

# On the Hierarchical Modeling of Spatial Measurements from Different Station Networks

Jieun Choi<sup>a</sup> · Man Sik Park<sup>a,b,c,1</sup>

<sup>a</sup>Department of Statistics, Sungshin Women's University

<sup>b</sup>Institute of Statistics, Sungshin Women's University

<sup>c</sup>Basic Sciences Research Institute, Sungshin Women's University

(Received December 26, 2012; Revised December 31, 2012; Accepted January 3, 2013)

---

## Abstract

Geostatistical data or point-referenced data have the information on the monitoring stations of interest where the observations are measured. Practical geostatistical data are obtained from a wide variety of observational monitoring networks that are mainly operated by the Korean government. When we analyze geostatistical data and predict the expectations at unobservable locations, we can improve the reliability of the prediction by utilizing some relevant spatial data obtained from different observational monitoring networks and blend them with the measurements of our main interest. In this paper, we consider the hierarchical spatial linear model that enables us to link spatial variables from different resources but with similar patterns and guarantee the precision of the prediction. We compare the proposed model to a classical linear regression model and simple kriging in terms of some information criteria and one-leave-out cross-validation. Real application deals with Sulfur Dioxide(SO<sub>2</sub>) measurements from the urban air pollution monitoring network and wind speed data from the surface observation network.

Keywords: Sulfur Dioxide, wind speed, hierarchical model, kriging, spatial association, cross-validation.

---

## 1. 서론

18세기에 시작된 산업혁명 이후로 화석에너지의 사용이 증가하면서 대기오염이 심화되었다. 세계보건기구(WHO)에서는 대기오염을 “대기 중에 인위적으로 배출된 오염물질이 한 가지 또는 그 이상 존재하여 오염물질의 양, 농도 및 지속시간이 어떤 지역의 불특정 다수인에게 불쾌감을 일으키거나 해당지역에 공중보건상 위해를 끼치고, 인간이나 동물, 식물의 활동에 해를 주어 생활과 재산을 향유할 정당한 권리를 방해받는 상태”로 정의하고 있다. 우리나라도 1950년 이후 급속한 산업화와 도시화가 진행되면서 대기오염이 심각해지고 있다. 우리나라는 대기오염물질을 가스상 물질인 이산화황(Sulfur Dioxide; SO<sub>2</sub>), 일산화탄소(CO), 이산화질소(NO<sub>2</sub>)과 입자상물질을 포함한 총 61종의 물질로 정하며, 이 중 35종을 특정대기유해물질로 정하여 관리하고 있다(대기환경보전법).

대표적인 대기오염물질인 이산화황(SO<sub>2</sub>)은 무색의 자극성이 있으며 물에 매우 잘 녹는 타지 않는 가스

---

This work was supported by the Sungshin Women's University Research Grant of 2010.

<sup>1</sup>Corresponding author: Assistant Professor, Sungshin Women's University, Dongseon-dong 3-ga, Seongbuk-gu, Seoul 136-742, Korea. E-mail: mansikpark@sungshin.ac.kr

로, 황을 포함하는 연료(주로 석탄과 석유)가 연소되거나 금속 제련과정, 기타 산업과정 등에서 발생한다. 이는 노출되어 있는 인체의 점막을 자극하고 진한 이산화황 기체를 흡입하면 콧물, 담, 기침 등이 나오고 호흡곤란을 일으킨다. 일부의 경우에는 기관지염, 폐수종, 폐렴 등에 걸릴 수도 있는 것으로 파악되었다. 또한 이산화황은 질소산화물과 함께 산성비의 주요 원인물질로 토양, 호수, 하천의 산성화에 영향을 미치며, 식물의 잎맥 손상, 성장저해 및 빌딩이나 기념물 등 각종 구조물의 부식을 촉진시키는 것으로 알려져 있으며 시정장애를 일으키는 미세먼지의 주요 원인물질이기도 하다 (Ministry of Environment, 2007). 과거 이산화황으로 인한 큰 피해를 살펴보면, 1930년 12월, 벨기에 뮤즈 계곡에 위치한 대규모 공업단지에서 배출된 가스가 지면에 오래 머무르며 대기 중 이산화황 농도가 9.6ppm에서 38.4ppm까지 높아졌으며 이에 심장병과 급성폐렴으로 63명이 사망하고 수많은 급성 호흡기질환자가 발생하였다. 또한 동물, 식물에도 그 피해는 매우 컸다. 그 후, 1952년 12월 4일부터 12월 10일 사이에 발생한 런던스모그사건은 석탄이 연소되면서 발생하는 이산화황의 배출로 스모그가 발생하여 런던 시민의 약 12,000명이 만성 폐질환과 호흡장애로 사망했다. 이 이후로 전 세계에서는 대기오염이 중요한 문제로 언급되었다. 그러나 급격한 산업화와 도시화로 인해 대기오염의 범위는 더욱 넓어지고 심각해지고 있다. 이와 같이 대기와 인체에 심각한 영향을 끼치는 이산화황에 대한 관심과 연구가 필요하다.

최근의 공간통계를 이용한 연구를 보면, Kim과 Choi (2000)은 공간상관이 존재하는 환경자료를 이용하여 일반통계 분석법과 공간통계 분석법을 비교하였고, Cho 등 (2001)은 공간통계학에서 다루는 자료에 대하여 일반통계와 공간통계 방법에 대해 예측오차제곱합(predicted residual sums of squares; PRESS)을 사용하여 예측성능을 비교하였다. Huh 등 (2004)은 남한의 강수량 자료의 공간분석을 하는데 가중최소방법, 최대우도방법, 잔차최대우도방법을 사용하여 각 방법을 비교하였고, Cho와 Jeong (2006)은 강우관측소의 강우량에 대하여 공간보간을 실시하여 일반 크리깅을 실시하여 정확성을 검증하였다. Shin 등 (2007)은 계층적 베이지안 추정법과 공간통계량을 활용해 소지역 추정법을 평가하였고, Heo 등 (2007)은 연평균 일교통량을 최단경로를 기반으로 예측하기 위해 공간적회귀모형을 제시하였다. Jung 등 (2008)은 강수량 자료를 이용해 최소제곱법과 최대우도추정방법으로 크리깅을 하였으며 신뢰성 검증을 위해 교차검증방법을 사용하였고, Park과 Jang (2008)은 공간적 분포도를 작성하는데 기온과 강수량의 관측치를 이용하여 다변량 크리깅의 적용 가능성을 검토하였다. Choi와 Park (2009)은 공간분석을 하는데 자료의 정규성을 만족시키기 위하여 Box-Cox 변환을 이용하였고, 공간적 패턴을 살펴보기 위하여 베이지안 계층 모형을 이용했고, Heo와 Park (2009)은 공간분석을 하는데 베이지안 계층 모형을 이용하여 모수를 추정하고 강우량을 예측하였다. Jeong 등 (2010)은 보정된 80m에서의 풍속자료를 가지고 크리깅 기법을 이용해 남한의 풍속 예측지도를 구현하였다. Kim 등 (2010)은 인구 분포를 추정하기 위한 기존의 방법인 위성영상과 회귀분석 모형을 사용하는 것과 공간통계 기법을 적용하여 크리깅을 이용하는 것과 비교하였고, 크리깅을 사용하는 것이 인구분포를 추정하는데 정확도가 더 높다고 하였다. Kim (2010)은 일사량 및 일조시간의 추정 정확성을 향상시키기 위하여 공간적 상호관계를 파악할 수 있는 베리오그램을 이용하여 자료 사이의 관계를 파악해 공간통계분석방법인 정규 크리깅을 적용해 일사량과 일조시간을 추정하였다.

Han 등 (2004)에 따르면 이산화황이 기류, 풍속 등에 의해 보통 수백에서 수천 km까지 이동이 가능하다. 즉 이산화황이 거의 발생하지 않는 지역에서도 기류, 풍속 등에 의해 영향을 받을 수 있게 된다. 결국 이산화황을 분석함에 있어서 이 변수만을 연구하는 것보다 기류, 풍속 등의 외부요인들을 이용하여 이산화황을 분석하는 것이 바람직하다. 또한 이러한 외부 요인들이 하나의 공간적 상관성을 내포하는 자료가 될 수 있음을 고려한다.

본 논문에서는 서로 다른 관측망에서 얻어진 두 자료를 이용하여 공간 분석하고자 한다. 하나의 자료를 예측하는 데 있어서 그 자료와 관련성이 높으며 다른 관측망에서 얻어진 자료를 이용하여 공간 분석을

한다면 예측력을 높일 수 있다. 그러나 서로 다른 관측망의 자료를 이용할 때, 두 자료 사이의 관련성의 강도가 낮으면 모형의 효용성이 떨어지게 된다. 따라서 서로 다른 관측망의 자료를 이용하여 분석하는 경우 두 자료의 관련성에 대한 평가가 모형적합, 공간예측과 함께 이루어져야 한다. 이 경우 두 자료의 관측망이 서로 다르므로 하나의 자료를 예측하는데 다른 자료를 바로 적용할 수 없다. 이를 해결하기 위해 계층모형을 고려하고자 한다. 이를 적용하기 위하여 이산화황자료와 풍속자료를 이용한다. 이산화황이 기류, 풍속 등 여러 가지 요인들에 의해 이동하는데, 여러 요인 중 풍속자료를 얻을 수 있었으며, 가장 간단한 계층모형을 구축하기 위하여 풍속자료만을 사용한다. 주로 도시지역에서만 관측되는 이산화황을 예측하는 데 예측력을 높이기 위해 우리나라 전역에 걸쳐 측정되는 풍속자료를 이용하여 모형을 구축하였다. 이산화황과 풍속의 관측망이 같지 않아 이산화황을 예측하는 데 풍속자료를 바로 이용할 수 없으므로 계층모형을 고려한다. 모형을 추정하고 적합하는데 잔차최대우도추정법(restricted maximum likelihood estimation; REML), 최대우도추정법(maximum likelihood estimation; ML)을 사용하고 오차항의 공분산 함수에는 지수모형(exponential model), 가우시안모형(Gaussian model), 구형모형(spherical model) 등을 가정하였다.

본 논문의 순서는 다음과 같다. 2장에서는 사례연구를 위한 두 가지의 자료에 대한 관측망에 대한 설명을 하였다. 3장에서는 공간분석을 위한 기본적인 공간통계학의 개념을 설명하고 다양한 관측망에서 얻은 자료를 이용하여 계층모형을 구축하는 방법에 대하여 기술하였다. 4장에서는 풍속자료와 이산화황자료를 이용한 분석결과에 대해 기술하였으며, 이산화황만을 이용한 결과와 풍속에 대한 정보를 추가하여 얻은 결과를 비교하였다. 5장에서는 결론 및 향후 연구 과제에 대하여 논하였다.

## 2. 관측망

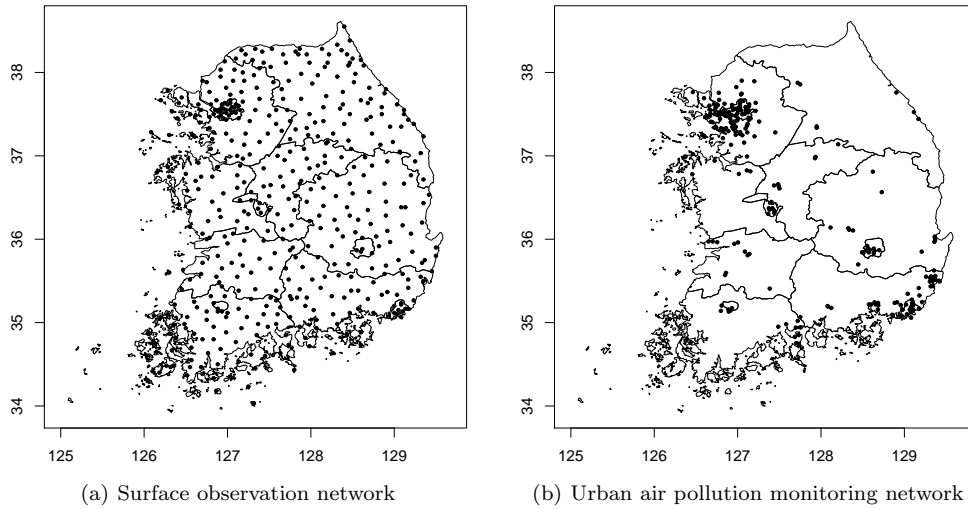
제 2장에서는 계층모형의 구축을 위해 사용한 두 가지 관측망(monitored network)들에 대하여 소개하고자 한다. 본 논문에서는 다음과 같은 두 개의 관측망을 고려하였는데, 첫 번째는 지상기상관측망에서 측정된 풍속 및 풍향자료이고, 두 번째는 도시대기측정망에서 측정된 이산화황이다.

### 2.1. 지상기상관측망(Surface observation network)

본 연구에서 사용한 자료 중 첫 번째는 2003년부터 2007년까지 5년 평균 풍속자료이다. 이는 주요 도시지역을 제외한 남한지역 460개 기상관측지점(Figure 2.1(a))에서 측정되었으며 기상청(<http://www.kma.go.kr>)에서 얻은 자료이다. 기상청은 지상기상관측망에서 종관기상관측장비와 무인으로 운영되는 자동기상관측장비를 이용하여 지상관측업무를 수행하고 있다. 종관기상관측장비는 지방청, 기상대, 관측소에 설치되어 기상상태를 관측하고 국제전문 작성 및 통계표 작성과 같은 관측업무를 자동으로 처리한다. 기상청에서 기상예보를 위해 관측하는 바람고도는 10m이다. 풍속을 풍력발전 등에 활용하려면 50m-100m의 고도에서의 풍속이 사용되는데, 본 연구에서는 변환된 풍속이 아닌 관측된 위치인 10m의 바람고도의 풍속자료를 사용하고자 한다. 그리고 지상기상관측망에서는 풍속만을 관측한 것이 아니라 주풍향(main wind direction)도 관측하였다. 또한, 모형에 각 지점의 고도를 포함시키기 위해 Google Earth 를 이용하여 킬로미터(km) 단위의 고도를 사용한다. 위도와 경도는 지구의 곡면의 정보가 포함되었기 때문에 위치정보를 보정하여 고도와 마찬가지로 킬로미터(km) 단위를 사용한다.

### 2.2. 도시대기측정망(Urban air pollution monitoring network)

본 연구에서 사용한 자료 중 두 번째는 2005년부터 2007년까지 3년 평균 이산화황(SO<sub>2</sub>)의 자료이다. 이는 주요 도시지역을 제외한 남한지역 224개 대기오염관측지점(Figure 2.1(b))에서 측정되었으며 국



**Figure 2.1.** Two different monitoring networks (460 monitoring stations(a) and 224 monitoring stations(b)).

립환경과학원(<http://www.nier.go.kr>)에서 얻은 자료이다. 도시대기측정망은 도시지역의 평균 대기 질 농도를 파악하여 환경 기준 달성 여부를 판정한다. 그러나 측정지점이 대부분 도시지역으로 한정되어 있어서 농촌지역에서는 측정되지 않는 한계가 있다. 이산화황은 자외선형광법(Pulse U.V Fluorescence Method)을 이용해 측정된다. 자외선형광법은 비교적 짧은 파장 영역의 자외선에 의해 나타나는 이산화황( $\text{SO}_2$ ) 분자로부터 발생하는 형광강도를 측정하여 이산화황의 농도를 연속적으로 측정하는 방법이다(대기환경연보). 풍속자료와 마찬가지로 대기오염자료에서도 모형에 각 지점의 고도를 포함하기 위해 Google Earth를 이용하여 킬로미터(km) 단위의 고도를 구하였다. 대기오염자료의 위도와 경도 또한 지구의 곡면의 정보가 포함되었기 때문에 위치정보를 보정하여 고도와 마찬가지로 킬로미터(km) 단위를 사용한다.

두 관측망의 지점들은 모두 상이하며, 두 관측망에서 얻은 관측값의 측정기간 역시 상이하다. 풍속자료는 2003년부터 2007년까지의 평균자료이고 이산화황자료는 2005년부터 2010년까지의 일별자료를 얻었다. 이에 두 자료의 기간을 최대한 맞추기 위하여 이산화황자료를 2005년부터 2007년까지의 3년 자료의 평균을 구하여 분석에 이용했다.

### 3. 공간자료들을 활용한 계층모형

공간통계학에서 사용되는 데이터는 얻어진 형태에 따라 지리통계자료(geostatistical data), 격자자료(lattice data), 공간점 패턴자료(spatial point-pattern data)의 세 가지로 분류된다. 지리통계자료는 고정된 위치에서 얻어진 측정값으로 관측지점이 지도상에 점으로 표현되고 그 지점에서만 자료를 관측하게 된다. 대표적인 지리통계자료는 기상관측소에서 측정한 기온 및 강수량, 특정 지점에서 관측된 석탄의 매장량 등이 있다. 이 자료의 분석목적은 미관측지점에서의 예측에 있다.

#### 3.1. 크리깅(Kriging)

지리통계자료는 다음과 같이 표현할 수 있다.

$$\left\{ Z(\mathbf{s}) : \mathbf{s} \equiv (s_1, \dots, s_d)^T \in \mathbb{D} \subset \mathbb{R}^d \right\}.$$

위도, 경도, 고도를 고려하는 지리통계자료는  $\{Z(\mathbf{s}_i); \mathbf{s}_i = (s_{i1}, s_{i2}, s_{i3})^T, i = 1, 2, \dots, n\}$ 으로 표현되어진다. 이러한  $\mathbf{z} \equiv (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))^T$ 의 확률과정(stochastic process)은 다음과 같은 선형회귀모형으로 표현할 수 있다.

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (3.1)$$

여기서,  $\mathbf{X}\boldsymbol{\beta}$ 는 평균함수(mean function)로서, 전반적 추세 또는 경향을 나타내는 위치간의 특성을 의미한다. 일반적으로  $\mathbf{X}$ 는 그 지점에서 관측되는 외부요인들 뿐만 아니라 그 지점의 위치정보를 포함한다. 또한 식 (3.1)의  $\boldsymbol{\epsilon}$ 는 위치에 내재된 고유한 변동을 의미하며 2차 정상성(second-order stationarity)을 만족하는 확률변수로 가정한다. 즉, 오차항들의 평균은 0이고, 오차항들은 유한한 분산을 가지며 오차항 사이의 공분산은 두 지점사이의 거리에 의존함을 가정한다. 공간적 연관구조를 의미하는 오차항의 분산-공분산 행렬에 공간적인 연관성을 포함시켜 모형화하기 위해서는 세미베리오그램(semivariogram)을 추정해야한다. 오차항에 대한 세미베리오그램은 다음과 같이 정의된다.

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2}E[\epsilon(\mathbf{s}_i) - \epsilon(\mathbf{s}_j)]^2 \quad (3.2)$$

세미베리오그램에는 공간적 연관구조를 나타내는 세 가지 모수( $\tau^2, \lambda^2, \phi$ )를 포함한다. 문치(nugget)이라 불리는  $\tau^2$ 는 측정오차로 인해 발생하는 분산의 일부분이고, 부분 문턱(partial sill)이라 불리는  $\lambda^2$ 는 임의의 두 관측지점의 거리가 멀어질수록 자료들 사이에 상관성이 작아져 상관성이 거의 0이 되는 순수한 분산의 일부분이다. 분산은  $\sigma^2 = \tau^2 + \lambda^2$ 으로 표현할 수 있고 문턱(sill)이라고도 불리운다. 그리고  $\phi$ 는 범위(range)로서, 문턱까지의 거리를 의미한다. 식 (3.2)를 근간으로 자료로부터 계산된 세미베리오그램을 이론적 세미베리오그램에 적합하여 위에서 언급한 세 가지 모수들을 추정하게 된다. 대표적인 이론적 세미베리오그램은 다음과 같다.

#### A. 지수모형(Exponential semi-variogram model)

$$\gamma(\mathbf{h}) = \tau^2 + \lambda^2 \left[ 1 - \exp\left(-\frac{\|\mathbf{h}\|}{\phi}\right) \right], \quad \|\mathbf{h}\| > 0.$$

#### B. 가우시안모형(Gaussian semi-variogram model)

$$\gamma(\mathbf{h}) = \tau^2 + \lambda^2 \left[ 1 - \exp\left(-\frac{\|\mathbf{h}\|^2}{\phi^2}\right) \right], \quad \|\mathbf{h}\| > 0.$$

#### C. 구형모형(Spherical semi-variogram model)

$$\gamma(\mathbf{h}) = \begin{cases} \tau^2 + \lambda^2 \left( \frac{3\|\mathbf{h}\|}{2\phi} - \frac{\|\mathbf{h}\|^3}{2\phi^3} \right), & \|\mathbf{h}\| \leq \phi, \\ \tau^2 + \lambda^2, & \|\mathbf{h}\| > \phi. \end{cases}$$

임의의 두 지점에 대하여  $\mathbf{h} = |\mathbf{s}_i - \mathbf{s}_j|$ 이고 등방성(isotropy)를 만족한다고 하자. 즉, 임의의 두 지점에서 관측한 값들 간의 상관성은 두 지점이 놓인 방향과는 무관하게 유클리드거리(Euclidean distance),  $\|\mathbf{h}\| = \sqrt{h_1^2 + h_2^2 + \dots + h_d^2}$ 에만 의존하게 된다. 이 모형들은 두 관측지점의 거리가 가까우면 강한 연관성을, 거리가 멀어지면 약한 연관성 혹은 무상관성을 가정한다. 그러나 각 모형은 구조적인 차이로 인해 거리에 따라 줄어드는 상관계수의 경향은 다르다. 본 논문에서는 선형회귀모형으로 표현되는 식 (3.1)과 공간적 연관구조를 의미하는 분산-공분산 행렬을 토대로 우도함수(likelihood function)를 최대화하는 추정방법(잔차최대우도추정법(REML), 일반최대우도추정법(ML))들을 고려하고자 한다. 추정

된 모형 하에서 크리깅을 통해 미관측지점에 대한 예측을 시행한다. 이는 이미 관측된 지점들에서 얻어진  $\{Z(\mathbf{s}_i)\}$ 의 선형결합으로 미관측지점의 예측이 표현되어짐을 의미한다. 공간자료의 모형 추정과 크리깅에 대한 보다 자세한 이론은 Goovaerts (1997), Banerjee 등 (2004), Schabenberger와 Gotway (2005), Cressie (1993)을 참조하기 바란다.

### 3.2. 계층모형(Hierarchical model)

이제, 두 관측망들로부터 얻은 자료들을 이용하여 계층모형을 구축하고자 한다. 먼저  $\mathbf{A}$ 라는 관측망의 관측지점,  $\{\mathbf{s}_k^a, k = 1, \dots, n_a\}$ 에서 관측된 자료  $\mathbf{z}_a \equiv (Z(\mathbf{s}_1^a), Z(\mathbf{s}_2^a), \dots, Z(\mathbf{s}_{n_a}^a))^T$ 에 대한 확률과정은 다음의 선형회귀모형으로 표현할 수 있다.

$$\mathbf{z}_a = \mathbf{X}_a \boldsymbol{\beta}_a + \boldsymbol{\epsilon}_a. \quad (3.3)$$

여기서,  $\mathbf{X}_a$ 는 독립변수들로 이루어진  $(n_a \times p)$  크기의 행렬이고  $\boldsymbol{\beta}_a = (\beta_1^a, \beta_2^a, \dots, \beta_p^a)^T$ 이다. 그리고  $\mathbf{B}$ 라는 관측망의 관측지점,  $\{\mathbf{s}_k^b, k = 1, \dots, n_b\}$ 에서 관측된 자료  $\mathbf{z}_b$ 에 대한 확률과정은 다음의 선형회귀모형으로 표현할 수 있다.

$$\mathbf{z}_b = \mathbf{X}_b \boldsymbol{\beta}_b + \boldsymbol{\epsilon}_b. \quad (3.4)$$

여기서,  $\mathbf{X}_b$ 는 독립변수들로 이루어진  $(n_b \times q)$  크기의 행렬이고  $\boldsymbol{\beta}_b = (\beta_1^b, \beta_2^b, \dots, \beta_q^b)^T$ 이다. 식 (3.3)과 식 (3.4)의 구성성분에 대한 가정은 식 (3.1)과 동일하다. 즉, 각 모형의 오차항에 대한 분산-공분산행렬은 공간 상관구조를 내포하는  $(\tau^2, \lambda^2, \phi)$ 에 의해 결정된다.

자료  $\mathbf{z}_b$ 를 분석함에 있어서 주어진 평균함수와 더불어 자료  $\mathbf{z}_a$ 로부터 크리깅을 통해 얻게 되는  $\hat{\mathbf{z}}_{ba} \equiv (\hat{Z}(\mathbf{s}_1^a), \hat{Z}(\mathbf{s}_2^a), \dots, \hat{Z}(\mathbf{s}_{n_b}^a))^T$ 에 대한 정보까지 포함한 모형을 고려할 수 있다. 따라서  $\mathbf{z}_b$ 의 확률과정은 다음의 선형회귀모형으로도 표현할 수 있다.

$$\mathbf{z}_b = \mathbf{X}_b \boldsymbol{\beta}_c + \delta \hat{\mathbf{z}}_{ba} + \boldsymbol{\epsilon}_c. \quad (3.5)$$

여기서,  $\mathbf{X}_a$ 와  $\mathbf{X}_b$ 의 변수는 동일하다고 가정한다. 이와 같이 서로 다른 관측망들로부터 얻었으나 공간적인 패턴이 유사하여 이를 계층적인 관계 하에서 구성할 수 있는 모형을 계층모형(hierarchical model)이라 한다. 이러한 계층모형은 다음과 같이 표현할 수도 있다.

$$\begin{aligned} \mathbf{z}_a &\sim N_{n_a}(\mathbf{X}_a \boldsymbol{\beta}_a, \boldsymbol{\Sigma}_a), \\ \hat{\mathbf{z}}_{ba} &= \mathbf{X}_b \hat{\boldsymbol{\beta}}_a + \hat{\boldsymbol{\epsilon}}_{ba}, \\ \mathbf{z}_b &\sim N_{n_b}(\mathbf{X}_b \boldsymbol{\beta}_c + \delta \hat{\mathbf{z}}_{ba}, \boldsymbol{\Sigma}_c). \end{aligned} \quad (3.6)$$

여기서,  $\boldsymbol{\Sigma}_a$ 와  $\boldsymbol{\Sigma}_c$ 는 각각 오차항의 분산-공분산행렬이고  $\hat{\boldsymbol{\epsilon}}_{ba} \equiv (\hat{\epsilon}(\mathbf{s}_1^a), \hat{\epsilon}(\mathbf{s}_2^a), \dots, \hat{\epsilon}(\mathbf{s}_{n_b}^a))^T$ 이다.

본 연구에서는 다음의 네 가지 모형들에 대해 비교 및 평가하고자 한다.

1. 공간상관성을 고려하지 않는 전통적 선형회귀모형(식 (3.4)에서  $\boldsymbol{\Sigma}_b = \sigma_b^2 \mathbf{I}_{n_b}$ ).
2. 공간상관성을 고려하지 않는 계층모형(식 (3.6)에서  $\boldsymbol{\Sigma}_c = \sigma_c^2 \mathbf{I}_{n_b}$ ).
3. 공간상관성을 고려하는 선형회귀모형(식 (3.4)).
4. 공간상관성을 고려하는 계층모형(식 (3.6)).

## 4. 사례연구

본 연구에서는 3장에서 제시한 모형구축방법을 실제 자료에 적용한 결과를 소개하고자 한다. 사용한 자료는 2장에서 제시한 풍속자료와 대기오염자료 중 하나인 이산화황(SO<sub>2</sub>)이다. 분석의 최종결과인 풍속자료와 이산화황자료의 예측지도(prediction map)를 구현하기 위해 935개의 일정한 거리를 가지는 격자지점들을 고려하였고, 이들 지점에 대한 위치정보 중 고도는 Google Earth 프로그램을 통해 구하였다. 사례연구를 위해 R(R Development Core Team, 2012) 프로그램을 사용하였다.

### 4.1. 풍속자료의 모형 적합

10m 높이의 평균풍속 자료를 분석함에 있어 Jeong 등 (2010)을 이용하였다. 로그변환한 풍속과 경도( $s_{i1}^a$ ), 로그변환한 풍속과 위도( $s_{i2}^a$ ) 사이의 관계를 살펴본 결과, 2차형의 모형식으로 표현할 수 있다. 그리고 로그변환한 풍속과 고도( $s_{i3}^a$ )와의 관련성에 대해서는 뚜렷한 관계가 있지는 않았지만 어느 정도 선형적 관계가 있다고 판단했다. 또한, 460개의 풍속 관측지점에서 주풍향은 7개 방위(북, 북동, 동, 남동, 남서, 서, 북서) 중 하나로 나타났으며, 남(s)풍이 주풍향인 관측지점은 없었다. 주풍향에 따라 평균 풍속의 차이를 살펴본 결과 북동(NE)풍과 북서(NW)풍이 주풍향인 관측지점에서의 평균 풍속이 다른 주풍향을 가진 지점들에 비해서 크다는 것을 알 수 있었으며, 주풍향에 의한 평균 풍속의 차이를 평균함수 모형 포함시키기 위해서 가변수 처리하였다. Figure 2.1(a)에서 나타난 460개의 풍속 관측지점들로부터 얻은 평균풍속에 대한 공간적 선형회귀모형은 식 (3.3)으로 표현되며 평균함수는 다음의 식과 같다.

$$\begin{aligned} \mathbf{X}_a \beta_a = & \beta_1^a + s_{i1}^a \beta_2^a + s_{i2}^a \beta_3^a + (s_{i1}^a)^2 \beta_4^a + (s_{i2}^a)^2 \beta_5^a + s_{i1}^a s_{i2}^a \beta_6^a + s_{i3}^a \beta_7^a \\ & + I_{NE} \beta_8^a + I_E \beta_9^a + \cdots + I_{NW} \beta_{13}^a. \end{aligned} \quad (4.1)$$

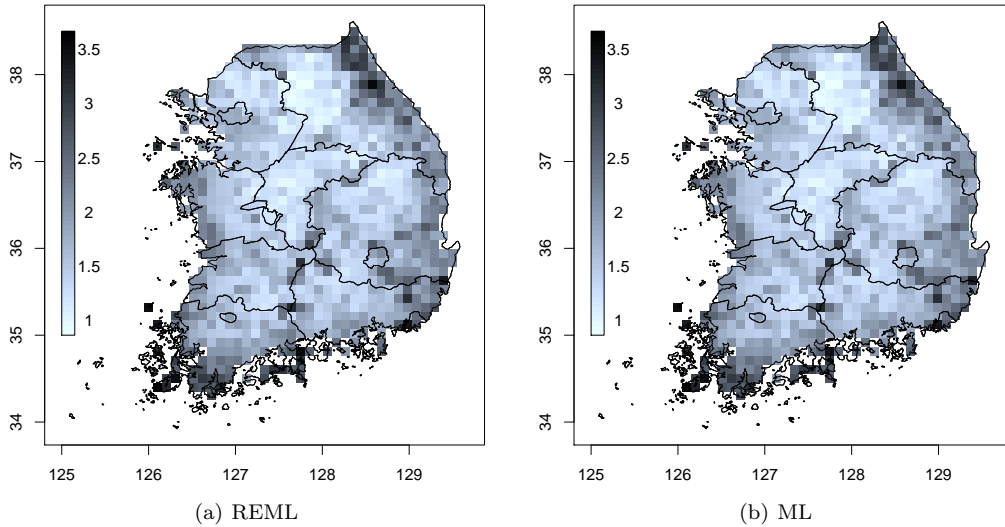
여기서,  $I$ 는 지시함수로서, 예를 들면 임의의 지점,  $s^a$ 의 주풍향이 북동(NE)풍이면  $I_{NE} = 1$ , 그렇지 않으면  $I_{NE} = 0$ 이다. 평균함수에 의해 설명되지 않는 오차항의 공간상관구조를 추정함에 있어서 2차 정상성을 가정한다.

식 (3.1)로 표현되는 선형회귀모형에 대해 여러 추정방법을 적용한 결과는 Table 4.1에 나타나 있다. 잔차최대우도추정(REML)방법 하에서는 지수모형, 가우시안모형, 구형모형을 공분산함수로 고려한 공간선형모형의 로그우도값이 각각  $-79.48, -80.45, -78.43$ 으로 구형모형의 로그우도값이 가장 컸으며, 일반최대우도추정(ML)방법 하에서도 세 가지 공분산모형들의 로그우도값이 각각  $-89.06, -83.02, -81.30$ 으로 구형모형을 고려한 경우가 가장 큰 로그우도값을 보였다. 또한, 공간선형모형과 전통적 선형회귀모형의 로그우도값( $-105.25$ )을 비교해보면, 공간선형모형의 설명력이 우수함을 알 수 있다. 공간선형회귀모형의 모수 추정에 사용한 방법들 간의 추정값에 대한 큰 차이는 보이지 않고 있다. 특히, 공간 상관구조를 설명하는  $(\tau^2, \lambda^2, \phi)$ 에 대한 추정값은 잔차최대우도추정방법과 일반최대우도추정방법이 거의 유사한 값으로 나타났다. 로그우도값으로 두 가지 공간 모형들을 비교해보면 잔차최대우도추정방법의 구형모형( $-78.43$ )이 일반최대우도추정방법의 구형모형( $-81.30$ )보다 크므로 계층모형을 적합하기 위한 풍속자료의 최적 모형은 잔차최대우도추정방법의 구형모형으로 한다. 임의의 격자지점의 추가적인 정보를 이용하여 잔차최대우도추정방법 하에서의 적합된 모형으로 예측한 결과는 Figure 4.1과 같다. Figure 4.1(a)와 (b)에서 알 수 있듯이, 영동지방, 호남지방, 영남지방의 해안가와 고도가 높은 덕유산, 지리산 일대의 풍속이 크게 예측되었고, 경기도 동부지방과 영서지방, 충청북도와 경상북도 내륙지방의 풍속은 비교적 작게 예측되었다.

**Table 4.1.** Parameter estimation based on linear regression models of wind speed data ( $a^b \equiv a \times 10^{-b}$ ).

	Classical linear model			Spatial linear model					
	Least Squares			REML(Spherical)			ML(Spherical)		
	Est. <sup>§</sup>	95% C.I		Est. <sup>§</sup>	95% C.I		Est. <sup>§</sup>	95% C.I	
$\beta_1^a$	-2.72 <sup>2</sup>	-9.25 <sup>2</sup>	3.80 <sup>2</sup>	-8.82 <sup>2</sup>	-2.95 <sup>1</sup>	1.19 <sup>1</sup>	-8.13 <sup>2</sup>	-2.62 <sup>1</sup>	9.99 <sup>2</sup>
$\beta_2^a$	-6.83 <sup>4</sup>	-1.08 <sup>3</sup>	-2.81 <sup>4</sup>	-9.55 <sup>5</sup>	-1.30 <sup>3</sup>	1.11 <sup>3</sup>	-1.90 <sup>4</sup>	-1.25 <sup>3</sup>	8.77 <sup>4</sup>
$\beta_3^a$	-9.24 <sup>4</sup>	-1.18 <sup>3</sup>	-6.65 <sup>4</sup>	-7.36 <sup>4</sup>	-1.57 <sup>3</sup>	1.04 <sup>4</sup>	-7.86 <sup>4</sup>	-1.52 <sup>3</sup>	-4.49 <sup>5</sup>
$\beta_4^a$	2.79 <sup>5</sup>	2.29 <sup>5</sup>	3.30 <sup>5</sup>	3.24 <sup>5</sup>	1.96 <sup>5</sup>	4.52 <sup>5</sup>	3.18 <sup>5</sup>	2.03 <sup>5</sup>	4.33 <sup>5</sup>
$\beta_5^a$	7.82 <sup>6</sup>	5.57 <sup>6</sup>	1.00 <sup>5</sup>	1.23 <sup>5</sup>	6.03 <sup>6</sup>	1.86 <sup>5</sup>	1.19 <sup>5</sup>	6.32 <sup>6</sup>	1.76 <sup>5</sup>
$\beta_6^a$	-3.39 <sup>6</sup>	-6.84 <sup>6</sup>	4.35 <sup>8</sup>	-2.32 <sup>6</sup>	-1.21 <sup>5</sup>	7.48 <sup>6</sup>	-2.50 <sup>6</sup>	-1.12 <sup>5</sup>	6.24 <sup>6</sup>
$\beta_7^a$	5.72 <sup>1</sup>	4.19 <sup>1</sup>	7.25 <sup>1</sup>	7.47 <sup>1</sup>	5.97 <sup>1</sup>	8.97 <sup>1</sup>	7.33 <sup>1</sup>	5.85 <sup>1</sup>	8.82 <sup>1</sup>
$\beta_8^a$	2.87 <sup>1</sup>	1.70 <sup>1</sup>	4.03 <sup>1</sup>	2.32 <sup>1</sup>	1.25 <sup>1</sup>	3.39 <sup>1</sup>	2.36 <sup>1</sup>	1.30 <sup>1</sup>	3.42 <sup>1</sup>
$\beta_9^a$	1.05 <sup>1</sup>	-6.37 <sup>2</sup>	2.73 <sup>1</sup>	1.10 <sup>1</sup>	-4.04 <sup>2</sup>	2.61 <sup>1</sup>	1.10 <sup>1</sup>	-3.91 <sup>2</sup>	2.60 <sup>1</sup>
$\beta_{10}^a$	1.86 <sup>1</sup>	5.02 <sup>2</sup>	3.23 <sup>1</sup>	1.63 <sup>1</sup>	3.96 <sup>2</sup>	2.87 <sup>1</sup>	1.62 <sup>1</sup>	3.96 <sup>2</sup>	2.85 <sup>1</sup>
$\beta_{11}^a$	2.04 <sup>1</sup>	9.38 <sup>2</sup>	3.15 <sup>1</sup>	1.20 <sup>1</sup>	1.92 <sup>2</sup>	2.22 <sup>1</sup>	1.23 <sup>1</sup>	2.33 <sup>2</sup>	2.24 <sup>1</sup>
$\beta_{12}^a$	3.02 <sup>1</sup>	2.15 <sup>1</sup>	3.90 <sup>1</sup>	2.22 <sup>1</sup>	1.41 <sup>1</sup>	3.03 <sup>1</sup>	2.26 <sup>1</sup>	1.46 <sup>1</sup>	3.06 <sup>1</sup>
$\beta_{13}^a$	2.68 <sup>1</sup>	1.80 <sup>1</sup>	3.56 <sup>1</sup>	2.06 <sup>1</sup>	1.24 <sup>1</sup>	2.87 <sup>1</sup>	2.09 <sup>1</sup>	1.28 <sup>1</sup>	2.89 <sup>1</sup>
$\tau^2$				0.06			0.06		
$\lambda^2$				0.06			0.05		
$\phi$				118.50			113.86		
log.lik.	-105.25			-78.43			-81.30		

Est., Estimate; log.lik., Log-likelihood.

**Figure 4.1.** Prediction maps of wind speed based on spherical model.

#### 4.2. 이산화황 자료의 모형 적합

3년간 이산화황 자료를 분석하기 위해 로그변환한 이산화황의 평균자료를 사용하였다. 이산화황에 관한 선형회귀모형을 적합하고 오차항의 공분산모형을 추정하기 위해 풍속자료에서 사용한 평균함수를 적용하고자 한다. Figure 2.1(a)에 나타난 460개의 풍속 관측지점에서는 평균풍속과 더불어 주풍향(main



**Table 4.2.** Parameter estimation based on linear regression models of Sulfur Dioxide(SO<sub>2</sub>) ( $a^b \equiv a \times 10^{-b}$ ).

	Classical linear model			Spatial linear model					
	Least Squares			REML(Spherical)			ML(Spherical)		
	Est. <sup>§</sup>	95% C.I		Est. <sup>§</sup>	95% C.I		Est. <sup>§</sup>	95% C.I	
$\beta_1^b$	-5.27 <sup>0</sup>	-5.42 <sup>0</sup>	-5.12 <sup>0</sup>	-5.23 <sup>0</sup>	-5.45 <sup>0</sup>	-5.01 <sup>0</sup>	-5.24 <sup>0</sup>	-5.45 <sup>0</sup>	-5.03 <sup>0</sup>
$\beta_2^b$	1.18 <sup>4</sup>	-6.69 <sup>4</sup>	9.04 <sup>4</sup>	5.57 <sup>4</sup>	-6.62 <sup>4</sup>	1.78 <sup>3</sup>	5.34 <sup>4</sup>	-5.88 <sup>4</sup>	1.66 <sup>3</sup>
$\beta_3^b$	3.40 <sup>4</sup>	-2.02 <sup>4</sup>	8.83 <sup>4</sup>	3.24 <sup>4</sup>	-5.77 <sup>4</sup>	1.23 <sup>3</sup>	3.09 <sup>4</sup>	-5.20 <sup>4</sup>	1.14 <sup>3</sup>
$\beta_4^b$	3.45 <sup>6</sup>	-6.17 <sup>6</sup>	1.31 <sup>5</sup>	-9.07 <sup>7</sup>	-1.74 <sup>5</sup>	1.56 <sup>5</sup>	-8.47 <sup>7</sup>	-1.60 <sup>5</sup>	1.43 <sup>5</sup>
$\beta_5^b$	5.25 <sup>6</sup>	-2.24 <sup>7</sup>	1.07 <sup>5</sup>	4.37 <sup>6</sup>	-5.10 <sup>6</sup>	1.38 <sup>5</sup>	4.55 <sup>6</sup>	-4.14 <sup>6</sup>	1.32 <sup>5</sup>
$\beta_6^b$	-5.37 <sup>6</sup>	-1.19 <sup>5</sup>	1.13 <sup>6</sup>	-3.77 <sup>6</sup>	-1.44 <sup>5</sup>	6.84 <sup>6</sup>	-3.92 <sup>6</sup>	-1.37 <sup>5</sup>	5.83 <sup>6</sup>
$\beta_7^b$	3.89 <sup>1</sup>	-1.15 <sup>0</sup>	1.93 <sup>0</sup>	-1.98 <sup>1</sup>	-1.81 <sup>0</sup>	1.41 <sup>0</sup>	-8.71 <sup>2</sup>	-1.63 <sup>0</sup>	1.46 <sup>0</sup>
$\beta_8^b$	-1.57 <sup>2</sup>	-1.52 <sup>1</sup>	1.21 <sup>1</sup>	7.75 <sup>2</sup>	-5.49 <sup>2</sup>	2.10 <sup>1</sup>	7.60 <sup>2</sup>	-5.23 <sup>2</sup>	2.04 <sup>1</sup>
$\beta_9^b$	1.33 <sup>1</sup>	-9.17 <sup>2</sup>	3.57 <sup>1</sup>	3.90 <sup>2</sup>	-1.83 <sup>1</sup>	2.61 <sup>1</sup>	4.46 <sup>2</sup>	-1.70 <sup>1</sup>	2.59 <sup>1</sup>
$\beta_{10}^b$	-1.33 <sup>1</sup>	-3.59 <sup>1</sup>	9.31 <sup>2</sup>	2.07 <sup>2</sup>	-2.17 <sup>1</sup>	2.59 <sup>1</sup>	5.44 <sup>3</sup>	-2.23 <sup>1</sup>	2.34 <sup>1</sup>
$\beta_{11}^b$	-5.57 <sup>2</sup>	-2.02 <sup>1</sup>	9.02 <sup>2</sup>	-5.23 <sup>2</sup>	-2.14 <sup>1</sup>	1.09 <sup>1</sup>	-5.29 <sup>2</sup>	-2.09 <sup>1</sup>	1.03 <sup>1</sup>
$\beta_{12}^b$	3.29 <sup>2</sup>	-7.25 <sup>2</sup>	1.38 <sup>1</sup>	-4.27 <sup>2</sup>	-1.52 <sup>1</sup>	6.63 <sup>2</sup>	-3.76 <sup>2</sup>	-1.43 <sup>1</sup>	6.77 <sup>2</sup>
$\beta_{13}^b$	-7.94 <sup>2</sup>	-1.93 <sup>1</sup>	3.41 <sup>2</sup>	-7.34 <sup>2</sup>	-1.91 <sup>1</sup>	4.46 <sup>2</sup>	-7.30 <sup>2</sup>	-1.87 <sup>1</sup>	4.10 <sup>2</sup>
$\tau^2$				0.03			0.03		
$\lambda^2$				0.08			0.06		
$\phi$				40.00			38.82		
log.lik.	-29.75			-0.08			0.51		

Est., Estimate; log.lik., Log-likelihood.

wind direction)에 관한 정보도 포함되어 있다. 그러나 Figure 2.1(b)에 나타난 224개의 이산화황 관측 지점에서는 주풍향의 정보가 없다. 따라서 각 이산화황 관측지점의 풍향을 모형에 포함시키기 위해 풍속자료에서 관측된 주풍향을 토대로 가장 가까이 위치한 풍속 자료의 관측지점의 주풍향을 사용하였다. 풍속자료의 관측 지점에서의 주풍향은 7개 방위(북, 북동, 동, 남동, 남서, 서, 북서) 중 하나로 나타났으며, 남(S)풍이 주풍향인 관측지점은 없었다. 224개 지점에서의 주풍향은 북(N)풍이 89개의 지점, 북동(NE)풍이 23개의 지점, 동(E)풍이 7개의 지점, 남동(SE)풍이 7개의 지점, 남서(SW)풍이 19개의 지점, 서(W)풍이 44개의 지점, 그리고 북서(NW)풍이 35개의 지점으로 간주하였다. 주풍향에 따라 평균 이산화황의 차이를 살펴본 결과 남동풍, 서풍이 주풍향인 관측지점에서의 평균 이산화황이 다른 지점들에 비해 큰 것으로 나타났다. 따라서, 주풍향에 의한 평균 이산화황의 차이를 평균함수에 포함시킬 필요가 있으며 이를 위해 북풍을 기준범주(reference group)로 하여 가변수 처리하였다. 따라서 이산화황에 관한 선형회귀모형의 평균함수는 식 (4.1)과 유사한 형태를 이루며 다음과 같이 표현된다.

$$\begin{aligned} \mathbf{X}_b \boldsymbol{\beta}_b &= \beta_1^b + s_1^b \beta_2^b + s_2^b \beta_3^b + \left(s_1^b\right)^2 \beta_4^b + \left(s_2^b\right)^2 \beta_5^b + s_1^b s_2^b \beta_6^b + s_3^b \beta_7^b \\ &+ I_{NE} \beta_8^b + I_E \beta_9^b + \cdots + I_{NW} \beta_{13}^b. \end{aligned}$$

이산화황에 관한 선형회귀모형의 회귀계수들과 오차항의 공간 상관구조를 의미하는 모수를 추정된 결과는 Table 4.2에 나타나 있다. 풍속자료의 분석결과와 마찬가지로, 제3장에서 소개한 세가지 공분산함수들 중에서 로그우도값을 비교해본 결과 구형모형이 오차항의 공간 상관구조를 가장 잘 설명하는 것으로 나타났다. 로그우도값을 살펴보면, 잔차우도추정법에서의 지수모형, 가우시안모형, 구형모형의 로그우도값은 각각 -0.61, -0.55, -0.08으로 구형모형의 로그우도값이 가장 크게 나왔으며, 일반우도추정법에서의 로그우도값은 각각 -1.01, 0.46, 0.51로 구형모형의 로그우도값이 가장 크게 나왔다. 회귀계수

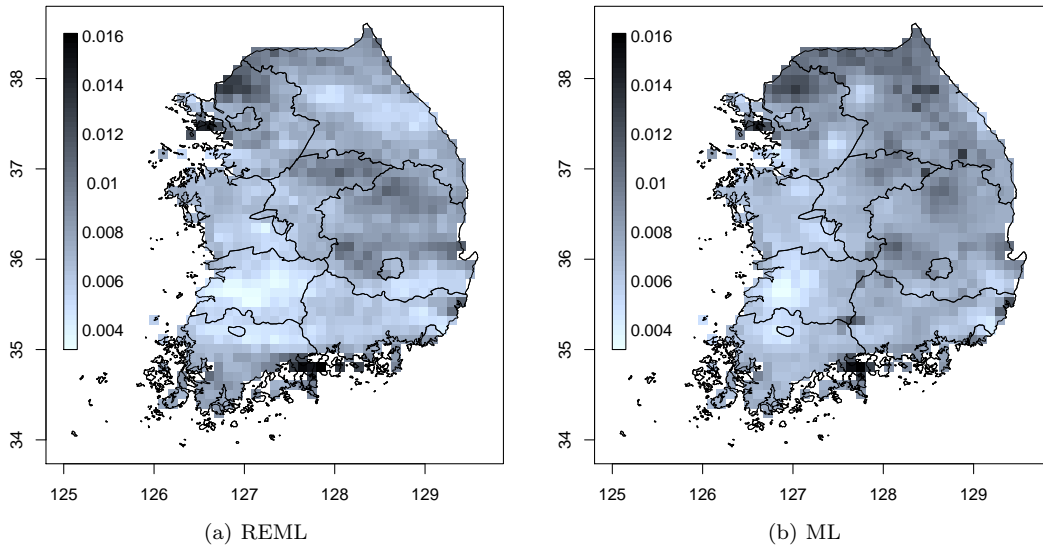


Figure 4.2. Prediction maps of Sulfur Dioxide(SO<sub>2</sub>) from spatial linear model based on spherical model.

의 추정에 있어서 우도를 바탕으로 한 두 가지 추정방법들 간에는 추정값에 대한 큰 차이는 보이지 않고 있다. 오차항의 분산-공분산 행렬에 포함된 모수들의 추정결과를 살펴보다도 방법론에 상관없이 거의 유사한 값으로 추정되었다. 로그 우도값으로 모형을 비교해보면, 잔차최대우도추정법(0.51)이 일반최대우도추정법(-0.08)보다 모형을 더 잘 설명한다. 그리고 전통적 선형회귀모형(-29.75)보다는 공간선형회귀모형이 보다 주어진 자료에 적합하다고 판단된다.

935개 격자지점의 추가적인 정보를 이용하여 각 추정방법 하에서의 적합된 모형으로 예측한 결과는 Figure 4.2에 나타나 있으며 이는 구형모형을 기반으로 적합된 모형들로부터 예측한 후 원 측정단위로 변환한 이산화황의 예측지도이다. 두 추정방법의 예측지도가 유사함을 알 수 있다. 예측지도를 자세히 살펴보면, 경기도 북부 지방과 충청북도 북부 지방, 태백산맥 일대, 경상북도 일대, 전라남도 여수, 경상남도 남해의 이산화황이 크게 예측되었고, 충청남도 내륙지방, 전라도 내륙지방의 이산화황은 낮게 예측되었다.

### 4.3. 풍속자료를 이용한 이산화황 자료의 계층모형 적합

본 절에서는 식 (3.6)에서 제시한 바와 같이 이산화황이 관측된 지점,  $\{s_i^b, i = 1, \dots, 224\}$ 에서의 풍속 예측값을 하나의 독립변수로 추가하여 이산화황의 공간 상관구조를 파악하기 위한 계층모형을 고려하고자 한다. 3.2절에서 설명한 이산화황의 크리깅에서 마찬가지로 3년간 이산화황의 평균 자료를 로그변환하였다. 먼저 이산화황이 관측된 지점  $\{s_i^b\}$ 에 대해 이산화황의 관측값과 풍속의 예측값에 대한 연관성을 살펴본 결과 피어슨 상관계수는 0.138로서 양의 선형관계가 있음을 확인할 수 있다(Figure 4.3). 즉, 풍속이 증가할수록 이산화황의 농도 역시 증가하는 경향이 있는 것으로 해석할 수 있다. 이에 대한 이론적인 해석이 가능한 지에 대한 부분은 논문의 범위를 넘어서는 것이라 판단하여 추후에 전문가의 자문을 받을 필요가 있겠다. 비교적 큰 표본크기로 인해 통계적인 유의성이 확보된 것이라 생각할 수도 있으나 주어진 자료를 분석한 결과로 판단컨대 이산화황을 예측함에 있어 풍속의 예측결과를 사용하는 것이 모

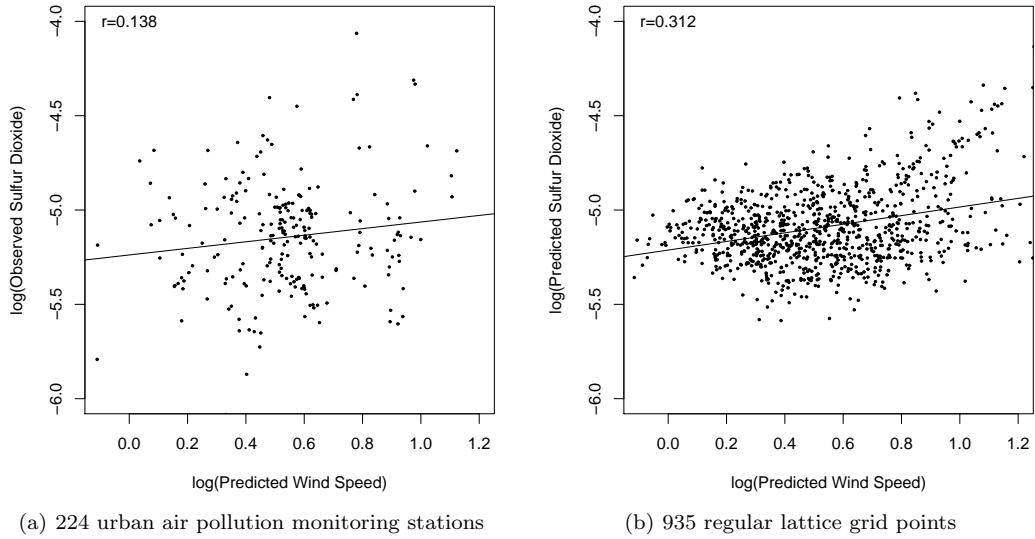


Figure 4.3. Scatter plots of observed(predicted) Sulfur Dioxide(SO<sub>2</sub>) versus predicted wind speed

형의 설명력을 높일 수 있으므로 공간선형회귀모형의 평균함수를 다음과 같이 고려하였다.

$$\begin{aligned} \mathbf{X}_b \boldsymbol{\beta}_c + \delta \widehat{\mathbf{z}}_{ba} = & \beta_1^c + s_1^b \beta_2^c + s_2^b \beta_3^c + \left(s_1^b\right)^2 \beta_4^c + \left(s_2^b\right)^2 \beta_5^c + s_1^b s_2^b \beta_6^c + s_3^b \beta_7^c \\ & + I_{NE} \beta_8^c + I_E \beta_9^c + \dots + I_{NW} \beta_{13}^c + \delta \widehat{\mathbf{z}}_{ba}. \end{aligned}$$

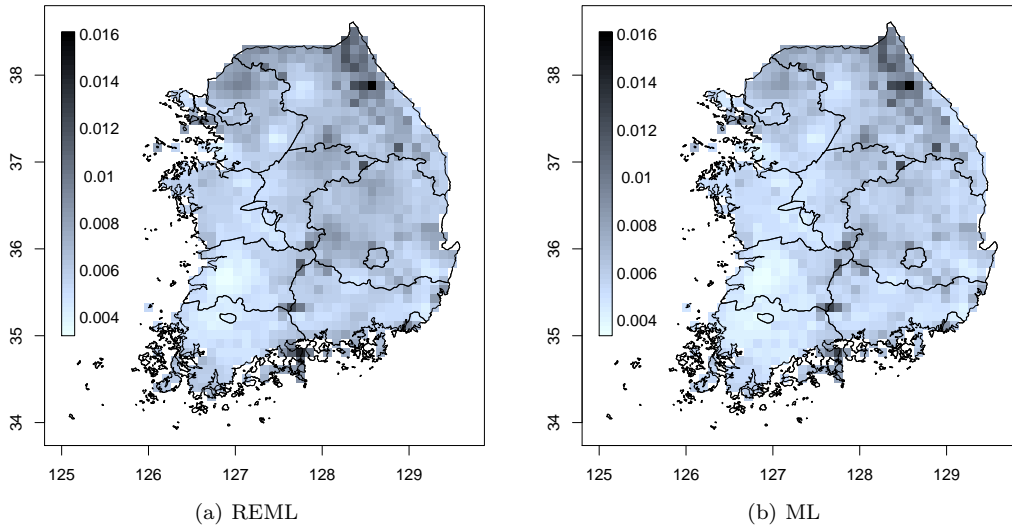
풍속자료를 이용해 이산화황의 계층모형을 구축한 결과는 Table 4.3에 나타나 있다. 앞서 설명한 이산화황만을 가지고 모형을 적합한 결과와 마찬가지로, 오차항의 공간 상관구조를 설명함에 있어서 세 가지 세미-베리오그램 모형들 중 구형모형이 가장 뛰어난 것으로 나타났다. 잔차최대우도추정법에서의 각 모형의 로그우도값을 비교해보면 지수모형, 가우시안모형, 구형모형의 로그우도값은 각각 0.98, 0.59, 1.23으로 구형모형의 로그우도값이 가장 크며, 일반최대우도추정법에서의 로그우도값은 각각 0.94, 1.94, 2.18로 역시 구형모형의 로그우도값이 가장 크다. 또한 잔차최대우도추정법과 일반최대우도추정법의 모수추정값에 대한 큰 차이가 없으며, 오차항의 분산-공분산 행렬에 포함된 모수들에도 큰 차이가 없음을 알 수 있다. 또한, 모형의 비교를 위하여 로그 우도값을 살펴본 결과, 잔차최대우도추정 방법(1.23)보다 일반최대우도추정방법(2.18)이 보다 우수한 것으로 판명되었으며, 전통적 선형회귀모형(-19.65)에 비해 공간선형모형이 모형을 더 잘 설명하는 것으로 나타났다. 공간 선형회귀모형의 적합에 대하여, 추정방법에 상관없이 북풍 대비 북동풍( $\beta_8^c$ ), 혹은 북풍 대비 남동풍( $\beta_{10}^c$ )은 다른 방향들에 비해 낮은 평균이산화황을 가지는 것으로 나타났다.

935개 격자지점의 추가적인 정보를 이용하여 각 추정방법 하에서의 적합된 모형으로 예측한 결과는 Figure 4.4와 같다. 구형모형으로 적합된 이산화황의 예측값은 원 측정단위로 변환된 것이다. 두 가지 추정방법 하에서 적합된 모형들을 기반으로 구성한 평균이산화황에 대한 예측지도는 거의 유사하다. 예측지도를 좀더 자세히 살펴보면, 두 모형 모두 경기도 북부 지방과 강원도 일대, 전라남도 여수, 경상남도 남해, 고도가 높은 덕유산, 지리산 일대의 이산화황이 크게 예측되었다. 또한 경기도 남부, 충청도 내륙지방, 전라도 내륙지방의 이산화황은 낮게 예측되었다.

**Table 4.3.** Parameter estimation based on the hierarchical model of Sulfur Dioxide(SO<sub>2</sub>) ( $a^b \equiv a \times 10^{-b}$ ).

	Classical linear model			Spatial linear model					
	Least Squares			REML(Spherical)			ML(Spherical)		
	Est.	95% C.I		Est.	95% C.I		Est.	95% C.I	
$\beta_1^c$	-5.80 <sup>0</sup>	-6.08 <sup>0</sup>	-5.53 <sup>0</sup>	-5.51 <sup>0</sup>	-5.91 <sup>0</sup>	-5.12 <sup>0</sup>	-5.55 <sup>0</sup>	-5.91 <sup>0</sup>	-5.18 <sup>0</sup>
$\beta_2^c$	3.96 <sup>4</sup>	-3.67 <sup>4</sup>	1.16 <sup>3</sup>	7.43 <sup>4</sup>	-4.59 <sup>4</sup>	1.94 <sup>3</sup>	7.23 <sup>4</sup>	-3.56 <sup>4</sup>	1.80 <sup>3</sup>
$\beta_3^c$	1.07 <sup>3</sup>	4.61 <sup>4</sup>	1.69 <sup>3</sup>	7.34 <sup>4</sup>	-2.74 <sup>4</sup>	1.74 <sup>3</sup>	7.50 <sup>4</sup>	-1.56 <sup>4</sup>	1.66 <sup>3</sup>
$\beta_4^c$	-1.54 <sup>5</sup>	-2.78 <sup>5</sup>	-2.94 <sup>6</sup>	-1.42 <sup>5</sup>	-3.70 <sup>5</sup>	8.48 <sup>6</sup>	-1.50 <sup>5</sup>	-3.53 <sup>5</sup>	5.32 <sup>6</sup>
$\beta_5^c$	-5.32 <sup>7</sup>	-6.37 <sup>6</sup>	5.31 <sup>6</sup>	8.87 <sup>7</sup>	-9.20 <sup>6</sup>	1.10 <sup>5</sup>	9.35 <sup>7</sup>	-8.06 <sup>6</sup>	9.93 <sup>6</sup>
$\beta_6^c$	-3.76 <sup>6</sup>	-1.00 <sup>5</sup>	2.51 <sup>6</sup>	-3.42 <sup>6</sup>	-1.37 <sup>5</sup>	6.83 <sup>6</sup>	-3.59 <sup>6</sup>	-1.27 <sup>5</sup>	5.57 <sup>6</sup>
$\beta_7^c$	7.28 <sup>1</sup>	-7.56 <sup>1</sup>	2.21 <sup>0</sup>	-2.08 <sup>1</sup>	-1.81 <sup>0</sup>	1.39 <sup>0</sup>	-2.84 <sup>2</sup>	-1.54 <sup>0</sup>	1.49 <sup>0</sup>
$\beta_8^c$	-2.74 <sup>1</sup>	-4.47 <sup>1</sup>	-9.99 <sup>2</sup>	-5.59 <sup>2</sup>	-2.61 <sup>1</sup>	1.49 <sup>1</sup>	-6.99 <sup>2</sup>	-2.63 <sup>1</sup>	1.23 <sup>1</sup>
$\beta_9^c$	-3.35 <sup>2</sup>	-2.60 <sup>1</sup>	1.94 <sup>1</sup>	-2.39 <sup>2</sup>	-2.58 <sup>1</sup>	2.10 <sup>1</sup>	-2.20 <sup>2</sup>	-2.46 <sup>1</sup>	2.02 <sup>1</sup>
$\beta_{10}^c$	-2.54 <sup>1</sup>	-4.78 <sup>1</sup>	-3.10 <sup>2</sup>	-6.54 <sup>2</sup>	-3.18 <sup>1</sup>	1.87 <sup>1</sup>	-9.19 <sup>2</sup>	-3.32 <sup>1</sup>	1.48 <sup>2</sup>
$\beta_{11}^c$	-2.33 <sup>1</sup>	-3.93 <sup>1</sup>	-7.28 <sup>2</sup>	-1.21 <sup>1</sup>	-3.01 <sup>1</sup>	5.84 <sup>2</sup>	-1.30 <sup>1</sup>	-3.02 <sup>1</sup>	4.21 <sup>2</sup>
$\beta_{12}^c$	-1.98 <sup>1</sup>	-3.41 <sup>1</sup>	-5.41 <sup>2</sup>	-1.54 <sup>1</sup>	-3.29 <sup>1</sup>	2.05 <sup>2</sup>	-1.57 <sup>1</sup>	-3.20 <sup>1</sup>	6.31 <sup>2</sup>
$\beta_{13}^c$	-2.90 <sup>1</sup>	-4.33 <sup>1</sup>	-1.47 <sup>1</sup>	-1.83 <sup>1</sup>	-3.57 <sup>1</sup>	-8.03 <sup>3</sup>	-1.91 <sup>1</sup>	-3.55 <sup>1</sup>	-2.73 <sup>2</sup>
$\delta$	4.86 <sup>1</sup>	2.71 <sup>1</sup>	7.02 <sup>1</sup>	2.79 <sup>1</sup>	-5.52 <sup>2</sup>	6.13 <sup>1</sup>	3.01 <sup>1</sup>	-2.98 <sup>3</sup>	6.04 <sup>1</sup>
$\tau^2$				0.03			0.03		
$\lambda^2$				0.07			0.05		
$\phi$				40.00			37.93		
log.lik.	-19.65			1.23			2.18		

Est., Estimate; log.lik., Log-likelihood.

**Figure 4.4.** Prediction maps of Sulfur Dioxide(SO<sub>2</sub>) from the hierarchical model based on spherical model.

#### 4.4. 선형회귀모형의 비교

본 절에서는 4.2절에서 구축한 이산화황 모형(식 (3.4))과 4.3절에서 예측한 풍속자료를 이용한 이산화황의 계층모형(식 (3.5))을 비교하고자 한다. Model 1은 이산화황의 전통적 선형회귀모형, Model 2는

**Table 4.4.** Model selection criteria for comparison of the models considered in this study.

	Classical linear model		Spatial linear model	
	Model 1	Model 2	Model 3	Model 4
AIC	87.50	69.29	34.98	33.64
BIC	132.27	120.47	96.39	98.46
CV(1)	94.20	92.41	84.38	85.27
ICC	0.03	0.15	0.49	0.48
PRESS( $\times 10^4$ )	8.91	8.25	5.92	6.10

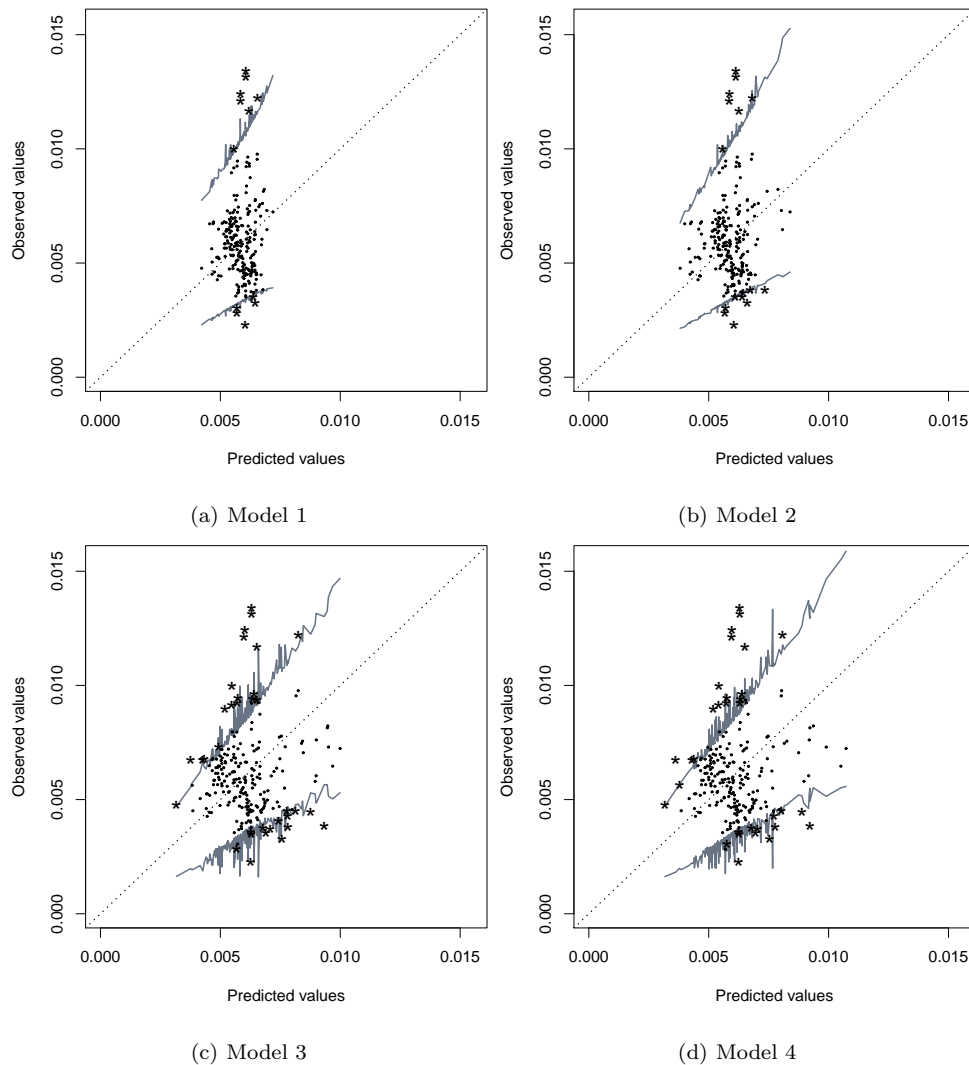
Model 1: classical linear regression model with  $\epsilon_b \sim N_{n_b}(\mathbf{0}, \sigma_b^2 \mathbf{I}_{n_b})$ ; Model 2: hierarchical model with  $\epsilon_c \sim N_{n_b}(\mathbf{0}, \sigma_c^2 \mathbf{I}_{n_b})$ ; Model 3: spatial linear regression model with  $\epsilon_b \sim N_{n_b}(\mathbf{0}, \Sigma_b)$ ; Model 4: hierarchical model with  $\epsilon_c \sim N_{n_b}(\mathbf{0}, \Sigma_c)$ .

CV(1): empirical coverage probability(%) from leave-one-out cross-validation method; ICC: intra-class correlation coefficient; PRESS: predicted residual sum of squares.

계층모형을 사용한 이산화량의 전통적 선형회귀모형, Model 3은 공간성 상관성을 고려하여 이산화량의 일반최대우도법을 이용한 구형모형, Model 4는 공간적 상관성을 고려하여 계층모형을 사용한 이산화량의 일반최대우도법을 이용한 구형모형이다. 모형들 간의 비교를 위해 사용된 방법은 먼저, AIC(Akaike information criterion)과 베이저안 정보기준(Bayesian information criterion)을 사용한다. 두 번째로는, 교차검증방법(leave-one-out cross-validation; CV(1))을 사용한다. 세 번째로, 교차검증방법으로 얻어진 예측값과 실제값 사이의 급내상관계수(intra-class correlation coefficient; ICC)를 구하였다. 급내상관계수는 쌍으로 측정된 두 변수의 상관성을 보는 척도이다. 마지막으로 예측오차제곱합(predicted residual sums of squares; PRESS)을 구하였다. PRESS는 교차검증방법으로 얻어진 예측값과 실제값의 차이의 제곱합으로서 다음과 같이 표현된다.

$$\text{PRESS} = \sum_{i=1}^{n_b} \left[ Z(\mathbf{s}_i) - \widehat{Z}(\mathbf{s}_{(i)}) \right]^2.$$

여기서,  $\widehat{Z}(\mathbf{s}_{(i)})$ 는  $\mathbf{s}_i$ 를 제외한  $n_b - 1$ 개의 지점들의 정보를 이용하여 모형을 구한 후  $\mathbf{s}_i$ 의 지점에서 구한 예측값을 의미한다. Table 4.4를 통해 알 수 있듯이 AIC와 BIC의 경우에 전통선형회귀모형에서는 Model 1에 비해 Model 2가 모형을 더 잘 설명하는 것으로 나타났으며, 공간적 연관성을 고려한 선형회귀모형에서는 Model 3과 Model 4가 큰 차이가 없다. 공간적 연관구조를 모형에 포함한 경우와 그렇지 않은 경우를 각각 고려하여 교차검증방법을 적용하여 각 지점에 대한 예측구간을 구하였으며 실제 이산화값이 해당 예측구간에 얼마나 많이 포함되어지는지를 비교 및 평가하였다(Figure 4.5 참조). Figure 4.5(a)의 전통적 선형회귀모형(Model 1)이 Figure 4.5(c)의 공간 선형회귀모형 (Model 3)보다 실제값과 예측값이 일치하는 것으로 나타났다. 전통적 선형회귀모형의 경우에는 95% 신뢰구간 내에 전체 자료의 94.20% 정도가 포함되었지만 잔차최대우도추정방법의 경우에는 84.38% 정도 포함되는 것으로 나타났다. 그러나 계층모형을 고려한 경우, Figure 4.5(b)의 Model 2는 Model 1에 비해 독립변수를 하나 더 추가했음에도 92.41%로 나타나 포함확률이 낮아졌다. 그리고 Model 4(Figure 4.5(d))의 경우 교차검증방법으로부터 얻은 포함확률이 85.27%로, Model 3의 공간 선형회귀모형(84.38%)보다 예측력이 높게 나타났다. 따라서, 교차검증방법으로 살펴본 결과 공간 선형회귀모형에서 다양한 관측망에서 얻은 자료를 활용하여 계층모형을 구축하는 것이 예측력을 높인다는 것을 알 수 있다. 교차검증방법으로 얻어진 예측값과 실제값 사이의 급내상관계수를 구한 경우에는 Model 1의 전통적 선형회귀모형의 경우에는 0.03이고 Model 3의 공간적 선형회귀모형의 경우에는 0.49로 공간적 선형회귀모형의 경우가 더 잘 추정되었다고 할 수 있다. 마찬가지로 계층모형에서도 Model 2의 전통적 선형회귀모형의 급내상



**Figure 4.5.** Calibration plots of Sulfur Dioxide(SO<sub>2</sub>) and its predicted intervals by leave-one-out cross-validation. Note that Model 1 classical linear regression model with  $\epsilon_b \sim N_{n_b}(\mathbf{0}, \sigma_b^2 \mathbf{I}_{n_b})$ ; Model 2 hierarchical model with  $\epsilon_c \sim N_{n_b}(\mathbf{0}, \sigma_c^2 \mathbf{I}_{n_b})$ ; Model 3 spatial linear regression model with  $\epsilon_b \sim N_{n_b}(\mathbf{0}, \boldsymbol{\Sigma}_b)$ ; Model 4 hierarchical model with  $\epsilon_c \sim N_{n_b}(\mathbf{0}, \boldsymbol{\Sigma}_c)$ . \* indicates the observations outside the prediction intervals.

관계수는 0.15이고 Model 4의 공간적 선형회귀모형의 경우에는 0.48으로 공간적 선형회귀모형의 경우의 모형의 설명력이 더 높다고 할 수 있다. 마지막으로 교차검증방법으로 얻어진 예측값과 실제값의 차이의 제곱합인 PRESS를 비교하였는데, 급내상관계수와 마찬가지로 Model 1 보다 Model 3의 PRESS 통계량 값이 더 작으며, Model 2 보다 Model 4의 값이 더 작다. Model 3과 Model 4의 PRESS 통계량의 값에는 큰 차이가 없다. 교차검증방법의 포함확률과 급내상관계수, PRESS 통계량의 결과가 상반된 이유는 교차검증방법의 포함확률은 단순히 예측값의 95% 신뢰구간에 실제값이 포함된 정도를 의미하지만, 급내상관계수나 PRESS 통계량의 경우에는 예측값과 실제값이 얼마나 일치하느냐를 살펴보는 척도

이기 때문이다. 특히, 본 논문의 경우에는 Model 1과 Model 2의 신뢰구간의 길이가 Model 3과 Model 4 보다 길기 때문에 이와 같은 결과가 나타난 것이다.

## 5. 결론

본 연구에서는 다양한 관측망에서 얻은 자료들을 이용하여 계층모형을 구축하는 방법을 제시하였다. 크리깅을 하고자 하는 관심 자료에 대하여 연관이 있는 다른 관측망에서 얻어진 자료가 있다면, 다른 관측망에서 얻어진 자료를 가지고 관심 자료를 분석하는데 이용하는 것이다. 일반적으로 선형회귀모형에서 독립변수가 하나 더 추가가 되면 모형의 설명력이 더 높아지게 된다. 다양한 관측망에서 얻은 두 자료의 연관성이 클 때, 계층모형을 구축하면 그렇지 않은 경우에 비해 예측력이 높아질 것은 명백하다. 즉 한 변수가 다른 변수를 설명할 수 있다면 계층모형을 구축해 보다 신뢰성 있는 예측을 할 수 있게 된다. 이를 확인하기 위하여 본 논문에서는 도시대기측정망에서 관측된 이산화황을 분석하는데 지상기상관측망에서 얻어진 풍속 및 풍향자료를 모형에 포함시켰다. 이산화황의 관측망, 즉 도시대기측정망은 주로 도시에서 관측이 이루어지므로 농촌지역에서는 관측이 되지 못하는 한계가 있다. 이에 이산화황을 단독으로 분석하는 것 보다는 이산화황과 양의 상관관계가 있으며, 우리나라 전역에 걸쳐 측정된 풍속 및 풍향의 관측망인 지상기상관측망을 이용하여 계층모형을 구축한다면 이산화황의 예측력을 높일 수 있다고 판단하였다. 이산화황의 관측 지점이 도시에 편중되어있어 이산화황을 단독으로 공간 선형회귀모형을 구축하는데 예측값이 다소 과소평가된 부분이 있었다. 이에 풍속을 이용해 계층모형을 고려하여 공간선형모형을 구축한 결과 예측값의 범위가 원 자료와 크게 다르지 않았다. 그러나 풍속은 고도가 높은 지역(태백산맥 일대, 지리산 일대, 덕유산 일대 등)에서 많은 영향을 받게 되므로 이산화황의 예측값이 고도가 높은 지역에서 크게 나타나게 되었다. 또한, 계층모형을 사용하는 것이 그렇지 않은 경우보다 얼마나 예측력이 좋은지 확인하기 위해 AIC, BIC, 교차검증방법, 급내상관계수, 예측오차제곱합을 비교하였다. 그 결과 공간적 연관구조를 포함한 경우가 그렇지 않은 경우에 비해 신뢰성 있는 예측성능을 보였다. 또한 공간 선형모형에서 계층모형을 사용하는 것이 그렇지 않은 경우보다 AIC와 교차검증방법에서는 예측력이 높음을 알 수 있었다.

본 논문에서는 풍속이 이산화황에 미치는 영향을 단순히 양의 상관관계로만 판단했다. 추후에 전문가의 자문을 구하여 이산화황에 직접적인 영향을 미치는 다른 관측망의 자료를 사용하여 계층모형을 구축한다면 좀 더 예측력이 높아질 것이다. 또한, 본 연구에서 지상기상관측망에서 관측된 2003년부터 2007년까지의 5년간 평균풍속자료와 도시대기측정망에서 관측된 2005년부터 2007년까지의 3년간 평균이산화황자료를 사용하였다. 두 관측망에서 얻은 자료의 관측 시기가 동일하지 못한 한계가 있으나 그럼에도 불구하고 두 자료를 이용해 계층모형을 구축해 보다 좋은 결과를 보였다. 서로 다른 두 관측망에서 관측된 자료의 관측 시기가 동일하다면 보다 예측성능이 높아질 것이라고 생각한다.

## References

- Banerjee, S., Gelfand, A. E. and Carlin, B. P. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman & Hall/CRC.
- Cho, H. L. and Jeong, J. C. (2006). Application of spatial interpolation to rainfall data, *The Journal of GIS Association of Korea*, **14**, 29–41.
- Cho, J. Y., Choi, S. B. and Kim, K. K. (2001). Comparative study on the prediction of geo-statistics with general statistics, *The Journal of Korea Data Analysis Society*, **3**, 41–49.
- Choi, J. S. and Park, M. S. (2009). Spatial prediction based on the Bayesian kriging with Box-Cox transformation, *Communications of the Korean Statistical Society*, **16**, 851–858.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*, John Wiley & Sons, New York.

- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York.
- Han, J. S., Kim, Y. M., Ahn, J. Y., Kong, B. J., Choi, J. S., Lee, S. U. and Lee, S. J. (2004). Spatial distribution and variation of long-range transboundary air pollutants flux during 1997~2004, *Journal of Korean Society for Atmospheric Environment*, **22**, 99–106.
- Heo, T. Y. and Park, M. S. (2009). Bayesian spatial modeling of precipitation data, *The Korean Journal of Applied Statistics*, **22**, 425–433.
- Heo, T. Y., Park, M. S., Eom, J. K. and Oh, J. S. (2007). A study on the prediction of traffic counts based on shortest travel path, *The Korean Journal of Applied Statistics*, **20**, 459–473.
- Hohn, M. E. (1999). *Geostatistics and Petroleum Geology*, Kluwer Academic Publishers.
- Huh, T. Y., Suh, E. H. and Kwon, W. T. (2004). Spatial analysis on the rainfall data by utilizing Variogram models, *The Journal of Korea Data Analysis Society*, **6**, 473–491.
- Jeong, S. H., Park, M. S. and Kim, K. W. (2010). Spatial prediction of wind speed data, *The Korean Journal of Applied Statistics*, **23**, 345–356.
- Jung, J. Y., Jin, S. H. and Park, M. S. (2008). Precipitation analysis based on spatial linear regression model, *The Korean Journal of Applied Statistics*, **21**, 1093–1107.
- Kim, B. S., Ku, C. Y. and Choi, J. M. (2010). Population distribution estimation using regression-kriging model, *The Korean Geographical Society*, **45**, 806–819.
- Kim, H. Y. (2010). A study on the improvement of the accuracy of photovoltaic facility location using the geostatistical analysis, *Journal of the Korean Association of Geographic Information Studies*, **13**, 146–156.
- Kim, K. K. and Choi, S. B. (2000). A study on the prediction of Geo-statistical methods using environmental data, *The Journal of Korea Data Analysis Society*, **4**, 499–510.
- Korea Meteorological Administration Homepage. <http://www.kma.go.kr>.
- Ministry of Environment (2007). Annual Report of Air Quality in Korea.
- National Institute of Environmental Research. <http://www.nier.go.kr>.
- Park, N. W. and Jang, D. H. (2008). Mapping of temperature and rainfall using DEM and multivariate kriging, *The Korean Geographical Society*, **43**, 1002–1015.
- Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods For Spatial Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- Shin, K. I., Choi, B. H. and Lee, S. E. (2007). Evaluations of small area estimations with/without spatial terms, *The Korean Journal of Applied Statistics*, **20**, 229–244.



# 다양한 관측네트워크에서 얻은 공간자료들을 활용한 계층모형 구축

최지은<sup>a</sup> · 박만식<sup>1,a,b,c</sup>

<sup>a</sup>성신여자대학교 통계학과, <sup>b</sup>성신여자대학교 통계연구소, <sup>c</sup>성신여자대학교 기초과학연구소

(2012년 12월 26일 접수, 2012년 12월 31일 수정, 2013년 1월 3일 채택)

---

## 요약

지리통계자료는 관측지점이 지도 상에 점으로 표현되고 그 지점에서만 자료가 관측되는 측정값이다. 이러한 지리통계자료는 매우 다양한 관측망에서부터 얻어진다. 지리통계자료를 분석하고 예측함에 있어서 하나의 자료만 이용하는 것보다는 유사한 패턴을 갖는 다른 관측망에서 얻어지는 여러 자료들을 함께 사용한다면 예측력을 향상시킬 수 있을 것이다. 본 논문에서는 서로 다른 관측망에서 얻은 두 가지의 공간자료를 이용하여 분석 및 예측하고 이를 위해 공간적 연관성을 파악할 수 있는 적절한 계층모형을 구축하였다. 그리고 선형회귀모형에 근간을 둔 크리깅 결과와 계층모형 하에서의 결과를 여러 검증방법을 통해 비교하였다. 이 논문에서는 도시대기측정망에서 측정된 이산화황과 지상기상관측망에서 측정된 풍속자료를 이용하여 계층모형을 구축하고 이산화황만을 이용한 선형모형과 비교하였다. 또한 각 모형에 의한 이산화황 예측지도를 구성하였다.

주요용어: 이산화황, 풍속, 계층모형, 크리깅, 공간적연관성, 교차검증방법.

---

---

이 논문은 2010년도 성신여자대학교 학술연구조성비 지원에 의하여 연구되었음.

<sup>1</sup>교신저자: (136-742) 서울특별시 성북구 보문로 34다길2, 성신여자대학교 자연과학대학 통계학과, 조교수.

E-mail: mansikpark@sungshin.ac.kr