Journal of the Faculty of Environmental Science and Technology.
Okayama University
Vol.13 No1, pp.51-56, March 2008

Detection of Hotspot for Korea Earthquake Data using Echelon Analysis and Seismic Wave Energy

Sanghoon HAN*, Fumio ISHIOKA* and Koji KURIHARA*

Echelon analysis (Myers et al., 1997) is a method to investigate the phase-structure of spatial data systematically and objectively. This method is also useful to prospect the areas of interest in regional monitoring of a surface variable. The spatial scan statistic (Kulldorff, 1997) is a method of detection and inference for the zones of significantly high or low rates based on the likelihood ratio. These zones are called hotspots. The purpose of this paper is to detect the hotspot area for spatial data using echelon. We perform echelon analysis for Korea earthquake data. We use ESRI's ArcGIS that is geographical information system (GIS) software to make the meshed areas and get contiguity information of these areas. With this contiguity information on the meshed areas, we detect the hotspot area using echelon analysis and spatial scan statistics. In addition, we compare with the result of analysis based on the total of number of times simply and the seismic wave energy.

Key words: *Hotspot, Echelon analysis, Spatial scan statistics, Seismic Wave Energy*

1 INTRODUCTION

Recently, the detection of areas which are out of other areas significantly, like incidence region of diseases in certain areas, is an important subject for study. Spatial scan statistic (Kulldorf and Nagarwalla, 1995; Kulldorf, 1997) is a method of detection and inference for the zones of significantly high and low rates based on the likelihood ratio. These zones are called hotspot. Kulldorf detected the hotspot using the method which scans the area based on circular form. This method is excellent to find the circular hotspot, but it is not appropriate for detection of the hotspot consisting of a line or other complex forms. To overcome this problem, Kurihara (2004) proposed a method which scans the area based on echelon. Echelon analysis (Myers et al., 1997, 2002) is the analyzing method to investigate the phase-structure of spatial data systematically and objectively, based on neighbor information between each cell. The echelon analysis is useful to prospect the areas of interest in regional monitoring of a surface variable. Using scanning method based on echelon, we have detected the hotspot which has the irregular shape's form.

The contemporary seismicity of the Korean peninsula is found to be low as compared to the past, and also surrounding Japan and China. Korean seismicity was considerably active the 13th century through the 17th century even though it is relatively calm in the temporal seismic gap. The Korean peninsula is also a region of less stress accumulation due to the tension axis from sea-floor spreading in and near the peninsula. It release of seismic energy not only at the hinge of subduction in the Japanese Island Arc, but also at the block edge of the Pacific Marginal Tectonic Domain in the Chinese continent.

Although the Korean Peninsula is usually believed to have little seismicity, this is not accurate when considering its historical seismicity (Kim and Hyun, 1978; Kim, 1980). In addition, the seismic intensity of earthquake becomes strong in South Korea.

In this paper, we will attempt to find the hotspots based on echelon analysis for South Korea earthquake data. We compare with the result of analysis that when using the total of number of times simply and using the seismic wave energy. First of all, we will get the each contiguity information from obtained data using ArcGIS. We will find the candidate of hotspot for Korea earthquake data from 1978 to 2007 based on echelon analysis.

2 ECHELON ANALYSIS

The echelon analysis for meshed spatial data is based on the areas of relative high and low values of response variables for spatial data. The echelon dendrogram is a graph which represents the surface topology of cellular data and hierarchical structure of these data. The echelon approach aggregates the areas in which the values have the same topological structure and makes hierarchically related structure of these areas. The

---

* Graduate School of Environmental Science, Okayama University, 3-1-1 Tsushima-naka Okayama 700-8530, Japan

spatial data like remote sensing data are given as the pixels of digital value (h) over the $n\times m$ array data $D=\{(x,y)\,|\,x_{i-1}<x<x_i,\ y_{j-1}<y<y_j\}$, i=1, 2,…, n, j=1, 2,…, m. The spatial data like population data have the numbers of population (h) over the meshed area D_{ij}. Thus, the form of such meshed spatial data is given by (i,j,h). For such meshed data, the function $h=f(i,j)$ is not simple function for digital value over a 5×5 array shown in **Figure 1**. The echelon dendrogram is made by the following steps.

Step1) Find the summits of first-order echelons

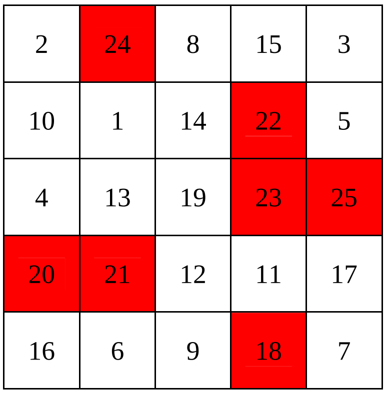
The values in the peak are bigger than that of connected area except the values in the same peak. There are four peaks in this array. The maximum value in the array is 25. The values of 25 belong to the first peak.

	A	B	C	D	E
1	2	24	8	15	3
2	10	1	14	22	5
3	4	13	19	23	25
4	20	21	12	11	17
5	16	6	9	18	7

Fig. 1 The digital data

Step2) Complete the first-order echelons

The maximum value among connected area to 25 is 23. The value of 23 is bigger than that of connected area to 25 and 23. Thus the value of 23 also belongs to the first peak. The maximum value among connected area to 25 and 23 is 22. The value of 22 is bigger than that of connected area to 25, 23, and 22. Thus the value of 22 belongs to the first peak. The maximum value among connected area to 25, 23, and 22 is 19. But the value of 19 is not larger than 21 which is connected to 19. Thus the value of 19 does not belong to the first peak. As a result, the first peak (the first order echelon) consists of the values of 25, 23, and 22, and then its echelon number becomes 1.The maximum value except the values of the first peak is 24. The value of 24 is larger than that of connected area to 24. The maximum value among connected to 24 is 14. But, the value of 14 is not larger than that of 23 connected to 14. Thus the value of 14 cannot belong to the second peak. Therefore the second peak consists of 24, and its echelon number becomes 2. In the same manner, we can find a result that the third peak consists of 21 and 20, the fourth peak consists of 18. These echelon numbers becomes 3 and 4, respectively (**Figure 2**).



2	24	8	15	3
10	1	14	22	5
4	13	19	23	25
20	21	12	11	17
16	6	9	18	7

Fig. 2 The peaks

Step3) Determine the foundations of peaks and foundations

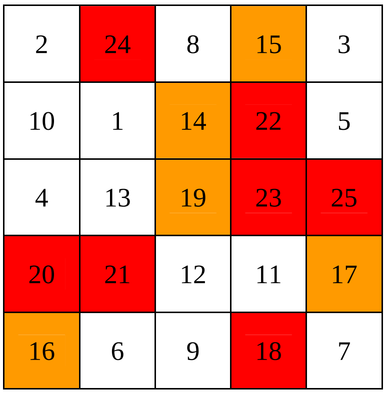
1) The maximum value except four 4 peaks is 19. The value of 19 is the foundation of the peaks whose echelon numbers are 1 and 3. The echelon number of this foundation becomes 5. The echelon number 5 becomes a parent of echelon number 1 and 3.

2) The maximum value except echelon number from 1 to 5 is 17. The value of 17 is the foundation of the echelon numbers 4 and 5. The echelon number of this foundation becomes 6. The echelon number 6 becomes a parent of echelon number 4 and 5.

3) The maximum value except echelon number from 1 to 6 is 16. But, the value of 16 belongs to the foundation of echelon number 6. The maximum value except the echelon number from 1 to 6 is 15. In same reason, the value of 15 belongs to the foundation of echelon number 6.

4) The maximum value except echelon number from 1 to 6 is 14. The value of 14 is the foundation of the echelon number 2 and 6. The echelon number of this foundation is 7. The echelon number 7 is parent of echelon number 2 and 6.

5) The maximum value except the echelon number from 1 to 7 is 13. The value of 13 belongs to the foundation of echelon number 7. The values smaller than 13 also belong to the foundation of echelon number 7 (**Figure 3**).



2	24	8	15	3
10	1	14	22	5
4	13	19	23	25
20	21	12	11	17
16	6	9	18	7

Fig. 3 The foundations

As a result, the structure of this 5-by-5 array data is represented by echelon dendrogram in **Figure 4**.

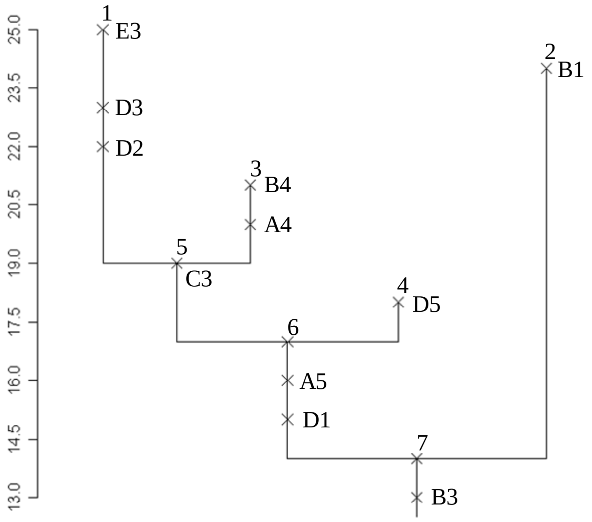


Fig. 4 Echelon dendrogram

3 SPATIAL SCAN STATISTICS

The scan statistic is a statistical method for point processes which is designed to detect clusters and to test whether such an excess have occurred by chance or not.

The spatial scan statistic is used to detect the areas of significantly high or low rates and to perform the regional features, such as hotspots. The spatial scan statistic is defined for circular window area on the map defined by Kulldorff(1997).

There is one area *Z*, which is a subset of whole area *G*. Each individual within the area *Z* has population probability *p* of the attribute, while the population probability for individual outside of the area is *q*. The probability for any individual is independent each other. The null hypothesis is $H_0 : p=q$. The alternative hypothesis to detect high rate is $H_1 : p>q$. Let *n*(*G*) be the total population in the whole area *G*, and *n*(*Z*) be the population within the area *Z*. The *c*(*G*) is the total number of the attribute in the whole area *G* and *c*(*Z*) is the number of the attribute within the area *Z* (**Figure 5**).

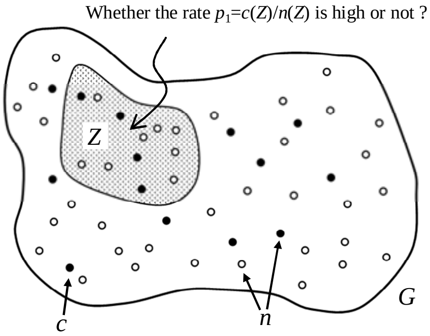

Whether the rate $p_1=c(Z)/n(Z)$ is high or not ?

Fig. 5 Population *n* (○, ●) and the number of the attribute *c* (●) in the whole area *G*.

The model based on Poisson distribution is considered. The probability of *c*(*G*) number of points in the study area is given by

$$\frac{\exp[-pn(Z)-q(n(G)-n(Z))][pn(Z)+q(n(G)-n(Z))]^{c(G)}}{c(G)!} \tag{1}$$

The density function *f*(*x*) of a specific point being observed at location *x* is

$$\begin{cases} \dfrac{pn(x)}{pn(Z)+q(n(G)-n(Z))} & \text{if} \quad x \in Z \\ \dfrac{qn(x)}{qn(Z)+q(n(G)-n(Z))} & \text{if} \quad x \notin Z \end{cases} \tag{2}$$

We can hence write the likelihood function as

$$L(Z,p,q) = \frac{\exp[-pn(Z)-q(n(G)-n(Z))]}{c(G)!} \times p^{c(Z)} q^{c(G)-c(Z)} \prod_{x_i} n(x_i) \tag{3}$$

In order to maximize the likelihood function, we calculate the maximize likelihood function conditioned the area *Z*. The maximum likelihood estimator $\hat{p} = c(Z)/n(Z)$ and $\hat{q} = (c(G)-c(Z))/(n(G)-n(Z))$ are substituted.

$$L(Z) = \frac{\exp[-c(G)]}{c(G)!} \times (\frac{c(Z)}{n(Z)})^{c(Z)} (\frac{c(G)-c(Z)}{n(G)-n(Z)})^{c(G)-c(Z)} \prod_{x_i} n(x_i) \tag{4}$$

The likelihood ratio λ is maximized over all subset area of whole area to detect the hotspots. The test statistic λ of the likelihood ratio test can now written as

$$\lambda = \frac{\max_z L(Z)}{L_0} = \frac{(\frac{c(Z)}{n(Z)})^{c(Z)} (\frac{c(G)-c(Z)}{n(G)-n(Z)})^{c(G)-c(Z)}}{(\frac{c(G)}{n(G)})^{c(G)}} \tag{5}$$

Where, L_0 is the following likelihood function under the null hypothesis.

$$L_0 \overset{def}{=} \sup_{p=q} L(Z,p,q) = (\frac{c(G)}{n(G)})^{c(G)} (1-\frac{c(G)}{n(G)})^{n(G)-c(G)} \tag{6}$$

4 DETECTION OF HOTSPOTS FOR EARTH QUAKE DATA

For analysis, we use acquired Korea earthquake data from 1978 to 2007. For the aggregated 24 zones, a total of 755 cases were reported. In the data, we only use the data of South Korea (613, 81%) except North Korea and missing value (23) in these data. Procedure of analysis is as follows.

Step1) Plot the epicenter of an earthquake on the map using latitude and longitude.

Step2) Create the mesh including all points (nearly 1600km^2, Figure 6).

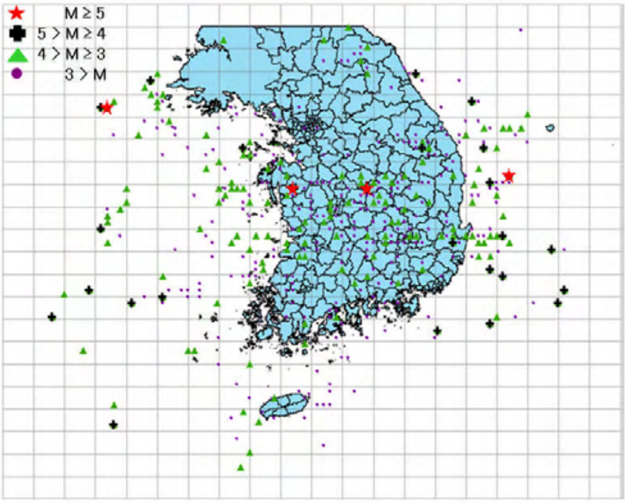


Fig. 6 Plotting on the mesh

Step3) Count frequency of the point in the each mech.

Step4) Perform echelon analysis, and calculate spatial scan statistics and find the candidate of hotspot.

Using contiguity information in the each mesh, we perform echelon analysis. The candidates of hotspot will be located on the top echelon in the dendrogram (**Figure 7**).

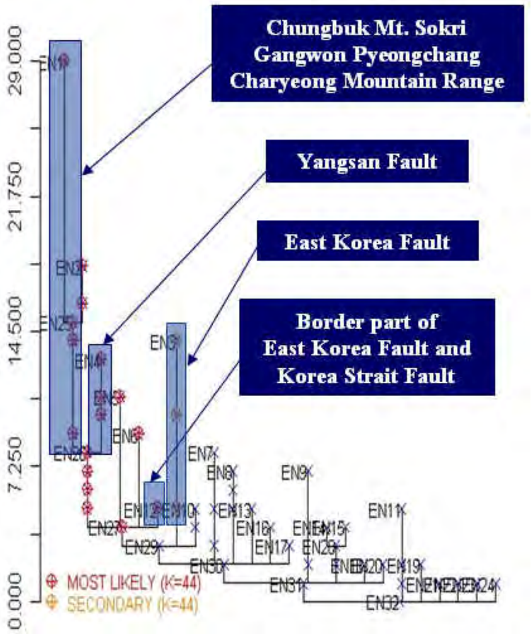


Fig. 7 Echelon dendrogram (The total number of times)

Figure 8 is drawn on the basis of the hotspots detected from the echelon dendrogram of **Figure 7** and is drawn separately for the comprehension. From **Figure 8** we can see the fact that the hotspots are included to middle and southern district of South Korea. That is, hotspot areas are Chungnam, Chungbuk and Gyeongbuk (logλ =354.9508, p-value=0.001). And this hotspot area includes capable fault (Yangsan Fault), fault (East Korea Fault and Korea Strait Fault) and mountain range (Charyeong Mountain Range).

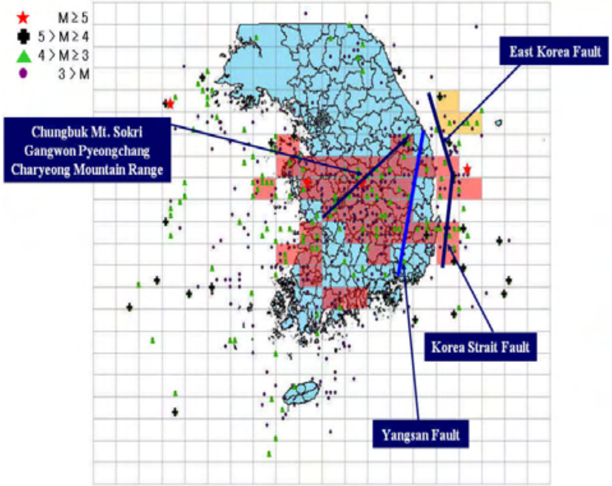


Fig. 8 Hotspots on map (The total number of times)

5 SEISMIC WAVE ENERGY

When does research about earthquake, have better and important use variable including seismic scale (magnitude) than compare with using number of times that earthquake happens simply.

Magnitude is a measure of the strength of an earthquake or strain energy released by it, as determined by seismographic observations. This is a logarithmic value originally defined in 1935 by Charles Richter.

An increase of one unit of magnitude (for example, from 4.6 to 5.6) represents a 10-fold increase in wave amplitude on a seismogram or approximately a 30-fold increase in the energy (seismic wave energy) released. In other words, a magnitude 6.7 earthquake releases over 900 times (30 times 30) the energy of a 4.7 earthquake - or it takes about 900 magnitude 4.7 earthquakes to equal the energy released in a single 6.7 earthquake. There is no beginning or end to this scale. However, rock mechanics seems to preclude earthquakes smaller than about -1 or larger than about 9.5. A magnitude -1.0 event releases about 900 times less energy than a magnitude 1.0 quake. Except in special circumstances, earthquakes below magnitude 2.5 are not generally felt by humans.

Relation with seismic wave energy Es and magnitude M is as follows.

$$\log Es = 11.8 + 1.5M \qquad (7)$$

M is no unit in equation (7), and Es has erg unit. Already explained by the definition of the magnitude, but we can see if magnitude increases as 1 then energy is grown by 30-folds according to this equation also.

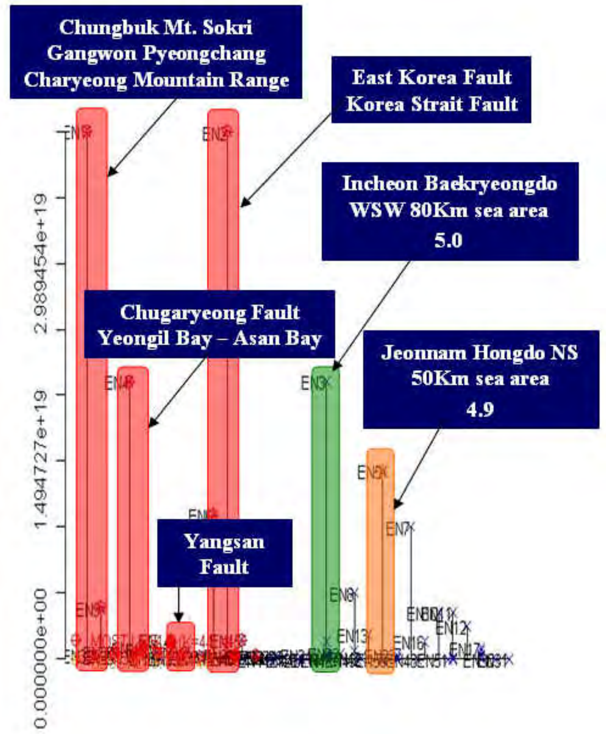


Fig. 9 Echelon dendrogram (Seismic wave energy)

Using this equation, perform echelon analysis once again. The method of the analysis is equal with the front. The candidates of hotspot will be located on the top echelon in the dendrogram (**Figure 9**).

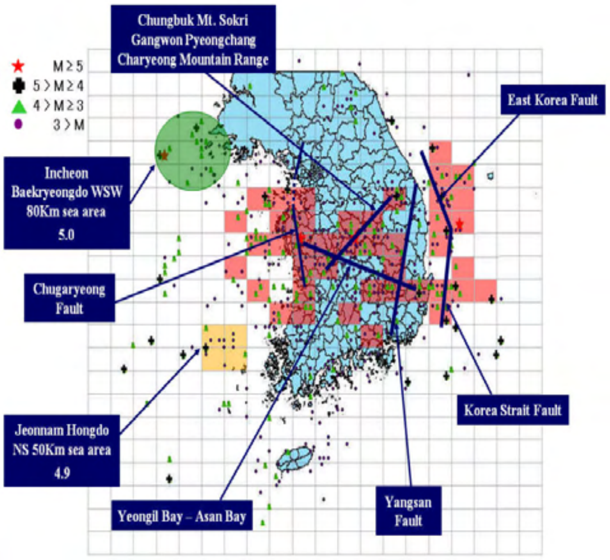


Fig. 10 Hotspots on map (Seismic wave energy)

Figure 10 is drawn on the basis of the hotspots detected from the echelon dendrogram of **Figure 9**. From **Figure 10**, we can see the fact that the hotspot areas are included to middle and southern district of South Korea and same the result of front analysis. But result of using the seismic wave energy (**Figure 10**) is more concrete than compare with result of using the total number of times (**Figure 8**). In this result, hotspot areas situated surrounding the fault and possible fault ($\log\lambda$ =1.7998e+20, p-value=0.001). And also we can see that the hotspot area that when using seismic wave energy includes more many faults and capable faults (Chugaryeong Fault, Yeongil Bay - Asan Bay, East Korea Fault, Korea Strait Fault and Yangsan Fault) than result of analysis that when using the total of number times. That is, the result that when using the seismic wave energy better explain about the earthquake phenomena than when using the total number of times simply. Because magnitude is important variable that explains the earthquake phenomena.

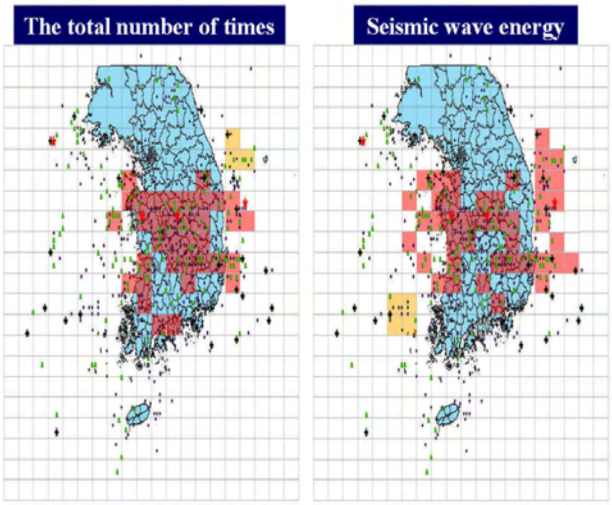


Fig. 11 Hotspots of the total number of times and seismic wave energy

Hotspot areas that overlap in the two results (result of using the total number of times and seismic wave energy) of **Figure 11** are as follows.

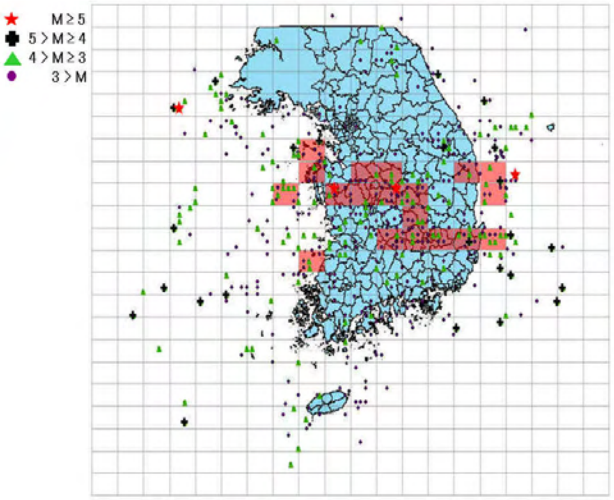


Fig. 12 The place where a hotspot is piled up

Through **Figure 12**, we can see that the place where a hotspot area is piled up is continual earthquake and big magnitude happens.

6 CONCLUSIONS

Echelon analysis enables expression of the phase-structure of space data. The hotspots are given as the upper echelon in the dendrogram. In this paper, we detected the hotspot area for Korea earthquake data using echelon analysis and spatial scan statistics. And after transform data using relation with seismic wave energy Es and magnitude M, analyzed once again, and then compared with the result of analysis that using the total number of times and using seismic wave energy. Result that compare two analyses, using seismic wave energy showed more concrete result than when using the total number of times simply.

A fault is a tectonic structure along which differential slippage of the adjacent earth materials has occurred parallel to the fracture plane. It is distinct from other types of ground disruptions such as landslides, fissures and craters. A fault may have gouge or breccia between its two walls and includes any associated monoclinal flexure or other similar geologic structural feature and that has a structural relationship to a capable fault.

A capable fault is a fault which has exhibited one or more of the following characteristics: (1) Movement at or near the ground surface at least once within the past 35,000 years or movement of a recurring nature within the past 500,000 years. (2) Macro-seismicity instrumentally determined with records of sufficient precision to demonstrate a direct relationship with the fault. (3) A structural relationship to a capable fault according to characteristics (1) or (2) of this paragraph such that movement on one could be reasonably expected to be accompanied by movement on the other.

There are several faults in the Korea. In the South Korea, possible faults are Yangsan fault, Chugaryeong fault and Yeongil bay - Asan bay as capable fault and there are areas in the hotspots that we detected using echelon analysis and spatial scan statistics.

REFERENCES

Kim, SG. and Hyun, BK (1978): Seismicity of the Korean Peninsula and Its Relation with plate tectonics, *Jour. Kor. Inst. Mining Geol.*, vol. 11, no. 2, pp. 69-80

Kim, SG. (1980): Seismicity of the Korean Peninsula and Its Vicinity, *Jour. Kor. Inst. Mining Geol.*, Vol. 13, no. 1, pp. 51-63

Kulldorff, M. and Nagarwalla, N.(1995) : Spatial disease clusters: Detection and inference, *Statistics in Medicine*, 14, pp. 799-810.

Kulldorff, M. (1997): A spatial scan statistics. *Communications in Statistics, Theory and Methods*, 26, pp. 1481-1496.

Kurihara, K. (2004): Classification of geospatial lattice data and their graphical representation. *Classification, Clustering and Data Mining Applications*, Springer, pp. 251-258.

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, pp. 131-152.

Myers, W.L. and Patil, G.P. (2002): Echelon analysis, *Encyclopedia of Environmetrics*, Volume 2, pp. 583-586.