

# On the Geometric Anisotropy Inherent In Spatial Data

Hye Ji Go<sup>a</sup> · Man Sik Park<sup>a,b,1</sup>

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

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

(Received August 8, 2014; Revised September 22, 2014; Accepted September 29, 2014)

---

## Abstract

Isotropy is one of the main assumptions for the ease of spatial prediction (named kriging) based on some covariance models. A lack of isotropy (or anisotropy) in a spatial process necessitates that some additional parameters (angle and ratio) for anisotropic covariance model be obtained in order to produce a more reliable prediction. In this paper, we propose a new class of geometrically extended anisotropic covariance models expressed as a weighted average of some geometrically anisotropic models. The maximum likelihood estimation method is taken into account to estimate the parameters of our interest. We evaluate the performances of our proposal and compare it with an isotropic covariance model and a geometrically anisotropic model in simulation studies. We also employ extended geometric anisotropy to the analysis of real data.

Keywords: spatial association, isotropy, geometric anisotropy, spherical covariance model, spatial prediction, maximum likelihood estimation.

---

## 1. 서론

일반적으로 인접한 위치에 놓인 지점들에서 얻은 측정값들은 멀리 떨어진 지점들의 그것에 비해 매우 유사한 경향을 보인다. 그리고 지점들 간의 거리가 늘어날수록 유사성의 정도가 감소하게 된다. 이러한 공간 상관성(spatial association)을 이용하여 미지의 지점(location)에 대한 예측을 가능케 하는 공간 통계방법은 기본적으로 분석의 용이성을 위해 몇 가지 중요한 가정을 고려하게 된다. 대표적으로는 정상성(stationarity)과 등방성(isotropy) 등을 들 수 있겠다. 시계열 자료분석에서의 약정상성(weak stationarity)은 추세(trend)가 일정하고 분산(variance)이 유한하며 임의의 두 시점의 관측값 사이의 공분산(covariance)은 오직 시차(time lag)에만 의존하는 성질을 의미한다. 공간통계학에서의 정상성은 추세가 일정하고 임의의 두 지점에서 관측한 측정값 사이의 공분산은 오직 상대적 위치(relative position of locations)의 차이, 즉 거리(distance)와 방향(direction)에만 의존하는 성질로 확장된다. 정상성 하에서 두 관측지점의 공분산이, 두 지점이 형성하는 방향과는 상관없이 오직 거리에만 의존하는 성질을 등방성이라 한다. 따라서 등방성의 가정 하에서는 특정 관측지점과 동일한 거리만큼 떨어진 모든 관측 지점들은 방향에 상관없이 동일한 정도의 공간 상관성을 가지게 된다.

---

This paper is a condensed version of the first author's thesis for the master's degree.

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

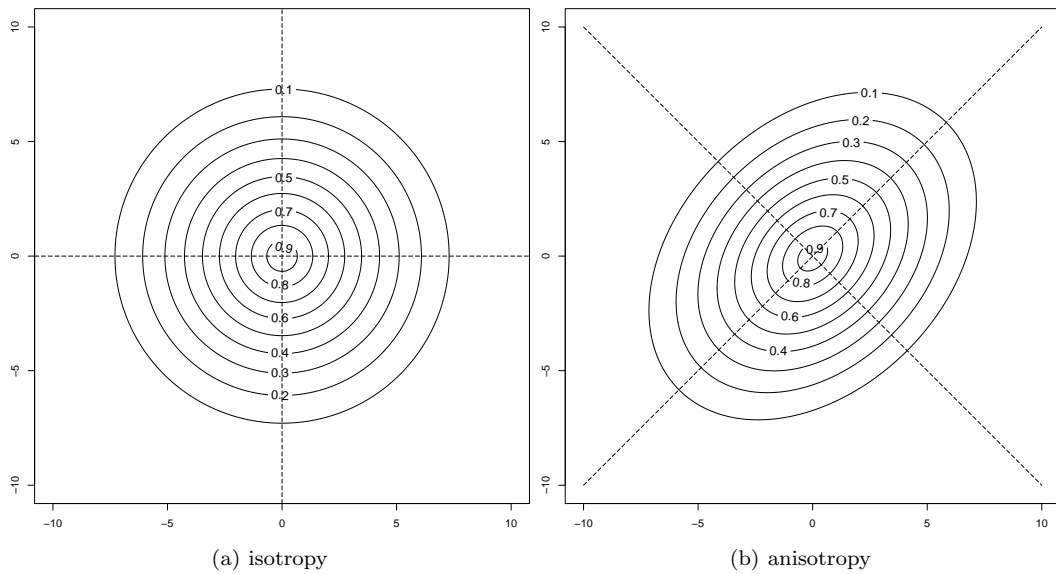


Figure 1.1. Contour plots of spatial association under isotropy and (geometric) anisotropy.

특정한 방향에서의 공간 상관성이 다른 방향에 비해 강하게 나타나는 경우를 고려해 보자. 예를 들면, 동일한 거리만큼 떨어져 있는 세 쌍의 관측지점이 있는데, 이들 중 두 쌍의 관측지점은 동-서로, 나머지 한 쌍의 관측지점은 남-북으로 위치해 있다고 하자. 만약 동-서로 위치한 각 쌍(두 지점)의 공간 상관성이 남-북으로 놓인 지점의 그것보다 강하다면 같은 방향을 형성하는 두 쌍의 관측지점은 동일한 거리 또한 유지하고 있으므로 동일한 공간 상관성을 가지게 되나 이외는 다른 방향을 가진 두 관측지점의 공간 상관성은 거리는 동일하다 하더라도 상이한 공간 상관성을 갖게 된다. 이와 같이 오직 거리에만 의존하는 공간 상관성이 실현되지 않는, 즉 등방성이 충족되지 않는 경우가 발생하게 된다. 이를 비등방성(anisotropy)이라 한다. 이러한 비등방성은 크게 기하학적 비등방성(geometric anisotropy)과 구역 비등방성(zonal anisotropy)으로 나뉜다. 기하학적 비등방성은 특정한 공간 상관성 구조 하에서 방향에 따라 상관성의 강도가 달라지는 현상을 의미한다. 하지만 구역 비등방성은 공간 상관성의 구조 자체가 방향에 따라 변화하는 복잡하고 다양한 현상을 의미하며 기하학적 비등방성 이외의 모든 비등방성을 포함한다. Figure 1.1은 등방성과 기하학적 비등방성 하에서의 공간 상관성을 시각적으로 표현한 것이다.

최근 공간 상관성과 관련하여 국내에서 이루어진 연구는 다음과 같다. Shin 등 (2003)은 충청남도 부여지방의 온실에서 얻어진 온실가루자료를 이용하여 공간 모형을 선택하는 방법을 연구하였고, Kim (2005)는 서울특별시 미세먼지 농도자료를 이용하여 공간 상관성에 관한 연구를 하였다. 또한, Jung 등 (2008)은 강수량 자료에 대해 공간적 상관 구조를 포함하는 선형회귀모형을 고려하여 최소제곱법과 최대우도추정방법으로 남한지역의 공간적 특성을 적절하게 파악할 수 있는 공간 모형을 탐색 및 비교하였고, Park 등 (2010)은 강우량 자료와 상관성이 높은 이차변수로 일평균풍속, 상대습도, 이슬점 온도 등을 고려하여 적절한 공간 모형을 선택하기 위한 방법을 연구하였다. 뿐만 아니라 Choi 등 (2010)은 각 지역별로 발생하는 범죄건수 자료에 관해 포아송분포, 음이항 분포, 일반화 포아송 분포 등을 이용하여 공간가산자료모형을 적합시키는 연구를 하였고, Jeong 등 (2010)은 기상청의 풍속자료의 선형회귀모형을 이용하여 공간적 특성을 파악할 수 있는 여러 가지 모형을 비교하였으며, Park 등 (2013)은 대중교통

수단과 출발지에 대한 공간 상관성에 관하여 실증분석을 실시하였다.

비등방성 모형 적용에 관한 연구는 다음과 같다. Budrikaite와 Ducinskas (2005)은 발틱해의 연안지역 염분자료를 이용하여 공간과정을 모형화하기 위해 기하학적 비등방성 베리오그램(anisotropic variogram)을 적용하였고, Kruminiene (2006)는 크로니안 석호 깊이(curonian lagoon depth)자료에 기하학적 등방성 베리오그램과 비등방성 베리오그램을 적용하여 비교하여 기하학적 비등방성 모형을 적용한 경우에 모형적합과 예측성능이 향상되었음을 밝혔다. 또한, Choi와 Kim (2006)은 방향성 세미베리오그램(directional semivariogram)의 계산 및 비등방성을 이용한 해수면 난류수 유동방향을 탐지하였고, Kim (2010)은 공간통계분석의 교통량 예측을 위해 비등방성을 적용한 분석을 실시하였으며 예측의 정확도 향상을 위한 새로운 방안을 제시하였다. Kubota와 Tarumi (2010)는 기하학적 비등방성을 탐지하고 보정하는 방법을 제안하여 모의실험을 수행하였고, Choi와 Um (2012)은 비등방성 베리오그램을 통해 이산화탄소의 유동 방향을 결정하여 계절별 변화추세를 확인하였다.

위의 연구와 같이 실제 공간 자료에서 특정한 방향에서 공간 상관성이 강하게 나타나는 경우에 이를 고려한 분석이 필요하고 공간 상관성을 내포하는 여러 모수들의 구간 추정을 위해 붓스트랩 방법(bootstrapping method)을 이용할 수 있다. 공간 분석에 붓스트랩을 적용한 연구들은 다음과 같다. Lahiri (2003)는 격자 자료의 블록 붓스트랩 방법을 논하였으며, Loh와 Stein (2004)은 공간점 패턴 자료에 붓스트랩을 이용하여 신뢰구간을 얻는 새로운 방법을 제안하여 블록 붓스트랩(block bootstrap)과의 구간추정 성능을 비교하였다. Liansheng 등 (2006)은 공분산행렬(covariance matrix)의 출레스키 분해(Cholesky decomposition)을 이용한 모수적(parametric) 붓스트랩을 자료에 적용하였다. 또한, Kang 등 (2008)은 크리깅 분산(kriging variance)의 과소추정 문제에 관하여 공간 붓스트랩 자료를 이용한 모의실험을 수행하였으며, Lin 등 (2010)은 선형회귀모형의 공간 상관성의 가설 검정을 위한 잔차기반의 공간 붓스트랩 검정을 하였다. 마지막으로, Choi와 Park (2013)은 공간 선형회귀모형에서 모수의 신뢰구간을 구하기 위해 붓스트랩과 잭나이프 방법을 이용하였다. 한편, 공간 상관성에 관한 추정 결과를 기반으로 공간예측에 관한 연구를 보면, Kim과 Choi (2000)는 일산화탄소 자료의 평균을 이용한 일반통계분석법과 공간통계분석법의 공간분석을 실시하여 예측력을 비교하였다. Park 등 (2011)은 기온 및 강수량자료를 이용한 공간 분포도 작성에 수치표고모형(digital elevation model; DEM)과 지구통계학적 공간예측기법의 적용 가능성을 연구하였다. 이처럼 공간 분석을 위해 붓스트랩을 적용한 연구 및 공간예측에 관한 연구가 다양하게 이루어지고 있다.

본 연구에서는 하나 이상의 방향에서 강한 공간 상관성이 나타나는 경우를 고려한 확장된 기하학적 비등방성 모형을 특정한 등방성 모형의 가중평균의 형태로 제안하고자 한다. 여러 다양한 등방성 모형들의 가중평균으로 비등방성 모형을 고려한다면 비등방성의 일반적인 형태인 구역 비등방성에도 적용할 수 있다. 본 논문에서는 주제의 명확성을 위하여, 확장된 형태의 기하학적 비등방성 모형을 적용하여 공간적 상관구조를 나타내는 여러 모수들을 추정하기 위해 최대우도추정법을 사용하였다. 이 방법은 모수의 점추정은 가능하나 구간추정이 불가능하여 신뢰구간을 산정하기 위해 붓스트랩 방법을 이용하였다. 또한, 이의 추정된 결과를 토대로 미지의 지점에 관한 공간예측을 통해 예측지도(predict map)와 교차검증(cross validation; CV)을 수행하여 등방성 및 비등방성 모형을 비교, 평가하고자 한다.

본 논문의 순서는 다음과 같다. 제 2장에서는 공간 상관성에 관한 등방성 및 비등방성 등의 특성에 관하여 설명하고, 확장된 기하학적 비등방성 모형을 적용하기 위한 방법을 설명한다. 제 3장에서는 본 연구에서 제안한 확장된 비등방성 모형 적용한 모의실험을 수행한다. 또한, 제 4장에서는 제안한 모형을 적용하기 위해 실증자료를 이용한 실증 분석을 실시하고, 마지막으로 제 5 장에서는 본 연구의 결론과 향후 연구 방향을 제시하고자 한다.

## 2. 확장된 비등방성 모형의 제안

### 2.1. 등방성 및 비등방성 모형

공간 상관성이 임의의 두 관측지점 사이의 거리에만 의존하는 것을 등방성이라 한다. 즉, 등방성은 모든 방향에서 동일한 공간 상관성을 가지고 있는 것을 의미한다. 등방성 하에서 공간 자료의 모형화를 하려면 너겟( $\theta_1$ : nugget), 부분문턱( $\theta_2$ : partial sill), 범위( $\theta_3$ : range)의 3 가지 모수(공통모수)를 추정해야 한다. 너겟은 측정오차에 기인하는 변동성을 의미하며, 부분문턱은 공간 자료 간의 거리가 멀어질수록 상관성이 작아지는 분산의 일부분이고, 범위는 각 관측값 사이의 공간 상관성을 띄는 최대거리를 의미한다. 여기서, 너겟과 부분문턱의 합은 문턱(sill)이 되며, 이는 자료의 분산을 나타낸다. 공간 상관성의 측도로 공분산 모형(covariance model)을 이용하여 3가지 모수를 추정할 수 있다. 본 연구에서는 구형(spherical) 공분산모형을 이용하여 분석을 진행하였다(공간통계학의 이론적 배경은 Banerjee 등 (2004), Cressie (1993), Haning (2003) 등을 참고하기 바란다).

$d$ -차원 공간 상에 존재하는 임의의 관측지점들로부터 얻은 자료는 다음과 같이 표현된다.

$$\left\{ Z(\mathbf{s}_i); \mathbf{s}_i = (s_{1i}, \dots, s_{di})^T, \quad i = 1, 2, \dots, n \right\}. \quad (2.1)$$

여기서, 임의의 두 지점  $\mathbf{s}_i = (s_{1i}, \dots, s_{di})^T$ ,  $\mathbf{s}_j = (s_{1j}, \dots, s_{dj})^T$ 의 상대적 위치 차이는  $\mathbf{h} = (h_1, \dots, h_d)^T = (s_{1i} - s_{1j}, \dots, s_{di} - s_{dj})^T$ 이다. 이러한 지점들 사이의 유클리드 거리(Euclidean distance)는  $\|\mathbf{h}\| = \sqrt{\mathbf{h}_1^2 + \dots + \mathbf{h}_d^2}$ 이고 등방성 하에서의 구형 공분산모형은 다음과 같이 나타낼 수 있다.

$$\text{Cov}_0(\mathbf{h}) = \theta_1 \times \mathbf{I}(\|\mathbf{h}\| = 0) + \theta_2 \times \left[ 1 - \frac{3}{2} \left( \frac{\min(\theta_3, \|\mathbf{h}\|)}{\theta_3} \right) + \frac{1}{2} \left( \frac{\min(\theta_3, \|\mathbf{h}\|)}{\theta_3} \right)^3 \right]. \quad (2.2)$$

여기서,  $\mathbf{I}(\cdot)$ 은 지시함수(indicator function)을 의미한다. 식 (2.1)로부터 위도와 경도를 고려하는 2-차원 공간자료에 대해 임의의 두 지점 간의 유클리드 거리는  $\|\mathbf{h}\| = \sqrt{h_1^2 + h_2^2}$ 이다.

하지만, 현실적으로 공간 상관성에 영향을 줄 수 있는 요인으로 방향성을 고려해야 하는 경우가 있다. 즉, 등방성이 위배되는, 비등방성을 고려해야 하는 경우를 말한다. 본 연구에서 다루는 비등방성 모형은 기하학적 비등방성 모형이며, 이는 공간 상관성이 강한 방향과 직각인 방향에서 같은 문턱값을 갖지만 범위가 다른 모형을 의미한다. Figure 1.1(b)에서 강한 공간 상관성이 나타나는 방향을 주축(major axis), 이와 수직인 방향을 보조축(minor axis)이라고 하자. 만약 주축의 범위가 6, 보조축의 범위가 2인 경우, 범위가 긴 방향과 짧은 방향의 비율을 계산하면  $2/6 = 1/3$ 이 된다. 이는 범위가 짧은 방향에 비해 긴 방향의 공간 상관성이 3배 더 강하다는 것을 뜻한다. 이처럼 기하학적 비등방성은 특정한 한 방향에서 다른 방향에 비해 길거나 짧은 범위를 갖는 경우를 의미한다.

2-차원 공간자료에 기하학적 비등방성이 존재하는 경우에 공간자료의 관측된 위치 지점에 대해 임의의 각도만큼 회전하여 주축과 보조축의 비율을 갖는 공간 상관성을 의미하는 타원(ellipse)이 그려질 수 있다. 이는 회전에 의해 기존의 좌표축이 재척도화 될 수 있음을 의미한다. 이를 이용하여 관측된 위치 지점 사이의 각도와 공간 상관성의 긴 범위와 짧은 범위의 비율에 관한 모수를 추정하여 비등방성 모형을 적용한 분석이 가능해지는 것이다. Figure 1.1(b)는 기존의 위치 지점에 관한 좌표값의 각도가 세로축(y-axis)을 기준으로  $\pi/4$ 만큼 회전하여 주축과 보조축의 비율이 1/2로 된 경우를 나타낸다.

방향에 따라 범위가 다른 기하학적 비등방성에서는 각도모수( $\psi$ : angle parameter)와 비율모수( $\lambda$ : ratio parameter)의 2가지 모수를 추정해야 한다. 이를 적용하기 위해, 각도 모수와 비율 모수를 추정하

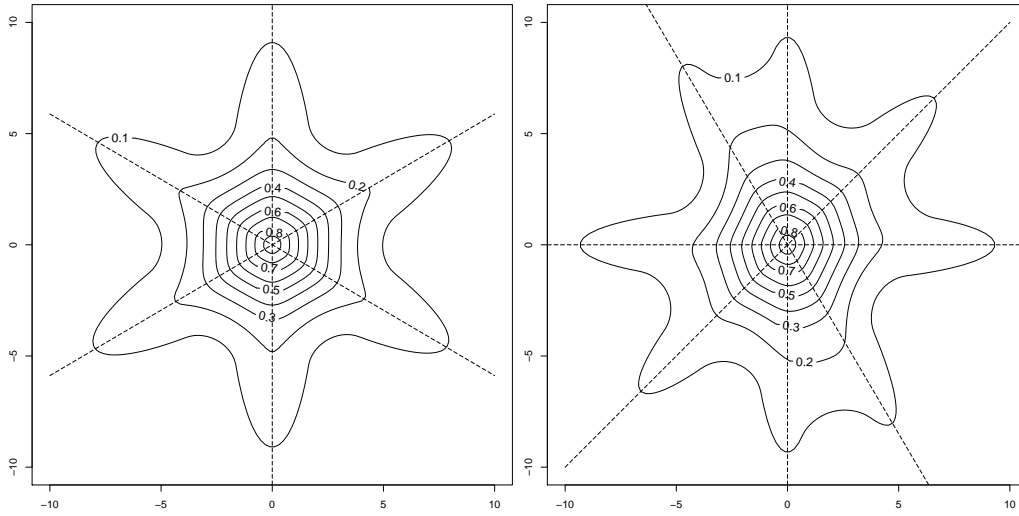


Figure 2.1. Contour plots of spatial association under extended geometric anisotropy.

여 좌표 변환을 해야 한다. 각도모수 행렬( $\mathbf{A}$ )과 비율모수 행렬( $\mathbf{R}$ )은 다음과 같이 표현된다.

$$\mathbf{A} = \begin{pmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}. \quad (2.3)$$

식 (2.3)에서의 행렬들을 이용하여 변환된 상대적 위치의 차이를 계산하면 다음과 같다.

$$\mathbf{h}^* = \mathbf{R}\mathbf{A}\mathbf{h} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}. \quad (2.4)$$

식 (2.4)으로부터 각도모수( $\psi$ )와 비율모수( $\lambda$ )를 통해 기하학적 비등방성 하에서의 공분산모형을 등방성 모형으로부터 얻을 수 있게 된다.

공간자료의 모형화 및 미관측지점의 예측을 수행하고자 할 때, 자료의 비등방성 유무를 확인하여 공간적 상관성 패턴을 보다 정확히 찾아야만 예측력을 높일 수 있다. 특히, 실제 공간자료에는 비정형화된 형태의 방향성을 고려해야 할 경우가 많으며 다양한 방향으로 상이한 패턴을 보이는 경우에는 비등방성 모형을 더욱 정밀하게 적용시켜야 할 것이다. 그러므로, Figure 2.1과 같이 여러 방향에서 다른 방향보다 공간 상관성이 길게 나타나는 실제자료의 공간 분석을 위해서는 이에 적합한 비등방성 모형을 적용해야 한다. 이를 위해, 기존의 한 방향에 관한 기하학적 비등방성 모형들의 선형결합으로 표현할 수 있는, 확장된 형태의 기하학적 비등방성 모형을 제안하고자 한다. 본 연구에서는 기하학적 비등방성 모수가 2개씩 존재하는, 특정한 두 방향으로 공간 상관성이 높게 나타나는 모형을 고려하였다. 공간자료에 여러가지 모형을 적용한 추정 결과의 비교를 위해서, 보편적인 등방성 모형 및 기하학적 비등방성 모형과 본 연구에서 제안하는 확장된 기하학적 비등방성 모형 등의 3가지 모형을 고려하여 연구를 진행하고자 한다.

### 2.2. 확장된 기하학적 비등방성 모형

수월한 설명을 위해 용어에 대한 간단한 정리를 하고자 한다. M.0는 모든 방향에서 공간 상관성이 동일하다고 가정하는 등방성 모형, M.1은 특정한 하나의 방향에 관한 기하학적 비등방성 모형, M.2는 본 연

구에서 제안한 임의의 두 방향에서의 확장된 기하학적 비등방성 모형이라고 하자. 제안한 모형(M.2)과의 비교를 위해 보편적인 등방성 모형(M.0) 및 비등방성 모형(M.1)을 고려하였다. M.0과 M.1, 그리고 M.2는 각 모형에 따라 특정 공분산모형을 가지게 되며, 이는  $\text{Cov}_0(\mathbf{h}), \text{Cov}_1(\mathbf{h}), \text{Cov}_2(\mathbf{h})$ 로 표현하자.

등방성 모형(M.0) 하에서는 모든 방향에서의 공간 상관성이 동일하므로 공분산 모형을 적합하기 위해 추정할 모수는  $\Theta = (\theta_1, \theta_2, \theta_3)^T$ 이다. 하나의 방향에 관한 비등방성 모형(M.1)에서는 공통모수에 기하학적 비등방성과 관련된 각도모수와 비율모수를 더하여 총 5가지 모수,  $\Theta = (\theta_1, \theta_2, \theta_3, \psi_a, \lambda_a)^T$ 를 추정해야 한다. M.1 하에서는 임의의 두 지점 사이의 상대적 위치와 더불어 각도 및 비율 모수에 관한 식 (2.3)를 이용하여 임의의  $i, j = 1, \dots, n$ 에 대한 거리는 다음과 같이 구할 수 있다.  $\mathbf{h}_a = (h_{a1}, h_{a2})^T$ 에 대하여

$$\|\mathbf{h}_a\| = \sqrt{\mathbf{h}_a^T \mathbf{h}_a} = \sqrt{(\mathbf{R}_a \mathbf{A}_a \mathbf{h}_a)^T (\mathbf{R}_a \mathbf{A}_a \mathbf{h}_a)} = \|\mathbf{R}_a \mathbf{A}_a \mathbf{h}_a\|. \quad (2.5)$$

여기서,

$$\mathbf{A}_a = \begin{pmatrix} \cos(\psi_a) & -\sin(\psi_a) \\ \sin(\psi_a) & \cos(\psi_a) \end{pmatrix}$$

이고  $\mathbf{R}_a = \text{diag}(1, \lambda_a)$ 이다. 위의 변환된 거리를 이용한 공분산 모형은 다음과 같이 표현한다.

$$\text{Cov}_1(\mathbf{h}) = \text{Cov}_0(\mathbf{h}_a)$$

만약 또다른 방향의 비등방성이 존재한다면, M.2는 공통모수에 각 방향에 따른 각도모수 및 비율모수와 가중치를 고려한, 8가지 모수,  $\Theta = (\theta_1, \theta_2, \theta_3, \psi_a, \lambda_a, \psi_b, \lambda_b, \omega)^T$ 를 추정하게 된다. 우선 한 방향의 비등방성 모형과 관련된 변환거리를 식 (2.5)라고 하면 또다른 방향의 비등방성 모형과 관련된 변환거리는 다음과 같다.  $\mathbf{h}_b = (h_{b1}, h_{b2})^T$ 에 대하여

$$\|\mathbf{h}_b\| = \sqrt{\mathbf{h}_b^T \mathbf{h}_b} = \sqrt{(\mathbf{R}_b \mathbf{A}_b \mathbf{h}_b)^T (\mathbf{R}_b \mathbf{A}_b \mathbf{h}_b)} = \|\mathbf{R}_b \mathbf{A}_b \mathbf{h}_b\|.$$

여기서,

$$\mathbf{A}_b = \begin{pmatrix} \cos(\psi_b) & -\sin(\psi_b) \\ \sin(\psi_b) & \cos(\psi_b) \end{pmatrix}$$

이고  $\mathbf{R}_b = \text{diag}(1, \lambda_b)$ 이다. 위의 2가지 변환된 거리를 이용한 공분산모형은 다음과 같이 표현된다.

$$\text{Cov}_2(\mathbf{h}) = \omega \text{Cov}_0(\mathbf{h}_a) + (1 - \omega) \text{Cov}_0(\mathbf{h}_b). \quad (2.6)$$

여기서,  $0 \leq \omega \leq 1$ 이다. 식 (2.6)을 통해 알 수 있듯이 각각의 비등방성 모수  $(\psi_a, \lambda_a), (\psi_b, \lambda_b)$ 를 이용해 계산된 공분산,  $\text{Cov}_0(\mathbf{h}_a)$ 과  $\text{Cov}_0(\mathbf{h}_b)$ 에 가중치를 부여하여 확장된 형태의 공분산모형인  $\text{Cov}_2(\mathbf{h})$ 를 만들 수 있다.

### 2.3. 모수 추정 방법

공간통계학에서의 대표적인 모수추정방법으로 최대우도추정법이 있다. 다변량 정규분포(multivariate normal distribution)를 따르는 공간확률과정,  $\mathbf{z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ 을 가정하여 평균함수를  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ 라고 하고 공간 상관성을 나타내는 모수벡터를  $\Theta$ , 공분산행렬을  $\boldsymbol{\Sigma}_\Theta$ 라고 하면 우도함수(likelihood function)를 기반으로  $\boldsymbol{\beta}$ 와  $\Theta$ 를 추정할 수 있으며  $\mathbf{z}$ 는 다음과 같이 나타낼 수 있다.

$$\mathbf{z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\Theta). \quad (2.7)$$

**Table 3.1.** Scenarios considered in the simulation studies.

Scenario	Truth	$\beta_1$	$\beta_2$	$\beta_3$	$\theta_1$	$\theta_2$	$\theta_3$	$\psi_a$	$\lambda_a$	$\psi_b$	$\lambda_b$	$\omega$
1	G.0	20	3	-2	0.5	2.0	4.0	.	.	.	.	.
2	G.1	20	3	-2	0.5	3.0	5.0	$\pi/8$	1/8	.	.	.
3	G.2	20	3	-2	0.5	2.0	4.0	$\pi/2$	1/10	$\pi$	1/9	0.4
4	G.2	20	3	-2	0.5	3.0	5.0	$\pi/8$	1/10	$3\pi/4$	1/8	0.5

Truth, true model;  $\beta_1, \beta_2, \beta_3$ , regression coefficients; G.0, isotropic data generated by M.0; G.1, geometrically anisotropic data generated by M.1; G.2, geometrically extended anisotropic data generated by M.2; M.0, isotropic model; M.1, geometrically anisotropic model; M.2, geometrically extended anisotropic model;  $\theta_1$ , nugget;  $\theta_2$ , partial sill;  $\theta_3$ , range;  $\psi_a, \psi_b$ , angle parameter;  $\lambda_a, \lambda_b$ , ratio parameter;  $\omega$ , weight parameter.

여기서,  $\mathbf{X}$ 는  $(n \times q)$ 의 예측인자행렬이고  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_q)^T$ 는 회귀계수 벡터이다. 따라서  $\mathbf{z}$ 의 우도함수는 다음과 같이 나타낼 수 있다.

$$L(\boldsymbol{\beta}, \Theta; \mathbf{z}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}_\Theta|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_\Theta^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right\}.$$

본 연구에서는 식 (2.7)의 우도함수 혹은 로그우도함수(log-likelihood function)를 최대화하는 모수값을 추정함에 있어서 식 (2.2)에서 소개한 구형 공분산모형을 기반으로 한 공분산행렬을 고려하였고, 주어진 공간 자료에 관하여 3가지 모형을 적용하였다.

### 3. 모의실험

본 장에서는 등방성 모형(M.0)과 기하학적 비등방성 모형(M.1), 그리고 확장된 형태의 기하학적 비등방성 모형(M.2)의 3가지 모형들을 고려하여 모의실험을 수행하였다. 모의실험을 위해 생성된 자료는 3 가지로 구분된다. G.0은 M.0에 의해 생성된 자료이고 G.1은 M.1에 의해 생성된 모의자료이며, G.2는 M.2에 의한 모의자료이다. 자료의 추세 또는 경향을 나타내는 위치정보에 관한 회귀계수,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$  또한 모의자료 생성에 고려하였다. 따라서 각 자료는 식 (2.7)을 따르는 다변량 정규분포 하에서 생성하였으며, 이에 3가지 모형(M.0, M.1, M.2)을 최대우도추정방법을 이용하여 적합하였다. 모의실험을 위한 공분산 모형은 구형모형을 이용하였으며 Table 3.1에 제시된 대로 4가지 시나리오를 고려하였다. 비등방성 모형의 적용을 위한 각도모수( $\psi$ )는 세로축(y-axis)을 기준으로 하여 시계방향으로  $(0, \pi)$ 의 구간 안에서 추정이 되도록 하였다. 또한, 장축(주축)과 단축의 비를 의미하는 비율모수( $\lambda$ )는  $(0, 1)$ 의 범위를 가지도록 하였다. 모의실험을 위해 R 프로그램(R program)을 사용하였으며 모수 추정을 위해 최적화 패키지(optim package)를 이용하였다.

#### 3.1. 모의실험 절차

모의실험의 구체적인 절차는 다음과 같다.

1. 균일분포  $U(0, 10) \times U(0, 10)$ 에서 200개(표본크기)의 지점,  $\{\mathbf{s}_i = (s_{1i}, s_{2i})^T, i = 1, \dots, 200\}$ 를 생성한다.
2. 단계 1.에서 구한 위치정보를 이용하여 각 모형 하에서 거리행렬을 계산한다.
3. 단계 2.에서 구한 거리행렬을 이용하여 구형 공분산모형을 이용한 공분산행렬을 계산한다.

4. 식 (2.6)의 다변량 정규분포 하에서 단계 3.에서 구한 공분산행렬을 갖는 자료를 각 모형 하에서 시나리오 별로 크기가 200인 자료를 각 299개 임의생성한다.
5. 단계 4.에서 생성된 모의자료에 각 모형(M.0, M.1, M.2)을 적합시킨다. 이 때 최대우도추정방법을 이용하여 모수를 추정한다.
6. 각 모형 하에서의 추정결과를 요약하고 모형들 간의 적합결과를 비교한다.

Table 3.1에 제시된 각 시나리오마다 299개의 모의자료를 생성하였으며 모의실험 결과로 평균(mean), 분산, 편차(bias), 평균제곱오차(mean squared error)를 계산하였다. 모의자료에 대한 모형의 적합도를 비교하기 위하여 로그우도함수값( $l$ )와 대표적인 모형 선택 기준(model selection criteria)인 AIC (Akaike, 1973)와 BIC (Schwarz, 1978), 그리고 AICc (Hurvich와 Tsai, 1989)를 활용하였다. 모의자료에 등방성 및 비등방성 모형을 적용한 결과를 토대로 분류율(classification rates)을 비교하였다. 여기서, AICc는 참 모형이 유한한 차원을 가지는 경우 다른 모형 선택 기준들에 비해 올바른 모형을 선택하는 경향이 강한 것으로 알려져 있다. 따라서 모수의 수 및 자료의 수에 관한 영향을 동시에 고려하므로 다양한 분야에서 대표적인 모형 선택 기준으로 활용되고 있다.

### 3.2. 모의실험 결과

모의실험을 위해 생성된 각 시나리오에 대한 모수의 추정결과는 Table 3.2와 같다. 모의실험 추정값들 중 평균값을 살펴보면 추정값이 모수의 참값 (Table 3.1 참조)과 유사한 경향을 보이고 있다. 먼저, 등방성모형(M.0)로부터 생성한 자료(G.0)에 대한 시나리오 1을 살펴보자. 이 경우 비율모수( $\lambda_a$ )가 1임을 가정하고 편차 및 평균제곱오차를 계산하였다. M.1을 적용한 결과 각도모수는 0.888로 추정되어 1에 가까운 값을 얻었다. M.2의 경우 한 방향에 관한 비율 모수( $\lambda_a$ )가 0.828으로 추정되었으며 이의 가중치( $\omega$ )가 0.744의 값을 가지는 것으로 나타났다. 이는 한 방향에 관한 가중치가 1에 가깝고, 가중치가 큰 방향에 관한 비율모수 또한 1에 가까우므로, 등방성 자료의 경향을 나타낸 추정결과가 나타난 것으로 볼 수 있다. 또한, 회귀계수와 공통모수의 변동성 관점에서 M.0이 다른 모형에 비해 평균제곱오차가 작은 경향을 보였다. 이는 분산이나 편차(bias)가 작은데 기인한다고 볼 수 있다. 다음으로 기하학적 비등방성모형(M.1)으로부터 생성한 자료인 G.1과 관련된 시나리오 2를 살펴보자. 공통모수에 관한 추정결과는 M.1을 적용한 경우의 평균제곱오차가 다른 모형에 비해 작은 경향을 보였으며, 그 중 너겟은 편차가 가장 작게 나타났다. 또한, 각도모수( $\psi_a$ )와 비율모수( $\lambda_a$ )의 추정결과에 대해 M.2에 비해 M.1의 분산과 평균제곱오차가 작거나 유사하게 나타났다. 또한 G.1을 M.2으로 추정한 가중치의 추정값을 살펴보면, 0.958으로 가중치가 1에 가까운 값을 가지는 것을 확인할 수 있다. 이는, 특정한 한 방향에서의 공간 상관성을 나타내는 공분산 모형으로도 자료의 비등방성을 충분히 설명할 수 있음을 의미하며, 결국 M.1의 경향이 잘 나타난 결과라고 할 수 있다. 임의의 두 방향에 관한 확장된 기하학적 비등방성모형(M.2)로부터 생성한 자료(G.2)에 관한 시나리오 3, 4를 살펴보자. 공통모수의 추정값들은 M.0이나 M.1을 적용한 경우보다 M.2을 적용하였을 때 평균제곱오차가 더 작았다. 또한, 각도모수와 비율모수에 관하여 M.2를 적용한 경우에 M.1을 적용했을 때보다 분산, 편차, 평균제곱오차가 모두 작았다. 결과적으로 G.0은 M.0이, G.1은 M.1이, 그리고 G.2는 M.2가 가장 작은 변동성을 가지는 것으로 나타났고 추정량의 불편성도 만족하는 것으로 나타났다.

모형별 추정 결과를 기반으로 공간 상관성을 등고선 그림(contour plot)을 이용하여 확인해 볼 수 있다. 등방성 모형의 경우에는 동심원 형태의 공간 상관성이 나타나고, 비등방성 모형의 경우에는 특정 방향에서 강한 공간 상관성을 보이는 형태가 나타난다. 확장된 기하학적 비등방성 모형에서는 두 방향에서의 비등방성이 나타난다. 각 자료에 어떤 모형을 적용하는지에 따라 참값에 관한 추정값이 나타나게 될

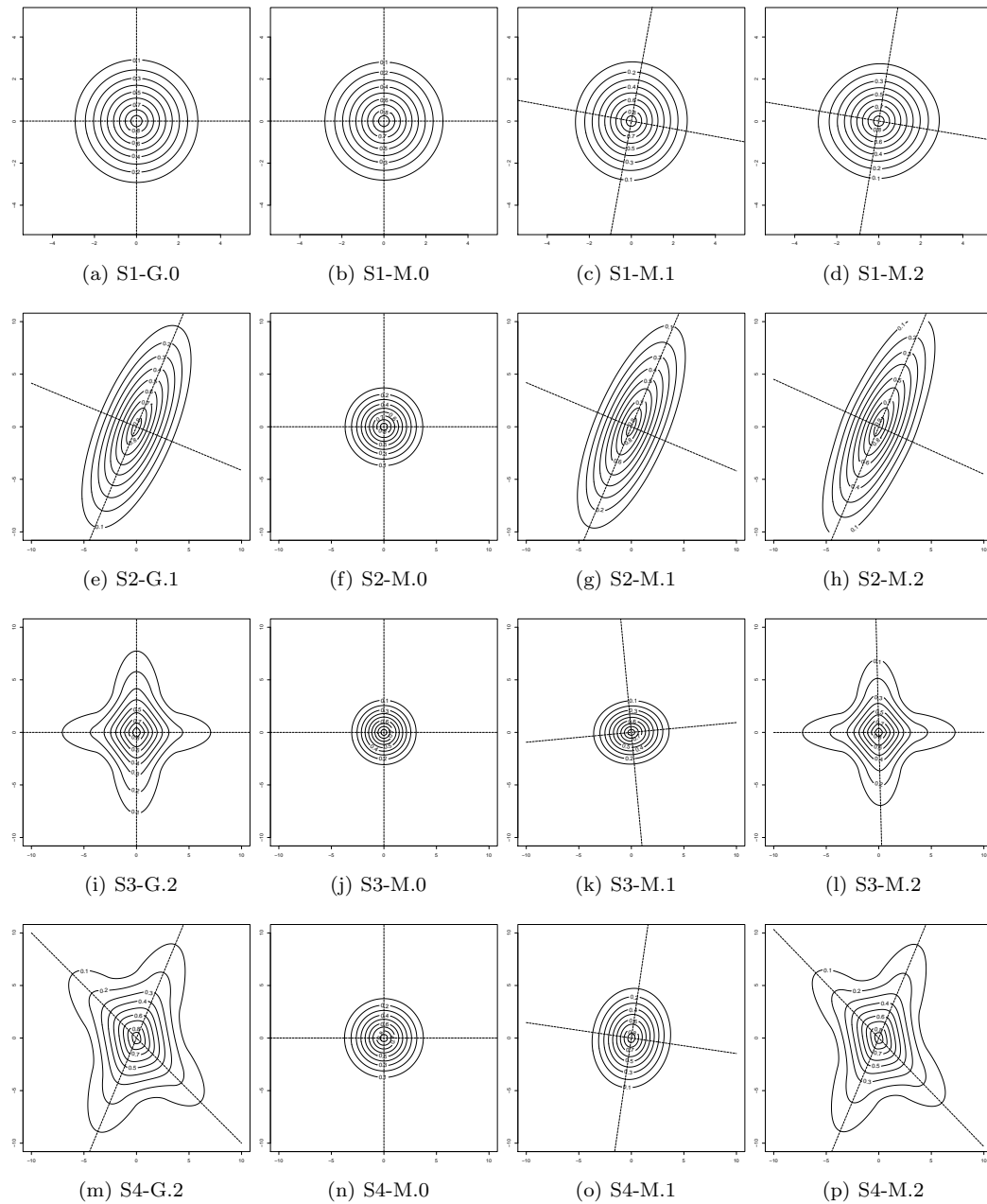


**Table 3.2.** Estimation results obtained from the true models

Scenario (Truth)	Parm	M.0				M.1				M.2			
		Mean	Var	Bias	MSE	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE
1(G.0)	$\beta_1$	19.900	0.679	-.100	0.687	19.893	0.678	-.107	0.687	19.905	0.681	-.095	0.688
	$\beta_2$	3.004	0.011	0.004	0.011	3.005	0.011	0.005	0.011	3.004	0.011	0.004	0.011
	$\beta_3$	-1.982	0.012	0.018	0.013	-1.982	0.012	0.018	0.013	-1.983	0.012	0.017	0.013
	$\theta_1$	0.505	0.012	0.005	0.012	0.500	0.012	-.001	0.012	0.495	0.012	-.005	0.012
	$\theta_2$	1.802	0.283	-.198	0.321	1.813	0.285	-.187	0.318	1.875	0.296	-.125	0.311
	$\theta_3$	3.851	0.380	-.149	0.400	3.644	0.445	-.356	0.571	3.486	0.464	-.514	0.728
	$\psi_a$					0.141	0.072			0.129	0.058		
	$\lambda_a$					0.888	0.018	-.112	0.031	0.828	0.031	-.172	0.060
	$\psi_b$									1.568	0.073		
	$\lambda_b$									0.433	0.044		
	$\omega$									0.744	0.092		
2(G.1)	$\beta_1$	19.985	1.559	-.015	1.554	20.017	1.336	0.017	1.332	20.030	1.360	0.030	1.356
	$\beta_2$	2.990	0.044	-.010	0.044	2.985	0.035	-.015	0.035	2.987	0.036	-.013	0.036
	$\beta_3$	-1.987	0.013	0.013	0.013	-1.989	0.011	0.011	0.011	-1.990	0.012	0.010	0.012
	$\theta_1$	0.537	0.017	0.037	0.019	0.518	0.008	0.018	0.009	0.508	0.009	0.008	0.009
	$\theta_2$	1.595	0.710	-1.405	2.682	2.429	0.814	-.571	1.137	2.404	0.770	-.596	1.122
	$\theta_3$	5.068	1.776	0.068	1.775	4.774	0.587	-.226	0.636	4.717	0.692	-.283	0.770
	$\psi_a$					0.398	0.003	0.005	0.003	0.396	0.003	0.004	0.003
	$\lambda_a$					0.113	0.004	-.012	0.004	0.093	0.004	-.032	0.005
	$\psi_b$									1.994	0.126		
	$\lambda_b$									0.391	0.072		
	$\omega$									0.963	0.005	-.037	0.006
3(G.2)	$\beta_1$	20.107	1.499	0.107	1.506	20.104	1.508	0.104	1.514	20.084	1.342	0.084	1.345
	$\beta_2$	2.990	0.022	-.010	0.022	2.989	0.022	-.011	0.022	2.990	0.021	-.010	0.021
	$\beta_3$	-2.007	0.015	-.007	0.015	-2.006	0.015	-.006	0.015	-2.004	0.014	-.004	0.014
	$\theta_1$	0.520	0.010	0.020	0.011	0.539	0.016	0.039	0.017	0.508	0.009	0.008	0.009
	$\theta_2$	1.202	0.242	-.798	0.877	1.316	0.301	-.684	0.767	1.668	0.354	-.332	0.463
	$\theta_3$	4.206	0.922	0.206	0.961	4.134	0.785	0.134	0.800	3.675	0.483	-.325	0.587
	$\psi_a$					1.477	0.491	-.094	0.498	1.570	0.013	-.001	0.013
	$\lambda_a$					0.705	0.082	0.605	0.447	0.083	0.011	-.017	0.011
	$\psi_b$									3.117	0.002	-.025	0.002
	$\lambda_b$									0.114	0.009	0.003	0.009
	$\omega$									0.418	0.026	0.018	0.026
4(G.2)	$\beta_1$	19.954	2.091	-.046	2.086	19.969	2.063	-.031	2.057	19.943	1.536	-.057	1.534
	$\beta_2$	3.004	0.033	0.004	0.033	3.002	0.033	0.002	0.033	3.008	0.026	0.008	0.026
	$\beta_3$	-2.003	0.017	-.003	0.017	-2.004	0.017	-.004	0.017	-2.003	0.015	-.004	0.015
	$\theta_1$	0.514	0.014	0.014	0.014	0.522	0.014	0.022	0.014	0.496	0.010	-.004	0.010
	$\theta_2$	1.853	0.562	-1.147	1.876	2.242	0.699	-.758	1.270	2.660	0.718	-.340	0.831
	$\theta_3$	5.141	1.339	0.140	1.355	5.087	0.819	0.086	0.824	4.684	0.589	-.316	0.687
	$\psi_a$					0.146	0.040	-.247	0.101	0.394	0.014	0.001	0.014
	$\lambda_a$					0.611	0.052	0.511	0.313	0.087	0.011	-.013	0.011
	$\psi_b$									2.372	0.011	0.016	0.011
	$\lambda_b$									0.112	0.013	-.013	0.013
	$\omega$									0.486	0.024	-.014	0.024

S, Scenario; Truth, true model; Parm, parameter; Var, variance; MSE, mean squared error; G.0, isotropic data generated by M.0; G.1, geometrically anisotropic data generated by M.1; G.2, geometrically extended anisotropic data generated by M.2; M.0, isotropic model; M.1, geometrically anisotropic model; M.2, geometrically extended anisotropic model.

것이며, 모수 추정값의 평균값들을 이용하여 공간 상관성의 형태를 파악할 수 있다. Figure 3.1은 각각의 모의실험 자료의 추정결과를 살펴보기 위해 시나리오 별 참값과 각 모형을 적용한 추정결과를 토대로 작성한 공간 상관성에 관한 등고선 그림이다. Figure 3.1의 (a)–(d)를 보면, Table 3.1의 G.0 및 M.0,



**Figure 3.1.** Contour plots of true and fitted spatial associations under each of the simulation scenarios. Note that S1, Scenario 1; S2, Scenario 2; S3, Scenario 3; S4, Scenario 4; G.0, isotropic data generated by M.0; G.1, geometrically anisotropic data generated by M.1; G.2, geometrically extended anisotropic data generated by M.2; M.0, isotropic model; M.1, geometrically anisotropic model; M.2, geometrically extended anisotropic model.

M.1, M.2 모두 동심원과 유사한 형태의 공간 상관성이 나타났다. Figure 3.1의 (e)–(h)는 G.1에 관한

**Table 3.3.** Empirical percentages of correctly classifying the true model

Scenario	M.0				M.1				M.2			
	log $L$	AIC	AICc	BIC	log $L$	AIC	AICc	BIC	log $L$	AIC	AICc	BIC
1(G.0)	0.110	0.906	0.926	1.000	0.298	0.064	0.060	0.000	0.592	0.030	0.013	0.000
2(G.1)	0.000	0.000	0.000	0.030	0.495	0.896	0.920	0.950	0.505	0.104	0.080	0.020
3(G.2)	0.003	0.097	0.134	0.709	0.020	0.047	0.050	0.064	0.977	0.856	0.816	0.227
3(G.2)	0.013	0.120	0.140	0.722	0.054	0.084	0.094	0.067	0.933	0.796	0.766	0.211

log  $L$ , log-likelihood; AIC, Akaike's information criterion; AICc, corrected AIC; BIC, Bayesian information criterion; G.0, isotropic data generated by M.0; G.1, geometrically anisotropic data generated by M.1; G.2, geometrically extended anisotropic data generated by M.2; M.0, isotropic model; M.1, geometrically anisotropic model; M.2, geometrically extended anisotropic model.

것으로 참값의 경우 특정한 한 방향으로 강한 공간 상관성을 나타내는 것을 확인할 수 있다. 이러한 특성을 가진 G.1에 M.0을 적용하는 경우, 각도 모수 및 비율 모수에 관한 추정이 불가능하므로 등방성 모형의 형태가 나타난다. M.1과 M.2의 추정법의 경우 참값과 유사한 패턴을 보이고 있다. Figure 3.1의 (i)-(l)은 G.2에 관한 것으로 수직을 이루는 특정한 두 방향에 관한 공간 상관성이 강하게 나타나는 모양을 가진다. 이러한 경향을 가진 G.2에 M.0을 적용하는 경우 등방성을 전제로 하기에 모든 방향에서 동일한 상관성이 나타난다. M.1의 경우 특정한 한 방향에 관한 비등방성 모수만을 추정하게 된다. 하지만 M.2를 적용한 경우 참값과 매우 유사한 공간 상관성을 보이는 것으로 나타났고 시나리오 4 (Figure 3.1의 (m)-(p))에서도 이와 유사한 결과를 확인할 수 있다.

Table 3.3에서는 모의실험에 사용한 각 시나리오별 299개 자료에 대해 각 후보 모형들의 적합성을 비교하기 위해 AIC, AICc, BIC, 로그우도값을 계산한 후, 각 시나리오 별 정분류율을 계산하였다. G.0에 M.0을 적합한 결과를 살펴보자. M.0 으로부터 모의생성한 자료 G.0의 경우(시나리오 1) M.0의 로그우도값이 M.1과 M.2에 비해 큰 자료의 비율이 전체 299개 중 11%로 나타났고 모형선택기준의 관점에서 M.0이 다른 모형에 비해 잘 적합한 자료의 비율이 80%이상으로 나타났다. 이는 등방성 자료인 G.0에 적합한 모형은 M.0라는 것을 의미한다. M.1으로부터 모의생성한 자료 G.1의 경우(시나리오 2) M.0과 M.2에 비해 M.1의 AIC, AICc, BIC가 작게 계산된 자료가 전체 299개 중 각각 92.6%, 93.6%, 93.6%로 나타났다. 이는 M.1이 G.1에 가장 적합한 모형임을 의미한다. 마지막으로 M.2으로부터 모의생성한 자료 G.2의 경우(시나리오 3, 4)를 살펴보자. 확장된 비등방성 모형인 M.2가 M.0과 M.1에 비해 로그우도값이 큰 비율이 전체 299개 자료들 중 97.7%, 93.3%로 매우 높게 나타났고 AIC와 AICc의 관점에서도 이와 유사한 결과를 보였다. 하지만 BIC는 M.2보다 단순한 모형(M.0과 M.1)을 선호하는 것으로 나타났다. 이는 BIC가 추정해야 할 모수의 수에 민감한 영향을 받게 됨에 따라 가급적 단순한 모형을 선택하려는 경향을 갖는 특성을 보여주는 결과라 판단된다.

결과적으로 참모형에 상관없이 모형이 복잡할수록 로그우도값에 의한 모형선택비율이 증가하였다. 그리고 모형선택기준의 관점에서 등방성 자료에는 등방성 모형이, 비등방성 자료에는 그에 맞는 비등방성 모형이 가장 적합한 모형으로 판명되었다. 결과적으로 공간 자료에 내재된 특성을 잘 파악할 수 있는 모형에서 모형적합이 잘 이루어졌음을 확인할 수 있다.

## 4. 실증분석

### 4.1. 자료설명 및 분석방법

실증연구에서 비등방성 모형의 적용을 위해 사용한 자료는 미세먼지(PM<sub>10</sub>)자료이다. 미세먼지는 눈에

보이지 않을 정도로 작은 물질로 한 곳에 체류하는 것이 아니라 대기 중에 오래 떠다니거나 훑날려 오는 대기오염물질이다. 본 연구에서는 미세먼지의 예측에 보다 정확성있는 분석을 위해 등방성, 비등방성, 그리고 확장된 비등방성 모형을 적용하여 분석을 실시하였다. 모형적합을 위하여 위치정보인 위도, 경도 이외에 이산화황(SO<sub>2</sub>)자료와 도시규모변수를 사용하였다. 도시규모변수는 서울특별시와 6개 광역시(인천, 광주, 울산, 대구, 대전, 부산)이면 1, 아니면 0의 값을 갖는다. 미세먼지 및 이산화황 자료는 환경부 대기환경정보(<http://www.airkorea.or.kr/>)에서 얻은 2011년 연평균 자료이다. 지역의 동질성을 확보하기 위해 제주도 및 울릉도에 있는 측정망을 제외한 총 296개 관측지점의 측정 자료를 사용하였다.

자료의 모형화를 위해 평균함수가 앞서 언급한 변수들의 선형결합으로 이루어진 범용크리깅(universal kriging)을 고려하였다. 위도, 경도로 표현되는 지점의 정보로부터 킬로미터 단위의 거리를 산출하기 위해 지오데식 변환(geodesic transformation)을 적용하였다. 자료가 정규성을 만족하도록 박스-콕스 변환(the Box-Cox transformation)을 이용하여 미세먼지(PM<sub>10</sub>) 자료에 제공된 변환을 취한 후 분석을 실시하였으며, 6365개의 임의 격자지점에 관한 예측지도를 작성하고, 예측성능을 확인하기 위해 교차검증(leave-one-out CV)을 이용하였다. 본 연구에서 제안한 모형을 적용하여 모수 추정을 수행하였으며, 구간추정을 위해서 붓스트랩 방법을 적용하였다. 여기서 붓스트랩 표본은 다변량 정규분포에서 생성하였고 생성한 붓스트랩 표본의 수는 199개이다(이에 관한 자세한 설명은 Go (2014)를 참조하기 바란다).

#### 4.2. 분석결과

Table 4.1는 최대우도추정법을 기반으로 각 모형 하에서 모수적 붓스트랩방법을 적용하여 얻은 모수 추정 결과이다. 우선, 회귀계수와 공간 상관성을 나타내는 모수의 붓스트랩 추정값이 전체적으로 원자료의 추정값(참값)과 매우 유사하다는 것을 알 수 있다. 그러나, 너겟( $\theta_1$ )의 경우 기존 자료를 이용한 경우와 비교하여 붓스트랩으로 생성된 자료를 이용한 경우에서 과대 추정된 결과를 보였다. 그러나, 부분 문턱( $\theta_2$ )의 경우에는 붓스트랩 추정결과가 과소 추정되었으며 이의 평균제곱오차와 신뢰구간이 다른 공통모수에 비해 큰 경향으로 나타났다. 각 모형을 적용한 결과를 살펴보면, M.0의 부분문턱에 관한 평균 제곱오차가 M.1과 M.2에 비해 크게 나타났고 M.1를 적용한 경우 각도모수의 추정값이 3.000으로 나타나 기존 자료를 이용한 결과인 3.013와 유사한 것으로 나타났다. 제안한 확장된 기하학적 비등방성 모형(M.2)을 적용했을 때 가중치 모수에 관한 참값이 1, 붓스트랩 자료를 이용한 추정값도 0.976으로 1에 가까운 결과를 얻었다. 이를 통해 한 방향에 관한 비등방성 모형이 주어진 자료에 적합하다고 해석할 수 있다. 그러나, 여기서 유효한 한 방향에 관한 비율모수의 구간추정 결과를 확인해보면 신뢰 상한이 1을 포함하는 것으로 나타나, 주어진 자료가 등방성 가정을 만족할 수 있음을 알 수 있다.

각 모형을 비교하기 위해 원 자료를 이용하여 계산된 PRESS(prediction residual error sum of squares; Allen, 1974) 통계량에서는 M.2가 0.336으로 가장 작은 값을 나타냈으나 M.1의 값인 0.337과 차이가 크지 않으므로 M.1의 모형을 선택하는 것이 더 적절하다고 볼 수 있다. 여기서, M.1의 구간추정 결과를 살펴보면 비율모수의 신뢰 상한이 1을 포함하는 넓은 신뢰구간을 나타내는 것을 확인할 수 있다. 또한, 모형선택기준의 관점에서 AIC, BIC, AICc에 관하여 등방성모형인 M.0을 선호하는 것으로 나타났다. 결과적으로 이 논문에서 사용한 미세먼지자료는 등방성을 가정한 공분산모형을 적용하는 것이 적절하다고 판단된다.

Figure 4.1는 앞서 언급한 미세먼지(PM<sub>10</sub>)자료의 모수 추정결과를 이용한 임의의 지점 6365개에 관한 공간 예측 지도이다. 이를 살펴보면, 서울을 둘러싸고 있는 경기지역 및 3개의 예측지도 모두 유사한 경향을 보이는 것을 확인할 수 있다. 그리고 대전, 대구, 울산, 부산, 광주 등의 광역시에서 다른 지역보다 더 낮은 미세먼지(PM<sub>10</sub>) 농도를 보이는 것을 확인할 수 있다.

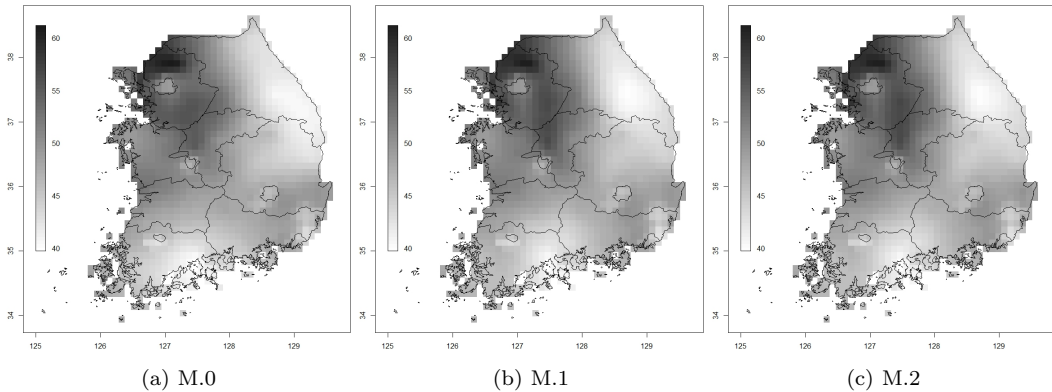
**Table 4.1.** Results of PM<sub>10</sub> data using spatial parametric bootstrap.

Model	Parm	Est.	Mean	Var	Bias	MSE	95% C.I.		AIC	BIC	AICc	PRESS					
M.0	$\beta_1$	6.442	6.437	0.016	-.005	0.016	6.168	6.691	474.9	475.4	504.4	0.359					
	$\beta_2$	-.002	-.002	< .001	< .001	< .001	-.005	0.001									
	$\beta_3$	0.001	0.001	< .001	< .001	< .001	-.001	0.003									
	$\beta_4$	94.89	94.89	< .001	< .001	< .001	94.88	94.89									
	$\beta_5$	-.227	-.235	0.007	-.008	0.007	-.377	-.076									
	$\theta_1$	0.230	0.235	0.001	0.005	0.001	0.189	0.279									
	$\theta_2$	0.148	0.121	0.003	-.027	0.004	0.013	0.231									
	$\theta_3$	170.2	170.2	< .001	-.001	< .001	170.2	170.2									
	M.1	$\beta_1$	6.446	6.456	0.015	0.010	0.015	6.190					6.644	477.4	478.2	514.3	0.337
$\beta_2$		-.002	-.002	< .001	< .001	< .001	-.004	0.001									
$\beta_3$		0.001	0.001	< .001	< .001	< .001	-.001	0.003									
$\beta_4$		97.88	97.88	< .001	< .001	< .001	97.87	97.88									
$\beta_5$		-.225	-.232	0.007	-.007	0.007	-.410	-.086									
$\theta_1$		0.230	0.234	0.001	0.004	0.001	0.188	0.281									
$\theta_2$		0.129	0.102	0.003	-.027	0.003	0.018	0.208									
$\theta_3$		110.6	110.6	< .001	-.001	< .001	110.6	110.6									
$\psi_a$		3.013	3.000	0.013	-.013	0.013	2.678	3.142									
$\lambda_a$		0.526	0.559	0.030	0.033	0.031	0.161	1.000									
M.2		$\beta_1$	6.443	6.454	0.011	0.011	0.011	6.270	6.670	483.6	484.9	531.5	0.336				
		$\beta_2$	-.002	-.002	< .001	< .001	< .001	-.004	0.001								
	$\beta_3$	0.001	0.001	< .001	< .001	< .001	< .001	0.003									
	$\beta_4$	97.88	97.88	< .001	< .001	< .001	97.87	97.88									
	$\beta_5$	-.229	-.223	0.007	0.006	0.007	-.382	-.046									
	$\theta_1$	0.229	0.233	0.001	0.004	0.001	0.190	0.283									
	$\theta_2$	0.130	0.101	0.002	-.029	0.003	0.019	0.226									
	$\theta_3$	110.6	110.6	< .001	< .001	< .001	110.6	110.6									
	$\psi_a$	3.109	3.073	0.008	-.036	0.009	2.825	3.142									
	$\lambda_a$	0.551	0.592	0.029	0.041	0.030	0.259	1.000									
	$\psi_b$	2.365	2.364	< .001	-.001	< .001	2.357	2.365									
	$\lambda_b$	0.709	0.709	< .001	< .001	< .001	0.704	0.720									
$\omega$	1.000	0.976	0.002	-.024	0.003	0.812	1.000										

Parm, Parameter; Est., parameter estimate; Mean, Bootstrap sample mean; Var, Bootstrap sample variance; Bias, Bootstrap bias; MSE, Bootstrap mean squared error; 95% C.I, Bootstrap 95% empirical confidence interval; AIC, Akaike's information criterion; AICc, corrected AIC; BIC, Bayesian information criterion; PRESS, prediction residual error sum of squares;  $\theta_1$ , nugget;  $\theta_2$ , partial sill;  $\theta_3$ , range;  $\psi_a$ ,  $\psi_b$ , angle parameter;  $\lambda_a$ ,  $\lambda_b$ , ratio parameter;  $\omega$ , weight parameter; M.0, isotropic model; M.1, geometrically anisotropic model; M.2, geometrically extended anisotropic model.

### 5. 결론

공간 자료가 특정한 하나 이상의 방향으로 긴 상관관계를 나타낼 때, 이를 고려한 분석을 함으로서 모형 적합도 및 공간예측력을 높일 수 있다고 판단하였다. 따라서, 본 논문에서는 하나 이상의 방향에서 공간 상관성을 나타내는 모수를 추정하기 위한 확장된 비등방성 모형을 제안하였다. 이를위해 공간 상관성을 나타내는 비등방성 모형 중에서도 기하학적 비등방성 모형을 중심으로 다루었다. 보편적인 기하학적 비등방성 모형의 선형결합형태로 이루어진 확장된 기하학적 비등방성 모형을 적용하여 공간 상관성을 나



**Figure 4.1.** Prediction maps based on the three covariance models. Note that M.0, isotropy model; M.1, geometrically anisotropic model; M.2, geometrically extended anisotropic model.

타내는 모수들을 추정하였다. 이의 모형 적합도를 확인해보기 위해 모의실험을 실시하였고, 구형모형의 이론적 공분산 모형을 이용한 4가지 시나리오에 따라 생성한 자료를 이용하여 3가지의 등방성 및 비등방성 모형을 적용시켜 보았다.

모의실험에 결과로 등방성을 따르는 자료에는 등방성 모형을 적용한 경우가, 한 방향으로 기하학적 비등방성 모형이 존재하는 자료에서는 하나의 기하학적 모형을 적용한 추정된 결과가, 두가지 방향에 관하여 확장된 기하학적 비등방성이 존재하는 자료에는 제안한 확장된 기하학적 비등방성 모형(M.2)을 적용하여 추정된 결과가 가장 효율적이었다고 할 수 있다.

다음으로, 실증연구를 위해 3가지 모형을 적용하여 붓스트랩 자료와 원자료의 추정결과를 비교하였다. 여기서, 원자료는 실제로 관측된 296개의 미세먼지( $PM_{10}$ ) 자료를 의미한다. 실증연구 결과로 본 연구에서 제안한 확장된 비등방성 모형의 추정 결과를 통해 특정한 하나의 방향에서의 비등방성만이 유효한 것으로 나타났다. 또한, 유효한 방향(1에 가까운 가중치에 해당하는 방향)에 관한 비율모수의 신뢰 상한이 1을 포함하였다. 이의 추정결과를 이용하여 구성된 예측지도에서는 서울을 둘러싸고 있는 경기지역에서 높은 농도를 보이는 것으로 모든 모형을 적용한 예측 지도가 유사한 경향을 보이는 것을 확인할 수 있다. 또한, 인천을 제외한 5개 광역시에서 다른 지역보다 더 낮은 미세먼지( $PM_{10}$ ) 농도를 보임을 확인하였다. 그러나, 미지의 지점에서 예측된 값은 실제값이 존재하지 않으므로 예측성능을 확인할 수 없게된다. 이를 위해서는 실제값이 존재하는 관측자료를 이용한 검증 과정이 필요하다. 따라서, 교차검증을 통해 실제값과 예측값 사이의 PRESS 통계량을 이용하여 예측성능을 확인하였다. PRESS 통계량 측면에서는 확장된 기하학적 비등방성 모형을 적용한 결과가 가장 작게 나타났다. 그러나, 이는 기존의 기하학적 비등방성 모형을 적용한 PRESS 추정값과 큰 차이가 없었고, 기하학적 비등방성 모형을 적용한 구간추정결과와 AIC, AICc, BIC 지표 등을 통해 본 연구에서 사용한 실증 자료에는 등방성 모형이 적합하다는 것을 알 수 있었다. 실증연구에서는 제안한 비등방성 모형을 적용하고, 기존의 등방성 및 기하학적 비등방성 모형을 적용한 것에 관한 모형 적합 정도를 비교할 수 있었다. 또한, 공간자료의 붓스트랩을 이용함으로써 모수에 관한 구간 추정을 실시할 수 있었다.

향후에는 공간자료에 기하학적 비등방성이 3개 이상이 존재하는 경우의 형태를 고려한 모형을 제안하고자 한다. 또한, 최대우도추정법 뿐 아니라 잔차최대우도추정법과 최소제곱법을 이용한 방법에 관해 연구할 필요가 있다고 판단된다. 또한 제안한 모형을 확장하여 구역 비등방성에도 적용해 볼 계획이다.

## References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, In: B. N. PETROV and F. CSAKI, eds, *Second International Symposium on Information Theory*, Budapest: Akademiai Kiado, 267–281.
- Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction, *Technometrics*, **16**, 125–127.
- Banerjee, S., Gelfand, A. E. and Carlin, B. P. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman & Hall/CRC.
- Budrikaite, L. and Ducinkas, K. (2005). Modeling of geometric anisotropic spatial variation, *Journal of Mathematical Modeling and Analysis*, 361–366.
- Choi, J. W. and Kim, H. W. (2006). Application of geo-statistical analysis method to detect the direction of sea surface warm flows, *Journal of the Korean Association of Geographic Information Studies*, **9**, 168–178.
- Choi, J. and Park, M. S. (2013). On the hierarchical modeling of spatial measurements from different station networks, *The Korean Journal of Applied Statistics*, **26**, 294–305.
- Choi, J., Park, M. S., Won, Y. B., Kim, H. Y. and Heo, T. Y. (2010). Analysis of total crime count data based on spatial association structure, *The Korean Journal of Applied Statistics*, **23**, 335–344.
- Choi, J. H. and Um, J. S. (2012). Comparative evaluation for seasonal CO<sub>2</sub> flows tracked by GOSAT in Northeast Asia, *Journal of Korea Spatial Information Society*, **20**, 1–13.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*, New York: John Wiley & Sons.
- Go, H. J. (2014). On the anisotropy inherent in spatial data, master's thesis, Sungshin Women's University.
- Hanig, R. P. (2003). *Spatial Data Analysis*, Cambridge: Cambridge University Press.
- Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples, *Biometrika*, **76**, 297–307.
- 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., Jin, S. and Park, M. S. (2008). Precipitation analysis based on spatial linear regression model, *The Korean Journal of Applied Statistics*, **21**, 1093–1107.
- Kang, C. W., Choi, S. B. and Yoo, S. H. (2008). A spatial bootstrap method for kriging variance, *Journal of the Korean data analysis society*, **10**, 1247–1254.
- Kim, B. (2005). Selection of variogram considering spatial correlation, master's thesis, Sungkyunkwan University.
- Kim, H. Y. (2010). A geo-statistical approach for improved prediction of traffic volume in urban area, *Journal of the Korean Association of Geographic Information Studies*, **13**, 138–147.
- Kim, K. K. and Choi, S. B. (2000). A study on the prediction of geo-statistical methods using environmental data, *Journal of the Korean Data Analysis Society*, **2**, 499–510.
- Kruminiene, I. (2006). Analysis of anisotropic variogram models for prediction of the curonian lagoon data, *journal of mathematical modelling and analysis*, **11**, 73–86.
- Kubota, T. and Tarumi, T. (2010). Simulation studies of geometric anisotropy detection methods for geo-statistical data, *Journal of the Japanese Society of Computational Statistics*, **25**, 13–24.
- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*, Springer.
- Liansheng, T., William, R. S., Wayne, A. W. and Richard, F. G. (2006). A parametric spatial bootstrap, Southern Methodist University.
- Lin, K. P., Long, Z. H. and Ou, B. (2010). The size and power of bootstrap tests for spatial dependence in a linear regression model, *Computational Economics*, **38**, 153–171.
- Loh, J. M. and Stein, M. L. (2004). Bootstrapping a spatial point process, *Statistica Sinica*, **14**, 69–101.
- Park, M., Park, C., Shin, K. I. and Yoo, C. (2010). On proper variograms of daily rainfall data, *Journal of the Korean Society of Civil Engineers*, **30**, 525–532.
- Park, M. S., Eom, J. and Heo, T. Y. (2013). The spatial correlation of mode choice behavior based on smart card transit data in Seoul, *The Korean Journal of Applied Statistics*, **26**, 623–634.

- Park, N. W. (2011). Time-series mapping and uncertainty modeling of environmental variables: a case study of PM10 concentration mapping, *Journal of the Korean earth science society*, **32**, 249–264.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461–464.
- Shin, K. I., Back, K. J. and Park, J. M. (2003). A study on the selection of variogram using spatial correlation, *The Korean Journal of Applied Statistics*, **10**, 835–844.



# 공간자료의 기하학적 비등방성 연구

고혜지<sup>a</sup> · 박만식<sup>a,b,1</sup>

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

(2014년 8월 8일 접수, 2014년 9월 22일 수정, 2014년 9월 29일 채택)

## 요약

등방성(isotropy)은 공분산 모형(covariance model)에 기반으로 공간 예측(spatial prediction)이라 불리는 크리깅(kriging)을 용이하게 수행하기 위한 주요 가정 중의 하나로 알려져 있다. 공간 과정에서 등방성이 충족되지 않는 경우에는, 보다 신뢰성 예측을 생성하기 위해 비등방성 공분산 모형(covariance model)과 관련된 모수들(각도 및 비율)을 추정해야 한다. 본 논문에서는 여러 방향의 기하학적 비등방성 모형(geometrically anisotropic covariance models)의 가중 평균으로 표현되는 확장된 형태의 기하학적 비등방성(geometrically extended anisotropic) 공분산모형을 제안한다. 연구에 관심이 되는 모수를 추정하기 위해 최대우도추정법(maximum likelihood estimation method)을 이용하였다. 제안한 모형의 성능을 평가하기 위해 등방성 공분산모형과 기하학적 비등방성 모형을 고려한 모의실험을 수행하였다. 또한 확장된 기하학적 비등방성 모형을 적용한 미세면지 농도자료 분석을 실시하였다.

주요용어: 등방성 모형, 비등방성 모형, 공간 상관성, 공분산 모형, 최대우도추정방법, 모수적 붓스트랩.

이 논문은 제1저자의 석사학위논문의 일부를 바탕으로 작성하였음.

<sup>1</sup>교신저자: (136-742) 서울특별시 성북구 보문로 34다길 2, 성신여자대학교 자연과학대학 통계학과.

E-mail: mansikpark@sungshin.ac.kr