Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 4, pp. 1141~1150

Spatial Cluster Analysis for Earthquake on the Korean Peninsula

Chang Wan Kang[1] · Sung Ho Moon[2] · Jang Sik Cho[3]
Jeong Hyeong Lee[4] · Seung Bae Choi[5] · Soo Gyun Beum[6]

Abstract

In this study, we performed spatial cluster analysis which considered spatial information using earthquake data for Korean peninsula occurred on 1978 year to 2005 year. Also, we look into how to be clustered for regions using earthquake magnitude and frequency based on spatial scan statistic. And, on the basis of the results, we constructed earthquake map by earthquake outbreak risk and gave a possible explanation for the results of spatial cluster analysis.

Keywords : 공간검색통계량, 공간 군집분석, 시-공간 조합 모형, 정규모형

1. 서론

최근 들어 동북아시아 일대의 빈번한 지진 발생으로 엄청난 자연 재해를 불러 일으켰다. 예를 들면, 1976년에 발생한 리히터 규모 7.6의 중국의 탕샹 지진은 242,000여명이 사망하였고, 1995년에 발생한 리히터 규모 7.2이상인 일본 고베 지진으로 6,433여명이 사망하였으며, 2004년에 발생한 리히터 규모 9.0인 인도네시아 수마트라 지진과

1) Corresponding Author : Associate Professor, Department of Data Information Science, Dongeui University, Busan, 614-714, Korea E-mail : cwkang@deu.ac.kr
2) Professor, Department of Statistics, Pusan University of Foreign Studies, Busan, 608-738, Korea E-mail : shmoon@pufs.ac.kr
3) Associate Professor, Department of Statistical Information Science, Kyungsung University, Busan, 608-736, Korea E-mail : jscho@star.ac.kr
4) Associate Professor, Division of Management Information Science, Dong-A University, Busan, 604-714, Korea E-mail : jeonglee@dau.ac.kr
5) Assistant Professor, Department of Data Information Science, Dongeui University, Busan, 614-714, Korea E-mail : csb4851@deu.ac.kr
6) Additional Job Professor, Department of Data Information Science, Dongeui University, Busan, 614-714, Korea E-mail : sgbeum@deu.ac.kr

함께 거대한 지진해일로 283,100여명의 인명피해를 입혔다. 최근 50년간(1941-1990) 세계 각지에서 발생한 규모 7이상의 지진은 약 500회에 이르며, 전 세계 지진의 약 15%가 일본 지역에서 발생하는 것으로 알려져 있다. 특히 2005년에 리히터 규모 7.0으로 발생한 일본 후쿠오카 지진은 인명피해는 없었지만, 울산 및 부산 경남일대에도 지진을 감지했을 정도로 우리나라에 지진의 영향을 간접적으로 주었다.

우리나라에서는 최근까지는 지진에 의한 직접적인 피해는 받고 있지 않지만 주변 국가, 특히 일본의 경우 지진의 빈도가 높을 뿐만 아니라 지진의 규모 또한 5이상의 강진으로 발생하고 있어 인접된 우리나라의 경우에도 일본과 같은 지진이 언제 어떤 규모로서 발생할 지에 대해서는 아무도 장담할 수 없는 상황이다. 이러한 최근의 지진 발생으로 인하여 국내의 지진에 대한 관심이 높아졌고, 지질분야뿐만 아니라 많은 지진관련 학문분야에서 지진에 대한 연구 또한 활발히 진행되고 있다. 대표적인 연구로서 김영석과 진광민(2006)은 경상북도 포항시 청하면 유계리에 위치한 유계단층의 트렌치를 통하여 얻은 최후기 단층운동에 의한 변위로부터 이 단층운동과 관련된 지진의 규모를 산정하고 있다. 김성균과 박민아(2002)는 1997년부터 2000년 사이에 발생한 143개 지진에 대한 총 822개의 디지털 지진기록에 대하여 한반도에서의 국지 지진 규모 공식을 경험적으로 유도하였고, 김성균 등(2001)은 한반도에서 발생 가능한 최대 지진빈도를 예측하기 위해서 확장된 최대우도법을 제안하고, 이 기법에 의해서 한반도에서 100년, 500년, 1000년 이내에 기대되는 최대 지진빈도가 각각 6.3, 6.8, 7.0이 될 것이라고 주장하였다. 그리고 홍태경 등(2000)은 진원거리의 소 구간 내의 자료에 대한 최소자승역산과 소구간 사이의 보간법을 이용하여 한반도 남부 지진의 지역 규모식을 제안하고 있다.

우리나라 국민들은 전 세계적으로 발생하고 있는 지진과 그 여파를 언론을 통하여 지진에 대한 피해의 심각성을 어느 정도는 인정하고 있지만, 아직까지는 우리나라는 안전하지 않는가라는 안일한 생각과 함께 지진의 피해를 절실히 인지하지 못하고 있는 실정이다. 이것은 최근 일본 지진의 여파로 인한 여진의 상황에서 전혀 상황대처를 하지 못했음을 볼 때 충분히 알 수 있다. 기상과 관련한 자연재해에 대한 대책은 어느 정도 마련되어 있으나 지진에 대한 대책은 전혀 없음을 여실히 보여 준 예라고 할 수 있다. 이러한 측면에서 우리나라 국민들의 지진에 대한 경각심을 불러일으키고, 지진 발생 시 지진에 대한 피해를 최소화시키기 위한 보다 체계적인 준비를 해야 할 필요가 있으며, 우리나라에서의 지진에 대해서 여러 각도에서 보다 많은 연구가 절실히 요구된다.

따라서 본 연구에서는 지난 과거 동안에 발생한 한반도의 지진들이 지역별로 어떤 규모로 발생되고 있으며, 어느 지역이 발생빈도가 높은지를 공간통계학적인 측면에서 알아보고자 한다. 이를 위하여 (1) 공간분석 도구인 SatScan을 이용하여 지진발생 지역의 공간 군집분석을 실시하고, (2) 지진규모별, 빈도별 자료를 가지고 공간검색통계량을 이용하여 공간 군집분석한 결과를 토대로 지진발생위험지도를 작성한다. 또한 (3) 공간 군집분석 결과와 지진발생위험지도를 토대로 하여 분석 결과에 대해서 해석하고, 본 연구의 결과를 차후 지진 연구의 기초자료로 활용하고자 한다.

본 연구의 2절에서는 공간 군집분석을 위한 알고리즘을 비롯하여 공간검색통계량과 모형에 대해서 소개한다. 그리고 공간 군집분석을 통하여 28년간(1978-2005) 발생한 한반도 지진자료를 이용한 실제 적용 사례를 3절에서 수행하고, 마지막 절에서 본 연구의 결론을 맺는다.

2. 공간검색통계량 및 모형

공간 통계학에서의 일반적인 관심사항은 일련의 점(point)들이 랜덤하게 분포되어 있느냐, 아니면 일정한 군집(cluster)성을 가지고 있느냐는 것이다. 이러한 군집의 위치를 찾아내고자 하는 경우에 사용할 수 있는 방법이 공간 검색통계량이다. 이러한 공간 검색 통계량은 지도위의 둥근 원(circular window) 혹은 둥근 원통(cylinder)으로 정의되며, 우도비(likelihood ratio)에 기초하여 둥근 원 혹은 원통을 관심지역의 모든 시공간 지점을 이동하며 검색한다. 이에 따라 통계적인 절차를 수행하고, 귀무가설 하에서의 결과와 비교하여 공간 군집분석을 수행한다. 공간 검색통계량를 이용한 공간 군집분석은 (1) 관측치가 이산형인 경우는 베르누이모형, 포아송모형 그리고 시-공간 조합 모형(space-time permutation model)을 가정하고, (2) 관측치가 연속형인 경우에는 지수모형와 정규모형를 가정하여 분석할 수 있다.

공간 검색통계량 분석에서 관측치가 이산형의 경우, 전체 지역을 G로 두고 그 부분집합 영역을 Z라 하자. 영역 Z의 내부에서 개체의 어떤 속성을 확률 p, 영역 Z 밖에서는 확률 q를 취하며 속성을 지닐 확률은 서로 독립이다. 이 때 가설은 다음과 같이 고려할 수 있다.

$$H_0 : p = q \quad v.s. \quad H_1 : p > q. \tag{1}$$

여기서 귀무가설이 기각이 되면 두 지역은 다른 군집으로 분류된다. 관측치가 이산형에 대한 베르누이모형, 포아송모형에 대해서는 문승호와 김종덕(2005)에서 자세히 소개되어 있다. 본 연구에서는 지진빈도에 대해서는 시-공간 조합 모형을 적용하고, 지진규모에 대한 분석은 정규 모형(normal model)을 고려하였다.

2.1 시-공간 조합 모형

시-공간 조합 모형에서의 가설은 식 (1)과 같고, 공간 위치정보와 시간 정보를 갖는 사례 수의 데이터를 필요로 한다. 시-공간 조합 모형의 알고리즘은 다음과 같다.

[단계 1] 검색 원(scanning window)으로 정의되는 중복 가능한 수많은 검색 원통(cylinders)을 활용하게 되는데 먼저 한정된 수의 지리적 격자점(위치)에서 시작하여 검색원 반경을 증가시키며 검색한다.

[단계 2] 검색 원통의 높이는 시간으로 설정하며 가장 최근의 시점은 반드시 포함하도록 한다.

[단계 3] 각 시-공간 위치에서 원통안의 관측도수와 기대도수를 계산하고 군집을 분류하는데 필요한 우도비는 계산하여 최대 우도비에 해당하는 원통이 군집이 될 가능성이 가장 높은 지역으로 분류한다.

시-공간 조합 모형에 의한 우도비 유도 과정은 다음과 같다(Kulldorff M. 외 4인, 2005). 지역번호 z와 시점 d에서의 관측 사례수를 c_{zd}라 하면 총 관측 사례 수는 다음과 같이 표현할 수 있다.

$$C = \sum_z \sum_d c_{zd}. \tag{2}$$

따라서 각 지역과 시점에 대해 기대도수는

$$\mu_{zd} = \frac{1}{C}(\sum_z c_{zd})(\sum_d c_{zd}) \tag{3}$$

이 된다. 그리고 특정 검색원통에서의 기대도수는 $\mu_A = \sum \mu_{zd}$이고, 검색 원통에서의 관측도수를 c_A라 하자. 여기서 시공간 연관성이 없다면 c_A는 초기하분포를 가정할 수 있다. 즉,

$$P(C_A) = \frac{\binom{\sum_z c_{zd}}{c_A}\binom{C - \sum_z c_{zd}}{\sum_d c_{zd} - c_A}}{\binom{C}{(\sum_d c_{zd})}}. \tag{4}$$

따라서 우도비(GLR : Generalized Likelihood Ratio)는 다음과 같이 정의된다.

$$GLR = (\frac{c_A}{\mu_A})^{c_A}(\frac{C - c_A}{C - \mu_A})^{(C - c_A)}. \tag{5}$$

검색을 통하여 후보 검색 원통에 대하여 우도비를 계산한 후 최대 우도비를 갖는 검색 원통에 대하여 군집을 구성한다. 참고로 우도비는 검정통계량으로 사용하게 되는데 몬테카를로에 기초한 모의실험 방법으로 유의확률 p-값을 계산한다.

2.2 정규 모형(normal model)

정규모형을 가정한 공간검색 통계량에서 지역 Z 내부의 개체들의 지진규모는 평균이 $\theta_{\in}$인 정규분포를 따른다고 하고, 반대로 지역 외부에 속하는 개체들의 지진규모는 평균이θ_{out}을 따르는 정규분포라고 가정하자.

정규분포를 가정할 때에는 다음의 가설을 고려할 수 있고, 검정 결과 귀무가설이 기각되면 지역 내의 개체들은 군집으로 형성된다.

$$H_0 : \theta_{\in} = \theta_{\text{out}} \quad v.s. \quad H_1 : \theta_{\in} > \theta_{\text{out}}. \tag{6}$$

이 가설에 대한 검정하기 위한 정규모형의 알고리즘은 다음과 같다.

[단계 1] 공간적으로 분포되어 있는 임의의 한 지역을 선정한다.
[단계 2] 단계 1의 지역으로부터 반경 r이 되는 원을 그린다.
[단계 3] 반경 r내의 평균값과 반경 r외의 평균값(본 연구에서는 지진의 규모)을 구한다.
[단계 4] 귀무가설과 대립가설 하에서 각각 우도함수를 구한다.

[단계 5] 단계 4에서 얻어진 두 우도함수에 의해서 우도비를 계산한다.
[단계 6] 단계 5에서 얻어진 최대우도비를 구하여 이를 검정통계량으로 사용하고, 몬테카를로(Monte Carlo)에 기초하여 p값을 계산한 후 검정을 실시한다.
[단계 7] 단계 6의 검정결과 귀무가설이 기각되면 반경 r내의 지역들은 군집으로 형성된다.
[단계 8] 반경 r을 0에서부터 공간거리 최대까지 움직여 가면서 단계 2부터 단계 7까지의 과정을 반복한다.

정규 모형에 의한 우도비 유도 과정은 다음과 같다. 단 여기서 지역 내부와 외부의 분산은 같다고 가정한다(즉, $\sigma^2_{\in} = \sigma^2_{\text{out}} = \sigma^2$). 먼저 i^{th}개체의 반응값(지진규모)을 X_i라 했을 때, 지역 Z 내부 개체들의 수를 $n_{\in}$, Z 외부 개체들의 수를 n_{out}, 그리고 전체 지역의 총 개체수를 $N = n_{\in} + n_{out}$이라 하자. 이때 임의의 지역 Z에 대한 우도(likelihood)는

$$L(Z, \theta_{\in}, \theta_{out}) = \prod_{i \in Z} (2\pi\sigma^2)^{-1/2} e^{-\frac{(x_i - \theta_{\in})}{2\sigma^2}} \times \prod_{I \ni nZ} (2\pi\sigma^2)^{-1/2} e^{-\frac{(x_i - \Theta_{out})}{2\sigma^2}} \quad (7)$$

이 된다. 그리고 우도비 검정통계량은

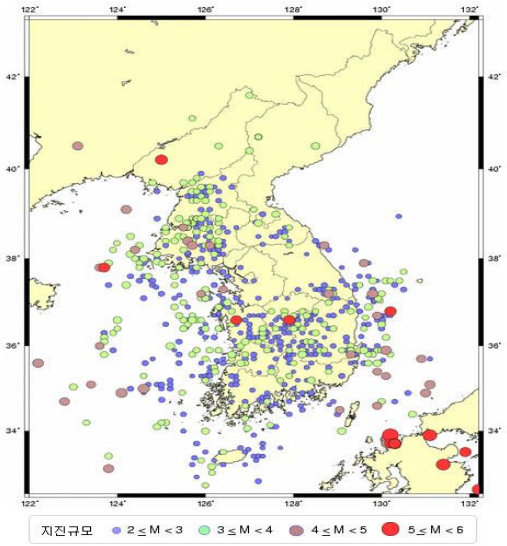
$$\lambda = \frac{\max_{Z,\ \theta_{\in} > \theta_{out}} L(Z, \theta_{\in}, \theta_{out})}{\max_{Z,\ \theta_{\in} = \theta_{out}} L(Z, \theta_{\in}, \theta_{out})} = \frac{L(\hat{Z})}{L_0} \quad (8)$$

와 같이 표현할 수 있다. 여기서 $\hat{Z}$는 대립가설 하에서 $L(Z, \theta_{\in}, \theta_{out})$을 최대화하는 지역을 의미하고, L_0는 귀무가설 하에서 $L(Z, \theta_{\in}, \theta_{out})$을 최대화하는 값이다.

3. 적용사례

3.1 분석자료 및 분석내용

본 연구의 분석에 사용된 자료는 한반도 지진 자료로서 기상청(http://www.kmaneis.go.kr)에서 제공하고 있는 663개 지진 데이터를 이용하였다. 이 데이터는 <표 1>과 <그림 1>에 주어져 있다.



<그림 1> 1978~2005 지진발생 분포도(자료출처 : 기상청)

<표 1> 원시데이터

연번	년도	월일	시분초	위도	경도	규모	지역
1	1978	08-30	2:29:43	39.1	124.2	4.5	평안남도
2	1978	09-16	2:07:06	36.6	127.9	5.2	충청북도
3	1978	10-07	18:19:52	36.6	126.7	5	충청남도
4	1978	11-23	11:06:05	38.4	125.6	4.6	황해도
5	1978	12-04	15:16:16	37.4	128.7	2.8	강원도
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
659	2005	06-15	7:07:02	33.15	126.14	3.7	제주도
660	2005	06-15	7:37:47	33	126.15	3	평양
661	2005	06-20	15:31:49	38.81	125.57	3	경상남도
662	2005	06-29	23:18:05	34.5	129.05	4	경상남도
663	2005	06-30	00:25:04	36.66	129.63	3.1	경상북도

본 연구에서는 1978년~2005년 사이에 한반도에서 발생한 663개의 지진자료를 이용하여 다음과 같은 두 종류의 공간 군집분석을 실시하였다.

(1) 한반도 지진빈도자료를 이용한 시공간적 지진 발생 지역의 군집화.

(2) 한반도 지진규모자료의 공간상관을 고려한 지진 발생 지역의 군집화.

이러한 두 종류의 공간자료에 대한 군집분석을 위하여 공간분석 툴인 SatScan 버전 7.0을 사용하였다(http://www.satscan.org).

3.2 분석 결과

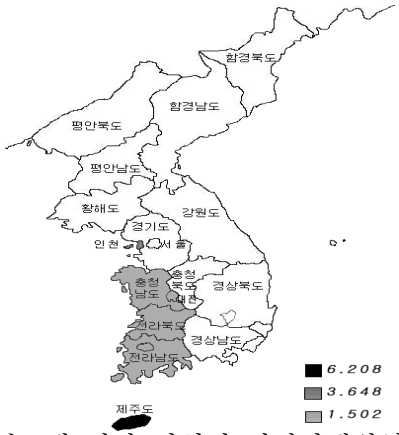
3.2.1 지진빈도에 의한 시공간 군집분석

지진빈도에 의한 시공간 군집분석은 시간과 공간을 동시에 고려하고, 한반도 지진빈도를 이용하여 실시하였다. 예를 들면, 한 시점과 한 지역에서 가능한 군집분석을 수행하는 방식으로 해서 모든 시점과 모든 지역에 대해서 시공간 군집분석을 수행하였다. 이에 대한 군집분석 결과는 <표 2>에 주어져 있다. <표 2>의 시공간 군집분석 결과를 보면 총 3개의 군집으로 구성되는데, 첫 번째 군은 제주도를 포함하고 있으며(p값=0.003), 인천이 두 번째 군으로 구성되고 있다(p값=0.003). 그리고 전라북도, 충청남도, 대전, 전라남도가 세 번째로 하나의 군집으로 형성되어 있다(p값=0.003).

<표 2> 지진빈도에 의한 시공간 군집분석 결과

구분	군집	case 수	time frame	지진발생위험도	p값
시공간 군집분석 (지진빈도)	제주도	16	2004. 1. 1-2005.12.31	6.208	0.003
	인천	14	2002. 1. 1-2005.12.31	3.648	0.003
	전라북도, 충청남도, 대전, 전라남도	82	2000. 1. 1-2005.12.31	1.502	0.003
	강원도, 경상북도	121	1996. 1. 1-2005.12.31	1.261	0.124
	평양	8	1995. 1. 1-2005.12.31	1.741	0.996

<표 2>에서 지진발생위험도는 관측빈도에서 기대발생빈도로 나눈 값으로 정의하였다. 이것은 군집으로 이루어진 지역들에서 기대되어진 빈도보다 더 많이 발생했다는 점에서 지진발생위험도가 높다면 그 군집의 지역들은 앞으로도 더 많이 발생할 수 있다고 해석할 수 있다. 이러한 측면에서 지진발생위험도가 높을수록 지진에 대한 위험도는 높다고 해석할 수 있다. 지진발생위험도의 결과 그림은 <그림 2>에 주어져 있다.



<그림 2> 지진빈도에 의한 지역별 지진발생위험도에 의한 그림

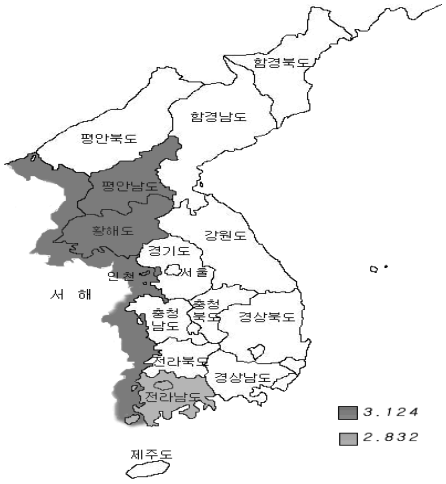
<그림 2>의 시공간 군집분석 결과의 그림에서 제주도가 가장 지진 발생위험 정도가 높고, 충청남도, 전라북도, 전라남도, 대전이 다음으로 지진 발생위험 정도가 높은 것으로 나타나고 있음을 알 수 있다.

3.2.2 지진규모에 의한 공간 군집분석

지진규모에 의한 공간 군집분석의 결과는 <표 3>에 주어져 있고, 지진발생위험도에 의한 결과의 그림은 <그림 3>에 주어져 있다. <표 3>을 보면 세 개의 군집으로 분류되었는데 마지막 군집은 p값이 0.996으로 유의하지 않은 것으로 나타났다. 첫 번째 군집으로는 대동강, 황해도, 평양, 평안남도, 인천, 서해로 구성되었다(p값=0.001). 그리고 두 번째 군집으로 전라남도가 하나의 군집으로 형성되었다(p값=0.003). <그림 3>으로부터 대동강, 황해도, 평양, 평안남도, 인천, 서해에서 지진발생이 높은 지역으로 나타나고 있음을 알 수 있다.

<표 3> 지진규모에 의한 공간 군집분석 결과

구분	군집	case 수	지진발생 위험도	p값
공간 군집분석 (지진규모)	대동강, 황해도, 평양, 평안남도, 인천, 서해	3390	3.124	0.001
	전라남도	1653	2.832	0.003
	부산, 울산	24		0.996



<그림 3> 지진규모에 의한 지역별 지진발생위험도에 의한 그림

4. 결론

한반도에서 발생된 연도별, 지역별 지진자료를 이용하여 지진빈도와 지진규모에 대해서 공간 군집분석을 실시하였다. 지진빈도에 의한 시공간 군집분석에서는 3개의 군집으로 형성하고 있으며, 제주도가 군집을 형성하는데 있어서 첫 번째의 군집으로 얻어져 가장 지진발생이 높은 지역으로 나타났다. 그리고 지진규모를 이용한 공간 군집분석에서 두드러지는 현상은 대동강, 황해도, 평양, 평안남도, 인천, 서해에서 지진발생이 높은 지역으로 나타나났다. 이러한 공간 군집분석 방법을 이용하여 군집화한 결과는 지진발생이 높은 지역이라고 판단할 수 있기 때문에 해당 지역에 대해서 지진에 대한 대비를 철저히 해야 할 것으로 판단된다. 지역별 지진빈도로 공간분석한 결과와 지진규모에 의한 분석 결과는 알고리즘 차이로 인하여 두 분석결과에서 지역에 있어서 다소 차이를 보이는 것은 당연하다고 할 수 있다. 이러한 경우에는 두 방법에서 동시에 포함된 지역을 지진에 위험한 군으로 판단해도 될 것으로 생각된다. 또한 본 연구에서 고려한 자료는 일본과는 달리 데이터의 75%가 규모 3.2 이하이기 때문에 분석된 결과가 실질적 위험군으로 해석하기보다 향후 지진에 대한 주의가 좀 더 필요한 지역으로 해석해야 할 것이다.

지진규모와 지진빈도를 이용하여 공간 군집분석한 결과를 보면 두 결과 모두 서해를 기준으로 하여 큰 규모 또는 높은 빈도의 지진이 발생한 것으로 나타났다. 이것은 서론에서 기술한 바와 같이 규모 7이상의 지진이 전 세계의 약 15%가 일본 지역에서 발생한다고 보았을 때, 동해 쪽의 지역에서 지진이 큰 규모로 많이 발생할 것이라는 생각과 반대의 결과를 보였다. 이러한 결과를 보이고 있는 것은 한반도는 판구조론적인 견지에서 볼 때 유라시아(Eurasia)판의 내부에 위치하며, 한반도의 지진활동은 일본과 같이 판 경계 지진활동이 아니고, 판 내부 지진활동의 범주에 속하기 때문이라고 생각된다.

참고문헌

1. 김성균, 박민아 (2002). 한반도의 국지 지진규모, *지질학회지*, 38권 2호, 217-235.
2. 김성균, 이정모, 김준경 (2001). 한반도에서 발생 가능한 최대지진, *지질학회지*, 37권 1호, 107-114.
3. 김영석, 진광민 (2006). 트렌치 단면의 단층변위 자료를 이용해 산정한 유계단층의 지진규모, *지질학회지*, 42권 1호, 79-94.
4. 홍태경, 박창업, 신진수 (2000). 한반도 남부 지진의 지역 규모식, *지질학회지*, 36권 4호, 545-558.
5. Sung-Ho Moon and Jong-Duk Kim (2006). Detection of Hotspots for Geospatial Lattice Data, *Journal of Korean Data and Information Science*, 17, 1, 131-139.
6. Kulldorff, M.(1997). A spatial scan statistics, *Communications in Statistics, Theory and Methods*, 26, 1481-1496.

7. Kulldorff, M.(2005). *SaTScan™ User Guide for version 7.0*.
8. Kulldorff, M., Heffernan R, Hartman J, Assuncao RM, Mostashari F. (2005). A space-time permutation scan statistic for the early detection of disease outbreaks. *PLoS Medicine*, 2, 216-224.
9. Myers, W. L. and Patil, G. P.(2002). Echelon analysis, *Encyclopedia of Environmetrics*, Vol. 2, 583-586.
10. Myers, W. L., Patil, G. P. and Joly, K.(1997). Echelon approach to areas of concern in synoptic regional monitoring, *Environmental and Ecological Statistics*, 4, 131-152.

[2006년 9월 접수, 2006년 11월 채택]