



저작자표시-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

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

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 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

공간 모형의 계층화와
붓스트랩에 관한 연구

2013

성신여자대학교 대학원

통계학과

최지은

공간 모형의 계층화와 붓스트랩에 관한 연구

박만식교수지도

이 논문을 석사학위 논문으로 제출함

2013년 5월

성신여자대학교 대학원

통계학과

최지은

인준서

최지은의 석사학위 논문으로 인준함.

심사위원 _____인

심사위원 _____인

심사위원 _____인

성신여자대학교 대학원

논문개요

다양한 관측망에서 얻어지는 지리통계자료는 관측지점이 지도 상에 점으로 표현되고 그 지점에서만 관측된다. 지리통계자료를 분석하고 예측할 때 하나의 자료만 이용하는 것보다 유사한 패턴을 갖는 다른 관측망에서 얻어지는 자료를 이용해 분석하면 예측력을 향상시킬 수 있다. 이에 본 논문에서는 서로 다른 관측망에서 얻은 두 가지의 자료인 도시대기측정망의 이산화황(SO₂)과 지상기상관측망의 풍속자료를 이용하여 분석 및 예측하고 공간적 연관성을 파악할 수 있는 계층모형을 구축하였다. 선형회귀모형에 근간을 둔 크리깅(kriging) 결과와 계층모형 하에서의 결과를 여러가지 검증방법을 통해 비교하고 예측 지도를 구성했다. 또한, 역거리가중치법의 결과와 크리깅 결과를 비교하였다.

공간 선형회귀모형에서는 오차항에 공간적 연관성을 포함시키기 위하여 세미베리오그램(semi-variogram)을 추정한다. 세미베리오그램에는 공간적 연관구조를 나타내는 여러가지 모수들을 포함한다. 이 모수들을 추정하기 위해 가중최소제곱법, 잔차최대우도추정법, 일반최대우도추정법 등이 사용된다. 그러나 이 방법들을 사용하면 모수의 점추정은 가능하나 구간추정은 불가능하다. 이에 본 논문에서는 이들 모수의 신뢰구간을 구하기 위해 붓스트랩(Bootstrap)과 잭나이프(Jackknife) 방법을 적용하였다. 붓스트랩 방법으로는 전통적 붓스트랩 방법인 공분산함수를 고유값 분해하는 방법과 다변량 정규분포에서 잔차를 난수 생성하는 방법을 사용하였다. 그리고 공간자료에서 자료들 간의 연관성을 유지시켜주기 위해 사용하는 블록 붓스트랩 방법을 제

안하고, 블록의 개수를 정하는 기준에 대하여 설명하였다. 제안한 방법들을 비교하기 위해 모의실험을 수행하였고, 풍속자료를 이용하여 제안된 방법들을 적용해 분석하였다.

주요어: 계층모형, 크리깅, 역거리가중치법, 공간적 연관성, 붓스트랩, 잭나이프

목 차

I.	서론	1
II.	공간모형의 계층화	4
2.1	서론	4
2.2	관측망	8
2.2.1	지상기상관측망 (Surface observation network)	8
2.2.2	도시대기측정망 (Urban air pollution monitoring network)	10
2.3	공간자료들을 활용한 계층모형	11
2.3.1	역거리가중치법(Inverse distance weighting)	11
2.3.2	크리깅(Kriging)	12
2.3.3	계층모형(Hierarchical model)	16
2.4	사례연구	18
2.4.1	풍속자료의 모형 적합	19
2.4.2	이산화황 자료의 모형 적합	22
2.4.3	풍속자료를 이용한 이산화황 자료의 계층모형 적합	25
2.4.4	선형회귀모형의 비교	28
2.4.5	역거리가중치법을 이용한 이산화황 적합	32
2.5	결론	35

III. 공간모형의 붓스트랩	37
3.1 서론	37
3.2 공간자료의 붓스트랩과 잭나이프	39
3.2.1 공간자료의 비모수적 붓스트랩	39
3.2.2 공간자료의 모수적 붓스트랩	41
3.2.3 공간자료의 블록 붓스트랩	42
3.2.4 공간자료의 잭나이프 방법	47
3.3 모의실험	48
3.3.1 모의실험 방법	48
3.3.2 모의실험 결과	49
3.4 실증연구	58
3.5 결론	61
 IV. 결론	 62
 참고 문헌	 64
 Abstract	 67

그림 목 차

그림 1.	서로 다른 관측망(460개 지점(a)과 224개 지점(b).)	9
그림 2.	세미베리오그램	13
그림 3.	대표적인 세미베리오그램 모형.	15
그림 4.	구형모형에 기반한 풍속의 예측지도.	21
그림 5.	공간상관성을 고려하는 선형회귀모형의 이산화황 (SO ₂) 예측지도(구형모형).	24
그림 6.	관측된(예측된) 이산화황과 예측된 풍속의 산점도.	25
그림 7.	공간상관성을 고려하는 계층모형의 이산화황(SO ₂) 예측지도(구형모형).	28
그림 8.	교차검증방법(CV(1))에 의한 이산화황 예측값과 실제값의 산점도 및 95% 예측구간	31
그림 9.	역거리가중치법에 의한 이산화황(SO ₂) 예측지도 .	33
그림 10.	격자 공간 블록 붓스트랩 1.	42
그림 11.	격자 공간 블록 붓스트랩 2.	43
그림 12.	지도상 관측지점 블록 붓스트랩	44
그림 13.	경계지점 제외한 블록 붓스트랩.	45

표 목 차

표 1.	풍속자료의 선형회귀모형 모수 추정 결과.	20
표 2.	선형회귀모형의 이산화황(SO ₂)의 모수 추정 결과. . .	23
표 3.	계층모형의 이산화황(SO ₂)의 모수 추정 결과.	27
표 4.	선형회귀모형 선택 기준.	30
표 5.	크리깅과 역거리가중치법 모형 선택 기준.	34
표 6.	모의실험 시나리오.	48
표 7.	지수모형의 모의실험 결과(Scenario 1 ~ 3).	52
표 8.	지수모형의 모의실험 결과(Scenario 4 ~ 6).	53
표 9.	가우시안모형의 모의실험 결과(Scenario 1 ~ 3).	54
표 10.	가우시안모형의 모의실험 결과(Scenario 4 ~ 6).	55
표 11.	구형모형의 모의실험 결과(Scenario 1 ~ 3).	56
표 12.	구형모형의 모의실험 결과(Scenario 4 ~ 6).	57
표 13.	풍속자료의 모수추정 결과(REML).	59
표 14.	풍속자료의 모수추정 결과(ML).	60

제 1 장

서론

공간분석(spatial analysis)은 공간상에서 얻어진 자료를 이용하여 분석하는 통계적 방법이다. 공간 분석에서 사용되는 공간 자료는 공간적 위치와 측정값의 형태로 이루어져 있으며, 측정값이 얻어진 위치 사이의 거리에 따라 상관관계가 다르게 구성되어진다. 즉 공간 자료에서는 임의의 두 지점 사이가 가까우면 가까울수록 두 지점 사이의 관측값에 대한 상관도가 크게 되며, 임의의 두 지점 사이가 멀어질수록 두 지점사이의 관측값에 대한 상관도가 작아진다. 이와 같은 공간 자료는 관측값들이 서로 독립적이지 않기 때문에 상관관계를 고려한 분석이 필요하다.

공간통계학에서 사용되는 자료는 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. 격자자료는 관심있는 지역을 임의의 공간적인 영역으로 나누어 각 영역 내에서의 합산된 정보로 표현된 자료이다. 대표적인 격자자료는 서울시 각 구에서 발생하는 교통사고 사망자 수, 각 동별 버스이용객수 등

이 있다. 공간점 패턴자료는 지리통계자료와 격자자료와 다르게 위치 정보가 고정되어있지 않고 유동적이다. 대표적인 공간점 패턴자료는 특정 지역 내의 동물의 관측지점, 지진의 진원지의 위치 등이 있으며, 자료가 임의적으로 위치해있는지, 군집 또는 규칙적으로 위치해있는지를 밝히는데 분석 목적이 있다.

본 논문에서는 세 가지 공간 자료 중 지리통계자료를 가지고 분석에 이용한다. 지리통계자료를 이용하는 본 연구에서는 크게 두 가지 주제를 다룬다. 첫 번째는 서로 다른 관측망에서 얻어진 두 가지의 공간 자료를 이용하여 계층모형을 구축한다. 이를 확인하기 위해 도시대기측정망의 이산화황을 분석하고 예측하는데 지상기상관측망의 풍속자료를 이용한다. 그리고 여러가지 검증방법을 이용해 모형들을 비교하며, 선형회귀모형에 근간을 둔 크리깅 기법 뿐만 아니라 역거리가중치법을 통해 이산화황을 분석하고 비교한다.

두 번째는 공간적 연관성을 나타내는 여러가지 모수들에 대한 구간추정을 하기 위해 붓스트랩 방법과 잭나이프 방법을 사용한다. 세미베리오그램에는 공간적 연관성을 나타내는 모수들을 포함하는데, 일반적으로 이 모수들에 대한 점추정은 할 수 있지만 구간추정은 불가능하다. 이에 붓스트랩과 잭나이프를 사용해 모수들의 신뢰구간을 구하고자 한다. 이 때, 붓스트랩 방법으로는 공간자료에서 기존에 사용된 전통적인 붓스트랩 방법 뿐만 아니라 블록 붓스트랩 방법을 제안하고, 블록의 개수를 정하는 방법을 설명한다. 그리고 모의실험과 실증연구를 통해 본 연구에서 제안한 방법을 확인하였다.

본 논문의 순서는 다음과 같다. 제 2장에서는 서로 다른 관측망에서 얻어진 공간 자료를 이용하여 계층모형을 구축하였다. 이 때, 풍속 자료를 이용해 이산화황 자료의 계층모형을 적합하였으며, 선형회귀

모형을 비교하고 설명하였다. 제 3장에서는 공간모형에서의 붓스트랩 방법과 잭나이프 방법을 소개하고 모의실험을 통해 비교 설명한 뒤, 풍속자료를 이용해 본 연구에서 제안한 방법을 토대로 분석하였다. 끝으로 제 4장에서는 결론 및 앞으로의 연구방향에 관하여 논의하고자 한다.

본 논문의 제 2장의 일부는 최지은과 박만식 (2013)의 내용이 포함되어있음을 밝힌다.

제 2 장

공간모형의 계층화

2.1 서론

영국에서 시작된 산업혁명 이후로 지구는 화석에너지의 사용의 증가로 대기오염이 심각해지고 있다. 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 20세기 이후 산업화와 도시화가 급속도로 진행되면서 대기오염이 발생하고 있다. 우리나라는 대기오염물질을 가스상 물질인 이산화황(Sulfur Dioxide; SO₂), 일산화탄소(CO), 이산화질소(NO₂) 과 입자상물질을 포함한 총 61종의 물질로 정하며, 이 중 35종을 특정대기유해물질로 정하여 관리하고 있다(대기환경보전법).

대표적인 대기오염물질인 이산화황(SO₂)은 무색의 자극성이 있으며 물에 매우 잘 녹는 타지 않는 가스로, 황을 포함하는 연료인 석탄과 석유가 연소되거나 금속 제련공정, 기타 산업공정에서 발생한다. 이는 인체의 점막을 자극할 뿐만 아니라 인간이 이산화황 기체를 흡입하면 콧물, 담, 기침 등이 나오고 호흡곤란을 일으키게 된다. 일부의 경우에는 이산화황으로 인해 기관지염, 폐수종, 폐렴 등에 걸릴 수도

있다고 한다. 또한 이산화황은 질소산화물과 함께 산성비의 주요 원인으로 산성화에 영향을 미치며, 식물의 잎맥 손상, 성장저해 및 빌딩이나 기념물 등 각종 구조물의 부식을 촉진시키는 것으로 알려져 있으며 시정장애를 일으키는 미세먼지의 주요 원인물질이기도 하다(대기환경연보, 2007). 과거 이산화황으로 인한 큰 피해를 살펴보면 다음과 같다. 1930년 12월 벨기에 뮤즈 계곡 근처에 있는 대규모 공업단지에서 배출된 가스가 지면에 오래 머무르며 대기 중 이산화황 농도가 9.6ppm에서 38.4ppm까지 높아져 심장병과 급성폐렴으로 63명이 사망하고 급성 호흡기질환자가 급증하였다. 또한 동·식물에도 피해는 매우 컸다. 그 후, 1952년 12월 4일부터 12월 10일 사이에 발생한 런던 스모그사건은 석탄이 연소되면서 발생하는 이산화황의 배출로 스모그가 발생하여 런던 시민의 약 12,000명이 만성 폐질환과 호흡장애로 사망했다. 이 이후로 세계적으로 대기오염이 중요한 문제로 대두되었다. 그러나 급격한 산업화와 도시화로 인하여 대기오염의 범위는 더욱 넓어지고 심각해지고 있다. 이와 같이 대기와 인체에 심각한 영향을 끼치는 이산화황에 대한 관심과 연구가 필요하다.

최근의 공간통계를 이용한 연구를 보면, 김규곤과 최승배(2000)는 공간상관이 존재하는 환경자료를 이용하여 일반 통계 분석법과 공간 통계 분석법을 비교하였고, 조재영 등(2001)은 공간통계학에서 다루는 자료에 대하여 일반통계와 공간통계 방법에 대해 예측오차제곱합(predicted residual sums of squares; PRESS)을 사용하여 예측성능을 비교하였다. 허태영 등(2004)은 남한의 강수량 자료의 공간분석을 하는데 가중최소방법, 최대우도방법, 잔차최대우도방법을 사용하여 각 방법을 비교하였고, 조홍래와 정종철(2006)은 강우관측소의 강우량에 대하여 공간보간(spatial interpolation)을 실시하여 일반 크리깅(ordinary

kriging)을 실시하여 정확성을 검증하였다. 신기일 등 (2007)은 계층적 베이지안 추정법과 공간통계량을 활용해 소지역 추정법을 평가하였다. 정지용 등 (2008)은 강수량 자료를 이용해 최소제곱법과 최대우도 추정방법으로 크리깅을 하였으며 신뢰성 검증을 위해 교차검증방법을 사용하였고, 박노욱과 장동호 (2008)는 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. 정승환 등 (2010)은 80m 높이에서의 보정된 풍속자료를 가지고 크리깅 기법을 이용해 남한의 풍속 예측지도를 구현하였다. 김병선 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. 김동휘 등 (2010)은 인천 송도지역 지층분포의 추정을 위해 크리깅과 역거리가중치법을 적용하였으며 크리깅의 추정결과가 역거리가중치법에 비해 신뢰성이 높은 것을 확인하였고, 주재희와 황인조 (2012)는 강하분진의 침적 특성 파악을 하기 위해 역거리가중치법을 사용했다.

한진석 등 (2006)에 따르면 이산화황이 풍속, 기류 등에 의해 보통 수 백에서 수 천 km까지 이동이 가능하다고 밝혔다. 즉 이산화황이 거의 발생하지 않는 지역에서도 풍속, 기류 등에 의해 이산화황의 영향을 받을 수 있게 된다. 결국 이산화황을 분석함에 있어서 이산화황만을 연구하는 것보다 풍속, 기류 등의 외부요인들을 이용하여 이산화

황을 분석하는 것이 바람직하다. 또한 풍속, 기류 등의 외부 요인들이 하나의 공간적 상관성을 내포하는 자료가 될 수 있음을 고려한다.

본 장에서는 서로 다른 관측망에서 얻어진 두 공간 자료를 이용하여 공간적 분석을 하고자 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 떨어질 수 있다. 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 적용하여 분석하기 위해 다른 관측망에서 얻어진 이산화황자료와 풍속자료를 이용한다. 이산화황이 풍속, 기류 등 여러 가지 요인들에 의해 이동하는데, 여러 요인 중 풍속자료를 얻을 수 있었으며, 가장 간단한 계층모형을 구축하기 위하여 풍속자료만을 사용한다. 우리나라 전역에 걸쳐 측정되는 풍속자료를 이용하여 주로 도시지역에서만 관측되는 이산화황을 예측하는 데 예측력을 높이도록 모형을 구축하였다. 풍속과 이산화황의 관측망이 같지 않아 이산화황을 예측하는 데 풍속자료를 바로 이용할 수 없으므로 계층모형을 고려한다.

또한 모형을 추정하고 적합하는데 잔차최대우도추정법(restricted maximum likelihood estimation; REML)과 최대우도추정법(maximum likelihood estimation; ML)을 사용하고 오차항의 공분산 함수에는 지수모형(exponential model), 가우시안모형(Gaussian model), 구형모형(spherical model)을 가정하였다. 그리고 크리깅 기법 뿐만 아니라 역거리가중

치법을 이용해 이산화황을 분석하였으며, 역거리가중치법의 결과와 크리깅의 결과를 비교한다.

본 장의 순서는 다음과 같다. 2절에서는 사례연구를 위한 두 가지의 자료의 관측망을 설명하였다. 3절에서는 공간분석을 위한 기본적인 공간통계학의 개념을 설명하고 다양한 관측망에서 얻은 자료를 이용하여 계층모형을 구축하는 방법에 대하여 기술하였다. 4절에서는 풍속자료와 이산화황 자료를 이용한 분석결과에 대해 기술하였으며, 이산화황만을 이용한 결과와 풍속자료를 이용한 계층모형의 결과를 비교하였다. 또한 이산화황을 역거리가중치법을 이용한 결과와 크리깅을 이용한 결과를 비교하였다. 5절에서는 결론 및 향후 연구 과제에 대하여 논하였다.

2.2 관측망

본 절에서는 계층모형의 구축을 위해 사용한 두 가지 관측망(mon-itoring network)들에 대하여 소개하고자 한다. 본 논문에서는 두 가지 관측망을 고려하였다. 첫 번째는 지상기상관측망에서 측정된 풍속 및 풍향자료이고, 두 번째는 도시대기측정망에서 측정된 이산화황이다.

2.2.1 지상기상관측망 (Surface observation network)

본 연구에서 사용한 자료 중 첫 번째는 2003년부터 2007년까지 5년 평균 풍속 및 풍향자료이다. 이 자료는 남한의 주요 도서지역 제외한 460개 기상관측지점(그림 1(a))에서 측정되었으며 기상청(<http://www.kma.go.kr>)에서 얻은 자료이다. 기상청은 지상기상관측망

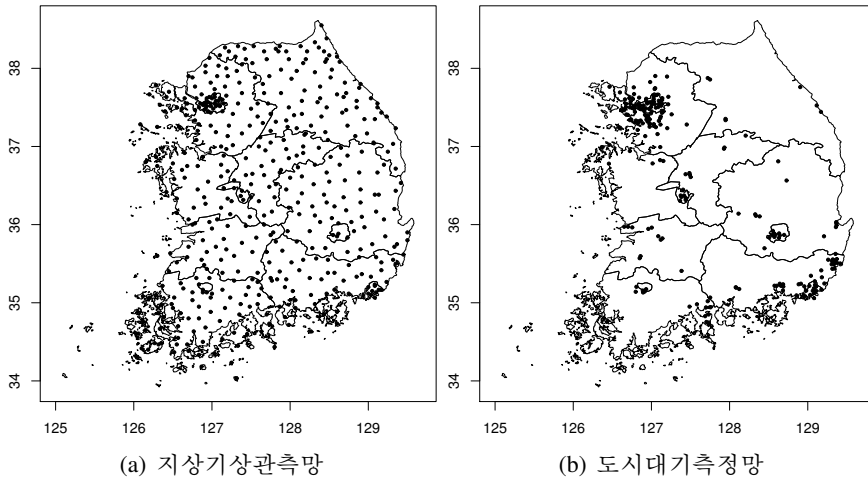


그림 1: 서로 다른 관측망(460개 지점(a)과 224개 지점(b).)

에서 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 풍력발전에 활용하려면 50m - 100m의 높이에서의 풍속이 사용되는데, 본 연구에서는 변환된 풍속이 아닌 관측된 위치인 10m 높이의 풍속자료를 사용하고자 한다.

그리고 지상기상관측망에서는 풍속 뿐만 아니라 주풍향(main wind direction)도 관측하였다. 또한, 모형에 각 지점의 고도(altitude)를 포함하기 위해 정승환 등 (2010)이 Google Earth 를 이용하여 킬로미터(km) 단위로 구한 고도를 이용한다. 위도와 경도는 지구의 곡면의 정보가 포함되었기 때문에 위치정보를 보정하여 고도와 마찬가지로 킬로미터 단위를 사용한다.

2.2.2 도시대기측정망

(Urban air pollution monitoring network)

본 연구에서 사용한 자료 중 두 번째는 2005년부터 2007년까지 3년 평균 이산화황(SO₂) 자료이다. 이 자료는 남한의 주요 도서지역을 제외한 224개 대기오염관측지점(그림 1(b))에서 측정되었으며 국립환경과학원(<http://www.nier.go.kr>)에서 얻은 자료이다. 도시대기측정망은 도시지역의 평균 대기질 농도를 파악하여 환경 기준 달성 여부를 판정한다. 그러나 도시대기측정망의 측정지점은 대부분 도시지역으로 한정되어 있어서 농촌지역에서는 측정되지 않는 한계가 있다. 이산화황은 자외선형광법(Pulse U.V Fluorescence Method)을 이용해 측정되며, 자외선형광법은 비교적 짧은 파장 영역의 자외선에 의해 나타나는 이산화황 분자로부터 발생하는 형광강도를 측정하여 이산화황의 농도를 연속적으로 측정하는 방법이다(대기환경연보). 풍속자료와 마찬가지로 대기오염자료에서도 모형에 각 지점의 고도를 포함하기 위해 Google Earth를 이용하여 킬로미터 단위의 고도를 구하였다. 대기오염자료의 위도와 경도 또한 지구의 곡면의 정보가 포함되었기 때문에 위치정보를 보정하여 고도와 마찬가지로 킬로미터 단위를 사용한다.

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

2.3 공간자료들을 활용한 계층모형

2.3.1 역거리가중치법(Inverse distance weighting)

역거리가중치법(Inverse distance weighting, IDW)은 Shepard(1968)에 의해 처음 적용되었으며, 공간적으로 가까운 지점 사이의 값은 유사성을 갖게 되고 두 지점 사이의 거리가 멀어질수록 유사성이 상대적으로 감소한다는 것에 기초한다. 이에 두 지점 사이의 거리가 가까울수록 큰 가중치가 적용되고, 반대로 두 지점 사이의 거리가 멀어질수록 가중되는 값의 영향력은 줄어들게 된다. 역거리가중치법은 식 (2.1)으로 표현 가능하다.

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i). \quad (2.1)$$

여기서, $\hat{Z}(\mathbf{s}_0)$ 는 위치 \mathbf{s}_0 에서 예측하고자 하는 값이고, n 은 관측지점의 수, $Z(\mathbf{s}_i)$ 는 관측지점(\mathbf{s}_i)에서의 관측값이며 λ_i 는 \mathbf{s}_i 에서의 가중치이며 다음과 같이 정의된다. 임의의 $i = 1, \dots, n$ 에 대하여

$$\lambda_i = \frac{w(\mathbf{s}_i)}{\sum_{i=1}^n w(\mathbf{s}_i)}, \quad \sum_{i=1}^n \lambda_i = 1.$$

이다. 여기서, $w(\mathbf{s}_i)$ 는 임의의 지점(\mathbf{s}_i)와 예측지점(\mathbf{s}_0) 사이의 거리가 증가할수록 줄어들게 되며, $p > 0$ 에 대하여 다음과 같이 정의한다.

$$w(\mathbf{s}_i) = \frac{1}{d(\mathbf{s}_i, \mathbf{s}_0)^p} = \frac{1}{\|\mathbf{s}_i - \mathbf{s}_0\|^p}. \quad (2.2)$$

식 (2.2)에서 $d(\mathbf{s}_i, \mathbf{s}_0)$ 는 관측지점(\mathbf{s}_i)과 예측하고자하는 지점(\mathbf{s}_0)사이의 거리를 나타낸다. 또한 p 는 거리에 따른 가중치 변화의 정도를 나타내며 양의 실수이다. 이 때, $0 < p < 1$ 이면, 전체적으로 보간된 거리가 좁고 날카로우며, $p > 1$ 이면, 보간된 거리가 넓고 부드럽게 퍼지게 된다.

2.3.2 크리깅(Kriging)

지리통계자료는 다음과 같이 표현 가능하다.

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

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 표현 가능하다.

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.3)$$

여기서, $\mathbf{X}\boldsymbol{\beta}$ 는 평균함수(mean function)로서, 전반적 추세 또는 경향을 나타내는 위치간의 특성을 의미한다. 일반적으로 $(n \times p)$ 인 행렬 \mathbf{X} 는 그 지점에서 관측되는 그 지점의 위치정보와 외부요인들을 포함하고 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ 는 회귀계수들로 구성된 모수벡터이다. 식 (2.3)의 $\boldsymbol{\varepsilon} \equiv (\varepsilon(\mathbf{s}_1), \varepsilon(\mathbf{s}_2), \dots, \varepsilon(\mathbf{s}_n))^T$ 는 위치에 내재된 고유한 변동을 의미하며 2차 정상성(second-order stationarity)을 만족하는 확률변수로 가정한다. 즉,

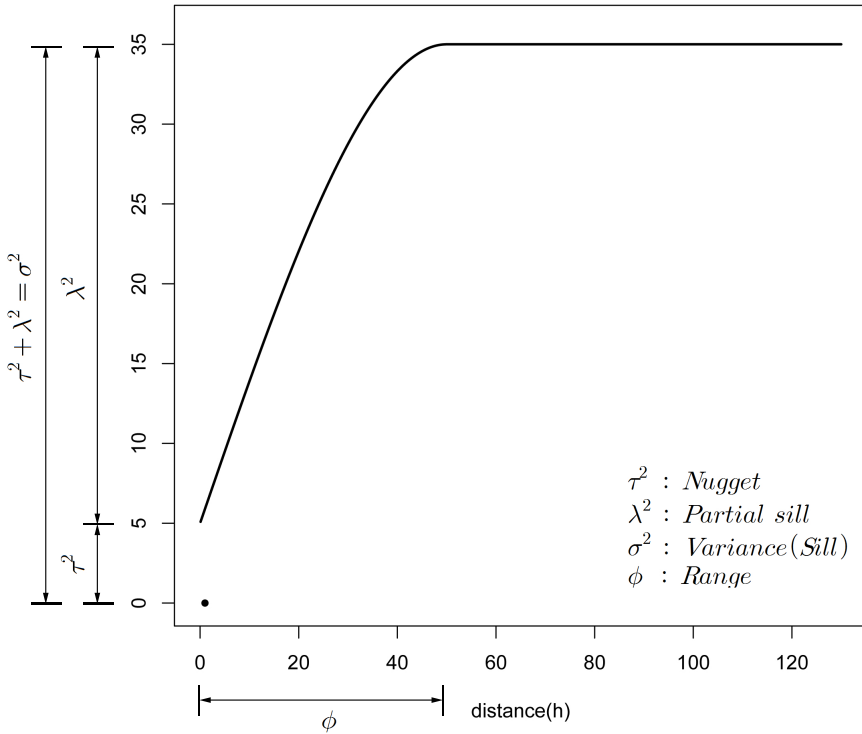


그림 2: 세미베리오그램

오차항들의 평균은 0이고, 오차항들은 유한한 분산을 가지며 오차항 사이의 공분산은 두 지점사이의 거리에 의존함을 가정한다. 공간적 연관구조를 의미하는 오차항의 분산-공분산 행렬(variance-covariance matrix)에 공간적인 연관성을 포함시켜 모형화하기 위해서는 세미베리오그램(semi-variogram)을 추정해야한다. 오차항에 대한 세미베리오그램은 다음과 같이 정의된다. 임의의 $i, j = 1, \dots, n$ 에 대하여

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2} E[\varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_j)]^2 \quad (2.4)$$

이다. 세미베리오그램(그림 2)에는 공간적 연관구조를 나타내는 세 가지 모수($\theta = (\tau^2, \lambda^2, \phi)$)를 포함한다. 뭉치(nugget)라 불리는 τ^2 는 측정 오차로 인해 발생하는 분산의 일부분이고, 부분 문턱(partial sill)이라 불리는 λ^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 ϕ 는 범위(range)로서, 문턱까지의 거리이다. 식 (2.4)을 근간으로, 자료로부터 계산된 세미베리오그램을 이론적(theoretical) 세미베리오그램에 적합하여 위에서 언급한 세 가지 모수들(θ)을 추정하게 된다.

대표적인 이론적 세미베리오그램은 다음과 같다.

A. 지수모형(exponential semi-variogram model)

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

B. 가우시안모형(Gaussian semi-variogram model)

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

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 지점

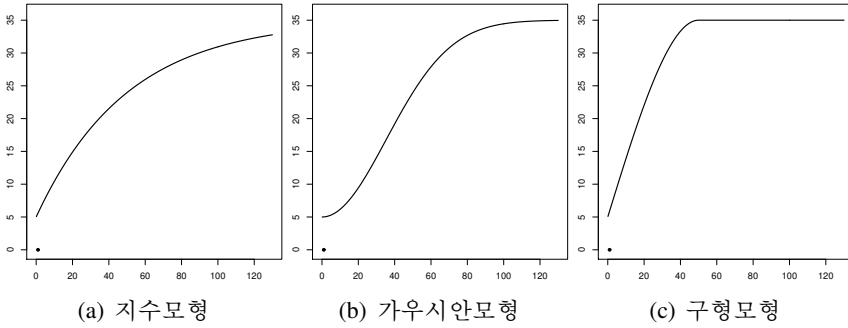


그림 3: 대표적인 세미베리오그램 모형.

이 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 가까울수록 연관성이 강하고, 두 지점 사이의 거리가 멀어질수록 연관성이 약하거나 상관이 없다고 가정한다. 그러나 각 모형은 구조적인 차이로 인해 거리에 따라 줄어드는 상관계수의 경향은 다르다. 대표적인 이론적 세미베리오그램의 형태를 보면 그림 3과 같다. 그림 3의 세미베리오그램은 멩치(nugget, τ^2)가 5, 부분 문턱(partial sill, λ^2)이 30, 범위(range, ϕ)가 50인 지수모형, 가우시안모형, 구형모형이다.

공간 자료를 이용해 경험적(empirical) 세미베리오그램은 다음과 같이 구해진다.

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2\|N(\mathbf{h})\|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} [\varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_j)]^2.$$

여기서, $N(\mathbf{h})$ 는 $\|(\mathbf{s}_i - \mathbf{s}_j)\| = \mathbf{h}$ 를 갖는 전체 쌍의 집합이고 $\|N(\mathbf{h})\|$ 는 $N(\mathbf{h})$ 에 속하는 쌍의 수이다. 여기에서 각 쌍들의 거리를 이용해 거리에 대한 구간의 개수와 구간의 길이를 결정한다. 각 쌍의 거리의 최솟값이 0이고 최댓값이 \mathbf{h}_M 이고 구간의 개수를 a 개라고 한다면, 구간은

$I_1 = (0, \mathbf{h}_1], I_2 = (\mathbf{h}_1, \mathbf{h}_2], \dots, I_a = (\mathbf{h}_{M-1}, \mathbf{h}_M]$ 으로 정의할 수 있다. 여기에서 $0 < \mathbf{h}_1 < \mathbf{h}_2 < \dots < \mathbf{h}_M$ 이다. 이렇게 구간을 나누는 뒤에 각 구간에 속하는 쌍들을 결정하고 각 구간의 대표값을 결정해 경험적세미베리오그램을 구하게 된다.

본 논문에서는 이론적 세미베리오그램으로 지수모형, 가우시안 모형, 구형모형을 고려하였다. 또한, 선형회귀모형으로 표현되는 식 (2.3)과 공간적 연관구조를 의미하는 분산-공분산 행렬을 토대로 우도함수(likelihood function)를 최대화하는 추정방법(잔차최대우도추정법과 일반최대우도추정법)을 고려하고자 한다. 추정된 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 (2008)을 참조하기 바란다.

2.3.3 계층모형(Hierarchical model)

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 표현 가능하다.

$$\mathbf{z}_a = \mathbf{X}_a \boldsymbol{\beta}_a + \boldsymbol{\varepsilon}_a. \quad (2.5)$$

여기서, \mathbf{X}_a 는 $(n_a \times p)$ 크기의 독립변수 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 표현 가능하다.

$$\mathbf{z}_b = \mathbf{X}_b \boldsymbol{\beta}_b + \boldsymbol{\varepsilon}_b. \quad (2.6)$$

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 (2.5)과 식 (2.6)의 구성성분에 대한 가정은 식 (2.3)과 동일하다. 즉, 각 모형의 오차항에 대한 분산-공분산행렬은 공간 상관구조를 내포하는 모수들 $(\boldsymbol{\theta} = (\tau^2, \lambda^2, \phi))$ 에 의해 결정된다.

자료 \mathbf{z}_b 를 분석함에 있어서 주어진 평균함수와 더불어 자료 \mathbf{z}_a 로 부터 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 표현 가능하다.

$$\mathbf{z}_b = \mathbf{X}_b \boldsymbol{\beta}_c + \boldsymbol{\delta} \hat{\mathbf{z}}_{ba} + \boldsymbol{\varepsilon}_c. \quad (2.7)$$

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 표현 가능하다.

$$\begin{aligned} \mathbf{z}_a &\sim N_{n_a}(\mathbf{X}_a \boldsymbol{\beta}_a, \boldsymbol{\Sigma}_a), \\ \hat{\mathbf{z}}_{ba} &= \mathbf{X}_b \hat{\boldsymbol{\beta}}_a + \hat{\boldsymbol{\varepsilon}}_{ba}, \\ \mathbf{z}_b &\sim N_{n_b}(\mathbf{X}_b \boldsymbol{\beta}_c + \boldsymbol{\delta} \hat{\mathbf{z}}_{ba}, \boldsymbol{\Sigma}_c). \end{aligned} \quad (2.8)$$

여기서, $\boldsymbol{\Sigma}_a$ 와 $\boldsymbol{\Sigma}_c$ 는 각각 오차항의 분산-공분산행렬이고 $\hat{\boldsymbol{\varepsilon}}_{ba}$ 는 다음과

같다.

$$\hat{\boldsymbol{\varepsilon}}_{ba} \equiv (\hat{\boldsymbol{\varepsilon}}(\mathbf{s}_1^a), \hat{\boldsymbol{\varepsilon}}(\mathbf{s}_2^a), \dots, \hat{\boldsymbol{\varepsilon}}(\mathbf{s}_{n_b}^a))^T.$$

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

M.1 공간상관성을 고려하지 않는 전통적 선형회귀모형(식 (2.6)에서

$$\boldsymbol{\Sigma}_b = \sigma_b^2 \mathbf{I}_{n_b})$$

M.2 공간상관성을 고려하지 않는 계층모형(식 (2.8)에서 $\boldsymbol{\Sigma}_c = \sigma_c^2 \mathbf{I}_{n_b}$)

M.3 공간상관성을 고려하는 선형회귀모형(식 (2.6))

M.4 공간상관성을 고려하는 계층모형(식 (2.8))

2.4 사례연구

본 연구에서는 2.3절에서 제시한 모형구축방법을 실제 자료에 적용한 결과를 소개하고자 한다. 사용한 자료는 2.2절에서 제시한 풍속 자료와 대기오염자료 중 하나인 이산화황이다. 여기서 풍속과 이산화황의 위치정보는 위도, 경도, 고도를 포함한다($\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3)$). 분석의 최종결과인 풍속자료와 이산화황자료의 예측지도(prediction map)을 구현하기 위해 935개의 일정한 거리를 가지는 격자지점들을 고려하였고, 이들 지점에 대한 위치정보 중 고도는 정승환 등 (2010)이 구한 값을 이용하였다. 사례연구를 위해 R(R Development Core Team, 2012) 프로그램을 사용하였다.

2.4.1 풍속자료의 모형 적합

평균풍속 자료를 분석함에 있어 정승환 등 (2010)을 이용하였다. 정승환 등 (2010)에서는 80m에서의 변환된 풍속자료를 이용하였지만, 본 논문에서는 10m 높이의 풍속자료를 사용했다. 로그변환한 풍속과 경도(s_{11}^a), 로그변환한 풍속과 위도(s_{12}^a) 사이의 관계를 살펴본 결과, 2차형의 모형식으로 표현 가능하며, 로그변환한 풍속과 고도(s_{13}^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 비해 크다는 것을 확인하였으며 주풍향에 의한 평균 풍속의 차이를 평균함수 모형 포함시키기 위해 북풍을 기준범주(reference group)로 하여 가변수 처리하였다. 그림 1(a)에 나타난 460개의 풍속 관측지점들로부터 얻은 평균풍속에 대한 공간적 선형회귀모형은 식 (2.9)으로 표현되며 평균함수는 다음과 같다.

$$\begin{aligned} \mathbf{X}_a \boldsymbol{\beta}_a = & \beta_1^a + s_{11}^a \beta_2^a + s_{12}^a \beta_3^a + (s_{11}^a)^2 \beta_4^a + (s_{12}^a)^2 \beta_5^a + s_{11}^a s_{12}^a \beta_6^a + s_{13}^a \beta_7^a \\ & + I_{NE} \beta_8^a + I_E \beta_9^a + I_{SE} \beta_{10}^a + I_{SW} \beta_{11}^a + I_W \beta_{12}^a + I_{NW} \beta_{13}^a. \end{aligned} \quad (2.9)$$

여기서, I 는 지시함수로서, 예를 들면 임의의 지점, \mathbf{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.

표 1: 풍속자료의 선형회귀모형 모수 추정 결과.

	Classical linear model			Spatial linear model					
	Least Squares			REML(Spherical)			ML(Spherical)		
	Est. [§]	95% C.I.		Est. [§]	95% C.I.		Est. [§]	95% C.I.	
β_1^a	-2.72 ²	-9.25 ²	3.80 ²	-8.82 ²	-2.95 ¹	1.19 ¹	-8.13 ²	-2.62 ¹	9.99 ²
β_2^a	-6.83 ⁴	-1.08 ³	-2.81 ⁴	-9.55 ⁵	-1.30 ³	1.11 ³	-1.90 ⁴	-1.25 ³	8.77 ⁴
β_3^a	-9.24 ⁴	-1.18 ³	-6.65 ⁴	-7.36 ⁴	-1.57 ³	1.04 ⁴	-7.86 ⁴	-1.52 ³	-4.49 ⁵
β_4^a	2.79 ⁵	2.29 ⁵	3.30 ⁵	3.24 ⁵	1.96 ⁵	4.52 ⁵	3.18 ⁵	2.03 ⁵	4.33 ⁵
β_5^a	7.82 ⁶	5.57 ⁶	1.00 ⁵	1.23 ⁵	6.03 ⁶	1.86 ⁵	1.19 ⁵	6.32 ⁶	1.76 ⁵
β_6^a	-3.39 ⁶	-6.84 ⁶	4.35 ⁸	-2.32 ⁶	-1.21 ⁵	7.48 ⁶	-2.50 ⁶	-1.12 ⁵	6.24 ⁶
β_7^a	5.72 ¹	4.19 ¹	7.25 ¹	7.47 ¹	5.97 ¹	8.97 ¹	7.33 ¹	5.85 ¹	8.82 ¹
β_8^a	2.87 ¹	1.70 ¹	4.03 ¹	2.32 ¹	1.25 ¹	3.39 ¹	2.36 ¹	1.30 ¹	3.42 ¹
β_9^a	1.05 ¹	-6.37 ²	2.73 ¹	1.10 ¹	-4.04 ²	2.61 ¹	1.10 ¹	-3.91 ²	2.60 ¹
β_{10}^a	1.86 ¹	5.02 ²	3.23 ¹	1.63 ¹	3.96 ²	2.87 ¹	1.62 ¹	3.96 ²	2.85 ¹
β_{11}^a	2.04 ¹	9.38 ²	3.15 ¹	1.20 ¹	1.92 ²	2.22 ¹	1.23 ¹	2.33 ²	2.24 ¹
β_{12}^a	3.02 ¹	2.15 ¹	3.90 ¹	2.22 ¹	1.41 ¹	3.03 ¹	2.26 ¹	1.46 ¹	3.06 ¹
β_{13}^a	2.68 ¹	1.80 ¹	3.56 ¹	2.06 ¹	1.24 ¹	2.87 ¹	2.09 ¹	1.28 ¹	2.89 ¹
τ^2				0.06			0.06		
λ^2				0.06			0.05		
ϕ				118.50			113.86		
log.lik.	-105.25			-78.43			-81.30		

Est., 추정값; log.lik., 로그우도값; $a^b \equiv a \times 10^{-b}$

식 (2.3)로 표현되는 선형회귀모형에 대해 여러 추정방법을 적용한 결과는 표 1에 나타나 있다. 잔차최대우도추정방법 하에서는 지수모형, 가우시안모형, 구형모형을 공분산함수로 고려한 공간선형모형의 로그우도값이 각각 -79.48, -80.45, -78.43으로 구형모형의 로그우도값이 가장 크게 나타났다. 일반최대우도추정방법 하에서도 세 가지 공분산모형들의 로그우도값이 각각 -89.06, -83.02, -81.30으로 구형모형을 고려한 경우가 가장 큰 로그우도값을 보였다.

또한, 공간선형모형(-78.43, -81.30)과 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 설명

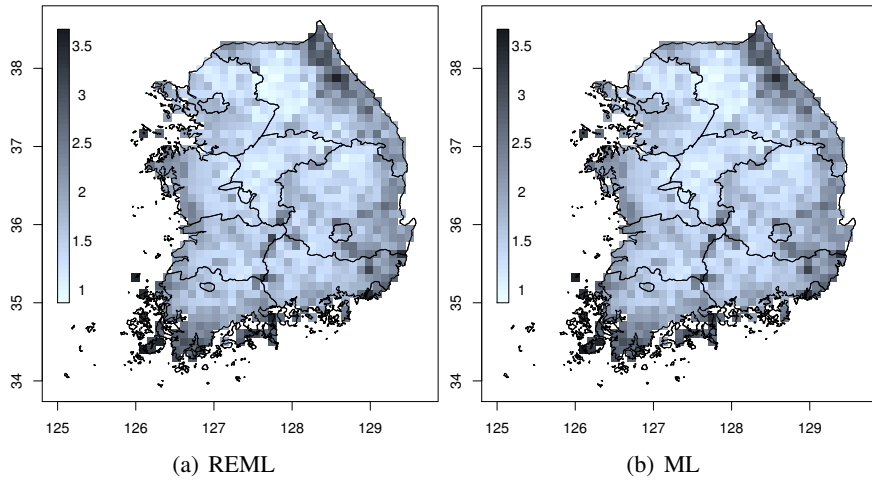


그림 4: 구형모형에 기반한 풍속의 예측지도.

하는 모수 (θ) 에 대한 추정값은 잔차최대우도추정방법과 일반최대우도추정방법이 거의 유사한 값으로 나타났다. 로그우도값으로 두 가지 공간 모형을 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 그림 4와 같다. 그림 4(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.

2.4.2 이산화황 자료의 모형 적합

2005년부터 2007년까지의 3년간 이산화황 자료를 분석하기 위해 로그변환한 이산화황의 평균자료를 사용하였다. 이산화황에 관한 선형회귀모형을 적합하고 오차항의 공분산모형을 추정하기 위해 풍속 자료에서 사용한 평균함수를 적용한다. 그림 1(a)에 나타난 460개의 풍속 관측지점에서는 평균풍속과 더불어 주풍향에 관한 정보도 포함되어 있다. 그러나 그림 1(b)에 나타난 224개의 이산화황 관측지점에서는 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 것을 확인할 수 있었다. 이에 주풍향에 의한 평균 이산화황의 차이를 평균함수에 포함시킬 필요가 있으며 이를 위해 북풍을 기준범주로 하여 가변수 처리하였다. 따라서 이산화황에 관한 선형회귀모형의 평균함수는 식 (2.7)과 유사한 형태를 이루며 다음과

표 2: 선형회귀모형의 이산화황(SO₂)의 모수 추정 결과.

	Classical linear model			Spatial linear model					
	Least Squares			REML(Spherical)			ML(Spherical)		
	Est. [§]	95% C.I.		Est. [§]	95% C.I.		Est. [§]	95% C.I.	
β_1^b	-5.27 ⁰	-5.42 ⁰	-5.12 ⁰	-5.23 ⁰	-5.45 ⁰	-5.01 ⁰	-5.24 ⁰	-5.45 ⁰	-5.03 ⁰
β_2^b	1.18 ⁴	-6.69 ⁴	9.04 ⁴	5.57 ⁴	-6.62 ⁴	1.78 ³	5.34 ⁴	-5.88 ⁴	1.66 ³
β_3^b	3.40 ⁴	-2.02 ⁴	8.83 ⁴	3.24 ⁴	-5.77 ⁴	1.23 ³	3.09 ⁴	-5.20 ⁴	1.14 ³
β_4^b	3.45 ⁶	-6.17 ⁶	1.31 ⁵	-9.07 ⁷	-1.74 ⁵	1.56 ⁵	-8.47 ⁷	-1.60 ⁵	1.43 ⁵
β_5^b	5.25 ⁶	-2.24 ⁷	1.07 ⁵	4.37 ⁶	-5.10 ⁶	1.38 ⁵	4.55 ⁶	-4.14 ⁶	1.32 ⁵
β_6^b	-5.37 ⁶	-1.19 ⁵	1.13 ⁶	-3.77 ⁶	-1.44 ⁵	6.84 ⁶	-3.92 ⁶	-1.37 ⁵	5.83 ⁶
β_7^b	3.89 ¹	-1.15 ⁰	1.93 ⁰	-1.98 ¹	-1.81 ⁰	1.41 ⁰	-8.71 ²	-1.63 ⁰	1.46 ⁰
β_8^b	-1.57 ²	-1.52 ¹	1.21 ¹	7.75 ²	-5.49 ²	2.10 ¹	7.60 ²	-5.23 ²	2.04 ¹
β_9^b	1.33 ¹	-9.17 ²	3.57 ¹	3.90 ²	-1.83 ¹	2.61 ¹	4.46 ²	-1.70 ¹	2.59 ¹
β_{10}^b	-1.33 ¹	-3.59 ¹	9.31 ²	2.07 ²	-2.17 ¹	2.59 ¹	5.44 ³	-2.23 ¹	2.34 ¹
β_{11}^b	-5.57 ²	-2.02 ¹	9.02 ²	-5.23 ²	-2.14 ¹	1.09 ¹	-5.29 ²	-2.09 ¹	1.03 ¹
β_{12}^b	3.29 ²	-7.25 ²	1.38 ¹	-4.27 ²	-1.52 ¹	6.63 ²	-3.76 ²	-1.43 ¹	6.77 ²
β_{13}^b	-7.94 ²	-1.93 ¹	3.41 ²	-7.34 ²	-1.91 ¹	4.46 ²	-7.30 ²	-1.87 ¹	4.10 ²
τ^2				0.03			0.03		
λ^2				0.08			0.06		
ϕ				40.00			38.82		
log.lik.	-29.75			-0.08			0.51		

Est., 추정값; log.lik., 로그우도값; $a^b \equiv a \times 10^{-b}$

같이 표현된다.

$$\begin{aligned} \mathbf{X}_b \beta_b = & \beta_1^b + s_1^b \beta_2^b + s_2^b \beta_3^b + (s_1^b)^2 \beta_4^b + (s_2^b)^2 \beta_5^b + s_1^b s_2^b \beta_6^b + s_3^b \beta_7^b \\ & + I_{NE} \beta_8^b + I_E \beta_9^b + I_{SE} \beta_{10}^b + I_{SW} \beta_{11}^b + I_S \beta_{12}^b + I_{NW} \beta_{13}^b. \end{aligned}$$

이산화황에 관한 선형회귀모형의 회귀계수들과 오차항의 공간 상관구조를 의미하는 모수를 추정한 결과는 표 2에 나타나 있다. 풍속자료의 분석결과와 마찬가지로, 2.3장에서 소개한 세가지 공분산함수들 중에서 로그우도값을 비교해본 결과 구형모형이 오차항의 공간 상관구조를 가장 잘 설명하는 것으로 나타났다. 로그우도값을 살펴보면, 잔차우도추정법에서의 지수모형, 가우시안모형, 구형모형의 로그

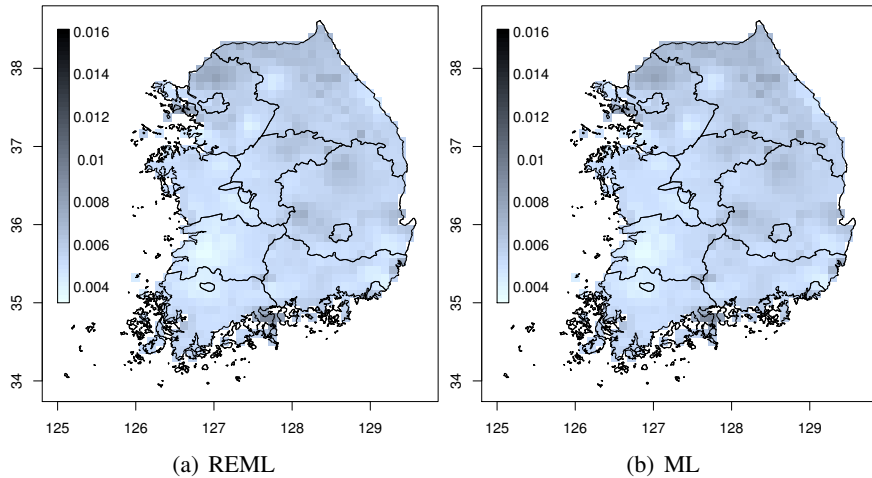


그림 5: 공간상관성을 고려하는 선형회귀모형의 이산화황(SO₂) 예측 지도(구형모형).

우도값은 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. 회귀계수의 추정에 있어서 우도를 바탕으로 한 두 가지 추정방법들 간에는 추정값에 대한 큰 차이는 보이지 않으며 오차항의 분산- 공분산 행렬에 포함된 모수들의 추정결과를 살펴보더라도 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 그림 5에 나타나 있으며 이는 구형모형을 기반으로 적합된 모형들로부터 예측한 후 원 측정단위로 변

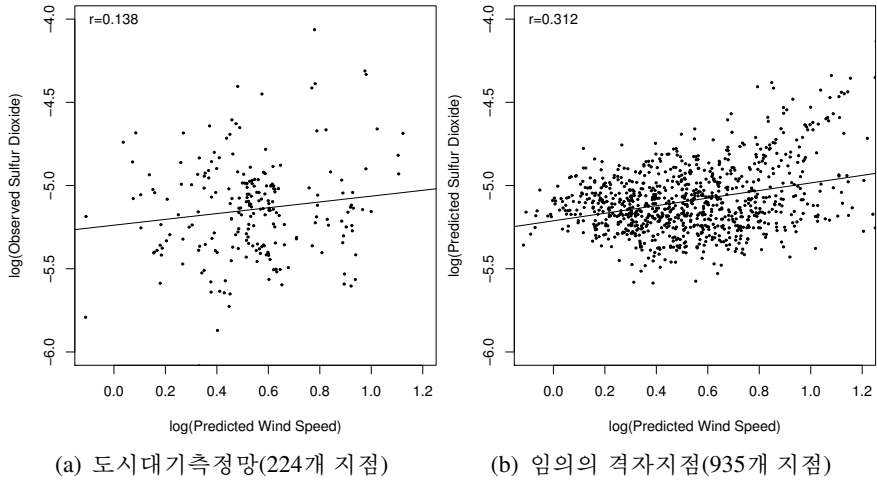


그림 6: 관측된(예측된) 이산화황과 예측된 풍속의 산점도.

환한 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.

2.4.3 풍속자료를 이용한 이산화황 자료의 계층모형 적합

본 절에서는 식 (2.8)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. 2.3.2절에서 설명한 이산화황의 크리깅에서 마찬가지로 3년간 이산화황의 평균 자료를 로그변환 하였다. 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(그림 6). 즉, 풍속이 증가할수록 이산화황의 농도 역시 증가하는 경향이 있다고 판단된다. 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 모형의 설명력을 높일 수 있으므로 공간선형회귀모형의 평균함수를 다음과 같이 고려하였다.

$$\begin{aligned} \mathbf{X}_b\beta_c + \widehat{\delta z}_{ba} = & \beta_1^c + s_1^b\beta_2^c + s_2^b\beta_3^c + (s_1^b)^2\beta_4^c + (s_2^b)^2\beta_5^c + s_1^b s_2^b\beta_6^c + s_3^b\beta_7^c \\ & + I_{NE}\beta_8^c + I_E\beta_9^c + I_{SE}\beta_{10}^c + I_{SW}\beta_{11}^c + I_S\beta_{12}^c + I_{NW}\beta_{13}^c + \widehat{\delta z}_{ba}. \end{aligned}$$

풍속자료를 이용해 이산화황의 계층모형을 구축한 결과는 표 3에 나타나 있다. 앞서 설명한 이산화황만을 가지고 모형을 적합한 결과와 마찬가지로, 오차항의 공간 상관구조를 설명함에 있어서 세 가지 세미베리오그램 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 크게 나타났다.

또한 잔차최대우도추정법과 일반최대우도추정법의 모수추정값에 대한 큰 차이가 없으며, 오차항의 분산-공분산 행렬에 포함된 모수

표 3: 계층모형의 이산화황(SO₂)의 모수 추정 결과.

	Classical linear model			Spatial linear model					
	Least Squares			REML(Spherical)		ML(Spherical)			
	Est.	95% C.I.		Est.	95% C.I.	Est.	95% C.I.		
β_1^c	-5.80 ⁰	-6.08 ⁰	-5.53 ⁰	-5.51 ⁰	-5.91 ⁰	-5.12 ⁰	-5.55 ⁰	-5.91 ⁰	-5.18 ⁰
β_2^c	3.96 ⁴	-3.67 ⁴	1.16 ³	7.43 ⁴	-4.59 ⁴	1.94 ³	7.23 ⁴	-3.56 ⁴	1.80 ³
β_3^c	1.07 ³	4.61 ⁴	1.69 ³	7.34 ⁴	-2.74 ⁴	1.74 ³	7.50 ⁴	-1.56 ⁴	1.66 ³
β_4^c	-1.54 ⁵	-2.78 ⁵	-2.94 ⁶	-1.42 ⁵	-3.70 ⁵	8.48 ⁶	-1.50 ⁵	-3.53 ⁵	5.32 ⁶
β_5^c	-5.32 ⁷	-6.37 ⁶	5.31 ⁶	8.87 ⁷	-9.20 ⁶	1.10 ⁵	9.35 ⁷	-8.06 ⁶	9.93 ⁶
β_6^c	-3.76 ⁶	-1.00 ⁵	2.51 ⁶	-3.42 ⁶	-1.37 ⁵	6.83 ⁶	-3.59 ⁶	-1.27 ⁵	5.57 ⁶
β_7^c	7.28 ¹	-7.56 ¹	2.21 ⁰	-2.08 ¹	-1.81 ⁰	1.39 ⁰	-2.84 ²	-1.54 ⁰	1.49 ⁰
β_8^c	-2.74 ¹	-4.47 ¹	-9.99 ²	-5.59 ²	-2.61 ¹	1.49 ¹	-6.99 ²	-2.63 ¹	1.23 ¹
β_9^c	-3.35 ²	-2.60 ¹	1.94 ¹	-2.39 ²	-2.58 ¹	2.10 ¹	-2.20 ²	-2.46 ¹	2.02 ¹
β_{10}^c	-2.54 ¹	-4.78 ¹	-3.10 ²	-6.54 ²	-3.18 ¹	1.87 ¹	-9.19 ²	-3.32 ¹	1.48 ²
β_{11}^c	-2.33 ¹	-3.93 ¹	-7.28 ²	-1.21 ¹	-3.01 ¹	5.84 ²	-1.30 ¹	-3.02 ¹	4.21 ²
β_{12}^c	-1.98 ¹	-3.41 ¹	-5.41 ²	-1.54 ¹	-3.29 ¹	2.05 ²	-1.57 ¹	-3.20 ¹	6.31 ²
β_{13}^c	-2.90 ¹	-4.33 ¹	-1.47 ¹	-1.83 ¹	-3.57 ¹	-8.03 ³	-1.91 ¹	-3.55 ¹	-2.73 ²
δ	4.86 ¹	2.71 ¹	7.02 ¹	2.79 ¹	-5.52 ²	6.13 ¹	3.01 ¹	-2.98 ³	6.04 ¹
τ^2				0.03		0.03			
λ^2				0.07		0.05			
ϕ				40.00		37.93			
log.lik.	-19.65			1.23		2.18			

Est., 추정값; log.lik., 로그우도값; $a^b \equiv a \times 10^{-b}$

들에도 큰 차이가 없음을 알 수 있다. 또한, 모형의 비교를 위하여 로그우도값을 살펴본 결과, 잔차최대우도추정방법(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(β_8^c), 혹은 북풍 대비 남동풍(β_{10}^c) 은 다른 방향들에 비해 낮은 평균이산화황을 가지는 것으로 나타났다.

935개 격자지점의 추가적인 정보를 이용하여 각 추정방법 하에서의 적합된 모형으로 예측한 결과는 그림 7와 같다. 구형모형으로 적합된 이산화황의 예측값은 원 측정단위로 변환된 것이다. 두 가지 추정

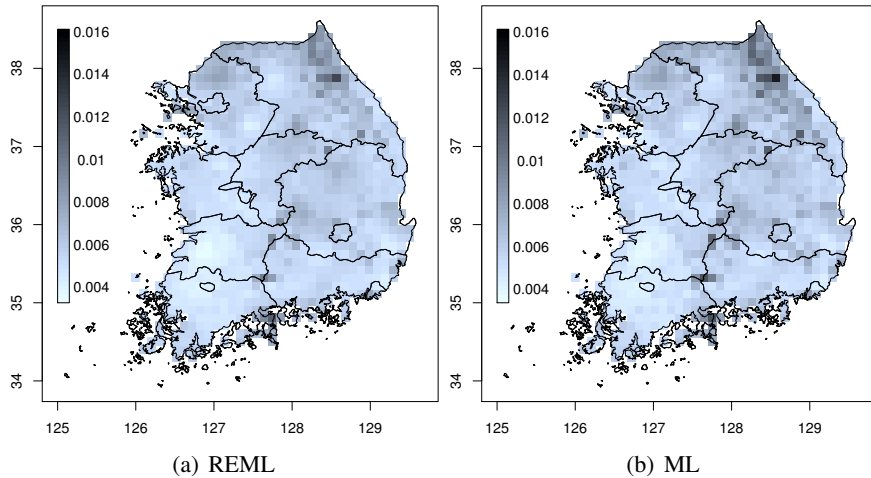


그림 7: 공간상관성을 고려하는 계층모형의 이산화황(SO₂) 예측지도 (구형모형).

방법 하에서 적합된 모형들을 기반으로 구성된 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. 이는 모형식에 풍속에 대한 정보가 포함되면서 이산화황을 예측할 때 풍속의 영향을 많이 받게되어 위와 같은 예측 결과가 나타났다. 또한 경기도 남부, 충청도 내륙지방, 전라도 내륙지방의 이산화황은 낮게 예측되었다.

2.4.4 선형회귀모형의 비교

본 절에서는 2.4.2절에서 구축한 이산화황 모형(식 2.6)과 2.4.3절에서 예측한 풍속자료를 이용한 이산화황의 계층모형(식 2.7)을 비교하고자 한다. M.1은 이산화황의 전통적 선형회귀모형, M.2는 계층모형

을 사용한 이산화황의 전통적 선형회귀모형, M.3은 공간성 상관성을 고려하여 이산화황의 일반최대우도법을 이용한 구형모형, M.4는 공간적 상관성을 고려하여 계층모형을 사용한 이산화황의 일반최대우도법을 이용한 구형모형이다. 모형들 간의 비교를 위해 사용된 방법은 먼저, AIC(Akaike information criterion)와 베이저안 정보기준(Bayesian information criterion; BIC)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. PRESS 는 교차검증방법으로 얻어진 예측값과 실제값의 차이의 제곱합으로서 다음과 같이 표현된다.

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

여기서, $\hat{Z}(\mathbf{s}_{(i)})$ 는 \mathbf{s}_i 를 제외한 $n_b - 1$ 개의 지점들의 정보를 이용하여 모형을 구한 후 \mathbf{s}_i 의 지점에서 구한 예측값을 의미한다.

표 4 를 통해 알 수 있듯이 AIC와 BIC의 경우에 전통선형회귀모형에서는 M.1에 비해 M.2의 값이 더 작으므로 모형을 더 잘 설명하는 것으로 나타났으며, 공간적 연관성을 고려한 선형회귀모형에서는 M.3 과 M.4가 큰 차이가 없다. 공간적 연관구조를 모형에 포함한 경우와 그렇지 않은 경우를 각각 고려하여 교차검증방법을 적용하여 각 지점에 대한 예측구간을 구하였으며 실제 이산화황값이 해당 예측구간에 얼마나 많이 포함되어지는지를 비교 및 평가하였다(그림 8 참조).

표 4: 선형회귀모형 선택 기준.

	Classical linear model		Spatial linear model	
	M.1	M.2	M.3	M.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

CV(1), 교차검증방법(CV(1))에 의한 실제값의 예측구간에서의 포함확률; ICC, 급내상관계수; PRESS, 예측오차제곱합.

그림 8(a)의 전통적 선형회귀모형(M.1)이 그림 8(c)의 공간 선형회귀모형(M.3)보다 실제값과 예측값이 일치하는 것을 확인할 수 있다. 전통적 선형회귀모형의 경우에는 95% 신뢰구간 내에 전체 자료의 94.20% 정도가 포함되었지만 잔차최대우도추정방법의 경우에는 84.38% 정도 포함되는 것으로 나타났다. 그러나 계층모형을 고려한 경우, 그림 8(b)의 M.2는 M.1에 비해 독립변수를 하나 더 추가했음에도 92.41%로 나타나 포함확률이 낮아졌다. 그리고 M.4(그림 8(d))의 경우 교차검증방법으로부터 얻은 포함확률이 85.27%로, M.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 M.1의 경우에는 0.03이고 M.3의 경우에는 0.49로 공간적 선형회귀모형의 경우가 더 잘 추정되었다고 할 수 있다. 마찬가지로 계층모형에서도 M.2의 급내상관계수는 0.15이고 M.4의 급내상관계수는 0.48으로 공간적 선형회귀모형의 경우의 모형의 설명력이 더 높다고 할 수 있다.

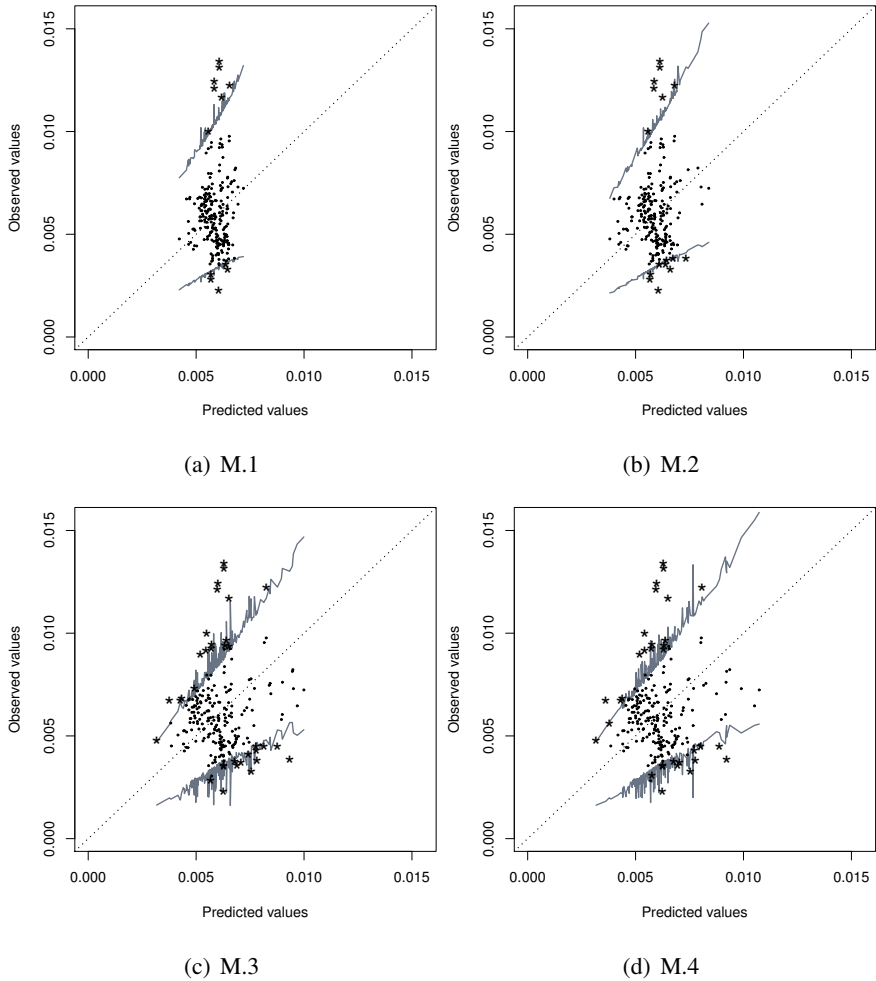


그림 8: 교차검증방법(CV(1))에 의한 이산화황 예측값과 실제값의 산점도 및 95% 예측구간

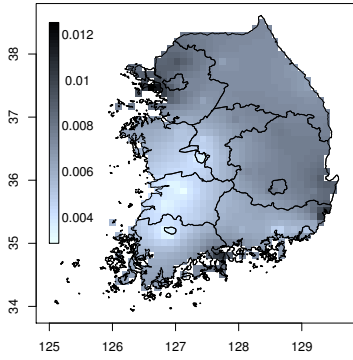
마지막으로 교차검증방법으로 얻어진 예측값과 실제값의 차이의 제곱합인 PRESS를 비교하였는데, 급내상관계수와 마찬가지로 전통적 선형회귀모형의 M.1보다 공간적 선형회귀모형의 M.3의 PRESS 통계량 값이 더 작으며, M.2보다 M.4의 값이 더 작은 것을 확인할 수 있다. 또한 M.3과 M.4의 PRESS 통계량의 값에는 큰 차이가 없다. 교차검

증방법의 포함확률과 급내상관계수, PRESS 통계량의 결과가 상반된 이유는 교차검증방법의 포함확률은 단순히 예측값의 95% 신뢰구간에 실제값이 포함된 정도를 나타내지만 급내상관계수나 PRESS 통계량의 경우에는 예측값과 실제값이 얼마나 일치하느냐를 살펴보는 척도이기 때문이다. 특히, 본 논문의 경우에는 M.1과 M.2의 신뢰구간의 길이가 M.3과 M.4보다 길기 때문에 이와 같은 결과가 나타난 것이다.

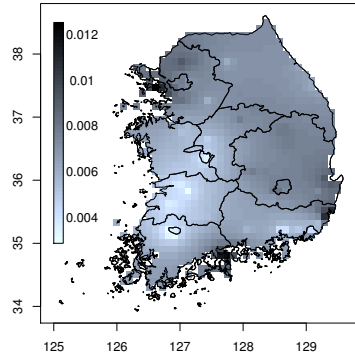
2.4.5 역거리가중치법을 이용한 이산화황 적합

본 절에서는 2.3.1절에서 소개한 역거리가중치법(Inverse distance weighting; IDW) 을 이용해 이산화황을 분석하고자 한다. 역거리가중치법은 가까이 있는 값에 가중치를 더 크게 주어 보간하고, 반대로 두 지점 사이의 거리가 멀어질수록 가중치를 작게 주는 것이다. 이는 앞 절에서 분석한 크리깅 방법과 달리 분석하고자 하는 자료만을 가지고 예측하게 된다. 이산화황을 역거리가중치법으로 분석하는데 있어서 p 값에 따른 보간 양상을 보기 위해 p 값을 0.2, 0.5, 0.8, 1.0, 2.0, 3.0 으로 하였다.

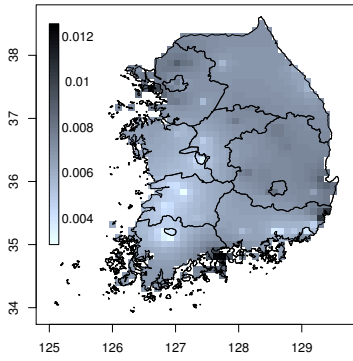
역거리가중치법을 이용해 이산화황을 예측한 지도는 그림 9와 같다. 크리깅에서와 마찬가지로 임의의 935개 격자지점에 이산화황을 예측하였다. p 값에 상관없이 이산화황에 대한 예측지도는 거의 유사하다. 예측지도를 좀더 자세하게 살펴보면 경기 북부지역, 강원도 남부지역, 충청북도 북부지역, 경상북도 일대와 전라남도 남해안 지역의 이산화황이 크게 예측되었고, 강원도 산간 일대, 충청남도 일대, 전라북도 일대와 경상남도 일대의 이산화황은 낮게 예측되었다. 이는 앞에서 살펴본 크리깅의 결과와 상반되는 부분이 있음을 확인할 수 있다.



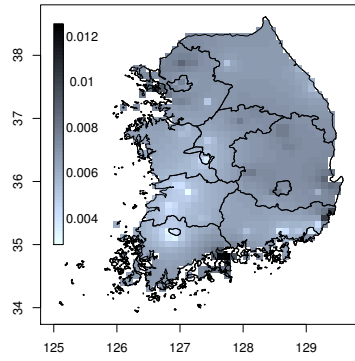
(a) $p = 0.2$



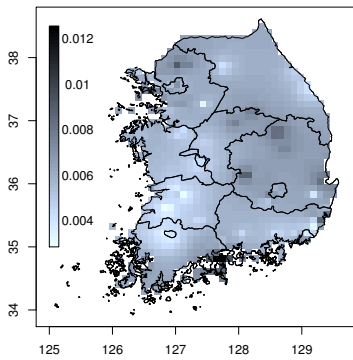
(b) $p = 0.5$



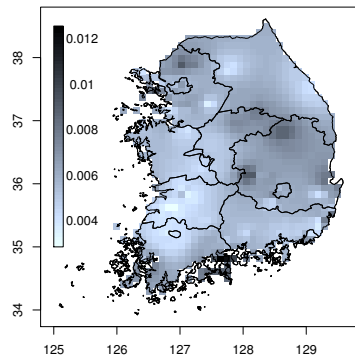
(c) $p = 0.8$



(d) $p = 1.0$



(e) $p = 2.0$



(f) $p = 3.0$

그림 9: 역거리가중치법에 의한 이산화황(SO₂) 예측지도

표 5: 크리깅과 역거리가중치법 모형 선택 기준.

	Spatial linear model		Inverse distance weighting					
	M.3	M.4	$p = 0.2$	$p = 0.5$	$p = 0.8$	$p = 1.0$	$p = 2.0$	$p = 3.0$
ICC	0.49	0.48	0.006	0.048	0.13	0.21	0.46	0.49
PRESS($\times 10^{-4}$)	5.92	6.10	8.48	8.15	7.57	7.11	6.54	7.02

ICC, 급내상관계수; PRESS, 예측오차제곱합.

또한 p 값이 0보다 작을 경우에는 보간된 양상이 좁고 날카로운 것을 확인할 수 있었고, p 값이 1이상인 경우에는 보간된 양상이 넓고 부드러운 것을 확인할 수 있었다.

또한, 본 절에서는 2.4.2절의 크리깅으로 이산화황을 예측한 결과, 2.4.3절의 크리깅으로 이산화황의 계층모형을 구축한 결과, 그리고 본 절의 역거리가중치법을 이용하여 이산화황을 예측한 결과를 비교하고자 한다. 모형들간의 비교를 위해 사용된 방법으로는, 앞절에서 설명한 급내상관계수(ICC)와 예측오차제곱합(PRESS)을 이용하였다.

여기서, M.3은 공간성 상관성을 고려하여 이산화황의 일반최대우도법을 이용한 구형모형, M.4는 공간적 상관성을 고려하여 계층모형을 사용한 이산화황의 일반최대우도법을 이용한 구형모형이다. 표 5를 살펴보면, 역거리가중치법에서 급내상관계수(ICC)의 경우 p 가 3.0일 때 가장 좋았으며, 이는 공간선형회귀모형의 결과와 크게 다르지 않다. 예측제곱오차합(PRESS)의 경우 역거리가중치법에서 p 가 2.0일 때 가장 작았으나 공간선형모형의 결과가 더 좋음을 확인할 수 있다. 이에 이산화황을 예측하는데 있어서 역거리가중치보다는 공간선형회귀모형을 사용하는 것이 더 적합하다는 것을 확인할 수 있다.

2.5 결론

본 연구에서는 서로 다른 두 가지 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 이용해 자료를 분석하는 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 풍향 자료를 모형에 포함시켰다. 이산화황의 관측망인 도시대기측정망은 주로 도시에서 관측이 이루어지므로 도시 이외의 지역에서는 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 편중되어 있기 때문에 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 되는 경향이 있었다.

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 AIC 와 교차검증방법에서는 예측력이 높음을 알 수 있었다. 또한, 본 논문에서는 공간예측을 하는데 선형회귀모형에 근간을 둔 크리깅 기법 뿐만 아니라 역거리가중치법을 사용했다. 이산화황을 예측하는 데 있어서 역거리가중치법의 p 값을 여러가지로 고려하였으며, 역거리가중치법과 크리깅의 결과를 비교하기 위해 급내상관계수와 예측오차제곱합을 비교하였다. 그 결과, 이산화황을 예측하는 데 있어서 역거리가중치법이 더 신뢰성 있는 예측성능을 보였다.

본 논문에서는 풍속이 이산화황에 미치는 영향을 두 변수간의 상관관계로만 판단했다. 추후에 전문가의 자문을 구하여 이산화황에 직접적인 영향을 미치는 다른 관측망의 자료를 사용하여 계층모형을 구축한다면 좀 더 예측력이 높아질 것이다. 또한, 본 연구에서 지상기상 관측망에서 관측된 2003년부터 2007년까지의 5년간 평균풍속자료와 도시대기측정망에서 관측된 2005년부터 2007년까지의 3년간 평균이산화황자료를 사용하였는데, 두 관측망에서 얻은 자료의 관측 시기가 동일하지 못한 한계가 있다. 하지만 두 자료를 이용해 계층모형을 구축해서 보다 좋은 결과를 보였다. 서로 다른 두 관측망에서 관측된 자료의 관측 시기가 동일하다면 보다 예측성능이 높아질 것이라고 판단된다.

제 3 장

공간모형의 붓스트랩

3.1 서론

공간 선형회귀모형에서는 오차항의 분산-공분산행렬에 공간적 연관성을 포함시키기 위하여 세미베리오그램을 추정한다. 즉, 이론적 세미베리오그램을 토대로 데이터에 적합한 세미베리오그램 모형을 추정하게 된다. 세미베리오그램 모형은 공간적 연관구조를 나타내는 모수인 뽀치(τ^2), 문턱(λ^2), 범위(ϕ) 등을 포함한다. 세미베리오그램의 모수들을 추정하기 위해서는 다양한 기법이 사용된다. 가장 많이 사용되는 방법으로는 가중최소제곱법(Weighted Least Square Method), 잔차최대우도추정법(Restricted Maximum Likelihood Method ; REML), 일반최대우도추정법(Maximum Likelihood Method ; ML) 등이 있다. 그러나 이 방법들을 사용하면 세미베리오그램의 모수는 추정할 수 있지만 검정을 위한 신뢰구간은 알 수 없다. 따라서 본 연구에서는 지수모형, 가우시안모형, 구형모형의 세 가지 모수($\theta = (\tau^2, \lambda^2, \phi)$)에 대한 구간추정을 하기 위하여 붓스트랩(Bootstrap)방법과 잭나이프(Jackknife) 방법을 이용하고자 한다.

공간자료를 이용한 붓스트랩에 관한 연구를 보면, Kang 등 (2008)은 과소추정되는 크리깅 분산을 보완하고자 공간 붓스트랩 방법을 이용해 크리깅 분산을 추정하였고, Loh와 Stein (2004)은 공간점 패턴 데이터에서 신뢰구간을 얻기 위해 재표집추출 또는 붓스트랩을 이용했

다. 특히 “Marked point method” 를 제안하고 이는 블록붓스트랩과 부분 표집추출보다 명목적 95% 신뢰구간에 더 일치함을 밝혔다. 또한 Loh와 Stein (2008)는 격자 데이터에서 공분산 함수를 Matérn 모형을 고려하여 관측치가 증가할 때 모수 ν 와 어떤 연관성이 있는지를 블록붓스트랩을 통해 밝혔고, Lillegard 등 (2004)은 공간점 패턴 데이터에서 모수적 붓스트랩을 이용해 공간적 연관성에 대해 명목적 95% 신뢰구간을 구했다. Lahiri (2003)는 격자 데이터에서 블록 붓스트랩의 방법을 소개했고, 특히 경계선이 위치한 블록을 정의하였다. Liansheng 등 (2006)은 출레스키 분해(Cholesky decomposition)를 이용해 모수적 붓스트랩을 하였고 가우시안 모형을 적용해 분석을 하였다. 또한 공간 자료를 이용한 잭나이프에 관한 연구를 보면, Lee 등 (1999)은 지질도, 지화학자료, 항공지구물리자료, 지형자료 및 원격탐사 영상 등 다양한 지질정보가 보고된 지역을 대상으로 광물 보존 예상도 작성 연구를 하여 공간 통합을 하고, 입력 자료와 통합결과에 대한 편의 분석에서 잭나이프 방법을 적용하였다. Varljen 등 (1999)은 휘발성유기화합물의 공간적 연관성을 살피고 잭나이프 방법을 기반으로 세미베리오그램의 신뢰한계를 추정하였다.

이처럼 공간 자료를 이용한 붓스트랩은 다양한 분야에서 이루어지고 있다. 블록 붓스트랩의 경우에는 격자 데이터나 공간점 패턴 데이터에 대해서는 비교적 활발한 연구가 이루어졌지만, 지리통계자료의 경우 블록 붓스트랩에 대한 연구가 활발히 이루어지지 않았다. 이에 본 연구에서는 세미베리오그램의 모수의 신뢰구간을 구하기 위해 기존의 지리통계자료의 붓스트랩 방법으로 사용된 전통적인 방법 뿐만 아니라 블록 붓스트랩 방법을 제안한다. 본 장의 순서는 다음과 같다. 2절에서는 세미베리오그램의 모수값의 신뢰구간을 추정하기 위한 공간

자료의 붓스트랩 방법과 잭나이프 방법을 소개한다. 3절에서는 제안한 방법을 모의실험을 통해 설명한다. 4절에서는 기상청의 풍속자료를 본 연구에서 제안한 방법으로 분석하였고, 마지막으로 5절에서는 결론 및 앞으로의 연구방향에 관하여 논의한다.

3.2 공간자료의 붓스트랩과 잭나이프

공간자료의 경험적 세미베리오그램을 추정할 때, 각 모수의 신뢰구간을 구하기 위하여 본 논문에서는 붓스트랩 방법을 이용하였다. 붓스트랩 방법은 Efron (1979)에 의해 처음 소개 되었으며, 표본자료를 복원으로 표집추출하여 추출된 붓스트랩 표본으로부터 통계적 추론을 하는 방법이다. 붓스트랩 방법은 크게 비모수적 붓스트랩과 모수적 붓스트랩으로 나누어진다. 비모수적 붓스트랩은 분포가 가정되지 않았을 때 사용하는 방법이며, 모수적 붓스트랩은 경험적 분포를 알고 있을 때 사용하는 방법이다. 또한, 블록 붓스트랩과 잭나이프 방법을 고려해 모수에 대한 추정을 한다.

3.2.1 공간자료의 비모수적 붓스트랩

공간자료의 분포를 가정하지 않고 사용하는 비모수적 붓스트랩의 방법은 다음과 같다. 이는 공간자료를 적합하여 얻어진 잔차의 분산-공분산 행렬을 분해하여 붓스트랩을 하는 것으로 구체적인 방법은 아래와 같다.

[단계 1] 공간자료를 이용해 선형회귀 모형을 적합한다.

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

$$\mathbf{z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \widehat{\boldsymbol{\Sigma}})$$

[단계 2] 적합된 모형의 잔차($\boldsymbol{\varepsilon}$)의 분산-공분산 행렬($\widehat{\boldsymbol{\Sigma}}$)을 고유값 분해(eigenvalue decomposition)한다.

$$\widehat{\boldsymbol{\Sigma}} = \Gamma\boldsymbol{\lambda}\Gamma'$$

$$\widehat{\boldsymbol{\Sigma}}^{1/2} = \Gamma\boldsymbol{\lambda}^{1/2}\Gamma'$$

[단계 3] $\widehat{\boldsymbol{\Sigma}}^{1/2}$ 의 역행렬을 구해서 잔차와 곱해준다.

$$\boldsymbol{\varepsilon}^* = \widehat{\boldsymbol{\Sigma}}^{-1/2} \boldsymbol{\varepsilon}$$

[단계 4] $\boldsymbol{\varepsilon}^*$ 를 중심화한 후 복원으로 자료의 개수만큼 재표집추출(resampling)한다.

$$\boldsymbol{\varepsilon}^{**} = \boldsymbol{\varepsilon}^* - \overline{\boldsymbol{\varepsilon}^*}$$

$$\boldsymbol{\varepsilon}_{B1} = \{\boldsymbol{\varepsilon}_1^{**}, \dots, \boldsymbol{\varepsilon}_n^{**}\}$$

[단계 5] $\boldsymbol{\varepsilon}_{B1}$ 를 이용해 새로운 붓스트랩 표본을 만든다.

$$\boldsymbol{\varepsilon}_* = \widehat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{\varepsilon}_{B1}$$

[단계 6] [단계 4]와 [단계 5]를 R 번 반복해 생성된 표본을 이용해 모수 추

정을 한다.

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_R)$$

3.2.2 공간자료의 모수적 붓스트랩

공간자료의 분포를 가정하고 사용하는 모수적 붓스트랩은 다음과 같다. 이는 공간자료를 적합하여 얻어진 잔차가 다변량정규분포(multivariate normal distribution)를 따른다고 가정하고 다변량정규분포에서 표본을 추출하는 방법이다.

[단계 1] 공간자료를 이용해 선형회귀 모형을 적합한다.

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

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

$$\boldsymbol{\varepsilon} \sim N_n(0, \hat{\boldsymbol{\Sigma}})$$

[단계 2] 다변량정규분포 $N_n(0, \hat{\boldsymbol{\Sigma}})$ 에서 자료의 개수만큼 잔차의 난수를 생성한다.

$$\boldsymbol{\varepsilon}_{B2} = \{\boldsymbol{\varepsilon}_1^*, \dots, \boldsymbol{\varepsilon}_n^*\}$$

[단계 3] [단계 2]를 R 번 반복해 생성된 표본을 이용해 모수 추정을 한다.

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_R)$$

3.2.3 공간자료의 블록 붓스트랩

시계열 자료, 공간 자료 등과 같은 의존성 자료는 자료들 간의 최소한의 연관성을 유지하며 붓스트랩을 할 필요가 있다. 따라서 시계열 자료, 공간자료에서는 자료들을 블록화하여 블록 내 자료들 간의 의존성을 유지시키고 블록 간에는 독립의 조건을 만족하도록 붓스트랩을 하는 블록 붓스트랩의 방법을 사용할 수 있다. 본 연구에서는 공간 자료에서 블록 붓스트랩을 하기 위하여 여러가지 방법을 제안한다. 첫 번째 방법은 다음과 같다.

[단계 1] 공간 자료(그림 10 (a))에 대하여 크기가 동일한 격자를 만들어 블록을 나누어준다. 아래의 예시에서는 300개의 지점이 표시되어 있으며, 총 25개의 블록을 나누어주었다(그림 10 (b)).

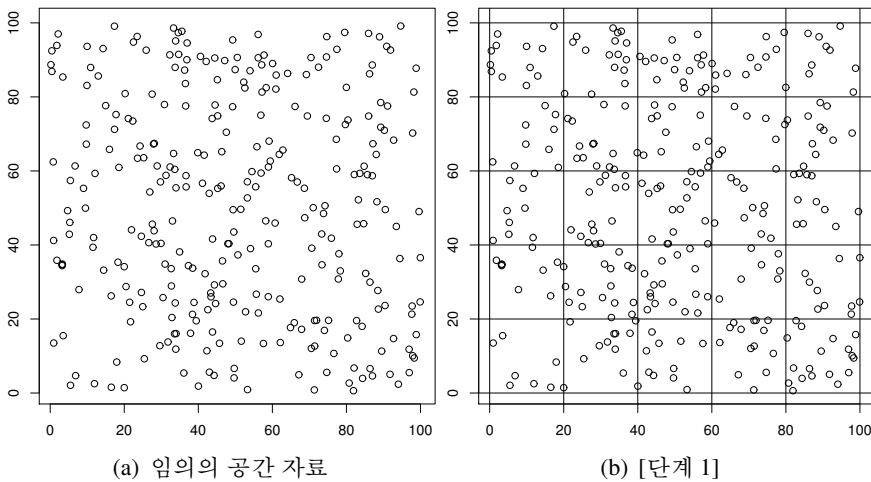


그림 10: 격자 공간 블록 붓스트랩 1.

[단계 2] 25개의 블록에 대하여 차례대로 번호를 매긴다(그림 11 (a)).

[단계 3] 1부터 25까지 복원으로 25개의 블록 붓스트랩 표본을 추출한다.

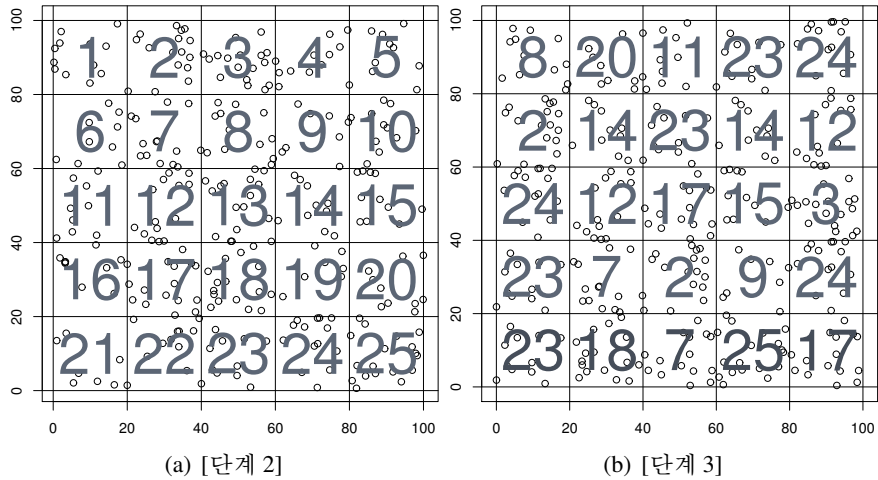


그림 11: 격자 공간 블록 붓스트랩 2.

[단계 4] [단계 3]에서 추출된 붓스트랩 표본을 차례대로 [단계 2]에서 지정한 블록의 순서대로 평행이동 시켜 새로운 붓스트랩 표본을 만든다(그림 11 (b)).

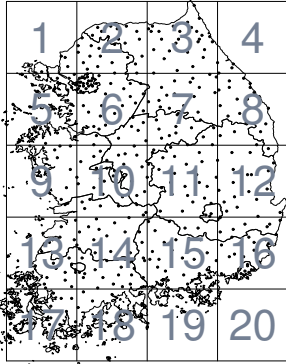
[단계 5] [단계 4]를 B번 반복해 생성된 표본을 이용해 모수 추정을 한다.

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_B)$$

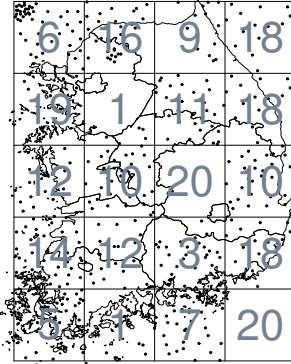
위의 방법으로 붓스트랩한 결과 그림 11 (b)와 같다. 이렇게 새로 만들어진 자료를 가지고 B번의 모수를 추정하여 모수의 평균, 분산, 편향 (Bias), 평균제곱오차(Mean Square Error, MSE), 신뢰구간 등을 구할 수 있다.

다음은 지도상의 관측된 지점에 대한 블록붓스트랩 방법이다.

[단계 1] 지도상의 지점에 대하여 크기가 동일한 격자를 만들어 블록을



(a) [단계 1]



(b) [단계 3]

그림 12: 지도상 관측지점 블록 붓스트랩

나누어준다. 아래의 예시는 2장의 풍속자료이고 460개의 지점이 표시되어 있으며, 총 20개의 블록을 나누어 차례대로 번호를 매긴다((그림 12 (a))).

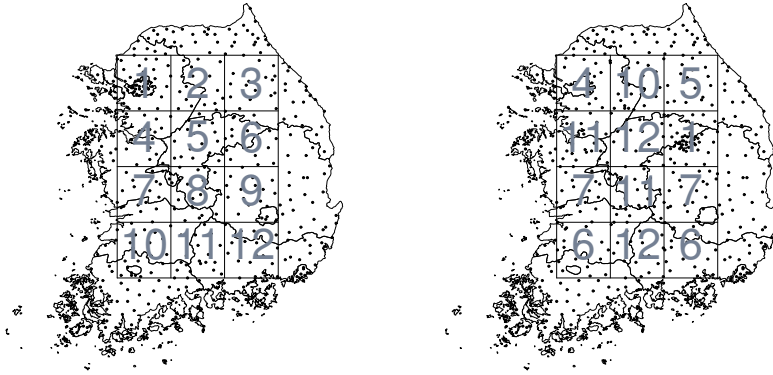
[단계 2] 1부터 20까지 복원으로 20개의 블록 붓스트랩 표본을 추출한다.

[단계 3] [단계 2]에서 추출된 붓스트랩 표본을 차례대로 [단계 1]에서 지정한 블록의 순서대로 평행이동 시켜 새로운 붓스트랩 표본을 만든다(그림 12 (b)).

[단계 4] [단계 3]를 B번 반복해 생성된 표본을 이용해 모수 추정을 한다.

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_B)$$

위의 방법으로 데이터를 생성한 결과 그림 12 (b)와 같다. 이는 육지가 아닌 바다에 붓스트랩 표본이 할당되어 추후 분석이 어렵다. 따라서 다



(a) [단계 1]

(b) [단계 3]

그림 13: 경계지점 제외한 블록 붓스트랩.

른 방법으로 블록 붓스트랩을 해야한다. 이에 다음 방법을 제안한다.

[단계 1] 지도상의 지점에 대하여 크기가 동일한 격자를 만들어 블록을 나누어준다. 이때, 육지와 바다의 경계지점을 포함하는 블록은 붓스트랩 모집단에서 제외한다. 그 결과, 총 12개의 블록을 만들었고 차례대로 번호를 매긴다(그림 13 (a)).

[단계 2] 1부터 12까지 복원으로 12개의 블록 붓스트랩 표본을 추출한다.

[단계 3] [단계 2]에서 추출된 붓스트랩 표본을 차례대로 [단계 1]에서 지정한 블록의 순서대로 평행이동 시켜 새로운 붓스트랩 표본을 만든다(그림 13 (b)).

[단계 4] [단계 3]를 B번 반복해 생성된 표본을 이용해 모수 추정을 한다.

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_B)$$

이와 같이 블록을 설정한 결과 그림 13 (b)와 같다. 이는 육지안의 관측 지점에서만 붓스트랩 표본이 뽑히므로 자료를 추후 분석에서 사용할 수 있다. 따라서 본 연구에서는 모의실험에서는 동일한 격자에서의 데이터를 생성하여 블록 붓스트랩 방법을 사용하고, 실제자료에 적용 할 때는 경계지점을 제외한 블록 붓스트랩 방법을 사용하고자 한다.

다음으로는 블록의 개수를 정하는 방법을 제안한다. 2.3.2절에서 정의한 경험적 세미베리오그램은 일반적으로 최대 거리에 대한 비율과 구간의 수, 하나의 구간 안의 평균적으로 들어있는 점의 수를 고려하여 만들어진다. 이에 한 블록 안의 관측값 수가 몇 개 이상이어야 위의 조건을 만족하는 지를 기반으로 블록의 수를 구하고자 한다. 블록의 수를 B 개, 전체 데이터 수를 n 개, 최대 거리 비율을 r , 구간의 수를 a 개, 한 개의 구간 안의 평균적 점의 수를 b 개, 한 블록 내 관측값 수를 x 개 라고 한다면 다음과 같이 한 블록 내 관측값 수(x)를 구하고 블록의 수(B)를 구할 수 있다.

$$\begin{aligned}
 r\binom{x}{2} &= ab. \\
 \Leftrightarrow \frac{rx(x-1)}{2} &= ab. \\
 \Leftrightarrow x^2 - x - 2ab/r &= 0. \\
 \Leftrightarrow x &= \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2ab}{r}}.
 \end{aligned} \tag{3.1}$$

$$B = \left\lceil \frac{n}{x} \right\rceil.$$

여기서, $[x]$ 는 실수 x 에 대하여 x 보다 크지 않은 최대의 정수이다.

3.2.4 공간자료의 잭나이프 방법

공간자료의 세미베리오그램 모수들의 신뢰구간을 구하기 위하여 붓스트랩 방법과 더불어 잭나이프 방법을 이용하였다. 잭나이프 방법은 Quenouille (1949)에 의해 처음 제안되었으며, 관측된 자료에서 하나의 값을 제외하는 것에 기반한다. 주어진 자료를 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 라고 할 때 아래의 방법으로 모수를 추정한다.

[단계 1] 관측된 자료에서 하나의 값을 제외해 $n-1$ 개의 자료를 갖는 $\mathbf{x}_{(-i)}$ 를 다음과 같이 생성한다. $i = 1, \dots, n$ 에 대하여

$$\mathbf{x}_{(-i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

[단계 2] 모든 $i = 1, \dots, n$ 에 대한 $\mathbf{x}_{(-i)}$ 를 이용해 모수 θ_i 를 추정 한다.

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n)$$

이에 본 연구에서는 아래의 세 가지 붓스트랩 방법과 잭나이프 방법을 고려한다.

[방법 1] 잔차의 공분산행렬을 고유값 분해하여 붓스트랩을 시행한다.

[방법 2] 잔차가 따르는 다변량 정규분포에서 난수를 생성하여 붓스트랩을 시행한다.

[방법 3] 블록을 나누어 붓스트랩을 시행한다.

[방법 4] 잭나이프 방법을 시행한다.

여기서, [방법 1]과 [방법 2]를 전통적인 붓스트랩이라고 하겠다.

표 6: 모의실험 시나리오.

Scenario	n	x	y	nugget	partial sill	range
1	300	$U(0, 100)$	$U(0, 100)$	0.1	0.9	10
2	300	$U(0, 100)$	$U(0, 100)$	0.2	0.8	10
3	300	$U(0, 100)$	$U(0, 100)$	0.3	0.7	10
4	300	$U(0, 100)$	$U(0, 100)$	0.1	0.9	20
5	300	$U(0, 100)$	$U(0, 100)$	0.2	0.8	20
6	300	$U(0, 100)$	$U(0, 100)$	0.3	0.7	20

n , 지점의 수; x , x 축 범위; y , y 축 범위; nugget, 뭉치; partial sill, 부분 문턱; range, 거리;

3.3 모의실험

3.3.1 모의실험 방법

세미베리오그램의 모수($\theta = (\tau^2, \lambda^2, \phi)$)의 다양한 통계량을 구하기 위해 3.2절에서 다룬 붓스트랩 방법과 잭나이프 방법으로 모의실험을 수행하였다. 본 연구에는 동일한 격자에서의 데이터를 생성하여 일반 최대우도추정법을 이용해 모수 추정을 한다. 또한 공분산함수는 지수 모형, 가우시안모형, 구형모형을 이용하여 여섯가지 시나리오(표 6)를 고려한다. 이 때, 블록 붓스트랩은 블록의 갯수가 9개일 때, 16개일 때, 25개일 때의 세 가지 경우를 고려한다. 이와 같은 시나리오를 다음 절차에 따라 모수의 통계량을 추정하였다. 여기서, 모형의 특성상 지수모형과 가우시안모형의 거리(ϕ)의 경우 구형모형의 거리(ϕ)의 1/3이므로 지수모형과 가우시안모형의 시나리오 1, 2, 3의 거리의 초기값은 10/3이고, 시나리오 4, 5, 6의 거리의 초기값은 20/3이다.

모의실험의 절차는 다음과 같다.

- (1) 균일분포 $U(0, 100)$ 에서 300개의 지점 $\mathbf{s} = (x, y)$ 를 생성한다.

- (2) (1)에서 구한 위치정보를 토대로 시나리오별, 모형별 자료를 각각 100개 생성한다(총 1,800개).
- (3) 각 데이터를 3.2절에서 소개한 분산-공분산 행렬의 고유값 분해 방법, 다변량 정규분포에서의 잔차의 난수 생성 방법, 블록 붓스트랩 방법을 이용해 각각을 99번 붓스트랩하여 표본을 만들고, 그 표본에 대하여 모수를 추정한다.
- (4) 각 자료에서 잭나이프 방법으로 300개의 잭나이프 표본에 대하여 모수를 추정한다.
- (5) 3, 4단계에서 추정된 값을 토대로 평균, 분산, 편향, 평균제곱오차, 95% 신뢰구간을 구한다.

3.3.2 모의실험 결과

표 7부터 표 12은 시나리오에 따른 모수($\theta = (\tau^2, \lambda^2, \phi)$)의 참값과 평균(mean), 분산(variance), 편향(bias), 평균제곱오차(MSE), 95% 신뢰구간의 결과를 각각 정리한 것이다. 여기서 붓스트랩 방법에서의 모수들의 평균, 분산, 편향, 평균제곱오차, 95% 신뢰구간을 구한 방법은 다음과 같다. 여기서 K 는 시나리오별, 모형별 데이터 수(100개)이고, R 은 반복수(99번)이다.

[1] 붓스트랩 평균(mean)

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{R} \sum_{r=1}^R \hat{\theta}_r^k = \frac{1}{N} \sum_{k=1}^K \bar{\theta}^k.$$

여기서, $\hat{\theta}_r^k$ 는 k 번째 데이터, $\mathbf{x}^k = (x_1^k, x_2^k, \dots, x_n^k)$ 로부터 생성한 r

번째 붓스트랩 표본에서의 추정량을 의미한다.

[2] 붓스트랩 분산(variance)

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r^k - \bar{\theta}_r).$$

[3] 붓스트랩 편향(bias)

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{R} \sum_{r=1}^R (\bar{\theta}^k - \theta).$$

[4] 붓스트랩 평균제곱오차(MSE)

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{R} \sum_{r=1}^R (\bar{\theta}^k - \theta)^2.$$

[5] 붓스트랩 95% 경험적 신뢰구간(empirical confidence interval)

$$\left(\frac{1}{K} \sum_{k=1}^K \hat{\theta}_{(0.025)}^k, \frac{1}{K} \sum_{n=1}^K \hat{\theta}_{(0.975)}^k \right).$$

여기에서, $\hat{\theta}_{(\alpha)}$ 는 $\{\hat{\theta}_r^k, r = 1, \dots, R\}$ 의 $(100 \times \alpha)$ 분위수이다.

또한 잭나이프방법에서는 모수 추정값의 과대평가를 막기 위해 다음의 방법으로 평균, 분산, 편향, 평균제곱오차, 신뢰구간을 구하였다. 여기서, n 는 관측된 자료의 표본 크기(300 개)이며, 경험적 신뢰구간이 아닌 정규분포에 의한 신뢰구간을 구한다.

[1] 잭나이프 평균(mean)

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i^k = \frac{1}{K} \sum_{k=1}^K \bar{\theta}^k.$$

여기서, $\hat{\theta}_i^k$ 는 k 번째 데이터 $\mathbf{x}_{(-i)}^k = (x_1^k, x_2^k, \dots, x_n^k)$ 로부터 생성한 i 번째 잭나이프 표본에서의 추정량을 의미한다.

[2] 잭나이프 분산(variance)

$$\frac{1}{K} \sum_{k=1}^K \frac{(n-1)}{n} \sum_{i=1}^n \left(\hat{\theta}_i^k - \bar{\theta}^k \right).$$

[3] 잭나이프 편향(bias)

$$\frac{1}{K} \sum_{k=1}^K (n-1) \left(\bar{\theta}^k - \theta \right).$$

[4] 잭나이프 평균제곱오차(MSE)

$$\frac{1}{K} \sum_{k=1}^K \left[\frac{(n-1)}{n} \sum_{i=1}^n \left(\hat{\theta}_i^k - \bar{\theta}^k \right) + \left\{ (n-1) \left(\bar{\theta}^k - \theta \right) \right\}^2 \right].$$

[5] 잭나이프 95% 신뢰구간(confidence interval)

$$\frac{1}{K} \sum_{k=1}^K \left(\bar{\theta}^k \pm 1.96 \sqrt{\frac{(n-1)}{n} \sum_{i=1}^n \left(\hat{\theta}_i^k - \bar{\theta}^k \right)} \right).$$

모의실험 결과를 살펴보면 전체적으로 멱차(τ^2), 부분 문턱(λ^2)의 경우에는 세 가지 모형과 모든 방법에서 비슷한 결과를 나타내고 있다

표 7: 지수모형의 모의실험 결과(Scenario 1 ~ 3).

	Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
Scenario 1	EVD	τ^2	0.10	0.10	0.01	-0.00	0.01	0.00 / 0.33
		λ^2	0.90	0.89	0.02	-0.01	0.02	0.60 / 1.15
		ϕ	3.43	3.43	0.60	0.01	0.60	2.20 / 5.33
	MVN	τ^2		0.09	0.01	-0.00	0.01	0.00 / 0.32
		λ^2		0.90	0.02	-0.01	0.02	0.60 / 1.15
		ϕ		3.44	0.59	0.01	0.59	2.20 / 5.39
	Blk9	τ^2		0.10	0.01	0.00	0.01	0.00 / 0.30
		λ^2		0.89	0.02	-0.01	0.02	0.62 / 1.11
		ϕ		3.18	0.88	-0.24	0.99	2.12 / 4.76
	Blk16	τ^2		0.10	0.01	0.00	0.01	0.00 / 0.34
		λ^2		0.89	0.02	-0.01	0.02	0.60 / 1.12
		ϕ		3.05	0.55	-0.38	0.77	1.96 / 4.74
	Blk25	τ^2		0.10	0.01	0.00	0.01	0.00 / 0.36
		λ^2		0.89	0.02	-0.01	0.02	0.57 / 1.13
		ϕ		2.89	0.57	-0.54	0.93	1.85 / 4.65
JK	τ^2		0.10	0.02	-0.03	0.03	-0.18 / 0.37	
	λ^2		0.90	0.03	0.02	0.04	0.57 / 1.23	
	ϕ		3.43	0.74	-0.06	1.07	1.79 / 5.06	
Scenario 2	EVD	τ^2	0.20	0.18	0.02	-0.02	0.02	0.00 / 0.46
		λ^2	0.80	0.81	0.02	0.01	0.02	0.48 / 1.11
		ϕ	3.40	3.38	0.92	-0.02	0.92	1.94 / 5.86
	MVN	τ^2		0.19	0.02	-0.02	0.02	0.00 / 0.46
		λ^2		0.81	0.03	0.01	0.03	0.47 / 1.11
		ϕ		3.40	1.00	0.00	1.00	1.92 / 5.90
	Blk9	τ^2		0.19	0.02	-0.01	0.02	0.01 / 0.46
		λ^2		0.80	0.02	0.00	0.02	0.48 / 1.08
		ϕ		3.17	1.29	-0.23	1.47	1.86 / 5.12
	Blk16	τ^2		0.19	0.02	-0.01	0.02	0.01 / 0.45
		λ^2		0.80	0.02	0.01	0.03	0.47 / 1.09
		ϕ		3.00	0.64	-0.39	0.93	1.80 / 4.87
	Blk25	τ^2		0.19	0.02	-0.01	0.02	0.00 / 0.50
		λ^2		0.81	0.03	0.01	0.03	0.44 / 1.10
		ϕ		2.87	1.48	-0.53	1.90	1.63 / 5.30
JK	τ^2		0.20	0.05	-0.04	0.13	-0.18 / 0.59	
	λ^2		0.80	0.06	0.04	0.15	0.37 / 1.23	
	ϕ		3.40	1.31	-0.07	3.63	1.35 / 5.44	
Scenario 3	EVD	τ^2	0.30	0.28	0.03	-0.03	0.03	0.00 / 0.60
		λ^2	0.70	0.72	0.03	0.02	0.03	0.35 / 1.07
		ϕ	3.47	3.50	1.58	0.03	1.58	1.71 / 6.63
	MVN	τ^2		0.27	0.02	-0.03	0.03	0.00 / 0.59
		λ^2		0.72	0.03	0.03	0.03	0.36 / 1.07
		ϕ		3.48	1.56	0.01	1.56	1.75 / 6.66
	Blk9	τ^2		0.28	0.02	-0.02	0.02	0.03 / 0.57
		λ^2		0.71	0.03	0.02	0.03	0.38 / 1.03
		ϕ		3.23	1.13	-0.23	1.26	1.79 / 5.42
	Blk16	τ^2		0.28	0.02	-0.03	0.03	0.02 / 0.58
		λ^2		0.72	0.03	0.02	0.03	0.36 / 1.05
		ϕ		3.10	1.78	-0.37	2.10	1.62 / 5.54
	Blk25	τ^2		0.27	0.03	-0.04	0.03	0.02 / 0.61
		λ^2		0.73	0.03	0.03	0.04	0.35 / 1.06
		ϕ		2.92	1.47	-0.55	1.92	1.52 / 5.54
JK	τ^2		0.30	0.06	-0.08	0.07	-0.12 / 0.73	
	λ^2		0.70	0.06	0.08	0.08	0.23 / 1.16	
	ϕ		3.47	1.50	-0.07	1.89	1.18 / 5.75	

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk9, 블록 9개; Blk16, 블록 16개; Blk25, 블록 25개; JK, 잭나이프방법;

는 것을 확인할 수 있다. 특히, 붓스트랩 방법에 비해 잭나이프 방법의 경우 뭉치와 부분 문턱의 편향과 평균제곱오차가 크다는 것을 확인할 수 있다. 거리(ϕ)의 경우에는 표 7과 표 8의 지수모형의 모의실험

표 8: 지수모형의 모의실험 결과(Scenario 4 ~ 6).

	Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
Scenario 4	EVD	τ^2	0.10	0.09	0.00	-0.01	0.00	0.00 / 0.22
		λ^2	0.90	0.89	0.02	-0.01	0.02	0.63 / 1.18
		ϕ	6.63	6.48	2.17	-0.15	2.20	4.08 / 10.08
	MVN	τ^2		0.09	0.00	-0.01	0.00	0.00 / 0.22
		λ^2		0.89	0.02	-0.01	0.02	0.63 / 1.18
		ϕ		6.45	2.19	-0.18	2.22	4.04 / 10.06
	Blk9	τ^2		0.09	0.00	-0.01	0.00	0.00 / 0.22
		λ^2		0.90	0.01	-0.00	0.01	0.66 / 1.13
		ϕ		5.48	1.21	-1.15	2.68	3.68 / 8.01
	Blk16	τ^2		0.09	0.00	-0.01	0.01	0.00 / 0.25
		λ^2		0.90	0.02	-0.00	0.02	0.64 / 1.15
		ϕ		5.14	1.17	-1.49	3.56	3.36 / 7.66
	Blk25	τ^2		0.08	0.00	-0.01	0.01	0.00 / 0.25
		λ^2		0.90	0.02	0.00	0.02	0.64 / 1.15
		ϕ		4.68	0.87	-1.95	4.88	3.13 / 6.90
	JK	τ^2		0.10	0.01	-0.02	0.01	-0.07 / 0.27
		λ^2		0.90	0.02	0.02	0.02	0.66 / 1.14
		ϕ		6.63	2.05	-0.02	3.65	3.96 / 9.30
Scenario 5	EVD	τ^2	0.20	0.19	0.01	-0.01	0.01	0.02 / 0.35
		λ^2	0.80	0.80	0.02	-0.00	0.02	0.53 / 1.09
		ϕ	6.73	6.50	2.82	-0.23	2.87	3.75 / 10.67
	MVN	τ^2		0.19	0.01	-0.02	0.01	0.03 / 0.35
		λ^2		0.80	0.02	-0.00	0.02	0.53 / 1.10
		ϕ		6.51	2.97	-0.21	3.02	3.77 / 10.91
	Blk9	τ^2		0.18	0.01	-0.02	0.01	0.03 / 0.35
		λ^2		0.80	0.02	0.00	0.02	0.54 / 1.08
		ϕ		5.50	1.64	-1.23	3.42	3.41 / 8.43
	Blk16	τ^2		0.18	0.01	-0.02	0.01	0.02 / 0.37
		λ^2		0.81	0.02	0.01	0.02	0.52 / 1.09
		ϕ		5.08	1.42	-1.65	4.34	3.13 / 7.84
	Blk25	τ^2		0.17	0.01	-0.03	0.01	0.01 / 0.39
		λ^2		0.82	0.02	0.02	0.02	0.52 / 1.11
		ϕ		4.62	1.32	-2.11	6.03	2.81 / 7.44
	JK	τ^2		0.20	0.01	-0.02	0.01	-0.00 / 0.41
		λ^2		0.80	0.02	0.02	0.02	0.53 / 1.07
		ϕ		6.73	2.61	-0.09	3.67	3.64 / 9.81
Scenario 6	EVD	τ^2	0.30	0.28	0.01	-0.02	0.01	0.07 / 0.47
		λ^2	0.70	0.70	0.02	0.00	0.02	0.42 / 1.00
		ϕ	6.56	6.37	3.69	-0.19	3.74	3.38 / 11.24
	MVN	τ^2		0.28	0.01	-0.02	0.01	0.07 / 0.47
		λ^2		0.70	0.02	0.00	0.02	0.43 / 1.01
		ϕ		6.32	3.56	-0.23	3.62	3.42 / 11.13
	Blk9	τ^2		0.27	0.01	-0.03	0.01	0.07 / 0.47
		λ^2		0.72	0.02	0.02	0.02	0.44 / 1.00
		ϕ		5.35	2.40	-1.21	4.17	3.11 / 8.79
	Blk16	τ^2		0.27	0.01	-0.03	0.02	0.05 / 0.49
		λ^2		0.72	0.02	0.02	0.03	0.43 / 1.02
		ϕ		5.02	2.00	-1.54	4.84	2.89 / 8.27
	Blk25	τ^2		0.27	0.01	-0.03	0.02	0.05 / 0.51
		λ^2		0.72	0.02	0.02	0.03	0.41 / 1.03
		ϕ		4.64	1.77	-1.91	5.94	2.69 / 7.89
	JK	τ^2		0.30	0.02	-0.03	0.02	0.04 / 0.55
		λ^2		0.70	0.03	0.02	0.03	0.39 / 1.00
		ϕ		6.56	3.47	0.01	5.11	3.04 / 10.07

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk9, 블록 9개; Blk16, 블록 16개; Blk25, 블록 25개; JK, 잭나이프방법;

결과, 시나리오 3을 제외하고는 전통적인 붓스트랩에서 편향과 평균
 제공오차가 가장 작게 나타났고, 블록 붓스트랩과 잭나이프의 편향과
 평균제공오차가 크게 나타났다. 또한, 표 8의 시나리오 4, 5, 6에서는 블

표 9: 가우시안모형의 모의실험 결과(Scenario 1 ~ 3).

	Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
Scenario 1	EVD	τ^2	0.10	0.10	0.00	-0.00	0.00	0.02 / 0.19
		λ^2	0.90	0.89	0.01	-0.01	0.01	0.69 / 1.11
		ϕ	3.40	3.39	0.11	-0.01	0.11	2.73 / 4.12
	MVN	τ^2		0.10	0.00	-0.00	0.00	0.02 / 0.19
		λ^2		0.90	0.01	0.00	0.01	0.70 / 1.12
		ϕ		3.39	0.12	-0.00	0.12	2.74 / 4.15
	Blk9	τ^2		0.11	0.00	0.01	0.00	0.02 / 0.26
		λ^2		0.88	0.01	-0.02	0.01	0.66 / 1.09
		ϕ		3.28	0.16	-0.12	0.19	2.56 / 4.14
	Blk16	τ^2		0.12	0.00	0.02	0.00	0.02 / 0.28
		λ^2		0.88	0.01	-0.02	0.01	0.64 / 1.09
		ϕ		3.23	0.17	-0.17	0.22	2.48 / 4.11
	Blk25	τ^2		0.13	0.01	0.03	0.01	0.02 / 0.32
		λ^2		0.87	0.01	-0.03	0.02	0.61 / 1.09
		ϕ		3.16	0.19	-0.23	0.27	2.39 / 4.11
	JK	τ^2		0.10	0.00	0.00	0.00	-0.02 / 0.22
		λ^2		0.90	0.01	-0.01	0.01	0.69 / 1.11
		ϕ		3.40	0.20	0.03	0.29	2.55 / 4.24
Scenario 2	EVD	τ^2	0.20	0.19	0.00	-0.00	0.00	0.07 / 0.34
		λ^2		0.80	0.01	-0.00	0.01	0.57 / 1.04
		ϕ	3.48	3.49	0.21	0.01	0.21	2.62 / 4.53
	MVN	τ^2		0.19	0.00	-0.01	0.00	0.06 / 0.34
		λ^2		0.80	0.01	-0.00	0.01	0.58 / 1.04
		ϕ		3.48	0.20	0.01	0.20	2.63 / 4.50
	Blk9	τ^2		0.21	0.01	0.01	0.01	0.07 / 0.37
		λ^2		0.79	0.01	-0.01	0.02	0.54 / 1.03
		ϕ		3.38	0.25	-0.09	0.28	2.53 / 4.41
	Blk16	τ^2		0.22	0.01	0.02	0.01	0.07 / 0.40
		λ^2		0.78	0.02	-0.02	0.02	0.53 / 1.03
		ϕ		3.30	0.26	-0.17	0.33	2.43 / 4.38
	Blk25	τ^2		0.22	0.01	0.02	0.01	0.06 / 0.41
		λ^2		0.78	0.02	-0.02	0.02	0.52 / 1.03
		ϕ		3.23	0.25	-0.24	0.34	2.35 / 4.34
	JK	τ^2		0.20	0.02	-0.01	0.02	-0.00 / 0.40
		λ^2		0.80	0.02	0.01	0.03	0.54 / 1.07
		ϕ		3.48	0.47	0.05	1.03	2.35 / 4.60
Scenario 3	EVD	τ^2	0.30	0.29	0.01	-0.01	0.01	0.10 / 0.49
		λ^2		0.70	0.02	0.01	0.02	0.45 / 0.98
		ϕ	3.40	3.42	0.33	0.02	0.33	2.35 / 4.77
	MVN	τ^2		0.29	0.01	-0.01	0.01	0.10 / 0.49
		λ^2		0.71	0.02	0.01	0.02	0.44 / 0.97
		ϕ		3.41	0.35	0.01	0.35	2.34 / 4.77
	Blk9	τ^2		0.30	0.01	0.00	0.01	0.11 / 0.50
		λ^2		0.70	0.02	-0.01	0.02	0.45 / 0.95
		ϕ		3.31	0.34	-0.09	0.38	2.33 / 4.60
	Blk16	τ^2		0.31	0.01	0.01	0.01	0.10 / 0.54
		λ^2		0.69	0.02	-0.02	0.02	0.41 / 0.96
		ϕ		3.31	0.51	-0.10	0.56	2.21 / 4.92
	Blk25	τ^2		0.32	0.01	0.01	0.02	0.09 / 0.56
		λ^2		0.68	0.02	-0.02	0.02	0.40 / 0.97
		ϕ		3.23	0.51	-0.17	0.59	2.07 / 4.82
	JK	τ^2		0.30	0.02	-0.02	0.02	0.03 / 0.57
		λ^2		0.70	0.03	0.02	0.03	0.39 / 1.02
		ϕ		3.40	0.55	0.03	0.88	2.03 / 4.77

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk9, 블록 9개; Blk16, 블록 16개; Blk25, 블록 25개; JK, 잭나이프방법;

록 붓스트랩의 경우에 블록의 개수가 많아질수록 세 가지 모수에 대한 편향과 평균제곱오차가 커지는 경향을 확인할 수 있다.

표 10: 가우시안모형의 모의실험 결과(Scenario 4 ~ 6).

	Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
Scenario 4	EVD	τ^2	0.10	0.10	0.00	-0.00	0.00	0.06 / 0.14
		λ^2	0.90	0.89	0.01	-0.01	0.01	0.67 / 1.16
		ϕ	6.62	6.61	0.19	-0.02	0.19	5.75 / 7.54
	MVN	τ^2		0.10	0.00	-0.00	0.00	0.06 / 0.14
		λ^2		0.90	0.01	-0.01	0.01	0.67 / 1.15
		ϕ		6.61	0.19	-0.01	0.19	5.76 / 7.56
	Blk9	τ^2		0.12	0.00	0.03	0.00	0.07 / 0.21
		λ^2		0.86	0.01	-0.04	0.01	0.66 / 1.07
		ϕ		6.04	0.26	-0.58	0.65	5.05 / 7.08
	Blk16	τ^2		0.14	0.00	0.04	0.00	0.07 / 0.23
		λ^2		0.85	0.01	-0.05	0.02	0.64 / 1.07
		ϕ		5.79	0.30	-0.83	1.06	4.69 / 6.94
	Blk25	τ^2		0.15	0.00	0.05	0.01	0.06 / 0.27
		λ^2		0.84	0.01	-0.06	0.02	0.62 / 1.08
		ϕ		5.52	0.35	-1.10	1.65	4.36 / 6.76
	JK	τ^2		0.10	0.00	-0.00	0.00	0.04 / 0.15
		λ^2		0.90	0.01	-0.00	0.01	0.74 / 1.07
		ϕ		6.62	0.26	0.00	0.40	5.65 / 7.60
Scenario 5	EVD	τ^2	0.19	0.19	0.00	-0.00	0.00	0.13 / 0.26
		λ^2	0.80	0.79	0.01	-0.00	0.01	0.58 / 1.04
		ϕ	6.65	6.64	0.33	-0.01	0.33	5.52 / 7.90
	MVN	τ^2		0.19	0.00	-0.00	0.00	0.13 / 0.26
		λ^2		0.80	0.01	-0.00	0.01	0.58 / 1.05
		ϕ		6.63	0.33	-0.02	0.33	5.52 / 7.89
	Blk9	τ^2		0.22	0.00	0.02	0.00	0.13 / 0.32
		λ^2		0.76	0.01	-0.04	0.01	0.55 / 0.98
		ϕ		6.10	0.43	-0.56	0.81	4.87 / 7.50
	Blk16	τ^2		0.23	0.00	0.04	0.00	0.13 / 0.35
		λ^2		0.75	0.01	-0.05	0.02	0.53 / 0.98
		ϕ		5.80	0.47	-0.85	1.30	4.49 / 7.24
	Blk25	τ^2		0.24	0.00	0.05	0.01	0.13 / 0.38
		λ^2		0.74	0.01	-0.06	0.02	0.52 / 0.98
		ϕ		5.55	0.48	-1.10	1.81	4.20 / 7.00
	JK	τ^2		0.19	0.00	-0.01	0.00	0.10 / 0.29
		λ^2		0.80	0.01	-0.00	0.01	0.61 / 0.99
		ϕ		6.65	0.50	-0.08	0.78	5.32 / 7.99
Scenario 6	EVD	τ^2	0.30	0.29	0.00	-0.00	0.00	0.20 / 0.40
		λ^2	0.69	0.69	0.01	-0.00	0.01	0.47 / 0.94
		ϕ	6.59	6.55	0.50	-0.03	0.50	5.18 / 8.11
	MVN	τ^2		0.29	0.00	-0.00	0.00	0.20 / 0.39
		λ^2		0.69	0.01	-0.00	0.01	0.48 / 0.93
		ϕ		6.56	0.50	-0.02	0.50	5.23 / 8.15
	Blk9	τ^2		0.31	0.00	0.02	0.00	0.20 / 0.43
		λ^2		0.67	0.01	-0.03	0.01	0.46 / 0.89
		ϕ		6.09	0.56	-0.49	0.86	4.69 / 7.68
	Blk16	τ^2		0.33	0.00	0.03	0.01	0.20 / 0.47
		λ^2		0.66	0.01	-0.04	0.02	0.44 / 0.89
		ϕ		5.81	0.69	-0.78	1.47	4.25 / 7.61
	Blk25	τ^2		0.33	0.01	0.04	0.01	0.19 / 0.48
		λ^2		0.65	0.01	-0.05	0.02	0.42 / 0.88
		ϕ		5.55	0.70	-1.04	1.93	4.00 / 7.37
	JK	τ^2		0.30	0.00	-0.01	0.01	0.17 / 0.43
		λ^2		0.69	0.01	0.00	0.01	0.49 / 0.89
		ϕ		6.59	0.73	-0.04	1.02	4.98 / 8.19

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk9, 블록 9개; Blk16, 블록 16개; Blk25, 블록 25개; JK, 잭나이프방법;

그리고 표 9의 가우시안 모의실험의 결과, 거리(ϕ)의 편향과 평균 제곱오차가 전통적인 붓스트랩에서 가장 작았으며, 잭나이프의 경우 가장 크게 나타났다. 거리의 참값이 더 긴 표 10의 결과에서도 거리(ϕ)

표 11: 구형모형의 모의실험 결과(Scenario 1 ~ 3).

	Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
Scenario 1	EVD	τ^2	0.10	0.10	0.00	-0.00	0.00	0.00 / 0.22
		λ^2	0.91	0.90	0.02	-0.00	0.02	0.66 / 1.16
	MVN	ϕ	10.10	10.09	1.47	-0.00	1.47	7.95 / 12.74
		τ^2		0.10	0.00	-0.00	0.00	0.00 / 0.22
		λ^2		0.90	0.01	-0.00	0.01	0.66 / 1.15
	Blk9	ϕ		10.11	1.25	0.01	1.25	8.00 / 12.72
		τ^2		0.12	0.00	0.02	0.00	0.02 / 0.26
		λ^2		0.89	0.01	-0.02	0.02	0.64 / 1.12
	Blk16	ϕ		9.62	1.55	-0.48	1.99	7.41 / 12.16
		τ^2		0.12	0.01	0.02	0.01	0.01 / 0.30
		λ^2		0.88	0.02	-0.03	0.02	0.61 / 1.13
	Blk25	ϕ		9.25	2.14	-0.85	3.26	6.64 / 12.21
		τ^2		0.13	0.01	0.03	0.01	0.01 / 0.32
		λ^2		0.87	0.02	-0.03	0.02	0.60 / 1.14
	JK	ϕ		8.95	2.18	-1.15	3.84	6.34 / 12.13
		τ^2		0.10	0.01	0.03	0.08	-0.06 / 0.27
		λ^2		0.91	0.02	-0.01	0.03	0.66 / 1.15
	Scenario 2	EVD	τ^2	0.20	0.20	0.01	-0.01	0.01
λ^2			0.80	0.80	0.02	-0.00	0.02	0.55 / 1.07
ϕ			10.07	10.08	1.76	0.01	1.76	7.58 / 13.20
MVN		τ^2		0.20	0.01	-0.01	0.01	0.05 / 0.35
		λ^2		0.81	0.02	0.00	0.02	0.55 / 1.08
		ϕ		10.08	1.90	0.01	1.91	7.44 / 13.37
Blk9		τ^2		0.22	0.01	0.02	0.01	0.07 / 0.38
		λ^2		0.78	0.02	-0.02	0.02	0.54 / 1.04
		ϕ		9.66	1.92	-0.40	2.41	7.17 / 12.49
Blk16		τ^2		0.22	0.01	0.02	0.01	0.05 / 0.40
		λ^2		0.78	0.02	-0.03	0.02	0.52 / 1.04
		ϕ		9.41	2.26	-0.66	3.23	6.77 / 12.75
Blk25		τ^2		0.23	0.01	0.03	0.01	0.05 / 0.44
		λ^2		0.77	0.02	-0.04	0.02	0.50 / 1.05
		ϕ		9.06	2.36	-1.00	3.87	6.33 / 12.54
JK		τ^2		0.20	0.02	-0.06	0.53	-0.01 / 0.41
		λ^2		0.80	0.02	0.03	0.17	0.53 / 1.08
		ϕ		10.06	9.16	-2.22	719.05	6.63 / 13.49
Scenario 3	EVD	τ^2	0.30	0.29	0.01	-0.01	0.01	0.10 / 0.48
		λ^2	0.70	0.70	0.02	0.00	0.02	0.44 / 0.98
		ϕ	10.16	10.17	2.73	0.00	2.74	7.00 / 14.16
	MVN	τ^2		0.29	0.01	-0.01	0.01	0.10 / 0.48
		λ^2		0.70	0.02	0.00	0.02	0.44 / 0.99
		ϕ		10.21	2.89	0.05	2.89	7.00 / 14.42
	Blk9	τ^2		0.31	0.01	0.01	0.01	0.11 / 0.51
		λ^2		0.68	0.02	-0.02	0.02	0.42 / 0.96
		ϕ		9.80	3.19	-0.37	3.90	6.87 / 13.67
	Blk16	τ^2		0.32	0.01	0.02	0.01	0.10 / 0.54
		λ^2		0.67	0.02	-0.03	0.02	0.39 / 0.96
		ϕ		9.57	4.10	-0.59	5.16	6.39 / 14.06
	Blk25	τ^2		0.32	0.01	0.02	0.02	0.10 / 0.55
		λ^2		0.68	0.02	-0.02	0.02	0.39 / 0.97
		ϕ		9.12	3.35	-1.05	5.26	6.00 / 13.36
	JK	τ^2		0.30	0.02	-0.02	0.10	0.04 / 0.56
		λ^2		0.70	0.03	0.02	0.08	0.39 / 1.01
		ϕ		10.17	8.30	0.36	76.34	6.13 / 14.20

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk9, 블록 9개; Blk16, 블록 16개; Blk25, 블록 25개; JK, 잭나이프방법;

의 편향과 평균제곱오차는 전통적인 붓스트랩에서 가장 작았으며, 블록의 개수가 25인 블록 붓스트랩에서 편향과 평균제곱오차가 가장 크게 나타났다. 또한 표 8의 지수모형에서의 결과와 같이 가우시안모형

표 12: 구형모형의 모의실험 결과(Scenario 4 ~ 6).

	Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
Scenario 4	EVD	τ^2	0.10	0.10	0.00	-0.00	0.00	0.03 / 0.17
		λ^2	0.90	0.89	0.02	-0.01	0.02	0.61 / 1.22
	MVN	ϕ	20.01	19.91	3.24	-0.10	3.25	16.18 / 24.06
		τ^2		0.10	0.00	-0.00	0.00	0.03 / 0.17
		λ^2		0.89	0.02	-0.01	0.02	0.62 / 1.22
	Blk9	ϕ		19.93	3.36	-0.08	3.38	16.19 / 24.12
		τ^2		0.12	0.00	0.03	0.00	0.04 / 0.22
		λ^2		0.89	0.03	-0.01	0.03	0.61 / 1.26
	Blk16	ϕ		17.70	7.09	-2.31	13.40	12.82 / 22.96
		τ^2		0.13	0.00	0.04	0.00	0.03 / 0.25
		λ^2		0.89	0.03	-0.02	0.03	0.60 / 1.29
	Blk25	ϕ		16.56	7.47	-3.45	21.28	11.81 / 22.07
		τ^2		0.14	0.00	0.04	0.01	0.03 - 0.28
		λ^2		0.85	0.03	-0.05	0.03	0.57 / 1.22
	JK	ϕ		14.95	7.03	-5.06	35.14	10.49 / 21.01
		τ^2		0.10	0.00	-0.02	0.03	0.01 / 0.18
		λ^2		0.90	0.03	-0.06	1.17	0.61 / 1.19
		ϕ		20.00	20.59	-3.42	1232.83	15.32 / 24.67
Scenario 5	EVD	τ^2	0.20	0.20	0.00	-0.00	0.00	0.11 / 0.29
		λ^2	0.80	0.80	0.02	-0.01	0.02	0.52 / 1.12
	MVN	ϕ	20.20	20.15	5.15	-0.05	5.15	15.64 / 25.49
		τ^2		0.20	0.00	-0.00	0.00	0.11 / 0.29
		λ^2		0.80	0.02	-0.01	0.02	0.52 / 1.12
	Blk9	ϕ		20.10	5.13	-0.09	5.14	15.57 / 25.42
		τ^2		0.22	0.00	0.02	0.00	0.11 / 0.33
		λ^2		0.78	0.02	-0.02	0.02	0.53 / 1.08
	Blk16	ϕ		17.29	9.01	-2.91	20.65	12.07 / 23.34
		τ^2		0.24	0.00	0.04	0.01	0.11 / 0.37
		λ^2		0.78	0.02	-0.03	0.02	0.51 / 1.09
	Blk25	ϕ		16.58	7.42	-3.61	23.14	11.60 / 22.22
		τ^2		0.24	0.01	0.04	0.01	0.11 / 0.39
		λ^2		0.76	0.02	-0.04	0.02	0.50 / 1.07
	JK	ϕ		15.31	7.39	-4.89	33.60	10.47 / 21.76
		τ^2		0.20	0.00	-0.03	0.01	0.09 / 0.31
		λ^2		0.80	0.04	0.18	1.60	0.53 / 1.07
		ϕ		20.02	31.10	2.87	1442.00	14.52 / 25.52
Scenario 6	EVD	τ^2	0.30	0.30	0.00	-0.00	0.00	0.19 / 0.41
		λ^2	0.70	0.69	0.02	-0.01	0.02	0.43 / 1.00
	MVN	ϕ	20.00	19.92	6.94	-0.08	6.96	14.79 / 26.29
		τ^2		0.30	0.00	-0.00	0.00	0.19 / 0.41
		λ^2		0.69	0.02	-0.01	0.02	0.43 / 1.00
	Blk9	ϕ		19.91	6.99	-0.09	7.01	14.79 / 26.24
		τ^2		0.32	0.00	0.02	0.01	0.19 / 0.44
		λ^2		0.67	0.02	-0.02	0.02	0.43 / 0.93
	Blk16	ϕ		17.54	9.80	-2.46	19.65	12.07 / 24.21
		τ^2		0.33	0.01	0.03	0.01	0.17 / 0.48
		λ^2		0.66	0.02	-0.03	0.02	0.40 / 0.94
	Blk25	ϕ		16.43	9.40	-3.56	25.42	10.74 / 22.66
		τ^2		0.34	0.01	0.04	0.01	0.17 / 0.50
		λ^2		0.65	0.02	-0.04	0.02	0.40 / 0.94
	JK	ϕ		15.23	9.29	-4.77	34.56	9.93 / 22.11
		τ^2		0.30	0.00	-0.01	0.01	0.17 / 0.43
		λ^2		0.70	0.02	0.00	0.04	0.46 / 0.93
		ϕ		20.00	20.00	-0.21	188.69	14.26 / 25.73

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk9, 블록 9개; Blk16, 블록 16개; Blk25, 블록 25개; JK, 잭나이프방법;

에서도 블록 붓스트랩에서 블록의 개수가 많아질수록 세 가지 모수에 대한 편향과 평균제곱오차가 커지는 경향이 있었다.

마지막으로 표 11과 표 12의 구형모형의 모의실험 결과, 거리(ϕ)의 편향과 평균제곱오차는 전통적 붓스트랩에서 가장 작았으며, 잭나이프 방법의 경우 평균제곱오차가 매우 큰 것을 확인할 수 있다. 다른 모형에서와 마찬가지로 구형모형에서도 블록 붓스트랩에서 블록의 개수가 많아질수록 편향과 평균제곱오차가 큰 것을 확인할 수 있다.

그리고 세 가지 모형에 대해 거리가 짧은 경우(시나리오 1, 2, 3)와 거리가 긴 경우(시나리오 4, 5, 6)에서 모두 뭉치(τ^2)가 커지고 부분 문턱(λ^2)가 작아질수록 세 가지 모수에 대한 편향과 평균제곱오차는 크며 신뢰구간의 길이도 더 긴 것을 확인할 수 있다. 또한 거리가 짧은 경우에 비해 거리가 긴 경우의 편향과 평균제곱오차, 신뢰구간의 길이가 더 큰 것을 확인할 수 있다.

또한 같은 시나리오 하에서 모형들간의 비교를 해보면, 뭉치(τ^2), 부분 문턱(λ^2)의 추정 결과는 모두 비슷한 것을 확인할 수 있다. 그러나 거리(ϕ)의 경우 지수모형과 가우시안모형에 비해 구형모형의 분산, 편향, 평균제곱오차가 큰 것을 확인할 수 있다. 이는 시나리오 설정에 있어서 각 모형의 특성에 따라 거리의 초기값을 다르게 주었기 때문이다. 지수모형과 가우시안모형의 거리의 초기값은 10/3이며 구형모형의 거리의 초기값은 10으로 구형모형의 거리의 초기값이 더 크기 때문에 구형모형의 거리가 지수모형과 가우시안모형에 비해 잘 추정이 되지 못하였다.

3.4 실증연구

3.2절에서 제안한 방법으로 2.2절에서 소개한 기상청의 풍속자료에 적용하였다. 이 때, 전통적인 붓스트랩 방법에서는 잔차의 붓스트랩

표 13: 풍속자료의 모수추정 결과(REML).

Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
EVD	τ^2	0.06	0.06	0.00	-0.00	0.00	0.05 / 0.07
	λ^2	0.06	0.06	0.00	0.00	0.00	0.02 / 0.10
	ϕ	118.50	113.94	374.53	-4.56	394.61	76.20 / 146.97
MVN	τ^2		0.06	0.00	-0.00	0.00	0.05 / 0.07
	λ^2		0.06	0.00	-0.00	0.00	0.03 / 0.10
	ϕ		114.51	401.91	-3.99	417.05	79.25 / 150.85
Blk	τ^2		0.08	0.00	0.02	0.00	0.05 / 0.12
	λ^2		0.03	0.00	-0.02	0.01	0.01 / 0.08
	ϕ		127.17	557177.59	8.68	556136.26	29.38 / 166.71
JK	τ^2		0.06	0.00	-0.00	0.00	0.04 / 0.08
	λ^2		0.06	0.00	-0.00	0.00	0.02 / 0.09
	ϕ		118.49	30.42	-4.57	51.26	107.68 / 129.30

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk, 블록 붓스트랩; JK, 잭나이프방법;

표본을 만든 뒤, 식 (2.9)의 평균함수부분과 더한 뒤 2.4.1절에서 분석한 방법으로 모형을 적합하여 공분산함수의 모수들을 추정하였다. 그러나 블록 붓스트랩의 경우 자료에서 블록을 나누게 되면 추세에 대한 효과가 사라지게 되므로 평균함수부분은 고려하지 않는 것이 바람직하다고 판단되어 잔차만을 가지고 공분산함수의 모수들을 추정하였다. 잭나이프 방법에서는 새로 만들어진 자료를 가지고 2.4.1에서의 방법으로 모수 추정을 하였다.

그리고 블록 붓스트랩의 블록의 수는 식 (3.1)에 의해 구했다. 이때, 전체 460개 지점 중에 육지와 바다의 경계를 제외한 나머지 249개의 지점을 사용하였으며 최대 거리 비율(r)은 0.5, 구간의 수(a)는 10개, 한 개의 구간 안의 평균적 점의 수(b)를 10개로 하여 한 블록 내 관측값의 수(x)를 20개로 하고 블록의 수를 12개로 하였다.

표 13는 잔차최대우도추정법을 이용해 풍속자료의 모수를 추정한 결과이다. 표 13의 뭉치(τ^2)와 부분 문턱(λ^2)은 전통적 붓스트랩 방법과 잭나이프 방법에서는 모수의 추정이 비교적 잘 되었다. 그러나 블록

표 14: 풍속자료의 모수추정 결과(ML).

Method	Param.	Truth	Mean	Var.	Bias	MSE	95% C.I.
EVD	τ^2	0.06	0.06	0.00	-0.00	0.00	0.05 / 0.07
	λ^2	0.05	0.04	0.00	-0.01	0.00	0.01 / 0.06
MVN	ϕ	113.86	95.39	395.59	-18.46	735.74	50.13 / 128.77
	τ^2		0.06	0.00	-0.00	0.00	0.05 / 0.08
	λ^2		0.03	0.00	-0.01	0.00	0.01 / 0.06
Blk	ϕ		95.25	404.21	-18.61	749.55	50.55 / 126.61
	τ^2		0.08	0.00	0.02	0.00	0.05 / 0.12
	λ^2		0.03	0.00	-0.02	0.00	0.00 / 0.07
JK	ϕ		67.43	1196.39	-46.43	3349.46	28.45 / 149.39
	τ^2		0.06	0.00	0.00	0.00	0.04 / 0.08
	λ^2		0.05	0.00	-0.02	0.00	0.02 / 0.07
	ϕ		113.75	420.77	-50.28	2948.84	73.54 / 153.95

Param., 모수; Var., 표본분산; EVD, 고유값분해방법; MVN, 다변량정규분포에서 난수생성방법; Blk, 블록 붓스트랩; JK, 잭나이프방법;

붓스트랩의 경우 뭉치는 다소 과대추정되었고, 부분 문턱은 다소 과소 추정 된 것을 확인할 수 있다. 그리고 거리(ϕ)는 전통적 붓스트랩에서 편향과 평균제곱오차가 가장 작으며, 블록 붓스트랩에서는 편향과 평균제곱오차가 가장 크게 나타났다.

그리고 표 14은 일반최대우도추정법을 이용해 풍속자료의 모수 추정을 한 결과이다. 표 14에서도 전통적 붓스트랩과 잭나이프 방법에서는 뭉치와 부분 문턱의 모수 추정이 비교적 잘 되었으나, 블록 붓스트랩의 경우에는 뭉치는 과대추정 되었고, 부분 문턱은 과소추정 되었다. 또한 거리에서도 전통적 붓스트랩에서 편향과 평균제곱오차가 작았으며, 잭나이프 방법의 편향이 가장 크고 블록 붓스트랩의 평균제곱오차가 가장 크게 나타났다. 두 추정법을 비교해보면, 전통적 붓스트랩 방법과 잭나이프 방법에서는 잔차최대우도추정법(표 13)의 편향과 평균제곱오차가 일반최대우도추정법(표 14)의 편향과 평균제곱오차보다 더 작으며, 블록 붓스트랩 방법은 일반최대우도추정법(표 14)의 편향과 평균제곱오차가 잔차최대우도추정법(표 13)보다 더 작게 나타

난 것을 확인할 수 있다.

3.5 결론

본 연구에서는 지리통계자료에 대하여 공간적 연관성을 나타내는 모수의 구간추정을 하였다. 모수의 구간추정을 하기 위하여 붓스트랩 방법과 잭나이프 방법을 사용하였다. 붓스트랩 방법으로는 분포의 가정 없이 사용하는 비모수적 붓스트랩을 이용해 잔차의 공분산행렬을 고유값 분해하여 붓스트랩을 하였고, 분포를 가정하고 사용하는 모수적 붓스트랩을 이용해 잔차가 따르는 다변량 정규분포에서 난수 생성을 통해 붓스트랩을 하였다. 또한, 공간 자료에서 공간적 연관성을 유지시키기 위해 블록 붓스트랩 방법을 이용하였다.

이를 확인하기 위하여 공분산함수는 지수모형, 가우시안모형, 구형모형을 이용하였으며, 6가지 시나리오에 따라 제안한 붓스트랩 방법과 잭나이프 방법을 토대로 모의실험을 하였다. 모의실험의 결과 공분산함수에 상관없이 모든 방법에서 멩치와 부분 문턱의 결과는 비슷했다. 그러나 거리의 경우에는 전통적 붓스트랩 방법이 블록 붓스트랩 방법보다 더 효율적임을 보였다. 또한, 블록의 개수가 많아질수록 거리에 대한 추정을 잘 하지 못하였으며, 원자료의 거리가 클수록 블록 붓스트랩은 적합하지 않았다. 이에 블록 붓스트랩은 원자료의 거리가 짧으며 블록의 개수가 클 때 적합하였다.

향후에는 블록 붓스트랩의 블록을 직사각형의 형태가 아닌 다른 형태를 고려하여 블록 붓스트랩을 하고자 한다. 또한, 본 연구에서는 모의실험을 일반최대우도추정법으로만 하였는데 잔차최대우도추정법과 가중최소제곱법으로도 모의실험을 해볼 필요가 있다고 판단된다.

제 4 장

결론

본 연구에서는 서로 다른 관측망에서 얻어진 자료를 이용하여 계층모형을 구축하고, 공간적 연관성을 나타내는 세 가지 모수에 대한 신뢰구간을 구하기 위해 붓스트랩과 잭나이프 방법을 사용했다. 먼저 공간자료의 계층모형은 두 관측망의 다른 자료가 서로 연관성이 클 때, 한 자료를 이용해 다른 자료를 분석하여 신뢰성 있는 예측을 할 수 있도록 하는 것이다. 이를 확인하기 위하여 지상기상관측망의 풍속자료를 도시대기측정망의 이산화황을 분석하는데 모형에 포함시켜 계층모형을 구축하였다. 이산화황이 측정되는 도시대기관측망은 우리나라의 도시지역에서만 관측되는 한계가 있는데, 우리나라 전역에서 관측되는 풍속자료를 이용한다면 모형의 예측력이 높아질 것은 명백하다고 판단했다.

두 자료를 이용한 계층모형의 결과 예측값이 다소 과소평가된 부분이 있었지만 예측값의 범위는 원자료와 비슷했음을 확인할 수 있었다. 그러나 풍속자료는 고도가 높은 지역에서 영향을 많이 받으므로 이산화황의 예측값 또한 고도가 높은 지역에서 크게 나타나는 경향이 있었다. 또한, 계층모형을 사용한 것과 사용하지 않은 모형의 비교를 위해 AIC, BIC, 교차검증방법, 급내상관계수, 예측오차제곱합을 사용했다. 그 결과 전통적 선형회귀모형보다는 공간적 선형회귀모형이 더 신뢰성 있는 예측 성능을 보였고, 계층모형을 사용하는 것이 그렇지

않은 경우에 비해 AIC와 교차검증방법에서는 예측력이 높음을 확인할 수 있었다. 또한, 이산화황을 예측하는데 크리깅 뿐만 아니라 역거리가중치법을 사용했다. 두 가지 방법을 이용해 예측된 이산화황을 비교하기 위해 급내상관계수와 예측오차제곱합을 사용했다. 그 결과 두 가지 측면에서 모두 역거리가중치법보다 크리깅에서 신뢰성 있는 예측을 하였다.

또한, 일반적으로 세미베리오그램을 추정할 때 공간적 연관성을 나타내는 모수의 신뢰구간을 구할 수 없기 때문에 이들 모수에 대한 구간추정을 위해 붓스트랩과 잭나이프 방법을 사용했다. 전통적 붓스트랩 방법 뿐만 아니라 블록 붓스트랩 방법을 제안했다. 특히, 블록 붓스트랩 방법에서는 자료의 형태가 직사각형일 때와 그렇지 않은 경우를 나누어 제안했다.

제안된 여러가지 방법을 확인하기 위해 일반최대우도추정법을 이용해 지수모형, 가우시안모형, 구형모형을 고려하여 6가지 시나리오에 따른 붓스트랩과 잭나이프 방법을 이용해 모의실험을 하여 모수들의 구간추정을 했다. 그 결과 공분산함수에 상관없이 모든 방법에서 멩치와 부분 문턱의 결과는 비슷했으나 거리는 블록 붓스트랩보다 전통적 붓스트랩의 결과가 더 효율적이었다. 원자료의 거리가 클수록 블록 붓스트랩 방법보다는 전통적 붓스트랩 방법이 더 적합한 것을 확인할 수 있었고, 블록 붓스트랩에서 블록의 개수가 많아질수록 효율성이 떨어지는 것을 확인할 수 있었다. 이에 추후에는 블록을 나눌 때, 보다 공간적 연관성을 유지시키기 위해 블록의 모양을 직사각형이 아닌 다른 형태로 고려하여 블록 붓스트랩을 할 필요가 있다.

참고 문헌

- [1] 김규곤, 최승배 (2000). 환경자료를 이용한 공간통계 분석법의 예측력 연구, 한국자료분석학회지, **4**, 499-510.
- [2] 김동휘, 류동우, 최영민, 이우진 (2010). 인천 송도지역 지층분포 추정을 위한 크리깅과 역거리가중치법의 적용, 한국지반공학회 논문집 **26**, 5-19.
- [3] 김병선, 구자용, 최진무 (2010). Regression-Kriging 모형을 이용한 인구분포 추정에 관한 연구, 대한지리학회지, **45**, 806-819.
- [4] 기상청 자료관리서비스팀 (2007). 풍력자원지도 개발 연구보고서, 기상청.
- [5] 기상청 홈페이지. <http://www.kma.go.kr>.
- [6] 박노옥, 장동호 (2008). 수치표고모델과 다변량 크리깅을 이용한 기온 및 강수 분포도 작성, 대한지리학회지, **43**, 1002-1015.
- [7] 신기일, 최봉호, 이상은 (2007). 공간 통계 활용에 따른 소지역 추정법의 평가, 응용통계연구, **20**, 229-244.
- [8] 조재영, 최승배, 김규곤 (2001). 일반통계에 대한 공간통계 방법의 예측성능에 관한 연구, 한국자료분석학회지, **3**, 41-49.
- [9] 정승환, 박만식, 김기환 (2010). 풍속 자료의 공간예측, 응용통계연구, **23**, 345-356.
- [10] 정지용, 진서훈, 박만식 (2008). 공간적 상관구조를 포함하는 선형회귀모형을 이용한 강수량 자료 분석, 응용통계연구, **21**, 1093-1107.
- [11] 조홍래, 정종철, (2006). 강우자료에 대한 공간보간 기법의 적용, The Journal of GIS Association of Korea, **14**, 29-41.

- [12] 주재희, 황인조 (2012). 강하분진의 침적 특성파악을 위한 통계학적 해석과 공간분포 분석, 한국대기환경학회지 **28**, 294-305.
- [13] 최지은, 박만식 (2013). 다양한 관측네트워크에서 얻은 공간자료들을 활용한 계층모형 구축, 응용통계연구, **26**, 93-109.
- [14] 한진석, 김영미, 안준영, 공부주, 최진수, 이상욱, 이석조 (2006). 장거리이동 대기오염물질 이동량의 공간적 분포와 변화 추이(1997 ~ 2004 년), 한국대기환경학회지 **22**, 99-106.
- [15] 허태영, 서의훈, 권원태 (2004). 세미베리오그램 모형을 적용한 남한지역 강수량 자료의 공간분석, 한국자료분석학회지, **6**, 473-491.
- [16] 환경부 (2007). 대기환경연보.
- [17] Banerjee, S., Gelfand, A. E., and Carlin, B. P. (2004). Hierarchical Modeling and Analysis for Spatial Data, Boca Raton: Chapman & Hall/CRC.
- [18] 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.
- [19] Cressie, N. A. C. (1993). Statistics for Spatial Data, New York: John Wiley & Sons.
- [20] Efron, B. (1979). Bootstrap methods : Another look at the jackknife, The Annals of Statistics, **7**, 1-26.
- [21] Goovaerts, P. (1997). geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
- [22] Heo, T. Y. and Park, M. S. (2009). Bayesian Spatial Modeling of Precipitation Data, The Korean Journal of Applied Statistics, **22**, 425-433.
- [23] J. M. Loh and M. L. Stein (2004). Bootstrapping a spatial point process, Statistica Sinica, **14**, 69-101.

- [24] J. M. Loh and M. L. Stein (2008). Spatial bootstrap with increasing observations in a fixed domain, *Statistica Sinica*, **18**, 667–688.
- [25] Mark, D. Varljen, Michael, J. Barcelona, and H. Allen Wehrmann (1999). A jackknife approach to examine uncertainty and temporal change the spatial correlation of a voc plume, *Environmental Monitoring and Assessment*, **59**, 31–46.
- [26] M. H. Quenouille (1949). Approximate Tests of Correlation in Time-Series, *Journal of the Royal Statistical Society. Series B (Methodological)*, **11**, 68–84.
- [27] Kang, C. W., Choi, S. B., and Yoo, S. H. (2008). A Spatial Bootstrap Method for Kriging Variance, *The Journal of Korea Data Analysis Society*, **10**, 1247–1254.
- [28] Lee, K. W., Park, N. W., Chi, K. H., and Kwon, B. D. (1999). GIS-based spatial integration and statistical analysis using multiple geoscience data sets : A case study for mineral potential mapping, *Journal of the Korean Society of Remote Sensing*, **15**, 91–105.
- [29] Liansheng. T, William R. S., Wayne A. W., and Richard F. G. (2006). *A Parametric Spatial Bootstrap*, Southern Methodist University
- [30] Lillegard, M., Engen. S., and Sæther. B. E. (2005). Bootstrap methods for estimating spatial synchrony of fluctuating populations, *OIKOS*, **109**, 345–356.
- [31] Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods For Spatial Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- [32] S. N. Lahiri (2003). *Resampling Methods for Dependent Data*, Springer.

Abstract

Hierarchical Model and Bootstrap of Spatial Data

Jieun Choi
Department of Statistics
The Graduate School
Sungshin Women's University

Geo-statistical data or point-referenced data are obtained from a wide variety of monitoring networks, where each of the measurements is only a marker of its monitoring location. In case that this kind of spatial dataset is to be analyzed, we can improve the prediction power much more when we combine various relevant spatial datasets measured from different sources of networks than we just utilize each of the datasets separately. In this paper, we consider the hierarchical spatial linear model, which can detect intra-network spatial association and make more reliable prediction by means of inter-network association between different datasets. For the modeling, we use two different sources of spatial measurements, Sulfur Dioxide(SO₂) from the urban air pollution monitoring and wind-speed data from the surface observation network. Also, we organize the prediction map with the results from the kriging and hierarchical model and compare them with the result of the inverse distance weighting method.

In the spatial linear regression model, we estimate the semi-variogram which contains several parameters representing the spatial association. In order to estimate the parameters, we use weighted least squares method (WLS), maximum likelihood estimation (ML), or restricted maximum likelihood estimation(REML). It is, however, impossible to obtain the confidence intervals of the parameters using the methods mentioned above. Here, we employ the parametric and the nonparametric bootstrapping methods and jackknife technique to obtain the confidence intervals of the parameters. Also we propose a block-based bootstrapping method and we provide some guidelines for determining how many blocks we consider. To compare the proposed methods with classical resampling methods, we do the simulation studies. We also perform a real data analysis by using wind speed data for the comparison.

Keywords : hierarchical model, kriging, inverse distance weighting, spatial association, bootstrap, jackknife