

강우강도 데이터를 이용한 세미베리오그램의 추정과 공간이상치에 관한 연구 *

유성모¹⁾ 엄익현²⁾

요약

서로 다른 위치에서 동시에 관찰된 자료들이 공간적인 변인에 의하여 영향을 받는다 면 공간적인 변인의 함수식에 의한 예측모형을 설정하는 것이 타당하다. 본 연구에서는 공간적인 변인으로 거리가 주어졌을 때, 공간자료에 대한 세미베리오그램 모형의 추정과 관측되지 않은 지점에 대한 공간예측기법을 정리하였으며, 또한 공간이상치 탐지를 위한 두 가지 방법론으로 분포론적 방법과 p-Deletion 방법을 제시하였다. 방법론의 예시를 위하여 강우강도 자료를 이용하였으며 서로 상관되어 있는 공간데이터에 대한 시뮬레이션을 통하여 두가지 방법을 비교하였다.

1. 서론

강우량, 일조량, 월평균기온 등 기후 관련자료와 같이 서로 다른 위치에서 동시에 관찰된 자료들은 공간적인 상관관계를 고려한 분석방법을 이용하는 것이 보다 현실적이다. 본 논문에서 이용한 강우강도 자료들은 서로 위치가 가까울수록 그 값이 높은 상관을 보이고, 멀어질수록 낮은 상관을 보이고 있으며, 따라서 강우강도 자료의 상관함수 또는 공분산함수는 어떠한 형태이든 거리에 대한 함수식으로 표현될 것이다. 본 연구에서는 공간적인 변인으로 거리가 주어졌을 때, 실제 공간자료인 강우강도 자료에 대한 세미베리오그램 모형의 추정과 관측되지 않은 지점에 대한 공간예측기법을 정리하였으며, 공간이상치 탐지를 위한 두 가지 방법론으로 분포론적 방법과 p-Deletion 방법을 제시하고 방법론의 예시를 위하여 강우강도 자료를 이용하였으며 서로 상관되어 있는 공간데이터에 대한 시뮬레이션을 통하여 두가지 방법론을 비교, 분석하였다.

이해를 돕기 위해서 본 논문에서 실증적 자료로 이용하고 있는 강우강도가 차지하는 의미와 강우강도의 특성을 살펴보도록 하겠다. 위성통신 등에서 사용되는 무선회선설계에서 밀리파와 같은 높은 주파수대에서는 대기중의 기체분자에 의한 흡수 손실이 자유공간 손실보다 크다. 이러한 특성은 현재까지 여러 연구에 의해 많이 규명되었다. 예를 들어, 산소 분자에 의한 최대 흡수는 60GHz 부근으로 1km당 전파손실이 10dB 이상이 되며 수증기에 의해서는 200GHz 부근에 최대 흡수가 일어나며 1km당 30dB의 손실값이 발생한다. 그러나 이러한 공기중의 감쇠는 일부 대역에 한정된 특성이며 실제 전파상에 커다란 영향을 미치는 것은 강우에 의한 감쇠이다. 강우로 인한 감쇠는 강우강도에 따라서 전파의 감쇠정도

* 이 논문은 1997년 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

1) (339-700) 충남 연기군 조치원읍 서창동, 고려대학교 정보통계학과, 부교수

2) (136-701) 서울시 성북구 안암동 5-1, 고려대학교 통계학과, 박사과정

표 1.1: 분석에 이용된 측후소의 위치와 강우강도(mm/Hr)자료

지역명	북위(°)	동경(°)	강우강도	지역명	북위(°)	동경(°)	강우강도
1. 서울	37.567	126.967	73.608	17. 온양	36.783	126.983	60.804
2. 강화	37.700	126.450	79.608	18. 전주	35.817	127.150	60.804
3. 이천	37.283	127.433	68.947	19. 군산	35.983	126.700	56.304
4. 춘천	37.900	127.733	59.268	20. 광주	35.167	126.883	91.824
5. 강릉	37.750	128.900	52.482	21. 목포	34.783	126.383	58.768
6. 대관령	37.683	128.750	49.608	22. 여수	34.733	127.733	67.902
7. 동해	37.500	129.117	33.590	23. 대구	35.883	128.617	40.152
8. 속초	38.250	128.567	35.813	24. 안동	36.550	128.717	47.826
9. 원주	37.333	127.950	63.216	25. 영덕	36.533	129.417	39.432
10. 홍천	37.683	127.883	67.608	26. 의성	36.350	128.683	49.608
11. 청주	36.633	127.433	59.801	27. 풍기	36.867	128.517	58.268
12. 제천	37.150	128.183	71.947	28. 부산	35.100	129.033	69.216
13. 추풍령	36.217	128.000	45.402	29. 거창	35.667	127.917	63.216
14. 대전	36.300	127.400	73.536	30. 밀양	35.483	128.750	58.608
15. 금산	36.100	127.467	59.982	31. 울산	35.550	129.317	47.768
16. 서산	36.767	126.467	56.561	32. 진주	35.200	128.100	72.216

가 다양하고, 강우강도는 공간적 시간적으로 그의 변동이 매우 다양하다. 강우감쇠가 증가하여 회선 품질이 정해진 규격을 벗어나면 시스템은 불가동 상태에 떨어지며 이러한 상태에 이르는 확률을 일정한 값이하로 억제하기 위하여 각 중계거리는 제약을 받게 된다. 따라서 강우강도에 의한 시간적 공간적 데이터의 확보를 토대로한 강우강도에 대한 신뢰성 있는 예측모형설정 및 분석은 강우로 인한 감쇠의 정확한 예측과 이를 고려한 회선설계를 위해서는 필수적이다.

또한, 비는 일년중 수 개월에 몰리는 경향이 있는데 대체적으로 연중 7~8월에 내리는 강우량이 전체 강우량의 41~56% 정도이며 6~9월 사이에 내리는 강우량이 전체의 76~86%에 달한다. 이러한 점을 감안할 때 우기 또는 최다우 수 개월 동안의 통화품질에 관심이 있는 경우에는 연중 통화품질이 요구되는 신뢰도보다 최다우 수 개월간의 신뢰도가 더 유용하다.

따라서 본 연구에서는 현재 우리 나라 남한전역의 72개 지상기상관측지점과 8개 항공기상관측지점 중 제주도를 제외한 8개도 32개 관측지점의 과거 10여년간(84~93년)의 최다우 2개월(7~8월) 상위 0.01%에 해당하는 20분 간격의 강우강도(mm/Hr)를 실증적 분석을 위한 데이터로 사용하였다(표 1.1). 32개 관측지점의 강우강도에 대한 정규성검정결과 Kolmogorov D통계량이 0.1이고 p-value는 0.15이상이었으며, 강우강도의 공분산관계를 제거한 다음에 정규성 결과를 시행하여도 강우강도의 정규성에 대한 가정은 무리가 없는 것

으로 판단된다. 하지만 강우강도 자료는 본 연구에서 제시한 방법론의 예시를 위하여 이용하였기 때문에 강우강도에 대한 정규성 여부는 크게 문제되지 않는다.

2. 범용 크리깅(ORDINARY KRIGING)

2.1. 공간변수와 세미베리오그램

본 논문에서 사용하고 있는 공간변수 $\{Z(s) : s \in D\}$ 는 다음과 같은 가정을 만족한다고 하자. 여기서, D 는 d 차원 공간 R^d 상의 부분집합이다.

- (가) $E(Z(s)) = \mu$, 모든 $s \in D$,
- (나) $Var(Z(s)) \equiv C(0) = \sigma^2$, 모든 $s \in D$,
- (다) $Cov(Z(s_i), Z(s_j)) \equiv C(s_i - s_j) < \infty$, 모든 $s_i, s_j \in D$,
- (라) 공간변수 $Z(\cdot)$ 는 정규분포를 따른다.

여기서, $s_i - s_j$ 는 두 지점 s_i 와 s_j 사이의 유클리드 거리라 표현한다면, 공간변수들 사이의 상관관계를 나타내는 공분산은 거리만의 함수인 $C(s_i - s_j)$ 형태로 표현된다. $C(s_i - s_j)$ 의 특별한 경우인 $s_i - s_j = 0$ 일 때의 $C(0)$ 는 공간변수의 분산 σ^2 이 된다.

앞의 가정들에 의하여 두 공간변수의 차 $Z(s_i) - Z(s_j)$ 는 평균이 0이고 분산이 $2C(0) - 2C(s_i - s_j) \equiv 2\gamma(s_i - s_j)$ 인 정규분포를 따른다. 여기서 $C(s_i - s_j)$ 와 $2\gamma(s_i - s_j)$ 를 각각 코베리오그램(Covariogram; 공변이도)과 베리오그램(Variogram; 변이도)이라고 부르며, $\gamma(s_i - s_j)$ 를 세미베리오그램(Semivariogram; 반변이도)이라고 부른다.

h 를 두 지점 사이의 거리의 차이를 나타내는 값이라고 할 때 $s_i = s, s_j = s + h$ 라고 놓을 수 있다. 이 때 코베리오그램 $C(\cdot)$ 와 세미베리오그램 $\gamma(\cdot)$ 사이에는 다음의 관계식이 성립한다.

$$\begin{aligned} \gamma(h) &= \frac{1}{2}E(Z(s) - Z(s+h))^2 \\ &= \frac{1}{2}Var(Z(s)) + \frac{1}{2}Var(Z(s+h)) - Cov(Z(s), Z(s+h)) \\ &= C(0) - C(h) \end{aligned} \quad (2.1)$$

직관적으로 코베리오그램의 값은 두 좌표의 공간상의 거리가 가까울수록 크고 멀수록 작아지며 마침내 0에 수렴한다고 생각할 수 있다. 따라서 $\lim_{h \rightarrow \infty} \gamma(h) = C(0)$ 라고 가정할 수 있다면 $C(0)$ 를 문턱(Sill)이라고 부른다. 또한, $\lim_{h \rightarrow 0+} \gamma(h) = C_0$ 라 하여 이를 멩치효과(Nugget Effect)라 정의하며, $Z(s)$ 와 $Z(s+h)$ 가 더 이상 공간적으로 상관관계에 있지 않는 최소거리를 범위(Range)라고 한다. 실질적인 자료들간의 최소거리보다 작은 거리에서는 세미베리오그램에 대한 정보가 없기 때문에 멩치효과가 존재하게 되며, 세미베리오그램은 거리가 0인 점에서 불연속점을 가지게 된다. 강우강도가 범위 이상의 거리만큼 떨어져 있을 때 강우강도가 독립적인 관계를 나타내게 되고, 이처럼 강우강도가 범위 이상 떨어져 있을 때의 세미베리오그램은 강우강도의 분산과 같게 되는 데 이때의 세미베리오그램의 값이 문턱이 된다.

2.2. 공간예측

공간통계에서는 공간예측을 나타내는 용어로 크리깅(Kriging)을 사용하고 있는데 이는 Matheron(1963)이 광물탐사에서 사용한 이래 공간예측의 일반적인 기법으로 자리잡고 있다. 크리깅에는 범용 크리깅(Ordinary Kriging), 일반 크리깅(Universal Kriging) 등 여러 가지가 있으며 범용 크리깅이 간단하여 계산이 용이하며 일반크리깅을 비롯한 다른 크리깅 결과와 큰 차이가 없기 때문에 일반적으로 많이 이용되고 있다. 범용 크리깅은 다음의 가정하에서 공간예측으로 기술된다.

범용 크리깅의 가정 :

$$Z(s) = \mu + \delta(s), \quad s \in D, \quad \mu \in R, \quad \mu \text{는 알려지지 않은 상수.}$$

크리깅에서 관심있는 새로운 지점 s_0 에서의 예측치 $\hat{Z}(s_0)$ 는 s_j ($j = 1, \dots, n$) 지점에서 관찰된 값 $Z(s_j)$ 의 선형결합으로 표현된다. 즉, $\hat{Z}(s_0) = \sum_{j=1}^n \lambda_j Z(s_j)$, $\sum_{j=1}^n \lambda_j = 1$ 이다. 이때 λ_j , $j = 1, \dots, n$ 를 크리깅 계수라 하고 이는 목적함수

$$Q = E \left(Z(s_0) - \sum_{j=1}^n \lambda_j Z(s_j) \right)^2 - 2m \left(\sum_{j=1}^n \lambda_j - 1 \right) \quad (2.2)$$

를 최소화시키는 값이다. 여기서 m 은 제약조건 $\sum_{j=1}^n \lambda_j = 1$ 하에서 라그랑주 승수(Lagrangian Multiplier)이다.

식 (2.2)의 목적함수 Q 를 최소화시키는 크리깅 계수는 목적함수 Q 를 크리깅 계수 λ_j , $j = 1, \dots, n$ 과 라그랑주 승수 m 으로 각각 편미분한 $n+1$ 개의 연립방정식의 해가 되며, 이 연립방정식을 행렬형태로 나타내면 다음의 식 (2.3)과 같다.

$$\Gamma_0 \underline{\lambda}_0 = \underline{\gamma}_0 \quad (2.3)$$

여기서, $\underline{\lambda}_0 = (\lambda_1, \dots, \lambda_n, m)'$ 이고, Γ_0 는 (i, j) 번째 원소가

$$\Gamma_{0ij} = \begin{cases} \gamma(s_i - s_j), & i = 1, \dots, n, \quad j = 1, \dots, n, \\ 1, & i = n+1, \quad j = 1, \dots, n, \\ 0, & i = n+1, \quad j = n+1. \end{cases}$$

인 $(n+1) \times (n+1)$ 대칭행렬이고, $\underline{\gamma}_0 = (\gamma_1, \dots, \gamma_n, 1)'$ 이다.

위의 식 (2.3)에서 $\underline{\lambda} = (\lambda_1, \dots, \lambda_n)'$ 이고, Γ 는 (i, j) 번째 원소가

$$\Gamma_{ij} = \gamma(s_i - s_j), \quad i = 1, \dots, n, \quad j = 1, \dots, n,$$

인 $n \times n$ 대칭행렬이며, $\underline{\gamma} = (\gamma_1, \dots, \gamma_n)'$ 이고 $\underline{1}_n = (1, 1, \dots, 1)'$ 로 표현한다면, 크리깅 계수는 다음의 식 (2.4)와 같이 구하여진다.

$$\underline{\lambda} = \Gamma^{-1} \underline{\gamma} + \frac{\Gamma^{-1} \underline{1}_n \underline{1}'_n \Gamma^{-1} \underline{\gamma} - \Gamma^{-1} \underline{1}_n}{-\underline{1}'_n \Gamma^{-1} \underline{1}_n} = \Gamma^{-1} \left(\underline{\gamma} + \underline{1}_n \frac{1 - \underline{1}'_n \Gamma^{-1} \underline{\gamma}}{\underline{1}'_n \Gamma^{-1} \underline{1}_n} \right) \quad (2.4)$$

이렇게 구한 크리깅 계수는 평균제곱예측오차 $E(Z(s_0) - \hat{Z}(s_0))^2$ 를 최소화시킨다. 이 최소평균제곱예측오차를 공간통계학에서는 크리깅 분산이라고 부르고 σ_k^2 로 표기하며 다음 식 (2.5)와 같다.

$$\sigma_k^2 = \underline{\gamma}'\Gamma^{-1}\underline{\gamma} - \frac{(1 - \underline{1}'_n \Gamma^{-1} \underline{\gamma})^2}{\underline{1}'_n \Gamma^{-1} \underline{1}_n} \quad (2.5)$$

또한 공간변수 $Z(s_0)$ 가 정규분포를 따르기 때문에 $Z(s_0)$ 의 95% 예측구간은 다음의 식 (2.6)과 같다.

$$(\hat{Z}(s_0) - 1.96 \sigma_k(s_0), \hat{Z}(s_0) + 1.96 \sigma_k(s_0)) \quad (2.6)$$

그런데 식 (2.3)에서 볼 수 있는 바와 같이 크리깅을 위해서는 크리깅 계수를 계산하여야 하며, 크리깅 계수를 계산하기 위해서는 세미베리오그램에 대한 추정이 필연적이다. 이의 추정은 일반적으로 수학적 모형을 선택하고 그의 타당성을 경험적으로 비교하는 과정을 선택적으로 반복함으로써 이루어진다. 따라서 연구자의 경험과 능력에 의존하게 된다. 보다 자세한 공간예측기법에 대한 이론적인 설명은 Cressie(1993)를 참조하기 바란다.

3. 세미베리오그램의 추정

세미베리오그램에 대한 정의는 본 논문의 2.1절에서 살펴본 바 있다. 본 절에서는 세미베리오그램에 대한 추정이 가지는 의의와 모수적인 세미베리오그램의 종류에 대하여 살펴 보았다.

앞의 2.2절에서 살펴본 바와 같이 새로운 위치에 대한 공간변수의 값을 추정하기 위해서는, 크리깅 계수 $\lambda_j, j = 1, \dots, n$ 의 정확한 추정이 필요하며 λ_j 의 추정을 위해서는 세미베리오그램의 모형추정이 필연적이기 때문에 공간변수에 대한 정확하고 효율적인 추정은 세미베리오그램의 정확한 추정에 달려있다고 보아도 과언이 아니다.

3.1. 세미베리오그램 모형의 종류와 추정

모형적합을 위하여 기존에 제시되어온 모수적인 세미베리오그램 모형에는 구형(Spherical)모형, 지수(Exponential)모형, 가우지안(Gaussian)모형, 파동(Wave)모형, 승수(Power)모형, 그리고 David와 Yoo(1993)에 의해서 소개된 구분선형(Piecewise Linear)모형 등이 있다.

본 소절에서는 간단한 형태의 세미베리오그램 모형인 구분선형모형과 구형모형을 소개하고자 한다. 식 (3.1)은 구분선형모형의 함수식을 식 (3.2)는 구형모형의 함수식을 나타내고 있으며, 그림 3.1의 (a)와 (b)는 각각의 모형 형태를 보여주고 있다.

$$\gamma(h; \theta) = \begin{cases} 0, & h = 0, \\ C_0 + \frac{C_s}{a_s} \|h\|, & 0 < \|h\| \leq a_s, \\ C_0 + C_s, & a_s < \|h\|. \end{cases} \quad (3.1)$$

$$\gamma(h; \theta) = \begin{cases} 0, & h = 0, \\ C_0 + C_s \left\{ \frac{3}{2} \left(\frac{\|h\|}{a_s} \right) - \frac{1}{2} \left(\frac{\|h\|}{a_s} \right)^3 \right\}, & 0 < \|h\| \leq a_s, \\ C_0 + C_s, & a_s < \|h\|. \end{cases} \quad (3.2)$$

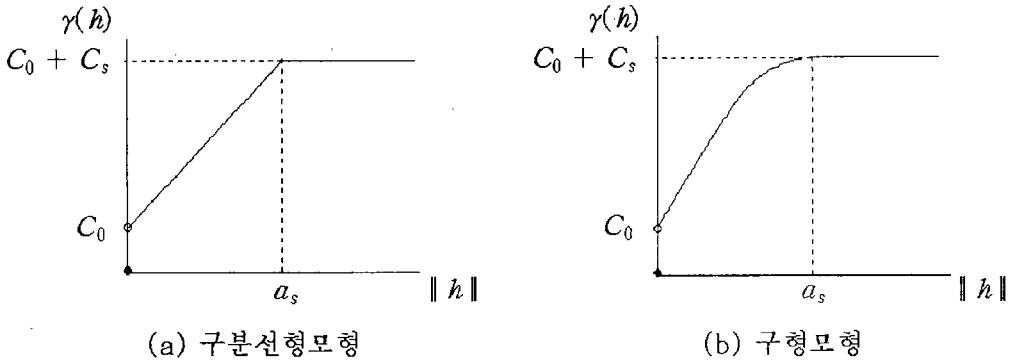


그림 3.1: 구분선형모형과 구형모형의 형태

여기서, 식 (3.1)의 구분선형모형과 식 (3.2)의 구형모형은 모두 $\theta = (C_0, C_s, a_s)'$, $C_0 \geq 0$, $C_s \geq 0$, $a_s \geq 0$ 를 만족하며, C_0 는 문턱효과를, $C_0 + C_s$ 는 문턱을, a_s 는 범위를 나타내며, 특히 C_s 를 부분문턱(Partial Sill)이라 부른다.

세미베리오그램의 추정은 세미베리오그램 모형이 가지고 있는 모수의 추정문제로 귀결된다. 본 논문에서 고려하고 있는 구분선형모형과 구형모형의 세 개의 모수를 추정하는 방법으로 보통최소제곱방법(Ordinary Least Square Method)을 채택하였다.

실증적 자료인 강우강도 자료를 이용하여 세미베리오그램 추정문제를 살펴보면, 2절의 가정에 의해서 세미베리오그램은 $\gamma(h) = \frac{1}{2}Var(Z(s_i) - Z(s_j)) = \frac{1}{2}E(Z(s_i) - Z(s_j))^2$ 이기 때문에 관측된 지점으로부터 얻은 $\frac{1}{2}(Z(s_i) - Z(s_j))^2$ 를 이용하여 세미베리오그램을 추정할 수 있다. 그림 3.2의 (a)는 32개 지역 강우강도 자료의 모든 가능한 조합으로부터 496($=_{32}C_2$)개의 $\frac{1}{2}(Z(s_i) - Z(s_j))^2$ 형태의 표본 세미베리오그램을 점으로 나타낸 후 구분선형모형을 적합시킨 결과를 선으로 보여주고 있다. 그림 3.2의 (b)는 구형모형을 적합시킨 결과이다. 여기서, 추정된 구분선형 세미베리오그램 모형의 모수는 $C_0 = 62.23$, $C_s = 232.19$, $a_s = 374$ 이

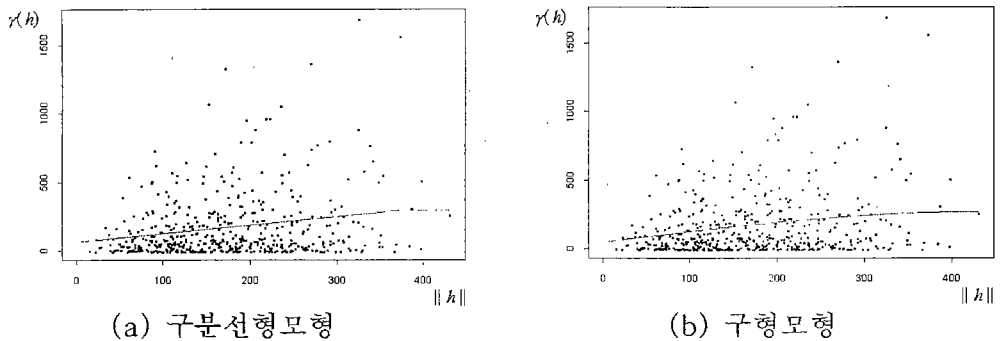


그림 3.2: 강우강도자료를 이용한 세미베리오그램의 모형적합

고, 구형모형의 모수는 $C_0 = 47.78$, $C_s = 222.42$, $a_s = 430.48$ 이다. 두 모형의 오차평방합을 계산한 결과 구분선형모형의 오차평방합이 2.586×10^7 로 2.591×10^7 인 구형모형의 오차평방합보다 작음을 알 수 있었다. 따라서 구분선형모형을 강우강도 자료의 세미베리오그램 모형으로 사용하였다.

이렇게 추정된 세미베리오그램 모형의 수치적 의미를 구분선형모형의 경우로 살펴보면, 강우강도의 세미베리오그램의 값은 거리가 0에서 0의 값을 가지며 멱치효과는 $62.23(=C_0)$ 이다. 또한 두 지점 사이의 거리가 증가함에 따라 두 지점 사이의 세미베리오그램의 값은 기울기 $0.62(=C_s/a_s)$ 를 가지는 선형증가추세를 보이며 이때의 두 지점의 강우강도 값 사이의 상관성이 점점 작아지다가, 두 지점 사이의 거리가 374km(범위) 이상일 때는 두 지점 사이의 세미베리오그램의 값은 294.42(문턱)로 이는 강우강도의 분산과 같고 이때의 두 지점의 강우강도 값은 독립이다.

그림 3.2에서 볼수 있듯이 본 논문에서 실증적 예로 사용하고 있는 강우강도 자료의 경우, 세미베리오그램 모형으로 구분선형모형을 사용한 경우와 구형모형을 사용한 경우에는 주목할 만한 차이가 나타나지 않았다. 하지만 세미베리오그램 모형에 대한 선택은 분석하고자 하는 자료의 종류에 따라 달라질 것이며 이의 선택은 당해분야의 실무적인 경험에 바탕을 두는 것이 바람직하다고 본다. 다시말해 강우강도자료의 분석을 위해 구분선형모형을 선택하였다고 하여 광물성분자료에 구분성분모형을 선택할 필요가 있다는 것은 아니다. 본 논문에서는 이후에 소개될 방법론들을 구현하기 위해서 보다 간단한 모형인 구분선형모형만을 사용하기로 하였으며 이는 본 논문에서 제시하는 방법론을 예시하는 측면에서 볼 때 구형모형을 사용하는 것과 전혀 차이가 없다.

그림 3.3의 (a)는 크리깅 기법을 이용한 강우강도에 대한 남한에서의 등고선그림이다. 제주도를 제외한 남한 전역을 1600개(40×40)의 지점으로 잘게 나누어 각 지점에 대해 크

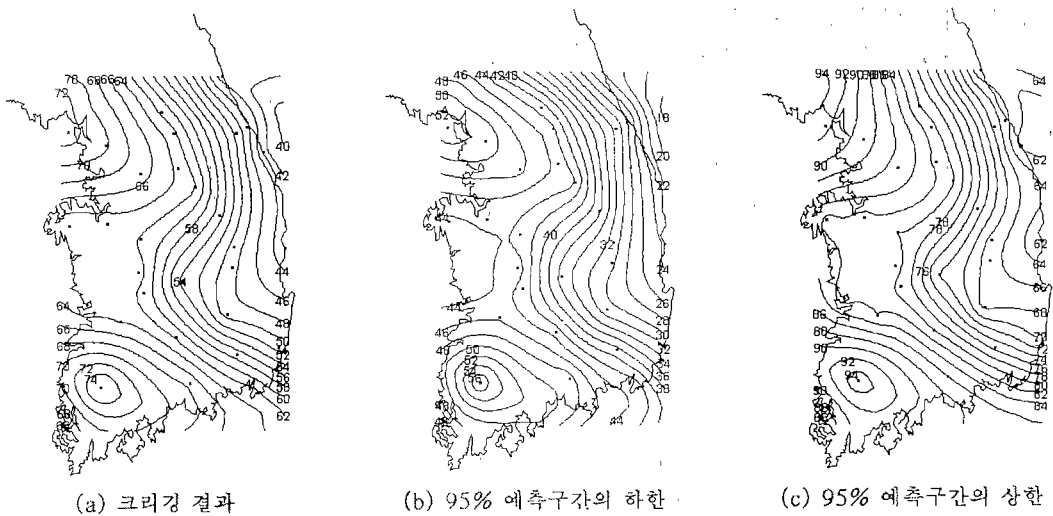


그림 3.3: 강우강도자료의 크리깅 결과와 95% 예측구간의 하한 및 상한

리깅한 자료를 보간한 후 등고선 그림으로 표현한 것이다. 크리깅을 위하여 사용한 세미베리오그램 모형으로 구분선형모형이 사용되었다. 그림 3.3의 (b)와 (c)는 (a)의 강우강도의 크리깅 자료에 대한 95% 예측구간의 하한과 상한을 각각 등고선 그림으로 표현한 것이다. 참고로 등고선 그림에서 작은 점은 강우강도가 관측된 32개 측후소의 위치를 나타내고 있다.

3.2. 래깅방법(LAGGING METHOD)

원자료를 이용하여 세미베리오그램 모형을 추정하는 것은 정확도면에서는 신뢰성있는 방법이나, 시간이 너무 많이 걸린다는 단점을 지니고 있다. 실제로 강우강도 자료의 이론적인 세미베리오그램을 추정하기 위해서 32개 지점으로부터 얻을 수 있는 496개의 점을 그래프상에 찍고 보통최소제곱방법을 이용하여 구분선형모형을 추정하는데 계산시간이 수십 초 정도(IBM Pentium 166기종 기준) 걸렸지만 고려하는 대상 지역이 많아질수록 소요되는 계산시간은 기하급수적으로 길어질 것이라고 전망된다. 따라서 원자료를 이용한 적합방법에 대한 단점인 시간문제를 해결하기 위하여 일반적으로 제시되는 방법이 래그(Lag)를 이용하는 방법이다.

Matheron(1962)은 고전적 세미베리오그램 추정량으로 식 (3.2)와 같은 방법을 제시하고 있다.

$$\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2 / 2, \quad h \in R^d. \tag{3.3}$$

여기서 h 는 고려하는 래그값에 의해 나뉜 각 래깅구간을 나타내는 벡터이고, $N(h) \equiv \{(s_i, s_j) : \|s_i - s_j\| \in h; i, j = 1, \dots, n\}$ 이고, $|N(h)|$ 는 벡터 h 에 의해서 나뉜 구간에 포함되는 서로 다른 쌍의 수이다.

그림 3.4는 강우강도 자료에서 몇몇 래그에 대한 실증적 세미베리오그램 모형을 구분선형모형에 적합시킨 결과를 보여주고 있으며, 표 3.1는 추정된 구분선형모형의 모수를 나타내고 있다.

이 방법은 앞 절의 원자료를 이용한 방법의 단점인 시간문제를 해결하기 위하여 거리에 래그를 주어 각 구간 안에 포함되는 표본 세미베리오그램의 산술평균을 해당 구간의 대

표 3.1: 래그를 이용한 강우강도자료의 세미베리오그램 모형적합 결과

	뭉치효과 (C_0)	부분문턱 (C_s)	문턱 ($C_0 + C_s$)	범위 (a_s)
(a) 원자료	62.2337	232.1870	294.4207	373.9999
(b) Lag = 10	12.2556	355.9131	368.1687	375.4790
(c) Lag = 20	34.1028	284.7120	318.8148	369.6407
(d) Lag = 30	44.9447	261.3464	306.2911	374.6407

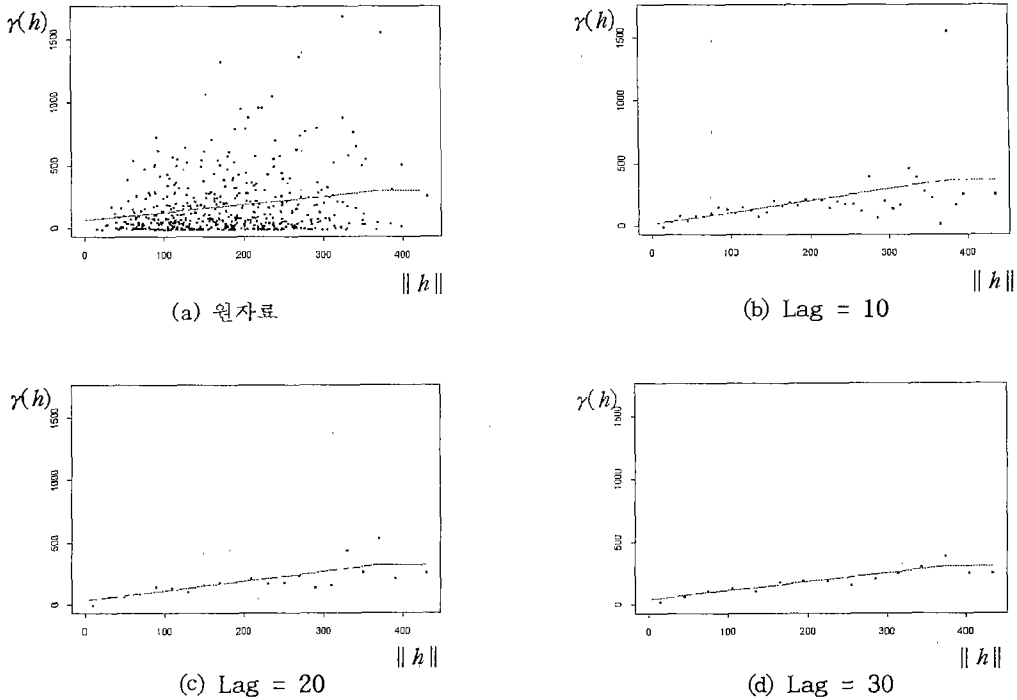


그림 3.4: 래그를 이용한 강우강도자료의 세미베리오그램 모형적합 결과

표값으로 사용하는 방법으로 원자료를 이용한 방법보다 계산시간을 많이 단축시킬 수 있다는 장점을 가지고 있다. 그러나 최근에는 컴퓨터의 성능 향상에 따라 시간문제는 어느정도 해결되었기 때문에 큰 장점이라고 보기는 어렵다. 또한 그림 3.4와 표 3.1를 통하여 원자료를 이용하는 방법과 래그를 이용하는 방법에는 다소 차이가 있음을 확인할 수 있었으며, 래그가 달라짐에 따라서 추정된 세미베리오그램의 형태에 차이가 있음을 확인할 수 있었다. 따라서 래그를 이용하는 방법을 이용할 때 원자료를 잘 설명해 주는 래그 선택의 기준에 대한 보다 집중적인 연구가 필요하다고 보지만, 연구자의 경험에 비추어 볼 때 원자료를 사용하는 방법이 타당하리라고 본다.

4. 공간이상치 탐지

본 절에서는 공간이상치 탐지 방법론으로 분포론적 접근방법과 p -Deletion 방법 2가지를 제시하고자 한다.

4.1. 분포론적 접근방법

2장에서 언급한 바와 같이 공간변수는 정규분포를 따른다. 따라서 $Z(s_i) - Z(s_j)$ 는 평균이 0이고 분산이 $2C(0) - 2C(s_i - s_j) \equiv 2\gamma(s_i - s_j)$ 인 정규분포를 따르고, 여기서, $\frac{|Z(s_i) - Z(s_j)|^2}{2\gamma(s_i - s_j)}$ 은

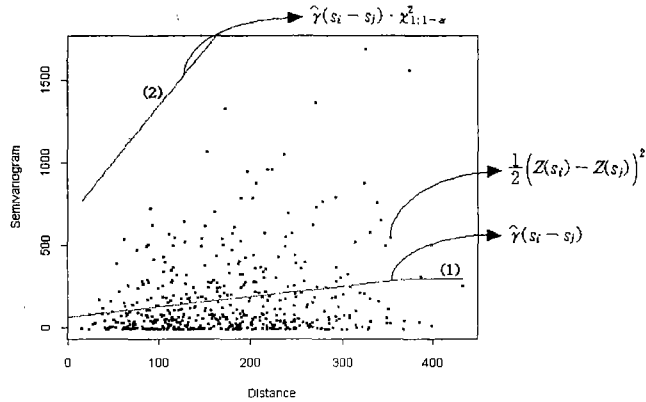


그림 4.1: 강우강도 자료에서 이상치 탐지를 위한 분포론적 접근방법 ($\alpha = 0.001$)

자유도가 1인 χ^2 분포를 따르게 된다.

또한 이론적인 세미베리오그램은 $\gamma(h) = \frac{1}{2}E(Z(s_i) - Z(s_j))^2$ 이기 때문에 관측된 지점으로부터 얻은 표본 세미베리오그램인 $\frac{1}{2}(Z(s_i) - Z(s_j))^2$ 를 이용하여 이론적인 세미베리오그램을 추정할 수 있다.

본 절에서 소개하고자 하는 공간 이상치 탐지 방법은 적절한 유의수준 α 에서 관측된 지점으로부터 얻은 표본 세미베리오그램 $\frac{1}{2}(Z(s_i) - Z(s_j))^2$ 이 $\hat{\gamma}(s_i - s_j) \times \chi^2_{1;1-\alpha}$ 보다 큰 경우를 공간이상치의 짝으로 하여 공간이상치를 가려내는 방법이다.

그림 4.1은 실증적 자료인 강우강도 자료를 이용하여 $\alpha = 0.001$ 하에서 분포론적 접근방법을 적용한 결과이다. 그림 4.1에서 그래프상의 점은 496($=_{32}C_2$)개의 표본 세미베리오그램 $\frac{1}{2}(Z(s_i) - Z(s_j))^2$ 을 나타내고 있고, (1)선은 표본 세미베리오그램을 토대로 추정된 구분선형 세미베리오그램 $\hat{\gamma}(s_i - s_j)$ 를 나타내고 있으며, (2)선은 유의수준 0.001%에서의 예측상한을 나타내고 있다. 분석결과 총 496($=_{32}C_2$)개의 표본 세미베리오그램 중 예측상한보다 큰 값을 가지는 경우는 발견되지 않음을 그림 4.1을 통해 알 수 있다. 이는 본 논문에서 방법론의 예시를 위하여 사용한 강우강도자료가 10여년간의 자료를 바탕으로 계산된 것임을 감안할 때 이상치로 간주될 지역별 강우강도가 나타나지 않음은 당연한 결과이다.

4.2. p-DELETION 방법

본 연구에서 소개하고 있는 공간이상치 탐지방법 중 두 번째 방법인 p-Deletion 방법은 특정 p 개 공간지점의 실제 데이터를 제외한 나머지 $n-p$ 개 공간지점의 데이터만을 이용하여 이론적인 세미베리오그램을 추정하고, 추정된 세미베리오그램을 이용하여 특정 p 개의 공간지점에 대한 크리깅을 시행한 후, 특정 p 개의 공간지점에 대한 크리깅 결과인 예측치와 이것에 대응되는 실제 데이터를 비교함으로써 이상치를 탐지하는 방법으로 Christensen(1992)에 의하여 제안된 방법이다.

p-Deletion 방법의 $Z(s_0)$ 에 대한 이상치 판단 기준은 유의수준 α 하에서 식 (4.1)의 예측

구간에 의존한다.

$$[\hat{Z}_{n-p}(s_0) - z_{1-\alpha/2} \cdot \sigma_{k;n-p}(s_0), \hat{Z}_{n-p}(s_0) + z_{1-\alpha/2} \cdot \sigma_{k;n-p}(s_0)] \quad (4.1)$$

식 (4.1)에서 $\hat{Z}_{n-p}(s_0)$ 는 총 n 개의 공간지점 중에서 s_0 공간지점을 포함한 p 개의 공간지점을 제외한 표본 세미베리오그램을 이용하여 s_0 공간지점의 변수값을 크리깅한 값 즉, 예측치를 나타내고, $\sigma_{k;n-p}$ 는 그 때의 크리깅 표준편차를 나타내며, $z_{1-\alpha/2}$ 는 표준정규분포의 $100(1 - \alpha/2)\%$ 누적확률값을 나타내고 있다.

p -Deletion 방법을 실증적 자료인 강우강도 자료에 적용한 결과 유의수준 0.001에서 1-Deletion 방법의 경우는 예측구간으로부터 벗어난 지역은 발견되지 않았으며, 2-Deletion 방법의 경우 광주지역이 한번 예측구간을 벗어난 것을 확인할 수 있었다. 또한 유의수준을 0.01로 설정하였을 경우는 1-Deletion 방법의 경우는 광주지역이 예측구간으로부터 벗어난 유일한 지역이었고, 2-Deletion 방법의 경우 역시 광주지역이 예측구간을 벗어난 31개 경우 중 30개 경우를 차지하고 있는 것을 확인할 수 있었다.

5. 시뮬레이션

앞에서 소개한 분포론적 접근방법과 p -Deletion 방법은 접근방법상으로는 서로 상이하나 공간이상치 탐지의 방법론이라는 점에서는 같은 맥을 가지고 있다. 본 연구에서는 이 두 방법론 중 어떤 방법론이 이상치 탐지에 효율적인지 알아보기 위하여 시뮬레이션을 시행하였다.

(1) 공간데이터 생성 : 제주도를 제외한 남한 전역을 400개(20×20)의 격자모양으로 나누어 이 중 40개의 지점을 임의로 선택한 후 선택된 40개의 지점에 대해 구분선형 세미베리오그램의 모수로 $C_0 = 50$, $C_s = 150$, $a_s = 250$ 을 가지는 공간변수를 발생시킨다. 이를 위하여 S-Plus의 `rfsim()` 함수를 이용하였다.

(2) 하나의 공간이상치를 가지는 데이터셋 생성 : 발생된 40개의 공간데이터를 토대로 구분선형 세미베리오그램의 모수를 추정하여 공간데이터의 분산 $\hat{\sigma}^2 = C_0 + C_s$ 을 추정한다. 이상치가 되기에 충분히 큰 값인 $4\hat{\sigma}$ (또는 $-4\hat{\sigma}$)를 40개의 데이터 중 i ($i = 1, \dots, 40$)번째 데이터에만 더하여 i 번째 데이터를 이상치로 만든다. 이렇게 하면 40개의 데이터 중 i 번째 데이터는 이상치이고, 이를 제외한 39개의 데이터는 정상치인 데이터셋이 생성된다.

(3) 공간이상치 탐지 : 하나의 공간이상치를 포함한 40개의 데이터로 이루어진 데이터셋을 이용하여 분포론적 접근방법과 p -Deletion 방법을 시행한다.

본 연구에서는 위의 전과정을 34번 반복시행하였으며, 그 결과를 표 5.1에 나타내었다. 공간이상치 판정기준에서 p -Deletion 방법의 경우에는 공간이상치로 탐지된 지역을 모두 공간이상치로 판정하면 되기 때문에 별다른 문제점이 없지만, 분포론적 접근방법의 경우에는 세미베리오그램의 예측상한을 벗어난 하나의 점에 대하여 두개의 지역이 관련되어 있기 때문에 이러한 경우 두 지역 모두 공간이상치라고 판정하기에는 무리가 있다. 따라서 분포론적 접근방법의 경우에는 세미베리오그램의 예측상한을 벗어난 점이 하나 일 경우에는 그 점에 해당하는 두 지역을 공간이상치로 판정하였으며, 세미베리오그램의 예측상한을 벗

표 5.1: 시뮬레이션 결과 (이상치가 1개인 경우)

탐지방법	분포론적 접근방법				1-Deletion 방법			
S.V.추정방법	방법 I		방법 II		방법 I		방법 II	
Outlier Status	Outlier	Non Outlier	Outlier	Non Outlier	Outlier	Non Outlier	Outlier	Non Outlier
Decision	O/N	O/N	O/N	O/N	O/N	O/N	O/N	O/N
1	33/ 7	12/1548	40/ 0	0/1560	40/0	71/1489	40/0	23/1537
2	19/21	7/1553	33/ 7	7/1553	40/0	6/1554	40/0	12/1548
3	20/20	4/1556	40/ 0	8/1552	40/0	17/1543	40/0	14/1546
4	40/ 0	0/1560	40/ 0	0/1560	40/0	47/1513	40/0	54/1506
5	36/ 4	7/1553	39/ 1	1/1559	40/0	40/1520	40/0	15/1545
6	40/ 0	0/1560	40/ 0	0/1560	40/0	40/1520	40/0	32/1528
7	37/ 3	13/1547	38/ 2	7/1553	40/0	42/1518	40/0	10/1550
8	39/ 1	4/1556	40/ 0	0/1560	40/0	33/1527	40/0	15/1545
9	32/ 8	16/1544	34/ 6	4/1556	38/2	11/1549	40/0	8/1552
10	39/ 1	0/1560	40/ 0	191/1369	40/0	44/1516	40/0	139/1421
11	32/ 8	7/1553	39/ 1	1/1559	40/0	29/1531	40/0	31/1529
12	38/ 2	9/1551	40/ 0	0/1560	39/1	29/1531	40/0	18/1542
13	39/ 1	1/1559	40/ 0	0/1560	40/0	25/1535	40/0	17/1543
14	23/17	13/1547	33/ 7	11/1549	40/0	23/1537	39/1	18/1542
15	34/ 6	12/1548	40/ 0	0/1560	40/0	17/1543	40/0	42/1518
16	32/ 8	23/1537	37/ 3	40/1520	39/1	13/1547	40/0	96/1464
17	40/ 0	1/1559	40/ 0	0/1560	40/0	51/1509	40/0	60/1500
18	31/ 9	17/1543	35/ 5	4/1556	40/0	19/1541	39/1	9/1551
19	31/ 9	16/1544	36/ 4	4/1556	40/0	21/1539	40/0	35/1525
20	22/18	11/1549	13/27	6/1554	40/0	28/1532	40/0	2/1558
21	34/ 6	9/1551	40/ 0	80/1480	40/0	71/1489	40/0	50/1510
22	29/11	5/1555	40/ 0	2/1558	40/0	40/1520	40/0	47/1513
23	36/ 4	7/1553	40/ 0	50/1510	40/0	41/1519	40/0	100/1460
24	40/ 0	2/1558	38/ 2	1/1559	40/0	41/1519	40/0	9/1551
25	31/ 9	9/1551	29/11	6/1554	40/0	25/1535	40/0	6/1554
26	23/17	17/1543	20/20	13/1547	37/3	43/1517	36/4	31/1529
27	30/10	10/1550	39/ 1	0/1560	40/0	29/1531	40/0	58/1502
28	39/ 1	7/1553	38/ 2	1/1559	40/0	27/1533	40/0	34/1526
29	21/19	9/1551	40/ 0	1/1559	40/0	21/1539	40/0	41/1519
30	23/17	12/1548	26/14	8/1552	37/3	46/1514	37/3	39/1521
31	27/13	10/1550	32/ 8	8/1552	39/1	21/1539	39/1	10/1550
32	35/ 5	5/1555	40/ 0	0/1560	40/0	23/1537	40/0	17/1543
33	39/ 1	9/1551	40/ 0	0/1560	40/0	46/1514	40/0	21/1539
34	40/ 0	0/1560	40/ 0	0/1560	40/0	39/1521	40/0	60/1500
평균	32.5/7.5	8.4/1551.6	36.4/3.6	13.4/1546.6	39.7/0.3	32.9/1527.1	39.7/0.3	34.5/1525.5

* Decision O : Outlier라고 판정하는 경우
 N : Non Outlier라고 판정하는 경우

표 5.2: 시뮬레이션 결과 (이상치가 3개인 경우)

탐지방법	분포론적 접근방법				1-Deletion 방법			
S.V.추정방법	방법 I		방법 II		방법 I		방법 II	
Outlier Status	Outlier	Non Outlier	Outlier	Non Outlier	Outlier	Non Outlier	Outlier	Non Outlier
Decision	O/N	O/N	O/N	O/N	O/N	O/N	O/N	O/N
01	37/83	0/1480	113/ 7	412/1068	106/14	12/1468	120/ 0	61/1419
02	49/71	0/1480	101/19	78/1402	92/28	8/1472	115/ 5	44/1436
03	36/84	2/1478	111/ 9	210/1270	97/23	7/1473	119/ 1	48/1432
04	36/84	3/1477	120/ 0	1399/ 81	119/ 1	10/1470	120/ 0	158/1322
05	43/77	4/1476	117/ 3	384/1096	114/ 6	6/1474	120/ 0	52/1428
06	44/76	6/1474	120/ 0	1375/ 105	118/ 2	16/1464	120/ 0	106/1374
07	43/77	2/1478	110/10	141/1339	108/12	8/1472	119/ 1	20/1460
08	44/76	3/1477	119/ 1	633/ 847	113/ 7	11/1469	120/ 0	69/1411
09	35/85	2/1478	106/14	74/1406	103/17	6/1474	119/ 1	24/1456
10	40/80	3/1477	120/ 0	1480/ 0	119/ 1	7/1473	120/ 0	300/1180
11	40/80	1/1479	119/ 1	739/ 741	108/12	3/1477	120/ 0	63/1417
12	43/77	3/1477	120/ 0	854/ 626	112/ 8	14/1466	120/ 0	56/1424
13	50/70	3/1477	120/ 0	1082/ 398	118/ 2	22/1458	120/ 0	69/1411
14	35/85	3/1477	92/28	67/1413	100/20	7/1473	114/ 6	45/1435
15	44/76	0/1480	120/ 0	865/ 615	112/ 8	12/1468	120/ 0	115/1365
16	39/81	5/1475	113/ 7	285/1195	97/23	5/1475	119/ 1	117/1363
17	32/88	4/1476	120/ 0	1356/ 124	119/ 1	23/1457	120/ 0	186/1294
18	42/78	4/1476	98/22	85/1395	97/23	15/1465	116/ 4	20/1460
19	43/77	5/1475	110/10	165/1315	110/10	12/1468	119/ 1	47/1433
20	40/80	2/1478	61/59	12/1468	104/16	6/1474	115/ 5	13/1467
21	41/79	3/1477	120/ 0	1116/ 364	111/ 9	26/1454	120/ 0	116/1364
22	32/88	1/1479	113/ 7	492/ 988	109/11	20/1460	120/ 0	49/1431
23	40/80	6/1474	120/ 0	1135/ 345	114/ 6	15/1465	120/ 0	151/1329
24	41/79	2/1478	111/ 9	217/1263	116/ 4	17/1463	120/ 0	32/1448
25	39/81	1/1479	95/25	35/1445	102/18	9/1471	117/ 3	19/1461
26	43/77	4/1476	67/53	20/1460	85/35	4/1476	109/11	36/1444
27	46/74	0/1480	119/ 1	693/ 787	111/ 9	6/1474	120/ 0	76/1404
28	48/72	5/1475	118/ 2	720/ 760	116/ 4	13/1467	120/ 0	51/1429
29	37/83	1/1479	120/ 0	245/1235	107/13	8/1472	120/ 0	70/1410
30	43/77	2/1478	76/44	16/1464	87/33	17/1463	106/14	36/1444
31	40/80	3/1477	94/26	49/1431	105/15	7/1473	118/ 2	33/1447
32	43/77	1/1479	120/ 0	550/ 930	109/11	6/1474	120/ 0	40/1440
33	47/73	2/1478	120/ 0	1108/ 372	107/13	21/1459	120/ 0	64/1416
34	40/80	5/1475	120/ 0	1468/ 12	118/ 2	14/1466	120/ 0	176/1304
평균	41.0/79.0	2.7/1477.3	109.5/10.5	575.3/904.7	107.7/12.3	11.6/1468.4	118.4/1.6	75.4/1404.6

어난 점이 두 개 이상일 경우에는 그 점에 두 번 이상 관계된 지역을 공간이상치로 판정하였다. 분포론적 접근방법에서 공간이상치 판정기준의 예를 들면, (4,14)와 같이 하나의 점만이 이상치로 탐지되었을 경우에는 4지역과 14지역을 모두 공간이상치로 판정하게 되며, (5,8) (5,16) (5,26) (8,26)과 같이 두 개 이상의 점이 이상치로 탐지되었을 경우에는 두 번 이상 탐지된 5, 8, 26지역을 공간이상치로 판정하게 된다.

(4) 3개의 공간이상치를 가지는 데이터셋 생성 및 공간이상치 탐지 : 과정 (1)에서 생성된 데이터를 가지고 과정 (2)와 비슷한 방법으로 하나의 데이터셋에 이상치 3개를 발생시킨 ${}_{40}C_3 = 780$ 개의 서로 다른 데이터셋 중 40개를 임의로 택하여 과정 (3)과 같은 방법으로 공간이상치를 탐지한다. 이를 34번 반복하였다. 이의 결과는 표 5.2에 정리되어 있다.

이러한 과정에서 세미베리오그램의 추정방법으로 두 가지를 사용하였다. 첫 번째 방법은 공간이상치 탐지를 위하여 발생된 데이터셋을 이용하여 추정한 세미베리오그램의 모수를 이용하는 방법(방법 I)이고, 두 번째 방법은 공간이상치탐지에 사용할 세미베리오그램의 모수로 최초 입력치인 $C_0 = 50, C_s = 150, a_s = 250$ 을 일률적으로 이용하는 방법(방법 II)이다.

표 5.1과 표 5.2의 결과를 이용하여 분포론적 접근방법과 p-Deletion 방법을 비교하기 위해서, 다음과 같은 총이득(Total Benefit) 함수를 고려할 수 있다.

$$(1 - \alpha)B_1 + (1 - \beta)B_2 = \text{총이득} \tag{5.1}$$

여기서, α 는 이상치를 이상치가 아니라고 판단할 제1종 오류(Type I Error)를 나타내고, β 는 이상치가 아닌데 이상치라고 판단할 제2종 오류(Type II Error)를 나타내고 B_1 은 이상치를 이상치라고 판단할 경우의 이득(Benefit)을 나타내고 B_2 는 이상치가 아닌 것을 이상치가 아니라고 판단할 경우의 이득을 나타낸다. 식 (5.1)에서 B_1 과 B_2 가 $B_1 = wB_2$ 의 관계를 가졌다고 할때, 예를들어, $B_1 = B_2 = 1$ 즉, w 가 1이면, 옳은 판단의 확률 $1 - \alpha$ 와 $1 - \beta$ 의 합이 꼴로 총이득 함수는 정의된다. 따라서 표 5.1에서 분포론적 접근방법의 경우, 방법 I의 총이득 값은 $\frac{32.5}{40} + \frac{1551.6}{1560} = 1.8071$ 이고, 방법 II의 총이득 값은 $\frac{36.4}{40} + \frac{1546.6}{1560} = 1.9014$ 이다. 마찬가지로, 1-Deletion 방법에서는 각각 1.9714, 1.9704로 계산된다. 표 5.3은 $B_1 = B_2 = 1$ 즉, w 가 1일 때의 총이득을 정리하였다.

그림 5.1은 w 에 따른 총이득을 나타내고 있다. 여기서 곡선 (a)와 (b)는 각각 분포론적 접근방법에서의 방법 I과 방법 II의 총이득을 나타내고, (c)와 (d)는 각각 1-Deletion 방법에서의 방법 I과 방법 II의 총이득을 나타낸다.

표 5.3: 시뮬레이션 결과에 대한 총이득 비교 ($B_1 = B_2 = 1$ 인 경우)

탐지방법	분포론적 접근방법		1-Deletion 방법	
	방법 I	방법 II	방법 I	방법 II
S.V.추정방법				
이상치 1개 발생	1.8071	1.9014	1.9714	1.9704
이상치 3개 발생	1.3398	1.5238	1.8897	1.9358

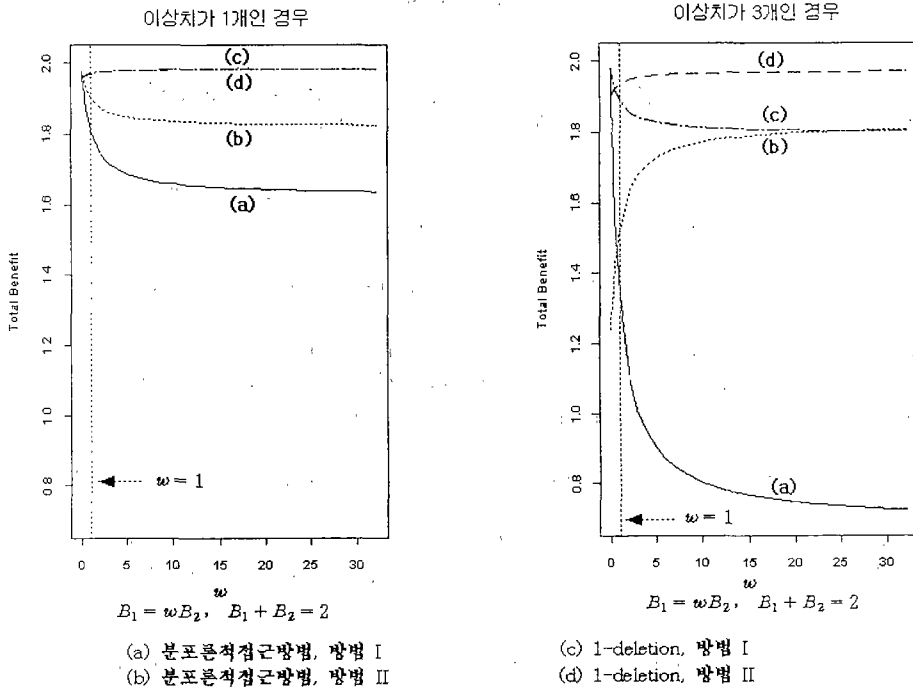


그림 5.1: 시뮬레이션 결과에 대한 총이익 비교

따라서 그림 5.1을 토대로 살펴 본 결과, 공간이상치 탐지에서 세미베리오그램 문제는 당연한 결과이겠지만 방법 I보다는 방법 II를 사용하는 것이 효율적인 것으로 나타났다. 또한 공간이상치 탐지 방법론의 비교결과 분포론적 접근방법 보다는 p-Deletion 방법이 공간 이상치 탐지에 효율적인 방법인 것으로 나타났다.

6. 결론

본 연구에서는 서로 다른 위치에서 동시에 관찰된 자료들이 공간적인 변인인 거리에 의해서 서로 상관되어 있을 때, 이들 공간 변수에 대한 세미베리오그램의 추정, 공간예측 및 공간이상치 탐지 등의 방법들을 정리 및 연구하였으며, 방법론의 실증적 예시를 위하여 강우강도 자료를 이용하였다.

세미베리오그램의 추정을 위해서는 원자료를 이용하는 것이 타당함을 제시하였으며, 공간이상치 탐지를 위해서 분포론적 접근방법과 p-Deletion 방법을 제시하였다. 시뮬레이션을 통하여 공간이상치 탐지를 위한 두 가지 방법론을 검토한 결과 p-Deletion 방법이 효과적임을 살펴볼 수 있었다.

참고문헌

- [1] 유성모, 엄익현 (1997). 강우강도 자료를 이용한 Semi-variogram 의 추정과 공간예측. <1997 춘계학술발표회논문집>. 65-70. 한국통계학회.
- [2] Christensen, R., Johnson, W. and Pearson, L. M. (1992). Prediction Diagnostics for Spatial Linear Models, *Biometrika*. Vol. 79, 3, 583-91.
- [3] Cressie, N. (1993). *Statistics for Spatial Data*. New York, John Wiley & Sons, INC.
- [4] David, H.T. and Yoo, S. (1993). The Best of Both Worlds: Integrating Statistical and Deterministic Approaches to Area Remediation, *Proceedings of the Fourth Annual International Conference on High Level Radioactive Waste Management*. 1418-1425.
- [5] Matheron, G. (1963). Principles of Geostatistics, *Economic Geology*. Vol. 58, 1246-1266.
- [6] McBratney, A.B. and Webster, R. (1963). Choosing Functions for Semivariograms of Soil Properties and Fitting Them to Sampling Estimates. *Journal of Soil Science*. Vol. 37, 617-639.
- [7] Yoo, S. and Um, I. (1997). Detection of Spatial Outliers with Rainfall Intensity Data. *The Ninth Korea and Japan Joint Conference of Statistics*, Korea Statistical Society and Japanese Society of Computational Statistics, 209-211.

[1998년 5월 접수, 1998년 10월 최종수정]

On the Estimation of Semivariogram and Spatial Outliers with Rainfall Intensity Data *

Seongmo Yoo ¹⁾ Ickhyun Um ²⁾

ABSTRACT

The closely located data in space are often more alike than those that are far apart. In this case, a spatial model incorporating this spatial correlation or dependence should be applied. We summarize a simple spatial prediction including Matheron's Kriging and we seek a seminal framework for the detection of spatial outliers. Two different methods for the detection of spatial outliers are proposed. One is based on classical distributional inference and the other is based on case deletion idea. Rainfall Intensity data and simulated data are applied for the application of our methodology.

* The authors wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997

1) Associate Professor, Department of Informational Statistics, Korea University, Jochiwon, 339-700, Korea

2) Graduate Student, Department of Statistics, Korea University, Anamdong, 136-701, Korea