



K-평균 군집분석에 의한 한반도에서의 지진지체구조구 구분

김성균¹ · 전정수^{2,*} · 전명순²

¹전남대학교

²한국지질자원연구원

요 약

지진활동이 활발하지 않은 판내부 지역에서는 확률론적 지진재해도 분석에 사용되는 지진원의 설정이 용이하지 않다. 지진원 모델을 설정하고 평가하는 유일무이한 공식적인 방법은 존재하지 않는다. 이 연구에서는 개개의 지진을 점지진원으로 가정한 지진활동 자료를 K-평균 군집분석에 적용하여 한반도의 지진지체구조도를 작성하였다. KL 지수법과 Elbow법에 의해 결정된 군집수 K는 각각 5개와 9개로 나타났다. 5개의 구조구로 구성된 지진지체구조구 모델과 9개로 구성된 대체 모델이 작성되었다. 각각의 구조구에 대하여 추정된 지진활동 상수들이 제시되었다.

주요어: K-평균 군집분석, 지진지체구조구 구분, 지진활동 상수

Sung Kyun Kim, Jeong Soo Jeon and Myung-Soon Jun, 2017, Seismotectonic zoning by K-means clustering analysis in the Korean Peninsula. Journal of the Geological Society of Korea. v. 53, no. 5, p. 703-714

ABSTRACT: It is not easy to identify seismic source zone for use in probabilistic seismic hazard analysis in the intraplate region. There is no unique formal procedure for developing and evaluating seismic source models. The K-means cluster analysis is applied to seismicity data as a point source to delineate seismotectonic model for the Korean Peninsula in this study. The number of clusters K determined by KL index and Elbow methods appears to be five and nine, respectively. A seismotectonic model composed of five source zones is developed and an alternative model with nine zones is also proposed. Seismicity parameters estimated in each zone are presented.

Key words: K-means cluster analysis, seismotectonic zoning, seismicity parameters

(Sung Kyun Kim, Department of Earth Systems and Environmental Sciences, Chonnam National University, Gwangju 61186, Republic of Korea; Jeong Soo Jeon and Myung-Soon Jun, Earthquake Research Center, Korea Institute of Geoscience and Mineral Resources, Daejeon 34132, Republic of Korea)

1. 서 론

지진재해 경감을 위하여 널리 사용되고 있는 확률론적 지진재해도 분석(Probabilistic Seismic Hazard Analysis: PSHA)에서는 지진원의 설정이 필요하며, 각각의 지진원에 대한 지진활동 상수들을 정의하지 않으면 안 된다. 지진원에 대한 개념은 Cornell (1968)에 의해 도입되었으며, 지진원은 일반적으로 점지진원, 선지진원이나 다각형의 형태를 갖는 면적지진원으로 나눌 수 있다(Anderson and Trifunac, 1978). 점지진원은 보통 과거에 지진이 발생했던 지점에서 같은 양상으로 지진이 발생할 것으로 가정할 때에

정의되며, 선지진원은 활성단층과 같이 어떤 선상을 따라 지진이 발생하는 것을 가정할 때 사용된다. 보통 면적지진원은 지진원내의 모든 지점에서 지진활동 특성(지진발생 깊이, 지진발생 기작, 규모별 지진 발생률 및 최대지진 등)이 균일한 범위의 공간으로 규정할 수 있다.

미국 원자력 규제위원회(United States Nuclear Regulatory Commission: US NRC)는 원자력 발전소의 지진영향 평가를 위하여, 지진활동 특성과 지체구조적 및 지질학적 특성이 동일한 지진원을 지진지체구조구(seismo-tectonic zone or seimogenic source)라는 용어로서 정의하였다(US NRC, 1997, 2007a).

* Corresponding author: +82-42-868-3341, E-mail: jsjeon@kigam.re.kr

지진지체구조구는 규모별 지진발생빈도(Gutenberg-Richter 관계식의 a, b값), 지질학적 역사와 현재 및 과거의 응력장 환경 등에 근거하여 설정되어야 하며, 특히 중요한 것은 현재의 지구조 환경과 지진활동 양상이다(US NRC, 2007b). 따라서 이러한 방법으로 설정된 지진지체구조구는 인접 지역과 상이한 지진 발생 특성을 가지며, 지진을 발생시키는 것으로 확인된 활성단층과 같은 선구조(선지진원)는 지진지체구조구에 해당되지 않는다. 지진원 설정에 있어서 지진활동이 활발하지 않은 판내부 지역에서는 선지진원의 설정이 용이하지 않으며, 지진지체구조구를 설정하는 유일무이한 방법은 없다(Bender, 1986). 지진원의 설정은 주관성이 강하여 이 분야의 전문가들 사이에도 의견이 일치하기 어렵다. 그 이유는 동일한 자료에 대한 서로 다른 해석 또는 기존자료의 불충분이 주된 원인이며, 특히 장기간의 지진활동과 관련된 지체구조운동에 대한 지식의 부족이라고 할 수 있다(Gupta, 2007). 이러한 불확실성을 보완하기 위한 방법으로 보통 PSHA에서는 다수의 전문가가 참여하여 다양한 지진지체구조 모델을 제안하며, 이 모델들에 가중치를 부여하여 사용하게 된다.

국내에서도 지진위험지도 작성 또는 특정 부지의 지진안전성 평가(원전 예비안전성 분석보고서)를 위한 다수의 지진지체구조도 모델이 제시되었다(예를 들면, Korea Institute of Energy and Resources, 1983; Lee and Kim, 2000). 국내의 여러 전문가들에 의해 제안된 지진지체구조도는 다양하나, 지진 및 구조지질 관련 자료의 불충분으로 그 대부분은 지질구조도와 유사하다고 할 수 있다. Korea Institute of Nuclear Safety (2010)는 한반도와 한반도 주변의 지각두께, 지열류, 해저지형 및 Lg Q 분포자료를 이용하여 5개의 구조구로 구성되는 지진지체구조도 모델을 제안하였다. 여기서 진앙분포와 현생응력(최소 축압계수) 분포는 단지 참고자료로 사용되었다. 이 모델은 과거의 모델과 달리 다수의 정량화된 자료를 사용한 점이 특징이라 할 수 있으나, 지진활동과 관련된 자료들이 단지 참고로만 사용되었다는 점에서 진정한 의미의 지진지체구조 모델로 보기 어렵다. 한편 Geological Society of Korea (GSK, 2016)는 원전의 지진 안전성을 평가할 목적으로 한반도와 인접 지역의 지진활동 특성과 지구물리학적 자료 및 구조지질학적 자료(단층자료)를 참조하여 3개의 지진지

체구조 모델을 작성하였으며, 각 구조구별로 발생 가능한 최대지진을 산정하였다.

위에서 지적한 바와 같이, 균일한 지진활동도를 갖는 면적지진원을 설정하는 데에 누구나 인정할 수 있는 객관적이고 합리적인 기법은 존재하지 않는다. 지진원 설정을 위해서는 지진학적, 지구물리학적 및 지질학적 자료 등 다수의 자료가 병합되어야 하나, 이들 자료가 모든 지역에서 지진원을 분할할 수 있을 정도로 충분치 않으며, 이들 자료에 대한 해석이 전문가에 따라 다르다. 이러한 여러 이유로 분할방법도 주관적일 수 밖에 없다(Weatherill and Burton, 2009). 지진원 설정에 있어서 가장 중요한 자료는 현재의 지진활동을 반영하는 진앙분포도이며, 이 진앙분포도에서 개개의 진앙들을 하나의 점 자료로 간주하여 이들을 그룹화하는 군집분석을 이용하여 지진원을 설정하는 연구들(Burton *et al.*, 2008; Weatherill and Burton, 2009; Ansari *et al.*, 2015; Matinez-Alvarez *et al.*, 2015)이 시도되었다. 이러한 군집분석을 이용하는 기법은 단지 진앙분포도만을 이용하지만, 방법 자체는 통계학적으로 객관성이 충분하다고 할 수 있다. 본 연구에서는 K-평균 군집분석을 이용하여 한반도와 주변해역에 대한 지진지체구조도를 작성, 제시하고자 한다. 지진지체구조도가 작성되는 범위는 경도 124°~131°, 위도 33°~43.2° 범위의 사각형 지역이다. 또한 이러한 방법으로 설정된 각각의 지진지체구조구에 대한 지진활동 상수(Gutenberg-Richter 관계식의 b값, 규모별 지진발생률, 최대지진 규모)들을 추정하게 되며, 지진활동 상수들의 추정방법에 대한 충분한 검토 역시 이루어진다.

2. K-평균 군집분석 방법

먼저 군집분석(Cluster analysis)은 일반적으로 하나의 객체가 여러 속성을 갖고 있고, 이러한 객체가 다수 있다고 할 때 군집분석이란 유사한 속성들을 갖는 객체들을 묶어 전체의 객체들을 몇 개의 그룹 또는 군집으로 나누는 것을 말한다(Wikipedia, 2017). 예를 들어, 회사에서 관리하는 고객들에 대하여 구매행태를 반영하는 속성들에 대한 데이터가 수집된다고 할 때 유사한 구매행태를 보이는 고객들을 서로 그룹핑하는 것을 군집분석이라 할 수 있겠다. 따라서 이 예의 경우 군집분석의 목적 중 하나는 서로

다른 그룹의 고객들에게 서로 다른 마케팅 전략을 수립하는 것이 될 것이다. 즉 군집분석을 수리적으로 표현한다면 전체 n 개의 객체가 있다고 할 때 이들을 서로 배타적인 K 개의 부분집합으로 나누는 것이다.

군집방법은 무수히 많으나 크게 계층적 방법(hierarchical method)과 비계층적 방법(non-hierarchical method)으로 구분된다. 계층적 방법은 사전에 군집 수 K 를 정하지 않고 단계적으로 서로 다른 군집결과를 제공하는 것이다. 비계층적 방법은 사전에 군집 수 K 를 정한 후 각 객체를 K 개 중 하나의 군집에 배정하는 것이다. 이러한 배정은 대체로 한 번에 이루어지는 것은 아니고 배정 후 새로운 군집 중심값을 산출하여 각 객체에 대한 새로운 배정을 하는 과정을 군집결과가 수렴할 때까지 반복적으로 시행한다. 군집방법의 기본 아이디어는 한 군집내의 객체들 간의 유사성은 가능한 한 크게, 서로 다른 군집간의 유사

성은 가능한 한 작도록 군집들을 형성하는 것이다. 이를 위해서는 객체간의 유사성 및 군집간의 유사성의 정의가 필요하다. K -평균 군집분석은 비계층적 군집분석중 하나로서 널리 사용되고 있다. K -평균 군집분석에서의 유사성의 척도로서는 유클리드 제곱거리가 가장 널리 이용된다. n -차원의 벡터 $A(a_1, a_2, \dots, a_n)$ 와 $B(b_1, b_2, \dots, b_n)$ 의 유클리드 제곱거리(Euclidean square distance)는 다음 식으로 나타낼 수 있다.

$$D_{AB} = (a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2 \quad (1)$$

이 연구에서는 Weatherill and Burton (2009)의 방법에 따라 한반도와 주변 지역의 진앙분포에 K -평균 군집분석을 적용하였다. 그림 1에 분석과정을 도시하였다. 먼저 몇 개의 그룹(면적지진원)으로 객체(진원)들을 나눌 것인가, 즉 그룹 수 K 를 입력한다. 그룹 수의 결정 방법에 대해서는 뒤에 설명하기로 한다. 다음으로 각 그룹 내에서 좌표 x_i 인 N 개의 객체에 대한 중심 \bar{X} 를 다음 식으로부터 결정한다.

$$\bar{X} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} \quad (2)$$

여기서 w_i 는 가중치로서 Weatherill and Burton (2009)에서와 같이 Wells and Coppersmith (1994)의 단층 파쇄길이(L)과 모멘트 규모(M_W) 사이의 관계식으로부터

$$\log L = -3.22 + 0.64 M_W \quad (3)$$

주어졌다. 즉 개개의 지진규모에 대하여 계산된 L 을 w_i 로 대치하였다. 단층길이가 아닌 다른 양(예를 들면 단층 면적 또는 지진에너지)으로 가중치를 부여할 수도 있다. 여기서 단층길이를 가중치를 준 것에 대한 정당성은 없다. 일반적으로 K -평균 군집분석에 있어서 가중치는 부여하지 않으며, 모든 점은 같은 중요도를 가지게 된다. 그러나 큰 지진에 큰 가중치를 부여함으로써 그룹의 중심은 큰 지진쪽으로 끌리게 되며, 결과적으로 그룹내에 많은 지진을 포함하

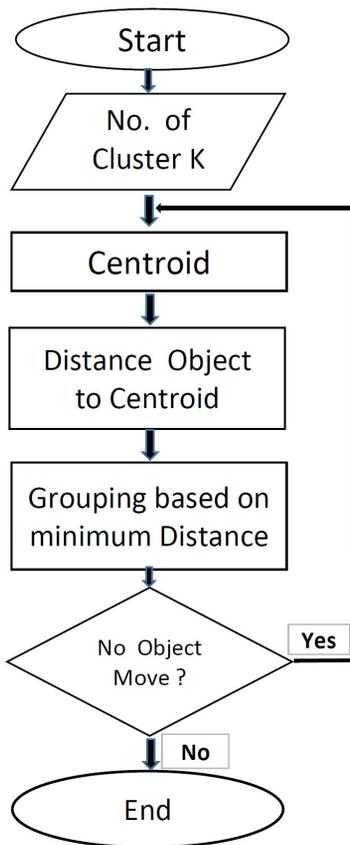


Fig. 1. A schematic flow diagram to explain K-means clustering.

게 된다(Weatherill and Burton, 2009).

다음 단계로서는 각 그룹의 중심으로부터 개개의 지진까지의 유클리드 제곱거리를 구한 후, 가장 가까운 그룹에 배정한다. 위 과정을 반복하여, 각각의 진앙들이 속한 그룹에 변화가 없으면 계산을 끝낸다. 위 과정 중 군집수 K와 초기의 중심위치 부여방법에 대해서는 다음 항에서 설명하기로 한다.

3. 지진지체구조도 설정

군집분석을 이용하여 지진지체구조도를 작성하기 위해서는 지진의 진앙 및 규모 등의 지진요소에 대한 불확실성이 적고 어느 정도의 완전성이 확보된 일정규모 이상의 관측된 지진목록이 필요하다. GSK (2016)는 역사지진과 계기지진 목록의 완전성과 불확실성에 대해 검토하였다. 그 결과에 따르면, 조선 왕조 이후의 역사 지진목록은 그 이전에 비하여 상대적으로 완전성 및 불확실성의 정도가 크게 개선된 편이나, 진앙지 오차(특히 해안지역), 진도평가 및 진도-규모 전환에 있어서의 오차는 상당히 클 것으로 결론지었다. 지진지체구조도를 작성함에 있어서는 가급적 긴 기간의 목록이 필요함은 자명하다. 그러나 역사지진목록은 불확실성 및 완전성에서의 문제점 외에도 진앙지에 있어서 근본적인 문제가 있다고 할 수 있다. 역사지진의 진앙지 분포와 계기지진의 분포를 비교해 보면, 육상에서의 분포는 어느 정도 유사하나, 해안 및 해양에서의 분포는 큰 차이를 보여준다. 역사지진 기록의 한계는 해양에서 발생한 지진에 대하여 그 진앙지를 해양이 아닌 육상으로 기술할 수밖에 없으며, 육상의 경우도 대부분 인구가 밀집된 지역이 진앙지로 기술된다는 점이다(Kim and Lee, 2017). 이 연구에서의 지진지체구조도의 구분에 이용되는 K-평균 군집분석은 각 객체(지진)의 공간적 위치에 민감하며, 구조구 구분은 해양까지도 포함되나 역사지진 자료는 육상자료에만 국한 된다. 따라서 역사지진 자료는 그 기간이 길어 지진활동을 잘 반영한다는 큰 장점에도 불구하고 지진지체구조도 작성에서 배제되었다.

다음으로 계기지진 목록에 포함된 규모의 불확실성에 대하여 고찰해 보기로 하자. 지진목록의 불확실성은 지진규모 산정과정 중의 불확실성과 어떤 규모로부터 다른 규모로 전환할 때의 불확실성으로 나

누어 생각할 수 있다. GSK (2016)는 계기지진 목록을 여러 기관에서 발췌하여, 중복된 목록의 제거 및 통합, 서로 다른 규모의 모멘트 규모로의 통일, 여진 및 전진 목록을 제거한 새로운 목록으로 편집하였다. 또한 이 목록의 완전성과 불확실성에 대해서도 검토하였다. 시대별 완전성에 대하여 Stepp (1972)의 방법으로 검토한 결과에 의하면, 1905년부터 1956년까지는 완전성 규모가 5.0이고, 1968년, 1976년까지는 각각 4.5, 4.0, 또한 1976년부터 2000년까지는 완전성 규모가 3.5, 2000년 이후는 2.5가 됨을 알아내었다. 규모에 대한 오차는 규모 산정시의 오차에 통일된 규모로 전환시의 오차가 가미되며, 시기에 따라 0.2~0.6 정도가 된다고 지적하였다. 한편 계기지진의 진앙지 오차는 관측망이 충실해지고 관측계기의 정밀도가 높아지는 최근에 이르러 크게 향상되었다. 그러나 초기의 오차는 최소한 진앙지의 경도, 위도에서 0.1° 이상이 될 것으로 짐작된다. 왜냐하면 기상청의 지진목록에서 2005년 이전에는 진앙지를 경위도의 소숫점 첫째자리까지만 공표하였고, 그 이후 둘째자리까지로 향상하였기 때문이다. 이러한 오차에 지진기록으로부터 위상을 읽어 수식적으로 진앙을 결정할 때까지 개입되는 오차가 가산된다고 볼 수 있다.

이 연구에서는 가급적 긴 기간의 많은 수의 지진이 포함된 목록을 이용하기 위하여, GSK (2016)가 작성한 1913-2012년 사이의 지진목록을 선택하였다. 계기지진 자료는 1906년부터 존재하지만, 1913년 이전에는 목록의 지진 수가 1개에 불과하기 때문에 본격적인 목록이 시작되는 1913년부터의 목록을 채택하였다. 여기에 2013년부터 2016년까지의 기상청 지진목록중 규모 3.0이상의 지진 중 전진 및 여진을 제거한 목록을 추가하여 지진지체구조도 작성에 사용할 관측지진목록을 만들었다. 이 지진목록은 1913년부터 2016년까지 규모 3.0이상의 지진(전진 및 여진 제외) 781개로 구성되어 있다. 지진목록에서 최소규모(또는 한계규모)를 3.0으로 결정한 것은 GSK (2016)의 완전성 검토결과를 참조한 것이며, 지진지체구조도에 가급적 긴 기간, 많은 수의 지진활동이 반영되도록 하기 위함이다.

지진목록에 의한 진앙분포도에 K-평균 군집분석을 적용하기 위한 다음 단계는 군집수 K를 어떻게 결정하느냐 하는 문제이다. Charrad *et al.* (2014)은

군집수 K를 결정하기 위한 여러 방법들을 비교 분석하였다. 이 연구에서는 이 방법들 중 널리 사용되고 있는 KL 지수법(Krzanowski and Lai, 1988)과 Kodinariya and Makwana (2013)의 Elbow법으로 군집수 K를 결정하였다. KL 지수법은 다음 식으로 정의되는 KL 지수중 최대값을 갖는 K를 찾는 방법이다.

$$KL(K) = \left| \frac{DIFF(K)}{DIFF(K+1)} \right| \quad (4)$$

$$DIFF(K) = (K-1)^{2/d} SSW(K-1) - K^{2/d} SSW(K)$$

위 식에서 d 는 식 (1)에서의 차원을 나타내며, SSW 는 군집사이 거리의 제곱평균(within-cluster sum of square)로서 다음 식으로 정의된다.

$$SSW = \sum_{j=1}^K \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad (5)$$

위에서 \bar{x}_j 는 j 번째 그룹에 속한 객체들의 중심좌표를 나타내며, x_{ij} 는 j 번째 그룹내에서 i 번째 객체의 좌표를 나타낸다.

한편 Elbow법에서는 군집수 K를 가로축으로 목적함수를 세로축으로 하여 나타내었을 때, 목적함수의 변화가 급격히 줄어들기 시작하는 변곡점에서의 K를 최적 군집수로 선택하는 기하학적 방법이다. 이 연구에서는 목적함수를 SSW/SST 로 하여 군집수 K에 대하여 조사하였다. SST 는 객체 전체의 중심과

개개의 객체사이의 유클리드 거리제곱의 합으로서 다음 식으로 표현된다.

$$SST = \sum_{j=1}^K \sum_{i=1}^{n_j} (x_{ij} - \hat{x})^2 \quad (6)$$

\hat{x} : grand mean

KL 지수법으로 결정된 군집수 K는 5이며, Elbow법으로 결정된 수는 9이다. 그림 2에 Elbow법에 의해 목적함수 SSW/SST 를 군집수 K에 대하여 나타낸 결과를 도시한다. 그림에서 군집수 9를 경계로 목적함수 값이 거의 수평으로 바뀌는 것을 관찰할 수 있다. 그림 3과 그림 4에 군집수 5개(모델 A)와 9개(모델 B)에 의한 군집분석 결과에 따라 설정된 지진지체구조도를 도시하였다. 지진지체구조도에서 각 군집내의 지진은 다른 색깔로 표시되어 있어 쉽게 구별이 가능하다. 2개의 지진지체구조구 모델에서 비교적 단순한 5개의 구조구로 구성된 모델을 추천 모델로 하며, 9개의 구조구로 구성된 모델을 대체 모델로 한다.

4. 지진지체구조구별 지진활동 상수

PSHA를 위한 지진활동 상수에는 보통 지진의 규모별 빈도분포를 나타내는 Gutenberg- Richter의 관계식에서의 b 값, 최소규모와 최대규모 및 최소규모 이상 지진의 발생률 등이 있다. 이하에서 이들 지진활동 상수를 구하는 기법에 대하여 설명하기로 한다. 지진의 규모별 빈도분포에 관한 Gutenberg-Richter의 관계는 다음 식으로 표현된다.

$$\log N(N \geq M) = a - bM \quad (7)$$

여기서 N 은 지진규모 M 보다 크거나 같은 규모의 누적빈도를, a 와 b 는 상수를 나타낸다. 상수 a 는 대상 면적과 기간, 지진활동도에 따라 달라지며, b 는 지역적인 구조적 환경과 관련이 있는 것으로 알려져 있다(Scholz, 2015).

b 값은 통상 최소제곱법(Least Square Method) 또는 최대우도법(Maximum Likelihood Method)에 의해 구할 수 있다. 그러나 선형식의 계수추정에

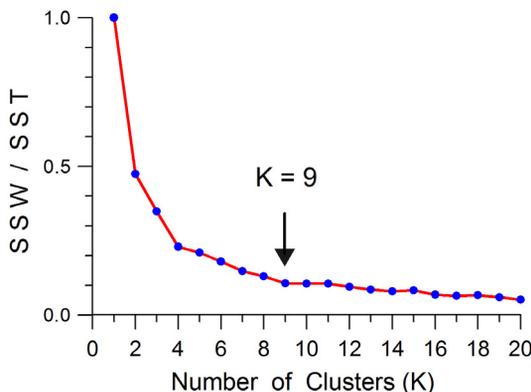


Fig. 2. SSW/SST versus number of clusters. SSW and SST denote the within cluster sum of square and the total sum of squares, respectively. The arrow indicates the selected number of clusters by the Elbow method.

널리 사용되고 있는 통상적인 최소제곱법은 그 가정 자체가 b 값 추정에 적절하지 않다(Weichert, 1980). 따라서 최소제곱법보다는 최대우도법이 좀 더 실제에 가까운 b 값을 추정할 수 있는 것으로 알려져 있다. 최대우도법에 근거한 b 값을 추정하기 위한 여러 방법들(Utsu, 1965; Page, 1968; Weichert, 1980)이 제안되어 있다. 추정된 b 값은 사용한 방법에 따라 그 값에 차이가 크게 발생하며, 이러한 통계학적인 방법론 외에도 지진목록에서의 자료의 수나 여진의 제거 여부 등이 b 값에 크게 영향을 주는 것으로 알려져 있다(Bender, 1983; Amorese *et al.*, 2009). GSK (2016)는 모의실험을 통하여 Utsu (1965)의 방법이 다른 방법

에 비하여 좀 더 안정적인 b 값을 주게 됨을 입증하였다. 따라서 이 연구에서는 이 방법을 채택하였다.

Utsu (1965)에 의해 제안된 b 값의 추정치 b^* 는 다음 식으로부터 얻을 수 있다.

$$b^* = \frac{1}{\ln 10 (\bar{m} - m_c)} \quad (8)$$

여기서 \bar{m} 은 지진규모의 평균치를, m_c 는 최소지진 즉 지진목록에서 완전성이 보장되는 절단규모를 의미한다. 또한 b 값의 표준오차는 다음 식으로부터 구할 수 있다(Shi and Bolt, 1982).

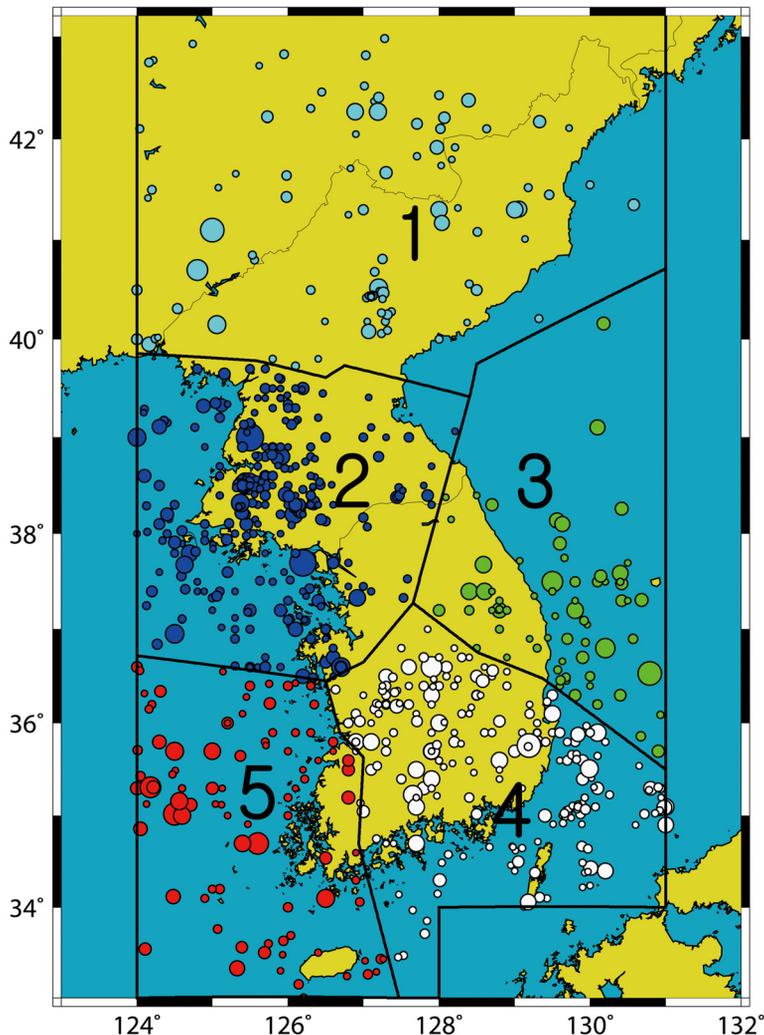


Fig. 3. The seismotectonic model composed of five source zones.

$$\hat{\sigma}_b = 2.30(b^*)^2 \sqrt{\frac{\sum_{i=1}^N (m_i - \bar{m})^2}{N(N-1)}} \quad (9)$$

위에 기술한 방법으로 b값이 구해지면, 식 (7)로부터 일정규모 이상 지진의 발생확률을 구할 수 있다.

최대지진 규모(M_{max})는 어떤 지진구역내에서 발생가능한 최대 규모의 지진을 의미하며, 그 이상을 초과하는 지진은 발생하지 않는다(Kijko, 2009). 또한 그 이상의 규모에서는 Gutenberg-Richter 관계 곡선에서 지진규모에 대한 누적빈도가 더 빨리 감소한다. 최대지진규모에 대한 추정은 주요구조물의 내

진설계 등의 여러 공학적인 응용에 필요하지만, 이에 대한 연구는 크게 진전되어 있지는 않다.

최대지진 규모의 추정법은 크게 결정론적인 방법과 확률론적인 방법의 두 가지로 나눌 수 있다. 결정론적인 방법은 지진규모와 지체구조 또는 단층관련 상수들 사이의 경험적 관계로부터 M_{max} 를 추정하는 방법이다(예를 들면, Wells and Coppersmith, 1994). 또는 지진모멘트나 지각변형률- M_{max} 관계 또는 coda Q_0 - M_{max} 관계(Jin and Aki, 1988)로부터 최대지진 규모를 추정하게 된다. 그러나 이러한 방법으로 추정된 M_{max} 의 불확실성은 상당히 크다(Kijko, 2009). 현재까지 조사, 연구된 한반도의 제 4기 단층

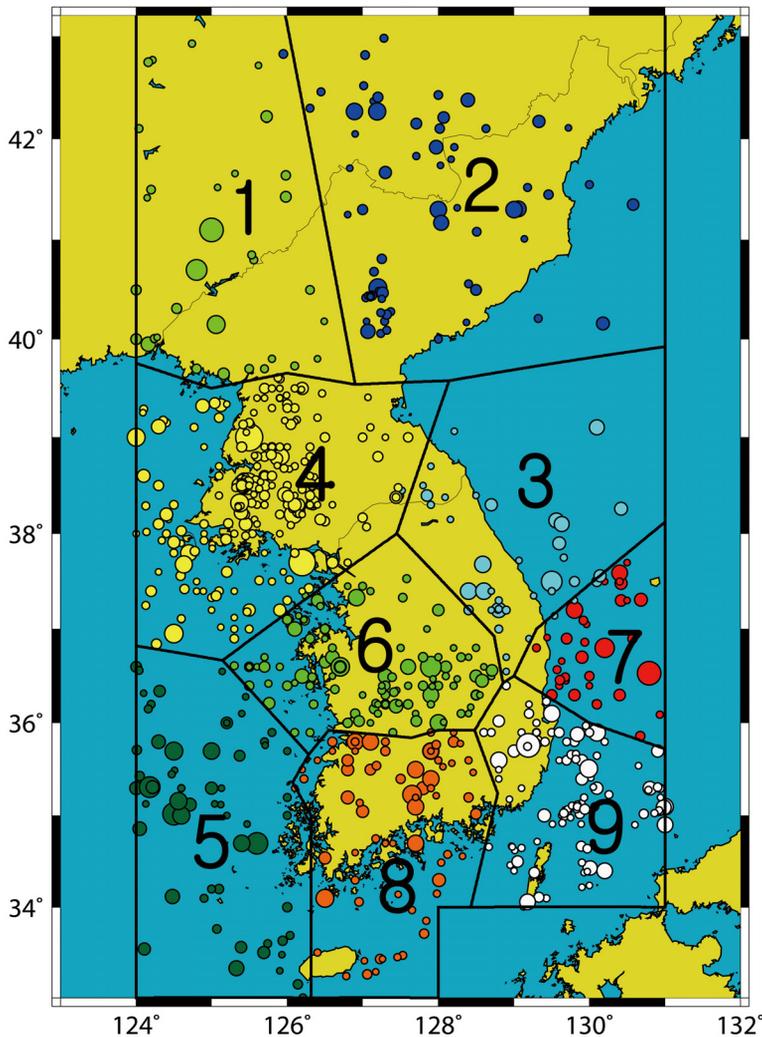


Fig. 4. The alternative seismotectonic model composed of nine source zones.

에 대한 자료는 결정론적인 방법에 의하여 M_{max} 를 추정하기에 충분하지 않다. 따라서 이 연구에서는 확률론적인 방법만으로 M_{max} 를 추정하였다.

확률론적인 방법은 어떤 지역내에서 발생한 지진의 목록을 통계학적인 방법으로 분석하여 M_{max} 를 추정하는 것이다. 통계학적인 방법으로 발생 가능한 최대지진의 규모를 추정하기 위한 방법론에 대하여 Kijko (2009)는 자신의 연구를 포함하여 다른 연구자들의 연구들을 폭 넓게 조사하여 최대지진을 추정하는 방법론을 상세히 검토하였다. Kijko (2009)에 의하면 최대지진 규모 M_{max} 를 추정하는 식은 지진활동 상수만을 입력자료로 하는 방법과 지진목록 전체를 필요로 하는 방법으로 나눌 수 있다. 결과적으로 말하면, 지진목록을 이용하는 방법도 자체적으로 지진활동 상수를 구하여 수식에 적용하게 된다. 지진활동 상수를 입력자료로 하는 M_{max} 의 추정법은 다음과 같은 간단한 식으로 나타낼 수 있다.

$$M_{max} = M_{max}(obs) + \Delta M \quad (10)$$

위에서 $M_{max}(obs)$ 는 관측된 최대지진을 나타내며, ΔM 은 b 값, 완전성이 보장되는 최소규모 이상 지진의 연간발생률과 관측기간의 함수로서 M_{max} 를 구하는 Kijko (2009)의 대표적인 방법은 다음과 같다.

- Tate - Pisarenko (방법 1)
- Kijko - Sellevoll (방법 2)
- Tate - Pisarenko - Bayesian (방법 3)
- Kijko - Sellevoll - Bayesian (방법 4)

위의 4가지 방법들에 의한 M_{max} 는 자료수 (=관측기간×최소규모 이상 지진의 연간 발생률)가 많은 경우에는 그 결과들이 거의 비슷하지만 자료수가 적은 경우에는 그 편차가 크다. 이 연구에서는 다음과 같은 수치 모의실험을 통하여 위의 4가지 방법들을 비교하였다. 최소규모와 최대규모를 각각 m_{min} , m_{max} 라 할 때, 규모 m 에서의 누적분포확률 $F_M(m)$ 은 다음 식으로 나타낼 수 있다(Page, 1968).

$$F_M(m) = \frac{1 - 10^{-b(m - m_{min})}}{1 - 10^{-b(m_{max} - m_{min})}} \quad (11)$$

이 연구에서는 최소규모와 최대규모를 각각 4.0, 8.0으로 하고, b 값을 0.85로 한 누적확률분포(식 11)에 따라 무작위수를 이용하여 규모를 결정한다. 이러한 방법으로 자료수가 50, 100, 200, 500, 1000개로 구성된 지진목록을 만든다. 다음으로 이 목록에서 관측된 최대지진 규모($M_{max}(obs)$)를 조사한다. 이렇게 얻어진 $M_{max}(obs)$ 와 최소규모, b 값 및 자료수를 입력하여 Kijko (2009)의 4가지 방법으로 M_{max} 를 구한다. 이 연구에서 M_{max} 의 추정에는 Matlab으로 작성된 Kijko (2010)의 컴퓨터 프로그램을 사용하였다.

이와 같은 과정을 1000회 반복하여 4가지 방법에 의해 추정된 M_{max} 와 표준편차의 평균치를 계산하여 그림 5에 도시하였다. 그림에서 방법 1과 3에 의해 추정된 M_{max} 는 미리 정한 8.0에 가까우나, 특히 지진목록의 자료수가 적은 경우에 표준편차가 지나치게 크다는 것을 알 수 있다. 자료수 50개의 경우에 그 표준편차는 3.0 정도로 나타난다. 이에 비하여 방법 2와 4에 의해 추정된 M_{max} 는 자료수가 적을 때 8.0에 비해 지나치게 작으나, 표준편차는 방법 1과 3에 비해 작다는 것을 알 수 있다. 따라서 추정되는 M_{max} 와 표준편차를 적절히 조화시키기 위해 이 연구에서는 방법 3과 4에 의해 얻어진 결과들의 평균치를 사용하였다.

이상에서 설명한 b 값과 M_{max} 의 추정법을 이용하여 지진지체구조구 모델 A의 구조구별 지진활동 상수들을 표 1에 제시하였다. M_{max} 는 역사지진과 계기지진에 대하여 각각 추정되었다. 관측된 최대규모 $M_{max}(obs)$ 의 오차는 역사지진과 계기지진에 대하여 각각 0.5와 0.3으로 가정하였다. 또한 역사지진의 M_{max} 추정에는 계기지진에서 얻어진 b 값을 사용하였다. b 값은 0.78에서 1.0까지의 범위로 나타났으며, Zone 2와 4에서 큰 값을 보여준다. 추정된 M_{max} 는 대략 5.9에서 7.0의 범위에 있는 것으로 나타났으며, Zone 2에서 가장 큰 값을 보여준다. Zone 1과 5에서는 기록된 역사지진의 수가 적어 M_{max} 추정을 하지 않았다.

5. 토론 및 결론

그림 1에서 초기에 군집수 K가 입력되었을 때, 각 그룹의 중심좌표를 어떻게 부여할 것인가 하는 문제점이 있다. 초기의 중심좌표는 보통 무작위적으로 부여한다(Singh and Kaur, 2013). 그러나 초기값을 어

평균 주느냐가 최종결과에 영향을 미치게 된다고 알려져 있다(Kodinariya and Makwana, 2014; Singh and Kaur, 2013). 이 연구에서는 각 그룹 중심좌표의 초기값을 무작위 수를 이용하여 부여하지만, 중심좌표 사이의 간격 I 가 다음 조건을 만족하도록 조정하였다.

$$I \geq d/\sqrt{K} \tag{12}$$

여기서 d 는 지진지체구조도를 작성하는 사각형 지역에서 세로방향의 거리이다. 이 방법은 각 그룹의 중심들이 가급적 멀리 떨어지도록 하기 위한 하나의 방안이다. 이 연구에서는 초기값을 아무 조건

없이 부여한 경우와 식 (12)와 같은 조건을 부여한 경우를 비교하였다. 그 결과 최종결과에는 큰 차이가 없음이 밝혀졌으며, 아무 조건 없이 초기값을 부여한 경우는 수렴이 늦어져, 최종결과가 얻어질 때까지 단지 반복횟수가 증가할 뿐이었다.

Houng and Hong (2013)는 Lee and Kim (2002)의 진도-거리 감쇠식과 Kim (2012)의 진도 - 지진규모 M 사이의 전환식을 결합하여 다음의 식을 제안하였다.

$$M = 0.58I + 1.02\ln(0.59I + 2.87) - 0.86 \tag{13}$$

여기서 I, l 은 각각 관측된 진도 및 진앙거리를 나

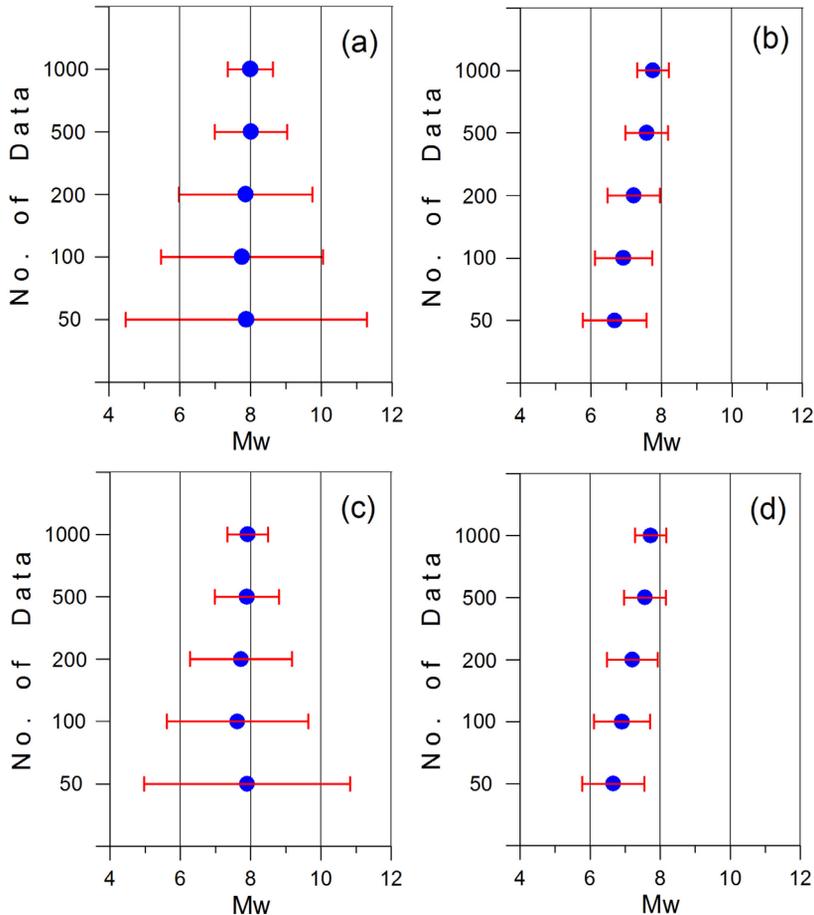


Fig. 5. Comparison of the methods to estimate the maximum earthquake magnitude. (a), (b), (c), and (d) denote results by the method for Tate-Pisarenko, Kijko-Sellevoll, Tate-Pisarenko-Bayesian, and Kijko-Sellevoll-Bayesian, respectively. The closed circle and horizontal bar represent the mean and standard deviation in each diagram, respectively.

Table 1. Seismicity parameters for the seismotectonic model A delineated by the K-means cluster analysis. SD and λ_5 denote the standard deviation and the mean occurrence probability per year for the earthquake of which magnitude is equal to or greater than Mw=5.0, respectively. $M_{max}(obs)^*$ and $M_{max}(obs)^{**}$ represent the observed maximum earthquake magnitude for instrumental and historical earthquakes, respectively. $M_{max}(\pm SD)^*$ and $M_{max}(\pm SD)^{**}$ represent the maximum probable earthquake magnitude estimated from instrumental and historical data, respectively.

Parameters	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
$b(\pm SD)$	0.78±0.19	1.00±0.17	0.87±0.21	0.98±0.13	0.83±0.13
λ_5	0.0360	0.0526	0.0362	0.0568	0.0456
$M_{max}(obs)^*$	5.8	5.0	5.4	6.2	5.6
$M_{max}(\pm SD)^*$	6.27±0.56	7.03±0.89	6.28±0.57	6.15±0.47	5.87±0.41
$M_{max}(obs)^{**}$	-	6.2	6.2	5.7	-
$M_{max}(\pm SD)^{**}$	-	6.45±0.56	6.49±0.58	6.43±0.55	-

타내며, 지진이 관측 또는 보고된 지점으로부터 진앙까지의 거리 l 과 관측된 진도 I 로부터 규모를 구할 수 있다. Houg and Hong (2013)에 의하면, 조선시대 인구가 밀집된 주요 도시간의 평균거리는 43.8 km이며, 이 시대의 역사지진 자료에서 관측된 최대진도는 반드시 진앙의 진도를 나타내는 것은 아니라고 지적하였다. 즉 실제의 진앙에서 최대 43.8 km 떨어진 곳에서 관측된 진도로 볼 수 있다는 것이다. 그 결과 실제의 진앙지 진도는 관측된 진도보다 최대 2.47만큼 증가함을 지적하였다. 그러나 지진이 두 도시의 중간에서 일어난 경우를 진앙거리의 최대값으로 보는 것이 더욱 합리적일 것이다. 이 경우 진앙에서 지진이 관측 또는 보고된 지점까지의 최대거리는 43.8 km의 1/2인 21.9 km가 되며, 이에 따른 진도의 증가는 1.42 정도가 된다. 진도 1.42의 증가는 Kim (2012)의 진도-규모 전환식에 의하면 규모에 0.82만큼의 증가를 가져온다. 따라서 지진지체구조 모델 A의 각 구조구에서 역사지진에 의해 추정된 최대지진 값에 0.82만큼씩을 더한 값을 발생가능한 최대지진 값으로 볼 수도 있다.

이 연구에서는 K-평균 군집분석을 한반도의 진앙 분포도에 적용하여 지진지체구조도 작성을 시도하였다. KL 지수법과 Elbow 법에 의해 결정된 그룹의 수는 각각 5개와 9개로 나타났으며, 5개의 구조구로 구성된 지진지체구조 모델 A와 9개로 구성된 모델 B를 작성하였다. 각각의 구조구에 대한 지진활동상수 b 값은 Utsu (1965)의 방법으로 추정하였다. 그 값은 0.78에서 1.0의 범위로 나타났으며, Zone 2와 4에서 비교적 큰 값을 보여준다. 한편 각 구조구

에서의 최대지진 값을 추정하기 위하여 수치 모의실험을 통하여 Kijko (2009)가 제안한 방법 중 가장 적절한 방법을 찾았다. 결과적으로 Tate - Pisarenko - Bayesian법과 Kijko - Sellevoll - Bayesian 방법으로 추정된 값의 평균을 취하는 방법을 선택하였다. 이 방법으로 지진지체구조 모델 A의 각 구조구에 대해서 역사지진과 계기지진 목록으로부터 최대지진 값을 추정하였다. 추정된 최대지진 값은 대략 5.9에서 7.0의 범위에 있는 것으로 나타났으며, Zone 2에서 가장 큰 값을 보여준다.

이 연구에서는 K-평균 군집분석에 의해 지진지체구조구 구분을 시도하였다. 기존의 지진지체구조구 구분은 대부분 주요 지질구조선 또는 주요 단층선(선구조선)을 경계로 하여 이루어졌다. 여기에 비하여 이 연구에서의 구조구 구분은 단순히 지진분포에 따라 이루어졌으므로 단층이나 지질경계와는 무관하며, 큰 지진을 핵으로 하여 나누어져 있다고 할 수 있다. 정량적인 방법으로 기존의 지진지체구조구 모델과의 비교를 위해서는 특정의 부지를 상정하여 기존의 모델과 이 연구에서의 모델에 의한 최대지진가속도 값 등에 대한 비교가 수행되어야 할 것으로 생각된다. 이 연구에서는 지진지체구조구 구분에 지진자료 중 계기지진의 진앙분포만을 사용하였으나, 다른 지진학적, 지구물리학적 및 구조지질학적 다양한 자료를 정량적으로 결합시키는 방법이 요구된다.

사 사

이 연구는 한국지질자원연구원 2017년 주요사업

의 하나인 “HLW 지층처분 후보부지 선정을 위한 전국규모 지질환경정보도 작성(17-3424)” 사업의 지원을 받아 수행하였습니다.

REFERENCES

- Amorese, D., Grasso, J.R. and Rydelek, P.A., 2009, On varying b-values with depth: Result from computer-intensive tests for southern California. *Geophysical Journal International*, 180, 347-360.
- Anderson, J.G. and Trifunac, M.D., 1978, Uniform risk functionals for characterization of strong earthquake ground motion, *Bulletin of the Seismological Society of America*, 68, 205-218.
- Ansari, A., Firuzi, E. and Etemadseed, L., 2015, Delineation of seismic sources in probabilistic seismic-hazard analysis using fuzzy cluster analysis and Monte Carlo simulation. *Bulletin of Seismological Society of America*, 105, 2174-2191.
- Bender, B., 1983, Maximum likelihood estimation of b values for magnitude grouped data. *Bulletin of the Seismological Society of America*, 73, 831-851.
- Bender, B., 1986, Modelling source zone boundary uncertainty in seismic hazard analysis. *Bulletin of Seismological Society of America*, 76, 329-341.
- Burton, P.W., Weatherill, G., Karnawati, D. and Pramumijoyo, S., 2008, Seismic hazard assessment and zoning in Java: New and alternative probabilistic assessment models. *International conference on Earthquake Engineering and Disaster Mitigation (ICEEDM08)*, Jakarta, Indonesia, 1-8.
- Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A., 2014, NbClust: An R determining the number of clusters in a data set. *Journal of Statistical Software*, 61 (Issue 6), 1-35.
- Cornell, C.A., 1968, Engineering seismic risk analysis. *Bulletin of Seismological Society of America*, 58, 1583-1606.
- Geological Society of Korea, 2016, Investigation and research of maximum earthquake for nuclear power plant sites (I). Technical Report to Korea Hydro & Nuclear Power Co., Seoul, Korea, 427 p (in Korean).
- Gupta, I.D., 2007, Probabilistic seismic hazard analysis method for mapping of spectral amplitudes and other design-specific quantities to estimate the earthquake effects on man-made structures. *ISSET Journal of Earthquake Technology*, 44, 127-167.
- Houng, S.E. and Hong, T.K., 2013, Probabilistic Analysis of the Korean Historical Earthquake Records. *Bulletin of the Seismological Society of America*, 103, 2782-2796.
- Jin, A. and Aki, K., 1988, Spatial and temporal correlation between coda Q and seismicity in China. *Bulletin of Seismological Society of America*, 78, 741-769.
- Kijko, A., 2009, Different techniques for estimation of the maximum earthquake magnitude Mmax, File: Mmax_Rev 20 May 2009. Aon-Benfield Natural Hazard Centre Pretoria University, Pretoria 0002, South Africa.
- Kijko, A., 2010, Computer code to estimate maximum earthquake magnitude.
- Kim, S.K., 2012, Korean historical and instrumental earthquake catalogues. in *International Workshop on KHNP SSHAC Project*, Gyeongju, South Korea (October 11-13, 2012).
- Kim, S.K. and Lee, J.M., 2017, Probabilistic seismic hazard analysis using a synthetic earthquake catalog: Comparison of the Gyeongju City Hall site with the Seoul City Hall site in Korea. *Geosciences Journal*, 21, 523-533.
- Kodinariya, T.M. and Makwana, P.R., 2013, Review on determining number of cluster in K-means clustering. *International Journal of Advance Research in Computer Science and management studies*, 1, 90-95.
- Korea Institute of Energy and Resources, 1983, Seismic risk map of Korea, 295 p (in Korean).
- Korea Institute of Nuclear Safety, 2010, Technological development for geological seismic safety evaluation basis of NPP sites, 68 p (in Korean).
- Kodinariya, T.M. and Makwana, P.R., 2013, Review on determining number of cluster in K-means clustering. *International Journal*, 1(6), p. 90-95.
- Kraznowski, W.J. and Lai, Y.T., 1988, A criterion for determining the Number of Groups in a data set using sum-of-squares clustering. *International Biometric Society*, 44(1), pp. 23-34.
- Lee, K. and Kim, J.-K., 2000, Seismic characteristics of tectonic province of the Korean Peninsula. *Journal of Korean Geophysical Society*, 3, 91-98 (in Korean with English abstract).
- Lee, K. and Kim, J.-K., 2002, Intensity attenuation in the Sino-Korean craton. *Bulletin of the Seismological Society of America*, 92, 783-793.
- Page, R., 1968, Aftershock and microaftershocks of the Great Alaska earthquake of 1964. *Bulletin of Seismological Society of America*, 58, 1131-1168.
- Martinez-Alvarez, F., Gutierrez-Aviles, D., Morales-Esteban, A., Reys, J., Amaro-Mellado, J.L. and Rubio-Escudero, C., 2015, A novel method for seismogenic zoning based on triclustering: Application to the Iberian Peninsula. *Entropy*, 17, 5000-5021.
- Scholz, C.H., 2015, On the stress dependence of the earthquake b value. *Geophysical Research Letters*, 42, 1399-1402.
- Shi, Y. and Bolt, B.A., 1982, The standard error of the magnitude-frequency b value. *Bulletin of Seismological*

- Society of America, 72, 1677-1687.
- Singh, H. and Kaur, K., 2013, New method for finding initial cluster centroids in K-means algorithm. *International Journal of Computer Applications*, 74, 27-30.
- Stepp, J.C., 1972, Analysis of completeness of the earthquake sample in the Puget Sound area and its effect on statistical estimates of earthquake hazard. *Proceedings of First Microzonation Conference*, Seattle, WA, U.S.A., 897-909.
- United States Nuclear Regulatory Commission, 1997, Identification and characterization of seismic sources and determination of safe shutdown earthquake ground motion. *Regulatory Guide 1.165*.
- United States Nuclear Regulatory Commission, 2007a, A performance-based approach to define the site-specific earthquake ground motion. *Regulatory Guide 1.208*.
- United States Nuclear Regulatory Commission, 2007b, Standard review plan (Section 2.5.2 Vibratory ground motion) in NUREG-0800.
- Utsu, T., 1965, A method for determining the value of b in a formula $\log n=a-bM$ showing the magnitude-frequency relation for earthquakes. *Geophysical Bulletin of Hokkaido University*, 13, 99-103.
- Weatherill, G. and Burton, P.W., 2009, Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region. *Geophysical Journal International*, 176, 565-588.
- Weichert, D.H., 1980, Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of Seismological Society of America*, 70, 1337-1346.
- Wells, D.L. and Coppersmith, K.J., 1994, New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of Seismological Society of America*, 84, 974-1002.
- Wikipedia, 2017, https://en.wikipedia.org/wiki/K-means_clustering (June 15, 2017).

Received : July 3, 2017

Revised : September 6, 2017

Accepted : September 6, 2017