

Lattice형 공간정보의 선형모형 추정방법

권 오 룡* 염 준 근**

요 약

공간자료의 선형모형은 지역적 특성과 해당지역의 격자형태의 통계량을 처리하여 추정하는 방법을 제시하고, 이에 대한 사례연구를 통하여 공간모형을 추정하여, 회귀분석방법과의 효율성을 비교 분석하였다. 공간정보의 모형추정은 최근에 국내 소개되었으며, 격자형 공간정보의 자기상관 여부는 Moran의 지수 및 이에 대한 Correlograms를 통하여 자료간의 인접성 유무를 식별할 수 있다. 따라서 공간모형의 회귀추정에서 상관계수 ρ 의 최우추정방법은 많은 컴퓨터 시간 및 경비가 소요되어 eigenvalue에 의한 방법으로 개선하였다. 캘리포니아주의 해안선과 육지지역의 식물종수에 대한 공간정보의 회귀모형을 추정하였다.

Estimation Methods for Linear Spatial Model on Lattice

Kwon, Oh-Ryong,* Yum, Joon-Keun**

ABSTRACT

Linear models for spatial data are proposed by example in this paper. This method was introduced to Korea for the first time in the early part of 1990's. The correlation of spatial patterns is computed by Moran Index, and then correlogram is proposed as the method to identify correlation of spatial patterns. Due to computational difficulties with ML, an alternative estimator has been used as an eigenvalue method.

1. 서 론

여러 분야의 자료들은 공간과 지역적 특성을 내포하고 있다. 즉 모든 자료는 지역적 위치와 관련되어 관찰되고, 이들은 지역과 공간의 특성을 함께 나타내고 있다. 이런 공간 정보에서 서로 인접한 지역간의 자료들은 떨어져 있는 자료들보다 연관성이 매우 높은 것처럼 보인다. 즉 자료들 간의 독립성의 가정이 충족되지 못한 상태의 자료를 모형화에 이용하고 있다.

공간 모형은 서로 다른 위치에서 측정된 자료들 사이에 존재하는 지역특성이 있는 자료를 이용하고 있다.

공간 자료를 이용한 모형 추정은 관찰지역의 평균치 추정뿐 아니라 관찰되지 않은 지역에 대

한 예측을 가능하게 해 주고 있다.

이런 공간 정보분석은 1950년대 지리학 분야에 계량적 통계기법의 적용으로 시작되어 1963년 Metern에 의해 공간예측방법이 개발되고 최근 Cressie, Rieply, Ord 등에 의해 공간 정보분석 과정에 대한 연구가 계속되어 공간자료를 이용한 예측방법을 실제 현상에 적용시키고 있다 [6].

본 논문에서는 공간 자료의 회귀모형에 대한 추정방법을 기술하고자 한다.

2. 공간정보의 특성치

공간정보모형에 대한 추정방법을 설명하기에 앞서 공간정보의 특성 및 관련된 특성치를 설명한다.

2.1 공간정보시스템의 특성

공간정보시스템의 확률모형에 대한 특성은 다

* 중신회원 : 한국건설기술연구원 기술정보센터장

** 비회원 : 동국대학교 통계학과 교수

논문접수 : 1995년 4월 6일, 심사완료 : 1995년 12월 29일.

음과 같다[9].

첫째, 공간자료는 지역 자체의 확률분포를 의미하는 것이 아니고, 관련된 변수들 간에 상호 의존성을 표시하기 위한 지역의 상대적 위치에 관한 특정한 정보를 필요로 한다.

둘째, 공간정보시스템이 계속 개발됨에도 불구하고 지역의 관찰자료는 한 순간에 이용 가능함을 가정한다. 즉, 시간적으로 한 시점에서 변수를 관찰할 수 있기 때문이다.

공간정보 모형에 사용되는 용어의 정의를 설명한다[5].

정의1) 격자(Lattices)는 공간지역의 셀 수 있는 규칙적 혹은 불규칙적 표현으로 식(2.1)로 정의할 수 있다. 이는 지도가 없으면 공간정보가 아니므로 좌표를 포함하여 표현한다.

$$D \equiv \{(x_n, y_i); i=1, \dots, 10 \dots \dots \dots (2.1)$$

x : 경도, y : 위도

정의2) 공간격자 "D"에서는 인접성(Neighborhood)을 식(2.2)로 지정할 수 있다.

D_n 는 공간격자라 하며 이는 공간적 의존성을 모델링하는데 있다.

공간격자 (D)에서는 인접성정보(Neighborhood information)을 지정할 수 있다. 예로서 모든 격자는 그 지역에서 인접을 i 번째 지역까지 30km 거리로 말할 수 있다.

$$N_i = \{k : k \text{는 } i \text{의 인접}\} \quad i=1, \dots, n, \\ D_n = \{(I, N); i=1, 2, \dots, n\} \dots \dots \dots (2.2)$$

N_i 는 격자 i 에 인접함.

정의3) 공간분포에 사용되는 거리는 동질성의 거리를 말한다.

(x, y) 와 (x_n, y_i) 의 거리는 $|x_i - x| + |y_i - y|$ 로 산출한다.

2.2 공간정보 자기상관

공간 자료는 지형상 지역 자료들이 조직적인 유형을 가질 때나 지역상의 변수값들이 체계적인 공간 변동을 할 때 자기상관(Autocorrelation)을 갖는다. 이는 지역상의 위치가 아닌 변수의 값에 의존된 정의이다. 즉 어떤 지역에서 관측된 변수들의 값은 일정한 유형을 갖음을 말한다.

공간 자료의 자기 상관 통계량으로 정의된 Moran의 통계량을 식 (2.2)로 정의할 수 있다.

$$I = \frac{n}{s_0} \frac{\sum \sum W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2} \dots \dots \dots (2.3)$$

$$s_0 = \sum \sum W_{ij} \quad (i \neq j)$$

지역 i 의 관찰치는 x_n , n 지역에 대한 $\{x_i\}$ 의 평균은 \bar{x} 로 표시된다. 여기서 W_{ij} 는 지역 (j, i)의 공간적 인접성을 나타내는 척도이다. 따라서 지역들이 연속이면 $W_{ij}=1$, 연속이 아니면 $W_{ij}=0$ 으로 표시한다.

2.3 Correlograms

Correlogram은 거리의 변화에 따라 발생하는 자기상관의 변화과정을 표나 도표형태로 표시한 것으로 이는 지도상의 거리에 따라, 가까운 거리에서 자기상관이 높아지고 거리가 멀어짐에 따라 자기상관이 작아짐을 보여준다. 거리간격은 Wd , $d=1, 2, \dots$ 로 표시하며, 두 지역간의 거리는 가까운 거리에서 계속 연결되는 먼거리까지 변화시킬 수 있다.

Moran의 지수 I에 대한 Correlogram은 두 지역간 거리 d 의 변화에 따른 거리간격을 적용하여 Moran의 지수(I)의 산정공식은 식(2.4)와 같다.

$$I = \frac{n}{s_0} \frac{\sum \sum W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2} \dots \dots \dots (2.4)$$

w_{ij} 는 d 의 변화에 따라 변화하며, 여기서 Moran의 지수 I공식에서 $d=1, \dots, n$ 로 변화됨에 따라 w_{ij} 및 s_0 의 값이 변한다.

2.4 공간처리 조건 확률

공간확률 변수의 처리는 최인접성 (nearest-neighbor)의 비동등성 정의에서 유래된다. 이는 1963년 Whittle이 제안하여 공식화 하였다. 만약 좌표(i, j)로 지정된 확률변수 $\{X_{ij}\}$ 와 관련된 사각격자가 존재할 경우 Lattice의 다른 것들은 관련성을 무시한다. Whittle는 변수 X_{ij} 의 결합 확률은 식(2.5)와 같다[2].

$$\prod Q_{ij}(X_{ij}; x_{i-1}, x_{i+1}, x_{j-1}, x_{j+1}) \dots (2.5)$$

x_i 는 X_i 의 값이다.

다시말하면 이는 사각격자의 관련성이 최인접성에만 의존함을 말해주고 있다.

3. 공간 정보의 인접성 모형

확률 변수 Y 의 자료는 n 개 지역에서 조사되었으며, 지역은 지점(Points) 혹은 지방(행정구역), 정규적이거나 비정규적 유형일 수도 있다. 표본 자료는 지역 자체를 정의하는 것이 아니고 특정 지역의 속성(인구, 식물의 키)을 지정한 것이다. 근접(Neighbor)은 엄밀하게 물리적 인접성을 표현하기 위해서 사용한다. 지역집합의 i 번째에 위치하는 Y_i 는 이들 변수들과 상호 관련이 있는 변수 Y_j 에 관한 모형을 공식으로 표현할 수 있다. 이런 변수 Y 의 모형은 Whittle의 결합할공식에 따른 제1계차 자기회귀모형을 식 (3.1)과 같이 표시할 수 있다.

$$Y_i = \alpha + \rho \sum W_{ij} Y_j + e_i, (i=1, \sim 2, \dots, n) \dots \dots (3.1)$$

여기서, α, ρ : 모수, $\{W_{ij}\}: \geq 0$: 지역 i 와 j 의 상호작용, \sum : over $j \in \mathcal{N}(i)$ 를 의미하고, 오차항 e_i 는 평균이 "0"이고 동일 분산을 갖는다.

3.1 최우추정법에 의한 추정

자기회귀모형 식 (3.1)의 모형을 행렬 형태로 표현하면 식 (3.2)과 같이 표시된다.[8]

$$Y = \rho W Y + \epsilon \dots \dots \dots (3.2)$$

$$\alpha = 0$$

$$W: n \times n \text{ 가중치 행렬}$$

$$Y, \epsilon: n \times 1 \text{ Vector}$$

$$\epsilon = AY, A = I - \rho W \dots \dots \dots (3.3)$$

만약 $\epsilon \sim N(0, \delta^2 I)$ 로 주어지면 $Y = y$ 조건하에 ρ 와 δ^2 의 지수형 우도 함수는 식 (3.4)과 같다.

$$L(\rho, \delta^2) = -(n/2) \ln (2\pi \delta^2) - (1/2\delta^2) Y A A Y + I_n |A| \dots \dots \dots (3.4)$$

식 (3.4)에서 σ^2 의 최우추정치는 식 (3.5)와 같다.

$$\delta^2 = n^{-1} Y A A Y \dots \dots \dots (3.5)$$

식 (3.6)을 최대화하는 ρ 의 값을 $\hat{\rho}$ 로 취할 수 있다.

$$L(\rho, \delta^2) = \text{const} - (n/2) \ln (|A|^{-2/n}) \dots \dots \dots (3.6)$$

여기서 $|A|$ 는 ρ 의 제 n 계차 다항식으로 이는 표본의 규칙적 형태를 가진다. 그러나 ρ 의 계산은 n 가 커지면 컴퓨터 시간이 많이 소요된다.

3.2 개선된 계산 방법

우도함수

$L(\rho, \sigma^2) = \text{const} - (n/2) \ln (\sigma^2, |A|^{-2/n})$ 에서 A 는 $|A| = |I - \rho W|$ 의 계산에 의존함으로 $\hat{\rho}$ 의 결정에 어려움이 많다. 그러므로 $\hat{\rho}$ 의 계산의 대체방안을 모색할 수 있다.

만약 ω 가 eigenvalue $\lambda_1, \dots, \lambda_n$ 을 가지면, 이는 식 (3.7)과 같이 기술할 수 있으며,

$$|\lambda - \omega| = \prod_{i=1}^n (\lambda - \lambda_i) \dots \dots \dots (3.7)$$

이에 따라 $|A|$ 는 $|A| = \prod_{i=1}^n (1 - \rho \lambda_i)$ 로 표현가능하다. 식 (3.7)에서 $\{\lambda_i\}$ 는 계산될 수 있고 $\hat{\rho}$ 는 식 (3.8)을 최소화하는 ρ 의 값이 된다.

$$\left\{ \prod_{i=1}^n (1 - \rho \lambda_i) \right\}^{-2/n} (y' y - 2 \rho y' y L + \rho^2 y L y L) \dots \dots \dots (3.8)$$

4. 자기회귀모형

공간정보는 지역적 인접성으로 인하여 공간적 자기상관 관계가 존재함으로서 일반 회귀모형의 기본 가정인 오차항 (ϵ)의 독립성이 존재하지 않는다. 그러므로 오차항의 Autocorrelation을 알고 있다는 가정하에 자기회귀모형의 추정방법을 고려하고자 한다.[4]

$$Y = X B + e \dots \dots \dots (4.1)$$

오차항 e_i 에 Autocorrelation 함수관계는 식 (4.2)와 같다.

$$e = \rho \omega e + \mu$$

$$\mu = (I - \rho \omega) e = A e \dots \dots \dots (4.2)$$

여기서 μ 는 $E(\mu)=0, var(\mu)=\delta^2 I$ 이다. 오차항 e_i 와 e_j 의 함수관계는 식 (4.2)와 같으므로 공분산은 행렬 μ 의 요소이다. 만약 $e \sim N(0, \delta^2 D)$ 이면 식 (4.2)로부터

$$D^{-1} = (I - \rho\omega)' (I - \rho\omega) = A' A$$

$$V = \delta^2 ((I - \rho\omega)' (I - \rho\omega))^{-1} \text{이 된다.}$$

오차항이 공간적 자기상관을 가지는 모형의 계수 추정을 위하여 최우추정법을 적용한다. 이를 위해 일반적 최소자승법의 우도함수는 식 (4.3)과 같다.

$$L = \frac{1}{\delta^n (2\pi)^{n/2}} \exp\left(\frac{-e' e}{2\delta^2}\right) \dots\dots\dots (4.3)$$

식 (4.3)의 e 는 공간적 자기상관이 존재함으로서 식 (4.2)를 대입하면 식 (4.3)은 식 (4.4)와 같이 표시할 수 있다.

$$L = |A| \frac{1}{\delta^n (2\pi)^{n/2}} \exp\left\{\frac{-(Ae')(Ae)}{2\delta^2}\right\} \dots\dots\dots (4.4)$$

식 (4.4)의 지수를 취하면

$$\log L = const - \frac{n}{2} \log \delta^2 - \frac{1}{2\delta^2} (e' A' Ae) + \log |A| \dots\dots\dots (4.5)$$

식 (4.5)의 e 를 식 (4.1)로 대체하면

$$\log L = const - \frac{n}{2} \log \delta^2 - \frac{1}{2\delta^2} (Y' A' AY) - 2bX' A' AY + (b' X' A' AXb) + \log |A| \dots\dots\dots (4.6)$$

식 (4.6)에서 ρ 를 알고 있다는 가정하에서 b 와 δ^2 의 최우추정치는 $\log L$ 의 1차도함수를 0으로 놓고 풀면 된다.

$$\frac{\partial L}{\partial b} = -\frac{1}{\delta^2} (-2X' A' AY) + 2X' A' AXb$$

$$\frac{\partial L}{\partial \delta^2} = -\frac{n}{2\delta^2} + \frac{1}{2\delta^2} (u' u) \dots\dots\dots (4.7)$$

$$b = (X' A' AX)^{-1} X' A' AY$$

$$\delta^2 = \frac{1}{N} (Y' A' AY - 2bX' A' AY + bX' A' A' Xb) \dots\dots\dots (4.8)$$

식 (4.8)의 \hat{b} 는 최소자승 추정치의 형태는 동일하다. 그러나 $A=(I-\rho\omega)$ 라 정의 하였으므로 먼저 ρ 의 추정치를 얻어야 한다. ρ 의 추정도 최우추정방법을 적용할 수 있다. 왜냐하면 행렬 A 에 ρ 가 포함되어 있기 때문이다. 그러므로 \hat{b} 는 $\log L$ 을 ρ 에 관하여 최대화할 때 얻을 수 있다. 즉, 식 (4.8)의 δ^2 을 식 (4.6)에 대입하여 정리하면

$$\log L = const - \frac{n}{2} \log \delta^2 + \log |A| \dots\dots\dots (4.9)$$

식 (4.9)에 $\hat{\delta}, \hat{b}$ 로 $\hat{\rho}$ 의 값을 계산하기 위하여 $\log L$ 을 최대화하는 것은 식 (4.10)을 최소화하는 것과 동일하다.

$$M = \log \delta^2 - \frac{2}{n} \log |A| \dots\dots\dots (4.10)$$

$$M' = \log(Y' A' PAY) - \frac{2}{n} \log |A| \dots\dots\dots (4.11)$$

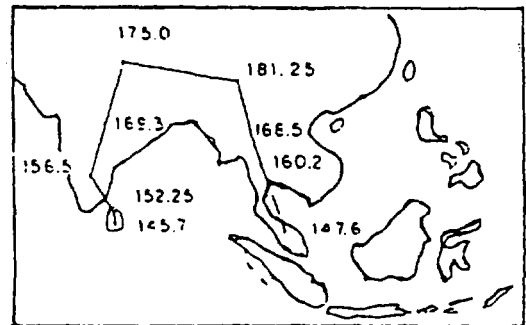
$\hat{\rho}$ 는 식 (4.11)의 값을 최소화하는 값으로 얻을 수 있다.

$$\rho = \min M' \dots\dots\dots (4.12)$$

M' 의 최소화는 직선적 방법인 시행착오 방법으로 선택할 수 있으며, M' 는 $-1 \leq M' \leq 1$ 사이의 값이 된다.

5. 공간모형추정

공간정보모형에서 자기상관 및 회귀모형을 추정한다. 여기에 사용되는 통계는 「Spatial Data



(그림 5.1) 새의 이동도
(Fig 5.1) Moving Map of the Bird

Analysis」 [5]에서 사용한 자료를 인용한다.

5.1 공간자기상관 추정

관찰지역내 새의 이동을 연결한 자료는(그림 5.1)을 근간으로 한다. 여기서 공간정보의 자기상관을 추정하기 위해 식 (2.3)으로 계산한다.

Moran의 자수를 계산하기 위하여 먼저 W행렬을 <표 5.1>과 같이 표시할 수 있다.

(표 5.1) W 행렬
(Table 5.1) W matrix

지역	A	B	C	D	E	F	G	H	I
A	0	1	0	0	0	0	0	0	0
B	1	0	1	0	0	0	0	0	0
C	0	1	0	1	0	0	0	0	0
D	0	0	1	0	1	0	0	0	0
E	0	0	0	1	0	1	0	0	0
F	0	0	0	0	1	0	1	0	0
G	0	0	0	0	0	1	0	1	0
H	0	0	0	0	0	0	1	0	1
I	0	0	0	0	0	0	0	1	0

여기서 $s_0 = \sum \sum W_{ij} = 16$

$$I = \frac{n}{s_0} \frac{\sum \sum W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2} = 0.603$$

여기서 Moran의 자수 I=0.603은 새의 이동과 날개 길이에 대한 공간자기 상관이 존재함을 알 수 있다.

또한 Moran 지수 I 공식에서 d=1, 2, 3, ..., n로 변화시키면서 W_{ij} , S_0 의 값을 계산하면 <표 5.2>와 같은 Correlogram을 계산할 수 있다.

(표 5.2) Moran의 correlogram
(Table 5.2) Correlogram of Moran

d	s_0	s_1	s_2	L_d	$var(L_d)$	z	r_d
1	960	1920	14624	0.064	0.002	1.5	0.113
2	896	1792	12928	0.100	0.002	2.2	0.191
3	832	1664	11296	0.113	0.002	2.4	0.195
4	768	1536	9728	-0.093	0.003	-1.7	-0.160
5	704	1408	8,224	0.028	0.003	0.6	0.048
6	640	1280	6784	-0.055	0.003	-0.9	-0.082
7	576	1152	5408	-0.013	0.003	-0.2	-0.019

<표 5.2>에서는 d가 4보다 적으면 자기상관이 분명히 존재하고 d>4는 자기상관이 없음을 알 수 있다.

5.2 공간모형 추정

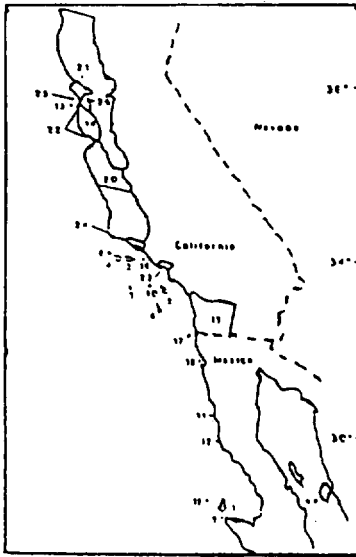
공간정보의 선형모형 추정 미국, 영국에서는 많은 사례를 찾을 수 있다. 그러나 국내에는 아직 공간의 특성을 배제한 일반선형모형 추정방법을 대부분 이용하고 있는 실정이므로, 공간의 특성을 지닌 모형의 추정은 설명변수의 모든 성질을 반영할 수 없다. 그러므로 공간회귀모형추정은 Johnson et al (1968)과 Johnson and Raven (1970)이 공간정보모형의 연구에 사용한 미국 캘리포니아주의 섬들과 해안선에 관련된 식물종수에 관련된 통계 자료[5]를 이용한다.

(표 5.3) 캘리포니아지역 자료
(Table 5.3) The data of California

지역	Y	X_1	X_2	X_3
1 Cedros Island	205	134	3950	28.2
2 Guadalupe Island	163	98	4600	29.0
3 Santa Cruz Island	420	96	2470	34.0
4 Snta Rosa Island	340	84	1560	34.0
5 Island Catalina Island	392	75	2125	33.3
6 San Clemente Island	235	56	1965	32.9
7 San Nicolas Island	120	22	910	33.2
8 San Miguel Island	190	14	830	34.0
9 Natividad Island	42	2.8	490	27.9
10 Santa Barbara Island	40	1.0	635	33.4
11 San Martin Island	62	0.9	470	30.5
12 San Geronimo Island	4	0.2	130	29.8
13 South Farallon Islan	12	0.1	360	37.7
14 Ano NUEVO	40	0.02	60	37.1
15 San Benito Island	39	2.5	660	28.3
16 Anacapa	70	1.1	930	34.0
17 Los Coronadas	83	1.0	670	32.4
18 Todos Santa Island	72	0.5	315	31.8
19 Monterey County	1450	4260	6535	33.0
20 Monterey County	1400	3324	5860	36.2
21 Marin County	1060	529	2610	38.1
22 Santa Cruz Mountains	1200	1386	3810	37.3
23 Santa Monica Mountains	640	320	3110	34.1
24 Santa Barbara Area	680	110	3985	34.4
25 San Francisco County	640	45	930	37.8
26 Tiburon Peninsula	370	5.9	750	37.9

Y=식물종수 X_1 =면적 X_2 =고도 X_3 =위도

종속변수 Y는 식물의 종수, 설명변수 x_1 : 지역면적, x_2 : 지역의 고도, x_3 : 위도 등의 자료는 <표 5.3>과 같고 지역도는 (그림 5.2)에 나타나 있다. 이들 자료를 이용한 공간 선형모형을 추정하고자 한다.



〈그림 5.2〉 캘리포니아 지역도
(Fig 5.2) location map of California

5.2.1 W행렬

W행렬은 표본지역의 공간적 인접성의 성질을 표시하며, 본 자료에서는 육지간의 고립성을 나타내는 거리에 따라 W행렬을 표시한 것이다.

〈표 5.4〉 가중치 W행렬
(Table 5.4) W Weighted matrix

i	$w_{i,j}$ [J는 괄호안 수]
19 (23)	1,0000
20 (21)	0.1162; (22) 0.3285; (23) 0.0752; (24) 0.1708; (25) 0.1666; (26) 0.1427
21 (20)	0.0082; (22) 0.0493; (25) 0.3016; (26) 0.6409
22 (20)	0.0797; (21) 0.0933; (24) 0.6928
23 (19)	0.2139; (21) 0.1693; (25) 0.4759; (26) 0.2651
24 (20)	0.2342; (23) 0.7658;
25 (20)	0.0080; (21) 0.2041; (22) 0.0939; (26) 0.6941
26 (20)	0.0057; (21) 0.3649; (22) 0.0456; (25) 0.5838

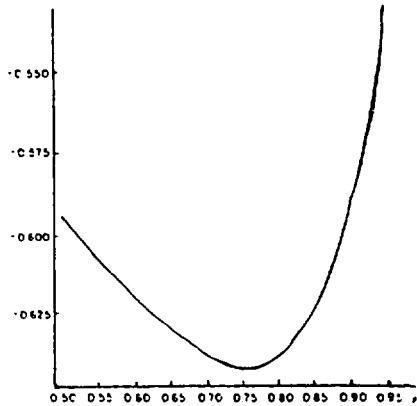
섬지역이 육지와 떨어진 거리가 187.5mile 이상 떨어져 있으면 고립성, 그렇지 않으면 상대비율 즉, 거리/187.5 비율로 고립성을 표시하는 가중치 행렬을 의미한다. 따라서 섬지역은 $w_{i,j}=1, i=j$ $w_{i,j}=0$ $i \neq j$ 이 된다. 그러므로 육지지역은 다음과 같은 조건으로 가중치를 결정할 수 있다.

- cut-off 거리 (k)가 187.5mile로 결정되었으므로
- $w'_{i,j}=d$ $b_{i,j} \leq k$
- $w'_{i,j}=0$ $b_{i,j} > k$

○ $w_{i,j}=w'_{i,j}/\sum w_{i,j}$ 로 가중치를 계산하며, 그 값은 〈표 5.4〉와 같다.

5.2.2 ρ 계산

공간모형의 지역간의 자기 상관관계를 표시하는 ρ 의 계산은 식 (4.11)에 따라 ρ 의 값을 $0 < \rho < 1$ 사이에서 변화시키면서 M^2 을 계산하고 이들 값중 최소치를 ρ 의 추정치로 결정한다. 본 모형에서 공간적 특성을 나타내는 W행렬과 캘리포니아주의 자료를 이용하여 식 (4.11)을 계산하기 위하여 SAS의 공간모형 모듈[10]을 사용하여 계산된 ρ 와 M^2 의 그래프 (그림 5.3)에서 $\rho = 0.75$ 시 M^2 가 최소치가 된다.



(그림 5.3) 캘리포니아 주 자료에 대한 M^2 와 ρ 의 그래프
(Fig 5.3) Graph of M^2 and ρ [California]

5.2.3 회귀계수 추정

미국 New York주 Syracuse University 지리학자 Daniel Griffith 교수팀에서 개발한 SAS의 공간모형 모듈[10]을 사용하여 캘리포니아주의 식물종수를 추정하는 선형모형은 일반회귀

〈표 5.5〉 회귀모형 추정치
(Table 5.5) Estimate of Regression Model

모수	공간회귀분석			일반회귀		
	추정치	표준오차	t-값	추정치	표준오차	t-값
b_0	-860.9	382.4		-1668.2	370.0	4.51
b_1	0.1479	0.0374	3.95	0.1642	0.0502	4.51
b_2	0.1017	0.0247	4.12	0.1165	0.2801	3.87
b_3	27.0610	11.4761	2.36	52.5103	10.8296	4.84
ρ	0.75	0.1342	5.19	-	-	-
δ_2	0.0102	0.0052		-	-	-
γ^2	0.976			0.816		
F				51.63		

(OLS)방법과 공간모형방법으로 추정한 결과치는 <표 5.5>와 같다.

공간회귀방법에 의한 회귀선 결정계수는 $r^2=0.976$ 이고 최소 자승법 회귀선의 결정 계수 $r^2=0.816$ 로 공간정보모형 방법이 설명력이 훨씬 높다.

또한 두 가지 분석방법에 의한 잔차를 비교하면 OLS방법은 3δ을 벗어나는 이상치가 2개(13, 21), 2δ을 벗어나는 이상치는 3개(3, 14, 21)을 판별되나 공간분석, 공간분석 방법으로는 이상치를 발견할 수 없다.

6. 결 론

Lattice형 공간자료의 분석시 일반 회귀분석과 상이점은 공간자기상관계수에 필요한 가중치행렬 w 의 결정이 가장 중요한 역할을 한다. w 의 결정은 지역적 동질성에 따라 여러가지 방법으로 선정할 수 있다. 즉 인접성의 성질에 따라 동질성의 거리, 고립도, 방향성, Lattice 간의 접근성 등 여러 방법중 선정하여야 하나 Moran의 I지수의 Correlograms에 따라 자기상관계수의 유무를 판정한 후 d 의 값을 선정하여 w 의 행렬을 결정할 수 있다.

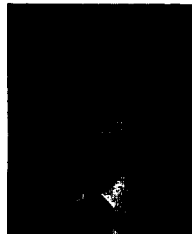
본 논문에서 공간정보의 선형모형추정에서 예시된 것 같이 지역적 모형은 지역상관계수 ρ 를 포함시킨 선형모형의 추정은 최량의 모형을 선정하는 것이 가능하나 향후 이들의 컴퓨터 응용을 위한 Program의 개발적용으로 공간정보의 모형개발이 실용화 할 수 있는 연구가 절실히 요구된다.

참 고 문 헌

- [1] 남궁평, 변종석, 공간표본자료를 이용한 예측방법 pp33-47, 1994.
- [2] Besag. J.E, Spatial Interaction and the Statistical analysis of Lattice Systems, JRSS, B, 36, No. 2, 1974.
- [3] Besag.J.E, Nearest Neighbour Systems and the Auto Logistic Model for Binary Data, JRSS B Vol. 34, pp75-83, 1975.
- [4] Cressie, Statistics for Spatial Data, New

York, Wiley

- [5] G.UPTON and B Fingleton, Spatial Data Analysis, U.K, John Wiley, pp172, pp273
- [6] Gleeson A.C. and Cullis B.R., Residual Maximum likelihood Estimation of neighbour model for field experiment, Biometrics Vol.43 pp277-288, 1987.
- [7] Green P. Jennison. C, and Seheult. A, Analysis of Field Experiments by least squares smoothing. JRSS B Vol. 47 pp299-315, 1985.
- [8] Keight Ord, Estimation Methods for models of spatial interaction, JASA pp120-126, 1970.
- [9] Ripley. B.D, Spatial Statistics, New York, Wiley
- [10] Daniel A. Griffith Spatial Regression Analysis on the PC : Spatial Statistics Using SAS, Institute of Mathematical Geography, 1992.



권 오 통

1977년 고려대학교 정경대학 통계학과졸업
 1984년 고려대학교 대학원 통계학석사
 1994년 동국대학교 대학원 통계학과 박사과정 수료
 1977년~79년 한국과학기술연

구원 전산개발센터

1979년~84년 산업연구원 전산실 책임연구원
 1984년~94년 한국건설기술연구원 정보관리실장
 1995년~현재 한국건설기술연구원 건설기술정보센터장

관심분야 : 전산수학, 데이터베이스, GIS, 회귀분석



염 준 군

1975년 동국대학교 통계학과 졸업
 1979년 동국대학교 대학원 통계학(이학석사)
 1985년 서울대학교 대학원 통계학(이학박사)
 1979년~현재 동국대학교 통계

학과 교수

관심분야 : 회귀분석, 표준조사론