Journal of the Korean Society of Remote Sensing, Vol. 11, No.3, 1995

Spatial Integration of Multiple Data Sets regarding Geological Lineaments using Fuzzy Set Operation

Kiwon Lee* and Kwang-Hoon Chi**

Currently, with Computer Vision Research Lab., AI Div., Systems Engineering Research Inst.,
P.O. Box 1, Yusung, Taejon, 305-600*
Korea Inst. of Geology Mining and Materials (KIGAM), Taejon, 305-350**

퍼지집합연산을 통한 다중 지질학적 선구조 관련자료의 공간통합

이기원 · 지광훈

시스템공학연구소 · 한국자원연구소

Abstract

Features of geological lineaments generally play an important role at the stage of data interpretation concerned geological processes, mineral exploration or natural hazard risk estimation. However, there are intrinsically discordances between lineaments-related features extracted from surficial geological survey and those from satellite imagery; nevertheless, any data set contained those information should not be considered as less meaningful within their own task. For the purpose of effective utilization task of extracted lineaments, the mathematical scheme, based on fuzzy set theory, for practical integration of various types of rasterized data sets is studied. As a real application, the geological map named Homyeong sheet (1:50,000) and the Landsat TM imageries covering same area were used, and then lineaments-related data sets such as lineaments on the geological map, lineaments extracted from a false-color image composite satellite, and major drainage pattern were utilized. For data fusion process, fuzzy membership functions of pixel values in each data set were experimentally assigned by percentile, and then fuzzy algebraic sum operator was tested. As a result, integrated lineaments by this well-known operator are regarded as newly-generated reasonable ones. Conclusively, it was thought that the implementation within available GISs, or the stand-alone module for general applications of

---

Received 15 November 1995 ; Accepted 8 December 1995.

this simple scheme can be utilized as an effective scheme for further studies for spatial integration task for providing decision-supporting information, or as a kind of spatial reasoning scheme.

요 약

지질학적 선구조선에 관한 특성은 일반적인 지질작용, 광물탐사 또는 자연재해 예측과 관련된 자료의 해석 및 분석단계에서 중요한 역할을 한다. 그러나 지질조사에 의하여 드러난 선구조와 위성자료로부터 정량적으로 판독된 선구조정보는 이들 자료 각각의 고유한 유용성을 가지고 있으면서도 여러가지 이유로 인하여 실질적인 차이를 보이는 경우가 많다. 본 연구에서는 선구조와 연관된 위성영상자료의 효과적인 지질학적 활용으로서의 자료통합문제를 해결하고자 퍼지집합연산방법을 실제 사례연구를 통하여 고찰하고자 하였다. 실제 응용으로서, 호명지질도(1:50,000)가 포함하는 지역과 같은 지역의 Landsat TM자료를 이용하였고, 자료 통합과정에서는 지질도상에 있는 선구조선, 위성자료로 부터 도출한 선구조선 및 주요 배수 양상과 같은 자료를 시험적으로 이용하였다. 자료 통합단계에서는 각각의 자료에 있는 영상요소에 대하여 백분위수계념을 응용한 퍼지소속함수를 정의한 뒤, 퍼지 산술합연산과정을 실시되었다. 비록 현재까지의 결과로 본 방법의 유용성에 관한 일반적인 결론을 얻기에는 다소 미흡하지만, 본 사례연구지역의 경우에는 통합된 정보에서 다른 분석방법과는 구별되는 비교적 의미있는 새로운 선구조관련 지질 정보가 내포되어 있는 것으로 판단된다. 결론적으로 본 통합방법은 앞으로 보다 많은 기초및 관련 연구를 통하여 결정판단 보조자료를 얻기 위한 효과적인 공간사고방법으로 간주될 수 있을 것으로 생각된다.

1. Introduction

Currently, most commercial or public-domain GISs (Geographic Information Systems) are increasingly utilized to various geological problems. Related to this computer-oriented geological approach, new technologies such as fuzzy logic, neural network and fractal are also applied to geological sciences due to their powerful potential. As well, the geology-related information extracted from satellite remote sensing have been considered as the significant supporting information to interpret the geological and geophysical processes at the large area. In this study, a model study with consideration of these aspects is presented for the wide utilization for actual geological problem such as the determination of geological lineaments.

O'Leary *et al.*(1976) term 'lineament' as *a mappable, simple, or composite linear feature of a surface, whose parts are aligned in rectilinear or slightly curvilinear relationship and which differs distinctly from the patterns of adjacent features and presumably reflects a subsurface phenomenon.*

Geological lineaments which are revealed on ground surface are generally composed of lithological boundaries, straight stream lines, folds, joints, fault lines, aligned volcanoes, bedding trace, and fracture lines. These features at a small area can be investigated by the detailed survey, but the determination of those features in case of large area with the rugged and mountainous terrain is somewhat difficult. Therefore, the remotely sensed data covering the large study area is necessary as the crucial supporting data for this task.

Proximity analysis (Aronoff, 1989; Legg, 1992), one of GISs' operations, is regarded as the useful method to define fracture zone related to geological lineaments, i.e., fault corridor. However, this analysis is focused on the visualization without the reasoning process. If two or more data sets including vector or raster format data sets are available for an integrated interpretation, the spatial reasoning scheme with mathematical background is necessary: spatial reasoning using evidential belief function approach (Lee *et al.*, 1987), fuzzy logic-based approach (Burrough *et al.*, 1992; Davidson *et al.*, 1992; Gettings and Bultman, 1993; Gong, 1993; Moon, 1993; Lee and Kwon, 1995a; Lee and Chi, 1995), and neural network approach (Lee and Kwon, 1995b). Especially, as for the utilization of fuzzy set theory, Burrough et al. (1992) and Davidson et al. (1992) applied fuzzy set theory to land suitability mapping task, and Gong (1993) studied change detection task through fuzzy set operation to combine change information from different image channels into single-image channel. While, for raster-based data integration task for mineral exploration, Moon (1993), Gettings and Bultman (1993), Lee and Chi (1995), and Lee and Kwon (1995a) presented the related-theoretical basis or actual application with real data sets for fuzzy set operation.

For these approaches, vector data set as one of input layers, is first rasterized. Among them, fuzzy logic scheme is tentatively adopted in this study due to its strength of flexibility related to the rasterizing process. Actually, the conversion process of data structure such as vector-to-raster or raster-to-vector is also a basic procedure within GISs. This can be regarded as the initial step for topological analysis related to the ways in which geographical elements are linked together (Burrough, 1986). On the other hand, semantic analysis of spatial data is more reasonable such as the determination of geological lineaments. In semantic analysis (Lam, 1992), both characteristics of attribute and format of geographical elements are considered during spatial analysis.

The fuzzy membership functions for semantic analysis can be determined either normatively or empirically. The derivation of membership functions is crucial in fuzzy information processing and the lack of simple and generally acceptable methods to build membership functions cause it compare less favorably with other information processing methodologies. The normative approach is commonly used for deriving membership functions for linguistic values because impreciseness inherent to these values are subjective. The empirical approach used in this study follows the objective experimental procedures of the scientific methods found in the measurement theory. Despite the lack of scientific foundation, many fuzzy systems have demonstrated satisfactory performance when compared with two-valued logic system composed of crisp set theory.

While, satellite remote sensing imagery in the geological applications is often of great value in such studies by providing images of large areas which reveal features previously unknown when only field observations or large-scale aerial photography are used. Missallati *et al.* (1979) used various geological data and satellite remote sensing data for uranium exploration.

For a model study, the area covering 1:50,000 Homyeong geological map (1962) is studied. Despite the rugged and mountainous environment, several zones within this study area are economically known to precious and base metal deposit, as well as tens of coal mines. Three kinds of vector data were used: lineaments marked on the geological map, linear features extracted from Landsat TM imagery, and major drainage patterns drawn from image processing of Landsat TM imagery.

2. Fuzzy Set Operation

Both probability and possibility are generally used to describe uncertainty. Probability theory is a tool to study randomness and possibility theory is a tool to study imprecision. While the uncertainty in probability models is caused by randomness, and the uncertainty in possibility models is due to the incompleteness and imprecision in information. Practically, while probability and information theory measure the quantity of information, possibility and fuzzy set theory deal with the semantics of information (Lam, 1992).

Figs. 1 is the schematic diagram to illustrate the conceptual difference between a crisp set and a fuzzy set to represent a certain rasterized data. Fig.1(a) is the original vector data set which is assumed its superimposition on an artificial base map. In this case, this vector layer can be

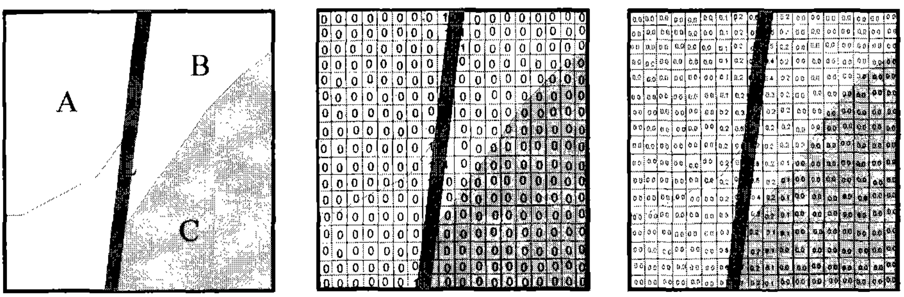

Fig. 1. Schematic illustration of a rasterized vector layer: (a) an arbitrary vector layer - a thick line located at center part on the base map, (b) rasterization using crisp or hard set concepts, and (c) rasterization using fuzzy set concepts.

regarded as one of previously identified lineaments. For information representation for integration task of this data layer with other available ones, rasterization is the initial process. As for this step, either 1 or 0, for the indicative whether existence of information or not, can be easily assigned whether the lineaments cross each cell within a given base map or not, respectively. This represent a crisp set (Fig. 1(b)). While, Fig. 1(c) is an example of fuzzy set representation, given same environment; in this case, contents of information with consideration of certainty/uncertainty of data set used, of possibility of threshold cell, and inter-relationship with the given base map can be quantified as fuzzy membership function, though they are somewhat subjective according to an expert expertise. Sometimes, this subjectiveness is criticized at real world application utilizing original fuzzy set theory by Zadeh(1965), despite its wide potential towards engineering and scientific problems (Lam, 1992). In other words, fuzzy membership function terms possibility measure or compatibility towards a given target in the range of [0,1].

In this study related to this shortage problem of membership function, the following monotonical assignments are attempted as one of possible approaches for generalization of fuzzy membership functions, μs,

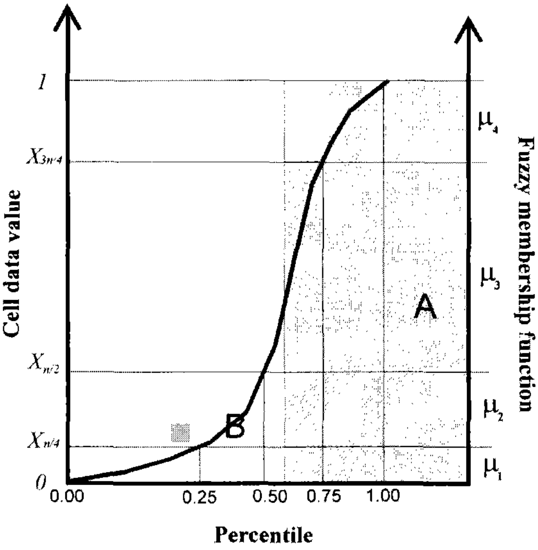

Fig. 2. Relationships between cell data value, fuzzy membership function, and percentile. The filled box A and the curve B represent the original binary-type data and the resampled and interpolated data, respectively. The denotation n mean the class number on histogram of cell data value.

$$
\begin{aligned}
&0 < \mu_1 < (x_{n/4} + x_{n/4+1})/2 \\
&(x_{n/4} + x_{n/4+1})/2\ \mu_2 < (x_{n/2} + x_{n/2+1})/2 \quad \text{.........(eq. 2-1)} \\
&(x_{n/2} + x_{n/2+1})/2\ \mu_3 < (x_{3n/4} + x_{nn/4+1})/2 \\
&(x_{3n/4} + x_{3n/4+1})/2 < \mu_4\ ,
\end{aligned}
$$

where $x_{n/4}, \cdots, x_{3n/4+1}$ are percentiles when n and x are total cell number and pixel value, respectively, as shown at Fig. 2. While, so-called γ operator is effectively utilized (Zimmermann and Zysno, 1980; Moon, 1993) for the integration task of multiple fuzzy sets with respect to a pixel position p,

$$\pi_{\mathrm{p}} = [\prod_{k=1}^{n} (\mu_{1,\cdots,\mathrm{n}})^{1-\gamma} - \ [1 - \prod_{k=1}^{n} (1\ \ (\mu_{1,\cdots,\mathrm{n}}))]^{\gamma}\,{}_{\mathrm{p}}, \quad \text{.........(eq. 2-2)}$$

where π_{p} represents the possibility function. Rationale or relationship between possibility function and fuzzy set were fully discussed in Zadeh(1978), and the meaning of possibility in this study terms compatibility or favorableness of integrated information towards a given task; thus, each possibility function can be interpreted as fuzzy membership function which is composed of newly-generated fuzzy set. Therefore, $\mu_{A \cup B}(x)$ and $_{A \cap B}(x)$ shown in Appendix can be regarded as measure of possibility, though it is not strictly followed by possibility theory distinguished from probability theory, as long as they are utilized to aggregate from subjective-based membership functions for a given integration task.

Meanwhile, the determination of optimum value is closely associated with degree of compensation between the two extreme confidence levels; therefore, in cases of γ=1 (full compensation) or γ=0 (no compensation), these γ operators are equivalent to algebraic sum operator or product sum operator, respectively.

3. Real Application

From the general geologic map published in 1960s(Fig. 3(a)), the southern part of the study area is mostly composed of sedimentary rocks such as sandstone and shale of Carboniferous to Cretaceous. Most mines located below 37°15' N in latitude are coal mines, especially in the sedimentary formations named Rn, Rg, and Ps. However, the area above 37°15' N in latitude is mainly composed of Cambrian sedimentary rocks, showing great limestone zones, e.g. Pungchon limestone and Maggol limestone. Unlike dominant coal mines in the southern area, the polymetallic orebodies in the northern part contain trace elements such as Fe, Pb, Zn, Mo, Ag, and Au. Major lineaments are developed along with boundaries between sedimentary formations, as expected. As for the main interest of this study, geological lineaments in the geological map (Fig. 3(a)) are composed of lithological boundaries, straight stream lines, folds, joints, and fault lines. While, geochemical survey and airborne geophysical survey within this study area were carried out by Jin et al(1983) and Korea Inst. of Energy and Resources (1983), respectively.

A false-color composite (753 bands) of the geometric-corrected Landsat TM image(mid-Oct., 1989) of the same covered area is presented with extracted lineaments in Fig. 3(b). In lineament extraction process, Chi(1994)'s method based on Wang and Howarth(1987)'s algorithm which can be perform by calculating acuteness and size of an arbitrary point on two

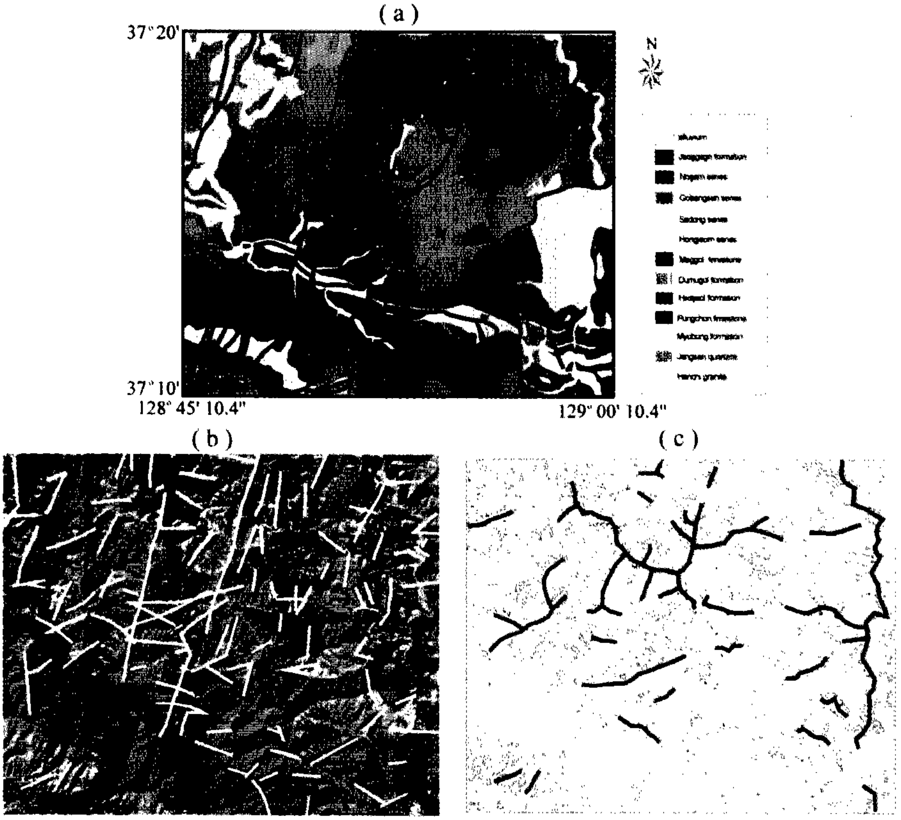

Fig. 3. Input data sets: (a) Geological map and the major lineaments (Homyeong sheet, Taebaek area, Geological Survey of Korea, 1962), (b) Extracted lineaments superimposed on a false-color composite image (431) of Landsat TM data, and (c) Contour map, related to major drainage pattern, traced above 85 percentile of Landsat TM imageries.

dimensional right angled coordinate was previously used. For instance, a false-color composite image with 754 or 753 bands is known to be useful for the geological application such as discrimination of mineral and rock types in arid or semi-arid areas, while a false-color composite with 431 or 543 bands is for mapping of urban features and vegetation types, though just general case excluded areal locality and geobotany. In this area, a composite image with 431 band instead of a composite image 754 band is presented as the base image due to geological features showing mostly sedimentary rocks or formations with rare igneous or volcanic rocks. Extracted lineaments were superimposed on the false color composite image as white thick lines (Fig. 3(b)).

Fig. 3(c) is the result of automatic contouring above specified color value in 8-bit color value (0-255), and the color values above 85 percentile on the histogram of pixel frequency with regard to a grey-converted single channel image of 753 bands. On the contour lines, major drainage pattern was overlaid as thick black vectors, and these vectors were checked with the topographic map in this area. From this figure, the mixed pattern of deranged and dendritic pattern is revealed in this area. This feature can be regarded as the subtle geological lineament somewhat concerned further watershed problems. As initial data sets, these vector data sets drawn from Fig. 3(a), Fig. 3(b), and Fig. 3(c) are used. Fig. 4(a) is a vector layer superimposed Fig. 3(a) with Fig. 3(b).

As an initial step in this study, all vector data sets are converted as raster format of the size of 52 112. This practical rasterization process in this study was manually performed with GISs concept. Computerized rasterization process is beyond the scope of this study, though it is effective. In this process, logical numbers such as 0 or 1 are assigned at each cell indicating non-

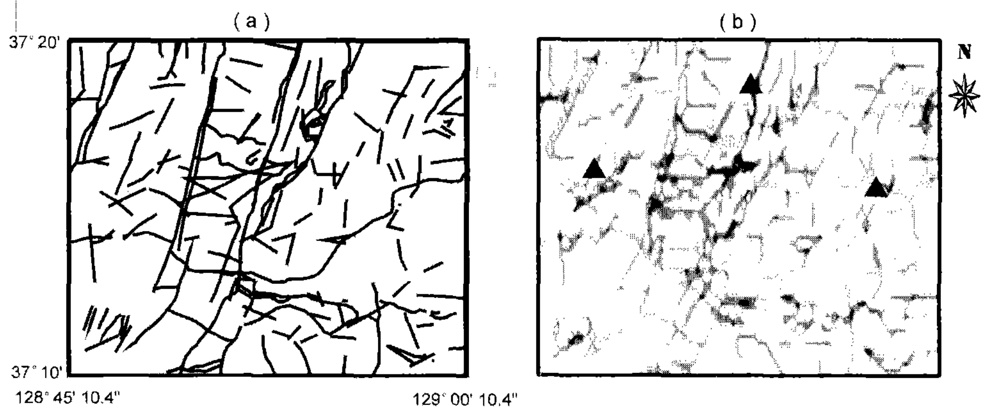

Fig. 4. (a) Overlapped lineament of Figs. 3(a) and (b) by simple vector adding operation without fuzzy concepts, (b) Integrated lineament map by fuzzy algebraic sum operation with differential weight in each data set as newly generated decision supporting layer.

crossing or crossing of lineament with respect to each cell, according to crisp set form mentioned previously, and then each raster data set composed of logical number of 0 or 1 was reproduced by resampling and interpolation. Therefore, at each cell located nearby 1, the floating number gradually closed to 1.0 is automatically assigned.

As the following step, fuzzy membership function with respect to each cell was assigned by eq. 2-1 utilizing percentile concept. Percentile value depends on total class number of histogram. In this study, total 16 classes is firstly setup, and the membership grades of μ_4, μ_3, μ_2, μ_1 are used as 0.4, 0.3, 0.2, 0.0, respectively (Fig. 2), and these term differential weighting membership function. While, it is assumed that all input data sets are noise-free. Actual converted data sets are not presented because assignment processes of fuzzy membership functions were internally performed within integrated systems.

For fuzzy set aggregation, algebraic sum operator (γ=1) of eq. 2-2 is used. Fig. 4(b) is the result of algebraic sum operation with consideration of different weights: Fig. 3(a) of 0.3 as maximum, Fig. 3(b) of 0.4 as maximum, Fig. 3(c) of 0.2 as maximum. This weighting factor appears to be somewhat subjective, but the result contains the semantics about the certainty for previously identified geological lineaments, in some extents. Actually, weighting function problems related to fuzzy set theory were discussed and adopted in Burrough *et al.* (1992) and Davidson *et al.* (1992); however, selection of weighting factor in this study is not followed by those pervious works because the purpose of each integration task is different due to different features of original data such as vector and raster.

Interpretatively, the resultant layer (Fig. 4(b)) can be regarded as decision-supporting indicator for the lineament intensity of existence/non-existence and certainty/uncertainty with regards to a given base map. Especially, filled triangular marks in Fig 4(b) represent the location of known polymetallic ore deposits related to mineral exploration task, and several potential sites of prospect are delineated on this base map. However, it is not always necessary that all information contained within newly-generated layer indicate sites of prospect because any kinds of information about lineaments are just one important factor for this geological problem.

4. Concluding Remarks

1. In aggregation of fuzzy membership function, algebraic sum operator is tested to obtain significant newly-generated information about overall geological lineaments. The resultant layer of this case study by using differential weighting membership functions and algebraic sum operation tentatively provides high possible area for lineament intensity or certainty/uncertainty of existence. Though several sites showing active lineaments are closely related to geological processes such as ore deposits or natural hazard, high intensity which

are generated by overlapping of independent inputs may be misconducted as non-reasonable results.

2. Fuzzy set operation, one of spatial reasoning with the perspectives of GIS/RS, can be easily implemented as GISs operation handling RS data sets and other raster-based data, but the evaluation stage of target-dependent fuzzy membership function should be preliminarily carried out. While, it is thought that the implementation within available GISs, or the stand-alone module for general applications of this simple scheme can be utilized as an effective scheme for spatial reasoning process for spatial integration task.
3. As further research in conjunction with the proposed scheme of this study, it is thought that the rule-base approach of AI/Expert systems to identify attributes of each lineament is possible; furthermore, this study is still in progress with other available data sets such as DEMs (Digital Elevation Models) and airborne geophysical data as well as rasterized data used in this study.

5. Acknowledgments

The authors are grateful to the anonymous reviewers for comments on the earlier version of manuscript. Especially, the first author must express many thanks to Prof. Moon, W. M. at Geological Sciences, University of Manitoba, Canada and Prof. Kwon, B.-D. at Dept. of Earth Sciences, Seoul National University for their valuable advises.

6. References

Anoroff, S., 1989, Geographic Information Systems: A Management Perspective. WDL Publications, Ottawa, 294p.

Burrough, P. A., 1986, Principles of geographical information systems for land resources assessment. Oxford University Press.

Burrough, P. A., MacMillan, R. A., and van Deursen, W., 1992, Fuzzy classification methods for determining land suitability from soil profile observations and topography. Jour. of soil science, v. 43, p.193-210.

Chi, K.-H., 1994, A study on the extraction of groundwater potential area utilizing the remotely sensed data, Jour. of the Korean Soc. of Remote Sensing, vol.10, p.109-120.

Davidson, D. A., Theocharopoulos, S. P., Bloksma, R. J., 1992, A land evaluation project in Greece using GIS and based on Boolean and fuzzy set methodologies. Int. Jour. of GISs. v.8, p.369-384.

Gettings, M. E. and Bultman, M. W., 1993, Quantifying favorableness for occurrence of a mineral deposits type using fuzzy logic- an example from Arizona. USGS Open file report 93-392, 24p.

Gong, P., 1993, Change detection using principal component analysis and fuzzy set theory. Canadian Jour. of Remote Sensing, v. 19, p.22-29.

Harris, R., 1987, Satellite Remote Sensing- An Introduction. Routledge & Kegan Paul, London, 220p.

Jin, M. S., Lee, J. S., Kim, S. J., and Lee, C. Y.,1983, Geochemical maps for Homyeong sheet in the Taebaeksan mineralized belt. Korea Inst. of Energy and Resources.

Korea Inst. of Energy and Resources, 1983, Aerial gamma ray and magnetic survey map: Taebaek, Jincheon, and Imgye (1:50,000).

Lam, S. S. M., 1992, Uncertainty management with fuzzy sets in a commercial vector GIS. UCGE Report # 20059, Univ. of Calgary, 99p.

Lee, K. and Kwon, B.-D., 1995a, Mineral exploration using GRASS-GIS and fuzzy spatial reasoning (Abstract). Proc. of Economic and Environmental Geology, p.58.

Lee, K. and Kwon, B.-D., 1995b, Application of artificial neural networks to spatial integration of raster-based geological data. Jour. of the Korean Earth Sciences, v.16, p.358-367.

Lee, K. and Chi, K.-H., 1995, Theoretical background for data-driven integration of raster-based geological information. Jour. of the Korean Society of Geo-spatial Information systems, v.3, p. 115-121.

Lee, T., Richards, J. A., and Swain, P. H., 1987, Probabilistic and evidential approaches for multisource data analysis. IEEE Trans. Geosci. Remote Sensing, v.25, p.283-293.

Legg, C., 1992, Remote Sensing and Geographic Information Systems: Geological mapping, mineral exploration and mining. Ellis Horwood, New York.

Moon, W. M., 1993, On mathematical representation and integration of multiple spatial geoscience data sets. Canadian Jour. of Remote Sensing, v.19, p.63-67.

Missallati, A., Prelat, A. E., and Lyon, R. J. P., 1979, Simultaneous use of geological, geophysical, and Landsat digital data in uranium exploration. Remote Sensing of Environment. v.8, p.189-210.

OLeary, D.W., Friedman, J. D., Pohn, H. A., 1976, Lineament, linear, lineation: some proposed new standards for old terms. Geological Society of America Bulletin, v. 87, p. 1463-1469.

Penn, B. S., Gordon, A. J., and Wendlandt, R. F., 1993, Using neural networks to locate edges and linear features in satellite images. Computer & Geosciences, v. 19. p.1545-1565.

Wang, J. F. and Howarth, P. J., 1987, Automated road network extraction from Landsat TM imagery, ASPRS, v., p. 429-438.

Zadeh, L. A., 1965, Fuzzy sets. IEEE Information and Control, v.8, p.338-352.

Zadeh, L. A., 1978, Fuzzy sets as a basis for the theory of possibility. Fuzzy Sets and Systems, v.1, p.3-28.

Zadeh, L. A., 1984, Making computers think like people. IEEE spectrum, v.8, p.26-32.
Zadeh, L. A., 1988, Fuzzy logic, IEEE Computer, v.8, p.83-93.
Zimmermann, H. and Zysno, P., 1980, Latent connectives, human decision making. Fuzzy Sets and Systems, v.4, p.37-51.

Appendix: Fundamentals of Fuzzy Set Theory

The fuzzy sets (Zadeh, 1984) are defined as

> *sets that do not have a crisply defined membership, but rather allow objects to have grades of membership from 0 to 1.*

A fuzzy set of attribute values is defined mathematically as follows (Zadeh, 1965). If $X=\{x\}$ denotes a universe of the attribute values, then the fuzzy set A in the X is the set of ordered pairs

$$A = \{ x, \mu_A(x)\}, x \in X$$

where $\mu_A(x)$ is known as grade of membership of x in the A and $x \in X$ means that x is a value contained in X. Usually, $\mu_A(x)$ is the number in the range [0,1] with 1 representing full membership and 0 non-membership. The grade of membership x in A reflect a kind of ordering that is not based on probability but on admitted possibility. The value of $\mu_A(x)$ of the attribute value *x* in *A* can be interpreted as the degree of compatibility of the predicate associated with set *A* and attribute value *x*.

While, fuzzy logic terms as (Zadeh, 1984)

> *a kind of logic using graded or qualified statements rather than ones that are strictly true or false. The results of fuzzy reasoning are not as definite as those derived by strict logic, but they cover a large field of discourse.*

The Max and Min operators are the basic tools used for the aggregation of fuzzy sets. Let A and B be two fuzzy subsets of the universe, U:

$$\mu_{A \cup B}(x) = Max\ (\mu_A(x), \mu_B(x))$$
$$\mu_{A \cap B}(x) = Min\ (\mu_A(x), \mu_B(x))$$

Spatial Integration of Multiple Data Sets regarding Geological Lineaments
using Fuzzy Set Operation
by Kiwon Lee and Kwang-Hoon Chi