응용통계연구 (2008)
21(6), 1093–1107

공간적 상관구조를 포함하는 선형회귀모형을 이용한 강수량 자료 분석

정지용[1] · 진서훈[2] · 박만식[3]

[1]고려대학교 정보통계학과; [2]고려대학교 정보통계학과; [3]고려대학교 의학통계학교실

(2008년 8월 접수, 2008년 9월 채택)

요 약

매년 전 세계는 여러 자연재해로 인하여 많은 피해를 받고 있다. 그 중에서도 강수와 관련한 집중호우와 가뭄, 홍수, 상수원 부족 등으로 많은 손실을 입고 있다. 이러한 재해에 의한 피해를 줄이기 위해서는 기상에 대한 정확한 예측이 필요하다. 따라서 강수량에 대한 정확한 예측을 실시하여 수자원을 적절하게 이용하고 재해에 의한 피해를 줄이기 위하여 많은 연구가 진행되고 있다. 본 연구에서는 강수량을 측정하는 지상기상관측지점자료에 대해 공간적 상관구조를 포함하는 선형회귀모형(크리깅)을 고려하여 세미베리오그램을 기반으로한 최소제곱법과 코베리오그램을 기반으로한 최대우도추정방법으로 남한지역의 공간적 특성을 적절하게 파악할 수 있는 모형들을 찾고 이 모형들을 비교하였다. 공간적 선형회귀모형들에 대한 신뢰성을 검증하기 위하여 자동기상관측지점과 항공기상관측지점에서 측정된 실제값과 예측값을 비교하고 이를 바탕으로 강수량 예측에 관한 발전 및 개선방향에 대해 알아보았다.

주요용어: 세미베리오그램, 강수량, 범용크리깅, 공간적 선형회귀모형, 교차검증방법.

1. 서론

우리나라는 강수피해에 관한 대비와 적절한 수자원 활용을 위해 전국 76개 지상관측지점 및 자동기상관측지점과 항공기상관측지점을 이용하여 기상관측을 실시하고 있다. 자동기상관측지점은 기상관측소가 없는 곳에 설치되어 집중호우, 우박, 뇌우, 돌풍 등과 같은 국지적인 악기상 현상을 실시간으로 감시하고 있는데, 특히 산악지역이나 섬지역처럼 사람이 관측하기 어려운 곳에 설치되어 집중호우와 같은 돌발 악기상을 감시하며, 관측된 자료는 수치예보모형의 초기 입력 자료로 유용하게 사용된다. 항공기상관측지점은 항공기의 안전운항을 위해서 각 공항이 항공기상종합자동관측장비를 설치하여 비행장과 비행정보구역의 기상상태를 관측, 감시하는 역할을 하는 곳이다.

우리나라의 강수는 1년 중 수개월에 집중되는 경향이 있는데 대체적으로 연중 7–8월에 내리는 강수량이 전체 강수량의 41%–56% 정도이며 6–9월 사이에 내리는 강수량이 전체의 76%–86%에 달한다. 이러한 점을 감안할 때 우기가 아닐 때는 가뭄이 발생하고 우기에 해당하는 수개월 동안에는 홍수, 산사태 등이 집중적으로 발생되고 있다. 이러한 지리학적 요건으로 우리나라는 집중호우에 의한 홍수 피해와 물 부

[1](339-700) 충청남도 연기군 조치원읍 서창리 208, 고려대학교 자연과학대학 정보통계학과, 석사과정.
E-mail: jjiyongi@korea.ac.kr
[2](339-700) 충청남도 연기군 조치원읍 서창리 208, 고려대학교 자연과학대학 정보통계학과, 조교수.
E-mail: seohoon@korea.ac.kr
[3]교신저자: (136-701) 서울시 성북구 안암동 5가1 고려대학교 의과대학 의학통계학교실 및 의과학연구원, (유전체 및 단백체 독성연구소), 연구조교수. E-mail: bayesia@korea.ac.kr

족이라는 큰 과제를 가지고 있다. 현재 환경자료나 지리자료의 공간적 분석에 대한 연구는 활발하게 이루어지고 있으나 기상과 관련한 자동관측지점에 대한 공간적 분석에 대한 연구는 미미한 실정이다. 따라서 본 연구에서 집중호우로 인해 많은 피해를 받고 있는 7-9월을 중심으로 지상기상관측지점의 강수량 자료를 공간통계학에서의 선형모형으로 분석하고 이를 바탕으로 전국에 퍼져 있는 자동관측지점들의 강수량을 예측한 후 실제값과 비교하여 보겠다.

오광중 등 (1998)는 대기오염농도와 기상인자와의 관련성을 다변량 통계기법 중 인자분석 및 다중회귀분석을 계절별로 실시하여 상호관련성을 규명하기 위한 다중회귀모형을 연구하였고, 유성모와 엄익현 (1999)은 강우량 자료를 이용하여 세미베리오그램(semi-variogram)에 의한 모수 추정과 공간이상값 탐지방법에 대하여 연구하였다. 신만용 등 (1999)은 거리역산 가중법을 사용하여 미관측 지점에 대한 기상정보(온도)를 예측하였으며 전국의 약 500개소의 자동기상관측지점에서 수집된 자료와 위성자료를 이용하여 모형의 실용성을 검정하였다. 조재영 (2001)은 공간통계 분석법과 일반통계 분석법을 적용하여 미지의 위치에 대한 예측력을 비교하였다. 조재영 등 (2001)은 공간 통계적 방법인 범용크리깅 방법(universal kriging)을 일반 통계적 방법 중 비모수적 방법인 국소적 회귀분석방법과 일반화가법모형(generalized additive model: GAM)과 비교하여 공간적 상관성이 있는 환경자료에 대하여 공간 통계적 방법이 일반 통계적 방법보다 공간예측력이 좋다는 것을 보였다. 이지영과 황철수 (2002)는 다중회귀분석을 사용하여 속성들의 분포를 계산하되 입지적 특성을 통합하기가 어렵기 때문에 공간통계분석으로 이 모형을 개선하고자 하였으며 장지희 (2003)는 거리역수 크리깅 방법보다는 일반크리깅(ordinary kriging) 방법의 결과가 좋은 결과가 나타났다는 연구를 발표하였고 최승배 등 (2004)은 동일한 상황과 목적하에 있다고 가정하고 공간변이성에 대한 추정량들의 예측력을 비교하는 연구를 하였다. 허태영 등 (2004)는 강수량 자료를 이용하여 공간 예측 성능을 향상시킬 수 있는 로버스트(robust)한 세미베리오그램에 대해 소개하고 특정한 세미베리오그램 모형 하에서 일반크리깅 방법을 사용하여 공간예측을 실시하였다. 김선우 등 (2005)은 일산화탄소 자료를 이용하여 크리깅 방법과 지리적 가중회귀 분석방법과의 예측정도를 비교하는 연구를 실시하였다.

본 연구에서는 세미베리오그램 모형의 모수를 추정하는 여러 가지 방법을 소개하고, 실제 강수량 자료를 사용하여 추정방법들을 비교하며 실증적 검정을 위해 자동기상관측지점에 대해 공간 예측하여 실제값과 비교연구를 수행하고자 하였다. 모수의 추정방법으로서 가중최소제곱방법(weighted least squares estimation)과 잔차최대우도추정방법(restricted maximum likelihood estimation)을 고려하였다. 그리고 공간적 상관구조를 적합하기 위한 세미베리오그램 혹은 공분산(covariance)모형으로는 공간통계학에서 가장 많이 사용되는 가우시안모형(Gaussian model), 지수모형(exponential model), 구형모형(spherical model) 만을 고려하였다. 모수 추정방법들에 의해 가장 잘 적합되는 세미베리오그램 모형(혹은 공분산 모형)을 선택하여 자동관측지점에서의 최종적인 공간예측에 사용하였다. 모형의 선택기준으로서 최소제곱방법은 잔차제곱합(residual sum of squares)을 사용하였으며 우도함수방법은 로그우도값(log likelihood)을 사용하였다. 공간예측에 대한 성능을 평가하기 위해서 크리깅(kriging) 방법의 타당성을 검증하는 교차검증방법(leave-one-out cross-validation)을 이용하였으며 각 모형들의 비교는 예측오차 제곱합(prediction error sum of squares: PRESS)을 기준으로 이루어 졌다. 또한, 자동관측지점과 항공관측지점을 타당성 자료(validation data)로 간주하여 평균제곱예측오차(mean squared Errors of Prediction: MSEP)로 각 모형의 성능을 평가하였다.

본 논문에서는 2장에서 공간분석을 위한 공간통계학의 개괄적인 내용 및 세미베리오그램 모형과 추정방법들을 소개하였으며 공간예측방법으로서 범용크리깅 방법에 관하여 간략히 소개하였다. 3장에는 현재 우리나라 남한전역의 76개 지상기상관측지점 중 제주도 및 섬지역을 제외한 69개의 관측지점과 약 340개의 자동관측지점 및 9개의 항공기상관측지점에서 관측된 강수량 자료에 대한 기술통계적 분석을

하였으며, 4장에서는 실제 강수량자료를 가지고 추정방법들을 이용하여 공간예측력을 비교하고 실증적 차원에서 자동관측지점의 실제 관측치와 비교연구를 수행하였으며 공간적 변동성을 고려하지 않은 기법과의 비교를 실시하였다. 5장에서는 결론 및 향후 연구 과제에 대하여 논하였다.

2. 공간통계

공간자료는 지리통계 자료(geostatistical data), 격자자료(lattice data), 공간 점 패턴 자료(spatial point pattern data) 등 크게 3가지로 분류된다. 지리통계 자료는 연속적인 공간상의 고정된 위치에서 얻은 측정값들의 집합으로 그 위치는 일반적으로 공간상에서 연속이라고 가정한다. 이에 대한 예로 '강수량'과 '광산내에 있는 관심위치에서 측정된 무기물 농도' 등을 들 수 있다. 격자 자료는 일반적으로 연구자가 정한 공간상의 고정된 부분 지역들에서 측정된다. 부분 지역들의 형태는 동일하거나 동일하지 않을 수 있다. 동일한 형태의 격자자료는 위성으로부터 원격감지에 의해서 얻어진 자료를 들 수 있고 불규칙적인 격자데이터는 어느 한 도시에 있는 각 구에서 발생하는 암 발생률로 얻어진 자료를 예로 들 수 있다. 그리고 공간 점 패턴 자료는 위치 자체가 관심의 대상이 되는 변수로서 공간지역 내에서 관측된 유한한 위치의 집합으로 구성되며 그 예로 '산림지역에서 나무종류의 위치', '발생한 지진의 진원지 위치' 등을 예로 들 수 있다. 본 논문에서는 세 가지 자료형태 중에서 지리통계자료만을 다루기로 한다.

서로 다른 위치에서 측정된 자료는 일반적으로 자료의 독립성을 가정하게 된다. 그러나 공간적인 변인에 의해서 측정값들이 서로 영향을 받는다면 공간적인 상관구조를 고려한 분석방법을 이용하는 것이 보다 현실적일 것이다. 공간적인 변인으로는 관측지점 사이의 거리와 방향 등을 생각할 수 있는데 대부분의 공간자료들에 영향을 주는 변인 중 가장 영향력 있는 변인은 관측된 자료들 사이의 유클리디안 거리(Euclidean distance)이다. 따라서, 공간적인 변인인 거리에 의하여 영향을 받는다면 측정값들은 서로 위치가 가까울수록 강한 연관성을 보이고 멀어질수록 약한 연관성을 보이게 될 것이다. 그러므로 서로 다른 위치에서의 공간자료들이 서로 독립관계에 있지 않고 상관관계에 있을 때 사용할 수 있는 공간자료의 예측모형의 이론적인 기반이 필요하다고 볼 수 있다. 이 장에서 공간자료를 분석하는 데 주로 사용되는, 공간적 상관구조를 포함하는 선형회귀모형을 제시하고자 한다. 이 모형에 의한 접근방법은 흔히 크리깅이라고 불리운다.

2.1. 공간변수와 세미베리오그램

공간자료는 확률과정(stochastic process)의 $Z(\mathbf{s})$의 실현치이며, $\mathbf{s}$가 $\mathbb{D}$의 위치이고, $\mathbb{R}^d$는 d차원의 유클리드 공간이라 할 때, 확률벡터 $Z(\mathbf{s})$는 다음과 같이 표현된다.

$$\left\{ Z(\mathbf{s}) : \mathbf{s} \equiv (s^1, \ldots, s^d)' \in \mathbb{D} \subset \mathbb{R}^d \right\},$$

여기서 d는 일반적으로 1, 2, 3의 값을 가진다. 즉 공간자료는 $\{(Z_i, \mathbf{s}_i);\ i = 1, 2, \ldots, n\}$으로 표현될 수 있다. 공간자료 분석에서 관측된 자료는 다음과 같은 모형으로 표현할 수 있다.

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \quad i = 1, 2, \ldots, n, \tag{2.1}$$

여기서, $\mu(\mathbf{s})$는 방향(direction) 또는 추세(trend)와 같은 큰 변동(large-scale variation)을 의미하고, $\epsilon(\mathbf{s})$는 작은 변동(small-scale variation)을 의미하며, 오차항으로 생각할 수 있다. 즉 $\epsilon(\mathbf{s})$는 평균이 0이고 2차 정상성(second-order stationarity)을 만족하는 확률변수로서 공분산과 세미베리오그램를 가진다. $Z(\mathbf{s})$의 모델링은 $\mu(\mathbf{s})$와 $\epsilon(\mathbf{s})$를 각각 모형화하여 구할 수 있다.

본 논문에서 사용하고 있는 공간자료 $\{Z(s) : s \in D\}$가 다음과 같은 가정을 만족한다고 하자. 모든

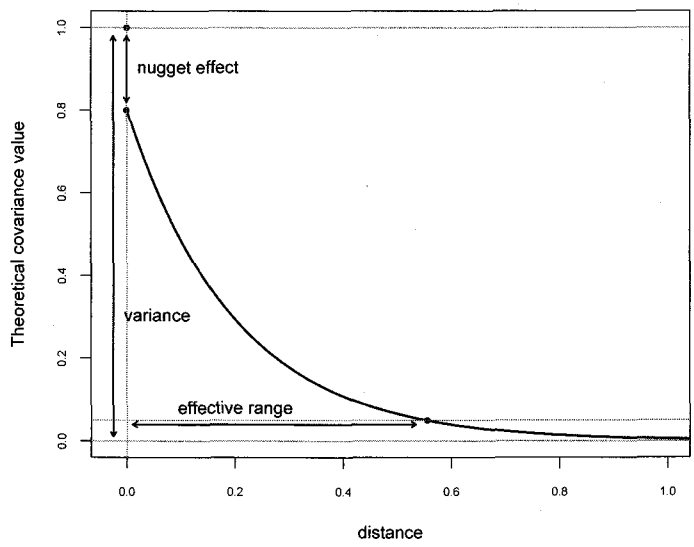


그림 **2.1.** 코베리오그램

$\mathbf{s}_i, \mathbf{s}_j \in \mathbb{D}$에 대하여

(1) $E(Z(\mathbf{s}_i)) = \mu$,

(2) $\mathrm{Var}(Z(\mathbf{s}_i)) = C(\mathbf{0}) = \sigma^2$,

(3) $\mathrm{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \equiv C(\mathbf{s}_i - \mathbf{s}_j) = C(\|\mathbf{s}_i - \mathbf{s}_j\|) < \infty$,

(4) 공간변수 $Z(\cdot)$는 정규분포를 따른다.

여기서, $\|\mathbf{s}_i - \mathbf{s}_j\|$는 두 지점 $\mathbf{s}_i$와 $\mathbf{s}_j$사이의 유클리디안 거리를 표현하고, 관측값들 사이의 상관관계를 나타내는 공분산은 거리만의 함수인 $C(\|\mathbf{s}_i - \mathbf{s}_j\|)$형태로 표현된다. $C(s_i - s_j)$의 특별한 경우인 $\|s_i - s_j\| = 0$일 때의 공분산은 주어진 자료의 분산 σ^2이 된다.

앞의 가정들에 의하여 두 공간변수의 차 $Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$는 평균이 0이고 분산이 $2C(0) - 2C(\|\mathbf{s}_i - \mathbf{s}_j\|) \equiv 2\gamma(\|\mathbf{s}_i - \mathbf{s}_j\|)$인 정규분포를 따른다. 여기서 $C(\|\mathbf{s}_i - \mathbf{s}_j\|)$와 $2\gamma(\|\mathbf{s}_i - \mathbf{s}_j\|)$를 각각 코베리오그램(covariogram)과 베리오그램(variogram)이라고 부르며, $\gamma(\|\mathbf{s}_i - \mathbf{s}_j\|)$를 세미베리오그램이라고 한다. 만약 세미베리오그램이 $\mathbf{s}_i - \mathbf{s}_j$의 거리 뿐만 아니라 방향에도 의존한다면, 불등방성(anisotropy)이라고 하며 $\gamma(\|\mathbf{s}_i - \mathbf{s}_j\|) \neq \gamma(\mathbf{s}_i - \mathbf{s}_j)$가 된다. $\mathbf{h}$를 두 지점 사이의 위치 차이를 나타내는 값이라고 할 때 $\mathbf{s}_j = \mathbf{s}_i + \mathbf{h}$라고 놓을 수 있다. 이때 코베리오그램, $C(\cdot)$과 세미베리오그램, $\gamma(\cdot)$ 사이에는 다음의 관계식이 성립한다. 모든 $i = 1, \ldots, n$에 대해서

$$\gamma(\mathbf{h}) = \frac{1}{2}E(Z(\mathbf{s}_i) - Z(\mathbf{s}_i + \mathbf{h}))^2 = C(0) - C(\mathbf{h}).$$

이제, 코베리오그램(혹은 세미베리오그램)이 가지는 주요 모수들을 정의하고 그 의미를 살펴보기로 하자. 일반적으로 코베리오그램은 그림 2.1과 같이 문턱(sill = variance + nugget effect), 뭉치효과(nugget effect), 범위(range 또는 effective range) 등 3개의 모수를 가지는 거리의 함수식 형태를 가진다. 직관적으로 코베리오그램의 값은 두 좌표의 공간상의 거리가 짧을수록 크고, 멀수록 작아지며 마침내 0에 수렴한다고 생각할 수 있다. 따라서 $\lim_{\|h\|\to\infty}\gamma(\mathbf{h}) = C(\mathbf{0})$이라고 가정할 수 있고 이

를 문턱(sill)이라고 부른다. 즉, 문턱은 관측변수의 분산이 된다. $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$이고, $\gamma(\mathbf{0}) = 0$이다. $\lim_{\|h\|\to 0}\gamma(\mathbf{h})$를 뭉치효과(nugget effect)라 정의하며, $Z(s)$와 $Z(s+h)$가 더이상 공간적인 상관관계에 있지 않는 최소거리를 범위(range)라고 한다. 주어진 자료에서 얻게 되는 최소거리보다 짧은 거리에서는 세미베리오그램에 대한 정보가 존재하지 않으므로 뭉치효과가 나타나게 되며, 세미베리오그램은 거리가 0인 점에서 불연속점을 가지게 된다. 강수량이 범위 이상의 거리만큼 떨어져 있을 때 강수량이 독립적인 관계를 나타내게 되고, 이처럼 범위보다 먼 거리에서의 세미베리오그램 값은 분산과 같게 되는데 이 값을 문턱이라고 한다.

2.2. 세미베리오그램의 추정

관측된 자료로부터 계산된 세미베리오그램은 일반적으로 경험적 세미베리오그램이라고 하며 공간적 보간(spatial interpolation)을 위해서는 관측되지 않은 지점에 대한 세미베리오그램의 계산이 필요하다. 따라서 경험적 세미베리오그램을 함수화한 이론적 세미베리오그램을 고려하게 된다. 공간자료 분석에서 자주 사용되는 이론적 세미베리오그램 모형에는 가우시안모형, 지수모형, 구형모형 등이 있다.

(1) 구형모형(spherical model)

$$\gamma(\mathbf{h}:\boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = \mathbf{0}, \\ \theta_0 + \theta_1 \left\{ \dfrac{3\|h\|}{2\theta_2} - \dfrac{1}{2}\left(\dfrac{\|h\|}{\theta_2}\right)^3 \right\}, & 0 < \|\mathbf{h}\| \le \theta_2, \\ \theta_0 + \theta_1, & \text{otherwise.} \end{cases}$$

(2) 가우시안모형(Gaussian model)

$$\gamma(\mathbf{h}:\boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = \mathbf{0}, \\ \theta_0 + \theta_1 \left\{ 1 - \exp\left(-\dfrac{\|h\|^2}{\theta_2^2}\right) \right\}, & \text{otherwise.} \end{cases}$$

(3) 지수모형(exponential model)

$$\gamma(\mathbf{h}:\boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = \mathbf{0}, \\ \theta_0 + \theta_1 \left\{ 1 - \exp\left(-\dfrac{\|h\|}{\theta_2}\right) \right\}, & \text{otherwise.} \end{cases}$$

여기서, $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)'$이고 양의 값을 가진다. θ_0는 뭉치효과를 $\theta_0 + \theta_1$은 문턱을, θ_1은 부분 문턱(partial sill)을 그리고 θ_2는 범위를 나타낸다. 세미베리오그램의 추정은 세미베리오그램이 가지고 있는 모수들의 추정이며, 적절한 이론적 세미베리오그램의 모수들을 가지고 크리깅 예측의 가중치를 결정하기 때문에 이론적 베리오그램의 모형적합은 공간예측의 단계에서 매우 중요하다.

2.3. 크리깅

공간자료 분석에서 가장 중요하게 다루는 문제는 관측된 지점의 자료를 근거로 하여 관측되지 않은 지점의 값을 예측하는 것이다. 공간상의 상관관계를 고려하는 예측 방법은 공간통계학에서 크리깅으로 알려져 있다. 크리깅의 종류에는 단순 크리깅(simple kriging), 일반 크리깅, 범용 크리깅, 지시 크리깅(indicator kriging) 등이 있다. 본 연구에서 고려한 일반 크리깅은 주어진 공간 영역에서 평균이 일정하다고 보기 어려운 경우에 적합한 예측방법으로서 설명변수를 고려한 방법이다. 일반 크리깅에서는

평균이 갖는 공간적 추세를 제거한 잔차를 이용하여 세미베리오그램을 추정한 후 미관측 지점의 자료값을 예측하게 된다. 따라서 일반 크리깅에서 자료는 식 (2.1)과 같은 모형으로 표현할 수 있다. 평균함수 $\mu(s)$는 보통 추세와 같은 큰 변동을 나타내며 아래와 같이 $(p+1)$개의 설명변수의 선형 결합으로 나타내어질 수 있다.

$$\mu(\mathbf{s}) = \beta_0 X_0(\mathbf{s}) + \beta_1 X_1(\mathbf{s}) + \cdots + \beta_p X_p(\mathbf{s}).$$

여기서, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$는 회귀계수이며, $X_0(\mathbf{s}) = 1$이라면 평균함수는 절편을 가지고 있는 모형이 된다. 설명변수들 $\{X_k(s);\ k = 1, \ldots, p\}$와 공간적 상관구조를 가지고 있는 오차항이 존재하므로 $\{Z(\mathbf{s})\}$에 대한 식 (2.1)의 모형을 공간적 선형회귀모형이라 부른다. 따라서 이 연구에서 고려한 모형식은 식 (2.2)와 같이 표현된다. $i = 1, \ldots, n$에 대하여

$$Z(\mathbf{s}_i) = \beta_0 + \beta_1 X_1(\mathbf{s}_i) + \cdots + \beta_p X_p(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \tag{2.2}$$

여기서, $\{X_k(\mathbf{s});\ k = 1, \ldots, p\}$는 주로 위치좌표들에 대한 선형함수(linear function) 또는 2차형함수(quadratic function)이거나, 지역의 여러 가지 측정변수들(온도, 바람의 세기, 바람의 방향 등)의 함수일 수 있다. 이러한 공간적 선형회귀모형에 의한 접근방법을 일반적으로 범용 크리깅이라고 한다.

범용 크리깅에서 미관측지점인 $\mathbf{s}_0$에서의 확률과정은 $\{Z(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)'\boldsymbol{\beta} + \epsilon(\mathbf{s}_0)\}$이고, 범용 크리깅에 의한 예측값 $\widehat{Z}(\mathbf{s}_0)$은 관측된 n지점들에서 측정한 $\{Z(\mathbf{s}_i);\ i = 1, \ldots, n\}$의 선형결합으로 다음과 같이 표현될 수 있다.

$$\widehat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i), \tag{2.3}$$

여기서, 선형 계수 $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$는 범용 크리깅의 가중치들을 나타내며 균일적 불편 예측값의 조건인 $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}(\mathbf{s}_0)'$을 만족하여야 한다. 식 (2.3)은 확률과정이 2차 정상성(second-order stationarity)이라는 가정하에서 최량선형불편성예측량(best linear unbiased predictor: BLUP)이다. 이러한 선형계수를 얻기 위해 필요한 모수들의 추정량은 다음의 일반화최소제곱 추정방법에 의해 계산되어 진다.

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{z},$$

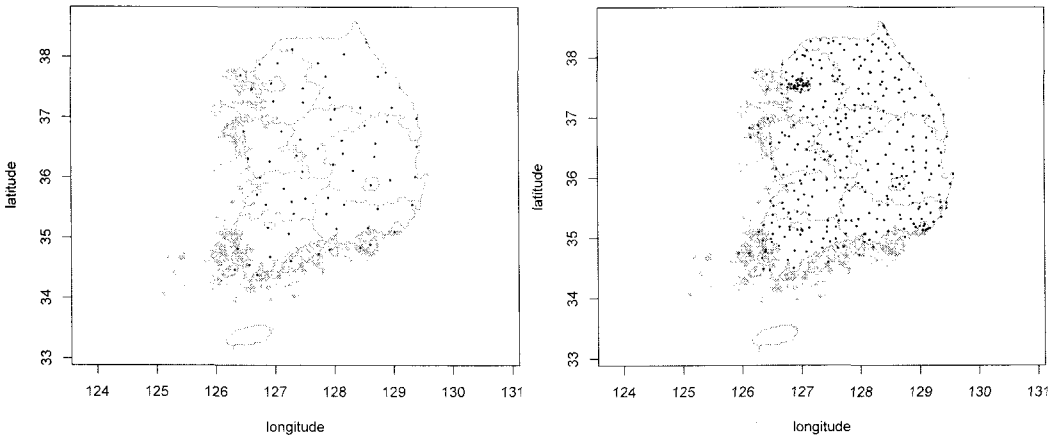
여기서 $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$는 세미베리오그램 혹은 코베리오그램에 관련된 모수 $\boldsymbol{\theta}$를 포함하는, 오차항의 분산-공분산 행렬이고 $\mathbf{z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n))'$이다. 보다 자세한 이론적 배경은 Cressie (1993)를 참조하기 바란다.

2.4. 모형비교기준

공간자료에 대하여 2가지 분석방법의 예측방법을 비교하고자 한다. 다양한 비교 기준들 중에서 한 관측지점씩을 제거한 후 나머지 $n-1$개의 지점에서 얻은 자료들만을 이용하여 모형을 적합하고 제거된 각 지점에 공간 예측하여 실제값과 비교하는 접근방법인 교차검증방법을 고려하였다. 또한 추정된 세미베리오그램 모형의 예측력 비교 기준으로는 식 (2.4)에서 정의된 예측잔차제곱합(PRESS)을 고려하였다.

$$\text{PRESS} = \sum_{i=1}^{n} \left\{ Z(\mathbf{s}_i) - \widehat{Z}(\mathbf{s}_{-i}) \right\}^2, \tag{2.4}$$

여기서, $Z(\mathbf{s}_i)$는 i번째 위치에서 얻은 실제값이고, $\widehat{Z}(\mathbf{s}_{-i})$은 i번째 위치에 있는 관측값을 제외하고 나머지 관측값으로 i번째 위치에 공간 예측한 값이다.



(a) 지상기상관측지점($n = 69$) (b) 자동(항공)기상관측지점($m = 349$)

그림 3.1. 모형화 자료와 타당성 자료의 관측지점들에 대한 위치좌표

각 추정방법에서 얻은 예비모형들의 예측력을 평가하기 위해서 지상기상관측지점에서 얻어진 자료를 모형화 자료(modeling data)로, 자동기상관측지점과 항공기상관측지점에서의 자료를 타당성 자료로 간주한다. 모형화 자료로부터 얻은 모형을 이용하여 타당성 자료의 관측지점들에서의 예측값을 구하고 예측오차를 구함으로써 예비모형들 간의 예측력을 비교한다. 예측력에 대한 기준은 모형적합에 사용되지 않은 타당성 자료의 예측값을 실제값과 비교, 평가하기 위하여 평균제곱예측오차(MSPE)척도를 사용하였다.

$$\text{MSPE} = \frac{1}{m}\sum_{j=1}^{m}\left\{\widehat{Z}(\mathbf{s}_j) - Z(\mathbf{s}_j)\right\}^2, \tag{2.5}$$

여기서, m은 타당성 자료의 표본크기이고, $\widehat{Z}(\mathbf{s}_j)$는 모형화 자료를 이용하여 $\mathbf{s}_j$에 공간 예측한 값이다.

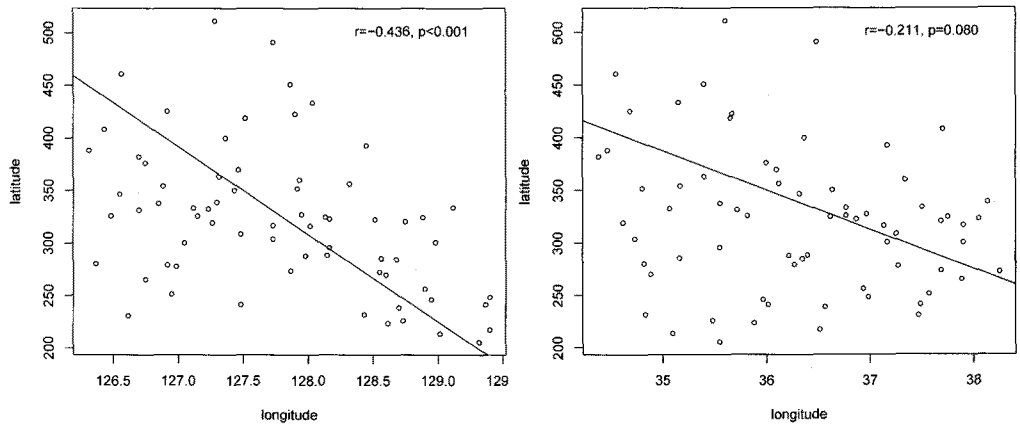
3. 실증적자료분석

2장에서 기술한 공간통계학에서의 예측모형 (식 (2.2)를 참조)을 기반으로 실제 자료에 적용한 결과를 설명하고자 한다. 자료 분석은 통계 프로그램인 R (R Development Core Team, 2008)을 사용하였다.

3.1. 강수량자료

본 연구에서 사용한 모형화 자료는 기상청의 76개 관측소에 대한 2007년 3분기에 해당하는 7–9월의 월평균 강수량자료이다. 그림 3.1(a)에서 알 수 있듯이 76개 지상기상관측지점들 중에서 섬 지역을 제외한 69개 관측소 자료(모형화 자료)만을 이용하여 분석하였다. 실증적 검증을 위한 타당성 자료 또한 전국 자동기상관측지점과 항공기상관측지점들 중 육지의 공간변이만을 고려하기 위해 섬지역을 제외하고 349개의 관측소(자동기상관측지점 340개, 항공기상관측지점 9개)의 지리좌표와 강수량으로 구성되어 있다(그림 3.1(b)). 모든 관측지점의 지리좌표는 위도(latitude)와 경도(longitude)이고 강수량 측정단위는 밀리미터(mm)이다.

표 A.1(부록)은 각 지상기상관측소별 좌표(위도, 경도)와 3분기 월평균 강수량을 나타내고 있다. 월평균 강수량이 가장 많은 곳은 지리산 일대와 남해안 지역의 해남, 산청, 진주, 장흥, 거창으로 나타났고



(a) 경도와 강수량 (b) 위도와 강수량

그림 **3.2.** 지리좌표와 월평균 강수량

3분기 강수량이 가장 적은 곳은 울산, 부산, 영덕, 대구 등으로 동해안의 경상도 지역으로 나타났다.

위도와 경도별 강수량의 분포를 살펴보면 그림 3.2와 같다. 그림 3.2(a)를 살펴보면 경도에 따라서는 동해안 지역보다는 서해안과 중부지역의 강수량이 높고 1차원의 패턴을 보이고 있는 것을 알 수 있으며 위도와 강수량의 상관계수는 -0.4358로서 유의한 연관성을 보이고 있다. 그림 3.2(b)로부터 북쪽 지역보다는 남쪽지역의 강수량이 높음을 알 수 있다. 이 두 그림을 종합적으로 살펴보면, 동해안의 경상도 지역이 가장 강수가 낮게 나타났는데 전라도 및 남해안 지역의 강수가 높은 지역이 많아서 위도로 살펴보았을 때 남쪽지역의 강수량이 높게 나타나고 있다. 따라서 본 연구에서는 $\mu(\mathbf{s})$를 위도와 경도를 이용한 선형 함수로 고려하였다. 그리고 모형의 단순화를 위하여 주어진 자료가 2차 정상성을 만족한다고 가정하였다. 본격적인 분석에 앞서 지구의 곡면구조에서 발생할 수 있는 편의(bias)를 보정하기 위한 작업의 일환으로 지리좌표를 킬로미터(km) 단위의 지오데식 거리(geodesic distance)로 변환하였다. 따라서 추정된 범위 값의 단위는 킬로미터이다.

3.2. 공간적 선형모형의 적합

공간적 선형모형의 모수들을 추정하는 방법은 세미베리오그램 모형을 기반으로 한 최소제곱방법과 분산-공분산 행렬(혹은 코베리오그램)을 기반으로 한 최대우도추정방법으로 나누어지며 각각 장단점을 가지고 있다. 최소제곱방법은 모집단 분포에 대한 가정이 없고 자료가 대용량이라도 계산이 용이하다는 장점을 가지고 있는 반면에 원 자료가 아닌 세미베리오그램 만으로 추정량을 도출한다는 단점을 가지고 있다. 이에 반해, 최대우도추정방법은 모든 자료값들을 직접적으로 사용하여 추정량을 구한다는 장점이 있지만 자료가 대용량일 경우에는 분산-공분산행렬의 역행렬로 인한 계산상의 어려움이 있다 (허태영 등, 2004). 본 연구에서는 가중최소제곱방법과 잔차최대우도추정방법 등 두 가지 모수추정방법들을 사용하고자 한다. 두 가지 모수 추정방법 각각에 대해, 세 가지 세미베리오그램(코베리오그램) 모형의 적합을 시도하였고 이들 모형 중에서 잔차제곱합이 최소인 그리고 로그우도값이 가장 큰 모형을 예비모형으로 선택하였다.

세미베리오그램(코베리오그램)을 가장 잘 설명하는 이론적 모형을 찾기 위하여 모수를 변화시키면서 구형모형, 지수모형, 가우시안 모형에 적합시켰다. 표 3.1에는 추정된 베리오그램 모수 추정값들과 잔차제

표 3.1. 세미베리오그램 모형의 모수 추정

추정방법	모형	모수 추정값			비교기준	
		뭉치	문턱	범위	잔차제곱합	로그우도
가중최소제곱	지수모형	1781.70	4380.86	55.46	6.70	
	가우시안	2591.69	4253.01	73.12	5.24	
	구형모형	2181.90	4221.82	138.99	4.99	
잔차최대우도	지수모형	1812.23	3388.10	35.98		−355.34
	가우시안	2289.09	3848.19	89.58		−354.59
	구형모형	2063.55	3481.81	142.99		−355.05

표 3.2. 평균함수에 포함된 모수 추정값

모수	구형모형(가중최소제곱)		가우시안(잔차최대우도)	
	추정값	(표준오차)	추정값	(표준오차)
β_0	319.43	(7.47)	308.33	(31.51)
β_1	−0.36	(0.10)	−0.44	(0.27)
β_2	−0.09	(0.06)	−0.03	(0.20)

곱합 그리고 로그우도값이 나타나 있다. 범위 추정값을 제외하고 문턱과 뭉치효과의 추정값들은 각 추정방법에 따라서 큰 차이를 보이고 있지 않음을 알 수 있다. 그리고 가중최소제곱법에서 잔차제곱합이 가장 작은 모형은 구형모형이었으며 잔차최대우도방법을 이용한 가우시안모형이 가장 큰 로그우도값을 가지고 있는 것으로 나타났다. 따라서 본 연구에서는 이 두 모형하에서의 모수추정값들을 이용하여 공간 예측을 실시하였다.

선택된 두 가지 예비모형을 이용하여 범용 크리깅을 실시하고 비교하여 보겠다. 크리깅에 적용하기 위한 최종모형들은 다음과 같이 표현된다.

(1) 가중최소제곱방법에 의한 구형모형

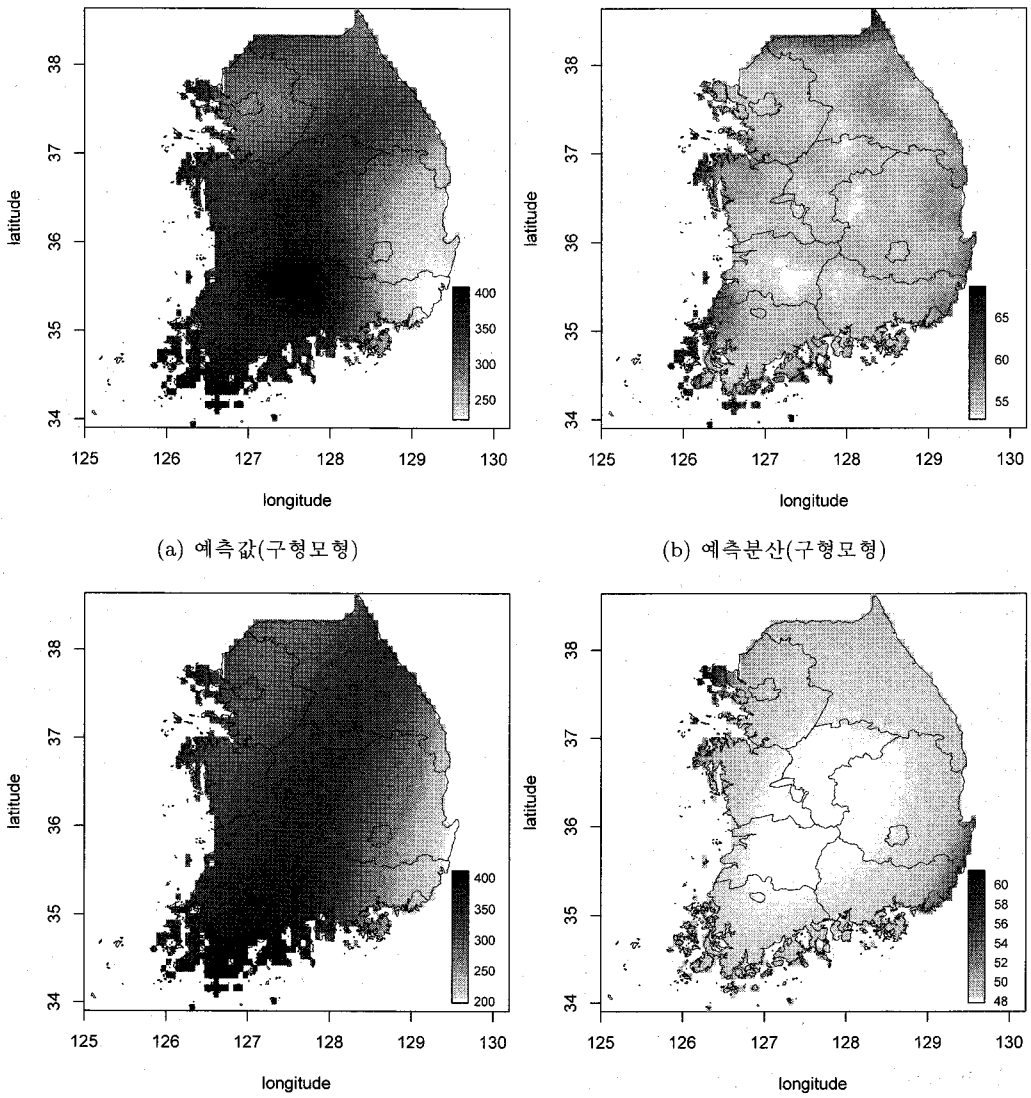
$$\gamma(\mathbf{h};\widehat{\boldsymbol{\theta}}) = \begin{cases} 0, & \mathbf{h}=\mathbf{0}, \\ 2181.90 + 2039.92 \times \left\{ \dfrac{3\|h\|}{2\times 138.99} - \dfrac{1}{2}\left(\dfrac{\|h\|}{138.99}\right)^3 \right\}, & 0 < \|\mathbf{h}\| \le 138.99, \\ \theta_0 + \theta_1, & \text{otherwise.} \end{cases}$$

(2) 잔차최대우도추정방법에 의한 가우시안모형

$$\gamma(\mathbf{h};\widehat{\boldsymbol{\theta}}) = \begin{cases} 0, & \mathbf{h}=\mathbf{0}, \\ 2289.09 + 1559.10 \times \left\{ 1 - \exp\left(-\dfrac{\|h\|^2}{89.58^2} \right) \right\}, & \text{otherwise.} \end{cases}$$

선택된 세미베리오그램 모형들의 평균함수에 대한 모수추정값들은 표 3.2에 제시되어 있다. 두 추정방법 하에서의 결과가 유사하나 경도의 효과가 가중최소제곱방법에서는 유의한 것으로, 잔차최소제곱방법에서는 유의하지 않은 것으로 나타났다.

그림 3.3은 표 3.1에서 선택된 예비 모형들 하에서 표 3.2에서 얻은 추정된 평균함수를 통해 계산된 남한지역의 강수량 예측지도(prediction map)이다. 가중최소제곱방법에 의한 구형모형으로부터 구한 예측값 지도(그림 3.3(a))를 살펴보면 남해 지역과 지리산 부근의 강수량이 다른 지역, 특히 경상도 지역에 비해 훨씬 높게 예측되었다.



(a) 예측값(구형모형) (b) 예측분산(구형모형)

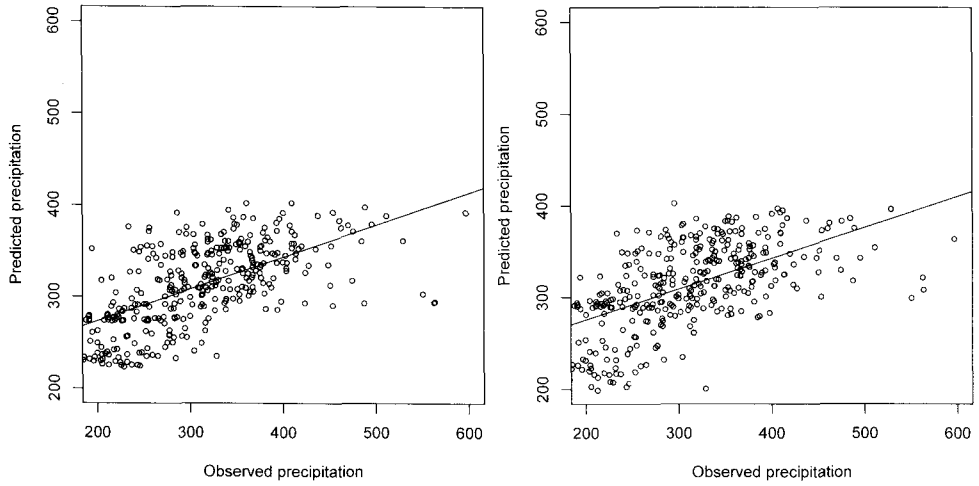
(c) 예측값(가우시안) (d) 예측분산(가우시안)

그림 3.3. 예비 모형들에 기반한 예측값과 예측분산

이러한 경향은 잔차최대우도추정방법에 의해 선택된 가우시안모형 하에서의 예측지도(그림 3.3(c))에서도 발견되어진다. 또한, 가우시안모형을 이용한 예측값 지도에서는 남한의 대각선 부분의 강수량이 다른 지역에 비해 높게 예측되었음을 알 수 있다. 예측분산의 관점에서는 가우시안모형이 구형모형에 비해 훨씬 작은 예측분산을 나타내고 있음을 알 수 있다(그림 3.3(b), (d)). 이는 모든 자료값들을 이용하여 모수를 추정하는 최대우도추정방법을 사용함으로써 얻게 되는 자연스러운 이점이다.

표 3.3. 예비 모형별 예측오차제곱합과 포함확률

추정방법	모형	예측오차제곱합	포함확률
가중최소제곱	구형모형	2757.94	97.10
잔차최대우도	가우시안	2733.00	92.75



(a) 구형모형 (b) 가우시안모형

그림 3.4. 타당성 자료의 실제값과 크리깅에 의한 예측값의 비교

3.3. 교차검증방법을 통한 예비 모형의 성능비교

표 3.1로부터 예비 모형으로 선택된, 가중최소제곱법을 이용한 구형모형과 잔차최대우도추정법을 이용한 가우시안모형을 이용하여 교차검증방법을 시행하였다. 모형 비교를 위해 각 예비 모형에서의 예측오차제곱합을 계산하였다. 예측값 성능의 지표로 사용되는 포함확률(coverage probability)을 계산하기 위하여 우선 교차검증방법을 이용하여 i번째 자료값들을 제외한 모형으로 제거된 관측지점의 예측값에 대한 95% 신뢰구간(confidence interval)을 구한 다음 그 관측지점에서 얻은 실제값이 그 신뢰구간 안에 포함되는 지의 여부를 살펴본다. 이 과정을 모든 관측값들에 적용하여 전체 관측값들 중 몇 %가 신뢰구간에 포함되는지를 보여주는 것이 바로 포함확률이다. 모형별 예측오차제곱합과 포함확률은 표 3.3과 같다. 예측오차제곱합 관점에서는 잔차최대우도추정방법을 이용한 가우시안모형이 미세하게 좋게 나타나고 있으나 전체 69개 중 67개의 관측값을 주어진 구간 안에 포함시킨 가중최소제곱법을 사용한 구형모형이 안정적인 성능을 나타내고 있음을 알 수 있다.

3.4. 타당성 자료를 이용한 예비 모형의 성능비교

예비 모형들의 예측 성능을 평가하기 위하여 타당성 자료의 실제값과 모형화 자료만으로 범용 크리깅한 예측값을 비교하였다. 그림 3.4는 강수량의 실제값과 예비 모형들을 이용한 예측값들 간의 산점도를 나타낸다. 이 그림을 통해 알 수 있듯이 전체적인 예측값들의 경향은 두 모형에서 비슷하게 나타나고 있으나 강수량이 200–300mm인 경우 구형모형이 가우시안모형에 비해 다소 높게 예측되었음을 알 수 있다.

좀 더 정확한 모형비교를 하기 위하여 식 (2.5)에 나타나 있는 평균제곱예측오차를 이용하여 모형을 비교하여 보았다. 표 3.4에 결과 값이 나타나있다. 타당성 자료를 예측지점으로 하여 크리깅을 실시하였

표 3.4. 관측지점별 평균제곱예측오차 비교

추정방법	모형	평균제곱예측오차
가중최소제곱	구형모형	59.60
잔차최대우도	가우시안	63.39

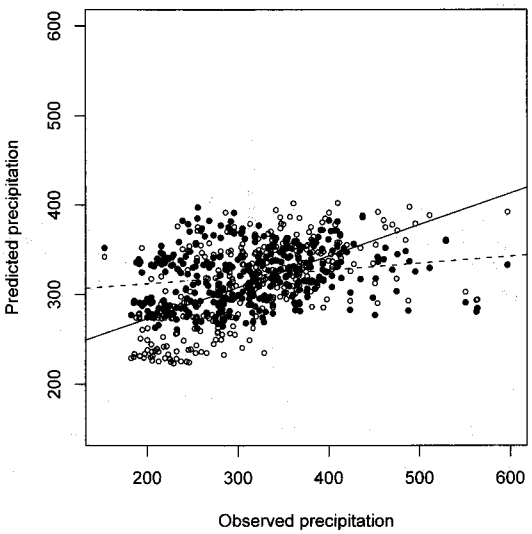


그림 3.5. 전통적 선형회귀모형과 공간적 선형회귀모형의 예측력 비교

을 때 가중최소제곱법을 이용한 구형모형의 평균제곱예측오차가 잔차최대우도추정법에 의한 가우시안 모형보다 상대적으로 작음을 알 수 있다.

타당성 자료를 이용한 두 예비 모형들의 비교 및 교차검증방법의 결과를 토대로 가중최소제곱법에 의한 구형모형이 잔차최대우도추정방법에 의한 가우시안모형보다 비교적 좋은 성능을 나타내고 있음을 알 수 있다. 위에서 소개한 두 가지 모형을 포함한 모든 모형의 평균제곱예측오차를 구하여 보았을 때도 같은 결과를 얻을 수 있었다. 따라서, 본 연구에서는 공간적 상관구조를 고려한 가중최소제곱법을 이용한 구형모형이 최적의 모형으로 판단된다.

마지막으로 본 연구에서 최적의 모형으로 판단한 구형모형을 이용한 공간적 선형회귀모형과 공간적 상관구조를 고려하지 않는 전통적인 선형회귀모형을 비교하고자 한다. 그림 3.5에서 검은 색의 점(solid point)은 공간적 상관구조를 고려하지 않은 선형회귀분석에 의한 예측값이고 비어 있는 점은 구형모형을 고려한 선형회귀분석에 의한 예측값이다. 절단선(dashed line)과 실선(solid line)은 각 선형회귀모형하에서 얻은 예측값과 실제값들 간의 기울기를 나타내고 있다. 그림 3.5에서 알 수 있듯이, 공간적 상관구조를 고려하지 않은 전통적 선형회귀모형보다 구형모형을 고려한 공간적 선형회귀모형의 예측력이 훨씬 우수함을 알 수 있다.

4. 결론

본 연구에서는 선택된 두 가지 예비 모형들을 가지고 공간적 상관구조를 이용하여 3분기 월평균 강수량을 예측하는 모형을 시도하였다. 그 실용성과 정확성을 검증하기 위하여 자동기상관측지점 340개와

항공기상관측지점 9개의 실제값과 비교하여 보았다. 전국 강수량의 공간적 변이를 가장 잘 설명해 내는 모형은 평균제곱예측오차(MSPE)와 교차검증방법의 관점에서 가중최소제곱방법에 의한 구형모형으로 나타났다. 또한, 공간적 변이를 고려하지 않은 전통적 선형회귀모형보다 공간적 선형회귀모형이 좋은 성능을 나타내고 있음을 알 수 있었다. 이러한 공간적 변이를 이용하여 강수량을 예측한다면 강수량에 대한 피해를 최소화하고 강수 관리의 효율화가 개선되리라고 생각된다. 지리좌표 만이 아닌 고도나 강수계의 높이 등을 고려한 연구와 강수량을 설명할 수 있는 새로운 변수를 이용하는 코크리깅(Co-kriging)방법을 이용한 더 세밀한 강수량의 공간적 분석은 차후과제로 남긴다.

부록

표 A.1. 모형화 자료의 구성(좌표. 3분가 월평균강수량)

관측지점	위도	경도	강수량	관측지점	위도	경도	강수량
속초	38°15'	128°33'	272.47	강화	37°42'	126°26'	408.03
철원	38°08'	127°18'	338.67	양평	37°29'	127°29'	241.17
동두천	37°54'	127°03'	300.23	이천	37°15'	127°29'	308.50
문산	37°53'	126°45'	265.13	인제	38°03'	128°10'	322.83
대관령	37°41'	128°45'	320.40	홍천	37°41'	127°52'	273.67
춘천	37°54'	127°44'	316.67	태백	37°10'	128°59'	300.63
강릉	37°45'	128°53'	324.47	제천	37°08'	128°01'	316.33
동해	37°30'	129°07'	333.77	보은	36°29'	127°44'	490.67
서울	37°34'	126°57'	251.20	천안	36°46'	127°07'	333.33
인천	37°28'	126°37'	230.63	보령	36°19'	126°33'	346.50
원주	37°20'	127°56'	359.90	부여	36°16'	126°55'	279.17
수원	37°16'	126°59'	277.83	금산	36°06'	127°28'	369.33
영월	37°10'	128°27'	392.60	부안	35°43'	126°42'	331.33
충주	36°58'	127°57'	327.13	임실	35°36'	127°17'	511.00
서산	36°46'	126°29'	325.80	정읍	35°33'	126°51'	337.50
울진	36°59'	129°24'	248.13	남원	35°24'	127°19'	362.83
청주	36°38'	127°26'	350.03	장수	35°39'	127°31'	418.50
대전	36°22'	127°22'	399.70	순천	35°04'	127°14'	332.33
추풍령	36°13'	127°59'	287.83	장흥	34°41'	126°55'	425.07
안동	36°34'	128°42'	238.63	해남	34°33'	126°34'	460.67
상주	36°24'	128°09'	288.23	고흥	34°37'	127°16'	319.17
포항	36°01'	129°22'	240.97	봉화	36°56'	128°54'	256.17
군산	36°00'	126°45'	375.83	영주	36°52'	128°31'	322.50
대구	35°53'	128°37'	223.43	문경	36°37'	128°08'	324.67
전주	35°49'	127°09'	325.70	영덕	36°31'	129°24'	217.00
울산	35°33'	129°19'	205.40	의성	36°21'	128°41'	284.67
마산	35°10'	128°34'	285.30	구미	36°07'	128°19'	356.17
광주	35°10'	126°53'	354.00	영천	35°58'	128°57'	245.83
부산	35°06'	129°01'	213.27	거창	35°40'	127°54'	422.67
통영	34°50'	128°26'	231.53	합천	35°33'	128°10'	295.67
목포	34°49'	126°22'	280.33	밀양	35°29'	128°44'	225.67
여수	34°44'	127°44'	303.87	산청	35°24'	127°52'	450.67
완도	34°23'	126°42'	381.77	거제	34°53'	128°36'	269.83
진도	34°28'	126°19'	387.90	남해	34°48'	127°55'	351.67
진주	35°09'	128°02'	433.43				

자료출처: 기상청 자료기록실

참고문헌

김선우, 정애란, 이성덕 (2005). 공간자료에 대한 지리적 가중회귀 모형과 크리깅의 비교, <응용통계연구>, **18**, 271–280.

신만용, 윤진일, 서애숙 (1999). 공간통계기법을 이용한 전국 일 최고/최저기온 공간변이의 추정, <대한원격탐사학회지>, **15**, 9–20.

오광중, 곽진, 정덕영, 손건태 (1998). 부산지역의 대기오염물질농도와 기상인자간의 통계분석-광안리지역을 중심으로, *Journal of Korean Society of Environmental Engineers*, **20**, 1235–1245.

유성모, 엄익현 (1999). 강우강도 데이터를 이용한 세미베리오그램의 추정과 공간이상치에 관한 연구, <응용통계연구>, **12**, 125–141.

이지영, 황철수 (2002). 공간통계분석을 이용한 지가의 입지값 측정에 관한 연구, <한국GIS학회지>, **10**, 233–246.

장지희 (2003). <크리깅 방법을 이용한 공간데이터 분석>, 성균관대학교, 석사학위논문.

조재영 (2001). <공간자료에 대한 공간통계와 일반통계 분석법에 의한 예측력 비교 연구>, 동의대학교, 석사학위논문.

조재영, 최승배, 김규곤 (2001). 일반통계에 대한 공간통계 방법의 예측성능에 관한 연구, *Journal of the Korean Data Analysis Society*, **6**, 473–491.

최승배, 조장식, 범수균 (2004). 공간변이성의 추정량들에 대한 예측력 비교 연구: 남한지역 일산화탄소 자료를 이용하여, *Journal of the Korea Data Analysis Society*, **6**, 279–291.

허태영, 서의훈, 권원태 (2004). 세미베리오그램 모형을 적용한 남한지역 강수량 자료의 공간분석, *Journal of the Korean Data Analysis Society*, **6**, 473–491.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, John Wily & Sons, New York.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*, Vienna, Austria, http://www.R-project.org

Precipitation Analysis Based on Spatial Linear Regression Model

Jiyoung Jung[1] · Seohoon Jin[2] · Man Sik Park[3]

[1]Dept. of Informational Statistics, Korea University;
[2]Dept. of Informational Statistics, Korea University;
[3]Dept. of Biostatistics, Korea University

(Received August 2008; accepted September 2008)

Abstract

In this study, we considered linear regression model with various spatial dependency structures in order to make more reliable prediction of precipitation in South Korea. The prediction approaches are based on semi-variogram models fitted by least-squares estimation method and restricted maximum likelihood estimation method. We validated some candidate models from the two different estimation methods in terms of cross-validation and comparison between predicted values and observed values measured at different locations.

Keywords: Semi-variogram, precipitation, universal kriging, spatial linear regression model, cross-validation.

[1]Graduate student, Dept. of Informational Statistics, Korea University, 208 Seochang-Ri, Jochiwon-Eup, Yeonki-Gun, Chungnam 339-700, Korea. E-mail: jjiyongi@korea.ac.kr

[2]Assistant professor, Dept. of Informational Statistics, Korea University, 208 Seochang-Ri, Jochiwon-Eup, Yeonki-Gun, Chungnam 339-700, Korea. E-mail: seohoon@korea.ac.kr

[3]Corresponding author: Research professor, Dept. of Biostatistics & Dept. of Preventive Medicine, Medical Research Center for Environmental Toxico-Genomics and Proteomics, College of Medicine, Korea University, 126-1 Anam-Dong, Sungbuk-Gu, Seoul 136-705, Korea. E-mail: bayesia@korea.ac.kr