



저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#) 

박만식교수지도

석사학위 청구논문

비등방성 공간자료의 연관성 연구

2014

성신여자대학교 대학원

통계학과

고혜지

비등방성 공간자료의 연관성 연구

박만식교수지도

이 논문을 석사학위 논문으로 제출함

2014년 5월

성신여자대학교 대학원

통계학과

고혜지

인준서

고혜지의 석사학위 논문으로 인준함.

심사위원 _____인

심사위원 _____인

심사위원 _____인

성신여자대학교 대학원

논문개요

공간통계학(spatial statistics)에서 다루는 공간 자료(spatial data)는 연구에 관심이 되는 변수에 관한 측정값과 이를 관측 또는 집계하는 위치정보(location information)로 구성되어 있다. 공간 자료를 기반으로 공간 상관성(spatial correlation)을 찾아 미지의 값을 예측하는 것이 가능하다. 자료의 공간 상관성 분석을 수행함에 있어 각 위치 지점(location point)사이의 상대적 위치(거리와 방향)를 고려하게 된다. 여기서, 공간 상관성이 거리에만 의존하는 경우에는 등방성 모형(isotropic model), 방향에 따라 공간 상관성이 달라지는 경우 비등방성 모형(anisotropic model)을 고려하게 된다.

본 논문에서는 비등방성 중에서도 기하학적 비등방성(geometric anisotropic)을 중심으로 다루어, 하나 이상의 방향에서 기하학적 비등방성이 존재하는 새로운 모형을 제안하였다. 이를 위해 모의실험 자료를 생성하여 보편적인 등방성 및 비등방성 모형을 포함하여 제안한 모형을 적용한 분석을 실시하였다. 마지막으로, 실증자료에 제안한 모형을 적합시켜 모수적 붓스트랩(parametric bootstrap)을 통해 생성된 자료들을 기반으로 구간추정을 실시하였다.

주요어: 등방성 모형, 비등방성 모형, 공간 상관성, 공분산 모형, 최대우도추정방법, 모수적 붓스트랩

목 차

I.	서론	1
II.	확장된 비등방성 모형의 제안	8
2.1	공간 상관성	8
2.1.1	정상성	8
2.1.2	공분산 모형	12
2.1.3	비등방성	17
2.2	구축한 모형에 관한 모수 추정 방법	22
III.	모의실험	26
3.1	모의실험 방법	26
3.2	모의실험 결과	31
IV.	실증연구	39
4.1	실증 자료	39
4.2	실증연구 방법	40
4.3	실증연구 결과	42
V.	결론	46
	참고 문헌	48
	Abstract	51

그림 목 차

그림 1.	등방성 및 비등방성	4
그림 2.	이론적 공분산 모형	14
그림 3.	이론적 공분산모형의 종류	15
그림 4.	비등방성 모형	18
그림 5.	등방성 및 기하학적 비등방성	19
그림 6.	여러 방향에서의 공간 상관성	21
그림 7.	등방성 자료(G.0) 결과	34
그림 8.	비등방성 자료(G.1) 결과	35
그림 9.	비등방성 자료(G.2) 결과	36
그림 10.	미세먼지 대기오염측정망과 6365개의 격자지점의 위치	40
그림 11.	미세먼지(PM ₁₀)자료 추정값 및 붓스트랩 추정결과	43
그림 12.	실증연구 예측지도	44

표 목 차

표 1.	모의실험 시나리오	27
표 2.	모의실험 추정결과	31
표 3.	오분류 평가 결과	38
표 4.	PM ₁₀ 자료 및 붓스트랩 표본자료의 최대우도추정법 에 의한 모수추정 결과	42

제 1 장

서론

공간 자료는 연구에 관심이 되는 변수에 관한 측정값과 이를 측정 또는 집계하는 위치정보로 구성되어 있다. 이러한 공간 자료에 내재된 공간 상관성(spatial correlation) 및 공간적 패턴(spatial pattern)을 찾아 미지의 값을 예측하는 것이 분석의 주요목적이 된다. 공간 자료는 자료가 되는 지점의 특성에 따라 지리통계학적 자료(geostatistical data or point-referenced data), 격자 자료(lattice data or areal data)와 공간점 패턴 자료(spatial point pattern data)로 구별된다.

지리통계학적 자료는 고정된 위치 지점(fixed location point)에서 수집되는 측정값을 의미한다. 이는 하나의 점으로 구성되어 있으며 기존의 정보를 이용하여 관측되지 않은 지점에 대한 예측을 시행하는 것이 분석의 목적이다. 대표적인 예로는 기상관측소에서 측정한 미세먼지 농도나 대기 측정망에서 얻은 대기 중의 중금속 농도 등이 있다. 격자 자료는 임의의 지역내에서 관심있는 특성에 관한 관측값으로 구성되며 관측지점은 임의의 공간지역을 대표할 수 있는 하나의 점으로 표현할 수 있다. 그러나 이는 지리통계학적 자료와 같이 고정된 한 점을 의미하는 것이 아니라 관심 대상이 될 수 있는 지역을 대표하는 지점을 의미한다. 또한, 격자자료에서의 공간 상관성은 분할된 지역(구역)들 간의 이웃정보(neighborhood information) 형태에 크게 좌우된다. 그리고 연구대상인 영역이 어떤 모양으로 분할되는지에 따라(비)정형 격

자자료로((ir)regular lattice data)로 나뉜다. 이러한 격자자료의 분석 목적은 임의의 지역(구역)의 측정값이 이웃지역들에 의해 얼마나 영향을 받는지를 공간 상관성의 형태로 추정함에 있다. 대표적 예로 한국의 각 도시 지역에서 일어난 범죄건수나 서울시의 각 구에서 발생한 암환자 수 등이 있다. 마지막으로 공간점 패턴자료는 자료의 위치가 고정되어 있지 않으며 임의적이고 확률적인 성격을 가진다. 이는 공간자료가 임의적으로 배치되어 있는지 혹은 군락(군집)을 형성하는지를 밝혀내는 데에 그 목적을 둔다. 대표적인 예로는 일본의 지진발생 진원지에 관한 공간적 패턴을 분석하거나 산림지역 내에 분포하는 여러가지 꽃의 군락 유무 등을 파악하는 것 등이 있다. 본 연구에서는 앞서 언급된 세 가지의 공간 분석 자료 중에서 지리통계학적 자료를 중심으로 공간 상관성에 관한 분석을 실시하였다.

최근 이루어진 공간 상관성에 관련된 연구는 다음과 같다. Shin 등 (2003)은 세미베리오그램에 포함된 모수 중 특히 범위와 공간 독립 검정통계량의 관계를 살펴보고, 충청남도 부여지방의 온실에서 얻어진 온실가루자료를 이용하여 세미베리오그램을 선택하는 방법을 연구하였고, 김봉조 (2005)는 서울특별시 미세먼지 농도자료를 이용하여 세미베리오그램 모형을 선택하는 방법을 연구하였다. 또한, 정지용 등 (2008)은 강수량을 측정하는 지상기상관측지점자료에 대해 공간적 상관 구조를 포함하는 선형회귀모형을 고려하여 세미베리오그램을 기반으로한 최소제곱법과 코베리오그램을 기반으로한 최대우도추정방법으로 남한지역의 공간적 특성을 적절하게 파악할 수 있는 모형들을 찾고 이 모형들을 비교하였고, 박민규 등 (2010)은 일 강수량 자료와 상관성이 높은 이차변수로 일평균풍속, 상대습도, 이슬점 온도를 함께 고려하여 적절한 세미베리오그램을 선택하기 위한 방법을 연구하였

다. 뿐만 아니라 최정순 등 (2010)은 각 지역별로 발생하는 범죄건수 자료에 관해 포아송분포, 음이항 분포, 일반화 포아송 분포등의 가산 자료 기반의 공간가산자료모형을 적합시키는 연구를 하였고, 정승환 등 (2010)은 기상청의 풍속자료의 선형회귀모형을 이용하여 공간적 특성을 파악할 수 있는 여러 가지 모형을 비교하였으며, 박만식 등 (2013)은 대중교통 통행자들이 이용하는 대중교통 수단과 출발지에 대한 공간적 상관성 여부를 실증적으로 분석하였다.

공간적으로 분포하는 자료에 관하여 공간적인 패턴을 찾는 과정에서 특정 방향에서 다른 방향보다 상관성이 강하게 나타날 때, 두 지점 사이의 거리 뿐 아니라 방향을 고려한 분석을 시행해야 한다. 만약 공간자료의 상관성이 방향에 관계없이 일정한 경우는 등방성 모형(isotropic model)을 다루게 되며, 방향에 따라 상이한 패턴을 보이는 경우에는 비등방성 모형(anisotropic model)을 적용해야 한다. 이에 관해 본 논문에서는 공간 분석의 측도가 되는 공분산 모형을 이용하여 비등방성이 존재하는 자료에 관하여 하나 이상의 방향을 고려하는 비등방성 모형을 제안하였다. 먼저, 공간통계분석에서 공간자료에 존재하는 상관성을 파악하기 위해서는 공간확률과정에서 오차항에 관한 연관성을 포함시키기 위하여 공간 모형화를 필요로 한다. 이는 임의의 두 지점의 거리가 가까울수록 높은 상관을 가지는 공간분석의 특성을 이용하여 모형화 할 수 있다.

공간 자료의 모형화를 위해서는 특정 방향으로 상관성이 강하게 나타나는 경우에는 방향성이 존재하는 지에 관한 고려가 필요하다. 만약 공간자료의 패턴이 방향에 관계없이 일정한 공분산값이 얻어질 경우 등방성 공분산모형(isotropic covariance model), 방향에 따라 상이한 공분산값이 얻어질 경우 비등방성 공분산모형(anisotropic covariance

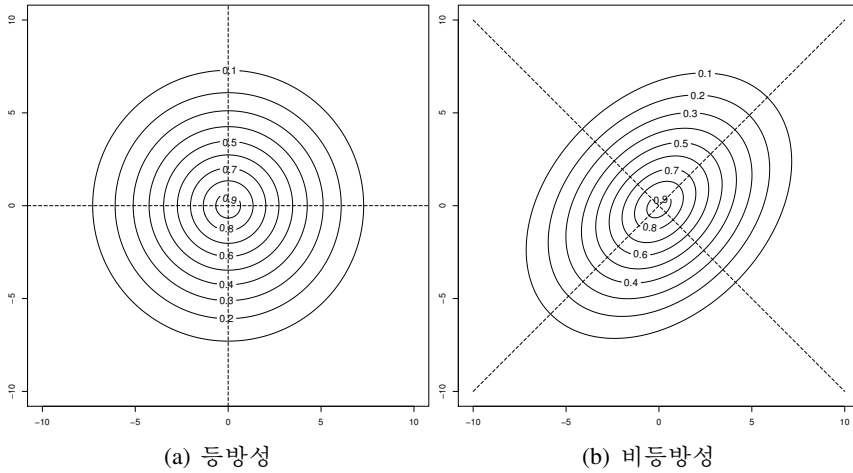


그림 1: 등방성 및 비등방성

model)을 다룬다. 즉, 등방성 모형은 그림 1(a)와 같이 공간 상관성이 모든 방향에서 같은 연관성의 강도를 나타내는 것으로, 공분산 값이 유클리드거리(Euclidean distance)에만 의존하게 되는 경우를 의미한다. 그러나 어떠한 방향에 대해 강한 공간 상관성을 나타내는 경우 방향성을 고려한 비등방성 모형을 고려해야 한다. 그림 1(b)에서는 북동쪽 및 남서쪽의 공간 상관성이 더 길게 나타나고 이와 수직인 방향인 북서쪽 및 남동쪽의 공간 상관성이 나타나는 길이와 다른 양상을 보임을 확인할 수 있다. 이와 동시에, 동일한 공분산 값을 갖게하는 두 지점 사이의 거리가 북동쪽 및 남서쪽에서 더 길다는 것을 의미한다.

비등방성 모형 적용에 관한 연구는 다음과 같다. Budrikaite *et al.* (2005)은 발틱해의 연안지역 염분자료를 이용하여 공간과정을 모형화하기 위해 기하학적 비등방성 베리오그램을 적용하였고, Kruminienė (2006)는 크로니안 석호 깊이(curonian lagoon depth)자료에 기하학적 비등방성 베리오그램과 등방성 베리오그램을 적용하여 비교하여 기하

학적 비등방성 모형을 적용한 경우에 모형적합과 예측성능이 향상되었음을 밝혔다. 또한, 최현우와 김현욱 (2006)은 방향성 반베리오그램(directional semivariogram)의 계산 및 비등방성을 이용한 해수면 난류수 유동방향을 탐지하였고, 김호용 (2010)은 공간통계분석의 교통량 예측을 위해 비등방성을 적용한 분석을 실시하였으며 예측의 정확도 향상을 위한 새로운 방안을 제시하였다. Kubota *et al.* (2010)는 기하학적 비등방성을 탐지하고 보정하는 방법을 제안하여 모의실험을 수행하였고, 최진호와 엄정섭 (2012)은 비등방성 베리오그램을 통해 이산화탄소의 유동 방향을 결정하여 계절별 변화추세를 확인하였다.

일반적인 기하학적 비등방성은 그림 1(b)와 같이 타원의 장축이 일정한 각도로 회전한 타원의 형태이다. 이처럼 보편적인 기하학적 비등방성의 경우 각도모수(ψ_A : angle parameter)와 비율모수(ψ_R : ratio parameter)를 하나씩 고려한 타원형태를 나타낸다. 본 연구에서는 장축의 각도를 달리하는 두개 이상의 타원을 겹쳐놓은 형태의 기하학적 비등방성을 제안한다. 즉, 실제 공간 자료에서 여러방향으로 공간 상관성이 강하게 나타나는 경우, 이에 적합한 모형을 적용하기 위하여 확장된 비등방성 모형을 제안하고자 한다. 모수 추정을 위해 각각의 비등방성 모수인 각도 모수와 비율 모수가 각각 2개씩 존재하는 모형에 관한 알고리즘을 구축하였으며, 여러가지 시나리오 자료를 생성하여 모의실험을 실시하였다. 또한, 제안한 모형을 적용하여 실제 자료를 이용한 공간 분석을 실시하였다. 여기서, 구간 추정을 위해 모수적 붓스트랩 방법을 이용하였다.

공간 분석에 붓스트랩을 적용한 연구들은 다음과 같다. Lahiri (2003)는 격자 자료의 블록 붓스트랩 방법을 논하였으며, Loh and Stein (2004)은 공간점 패턴 자료에 붓스트랩을 이용하여 신뢰구간을 얻는 새로

은 방법을 제안하여 블록붓스트랩과의 구간추정 성능을 비교하였다. 또한, Liansheng *et al.* (2006)은 공분산 행렬의 출레스키 분해(Cholesky decomposition)을 이용한 모수적 붓스트랩을 자료에 적용하였다. 또한, Kang 등 (2008)은 크리깅 분산의 과소추정 문제에 관하여 공간 붓스트랩 자료를 이용한 모의실험을 수행하였으며, Lin *et al.* (2010)은 선형 회귀 모형의 공간 연관성의 가설 검정을 위한 잔차 기반의 공간 붓스트랩 검정을 하였다. 마지막으로, 최지은 (2013)은 공간 선형회귀모형에서 모수의 신뢰구간을 구하기 위해 붓스트랩과 잭나이프 방법을 이용하였다. 한편, 공간 상관성에 관한 추정 결과를 기반으로 공간예측에 관한 연구를 보면, 김규곤과 최승배(2000)는 일산화탄소 자료의 평균을 이용한 일반통계분석법과 공간통계분석법의 공간분석을 실시하여 예측력을 비교하였고, 조재영과 최승배와 김규곤 (2001)은 자료들 간 서로 상관이 되어 있는 공간 자료를 이용하여 일반통계방법과 공간통계방법을 이용한 분석을 통해 각각의 예측성능을 비교하였다. 강창완 등 (2008)은 주성분분석을 이용하여 우리나라 전국의 208개의 관측소에서 측정된 대기오염 자료를 이용하여 공간예측을 하였고, 박노욱 등 (2008)은 기온 및 강수관측치 자료를 이용한 공간 분포도 작성에 수치 표고모델(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 논문에서는 공간 모수적 붓스트랩(spatial parametric bootstrap) 방법을 이용하여 제안한 확장된 형태의 기하학적 비등방성 모형을 미세먼지(PM_{10}) 자료에 적용하였고 신뢰구간을 계산하여 구간추정을 실시하였다. 또한, 이의 추정된 결과를 토대로 미지의 지점에 관한 공간예측을 통해 예측지도(predict map)를 작성하였다.

본 논문의 순서는 다음과 같다. 제 2장에서는 공간 상관성에 관한 정상성 및 비등방성등의 공간적 성질에 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 실시하였으며, 마지막으로 제 5장에서는 본 연구의 결론과 향후 연구 방향을 제시하고자 한다.

제 2 장

확장된 비등방성 모형의 제안

2.1 공간 상관성

2.1.1 정상성

특정 공간영역에서 관측한 측정값들 간에 공간 상관성이 존재한다면 공간 상관성의 구조를 모형화하여 모수를 추정하고 측정값들에 대한 유의미한 예측을 도출할 수 있다. 공간 자료의 모형화를 위해 공간적 확률과정(spatial stochastic process)을 이해해야 한다. 주어진 공간 자료에서 d -차원의 관심영역 \mathbb{D} 내에 존재하는 위치정보를 \mathbf{s} , 해당 위치 지점에 관측한 자료값은 $Z(\mathbf{s})$ 이고 식 (2.1)와 같이 표현할 수 있다.

$$\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{D} \subset \mathbb{R}^d\}. \quad (2.1)$$

여기서, 지리통계자료의 2차원 형태($d = 2$)를 고려할 때 공간 자료값은 $\{Z(\mathbf{s}_i); \mathbf{s}_i = (s_{1i}, s_{2i})^T, i = 1, 2, \dots, n\}$ 으로 표현된다. 여기서, d 차원 공간 상에 존재하는 고정된 부분집합으로 \mathbb{D} 내에 존재하는 모든 지점 $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 에 대한 확률변수를 $\mathbf{z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))^T$ 라고 표현하고, 이의 기대값과 분산이 존재한다고 가정하게 되면 \mathbf{z} 는 식 (2.2)와 같이 큰 규모변동과 작은 규모변동으로 구분하여 볼 수 있다.

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \quad (2.2)$$

식 (2.2)에서 큰 규모변동 부분인 $\boldsymbol{\mu} = (\mu(\mathbf{s}_1), \mu(\mathbf{s}_2), \dots, \mu(\mathbf{s}_n))^T$ 는 \mathbf{z} 에 관한 평균함수로 전체 지점의 전반적인 추세를 의미하며, 작은 규모변동 부분인 $\boldsymbol{\epsilon} = (\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2), \dots, \epsilon(\mathbf{s}_n))^T$ 는 각 지점(혹은 지역)내에서 내재된 측정오차를 의미한다. 따라서, 서로 상이한 위치 지점에서 얻은 측정값 사이에 공간 상관성에 관해서는 $\boldsymbol{\epsilon}$ 에 관한 모형화를 수행해야 한다. 임의의 지점 \mathbf{s}_i 에서의 $\epsilon(\mathbf{s}_i)$ 은 평균값이 0이고($E[\epsilon(\mathbf{s}_i)] = 0$), 등분산성(homoskedasticity)이 만족된다면, $\epsilon(\mathbf{s}_i)$ 의 분산은 σ^2 이 된다($\text{Var}[\epsilon(\mathbf{s}_i)] = \sigma^2$). 오차항에 관한 분석을 실시하기 위해서는 정상성 가정을 할 수 있게 되는데 시계열분석에서의 정상성 개념을 이용하면 공간통계학에서의 정상성 개념을 보다 쉽게 이해할 수 있다.

시계열분석에서의 정상성이란 시간의 흐름에 따라 확률변수의 확률적 성질이 변하지 않음을 의미한다. 즉, 임의의 시점 t 에 대한 관측값을 $Y(t)$ 라고 할 때 시계열 자료 $\{Y(t), i = (1, 2, \dots, n)\}$ 에 대하여 기대값 $E[Y(t)]$ 과 분산 $\text{Var}[Y(t)]$ 이 t 와 무관한 상수이고, $\text{Cov}(Y(t), Y(t) + h_t)$ 가 t 와 무관함과 동시에 시차 h_t 에만 의존할 경우 정상성을 갖는다고 할 수 있다. 즉, 시계열 자료가 추세를 갖지 않으면서 두 시점 사이의 공분산이 오직 시차에만 의존하는 것이 시계열 자료 분석에서의 정상성이다. 이와 관련하여 공간통계학의 정상성은 공간자료의 추세가 없으면서 공간 상관성이 임의의 두 지점 사이의 상대적 위치(거리와 방향)의 차이에만 의존하는 것을 의미한다. 즉, 임의의 지점 \mathbf{s}_i 에 대한 관측값을 $Z(\mathbf{s}_i)$ 라고 할 때, 이의 평균과 분산이 존재하는 경우에 $E(Z(\mathbf{s}_i))$ 와 $\text{Var}(Z(\mathbf{s}_i))$ 이 \mathbf{s}_i 와 무관한 상수이고, 임의의 두 지점사이의 공분산 $\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j))$ 이 상대적 위치의 차이인 $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ 에만 의존하는 경우에 정상성을 만족한다고 할 수 있다. 지리통계학적 자료의 모형화를 기반으로 미지의 지점에서 자료를 예측하기 위해 정상성 가정을

고려한다. 정상성은 크게 3가지로 강한 정상성(strict stationarity)과 약한 정상성(weak stationarity or second-order stationarity), 그리고 본질적 정상성(intrinsic stationarity)이 있다. 강한 정상성은 관측 지점에 따라 확률함수의 결합 분포(joint distribution)가 변화하지 않음을 의미하며 임의의 $\mathbf{h} \in \mathbb{R}^d$ 에 대하여 다음을 만족한다.

$$f(Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n)) = f(Z(\mathbf{s}_1 + \mathbf{h}), Z(\mathbf{s}_2 + \mathbf{h}), \dots, Z(\mathbf{s}_n + \mathbf{h})).$$

현실적으로 모집단의 분포가 알려져 있지 않으므로 주어진 자료가 강한 정상성을 따르는지를 확인하는 것은 불가능 하다. 약한 정상성은 1차 및 2차 적률(moment)이 유한한 경우로, 임의의 두 지점의 공분산이 오직 \mathbf{h} 에만 의존한다. 즉, 약한 정상성 하에서는 모든 $i, j = 1, \dots, n$ 와 $\mathbf{h} \in \mathbb{R}^d$ 에 대하여 $E[Z(\mathbf{s}_i)] = \mu$ 이며, 이의 공분산은 다음과 같이 표현할 수 있다.

$$\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_i + \mathbf{h})) = \text{Cov}(Z(\mathbf{s}_j), Z(\mathbf{s}_j + \mathbf{h})).$$

본질적 정상성은 기대값이 일정하고 임의의 두 위치 지점의 관측값의 차이에 관한 분산이 \mathbf{h} 에만 의존하는 경우로, 모든 $i, j = 1, \dots, n$ 에 대해서 다음과 같이 나타낼 수 있다.

$$\text{Var}[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)] = 2\gamma(\mathbf{s}_i - \mathbf{s}_j). \quad (2.3)$$

여기서, $\gamma(\cdot)$ 는 세미베리오그램이다. 본질적 정상성을 만족하는 경우 세미베리오그램을 정의내릴 수 있으며, 공분산이 존재하지 않을 경우가 있으므로 공분산모형의 구축을 위해서는 약한 정상성을 전제로 해

야 한다. 공분산과 세미베리오그램의 관계는 식 (2.3)을 통해 다음과 같이 정리할 수 있다.

$$\begin{aligned}
 2\gamma(\mathbf{s}_i - \mathbf{s}_j) &= E [(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2] \\
 &= E [Z^2(\mathbf{s}_i)] + E [Z^2(\mathbf{s}_j)] - 2E [Z(\mathbf{s}_i)Z(\mathbf{s}_j)] \\
 &= E [Z^2(\mathbf{s}_i)] + E [Z^2(\mathbf{s}_j)] - 2\{Cov [Z(\mathbf{s}_i), Z(\mathbf{s}_j)] + E [Z(\mathbf{s}_i)] E [Z(\mathbf{s}_j)]\} \\
 &= E [Z^2(\mathbf{s}_i)] - (E [Z(\mathbf{s}_i)])^2 + E [Z^2(\mathbf{s}_j)] - (E [Z(\mathbf{s}_j)])^2 - 2Cov [Z(\mathbf{s}_i), Z(\mathbf{s}_j)] \\
 &= 2Var [Z(\mathbf{s}_i)] - 2Cov [Z(\mathbf{s}_i), Z(\mathbf{s}_j)]
 \end{aligned}$$

위 식을 간단히 정리하면 다음과 같다.

$$\begin{aligned}
 \gamma(\mathbf{s}_i - \mathbf{s}_j) &= Var [Z(\mathbf{s}_i)] - Cov [Z(\mathbf{s}_i), Z(\mathbf{s}_j)] \\
 &= \sigma^2 - Cov (\mathbf{s}_i - \mathbf{s}_j).
 \end{aligned} \tag{2.4}$$

식 (2.4)는 두 지점의 상대적 위치의 차이에 관해 표현하면 다음과 같이 표현 된다.

$$\gamma(\mathbf{h}) = \sigma^2 - Cov(\mathbf{h}). \tag{2.5}$$

여기서, $Cov(\mathbf{h})$ 는 임의의 두 지점사이의 공분산을, 이에 대응되는 $\gamma(\mathbf{h})$ 는 세미베리오그램을 의미한다. 공분산 모형에 관한 설명은 2.1.2절에서 보다 자세히 소개된다. 식 (2.5)에 의해 상대적 위치의 차이가 작을수록 공분산이 증가하고 세미베리오그램은 감소하게 되며, 상대적 위치의 차이가 커질수록 공분산은 감소하고 세미베리오그램은 증가하게 된다는 것을 알 수 있다. 공간 자료를 이용하여 각 상대적 위치에

관한 경험적 공분산 모형을 계산하고, 이를 바탕으로 이론적 공분산모형을 구하게 되면 임의의 분리거리에 대한 공분산값을 계산할 수 있게 된다. 즉, 상대적 위치가 주어지게 되면 해당 공분산모형을 알 수 있으며 이미 알고 있는 표본의 공분산을 계산할 수 있게된다. 식 (2.5)는 지구 통계학적 자료 분석에 기본 가정이 되며, 미지의 관측지점에 관한 예측을 위한 크리깅에서도 매우 중요한 관계식이다.

2.1.2 공분산 모형

통계학에서의 공분산은 확률변수 간의 상관성 정도를 나타내는 값을 의미한다. 2개의 변수를 고려할 때 하나의 값이 상승하는 추세에 있는 경우에, 다른 값도 상승한다면 공분산의 값은 양수가 되며 반대로 다른 값이 하향하는 경우 공분산의 값은 음수가 된다. 공간 분석에서는 일정한 거리에 있는 자료들의 유사성을 나타내는 척도로 공분산 모형을 이용한다. 즉, 공분산 모형은 공간자료의 상관성 정도를 나타내는 척도로 지리통계학적 자료를 분석하는 과정 중 하나이다. 서로 다른 위치에서 관찰된 자료들이 거리에 영향을 받는다면, 공간 자료간에 위치가 가까울수록 높은 유사성을 가지게 되며 공분산 값이 크게 나타나며, 멀어질수록 낮은 상관을 보이게 되며 공분산 값은 작아진다. 공분산모형은 경험적 공분산모형(empirical covariance model)과 이론적 공분산모형(theoretical covariance model)으로 구분하여 살펴볼 수 있다. 먼저, 경험적 공분산 모형은 관측된 공간자료값을 이용하여 다음과 같이 계산된다.

$$\text{Cov}(\mathbf{h}) = \text{Cov}(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{N_h} \sum_{N(\mathbf{h})} \{Z(\mathbf{s}_i) - E[Z(\mathbf{s}_i)]\} \{Z(\mathbf{s}_j) - E[Z(\mathbf{s}_j)]\}.$$

여기서, $N(\mathbf{h})$ 는 상대적 위치의 차이인 $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ 를 갖는 전체의 쌍의 집합, N_h 는 $N(\mathbf{h})$ 에 속하는 쌍의 수를 의미한다.

즉, 임의의 두 지점에 해당하는 관측값 간의 공분산을 구할 수 있게 된다. 임의의 두 지점 $\mathbf{s}_i, \mathbf{s}_j \in \mathbb{D}$ 에 해당하는 자료값에 대한 경험적 공분산 모형은 다음과 같이 표현할 수 있다.

$$\text{Cov}(\mathbf{s}_i, \mathbf{s}_j) \equiv \text{Cov}\{Z(\mathbf{s}_i), Z(\mathbf{s}_j)\} \equiv \text{Cov}\{\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_j)\}.$$

공간 상관성을 의미하는 공분산 모형은 다음과 같은 대칭적인 성질을 가진다.

$$\text{Cov}(\mathbf{s}_i, \mathbf{s}_j) = \text{Cov}(\mathbf{s}_j, \mathbf{s}_i), \quad \forall \mathbf{s}_i.$$

또한, 양반정치(nonnegative definite)를 만족하며 다음과 같이 표현할 수 있다.

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(\mathbf{s}_i, \mathbf{s}_j) \geq 0.$$

여기서, a_i, a_j 는 $\sum_i a_i = 0$ 를 만족하는 모든 계열 $\{a_i : i = 1, \dots, n\}$ 을 의미한다.

공간 분석의 모형화를 위해 기본적으로 너겟, 부분문턱, 범위의 3가지 모수를 추정해야 한다. 이론적으로 거리가 0이면 공분산 모형의 정의에 의한 계산값은 분산이 되지만, 실제 자료값을 이용하여 계산된 경험적 공분산 모형을 이용하여 이론적 공분산 모형을 찾아내는 경우에 거리가 0인 곳에서 공분산 모형에 분산내에 상수값을 가지는 너겟(θ_1)이 존재하게 된다. 너겟은 짧은 거리에서 자료에 내재되어 측

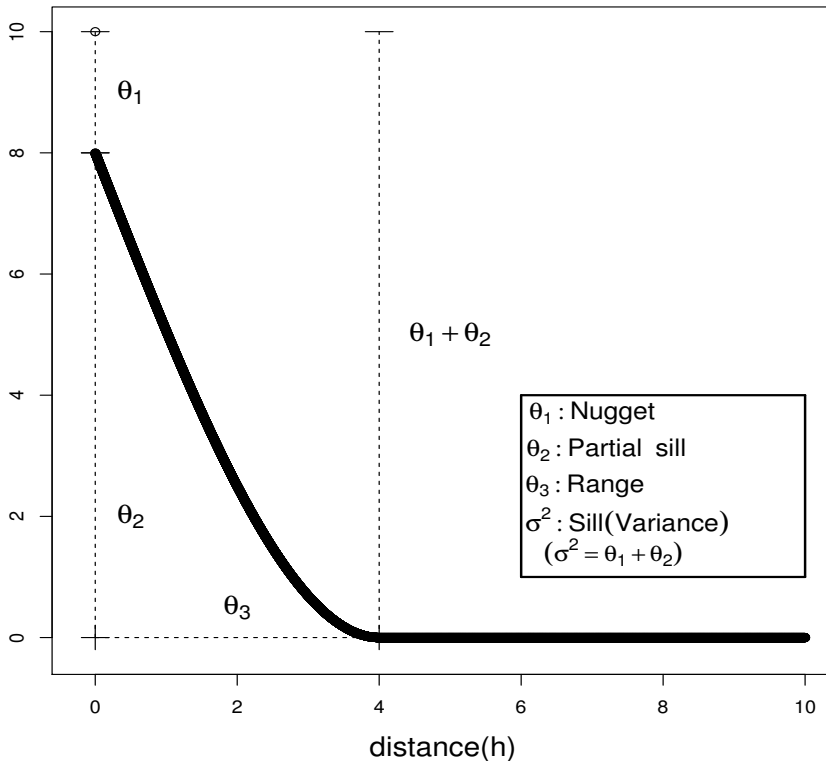


그림 2: 이론적 공분산 모형

정오차에 기인하는 변동성을 뜻한다. 실제 공간자료의 분석에서 넓게 분포하는 적은 양의 자료를 사용할 때 가까운 거리에서의 자료특성을 찾으려는 한계가 있어 불확실성이 커지게 된다. 또는 주어진 공간적 자료에 지구통계학적 기법을 적용하는 경우 거리가 작으나 면의 불연속적인 성질로 너겟이 발생하는 경우도 존재한다. 임의의 두 관측지점의 거리가 멀어질수록 자료들 사이에 상관성이 작아져 상관성이 0에 가까워지는 분산의 한 부분을 부분문턱(θ_2)이라고 한다. 여기서, 각 위치에 해당하는 관측값 쌍의 상관관계를 띄는 최대 거리를 범위(θ_3)라고 한다. 문턱은 자료의 분산(σ^2)을 나타내며, 너겟값과 부분 문턱의 합으로

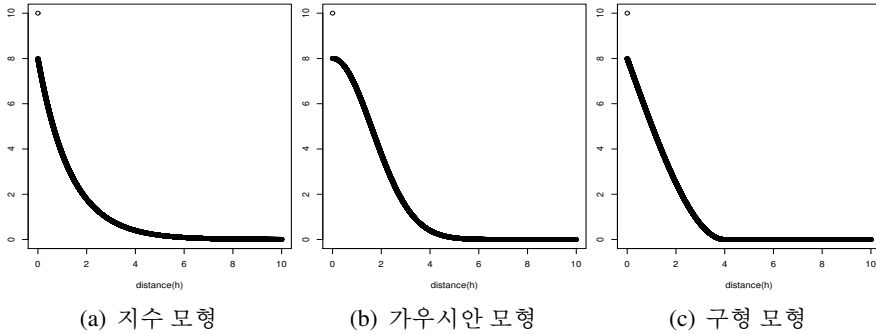


그림 3: 이론적 공분산모형의 종류

계산된다. 그림 2을 보면, 너겟이 2이며 부분 문턱이 8이고 범위가 4, 분산이 10인 이론적 공분산 모형을 나타낸다.

공간분석에서 사용되는 이론적 공분산 모형에는 지수모형(exponential model), 가우시안모형(gaussian model)과 구형모형(spherical model) 등이 있다. 그림3을 보면, 세가지 공분산 모형 모두 임의의 두 지점 사이의 거리($\|\mathbf{h}\|$)가 증가할수록 공분산 값은 작아지는 모습을 보인다. 그러나, 각각의 모형에서 $\|\mathbf{h}\| = 0$ 에 근접한 부분에 곡선형태가 다르다는 것을 확인할 수 있다. 이는 공간자료 분석에서 적절한 이론적 공분산 모형을 결정하는데에 중요한 부분이다. 가우시안 모형의 경우, 작은 거리 구간에서 역 S자 모양을 가지고 있는 것이 특징이다.

- 지수 모형의 수식은 다음과 같이 나타낼 수 있다.

$$\text{Cov}(\mathbf{h}) = \theta_1 \times \mathbf{I}(\|\mathbf{h}\| = 0) + \theta_2 \left[\exp\left(-\frac{\|\mathbf{h}\|}{\theta_3}\right) \right]$$

- 가우시안 모형의 수식은 다음과 같다.

$$\text{Cov}(\mathbf{h}) = \theta_1 \times \mathbf{I}(\|\mathbf{h}\| = 0) + \theta_2 \left[\exp\left(-\frac{\|\mathbf{h}\|}{\theta_3}\right)^2 \right]$$

- 구형 모형의 수식은 다음과 같다.

$$\text{Cov}(\mathbf{h}) = \begin{cases} \theta_1 + \theta_2, & \|\mathbf{h}\| = 0 \\ \theta_2 \left[1 - \frac{3}{2} \left(\frac{\|\mathbf{h}\|}{\theta_3} \right) + \frac{1}{2} \left(\frac{\|\mathbf{h}\|}{\theta_3} \right)^3 \right], & \|\mathbf{h}\| \leq \theta_3 \\ 0, & \|\mathbf{h}\| > \theta_3 \end{cases}$$

위의 모형식에서 나타나는 $\mathbf{I}(\cdot)$ 은 지시함수(indicator function), θ_1 은 너겟, θ_2 는 부분 문턱, θ_3 은 범위를 의미한다. 지구통계학적 자료값을 이용하여 경험적 공분산 모형을 계산하고 이를 바탕으로 이론적 공분산 모형을 선정한다. 일반적으로 너겟, 부분 문턱, 범위등 공간 상관성을 나타내는 모수를 추정하기 위해 최대우도추정법(maximum likelihood estimation method)이나 최소제곱추정법(minimized square estimation method)을 사용한다. 추정된 값은 공간적 모형이 어떻게 선정되느냐에 차이를 보이게 되므로 공분산 모형의 선정은 공간분석을 위한 중요한 과정 중 하나이다. 등방성 공분산 모형은 임의의 두 지점 간의 유클리드거리(Euclidean distance)인 $\|\mathbf{h}\| = \sqrt{h_1^2 + h_2^2 + \dots + h_d^2}$ 만을 공간적 변인으로 고려한다. 그러나, 등방성을 따르지 않은 경우, 방향에 관한 연관성을 고려해야 할 것이며 이에 관한 모수를 추정할 필요가 있다.

2.1.3 비등방성

공간통계분석에서 임의의 공간 과정들 간의 공간 상관성이 두 위치 지점 사이의 거리에만 의존하는 경우에 등방성을 가정하게 된다. 등방성 모형은 모든 방향에서 동일한 공간 상관성을 가지고 있는것을 의미한다. 즉, 등방성을 가정하면 공분산 모형은 각 위치 사이의 유클리드 거리에만 의존하게 되는 것이며 임의의 $\mathbf{h} \in \mathbb{R}$ 에 대하여 다음과 같이 나타낼 수 있다.

$$\text{Cov}(\mathbf{h}) = \text{Cov}(\|\mathbf{h}\|).$$

그러나 실제 공간자료분석을 수행할 때 공간 상관성에 영향을 줄 수 있는 공간적 변인으로 방향성을 고려해야 할 필요가 있는 경우가 많다. 그러므로, 공간자료 분석을 실시함에 있어 비등방성 모형에 관해 생각해보아야 한다. 비등방성 모형이란 공간자료의 패턴이 방향에 따라 달라지는 것을 뜻한다.

비등방성 모형은 크게 기하학적 비등방성과 구역 비등방성의 2 가지 유형으로 나누어 볼 수 있다. 기하학적 비등방성은 방향에 따라 범위가 달라지는 것이며 구역 비등방성은 범위 뿐 아니라 공간 모형 및 부분문턱도 변할 수 있는 다양한 비등방성을 포함한다. 본 연구에서는 기하학적 비등방성 모형을 중심으로 방향에 따라 공간적 상관거리가 달라질 수 있는 것을 고려하여 분석을 진행하였다. 즉, 임의의 어떤 방향의 공간적 연관성이 다른 방향에 비해 더욱 지속적으로 존재하는 경우의 비등방성 모수를 추정하여 공간적 상관 패턴을 찾고자 하였다. 기하학적 비등방성 모형은 공간 상관성이 강한 방향과 직각인 방향에서 같은 문턱값을 갖지만 범위가 다른 모형을 의미한다. 가장 길게

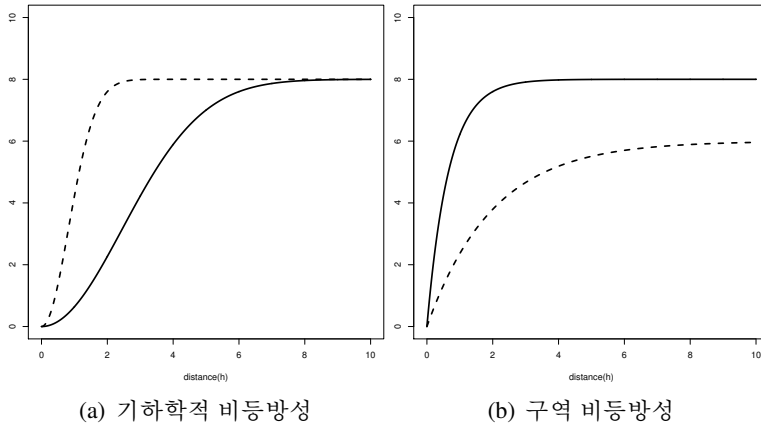


그림 4: 비등방성 모형

공간 상관성이 보이는 방향을 주축, 이와 수직인 방향을 보조축이라고 한다. 그림 4(a)에서 실선은 주축에 관한 세미베리오그램이며 점선은 보조축에 관한 세미베리오그램 값을 의미한다. 여기서, 주축의 범위는 6, 보조축의 범위는 2이므로 범위가 긴 방향과 짧은 방향의 비율은 $2/6=1/3$ 이 된다. 즉, 범위가 긴 방향에 비해 짧은 방향의 상관성이 3 배 더 길다는 것을 뜻한다. 이처럼 기하학적 비등방성은 특정한 한 방향에서 다른 방향에 비해 길거나 짧은 범위를 갖는 경우를 의미한다. 하지만 그림 4(b)와 같이 방향에 따라 범위 뿐 아니라 부분문턱이 변할 수 있는 구역비등방성 모형을 고려해볼 수 있다. 이는 비등방성의 일반적인 형태라고 볼 수 있으며 기하학적 비등방성 및 등방성 모형 이외의 모든 비등방성 모형을 포함한다.

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 좌표 축이 재척도화 될 수 있다는 것이다. 이

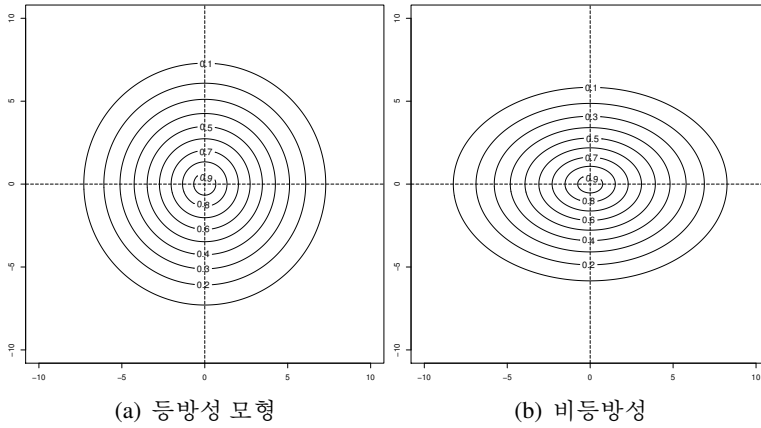


그림 5: 등방성 및 기하학적 비등방성

를 이용하여 관측된 위치 지점 사이의 각도와 공간 상관성의 긴 범위와 짧은 범위의 비율에 관한 모수를 추정하여 비등방성 모형을 적용한 분석이 가능해지는 것이다. 그림 5(b)는 기존의 위치 지점에 관한 좌표계의 각도가 y 축을 기준으로 $\pi/2$ 의 각도로 회전하여 주축과 보조축의 비율이 1/2인 경우를 나타낸다. 즉, 각도 $\pi/2$ 의 방향인 동-서쪽의 공간 상관성이 남-북쪽의 공간적 상관성에 비해 2배 긴 상관성을 나타낸다는 것을 의미한다. 방향에 따라 범위가 다른 기하학적 비등방성에서는 각도 모수(ψ_A)와 비율 모수(ψ_R)의 2가지의 모수를 추정해야 한다. 즉, 각도 모수와 비율 모수를 추정하여 좌표 변환을 해야 한다. 각도 모수에 관한 행렬은 \mathbf{A} 로 표현하면 식 (2.6)과 같이 나타낼 수 있다.

$$\mathbf{A} = \begin{pmatrix} \cos(\psi_A) & \sin(\psi_A) \\ -\sin(\psi_A) & \cos(\psi_A) \end{pmatrix}. \quad (2.6)$$

다음으로, 비율 모수에 관한 행렬은 \mathbf{R} 로 표현하며, 식 (2.7)과 같이 나

타낼 수 있다.

$$\mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & \Psi_R \end{pmatrix}. \quad (2.7)$$

지리통계학적자료의 2차원 형태인 위도, 경도를 고려하는 경우, 임의의 지점에 관한 자료값은 다음과 같다.

$$\{Z(\mathbf{s}_i); \mathbf{s}_i = (s_{1i}, s_{2i})^T, i = 1, 2, \dots, n\}$$

위에서 임의의 두 지점 $\mathbf{s}_i, \mathbf{s}_j$ 에 해당하는 좌표값을 $(s_{1i}, s_{2i}), (s_{1j}, s_{2j})$, 각각의 위치 차이를 $\mathbf{h}_1 = s_{1i} - s_{1j}, \mathbf{h}_2 = s_{2i} - s_{2j}$ 이라 하고, 이를 벡터 형태로 표현하면 $\mathbf{h} = (h_1, h_2)^T$ 이다. 2차원 공간 상에 존재하는 등방성 가정하에서의 임의의 두 지점 사이의 유클리드 거리는 $\|\mathbf{h}\| = \sqrt{h_1^2 + h_2^2}$ 이다. 각도 모수 행렬(\mathbf{A})과 비율 모수 행렬(\mathbf{R})을 이용하여 비등방성 모형 하에서 변환한 위치 차이를 계산하면 다음과 같다.

$$\begin{aligned} \mathbf{h}^* &= \mathbf{R}\mathbf{A}\mathbf{h} = \begin{pmatrix} 1 & 0 \\ 0 & \Psi_R \end{pmatrix} \begin{pmatrix} \cos(\Psi_A) & -\sin(\Psi_A) \\ \sin(\Psi_A) & \cos(\Psi_A) \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\Psi_A) & -\sin(\Psi_A) \\ \Psi_R \sin(\Psi_A) & \Psi_R \cos(\Psi_A) \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\Psi_A)h_1 - \sin(\Psi_A)h_2 & \Psi_R \sin(\Psi_A)h_1 + \Psi_R \cos(\Psi_A)h_2 \end{pmatrix}^T \end{aligned}$$

공간적 거리를 의미하는 상대적 위치의 차이인 \mathbf{h} 에 각도 모수(Ψ_A)와 비율 모수(Ψ_R)를 추정된 값을 이용한 행렬로 좌표를 변환할 수 있게 된다. 이를 이용하여 기하학적 비등방성 하에서의 경험적 공분산 모형을 얻을 수 있으며, 나아가 공간예측분석을 수행 하는데 필요한 이론적

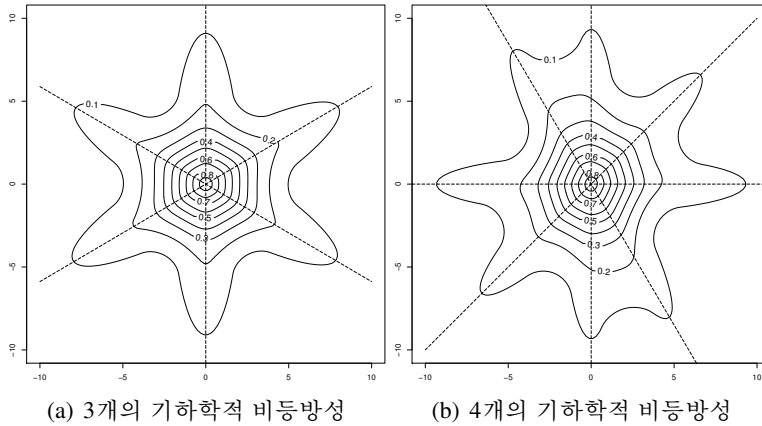


그림 6: 여러 방향에서의 공간 상관성

공분산 모형을 구축할 수 있다.

이와같이 주어진 공간자료를 이용하여 공간 상관성을 띄는 패턴을 찾고 미지의 지점의 관측값을 예측하고자 할 때, 자료의 비등방성 유무를 확인하여 공간적 연관성 패턴을 보다 정확히 찾아 예측력을 높일 필요가 있다. 특히, 실제 공간 자료에는 비정형화된 형태의 방향성을 고려해야 할 경우가 많으며 여러 방향으로 다양한 상관적 패턴을 보이는 경우에는 비등방성 모형을 더욱 정밀하게 적용시켜야 할 것이다. 그러므로, 그림 6과 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 모형을 선형 결합한 형태를 가진 확장된 기하학적 비등방성 모형을 제안하고자 한다. 본 연구에서는 하나 이상의 여러 방향을 고려하는 것을 단순화하여 기하학적 비등방성 모수가 2개씩 존재하는 특정한 두가지 방향으로 공간 상관성이 높게 나타나는 모형을 구축하여 모수 추정을 실시하였다. 공간자료에 여러가지 모형을 적용한 추정 결과의

비교를 위해서, 보편적인 등방성 모형 및 기하학적 비등방성 모형과 본 연구에서 제안한 두 가지 방향에 관한 확장된 기하학적 비등방성 모형의 3가지를 고려하여 연구를 진행하고자 한다.

2.2 구축한 모형에 관한 모수 추정 방법

2.1절에서는 공간자료의 분석에서 공간적 연관구조를 파악하기 위해 공분산 모형과 공간 상관성에 관한 이론을 살펴 보았다. 공간통계학에서 공간 상관성의 구성요소인 모수를 추정할 때 사용되는 일반적인 방법으로 최대우도추정법이 있다. 최대우도추정법이란 확률질량함수(probability mass function) 혹은 확률밀도함수(probability density function)의 결합으로 이루어진 우도함수(likelihood function ; L)와 우도함수에 로그를 취한 형태인 로그우도함수(log-likelihood function ; l)을 이용한 모수추정법이다. 다변량 정규 분포(multivariate normal distribution)를 따르는 공간확률과정 $\mathbf{z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))^T$ 을 가정하여 평균 함수 $\mu(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta}$ 라고 하고 모수 벡터를 Θ , 공분산 행렬을 $\text{Cov}(\mathbf{h}; \Theta) = \boldsymbol{\Sigma}_\Theta$ 라고 하면 우도함수를 기반으로 $\boldsymbol{\beta}$ 와 Θ 를 추정할 수 있으며 \mathbf{z} 는 다음과 같이 나타낼 수 있다.

$$\mathbf{z} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\Theta)$$

여기서, \mathbf{X} 는 $(n \times q)$ 의 예측인자 행렬이며, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_q)^T$ 는 예측인자에 관한 계수 벡터, $\boldsymbol{\Sigma}_\Theta$ 는 공분산 행렬을 의미한다. 결과적으로 우

도함수는 다음과 같이 나타낼 수 있다.

$$L(\Theta; \mathbf{z}) = (2\pi)^{-n/2} \|\Sigma_{\Theta}\|^{-1/2} \exp \left\{ -1/2 \sum (\mathbf{z} - X\beta)^T \Sigma_{\Theta}^{-1} (\mathbf{z} - X\beta) \right\}$$

이에 관한 로그우도함수는 식 (2.8)과 같이 표현할 수 있다.

$$l(\Theta; \mathbf{z}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_{\Theta}|) - \frac{1}{2} (\mathbf{z} - X\beta)^T \Sigma_{\Theta}^{-1} (\mathbf{z} - X\beta) \quad (2.8)$$

본 연구에서는 식 (2.8)를 최대화하는 모수값을 추정함에 있어서 구형 모형을 기반으로 한 분산-공분산 행렬을 고려하였고, 주어진 공간 자료에 관하여 3가지 모형을 적용하였다.

먼저, M.0는 모든 방향에서 공간 상관성이 동일하다는 전제하의 등방성 모형이고 M.1은 특정한 하나의 방향에 관한 기하학적 비등방성 모형을 의미한다. M.2는 본 연구에서 제안한 임의의 두 방향에서의 확장된 기하학적 비등방성 모형을 뜻한다. 제안한 모형(M.2)과의 비교를 위해 보편적인 등방성 모형(M.0) 및 비등방성 모형(M.1)을 고려하였다. M.0, M.1, M.2는 각 모형에 따라 특정 공분산 모형을 가지게 되며, 이는 $Cov_0(\mathbf{h}; \Theta_0), Cov_1(\mathbf{h}; \Theta_1), Cov_2(\mathbf{h}; \Theta_2)$ 로 표현하자. 3가지 모형(M.0, M.1, M.2)에서 공통적으로 추정해야 하는 3가지 모수인 너겟, 부분문턱, 범위를 공통 모수($\theta_1, \theta_2, \theta_3$)라고 하자. 나아가, 비등방성 모형에서의 각도 모수 및 비율 모수와 가중치 등은 비등방성 모수(Ψ_A, Ψ_R, ω)라고 하자.

등방성 모형(M.0) 하에서 모든 방향에서의 공간 상관성이 동일하므로 이론적 공분산 모형을 선정하기 위해 추정할 모수는 공통 모수이다. 그러므로 M.0 하에서 2차원 공간의 위치정보 $\mathbf{s}_i = (s_{1i}, s_{2i})^T$ 의 유

클리드 거리를 이용한 거리행렬(distance matrix)을 임의의 모든 $i, j = 1, \dots, n$ 에 대하여 다음과 같이 구하였다.

$$\|\mathbf{h}\| = \sqrt{(s_{1i} - s_{1j})^2 + (s_{2i} - s_{2j})^2} = \sqrt{\mathbf{h}^T \mathbf{h}}$$

여기서 h_1 와 h_2 는 기존의 좌표에 관한 임의의 두 지점 사이의 위치적 차이를 의미한다. 하나의 방향에 관한 비등방성 모형(M.1)에서는 공통 모수에 기하학적 비등방성 모형에 관한 모수인 각도모수와 비율모수를 더하여 총 5가지의 모수($\theta_1, \theta_2, \theta_3, \psi_A, \psi_R$)를 추정한다. M.1하에서는 임의의 두 지점 사이의 상대적 위치와 더불어 각도 및 비율 모수에 관한 행렬 식(2.7)과 식(2.8)을 이용하여 임의의 모든 $i, j = 1, \dots, n$ 에 대해 거리행렬은 다음과 같이 구하였다.

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

위의 변환된 거리를 이용한 공분산 모형은 다음과 같이 표현한다.

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

마지막으로, 두가지 방향에 관한 비등방성 모형(M.2)은 공통 모수에 두 방향에 관하여 각도 및 비율모수를 각각 2개씩 고려함과 동시에 이에 관한 가중치까지 8가지의 모수($\theta_1, \theta_2, \theta_3, \psi_{A1}, \psi_{R1}, \psi_{A2}, \psi_{R2}, \omega_1$)를 추정하였다. 우선 한 방향의 비등방성 모형과 관련된 변환거리를 $\mathbf{h}^{(1)} = (h_1^{(1)}, h_2^{(1)})$ 라고 하면 다음과 같이 나타낼 수 있다.

$$\|\mathbf{h}^{(1)}\| = \sqrt{\mathbf{h}^{(1)T} \mathbf{h}^{(1)}} = \sqrt{(\mathbf{R}_1 \mathbf{A}_1 \mathbf{h})^T (\mathbf{R}_1 \mathbf{A}_1 \mathbf{h})} = \|\mathbf{R}_1 \mathbf{A}_1 \mathbf{h}\|$$

여기서, $\mathbf{A}_1 = \begin{pmatrix} \cos(\Psi_{A1}) & \sin(\Psi_{A1}) \\ -\sin(\Psi_{A1}) & \cos(\Psi_{A1}) \end{pmatrix}$ 이고 $\mathbf{R}_1 = \text{diag}(1, \Psi_{R1})$ 이다.

또다른 방향의 비등방성 모형과 관련된 변환거리 $\mathbf{h}^{(2)} = (h_1^{(2)}, h_2^{(2)})$ 는 다음과 같다.

$$\|\mathbf{h}^{(2)}\| = \sqrt{\mathbf{h}^{(2)}\mathbf{h}^{(2)}} = \sqrt{(\mathbf{R}_2\mathbf{A}_2\mathbf{h})^T(\mathbf{R}_2\mathbf{A}_2\mathbf{h})} = \|\mathbf{R}_2\mathbf{A}_2\mathbf{h}\|$$

여기서, $\mathbf{A}_2 = \begin{pmatrix} \cos(\Psi_{A2}) & \sin(\Psi_{A2}) \\ -\sin(\Psi_{A2}) & \cos(\Psi_{A2}) \end{pmatrix}$ 이고 $\mathbf{R}_2 = \text{diag}(1, \Psi_{R2})$ 이다.

위의 2가지 변환된 거리를 이용한 공분산 모형은 다음과 같이 표현된다.

$$\text{Cov}_2(\mathbf{h}) = \omega_1 \text{Cov}_1(\mathbf{h}^{(1)}) + (1 - \omega_1) \text{Cov}_1(\mathbf{h}^{(2)}) \quad (2.9)$$

식 (2.10)을 통해 알 수 있듯이 각각의 비등방성 모수 $(\Psi_{A1}, \Psi_{R1}), (\Psi_{A2}, \Psi_{R2})$ 로 변환된 거리값을 넣어 계산된 공분산 $\text{Cov}_1(\mathbf{h}^{(1)})$ 과 $\text{Cov}_1(\mathbf{h}^{(2)})$ 에 $\omega_1, 1 - \omega_1$ 의 가중치를 주어 $\text{Cov}_2(\mathbf{h})$ 를 계산하였다.

이로써, 각 모형(M.0, M.1, M.2) 별로 이론적 공분산 모형을 구축할 때 고려해야 할 공간 모수의 수가 다름을 확인할 수 있다. 각 모형 별로 모수 추정을 통해 공간 자료의 모형화 및 모형적합을 비교하기 위하여 제 3장에서는 2차원의 공간자료를 생성하여 모의 실험을 시행 하였으며, 제 4장에서는 실제 자료를 이용하여 모수 추정을 위해 붓스트랩을 적용하여 분석을 실시하였다.

제 3 장

모의실험

3.1 모의실험 방법

본 장에서는 등방성 모형 및 비등방성 모형을 적용한 분석을 위해 2.2절에서 제시한 각 모형 별로 모수 추정을 하였다. 기본적인 등방성 모형 (M.0)에서 공간 상관성을 나타내는 너겟, 부분문턱, 범위의 3가지 모수 $\Theta_0 = (\theta_1, \theta_2, \theta_3)^T$, 하나의 주축을 갖는 기하학적 비등방성 모형(M.1)에서는 각도모수 및 비율모수를 포함한 5가지 모수 $\Theta_1 = (\theta_1, \theta_2, \theta_3, \psi_{A_1}, \psi_{R_1})^T$ 를 추정하였으며, 각도 모수와 비율 모수가 각각 2개를 고려하는 비등방성 모형(M.2)은 가중치를 포함한 8가지 모수 $\Theta_2 = (\theta_1, \theta_2, \theta_3, \psi_{A_1}, \psi_{R_1}, \psi_{A_2}, \psi_{R_2}, \omega_1)^T$ 를 추정하였다.

본 절에서는 등방성 및 비등방성을 갖는 자료를 가지고 여러가지 모형의 적용을 위한 추정방법에 관한 설명을 하고자 한다. 이를 위해 3가지의 다른 공간적 성질을 가진 자료를 2개씩 생성하여 구축한 3가지 모형(M.0, M.1, M.2)을 각각 적용시켜 보았다. G.0은 등방성을 따르는 자료이며 G.1은 하나의 방향에서 다른 방향과 다른 공간 상관성을 갖는 자료, 그리고 G.2는 두 방향에서 다른 방향과 다른 공간 상관성을 갖는 자료이다. 각 자료는 평균 0, 각 시나리오별 참값의 모수로 이루어진 공분산 행렬을 갖는 다변량 정규분포 하에서 생성된 자료를 이용하였으며, 3가지 모형(M.0, M.1, M.2)을 각각 적용시켰다. 모의실험 자료의 모수 추정을 위해 로그우도함수를 최대화 하는 모수값을 추정하였다.

표 1: 모의실험 시나리오.

S	Data	n	θ_1	θ_2	θ_3	Ψ_{A1}	Ψ_{R1}	Ψ_{A1}	Ψ_{R2}	ω
1	G.0	200	0.5	3.0	5.0
2	G.0	200	0.5	2.0	4.0
3	G.1	200	0.5	2.0	4.0	π	1/8	.	.	.
4	G.1	200	0.5	2.0	4.0	$\pi/8$	1/8	.	.	.
5	G.2	200	0.5	2.0	4.0	π	1/8	$\pi/2$	1/6	0.4
6	G.2	200	0.5	2.0	4.0	$\pi/8$	1/8	$5\pi/8$	1/5	0.5

S., 시나리오; n., 지점의 수; G.0., 등방성을 따르는 자료; G.1., 하나의 기하학적 비등방성이 존재하는 자료; G.2., 두가지 방향에서 비등방성이 존재하는 자료; θ_1 ., 너겟; θ_2 ., 부분문턱; θ_3 ., 범위; Ψ_{A1} ., 각도 모수; Ψ_{R1} ., 비율 모수; Ψ_{A2} ., 각도 모수; Ψ_{R2} ., 비율 모수.

모의실험을 위한 공분산 모형은 구형모형을 이용하였으며 총 6가지의 시나리오를 고려하였다. 표 1에서 시나리오 1,2는 등방성을 따르는 자료(G.0)로 공간 자료의 모수에 관한 참값이 너겟, 부분문턱, 범위의 3가지로 이루어져 있다. 또한, 시나리오 3,4는 기하학적 비등방성 모수인 각도 모수와 비율모수가 각각 1개씩 존재하는 자료(G.1)로 추정해야 할 모수가 5개이다. 또한, 시나리오 5,6은 기하학적 비등방성 모수가 각각 2개씩 존재하는 자료(G.2)로, 가중치를 포함하여 총 8개의 모수를 추정할 필요가 있다.

비등방성 모형의 적용을 위한 각도 모수(Ψ_A)는 y축을 기준으로 하여 시계방향으로 $(0, \pi)$ 의 구간 안에서 추정이 되도록 하였다. 또한, 장축(주축)과 단축의 비를 의미하는 비율모수(Ψ_R)는 $(0, 1)$ 의 범위를 가지도록 추정하였다. 추정을 위해 사용된 프로그램은 R프로그램(R program)이다. 모수 추정을 위하여 최대우도추정법을 기반으로 함수 알고리즘을 구축하였으며 최적화 패키지(optim package)를 이용하였다. 등방성 모형과 하나의 기하학적 비등방성 모형의 모수를 추정할 수 있는, R프로그램에 내장된 geoR 패키지의 함수의 추정결과와의 추정성능을 확인하였다. 또한, 등방성 모형(M.0)과 하나의 기하학적 비등방성 모

형(M.1), 그리고 두개의 기하학적 비등방성 모형(M.2)의 추정결과를 비교하기 위해 로그우도값(log-likelihood value)과 AIC(Akaike information criterion), BIC(Bayesian information criterion), AICc(corrected AIC)를 이용하였다.

모의실험의 절차는 다음과 같다.

- (1) 균일분포 $U(0, 10)$ 에서 200개의 지점 $\mathbf{s}_i = (s_{1i}, s_{2i})$ 를 생성한다.

$$\mathbf{x} \sim U(0, 10), \quad \mathbf{y} \sim U(0, 10).$$

- (2) (1)에서 구한 위치정보를 이용하여 각 모형(M.0, M.1, M.2)하에서 거리행렬을 계산한다.

- (3) (2)에서 구한 거리행렬을 이용하여 구형 공분산 모형식을 이용한 분산-공분산 행렬을 계산한다.

- (4) 다변량 정규분포하에서 평균이 0 이며 (3)에서 구한 분산-공분산 행렬을 갖는 자료를 각각 299개 생성한다.

- (5) (4)에서 생성된 자료값을 이용하여, 각 모형(M.0, M.1, M.2)하에서 다음과 같은 로그우도함수를 이용하여 모수를 추정한다.

$$l(\Theta; \mathbf{z}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_{\Theta}|) - \frac{1}{2}(\mathbf{z} - \mathbf{X}\beta)^T \Sigma_{\Theta}^{-1}(\mathbf{z} - \mathbf{X}\beta)$$

- (6) 각 자료에서 추정된 평균값 등을 계산해보고, 추정법 간의 추정 성능을 비교하기 위해 로그우도값, AIC, BIC, AICc값을 계산한다.

위와같은 절차에 따라 추정성능을 비교하기 위해 모의실험 결과의 추정값으로 평균(mean), 분산(variance), 편차(bias), 평균제곱오차(mean squared error)를 계산하였다. 각 시나리오별로 생성된 299개의 자료 중 i 번째 자료를 적합한 후 얻게되는 모수의 추정값을 $\hat{\Theta}^i$ 라고 하자. 여기서, M.0의 모수를 Θ_0^i , M.1방법에서는 Θ_1^i , M.2방법에서는 Θ_2^i 이라고 하자. 생성된 모의자료는 $R=299$ 이고, 다음과 같은 통계량들을 계산하였다.

$$\begin{aligned}\hat{\text{Mean}}(\hat{\Theta}^i) &= \frac{1}{R} \sum_{i=1}^R \hat{\Theta}^i, \\ \hat{\text{Var}}(\hat{\Theta}^i) &= \frac{1}{R} \sum_{i=1}^R \left(\hat{\Theta}^i - \overline{\hat{\Theta}}_p \right)^2, \\ \hat{\text{MSE}}(\hat{\Theta}^i) &= \frac{1}{R} \sum_{i=1}^R \left(\hat{\Theta}^i - \Theta_0 \right)^2.\end{aligned}$$

다음으로, 모의실험에서 생성된 자료 별로 각 모형 적합도를 비교하기 위하여 AIC와 AICc, BIC, 로그우도값(l)을 이용 하였다. AIC는 여러 통계 모형들로부터 최적의 모형을 선택하는 기준으로 사용되는 척도로, 추정할 모수의 수에 관한 벌점(penalty term)이 포함되므로 모형이 단순할수록 더 작게 나타나는 경향이 있다. 여기서, 추정해야 할 모수의 수를 k 라고 할 때, AIC는 다음과 같이 표현할 수 있다.

$$\text{AIC} = -2l + 2k.$$

이와 관련된 척도로 BIC가 있는데, 이는 AIC보다 벌점이 크게 고려되므로 AIC보다 단순한 모형을 선택할 확률이 높은 경향을 갖는다. 이는

다음과 같이 표현할 수 있다.

$$\text{BIC} = -2l + \log(n) \cdot k.$$

한편, AICc는 AIC의 단점을 보완한 수정된 AIC라 불리며, 모형 선택 기준으로 모수의 수 및 자료의 수에 관한 영향을 동시에 고려한 척도라고 할 수 있다. AICc는 다음과 같이 계산할 수 있다.

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}.$$

나아가, 각 모의자료에 등방성 및 비등방성 모형을 적용한 결과를 토대로하여 정분류율을 비교해보도록 한다.

표 2: 모의실험 추정결과

S	Param	Truth	M.0				M.1				M.2			
			Mean	Var	Bias	MSE	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE
1 (G.0)	θ_1	0.500	0.499	0.012	-0.001	0.012	0.496	0.012	-0.004	0.012	0.494	0.013	-0.006	0.013
	θ_2	3.000	2.931	0.528	-0.069	0.531	2.995	0.558	-0.005	0.556	3.061	0.623	0.061	0.625
	θ_3	5.000	4.956	0.332	-0.044	0.332	4.727	0.409	-0.273	0.482	4.503	0.441	-0.497	0.687
	Ψ_{A1}	1.000					2.948	0.096	-	-	2.950	0.094	-	-
	Ψ_{R1}		0.864	0.019	-0.136	0.037	0.771	0.049	-0.229	0.102				
	Ψ_{A2}		1.524	0.113	-	-								
	Ψ_{R2}		0.730	0.076	-	-								
ω_1	0.593	0.147	-	-										
2 (G.0)	θ_1	0.500	0.509	0.012	0.009	0.012	0.510	0.013	0.010	0.013	0.507	0.013	0.007	0.013
	θ_2	2.000	1.930	0.282	-0.070	0.286	1.960	0.306	-0.040	0.306	1.984	0.312	-0.016	0.311
	θ_3	4.000	3.967	0.322	-0.033	0.322	3.830	0.400	-0.170	0.428	3.569	0.363	-0.431	0.548
	Ψ_{A1}	1.000					2.949	0.125	-	-	2.944	0.082	-	-
	Ψ_{R1}		0.888	0.022	-0.112	0.034	0.784	0.054	-0.216	0.101				
	Ψ_{A2}		1.516	0.132	-	-								
	Ψ_{R2}		0.708	0.074	-	-								
ω_1	0.503	0.153	-	-										
3 (G.1)	θ_1	0.500	0.508	0.014	0.008	0.014	0.506	0.007	0.006	0.007	0.497	0.007	-0.003	0.007
	θ_2	2.000	1.523	0.566	-0.477	0.791	1.838	0.390	-0.162	0.415	1.848	0.408	-0.152	0.430
	θ_3	4.000	5.205	5.412	1.206	6.847	3.951	0.631	-0.048	0.631	3.920	0.662	-0.080	0.666
	Ψ_{A1}	3.142					3.117	0.001	-0.024	0.002	3.117	0.001	-0.024	0.002
	Ψ_{R1}	0.125					0.134	0.004	0.009	0.004	0.114	0.004	-0.011	0.004
	Ψ_{A2}	1.000									1.532	0.146	-	-
	Ψ_{R2}		0.373	0.074	-	-								
ω_1	0.959		0.005	-0.041	0.006									
4 (G.1)	θ_1	0.500	0.507	0.014	0.008	0.014	0.502	0.007	0.002	0.007	0.492	0.007	-0.008	0.007
	θ_2	2.000	1.485	0.412	-0.515	0.676	1.918	0.435	-0.082	0.441	1.924	0.461	-0.076	0.465
	θ_3	4.000	4.835	2.052	0.835	2.743	3.934	0.512	-0.066	0.515	3.921	0.565	-0.079	0.570
	Ψ_{A1}	0.393					0.394	0.003	0.002	0.003	0.391	0.003	-0.002	0.003
	Ψ_{R1}	0.125					0.125	0.004	-	0.004	0.101	0.005	-0.024	0.005
	Ψ_{A2}	1.000									1.490	0.156	-	-
	Ψ_{R2}		0.380	0.073	-	-								
ω_1	0.950		0.006	-0.050	0.008									
5 (G.2)	θ_1	0.500	0.492	0.012	-0.008	0.012	0.491	0.010	-0.009	0.010	0.500	0.008	-	0.008
	θ_2	2.000	1.552	0.311	-0.448	0.511	1.644	0.325	-0.356	0.450	1.886	0.381	-0.114	0.393
	θ_3	4.000	4.823	1.885	0.823	2.556	4.510	1.117	0.510	1.373	3.900	0.528	-0.100	0.536
	Ψ_{A1}	3.142					2.907	0.166	-0.234	0.221	3.094	0.009	-0.048	0.011
	Ψ_{R1}	0.125					0.820	0.048	0.694	0.530	0.130	0.019	0.005	0.019
	Ψ_{A2}	1.571									1.555	0.017	-0.016	0.017
	Ψ_{R2}	0.167									0.148	0.014	-0.018	0.014
ω_1	0.400									0.416	0.026	0.016	0.027	
6 (G.2)	θ_1	0.500	0.500	0.010	-	0.010	0.507	0.011	0.007	0.011	0.499	0.009	-0.001	0.009
	θ_2	2.000	1.518	0.278	-0.482	0.509	1.673	0.355	-0.327	0.460	1.916	0.415	-0.084	0.420
	θ_3	4.000	4.715	1.020	0.715	1.529	4.467	1.074	0.467	1.289	3.920	0.418	-0.080	0.423
	Ψ_{A1}	0.393					0.474	0.200	0.082	0.206	0.387	0.010	-0.005	0.010
	Ψ_{R1}	0.125					0.706	0.056	0.581	0.394	0.102	0.011	-0.023	0.012
	Ψ_{A2}	1.964									1.952	0.019	-0.011	0.019
	Ψ_{R2}	0.200									0.180	0.021	-0.020	0.021
ω_1	0.500									0.506	0.025	0.006	0.025	

S., 시나리오; Param., 모수; Truth., 참값; G.0., Mean., 평균; Var., 분산; Bias., 편차; MSE., 평균제곱오차; 등방성을 따르는 자료; G.1., 하나의 기하학적 비등방성이 존재하는 자료; G.2., 두가지 방향에서 비등방성이 존재하는 자료; M.0., 등방성 모형; M.1., 하나의 기하학적 비등방성 모형; M.2., 두가지 방향에서 비등방성 모형.

3.2 모의실험 결과

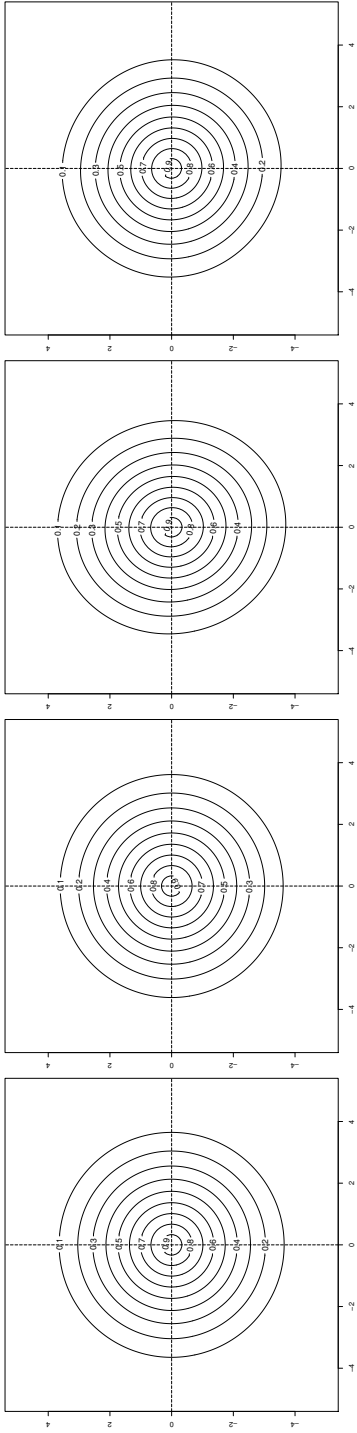
모의실험을 위해 생성된 각 시나리오에 대한 모수들의 참값 및 추정값은 표 2와 같다. 각 시나리오에 관한 추정값은 최대우도 추정법

을 기반으로한 M.0, M.1과 M.2의 각각의 모형을 적용하여 추정값들의 평균, 분산, 평균제곱오차 값을 계산하였다. 먼저, 모의실험 추정값들 중 평균값을 살펴보면 추정값이 모수의 참값과 유사한 경향을 보임을 확인할 수 있다. 등방성 모형을 전제로 한 자료(G.0) 시나리오 1,2의 참값에 관련된 모수는 너겟, 부분문턱과 범위의 3가지의 공통 모수가 존재하는 것으로 비등방성 모형을 적용한 추정법인 M.1, M.2에 관해서는 모든 방향에서 공간 상관성을 나타내는 거리가 같다는 의미의 비율모수(ψ_{R1})가 1을 참값으로 하여 편차 및 평균제곱오차를 계산하였다. 시나리오 1, 2에 관한 M.1모형의 각도모수의 추정값은 0.864, 0.888로 1에 가까운 경향을 보였으며 M.2모형의 경우에는 2개의 비율모수(ψ_{R1}, ψ_{R2})가 0.7과 0.8사이 값으로 추정되어 이의 가중치가 0.5와 0.6사이의 값을 가지는 것으로 나타났다. 이는 임의의 두 방향에 관하여 각각이 비슷한 비율 모수를 가짐을 나타내는 것이다. 또한 G.0와 관련된 시나리오 1,2에서 너겟, 부분문턱과 범위의 추정값에 대해 M.0 모형이 다른 모형에 비해 분산과 평균제곱오차가 작았다.

다음으로, G.1과 관련된 시나리오 3, 4를 보자. 3가지 공통모수에 관한 추정결과는 M.1모형을 적용한 경우의 분산과 평균제곱오차가 다른모형에 비해 작은 경향을 보였으며, 그 중 범위의 경우에는 편차도 가장 작았다. 또한, 각도모수(ψ_{A1})과 비율모수(ψ_{R1})의 추정결과에 대해 M.2모형에 비해 M.1모형의 분산, 편차와 평균제곱오차가 작거나 유사하게 나타났다. 또한 G.1을 M.2모형으로 추정된 가중치의 추정값을 살펴보면, 0.959 및 0.950으로 가중치가 1에 가까운 값을 가지는 것을 확인할 수 있다. 이는, 특정한 한 방향에서의 공간 상관성을 나타내는 공분산 함수 1개가 고려된 가중치가 1인 경우와 유사하므로 G.1의 경향이 잘 나타난 결과라 할 수 있다.

수직을 이루는 특정한 두 방향에 관한 비등방성을 고려한 자료 (G.2)인 5,6의 경우에는 너겟과 부분문턱, 범위의 추정값은 M.0, M.1 모형을 적용한 경우보다 M.2모형을 적용하였을 때 평균제곱오차가 더 작았다. 또한, 각도모수와 비율모수에 관하여 M.2를 적용한 경우에 M.1을 적용했을 때보다 분산, 편차, 평균제곱오차가 모두 작았다.

각 시나리오 별로 모의실험 추정결과를 살펴보면 G.0에 관하여 M.0이, G.1에 관하여 M.1이, 그리고 G.2에 관하여 M.2에 추정값에 관한 분산, 평균제곱오차 등이 작은 경향을 나타냈다. 생성된 자료(G.0, G.1, G.2)에 따라 적용된 모형(M.0, M.1, M.2) 별 추정 결과를 기반으로 각 시나리오 간의 추정법 별 등고선 그림을 이용하여 각각의 공간 상관성을 확인해 볼 수 있다. 등방성 모형의 경우에는 모든 방향에서 같은 거리를 가지는 경우 동심원 형태의 공간 상관성을 나타내며, 비등방성 모형의 경우에는 특정 방향에서 강한 공간 상관성을 보이는 형태를 가진다. 또한, 확장된 기하학적 비등방성 모형에서는 2가지 방향에 관한 비등방성 모수를 이용하여 2가지 방향에 관한 방향성을 가지게 된다. 각 자료에 어떤 모형을 적용했는지에 따라 참값에 관한 추정값이 나타나게 될 것이며, 추정 결과를 바탕으로 공간 상관성을 나타내는 그림을 그려볼 수 있다.

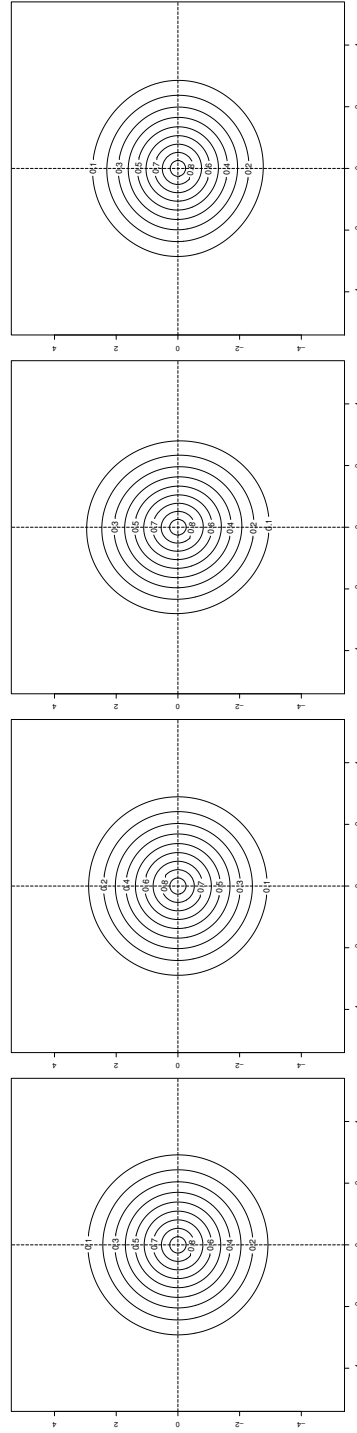


(a) S1-Truth

(b) S1-M.0

(c) S1-M.1

(d) S1-M.2



(e) S1-Truth

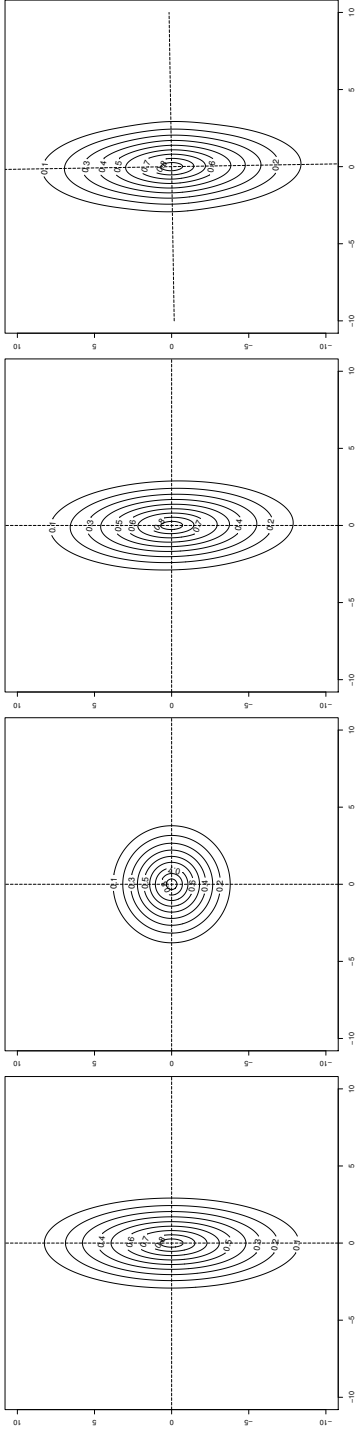
(f) S2-M.0

(g) S2-M.1

(h) S2-M.2

그림 7: 등방성 자료(G.0) 결과

S1., 시나리오1; S2., 시나리오2; Truth., 참값; M.0., 등방성 모형; M.1., 하나의 기하학적 비등방성 모형; M.2., 두개의 기하학적 비등방성 모형.

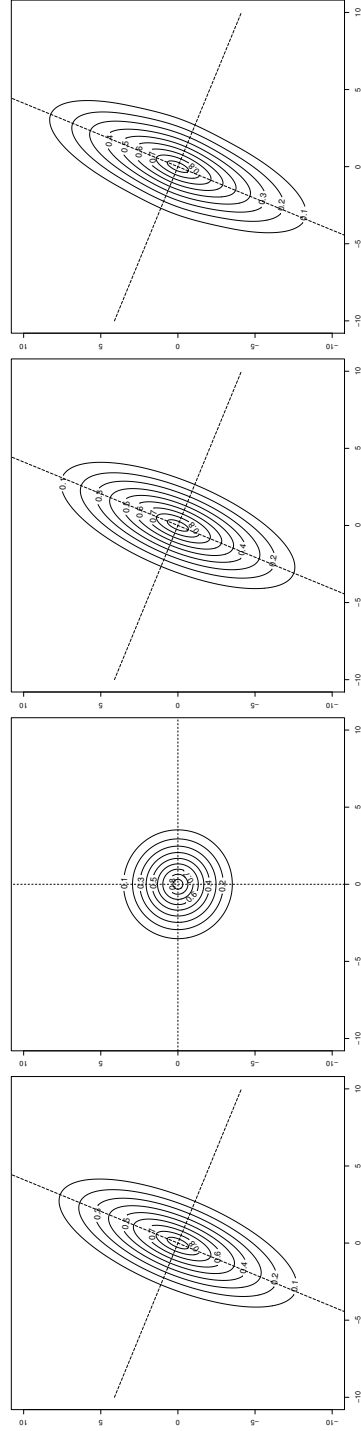


(a) S3-Truth

(b) S3-M.0

(c) S3-M.1

(d) S3-M.2



(e) S4-Truth

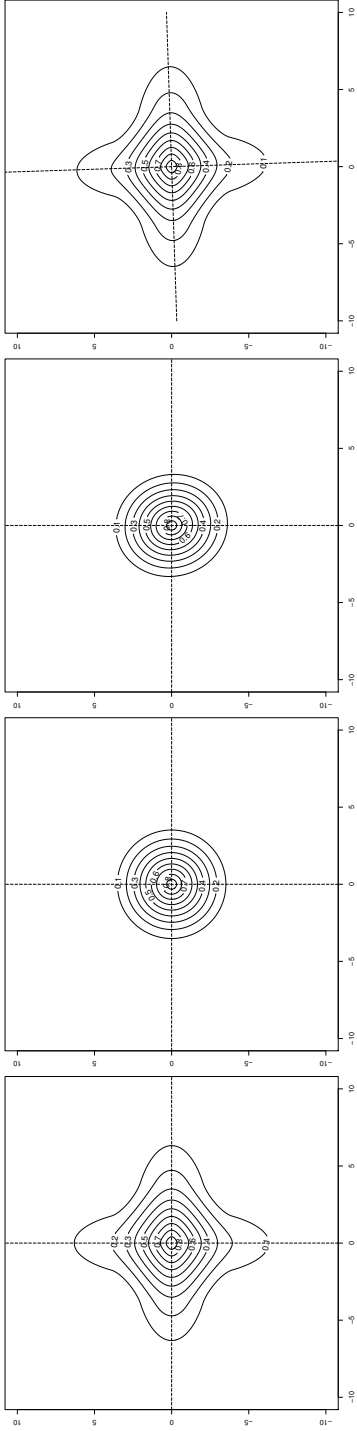
(f) S4-M.0

(g) S4-M.1

(h) S4-M.2

그림 8: 비등방성 자료(G.1) 결과

S3., 시나리오3; S4., 시나리오4; Truth., 참값; M.0., 등방성 모형; M.1., 하나의 기하학적 비등방성 모형; M.2., 두개의 기하학적 비등방성 모형.

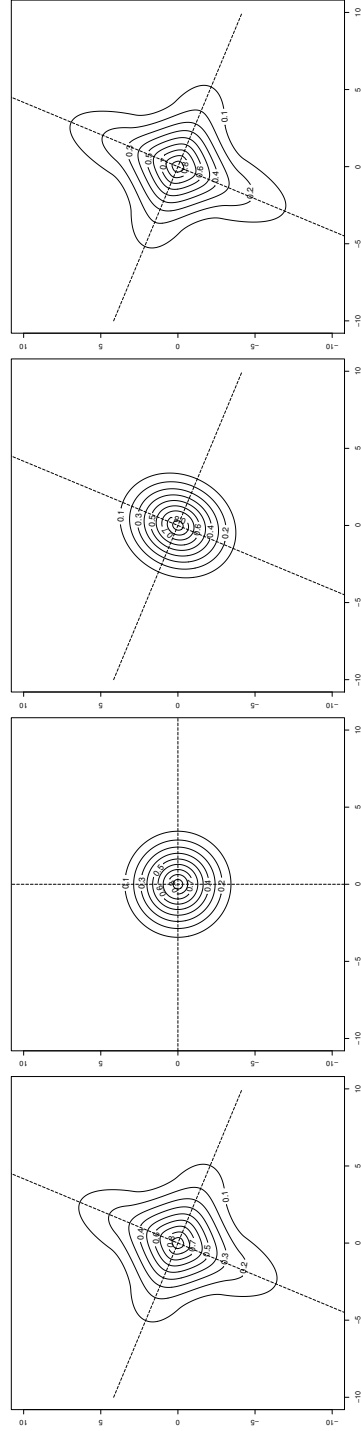


(d) S5-M.2

(c) S5-M.1

(b) S5-M.0

(a) S5-Truth



(h) S6-M.2

(g) S6-M.1

(f) S6-M.0

(e) S6-Truth

그림 9: 비등방성 자료(G.2) 결과

S5., 시나리오5; S6., 시나리오6; Truth., 참값; M.0., 등방성 모형; M.1., 하나의 기하학적 비등방성 모형; M.2., 두개의 기하학적 비등방성 모형.

그림 7부터 그림 9는 각각의 모의실험 자료의 추정결과를 살펴보기 위해 시나리오 별 참값과 각 추정법을 적용한 결과를 토대로 공간 상관성을 나타내는 등고선 그림(contour plot)이다. 이는 왼쪽부터 각 시나리오 별로 참값(true value), M.0, M.1, 그리고 M.2의 추정결과를 이용하여 공간 상관성을 나타내는 그림이다.

먼저, 그림 7을 보면 G.0에 관하여 참값 및 추정결과 모두 모든 방향에서 같은 상관거리를 보이는 경향을 나타냄을 확인할 수 있다. 시나리오 1,2 자료의 범위의 참값은 각각 5, 4로 시나리오 2에 비해 시나리오 1의 참값 및 추정결과가 더 긴 상관성을 보이는 결과를 확인할 수 있다. 그림 8은 G.1에 관한 것으로 참값의 경우 특정한 방향으로 긴 상관성을 나타내는 것을 확인할 수 있다. 이러한 특성을 가진 G.1에 M.0모형을 적용하는 경우, 각도 모수 및 비율 모수에 관한 추정이 불가능하므로 등방성 모형의 형태가 나타났다. M.1과 M.2의 추정법의 경우 참값과 유사한 패턴을 보이는 그림을 확인할 수 있다. 마지막으로, 그림 9는 G.2에 관한 것으로 참값의 경우 수직을 이루는 특정한 두 방향에 관한 공간 상관성이 길게 나타는 경향을 확인할 수 있다. 이러한 경향을 가진 G.2에 M.0을 적용하는 경우는 등방성을 전제로 하기에 모든방향으로 같은 상관성을 나타내게 된다. M.1의 경우에는 특정한 방향에 관한 비등방성 모수만을 추정하게 된다. 이를 M.2에 적용하는 경우 참값과 유사한 공간 상관성을 보였다.

모의실험에 사용한 자료와 추정법에 관하여 모형간 비교를 하기 위해 AIC, AICc, BIC, 로그우도값을 이용하였다. 각 시나리오 별 자료에 대해 M.0, M.1, M.2방법을 적용하여 얻은 모형 적합도와 관련된 통계량을 비교하여 정분류율을 계산하였다.

표 3: 오분류 평가 결과

	S	M.0				M.1				M.2			
		-2logL	AIC	AICc	BIC	-2logL	AIC	AICc	BIC	-2logL	AIC	AICc	BIC
G.0	1	0.027	0.883	0.896	1.000	0.304	0.070	0.067	0.000	0.669	0.047	0.037	0.000
	2	0.007	0.906	0.916	0.997	0.237	0.057	0.054	0.003	0.756	0.037	0.030	0.000
G.1	3	0.000	0.000	0.000	0.033	0.488	0.940	0.943	0.963	0.512	0.060	0.057	0.003
	4	0.000	0.007	0.007	0.030	0.472	0.896	0.913	0.970	0.528	0.097	0.080	0.000
G.2	5	0.000	0.134	0.151	0.803	0.003	0.047	0.043	0.043	0.997	0.819	0.806	0.154
	6	0.000	0.191	0.214	0.863	0.003	0.043	0.050	0.017	0.997	0.766	0.736	0.120

S., 시나리오; G.0., 등방성을 따르는 자료; G.1., 하나의 기하학적 비등방성이 존재하는 자료; G.2., 두가지 방향에서 비등방성이 존재하는 자료; M.0., 등방성 모형; M.1., 하나의 기하학적 비등방성 모형; M.2., 두가지 방향에서 비등방성 모형.

표 3에서 시나리오 1,2는 G.0, 시나리오 3,4이 G.1, 시나리오 5,6는 G.2에 속한다. 먼저, G.0을 M.0모형을 적용하여 추정했을 경우에 관하여 살펴보자. 등방성 자료(G.0)인 시나리오 1,2에 관하여 M.0의 -2logL이 M.1과 M.2에 비해 작은 경우는 매우 작은 값을 나타냈고 AIC, AICc, BIC 관점에서는 M.0이 최적의 모형임을 확인할 수 있다. G.1인 시나리오 3,4에서 하나의 기하학적 비등방성 모형인 M.1을 이용한 경우에 -2logL이 M.0과 M.2에 비해 작은 경우가 48.7%, 47.2%이었다. 시나리오 3에서 M.1이 M.0과 M.2에 비해 AIC, AICc, BIC가 작은 비율이 각각 94%, 94.3%, 96.3%였다. 이는 하나의 기하학적 비등방성이 존재하는 자료인 G.1에 관하여 M.1모형으로의 추정성능이 더 좋음을 확인할 수 있다. 또한, G.2의 시나리오 5,6에서 M.2모형이 M.0과 M.1에 비해 -2logL가 작은 비율이 99.7%로 나타냈으며, AIC, AICc의 경우에도 이와 유사한 결과를 보였다. 이는 각 자료의 공간 상관성을 잘 표현할 수 있는 모형을 적용하였을 때 추정성능이 높아진다는 것을 의미한다.

제 4 장

실증연구

4.1 실증 자료

본 장에서는 비등방성 모형의 모수 추정방법을 실제 자료에 적용한 결과를 소개하고자 한다. 실증연구에서 비등방성 모형의 적용을 위해 사용한 자료는 미세먼지(PM_{10})자료이다. 모형적합을 위하여 위치정보인 위도, 경도 이외에 이산화황(SO_2)자료와 도시규모변수를 사용하였다. 본 연구에서 사용한 미세먼지 및 이산화황 자료는 환경부 대기환경정보(<http://www.airkorea.or.kr/>)에서 얻은 2011년 연평균 자료이다. 도시규모변수는 서울특별시 외 6개광역시(인천광역시, 광주광역시, 인천광역시, 울산광역시, 대구광역시, 대전광역시)이면 0, 아니면 1의 값을 갖는다. 본 연구에서 사용한 대기환경자료는 도심(거주) 지역의 평균대기질 농도를 측정하는 도시대기측정망, 도시를 둘러싼 교외 지역의 배경농도를 측정하기 위한 교외대기측정망, 국가의 배경농도를 파악하고 외국으로의 오염물질 유입과 유출상태 등을 파악하는 국가배경농도 측정망, 자동차 통행량과 유동인구가 많은 도로변 대기 측정망 등에서 측정된 자료이다.

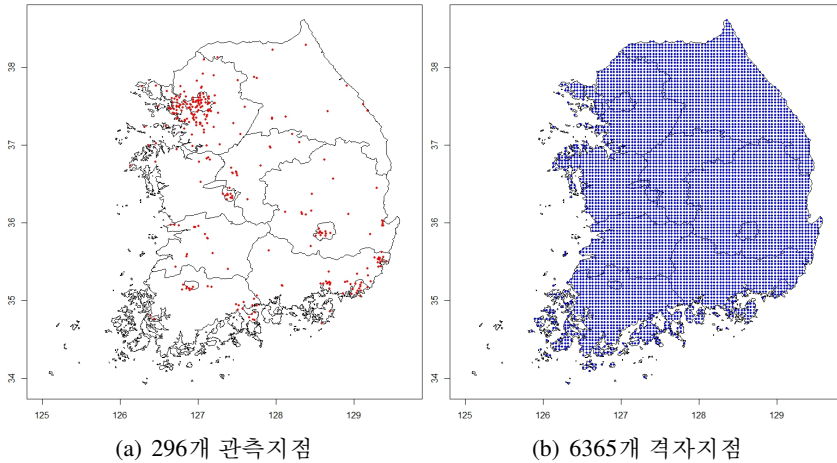


그림 10: 미세먼지 대기오염측정망과 6365개의 격자지점의 위치

4.2 실증연구 방법

본 연구의 분석을 위해 제주도 및 울릉도에 있는 측정망을 제외한 그림 10(a)에서 나타난 총 296개 관측지점의 측정 자료를 사용하였다. 또한, 자료의 모형화를 위해 공간자료의 확률과정 중 평균함수가 선형 결합으로 이루어진 일반크리깅을 고려하였다. 선형결합으로 이루어진 평균함수 부분에는 위치 정보인 위도, 경도 및 이산화황(SO₂)자료와 도시규모 변수가 포함된다. 자료가 정규성을 만족하도록 박스-콕스 변환(box-cox transformation)을 이용하여 미세먼지(PM₁₀) 자료에 제공된 변환을 하여 분석을 실시하였으며, 그림 10(b)에 표시된 6365개의 임의 격자지점에 관한 예측지도를 작성하였다. 실증 자료에 공간 상관성을 의미하는 공분산 모형의 모수들을 추정하기 위해 최대우도추정법을 기반으로 하여 구축한 3가지의 모형(M.0, M.1, M.2)을 적용하여 모수를 추정하였다. 이 때, 검정을 위한 구간추정을 위해서 실증자료에 모수적 붓스트랩을 통해 자료를 생성하여 실증 연구를 실시하였다.

본 연구에서는 지리통계자료의 모수적 붓스트랩을 통해 자료를 생성하였다. 공간자료의 모수적 붓스트랩을 수행하기 위해 다변량 정규분포를 가정하여 표본을 추출하였다. 공간자료의 모수적 붓스트랩 (parametric bootstrap)은 다음과 같은 절차를 갖는다.

- (1) 관측된 공간자료의 모형적합을 통해 선택된 모형과 추정된 값을 이용하여 분산-공분산 행렬을 계산한다.

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

- (2) 다변량 정규분포 $N_n(\mathbf{0}, \hat{\boldsymbol{\Sigma}}_{\Theta})$ 에서 표본 크기(n=296)인 잔차($\boldsymbol{\epsilon}_b^* = (\boldsymbol{\epsilon}_1^*, \boldsymbol{\epsilon}_2^*, \dots, \boldsymbol{\epsilon}_n^*)^T$)를 $B = 199$ 개 난수생성한다.
- (3) 단계(2)에서 생성된 B 개의 붓스트랩 표본을 각 모형(M.0, M.1, M.2)에 적용하여 해당 모수들을 추정한다.
- (4) 단계(3)에서 구한 각 모수의 추정값들을 이용하여 평균, 분산, 편향, 평균제곱오차, 그리고 다음과 같은 경험적 $100(1 - \alpha)\%$ 신뢰구간 등을 계산한다.

$$\left(\hat{\Theta}^{\alpha B}, \hat{\Theta}^{(1-\alpha)B} \right).$$

여기서, $\{\hat{\Theta}^{ab}, b = 1, \dots, B\}$ 의 $(100 \times \alpha)$ 분위수를 의미한다.

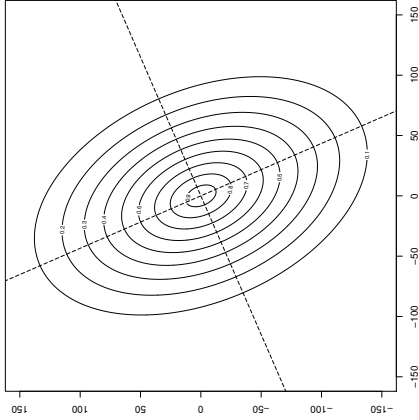
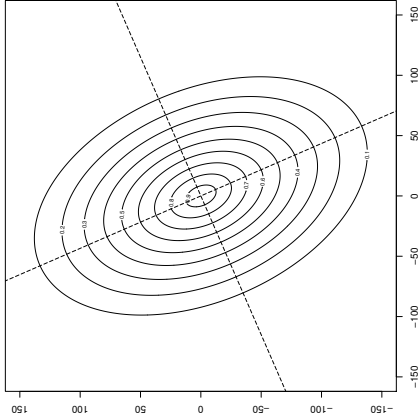
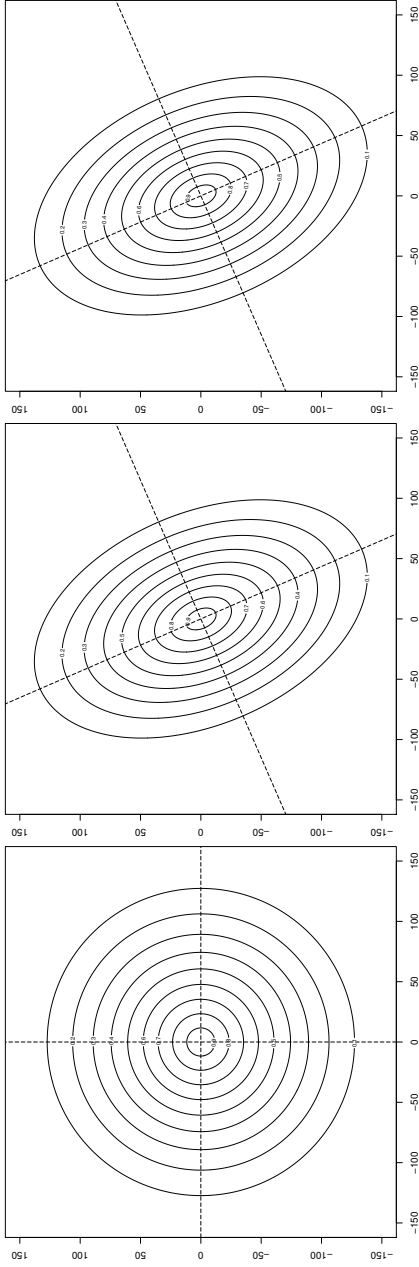
표 4: PM₁₀ 자료 및 붓스트랩 표본자료의 최대우도추정법에 의한 모수 추정 결과

Model	Param	Truth	Mean	Var	Bias	MSE	95% C.I.	
M.0	θ_1	0.230	0.234	0.001	0.005	0.001	0.186	0.285
	θ_2	0.172	0.108	0.006	-0.064	0.010	0.000	0.287
	θ_3	174.591	152.078	1646.491	-22.513	2145.056	56.296	212.437
M.1	θ_1	0.230	0.237	0.001	0.007	0.001	0.194	0.285
	θ_2	0.171	0.124	0.007	-0.046	0.009	0.000	0.377
	θ_3	120.177	103.343	774.222	-16.834	1053.723	36.328	143.046
	Ψ_{A1}	-0.404	-0.400	0.116	0.004	0.116	-1.140	0.176
	Ψ_{R1}	0.364	0.315	0.055	-0.049	0.057	0.010	0.999
M.2	θ_1	0.230	0.230	0.000	-0.000	0.000	0.189	0.279
	θ_2	0.171	0.134	0.007	-0.037	0.008	0.000	0.325
	θ_3	120.178	102.845	822.699	-17.333	1118.998	34.380	136.405
	Ψ_{A1}	-0.404	-0.412	0.096	-0.009	0.096	-1.377	0.201
	Ψ_{R1}	0.364	0.253	0.054	-0.111	0.066	0.010	0.878
	Ψ_{A2}	-0.544	-0.530	0.184	0.014	0.183	-1.571	0.739
	Ψ_{R2}	0.937	0.797	0.092	-0.140	0.111	0.010	0.999
	ω_1	0.999	0.813	0.075	-0.187	0.109	0.000	0.999

Param., 모수; Truth., 기존 미세먼지(PM₁₀)자료를 이용하여 추정된 결과; Mean., 붓스트랩 표본 평균; Var., 붓스트랩 표본분산; Bias., 붓스트랩 편향; MSE., 붓스트랩 평균제곱오차; 95% C.I., 붓스트랩 95% 경험적 신뢰구간; θ_1 ., 너겟; θ_2 ., 부분 문턱; θ_3 ., 범위; Ψ_{A1} ., 방향 모수; Ψ_{R1} ., 비율 모수; ω_1 ., 가중치; M.0., 등방성 모형 M.1., 하나의 기하학적 비등방성 모형, M.2., 두개의 기하학적 비등방성 모형.

4.3 실증연구 결과

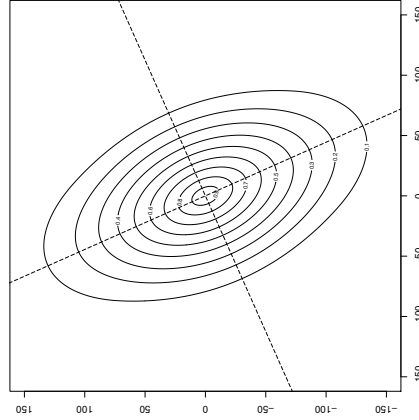
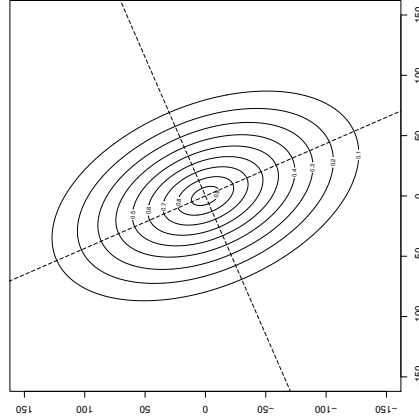
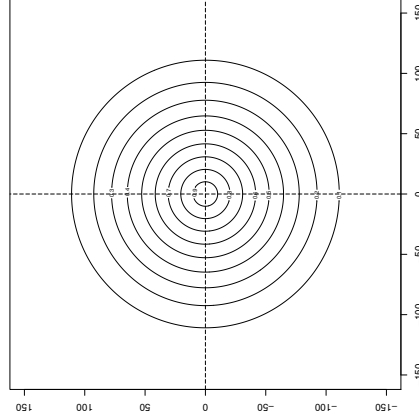
4.1절에서 소개한 미세먼지 자료를 이용하여 실증연구를 실시하였다. 모수 추정을 위해 추세를 나타내는 평균함수 부분에 위치 정보(위도, 경도)와 대기환경물질인 이산화황 자료와 도시규모변수를 고려하여 잔차를 더한 뒤 모형을 적합하는 방법을 이용하였다. 표 4는 최대우도추정법을 기반으로 3가지 모형을 적용하여 공간자료의 모수적 붓스트랩을 이용하여 모수추정을 한 결과이다. 먼저, 붓스트랩을 위한 참값은 이용한 원래의 자료를 이용하여 M.0, M.1과 M.2를 적용시킨 모수추정 결과를 이용하였다. 공간자료에 모수적 붓스트랩을 이용하여 추정된 평균값을 살펴보면 너겟(θ_1)과 부분 문턱(θ_2)은 비교적 잘 추정된 것으로 보인다.



(a) Truth-M.0

(b) Truth-M.1

(c) Truth-M.2



(d) BOOT-M.0

(e) BOOT-M.1

(f) BOOT-M.2

그림 11: 미세먼지(PM_{10})자료 추정값 및 붓스트랩 추정결과

Truth., 기존 미세먼지(PM_{10})자료의 추정값; BOOT., 붓스트랩 표본을 이용한 추정값; M.0., 등방성 모형; M.1., 하나의 기하학적 비등방성 모형; M.2., 두 방향에 관한 기하학적 비등방성 모형.

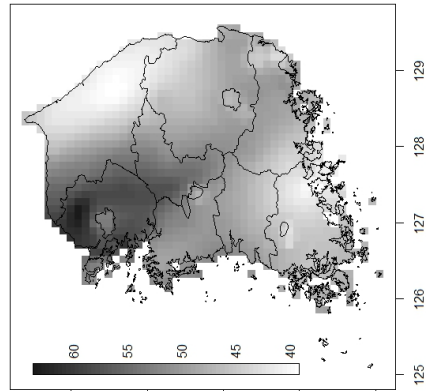
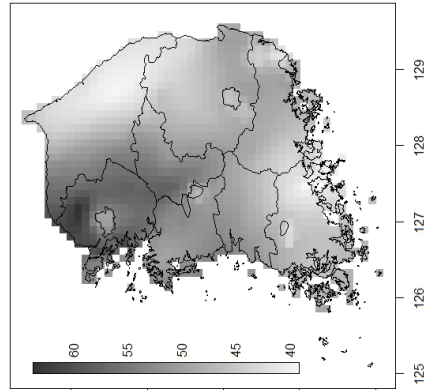
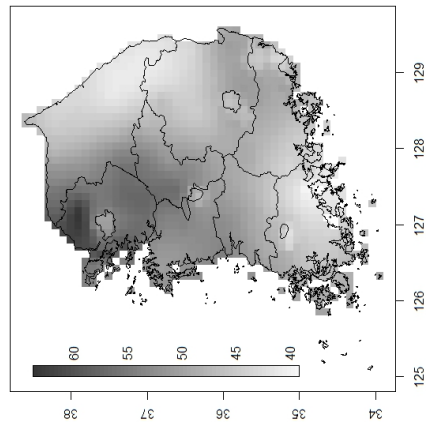
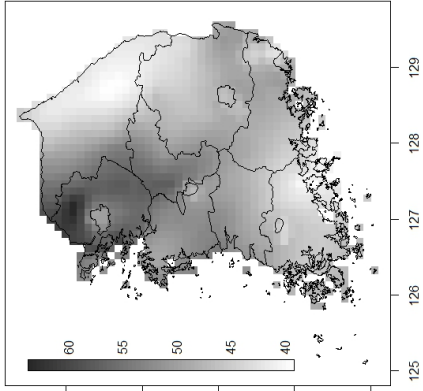
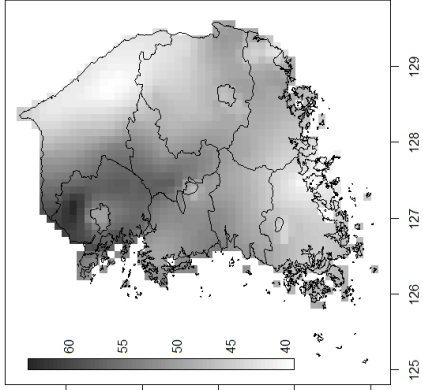
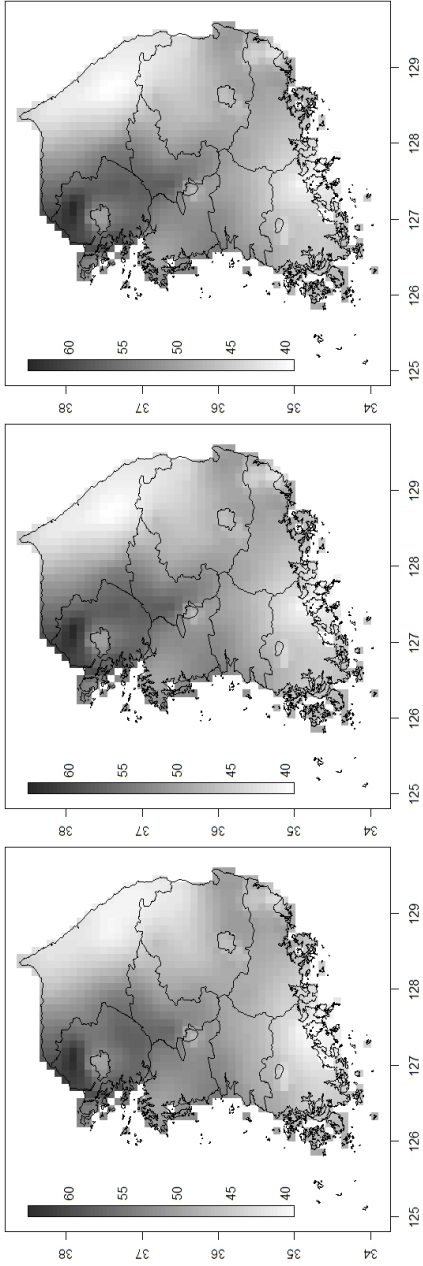


그림 12: 실증연구 예측지도

Truth., 기존 미세 먼지(PM_{10})자료의 추정값; BOOT., 붓스트랩 표본을 이용한 추정값; M.0., 등방성 모형; M.1., 하나의 기하학적 비등방성 모형; M.2., 두 방향에 관한 기하학적 비등방성 모형.

그러나, 범위(θ_3)의 경우 참값(기존의 미세먼지(PM_{10})자료의 추정값)에 비해 과소 추정된 결과를 보였으며 이에 관한 붓스트랩 자료에 관한 분산 및 편차, 평균제곱오차가 크게 나타났다. 붓스트랩 표본을 이용한 3가지 모형을 적용한 추정 결과를 살펴보면 M.0의 부분문턱 및 범위에 관한 편차와 평균제곱오차가 M.1과 M.2에 비해 크게 나타났다. M.1를 적용한 경우에는 각도 모수 및 방향 모수에 관한 추정값이 -0.400, 0.315로 참값인 -0.404, 0.364와 유사한 경향을 보였으나 각도 모수에 관한 분산 및 평균제곱오차가 비교적 큰 경향을 띄었다. 마지막으로 M.2의 경우, 비등방성 모형을 적용했을 때 가중치 모수에 관한 참값이 0.999로 1에 가까우므로 한 방향에 관한 비등방성 모형만이 유효하다고 해석할 수 있다.

그림 11의 (a), (b), (c)는 기존의 미세먼지(PM_{10})자료를 이용하여 추정한 값에 관한 그림으로, 붓스트랩 자료를 이용한 (d), (e), (f)와 비슷한 양상을 보임을 확인할 수 있다. 그림 12은 앞서 언급한 미세먼지(PM_{10})자료의 모수 추정결과를 이용한 임의의 지점 6365개에 관한 공간 예측 결과를 나타낸 그림이다. 그림 12의 (a), (b), (c)는 기존 자료를 이용하여 구축한 모형을 적용시켜 얻은 모수 추정 결과를 이용한 표 2의 참값에 관한 예측지도이다. 그림 12의 (d), (e), (f)는 붓스트랩을 이용한 추정 결과를 기반으로 그려진 예측지도이다. 6개의 예측지도 모두 비슷한 경향을 보이는 것을 확인할 수 있다. 특히, 서울을 둘러싸고 있는 경기지역 및 4개의 예측지도 모두 비슷한 경향을 보이는 것을 확인할 수 있다. 또한, 대전, 대구, 울산, 부산, 광주 등의 광역시에서 다른 지역보다 더 낮은 미세먼지(PM_{10}) 농도를 보이는 것을 확인할 수 있다.

제 5 장

결론

공간 자료가 특정한 하나 이상의 방향으로 긴 상관관계를 나타낼 때, 이를 고려한 분석을 함으로서 모형적합도 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 나타내는 모수들을 추정하였다. 이의 모형 적합도를 확인해보기 위해 모의실험을 실시하였고, 구형모형의 이론적 공분산 모형을 이용한 6가지 시나리오에 따라 제시한 3가지의 모형(M.0, M.1, M.2)을 적용시켜 보았다.

모의실험에 결과, 모수 추정결과는 참값과 유사한 경향을 보였다. 등방성을 따르는 자료(G.0)에 관해서는 등방성 모형(M.0)의 추정결과를 보면 시나리오 1에 관해서는 AIC가 작을 확률이 88.3%, 다른 모형을 적용했을 때 보다 AICc가 작을 확률이 89.6%였다. 한편, G.1자료 중 시나리오 3의 경우에는 M.1모형을 적용했을 경우 AIC, AICc가 작을 확률이 94%, 94.3%였으며, G.2자료 중 시나리오 5의 경우에는 $-2\log L$, AIC, AICc가 다른 모형을 적용했을 때보다 M.2를 적용했을 때 더 작을 확률이 99.7%, 81.9%, 80.6%로 가장 높았다. 즉, 등방성을 따르는 자료(G.0)

에는 등방성 모형(M.0)을 적용한 경우가, 한 방향으로 기하학적 비등방성 모형이 존재하는 자료(G.1)에서는 하나의 기하학적 모형(M.1)을 추정된 결과가, G.2자료를 두개의 기하학적 모형(M.2)으로 추정된 결과가 가장 효율적이었다고 할 수 있다.

실증연구에서는 구축한 비등방성 모형을 적용하고, 공간자료의 모수적 붓스트랩을 이용하였다. 공간자료의 붓스트랩을 이용함으로써, 모수에 관한 신뢰구간을 계산할 수 있었다. 실증연구에서 사용된 자료는 미세먼지(PM₁₀)의 2010년 연평균 농도이다. 실증연구를 위해 모의실험에서 적용했던 3가지 모형을 이용하여 붓스트랩 자료와 원자료의 추정결과를 비교하였다. 여기서, 원자료는 실제로 관측된 296개의 미세먼지(PM₁₀)자료를 의미하며, 이를 이용하여 얻은 추정결과를 참값으로 두었다. 실증연구 결과로 M.1의 경우 대체적으로 M.2의 경우보다 평균편차제곱값과 편차, 분산이 작은 경향을 나타냈다. 참값과 붓스트랩을 이용한 추정값을 비교해보면, 공간 모수인 범위의 경우 다소 과소 추정된 결과를 확인할 수 있었다. 한편, M.2의 추정 결과로 가중치의 추정값이 1에 가까운 값으로 하나의 방향에서의 비등방성만이 유효한 결과를 보였다. 이의 추정결과를 이용하여 구성된 예측지도에서는 서울을 둘러싸고 있는 경기지역에서 높은 농도를 보이는 것으로 모든 모형에서 비슷한 경향을 보이는 것을 확인할 수 있다. 또한, 대전, 대구, 울산, 부산, 광주 등의 광역시에서 다른 지역보다 더 낮은 미세먼지(PM₁₀) 농도를 보임을 확인하였다.

향후에는 공간자료에 기하학적 비등방성이 3개 이상이 존재하는 경우의 형태를 고려한 추정법을 제시하고자 한다. 또한, 최대우도추정법 뿐 아니라 잔차최대우도추정법과 최소제곱법을 이용한 방법에 관해 연구할 필요가 있다고 판단된다.

참고 문헌

- [1] 강창완, 최승배, 조장식 (2008). 다변량 공간자료에서의 코크리깅 기법 활용 연구, 한국자료분석학회, 제10권, 2661-2668.
- [2] 김규곤, 최승배 (2000). 환경자료를 이용한 공간통계 분석법의 예측력 연구, 한국자료분석학회지, 제4권, 449-510.
- [3] 김봉조 (2005). 공간 상관관계를 고려한 Variogram 선택, 성균관 대학교 대학원, 석사학위논문.
- [4] 김호용 (2010). 공간통계기법을 이용한 도시 교통량 예측의 정확성 향상, 한국지리정보학회지, 제13권, 4호, 138-147.
- [5] 박노욱 (2011). 시계열 환경변수 분포도 작성 및 불확실성 모델링: 미세먼지(PM-10)농도 분포도 작성 사례연구, 한국지구과학회지, 제32권, 제3호, 249-264.
- [6] 박만식, 엄진기, 허태영 (2013). 교통카드 자료를 이용한 서울시 지역별 대중교통 수단 선택 공간상관성 분석, 응용통계연구, 제26권, 제4호, 623-634.
- [7] 박민규, 박창열, 신기일, 유철상 (2010). 일강우량의 적정 베리오그램(On Proper Variograms of Daily Rainfall Data), 대한토목학회, 제30권, 525-532.
- [8] 정승환, 박만식, 김기환 (2010). 풍속 자료의 공간예측, 응용통계연구, 제23권, 345-356.
- [9] 정지용, 진서훈, 박만식 (2008). 공간적 상관구조를 포함하는 선형회귀모형을 이용한 강수량 자료 분석, 응용통계연구, 제21권, 1093-1107.
- [10] 조재영, 최승배, 김규곤 (2001). 일반통계에 대한 공간통계 방법의 예측성능에 관한 연구, 한국자료분석학회, 제3권, 41-49.

- [11] 최승배, 문승호, 강창완, 조장식, 이정형 (2008). SAS/STAT을 이용한 공간예측, 자유아카데미.
- [12] 최정순, 박만식, 원유복, 김학열, 허태영 (2010). 공간적 연관구조를 고려한 총범죄 자료 분석, 응용통계연구, 제23권, 제2호, 335-344.
- [13] 최종근 (2007). 지구통계학, (주)시그마 프레스.
- [14] 최지은, 박만식 (2013). 다양한 관측네트워크에서 얻은 공간자료들을 활용한 계층모형 구축, 응용통계연구, 제26권, 294-305.
- [15] 최진호, 엄정섭 (2012). GOSAT으로 추적된 동북아시아 이산화탄소 유동방향의 계절별 비교평가, 한국공간정보학회지, 제20권, 1-13.
- [16] 최현우, 김현욱 (2012). 해수면 난류수 유동방향 탐지를 위한 지구통계학적 분석기법 적용, 한국지리정보학회지, 제9권, 168-178.
- [17] 환경부 (2010), 대기환경연보.
- [18] Banerjee, S., Gelfand, A. E., and Carlin, B. P. (2004). Hierarchical Modeling and Analysis for Spatial Data, Boca Raton: Chapman & Hall/CRC.
- [19] Cressie, N. A. C. (1993). Statistics for Spatial Data, New York: John Wiley & Sons.
- [20] Haning., Robert. P. (2003). Spatial data analysis, Cambridge: Cambridge University Press.
- [21] Kruminienė, I. (2006). Analysis of anisotropic variogram models for prediction of the curonian lagoon data, *journal of mathematical modelling and analysis*, Vol. 11, 73-86.
- [22] Loh J. M., and Stein M. L. (2004). Bootstrapping a spatial point process, *Statistica Sinica*, Vol 14, 69-101.

- [23] 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*, Vol. 10, 1247-1254
- [24] Kubota, T., and Tarumi, T. (2010). Simulation studies of geometric anisotropy detection methods for geostatistical data, *Journal of the Japanese Society of Computational Statistics*, Vol. 25, 13-24.
- [25] Kuan-Pin Lin., Zhi-He Long., and Bianling Ou. (2010). The size and power of bootstrap tests for spatial dependence in a linear regression model, *Society for Computational Economics*, Vol. 38, 153-171.
- [26] Budrikaite, L., and Ducinkas, K. (2005). Modeling of geometric anisotropic spatial variation, *journal of mathematical modelling and analysis*, 361-366.
- [27] Liansheng, T., William, R. S., Wayne, A. W., and Richard, F. G. (2006). A parametric spatial bootstrap, southern methodist university.
- [28] Schabenberger, O., and Gotway, C. A. (2005). Statistical methods for spatial data analysis, Boca Raton: Chapman & Hall/CRC.
- [29] Shin, K. I., Back, K. J., and Park, J. Mo. (2003). A study on the selection of variogram using spatial correlation, *The Korean Journal of Applied Statistics*, vol.10, no.3,835-844.
- [30] Lahiri S. N. (2003). Resampling methods for dependent data, Springer.

Abstract

On the Anisotropy Inherent In Spatial Data

Hyeji Go

Department of Statistics

The Graduate School

Sungshin Women's University

Spatial data compose of measurement values of our interest and corresponding location information. It is possible to predict the arbitrary values by finding spatial correlation inherent in the spatial data. When modeling spatial correlation, we consider relative location(distance and direction) between two arbitrary location points. In case that spatial correlation only depends on the distance, an isotropic model is considered, and when spatial correlation depends not only distance but also direction, an anisotropic model is used.

In this paper, a geometrically anisotropic model is mainly discussed. We also proposed a new model of various shapes of geometric anisotropy, where more than one direction is realized. We examine the performance of the proposed models along with the common isotropic ones in the simulation studies. Finally, we employ the parametric bootstrapping approach to

the real data analysis in order to obtain the interval estimation.

Keywords : Isotropic model, Anisotropic model, Spatial correlation, Covariance model, Maximum likelihood estimation method, Parametric bootstrap