young KIM<sup>2\*</sup>

1)

### 요 약

미세먼지(PM<sub>10</sub>)의 건강영향에 대한 많은 연구들은 정부의 대기오염 측정자료를 이용해서 악영향을 보고했다. 정부 대기오염 측정자료가 제한된 수의 측정소에서 생산되는 반면, 사망률이나 유병률과 같은 정부생산 건강결과 자료는 지역별로 집계되어 공개된다. 따라서 정부에서 생산하는 건강통계자료를 이용해서 건강영향을 분석하기 위해서는, PM<sub>10</sub> 농도의 지역적인 대푯값을 산출할 필요가 있다. 본 연구에서는 서울특별시를 대상으로 이전 연구에서 개발된 점 사상에 대한 PM<sub>10</sub> 농도 예측 모형을 이용하여 구별 대푯값을 산정하였다. 이를 위해, 세 가지 종류의 위치들을 대상으로 지점들을 생성한 후, 그 지점들에 예측한 PM<sub>10</sub> 농도의 구별 평균으로 구별 대푯값을 구했다. 세 가지 위치는 16,230개 집계구 내 가장 넓은 주거지역의 중심점, 424개 동 주민센터, 610개 1km 격자의 중심점이었다. 위치별 구별 대푯값들을 비교하기 위하여 측정치와의 관련성 및 추정치 간 관련성을 탐색하였다. 측정치와의 비교 결과, 측정치와 세 가지 구별 대푯값 추정치들 간의 관련성은 위치의 인구 대표성이 높아짐에 따라 향상되었고( $R^2=0.06-0.59$ ), 상호비교에서는 격자 중심점을 이용한 추정치가 다른 추정치들과의 관련성이 상대적으로 낮았다(0.35-0.47). 본 연구는 PM<sub>10</sub>의 지역별 평균 농도를 추정함으로써 향후 정부 통계에 기반한 전국 규모의 지역 단위 건강영향분석 연구에 기여할 수 있을 것으로 기대된다.

주요어 : 미세먼지, 회귀크리깅, 지역 대푯값, 집계구, 서울

### ABSTRACT

Many epidemiological studies, relying on administrative air pollution monitoring

2016년 9월 27일 접수 Received on September 27, 2016 / 2016년 11월 3일 수정 Revised on November 3, 2016 / 2016년 11월 22일 심사완료 Accepted on November 22, 2016

\* 이 논문은 2013년 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 (과제번호: 2013R1A6A3A0405 9017)과 서울대학교 BK21플러스 4-Zero 지향 국토공간창조 사업단의 지원으로 수행되었음.

1 서울대학교 지리학과 Department of Geography, Seoul National University

2 서울대학교 보건환경연구소 Institute of Health and Environment, Seoul National University

\* Corresponding Author E-mail : puha0@snu.ac.kr

data, have reported the association between particulate matter (PM<sub>10</sub>) air pollution and human health. These monitoring data were collected at a limited number of fixed sites, whereas government-generated health data are aggregated at the area level. To link these two data types for assessing health effects, it is necessary to estimate area-level concentrations of PM<sub>10</sub>. In this study, we estimated district (Gu)-level PM<sub>10</sub> concentrations using a previously developed pointwise exposure prediction model for PM<sub>10</sub> and three types of point locations in Seoul, Korea. These points included 16,230 centroids of the largest census output residential areas, 422 community service centers, and 610 centroids on the 1km grid. After creating three types of points, we predicted PM<sub>10</sub> annual average concentrations at all locations and calculated Gu averages of predicted PM<sub>10</sub> concentrations as representative Gu-estimates. Then, we compared estimates to each other and to measurements. Prediction-based Gu-level estimates showed higher correlations with measurement-based estimates as prediction locations became more population representative ( $R^2=0.06-0.59$ ). Among the three estimates, grid-based estimates gave lowest correlations compared to the other two(0.35-0.47). This study provides an approach for estimating area-level air pollution concentrations and assesses air pollution health effects using national-scale administrative health data.

**KEYWORDS :** Particulate Matter, Regression Kriging, Area Prediction, Census Output Area, Seoul

## 서론

대기오염이 현대사회의 주요한 환경문제로 대두되면서, 대기오염이 건강에 미치는 영향에 관한 관심 또한 증가하고 있다. 특히 여러 역학연구에서 대기 중의 미세먼지(particulate matter less than or equal to 10 micron in a diameter: PM<sub>10</sub>)가 심혈관계질환(Brook *et al.*, 2010)과 호흡기계 질환(Atkinson *et al.*, 2014), 사망(Pope III and Dockery, 2006)에 영향을 미치는 것으로 보고했다.

한국에서도 1980년대 후반부터 대기오염측정망을 운영하여 대기오염 실측 자료를 축적할 수 있게 되었으며, 이에 따라 PM<sub>10</sub>을 포함한 대기오염과 건강 결과 간의 연관성을 실증적으로 분석할 수 있게 되었다. 환경부에서는 전국 약 300개 대기오염 측정소에서 매 시간 측정한다섯 가지 법정 대기오염물질(PM<sub>10</sub>, 이산화황, 이산화질소, 일산화탄소, 오존)자료를 제공하고

있다(Ministry of Environment, 2011). 이렇게 대기오염자료가 점 측정값으로 공개되는 반면, 사망률이나 유병률과 같이 국가통계로 생산되어 공개적으로 이용 가능한 건강 결과 자료는 주로 행정구역 단위로 통합되어 공개된다. 대부분의 건강 결과 자료가 공개되는 행정구역의 최소 단위는 시군구이다. 따라서 대기오염물질의 노출에 의한 건강영향을 공공 자료를 이용하여 분석할 때에는 대기오염물질의 지역적 대푯값, 즉 시군구의 대푯값을 추정해야 할 필요가 있다.

오염물질의 지역 대푯값을 추정하기 위해 고려할 수 있는 방법으로, 측정소의 측정값을 행정구역의 오염물질 대푯값으로 할당할 수 있다. 그러나 이러한 방법은 대푯값을 추정하기에 크게 두 가지 제한점을 가진다. 첫째로 측정 장소의 위치에 따른 왜곡이 발생할 수 있다. 서울특별시 경우 25개 구(區)가 있으나, 대기오염 측정망 측정소는 2010년 기준 37개소로 구당

2개 미만의 측정소가 설치되어 있어 각 구 인 구분포의 대표성을 확보하기 어렵다. 더욱이 37개 측정소 중 12개는 교통에서 야기되는 대기오염 정도를 파악하기 위해 주요 도로변에 설치한 도로변측정소이고, 사람들의 대기오염 노출 수준을 파악하기 위해 설치한 도시대기측정소는 25개에 불과하다. 둘째로, 분석 범위를 확대할 경우 대푯값을 가지지 못하는 다수의 행정구역이 발생한다. 2010년 기준 전국 대기오염측정망의 측정소 개수는 294개로 단순히 수효만을 고려하면 251개의 시군구 개수를 상회하지만, 측정소 유무를 기준으로 보면 97개 (39%) 시군구에 측정소가 설치되어 있지 않다.

지역의 대기오염 수준을 대표할 수 있는 지역 대푯값을 추정하는 방법으로, 점 사상에 대기오염 농도를 추정하기 위해 개발된 예측 모형들을 이용할 수 있다(Jeong, 2014). 이 때 지역농도를 대표할 수 있도록, 농도를 추정할 지점을 정하는 것이 중요하다. 먼저, 우편번호 지역이나 최소단위 행정구역과 같은 지역의 중심점 등으로 지역별 단일한 인구대표점을 결정한 후, 그 지점의 오염도를 추정하여 지역 대푯값으로 정할 수 있다(Brindley *et al.*, 2005). 그러나 지역크기가 넓을 경우 단일점을 이용하게 되면 지역 내 인구 분포와 오염물질 농도의 분포 및 공간 해상도에 따라 지역 대푯값이 왜곡될 수 있다. 이러한 경우, 지역 내 인구분포를 정밀한 수준으로 대표할 수 있는 있도록 세부지역으로 구분하고 각 세부지역 내에서 점을 생성한 후, 지역별로 점들에 추정된 농도의 평균을 취할 수 있다(Hao *et al.*, 2012; Parenteau and Sawada, 2012).

본 연구에서는 기존 연구에서 개발한 점 사상에 대한 회귀크리깅(regression kriging) 기반 PM<sub>10</sub> 농도 예측 모형의 예측농도를 이용하여 서울특별시 내 25개 구에 대한 PM<sub>10</sub>의 지역적 대푯값을 추정하는 방법을 제안한다. 이때, PM<sub>10</sub> 농도를 예측할 개별 지점 위치를 세 가지로 지정해서 위치 별로 비교하여 지역 대푯값의 차이를 조사하고자 한다.

## 연구방법

### 1. PM<sub>10</sub> 농도 예측 모형

PM<sub>10</sub> 농도는 이전 연구에서 개발된 위치에 따른 회귀크리깅 기반 PM<sub>10</sub> 농도 예측 모형을 이용하여 추정하였다(Kim, 2016). 이 예측 모형은 미국 전역을 대상으로 회귀크리깅 예측모형을 개발한 선행 연구(Mercer *et al.*, 2011; Sampson *et al.*, 2013)를 참고하여 한국 자료를 이용해 개발되었다(Kim, 2016). 크리깅은 관측치의 위치자료와 관측자료를 이용해서 공간적인 자기상관관계를 포함한 공간모형을 적합한 후, 관측되지 않은 지점에 값을 예측하는 방법이다. 회귀크리깅은 일반크리깅(universal kriging)의 하나로, 문헌에서는 외부 추이를 이용한 크리깅(kriging with external drift)이라고도 부르며, 지점별로 수집된 다변량 공간자료로 설명할 수 있는 종속변수의 경향을 회귀모형으로 적합한 후 그 오차의 변이를 베리오그램(variogram) 모형으로 적합하는 방법이다(Cressie, 1993; Wackernagel, 2003; Hengl, 2007; Park *et al.*, 2012). 베리오그램 모형은 지점 간 거리에 따른 공분산함수의 공간적 자기상관 및 비공간 변이에 대한 모수로 구성된다. 먼저, 측정소 위치  $\mathbf{s}$ 에 측정된 PM<sub>10</sub> 농도  $Y(\mathbf{s})$ 에 대해 적합한 공간모형은 아래의 식 (1)과 같다.

$$Y(\mathbf{s}) = \mathbf{m}(\mathbf{s}) + \epsilon(\mathbf{s}) \quad (1)$$

$$\mathbf{m}(\mathbf{s}) = \mathbf{q}(\mathbf{s})\boldsymbol{\beta}$$

$$\epsilon(\mathbf{s}) \sim (\mathbf{0}, \boldsymbol{\Sigma}(\phi, \sigma^2, \tau^2))$$

$Y(\mathbf{s})$ : 위치  $\mathbf{s}$ 의 PM<sub>10</sub> 농도

$\mathbf{m}(\mathbf{s})$ : 설명변수  $\mathbf{q}(\mathbf{s})$ 와 회귀모형 모수  $\boldsymbol{\beta}$ 를 이용하여 회귀모형으로 적합한 평균

$\epsilon(\mathbf{s})$ : 잔차 (공간적인 잔차와 비공간적 잔차의 합)

$\phi$ : 공간적 자기상관이 유효한 거리를 나타내는 베리오그램 모형의 모수

$\sigma^2$ : 공간적 변이를 나타내는 베리오그램 모형의 모수

$\tau^2$ : 비공간적 변이를 나타내는 베리오그램 모형의 모수

이러한 공간적인 모형과 적합한 모수들을 이용해서 농도가 측정되지 않은 지점들의 위치  $\mathbf{s}_o$ 에 PM<sub>10</sub> 농도( $\hat{Y}(\mathbf{s}_o|\mathbf{s})$ )를 예측하는 식은 식 (2)와 같다.

$$\hat{Y}(\mathbf{s}_o|\mathbf{s}) = \hat{\mathbf{m}}(\mathbf{s}_o) + \hat{\Sigma}_o \hat{\Sigma}^{-1} (Z(\mathbf{s}) - \hat{\mathbf{m}}(\mathbf{s})) \quad (2)$$

$\hat{\Sigma}$ : 베리오그램 모형으로 추정된 측정지점  $\mathbf{s}$ 에 대한 공분산 행렬

$\hat{\Sigma}_o$ : 베리오그램 모형으로 추정된 측정지점  $\mathbf{s}$ 와 예측지점  $\mathbf{s}_o$ 에 대한 공분산 행렬

위 식에서  $\mathbf{Y}(\mathbf{s})$ 는 2010년 기준 서울특별시 내 37개 측정소의 PM<sub>10</sub> 연평균 농도였다. 회귀 모형 ( $\mathbf{q}(\mathbf{s})\boldsymbol{\beta}$ )에서는 전진선택법(forward selection)<sup>1)</sup>과 5분할 교차검증법(5-fold cross validation)<sup>2)</sup>을 이용하여 325개 지리변수 중 모형의 설명능력이 높고(Cross-validated  $R^2=0.77$ ) 다른 변수와의 상관관계가 0.7 이하인 5개 지리변수를 설명 변수( $\mathbf{q}(\mathbf{s})$ )로 포함하였다(Min *et al.*, 2015). 325개 지리변수는 PM<sub>10</sub>의 오염원을 대표하는 교통, 인구, 수송, 지형, 토지이용, 식생, 배출량, 고도의 8개 범주로 생성되었다(Eum *et al.*, 2015). 예를 들어 교통에 관한 변수로는 가장 가까운 주요도로로부터의 거리나 일정 반경 내의 도로 길이의 합을 포함했다. 이들 중에서 선택된 5개의 지리변수는 개별 점의 근린에서 주요 도로 길이의 합, PM<sub>10</sub> 배출량, 수면으로 분류된 토지의 비율, 건설관련 사업체 수 및 연립주택 수에 대한 변수였다(표 1). 베리오그램 모형은 지수공분산함수(exponential covariance function)를 이용해서 적합했다. 이렇게 측정소 별 PM<sub>10</sub> 연평균 농도와 5개 지리변수를 이용하여 회귀모형 및 베리오그램 모형의 모수들을 추정하였다. 추

정된 모수와 서울시 내 임의의 지점에 대하여 계산된 5개 지리변수를 이용해서 임의 예측지점의 PM<sub>10</sub> 연평균 농도를 예측하였다.

## 2. 예측 위치 생성 및 구 대푯값 추정

지역 대푯값인 구 대표 PM<sub>10</sub> 농도는 각 구(區) 경계 내에 들어오는 지점들의 예측치를 평균하여 산정하였다. 이때, 앞서 설명한 회귀크리깅 모형에서 개별 위치에 대한 예측값을 산출하므로, 지역 대푯값인 구 대표 PM<sub>10</sub> 농도를 산정하기 위하여 지역을 대표하기 위한 세 가지 종류의 예측 위치를 고려하였다.

예측 위치로는 다음 세 가지를 생성하였다. 후술할 연구결과는 예측 위치 별로 명명된 약칭 EM(Estimation Method)으로 표기한다(그림 1). EM1은 인구분포를 가장 정밀하게 고려하여 대표점을 생성하는 방법이다. 내부적으로 일정한 인구학적, 사회경제적 특성을 가진 집계구 공간자료(Kang *et al.*, 2007)에 실제 거주지에 관한 정보를 분류한 토지피복도를 중첩하였다. 면적 중위값이 0.02km<sup>2</sup>로 인구통계자료 가운데 가장 정밀한 공간단위인 집계구(Yoo and Nam, 2014)와 중첩된 거주지역들 중 가장 넓은 거주지역에 대해 중심점을 생성하였다. 그 결과 16,230개의 점을 거주인구 대표점으로 얻었다. EM2에서는 주민센터가 행정 원리에 의해 각 동의 거주 인구에게 가장 효율적으로 서비스를 제공할 수 있도록 거주인구 위치의 대표성이 있는 지점에 입지해 있음을 가정하여 2010년 기준 총 422개 점을 생성하였다. EM3은 EM1과 EM2 등 여타 방법들과 결과값을 대조하기 위해 1km 간격의 격자점을 고려하여 총 610개 점을 생성하였다. 격자점은 공간좌표만을 이용하여 공간 내에 동일한 간격으로 생성되므로 공간

TABLE 1. Five selected variables of the regression-kriging model in Seoul, Korea

- Length of major roads multiplied by number of lanes within a 25 meter buffer area
- Estimated emissions of PM<sub>10</sub> within a 10 kilometer buffer area
- Proportion of water surface land use within a 500 meter buffer area
- Number of construction companies within a 1 kilometer buffer area
- Number of townhouses within a 1 kilometer buffer area

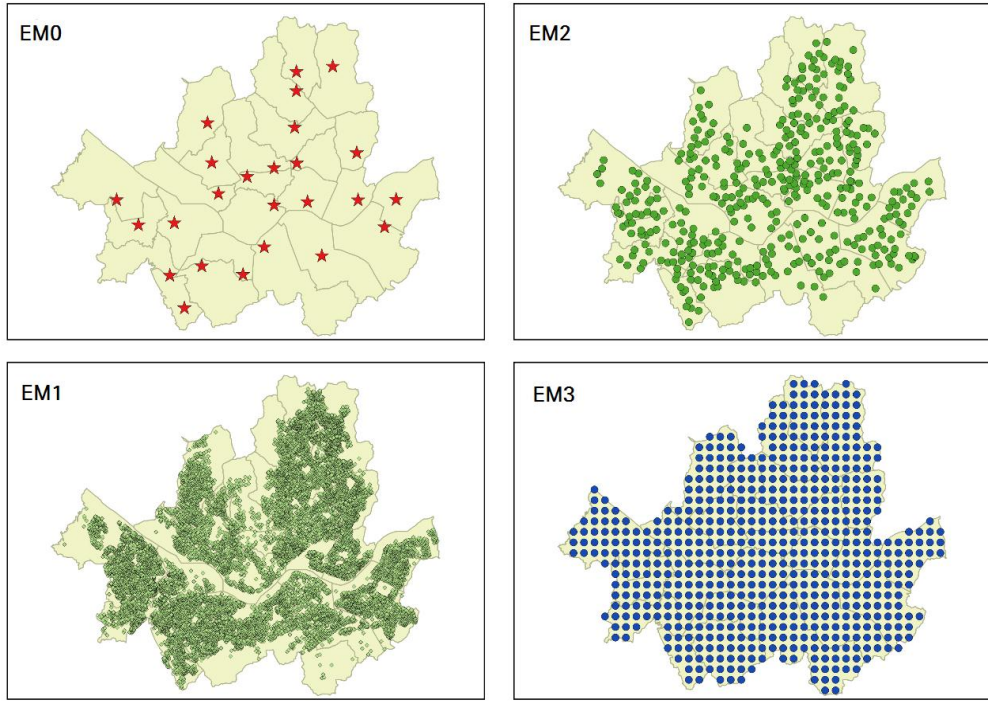


FIGURE 1. Locations of 25 air pollution regulatory monitoring sites (EM0), 16,230 centroids of the largest residential areas of census output areas (EM1), 422 community service centers (EM2), and 610 1-km grid coordinates (EM3) in Seoul, Korea

적인 대표성을 가진다. 반면 격자점이 인구분포와 무관하게 생성되므로, 해당 위치에 추정된 PM<sub>10</sub> 농도가 다른 방법에 의한 예측치와 달리 PM<sub>10</sub> 노출의 측면에서 무작위한 값을 보일 수 있다(Stehman and Overton, 1996).

### 3. 이용 자료

인구분포 재현에 필요한 공간 자료는 2010년 인구주택총조사 기준 집계구 경계 자료로 통계지리정보시스템으로부터 구득했다. 거주지역 식별을 위해 사용한 중분류 토지피복도<sup>3)</sup>는 환경부 환경지리정보시스템으로부터 구득했다. 이 자료는 2007년에서 2008년 사이에 촬영된 위성영상자료를 감독 분류한 후 실제 토지이용 현황과 대조하여 제작된 자료이다(Ministry of Environment, 2016). 본 연구에서는 실제 거주가 이루어지는 지역을 추출하기 위하여 용도지역 코드상 주거지

역으로 분류된 토지피복도를 이용했다.

회귀크리깅 모형을 이용해서 세 가지 위치에 PM<sub>10</sub> 농도를 예측하기 위해 각 위치에 대한 지리변수 자료를 계산했다. 5개 지리변수는 도로망, 토지피복도, 집계구별 주택 및 사업체 수, 격자별 배출량 등의 공간자료로부터 공간 연산<sup>4)</sup>으로 계산되었다(Eum *et al.*, 2015). 공간 연산은 ArcGIS 10.2 및 R 통계 패키지를 이용해서 수행했다. 대기오염측정자료는 국립환경과학원에서 2010년 1년 간 1시간 단위로 제공한 측정값을 정제한 후 계산한 연평균 PM<sub>10</sub> 농도를 이용했다(Yi *et al.*, 2016). 정제 시에는 측정장비 오류에 의한 결측 및 이상치를 제거한 후, 매일 18시간 이상 측정되지 않아 자료의 시간적 연속성을 만족하지 못하는 날을 제외했다. 또한 연간 측정결과가 연간 일수의 75% (243일) 이하로 수집되었거나 45일 이상 장비

작동이 중지된 측정소도 제외했다.

#### 4. 추정방법별 PM<sub>10</sub> 대푯값 비교

추정한 PM<sub>10</sub> 구별 대푯값을 실제 측정값 및 위치 종류 별로 비교하였다. 추정된 PM<sub>10</sub> 구별 대푯값은 세 가지 위치(EM1~EM3)를 이용해서 지점에 예측된 PM<sub>10</sub> 농도의 구별 평균값이었다. 실측 구별 대표값은 구별로 하나씩 설치된 도시대기측정소(EMO)의 측정 자료였다. 추정값과 측정값의 비교 외에, 세 가지 위치 생성 방법들 간 추정값들도 비교하였다. 구 대푯값들 간 비교는 산점도와 평균제곱오차를 이용한 결정계수(r-squared value: R<sup>2</sup>)를 이용했다. 평균제곱오차(mean square error: MSE)에 근거한 결정계수(MSE-based R<sup>2</sup>)는 평균제곱오차를 측정치의 분산으로 나눈 값을 1에서 차감해서 계산하며, 식 (3)와 같이 계산했다.

$$R^2 = \frac{1 - \text{MSE}(\hat{Y})}{\text{Var}(Y)} = 1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2} \quad (3)$$

상관계수의 제곱으로 계산하는 결정계수가 회귀선을 근거로 비교하는 반면, 평균제곱오차를 이용한 결정계수는 일치선을 근거로 비교한다(Keller *et al.*, 2015; Kim *et al.*, 2016).

구별 대푯값의 방법별 비교는 일치성을 탐색하기 때문에, 평균제곱오차에 근거한 결정계수를 이용하는 것이 합리적이다. 다만 추정치를 상호 비교할 경우, (3)에서 어떤 추정치를 기준으로 분산을 계산하는가에 따라 결정계수가 달라진다. 따라서 두 추정치에 대해 교대로 분산을 구해 계산한 두 결정계수를 평균하여 최종 결정계수를 산출했다. 또한 25개 구별로 예측위치 종류들 간 예측값 분포차이를 탐색하기 위해, 추정값과 측정값 및 추정값들 간의 농도 차이의 절댓값을 탐색했다.

## 연구 결과 및 고찰

### 1. 구 대푯값의 경향

25개 구별 대푯값들의 평균은 측정값인 EMO가 48.99 $\mu\text{g}/\text{m}^3$ , 추정값인 EM1, EM2, EM3가 각각 49.04, 49.76, 49.70 $\mu\text{g}/\text{m}^3$ 로, 측정값이 추정값에 비해 약간 낮았다(표 2). 표준편차는 EMO가 3.24 $\mu\text{g}/\text{m}^3$ 이고 EM1, EM2가 각각 2.26, 2.10 $\mu\text{g}/\text{m}^3$ 로 추정값이 더 낮은 변이를 보였으나, EM3은 3.54로 측정값보다 변이 정도가 더 컸다. 추정된 대푯값들(EM1~EM3)과 측정값(EMO) 간 차이의 절댓값의 최솟값과 최댓값은 0.06과 6.15 $\mu\text{g}/\text{m}^3$ 이었고, 평균과 표준편차는 1.79와 1.28 $\mu\text{g}/\text{m}^3$ 이었다.

TABLE 2. Summary statistics of Gu-average estimates of PM<sub>10</sub> annual average concentrations ( $\mu\text{g}/\text{m}^3$ ) based on predictions at three types of locations (EM1 to EM3) and observations at regulatory monitoring sites (EMO) in Seoul, Korea

Approach	Minimum	Median	Maximum	Mean	SD
EMO	40.65	48.88	54.53	48.99	3.24
EM1	43.80	49.81	51.77	49.04	2.26
EM2	45.02	50.65	52.42	49.76	2.10
EM3	42.26	50.06	55.06	49.70	3.54
EM1-EM0*	0.15	1.60	4.18	1.77	1.12
EM2-EM0*	0.06	1.89	5.29	1.93	1.32
EM3-EM0*	0.46	2.68	6.15	2.74	1.56
EM2-EM1*	0.06	0.65	1.67	0.75	0.50
EM3-EM1*	0.14	1.32	4.50	1.56	1.21
EM3-EM2*	0.30	1.27	4.12	1.78	1.07

\*Absolute values

### 2. 구 대푯값의 상관성

세 가지 위치에 근거해 예측한 PM<sub>10</sub> 농도의 구 대푯값(EM1~EM3)들은 모두 측정값 (EM0) 과 양의 관련성 ( $R^2=0.06-0.59$ )을 보였다(그림 2). EM3은 다른 예측값에 근거한 추정치에 비해 상당히 낮은 관련성을 나타냈다 ( $R^2=0.06$ ) (그림 2). 예측값에 근거한 대푯값들은 낮은 농도에서는 대체로 측정치에 비하여 높은 값을 가졌으나,  $50\mu\text{g}/\text{m}^3$  이상의 높은 농도에서는 측정값에 비해 낮은 값을 가지는 경향을 보였다.

세 가지 위치에 근거한 추정치들을 쌍으로 비교했을 때, 집계구 및 토지피복도를 이용한 EM1 과 동 주민센터 지점을 이용한 EM2 간의 관련

성은 결정계수가 0.8 이상으로 매우 강한 양의 관계를 보였다(그림 3). 동 주민센터와 격자를 이용한 EM2와 EM3는 집계구 및 토지피복도를 이용한 EM1에 비해 대체로 PM<sub>10</sub> 농도 대푯값을 크게 추정했다. EM1과 EM2는 EM3과 양의 관계를 나타냈으나, 결정계수 평균이 각각 0.35와 0.47로 연관성은 EM1과 EM2 사이의 연관성 ( $R^2=0.83$ )에 비해서는 상대적으로 낮았다.

### 3. 대푯값들의 구별 분포 특성

PM<sub>10</sub> 농도 대푯값은 구별로 차이를 보였다 (그림 4). 추정된 대푯값들(EM1~EM3)과 측정값(EM0) 간 차이의 절댓값은 관악구(EM3-

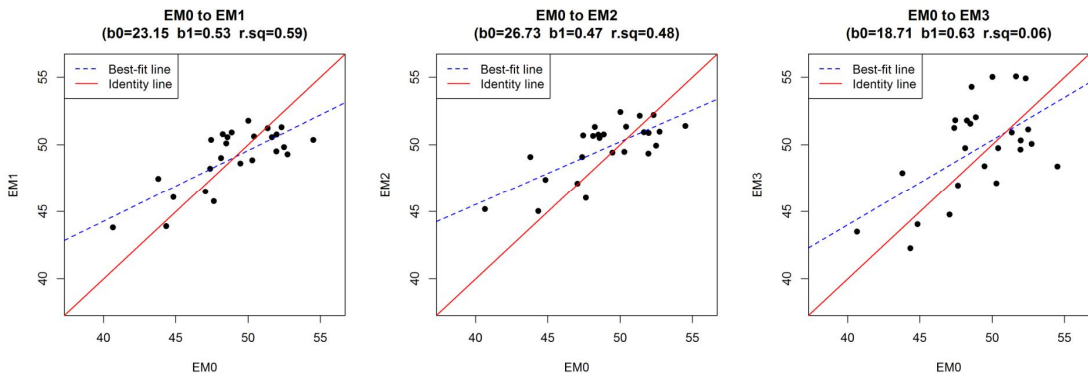


Figure 2. Scatter plots of Gu-average estimates of PM<sub>10</sub> annual averages ( $\mu\text{g}/\text{m}^3$ ) based on observations (EM0) and predictions at three types of locations (EM1 to EM3)

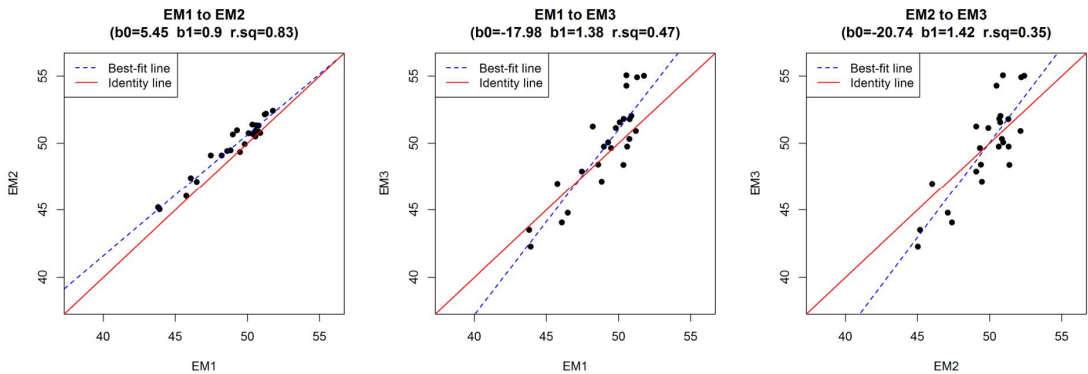


FIGURE 3. Scatter plots of pairs of three sets of prediction-based Gu-average estimates of PM<sub>10</sub> annual averages ( $\mu\text{g}/\text{m}^3$ ) (EM1 to EM3)



EM0), 성동구(EM3-EM0), 성북구(EM2-EM0), 관악구(EM2-EM0)에서 5 $\mu\text{g}/\text{m}^3$  이상으로 가장 크고, 구로구(EM2-EM0)와 서초구(EM2-EM0)에서 1 $\mu\text{g}/\text{m}^3$  이하로 가장 작았다. 세 추정방법과 측정값 간 차이 절댓값의 평균값은 관악구, 성북구, 광진구, 노원구, 성동구, 용산구에서 3 $\mu\text{g}/\text{m}^3$  이상으로 가장 높아, 이들 여섯 개 구에서는 어떤 추정방법을 사용하더라도 전반적으로 측정값과의 차이가 큼을 보였다. 반면, 구로구, 서초구, 중랑구, 은평구에서는 평균 절댓값 차이가 1 $\mu\text{g}/\text{m}^3$  이하로, 전반적으로 측정값과의 차이가 작았다.

추정된 대푯값들(EM1~EM3) 간 차이의 절댓값의 최솟값과 최댓값은 0.06과 4.50 $\mu\text{g}/\text{m}^3$  이었고, 최댓값은 영등포구에서 나타났다. 평균과 표준편차는 1.36과 1.06 $\mu\text{g}/\text{m}^3$ 로, 추정값들 간 차이는 측정값과의 차이에 비교해 작았다. 세 방법 간 차이의 절댓값의 평균 역시 성북구에서 1.52 $\mu\text{g}/\text{m}^3$ 로 가장 높았고, 양천구에서 0.31 $\mu\text{g}/\text{m}^3$ 로 가장 낮았다.

측정값(EM0)과 세 가지 추정방법(EM1~EM3) 간 구 대푯값의 차이는, 대푯값을 계산하기 위해 예측한 점 위치별 PM<sub>10</sub> 농도값들 간 비교에서도 나타난다(그림 4). 측정값은 대부분의 구에서 세 가지 추정방법에 의한 예측값들의 제3사분위수와 제1사분위수의 차이인 사분위범위(interquartile range) 밖에 있었다. 예측값들의 구별 평균인 대푯값을 비교할 때, 측정값은 17개 구에서 극단값(최솟값 또는 최댓값)이었다. 측정값이 최솟값인 구는 성동구, 성북구, 용산구, 노원구, 광진구, 송파구, 동작구, 중구, 강서구, 금천구 등 10개 구였으며, 최댓값인 구는 관악구, 서대문구, 종로구, 강남구, 동대문구, 강동구, 양천구 등 7개 구였다.

#### 4. 결과 해석

본 연구 결과는 측정치의 구 대표성에 대해서 의문을 제기한다. 서울특별시 각 25개 구에 하

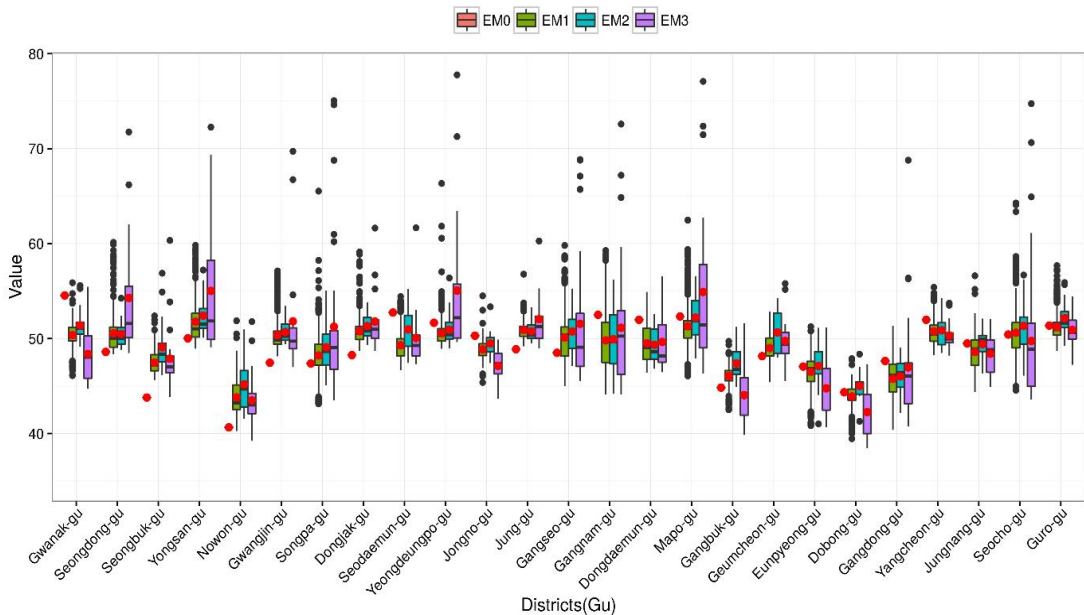


FIGURE 4. Distributions of PM<sub>10</sub> annual average concentrations measured at regulatory monitoring sites (EM0) and predicted at three types of locations (EM1 to EM3) across 25 districts in the order of the largest to smallest differences between prediction- and observation-based Gu-average estimates (red dots for Gu-average estimates)



나씩 설치된 상시측정망의 측정치는 세 가지 위치에 기반한 구 대푯값들과 대체로 관련성이 높았으나 구별로 편차가 있었다. 특히, 대부분의 구에서 측정치 대푯값은 추정치 대푯값들에 비해 최댓값이거나 최솟값이었다. 이는 서울특별시 내 대기오염 자동측정망의 측정소가 사람들의 거주지에 대한 대표성을 극대화할 수 있는 곳에 위치하지 않을 수 있는 가능성을 나타낸다. 대표성이 떨어지는 측정치를 지역의 PM<sub>10</sub> 노출 대푯값으로 이용할 경우 PM<sub>10</sub>의 건강영향 분석에서 부정확한 결과를 제공할 수 있을 것으로 예상된다.

인구 대표점을 생성하기 위해 본 연구에서 설정한 세 가지 위치 생성 방법들은 PM<sub>10</sub> 구별 대푯값 간 높은 관련성을 나타냈다. 집계구와 거주지역을 중첩한 EM1은 16,230개 지점을 포함해서 가장 정밀한 수준으로 거주 인구의 위치를 재현하였다. 이에 따라 개별 위치에 예측된 값들의 분포가 다른 방법들에 비해서 극단적인 이상치를 가지는 경우가 적었다(그림 4). 하지만, 동사무소 주소지를 이용한 EM2는 EM1에 비해서 지점 위치의 수가 적은 422개로 제한적인 공간 대표성을 가질 수 있음에도 EM1과 관련성이 높았다. 이 결과는 정밀한 수준으로 거주인구 분포를 고려하지 않더라도 행정 원리에 의한 주민센터 입지가 PM<sub>10</sub> 노출 산정을 위한 인구분포의 대표점으로서 의미가 있음을 보여 준다. 인구분포를 고려하지 않은 격자를 이용한 대푯값과의 낮은 상관성은, 서울특별시 내 인구가 적은 지역과 많은 지역 간 PM<sub>10</sub> 오염의 경향에 차이가 있음을 나타낸다. 따라서, 서울특별시와 같은 대규모 도시화 지역에서는 인구 거주 위치가 오염도의 소지역 분포와 밀접한 연관을 가짐을 알 수 있다.

한편, 공간적으로 볼 때 측정값과 예측 위치에 따른 대푯값 간의 차이는 동북-서남 방향(노원구, 성북구, 성동구, 동작구, 관악구 등)으로 크고, 이외의 방향에서는 외곽 지역(구로구, 중랑구, 양천구, 도봉구, 금천구 등)에서 차이가 작은 경향을 보였다. 또한 차이가 도심지보다는 비도심지역에서 더 크게 나타나, 이들 지역에서 대

기오염 측정소와 각 예측 위치들 사이의 토지이용 변수의 뚜렷한 차이가 대푯값의 차이에 영향을 미친 것으로 예상된다. 지역별로 측정소 및 예측 위치들의 공간적 특성이 EM0과 EM1-EM3 간 차이의 공간적 패턴에 영향을 미친 경로를 세밀하게 분석할 필요성이 제기된다.

## 결론

본 연구에서는 행정구역별로 집계되는 건강결과자료를 이용해서 PM<sub>10</sub>의 건강영향을 평가하는 연구에 필수적인 구별 PM<sub>10</sub> 대푯값 추정 방법을 제시하였다. 추정방법에서는 이전 연구에서 개발된 PM<sub>10</sub>에 대한 지점 예측모형을 이용하여 지점에 PM<sub>10</sub> 농도를 예측한 후, 구별로 평균값을 계산해서 구 대푯값으로 구했다. 지점은 연구자의 가정을 반영한 세 가지 위치로 집계구내 거주지역, 동 주민센터, 격자점을 고려하였다. 연구결과는 제한된 수의 대기오염 측정망 측정소에서 측정된 PM<sub>10</sub> 농도를 해당 구에 거주하는 인구의 PM<sub>10</sub> 노출 대푯값으로 사용하기에 무리가 있고, 인구 거주 위치를 대표할 수 있는 상대적으로 적은 수의 지점만으로도 구별 PM<sub>10</sub> 대표성을 가지는 농도값을 계산할 수 있음을 보였다.

본 연구는 서울특별시로 연구지역을 제한했으나, 연구에서 제시한 대기오염 대푯값 추정 방법은 전국규모로 확대 적용할 수 있다. 사망, 출산, 및 유병 등의 건강자료는 대부분 전국 시군구 단위로 제공된다. 그러나 약 40%에 해당하는 시군구에 대기오염 상시측정소가 없기 때문에, 상시측정망 자료를 이용해서 건강영향을 분석하는 경우 40%의 시군구는 분석에서 제외될 수밖에 없다. 그러므로 본 연구에서 개발한 대푯값 추정방법은 향후 전국규모의 건강영향 분석을 통해 한국에서 대기오염에 대한 건강취약지역을 식별하는 데 기여할 수 있다. 또한, 본 연구에서 제시한 대푯값 추정방법은 시군구 외 다른 지역 단위에도 적용될 수 있다. 현재 대부분의 건강자료가 시군구로 제공되고 있으나, 향후에는 소규모 지역 간 분포 파악이 가능

한 읍면동 등의 지역 단위 자료가 이용 가능할 수 있다. 본 연구에서 제시한 방법은 다양한 지역단위에 유용하게 적용될 수 있는 대푯값 추정방법을 제시한다.

본 연구는 몇 가지 제한점을 포함하고 있고, 이러한 제한점은 향후 연구 주제로 계속적인 탐색이 필요하다. 본 연구에서는 대푯값들 간의 차이가 공간적 경향이 나타나는 원인을 세밀하게 밝히지 못한 한계가 있다. 또한, 공간적인 대푯값 추정에 주목하기 위해 계절적인 변동의 영향을 받지 않는 연평균을 사용했다. 따라서 계절적 변동과 같은 PM<sub>10</sub> 농도의 시간적 변동의 측면을 고려하지 못했다. 일부 한계점에도 불구하고, 본 연구에서 제안하는 구 대푯값 추정방법은 향후 국가통계에서 제공하는 전국단위의 대규모 건강자료를 이용한 PM<sub>10</sub>의 건강영향분석 연구에서 유용하게 적용될 것으로 기대된다.

**KAGIS**

## 주

- 1) 통계 모델에서 변수를 선택하는 방법의 하나로, 상수항만을 포함한 모델을 기본 모델로 하여 가장 높은 부분상관계수를 가지는 순서대로 독립변수를 모델에 반영하는 방법이다(Montgomery *et al.*, 2001).
- 2) 이용 자료를 비복원추출 방식으로 5등분한 후, 네 집합의 자료를 추정자료(estimation data)로 나머지 한 집합의 자료를 예측자료(prediction data)로 총 5회에 걸쳐 모형을 적용하여 모형의 타당성을 검증하는 방법이다. 추정자료와 예측자료를 분리함으로써, 동일한 자료로 추정과 예측을 시행할 때 모형 예측력을 과도하게 추정하는 경향을 방지하고 적합한 모형 예측력을 추정할 수 있다(Montgomery *et al.*, 2001).
- 3) 2016년 현재 토지피복도는 세분류까지 제작(Ministry of Environment, 2016)되었으나, 2010년 기준으로는 중분류 토지피복도까지 완성되었다. 본 연구에서는 연구대상 시기에 가용한 중분류 토지피복도를 이

용하였다.

- 4) 공간 연산은 연산 방법 별로 크게 거리 계산과 버퍼(buffer) 연산으로 나누어 수행하였다. 거리 계산에 따라 도로, 항만, 공항 등 배출원과의 거리를 계산하였고, 버퍼 연산은 25, 50, 100, 300, 500, 1000, 5000m 등 총 7가지 반지름에 따라 위치 생성 방법 별로 얻은 점들을 기준으로 한 버퍼 면형 자료와 집계구 별 인구, 가구, 사업체자료, 토지피복도 등을 면형 자료를 교차(intersect) 하여 변수를 계산하는 연산과정이다. 이 때 벡터 자료의 경우 교차된 지역들의 변수 값들을 면적에 대하여 가중합하거나 버퍼 면적의 비율로 표준화하였고, 래스터 자료는 중첩된 셀들의 값을 총합산하는 방식으로 지리변수 값을 계산하였다.

## REFERENCES

- Atkinson, R. W., S. Kang, H. R. Anderson, I. C. Mills, and H. A. Walton. 2014. Epidemiological time series studies of PM<sub>2.5</sub> and daily mortality and hospital admissions: a systematic review and meta-analysis. *Thorax* 69(7):660-665.
- Brindley, P., S. M. Wise, R. Maheswaran, and R. P. Haining. 2005. The effect of alternative representations of population location on the areal interpolation of air pollution exposure. *Computers, Environment and Urban Systems* 29(4):455-469.
- Brook, R. D., S. Rajagopalan, C. A. Pope 3rd, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R.V. Luepker, M.A. Mittleman, A. Peters, D. Siscovick, S.C. Smith, L. Whitsel, and J.D. Kaufman. 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association.

- Circulation 121(21):2331-2378.
- Cressie, N. A. C. 1993. Statistics for spatial data(revised edition). John Wiley & Sons, New York, U.S. p.900.
- Eum, Y.S., I.S. Song, H.C. Kim, J.H. Leem, and S.Y. Kim. 2015. Computation of geographic variables for air pollution prediction models in South Korea. *Environmental Health Toxicology* 30 (e2015010).
- Hao, Y., H. Flowers, M. M Monti, and J. R Qualters. 2012. U.S. census unit population exposures to ambient air pollutants. *International Journal of Health Geographics* 11(3):1-9.
- Hengl, T. 2007. A practical guide to geostatistical mapping of environmental variables. JRC Scientific and Technical Reports. European Commission, Luxembourg. p.157.
- Jeong, J.C. 2014. A spatial distribution and time series change of PM<sub>10</sub> in Seoul City. *Journal of the Korean Association of Geographic Information Studies* 17(1):61-69 (정종철. 2014. 서울시 PM10 공간 분포 분석과 시계열 변화. *한국지리정보학회지* 17(1):61-69).
- Kang, Y.O., E.J. Yoon, and J.H. Jung. 2007. A study on the delineation of census output areas in Korea. *Journal of the Korean Urban Geographical Society* 10(1):15-36 (강영옥, 윤은주, 정재희. 2007. 소지역 통계구역 획정방안 연구. *한국도시지리학회지* 10(1):15-36).
- Keller, J.P, C. Olives, S.Y. Kim, L. Sheppard, P. D. Sampson, A. A. Szpiro, A.P. Oron, J. Lindström, S. Vedal, and J. D. Kaufman. 2015. A unified spatiotemporal modeling approach for prediction of multiple air pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution. *Environmental Health Perspectives* 123(4):301-309.
- Kim, S.Y. 2016. National exposure prediction approach for PM<sub>10</sub> and NO<sub>2</sub> in South Korea. *Proceedings of the Conference of International Society for Environmental Epidemiology and International Society of Exposure Science - Asia Chapter 2016*. Sapporo, Japan, June 26-29, 2016. p.93.
- Kim, S.Y., L. Sheppard, S. Bergen, A.A. Szpiro, P.D. Sampson, J.D. Kaufman, and S. Vedal. 2016. Prediction of fine particulate matter chemical components for the Multi-Ethnic Study of Atherosclerosis Cohort: a comparison of two modeling approaches. *Journal of Exposure Science and Environmental Epidemiology* 26(5): 520-528.
- Mercer, L.D., A.A. Szpiro, L. Sheppard, J. Lindström, S.D. Adar, R.W. Allen, E.L. Avol, A.P. Oron, T.Larson, L.-J.S. Liu, and J.D. Kaufman. 2011. Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen(NO<sub>x</sub>) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution(MESA Air). *Atmospheric Environment* 45(26):4412-4420.
- Min, K.D., Y.S. Eum, H.J. Kwon, and S.Y. Kim. 2015. Community air pollution monitoring design for epidemiological application in Seoul. *Proceedings of the International Conference of the Korean Society of Environmental Health and Toxicology*. Jeju, Korea, June 17, 2015. p.105.

- Ministry of Environment. 2011. Guidelines for installation and management of national air quality monitoring networks. p.409 (환경부. 2011. 대기오염측정망 설치 운영지침. 409쪽).
- Ministry of Environment. 2016. <http://egis.me.go.kr>. Accessed October 31, 2016. (환경부. 2016. 환경지리정보시스템 웹페이지).
- Montgomery, D.C., E.A. Peck, and G. G. Vining. 2001. Introduction to linear regression analysis (third edition). John Wiley & Sons, New York, U.S. p.310, 534, 535, 642.
- Parenteau, M.P. and M. C. Sawada. 2012. The role of spatial representation in the development of a LUR model for Ottawa, Canada. Air Quality, Atmosphere and Health 5(3):311-323.
- Park, H.J., H.S. Shin, Y.H. Roh, K.M. Kim, and K.H. Park. 2012. Estimating forest carbon stocks in Danyang using kriging methods for aboveground biomass. Journal of the Korean Association of Geographic Information Studies 15(1): 16-33 (박현주, 신휴석, 노영희, 김경민, 박기호. 2012. 크리깅 기법을 이용한 단양군의 산림 탄소저장량 추정-지상부 바이오매스를 중심으로-. 한국지리정보학회지 15(1):16-33).
- Pope III, C. A. and D.W. Dockery. 2006. Health effects of fine particulate air pollution : lines that connect. Journal of the Air & Waste Management 56(6):709-742.
- Sampson, P.D., M. Richards, A.A. Szpiro, S. Bergen, L. Sheppard, T.V. Larson, and J.D. Kaufman. 2013. A regionalized national universal kriging model using Partial Least Squares regression for estimating annual PM<sub>2.5</sub> concentrations in epidemiology. Atmospheric Environment 75:383-392.
- Stehman, S.V. and W. S. Overton. 1996. Spatial sampling. Practical Handbook of Spatial Statistics. CRC Press, Boca Raton, USA. pp.31-64.
- Wackernagel, H. 2003. Multivariate geostatistics : an introduction with applications. Springer-Verlag, Berlin, Germany. p.387.
- Yi, S.J., H. Kim, and S.Y. Kim. 2016. Exploration and application of regulatory PM<sub>10</sub> measurement data for developing long-term prediction models in South Korea. Journal of Korean Society for Atmospheric Environment 32(1):114-126 (이선주, 김호, 김선영. 2016. PM<sub>10</sub> 장기노출 예측모형 개발을 위한 국가 대기오염측정 자료의 탐색과 활용. 한국대기환경학회지 32(1): 114-126).
- Yoo, C.J. and K.W. Nam. 2014. Spatio-temporal distribution analysis of one-person household -the case of Busan City-. Journal of the Korean Association of Geographic Information Studies 17(2) :59-71 (유창주, 남광우. 2014. 1인가구의 시공간적 분포 분석 -부산시를 사례로-. 한국지리정보학회지 17(2):59-71) 