Jour. Korean Earth Science Society, v. 25, no. 1, p. 10–21, January 2004

Quantitative Assessment of Input and Integrated Information in GIS-based Multi-source Spatial Data Integration: A Case Study for Mineral Potential Mapping

No-Wook Park[1,*], Byung-Doo Kwon[2], Kwang-Hoon Chi[1], and Kiwon Lee[3]

[1]Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources,
30 Gajeong-dong, Yuseong-gu, Daejon 305-350, Korea
[2]Department of Earth Science Education, Seoul National University,
San 56-1, Shilim-dong, Kwanak-gu, Seoul 151-748, Korea
[3]Department of Software Systems, Information Engineering Division, Hansung University,
389 Samsun-dong 2-ga, Sungbuk-gu, Seoul 136-792, Korea

Abstract : Recently, spatial data integration for geoscientific application has been regarded as an important task of various geoscientific applications of GIS. Although much research has been reported in the literature, quantitative assessment of the spatial interrelationship between input data layers and an integrated layer has not been considered fully and is in the development stage. Regarding this matter, we propose here, methodologies that account for the spatial interrelationship and spatial patterns in the spatial integration task, namely a multi-buffer zone analysis and a statistical analysis based on a contingency table. The main part of our work, the multi-buffer zone analysis, was addressed and applied to reveal the spatial pattern around geological source primitives and statistical analysis was performed to extract information for the assessment of an integrated layer. Mineral potential mapping using multi-source geoscience data sets from Ogdong in Korea was applied to illustrate application of this methodology.

Keywords: data integration; multi-buffer zone; multi-source data; mineral potential

Introduction

New advances and rapid development of various data acquisition techniques and computer resources have fostered the use of integrated analysis of spatial data from multiple sources in order to make optimized decisions in the field of Geographic Information System (GIS). For efficient extraction of the higher level information from multi-source spatial data, GIS can be effectively used for spatial management and manipulation functionality within spatial databases. Though GIS provides some spatial analysis functionalities such as network analysis, proximity analysis and overlay analysis, traditional GIS functionality focuses on visual representation, not semantics and is developed with insufficient mathematical understanding of the data. Thus, insufficient consideration for geoscience data sets may result in a severely erroneous decision-making (Park et al., 2003).

Since the early 1990s, among GIS application issues dealing with geoscience data, computer-based methods for spatial data integration using multi-source geoscience data sets have been studied; Bayesian approaches (Chung and Fabbri, 1999; Park et al., 2003), weight of evidence (Bonham-Carter and Agterberg, 1988), evidential belief function theory (Moon, 1990), fuzzy set theory (An et al., 1991), artificial neural network (Lee and Kwon, 1995).

However, with regard to quantitative modeling framework of the data, the influence of input layers and error assessment for the integration task has not as yet been forthcoming. Most research has focused on the representation of data sets and interpretation of a resultant integrated layer, not concluding on the assessment of quality and spatial pattern of input layers in the scope of exploratory data

*Corresponding author: nwpark@kigam.re.kr
Tel: 82-42-868-3287
Fax: 82-42-867-0421

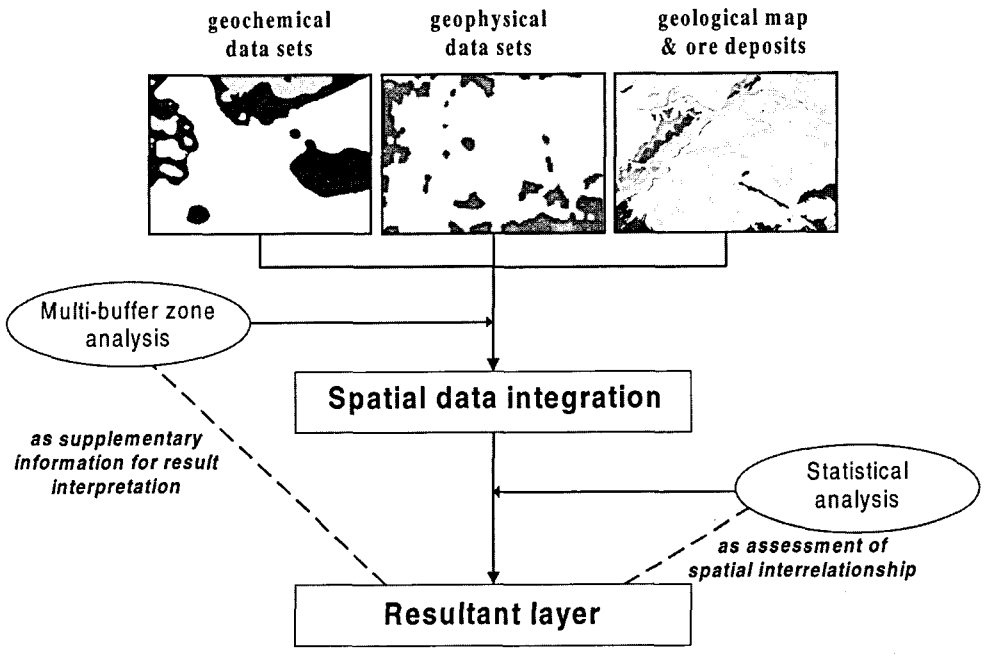

Fig. 1. Schematic diagram showing the processing flow used in this study.

analysis. Regarding exploratory data analysis, although schemes are newly proposed in geostatistics, they have not as yet been in conjunction with the spatial data integration.

The main purpose of this study is to propose a kind of hybrid-type spatial analysis scheme in a spatial data integration framework. This work consists of three parts. One part is the implementation of spatial integration for mineral potential mapping using multi-source geoscience data sets. Another part is the exploration of input layers, as a kind of pre-processing and/or post-processing. The third part is quantitative assessment between a resultant integrated layer and the input layers, as a kind of post-processing. As an exploration of input layers, or supplementary information with respect to spatially integrated information, "multi-buffer zone analysis", which is an extended version of the buffering operation, is newly proposed and applied to multiple geoscience data. Among the known spatial integration methodologies, certainty factor (CF) estimation was applied for mineral potential mapping. For quantitative assessment of a spatially integrated layer or extraction of statistical information, statistical analysis based on a contingency table was carried out.

This paper is structured as follows: in the next section, we give schemes and the rationale applied in this study. Then, we apply these schemes to real data sets from Ogdong in Korea and discuss the results. Finally, we conclude with a summary and remarks.

Applied Schemes

In this study, the CF estimation approach was applied for the data integration scheme. Multi-buffer zone analysis and statistical analysis, which do not depend on the algorithm used for integrating the data layers, were performed for the extraction of the supplementary information towards spatially integrated information (Fig. 1).

CF estimation

Among various spatial data integration schemes, CF estimation, one of data-driven methods which use known target information, was applied. Because

the main purpose of this study is the development of quantitative analysis in the spatial data integration task, rather than the application of the certain data integration method, short explanation of CF estimation is given. Detail theoretical background for the CF estimation is discussed in Shortliffe and Buchanan (1975) and Chung and Fabbri (1993).

CF estimation measures certainty level of conditional probability with respect to a priori probability given a certain evidence. A certainty factor (CF) at p for the kth layer, denoted by CF_k (p), is defined as the change in certainty that the proposition (a pixel p contains deposits of type D) is true, from without the evidence at p to given the evidence at p in the kth layer (Chung and Fabbri, 1993). CF's are reals ranging from −1 to 1. Positive numbers for CF correspond to an increase in certainty in a proposition after the evidence is observed, whereas negative numbers correspond to a decrease in certainty. The extreme values −1 and 1 represent "no" and "yes", respectively.

The definition discussed by Heckerman (1986) is followed.

$$CF_k(p) = \begin{cases} \dfrac{Prob_k(T_p|v_k(p)) - Prob_k(T_p)}{Prob_k(T_p|v_k(p))(1 - Prob_k(T_p))} & \text{if } Prob_k(T_p|v_k(p)) > Prob_k(T_p) \\ \dfrac{Prob_k(T_p|v_k(p)) - Prob_k(T_p)}{Prob_k(T_p)(1 - Prob_k(T_p))} & \text{if } Prob_k(T_p|v_k(p)) < Prob_k(T_p) \end{cases} \quad (1)$$

where,

$Prob_k(T_p)$: a priori probability that a pixel p contains certain deposit before any evidence is not given.

$Prob_k(T_p|v_k(p))$: the conditional probability that a pixel p contains certain deposit given the evidence.

The rules for combining CF values ($CF_i(p)$ and $CF_j(p)$) are as follows:

$$CF_{combine}(CF_i(p), CF_j(p)) = \begin{cases} CF_i(p) + CF_j(p) - CF_i(p) \cdot CF_j(p) & CF_i(p), CF_j(p) > 0 \\ (CF_i(p) + CF_j(p)) / (1 - \min(|CF_i(p)|, |CF_j(p)|)) & CF_i(p), CF_j(p) \text{ different sign} \\ CF_i(p) + CF_j(p) + CF_i(p) \cdot CF_j(p) & CF_i(p), CF_j(p) < 0 \end{cases} \quad (2)$$

Multi-Buffer Zone Analysis

A general schematic view for multi-buffer zone analysis is outlined in Fig. 2. This approach is not a technical improvement of the GIS processing algorithm, but a somewhat practical approach based on a GIS buffering operation. The buffering operation around features is a very useful and standard GIS capability. A buffer is a newly generated polygon enclosed in an area within a specified distance from a certain primitive source and is an expression of the influence zone with respect to a certain distance. Specifically, multi-buffer zone analysis focuses on the change of a spatially distributed pattern in buffered zones or buffering boundaries along with an equi-interval or a distance. In multi-buffering, a clump operator examines every cell to see if the dominant circular pattern with respect to a given source can be incrementally determined by cell counting. Through this scheme, spatial behaviors of multiple data layers around given source primitives can be extracted.

Generally, buffer zones can be created around three data types: 'point', 'line', and 'polygon'. In geo-based source primitives, features such as ore deposits, seismic epicenters or various types of point sources causing natural hazards are compatible with 'point' data types. Features such as surface fault lines and drainage networks are compatible with 'line' types and features such as mineral occurrence zones, alteration zones and geological patterns are compatible with 'polygon' data types.

In this study, the locations of mineral occurrence

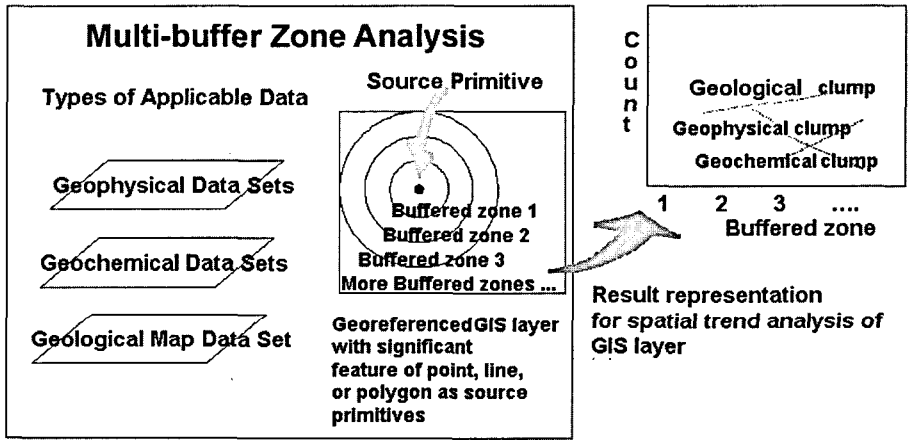

Fig. 2. General procedure for multi-buffer zone analysis.

zones are mainly dealt with as source primitives for assimilation into the spatial integration task. Spatial geological patterns around primitives of significant geological features (e.g. mineral occurrence zones) can be efficiently utilized to delineate complex geological behaviors, particularly when handing multiple data sets of multi-sources or from different backgrounds.

Measures of Spatial Interrelationship

Many different measures of association of data sets have been proposed. In the most general case of two thematic maps, each with multiple classes, the most common measures are based on a contingency table (Bonham-Carter, 1994). A contingency table for cross-tabulation is the table showing discrete frequency or cell-counting in the form of a matrix. After a contingency table is calculated, several statistics related to the measure of association can be obtained to quantitatively assess integrated results or layers.

In this study, entropy measures are used. Entropy measures use proportions and have an advantage over chi-squared measures by being unaffected by measurement units. For entropy measures, let the table between map A and B be called matrix T, with elements T_{ij}, where there are $i = 1,2,\cdots$,n classes of map B (rows) and $j = 1,2,\cdots$,m classes of map A (columns). The marginal totals of T are defined as $T_{i.}$ for the sum of the *i*th row, T_j for the sum of the *j*th column, and $T_{..}$ for the grand total summed over rows and columns.

To compute entropy measures, which vary between 0 and 1, the area proportions give the estimates of probabilities, by dividing each element by the grand total. Assuming that the proportions matrix for map A and map B has been determined from T, entropy measures (H) are defined by the relations below.

$$H = 2\left[\frac{H(A)+H(B)-H(A,B)}{H(A)+H(B)}\right] \tag{3}$$

$$\text{entropy of A: } H(A) = -\sum_{j=1}^{m}\frac{T_{.j}}{T_{..}}\ln\frac{T_{.j}}{T_{..}},$$

$$\text{entropy of B: } H(B) = -\sum_{i=1}^{n}\frac{T_{i.}}{T_{..}}\ln\frac{T_{i.}}{T_{..}}$$

$$\text{joint entropy of: } H(A,B) = -\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{T_{ij}}{T_{..}}\ln\frac{T_{ij}}{T_{..}}$$

In addition, the Yule coefficient and odds ratio are calculated to reveal the dominant layer that influences the integrated layer. The Yule coefficient and odds ratio are defined as follows:

$$\alpha = \frac{\sqrt{T_{11}/T_{21}}-\sqrt{T_{12}/T_{22}}}{\sqrt{T_{11}/T_{21}}+\sqrt{T_{12}/T_{22}}},\quad O_R = \frac{T_{11}T_{12}}{T_{12}T_{22}} \tag{4}$$

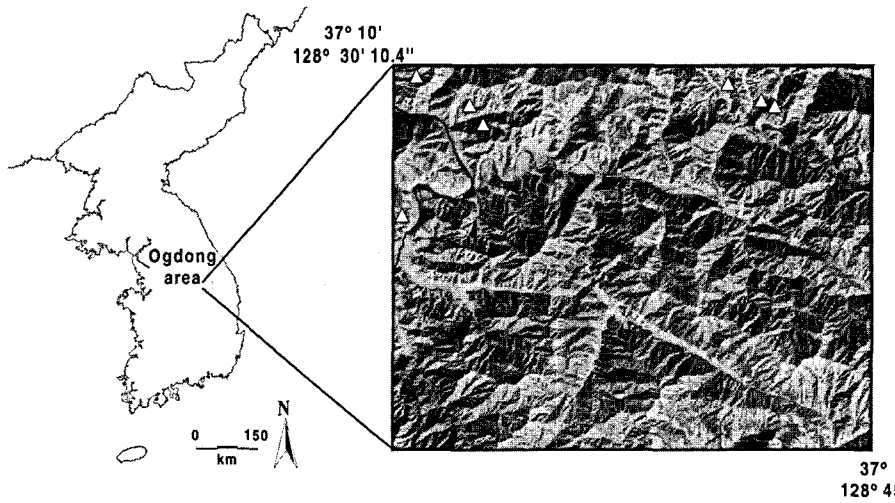

Fig. 3. The location map of the study area. Known polymetallic mines in the study area are represented as triangle marks, draped over Landsat Thematic Mapper band 5 imagery.

where, $A \cap B = T_{11}$, $A \cap \overline{B} = T_{21}$, $\overline{A} \cap B = T_{12}$, $\overline{A} \cap \overline{B} = T_{22}$

Although the statistics above are mainly utilized in the comparison of binary maps, measures of associations between binary patterns can be applied to the comparison of multi-class maps by treating each combination of map classes as a binary. The Yule coefficient ranges in value between −1 and +1, similar to a correlation coefficient. The odds ratio is always positive, being greater than 1 for patterns that are positively associated, 1 if the two patterns are independent and less than 1 if they are negatively associated.

The Study Area and GIS data layers

For the case study of this proposed scheme for mineral potential mapping, multiple geoscience data sets covering the Ogdong in Korea (Fig. 3) were used. It should be pointed out that the purpose of this case study is to illustrate the applicability of the proposed methodology in practice, not to reveal the local characteristics of the study area.

Geologically, the major part of the study area is composed of metasediments and most of the polymetallic mines are located in the western and eastern areas. Ore deposits are mainly located within the thick limestone series of the Joseon Supergroup and the granite; the rich elements of ore deposits are iron (Fe), lead (Pb), and zinc (Zn).

At this time, field or airborne surveyed data sets are in the process of being geocoded into an exchangeable format suitable to GIS environments. Therefore, most data sets in this study were originally text or raw format and pre-processing for geocoding and geo-registering process was carried out.

The GIS-derived data set lists are shown in Table 1. As for the geophysical data sets, airborne surveyed data were composed of: residual magnetic anomaly, radiometric anomaly of gamma ray, potassium (K), thorium (Th) and uranium (U). And as for the geochemical data sets, ground surveyed data of lead (Pb), zinc (Z), copper (Cu), silver (Ag), cadmium (Cd) and U were used. For U, both airborne and surface data sets were obtained, but they were separately processed. Geophysical data sets, originally in vector format, were converted to raster format and geochemical data sets, which were obtained from widely distributed stream rock samples in the study area, were interpolated to obtain grid format data. The geological map, one of basic

Table 1. GIS data layers used in this study

Class	Type	Airborne geophysical data sets (unit: cpu, gamma)					Geochemical data sets (unit: ppm)						Geologial map
		gamma ray	K	Th	U	Mag.	Ag	Cd	Cu	Pb	U	Zn	
1	min	200	10	5	4	-60	0.10	0.02	10.00	10.00	1.70	17.00	Alluvium
	max	250	15	6	5	-35	0.17	0.50	13.50	20.90	5.50	36.22	
2	min	251	16	7	6	-34	0.18	0.51	13.51	20.91	5.51	36.23	Daedong system
	max	350	26	10	7	-20	0.81	0.98	35.10	49.00	12.90	59.57	
3	min	351	27	11	8	-19	0.82	0.99	35.11	49.01	12.91	59.58	Pyeongan supergroup
	max	450	36	13	9	5	2.09	4.20	93.30	114.80	29.50	133.10	
4	min	451	37	14	10	6	2.10	4.21	93.31	114.81	29.51	133.11	Joseon supergroup (Ordovician)
	max	550	47	17	11	20	3.09	10.40	151.40	175.80	44.70	184.00	
5	min	551	48	18	12	21	3.10	10.41	151.41	175.81	44.71	184.01	Joseon supergrup (Cambrian)
	max	650	58	21	14	80	7.21	14.80	239.10	267.30	67.00	252.95	
6	min	651	59	22	15	81	7.22	14.81	239.11	267.31	67.01	252.96	Pegmatitic migmatite
	max	750	68	24	16	140	15.96	17.21	699.00	5695.00	69.48	495.00	
7	min	751	69	25	17	141							Precambrian schist & quartzite
	max	850	78	28	18	200							
8	min	851	79	29	19	201							Granitic gneisses
	max	950	90	30	22	300							
9	min												Acidic dike
	max												
10	min												Granite
	max												
11	min												Porphyrite
	max												

data sets for this approach, was fully geo-registered into the GIS with the geometric features and their database attributes.

After this pre-processing, all data sets were built into a cell-based database with 30 m resolution and were reclassified with respect to the feature of each data set. In addition, seven known mines in the study area were used as prior evidences in the spatial integration and as sources in the multi-buffer zone analysis.

Results

Spatial integration using CF estimation

Spatial integration using CF estimation was performed using all the data sets for favorable mapping of mineral occurrences. Firstly, in order to carry out the spatial integration, the study area was divided into a number of non-overlapping unique condition subareas by overlaying the input data layers specified in the GIS database (Chung and Fabbri, 1993). An image consisting of the unique condition subareas was generated, where the unique condition subareas mean that pixel values in a subarea have a unique identification number, unique to that subarea. After generating the unique condition subareas, we computed the CF based on the prior probability and the bivariate conditional probabilities of the known mineral deposit occurrences. Afterwards, the CF was computed by combining the certainty factor generated by the above process using a combination rule. As a result of the certainty factor estimation, CF values generally lay between −1 and 1. However, in this study, to emphasize the high potential zones, we classified the resulting image as follows: all the pixel values were sorted in descending order, then the ordered pixel values were classified per high rank 0.5%.

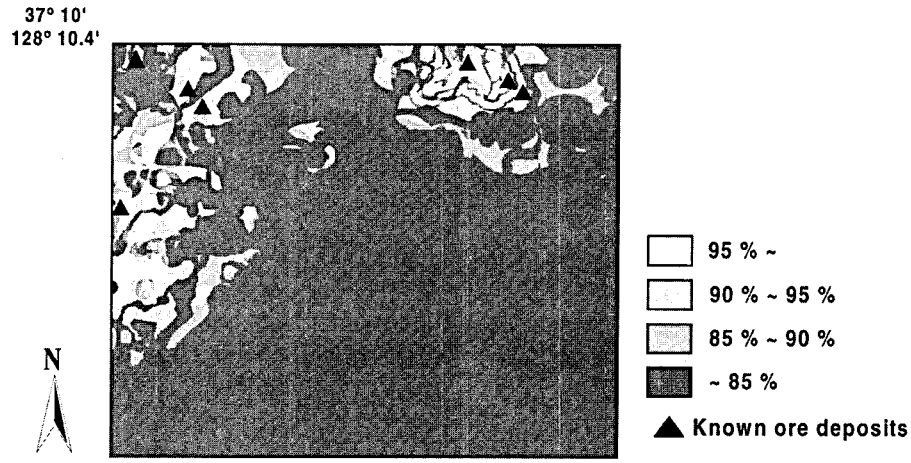

Fig. 4. Spatial integration result using CF estimation.

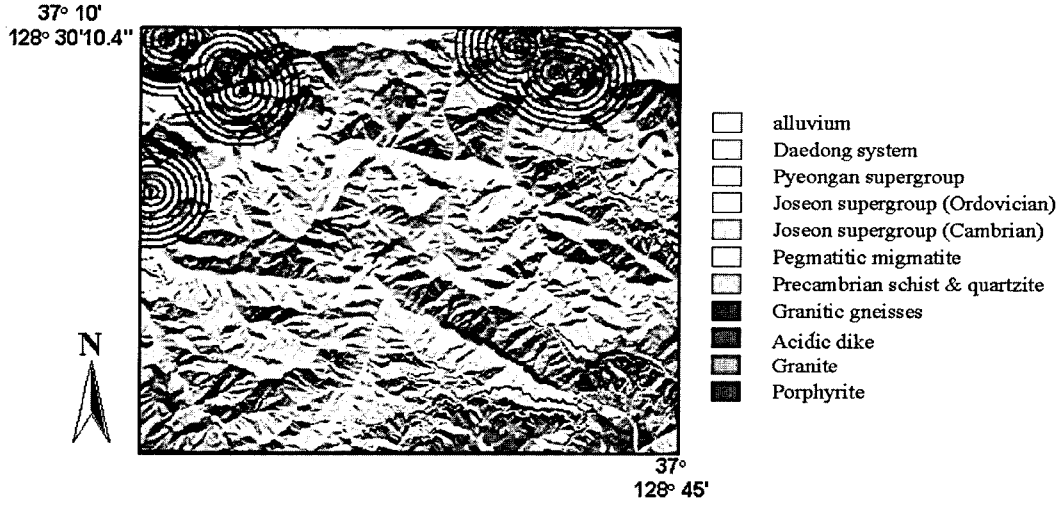

Fig. 5. Buffering zones on the surface relief image draped over the reclassified geological map.

Through this process, the final integrated result was generated as in Fig. 4. Overall, the resultant layer fitted the real situation, showing in situ mineral occurrences. This is basically caused by the intrinsic characteristics of the CF estimation method, which is based on probabilistic relations between a known occurrence event and input layers. For actual field exploration of this large area, the high ranked zone in CF values can be considered as likely new target zones for mineral exploration.

Multi-buffer zone analysis result

Fig. 5 represents buffered zones with intervals of 300 m from mine 1 to mine 7. Mine 1, 3, 4, 6 and 7 are Fe rich ore deposits and another mines are Pb- and Zn-rich ore deposits, respectively. Through consideration of the proximity of the mines, the interval of the buffered zone was selected. The whole buffered zone of each layer was obtained radially to 2400 m from the point source location. To carry out multi-buffer zone analysis, each clump in the buffered zone with an interval of 300 m from known ore deposits was represented in plot-style. In this context, the clump stands for the most dominant class value and is regarded as the representative value for a buffered zone. In following

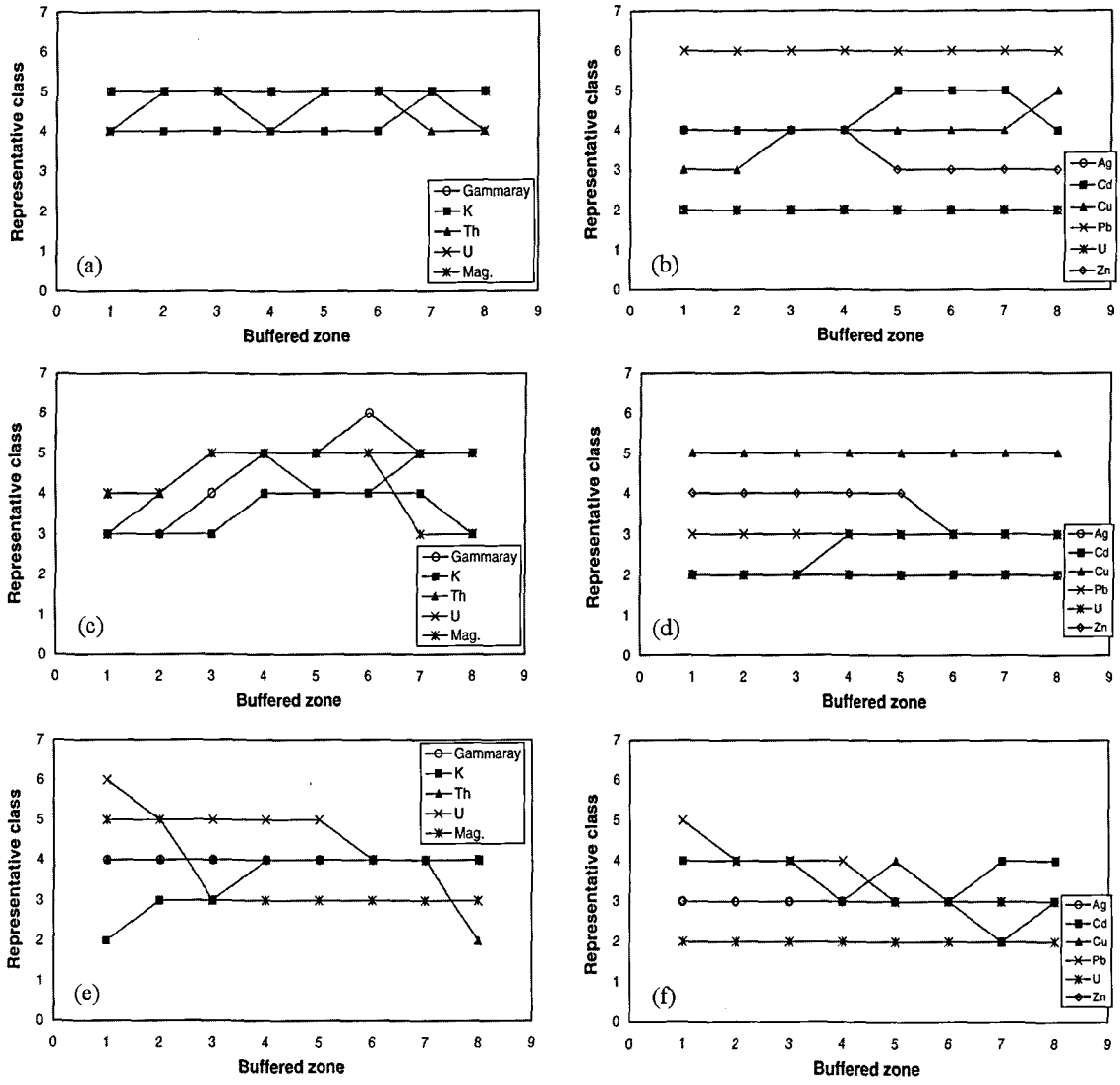

Fig. 6. Multi-buffer zone analysis results with respect to mine 1, mine 3 and mine 7. (a) result of airborne geophysical data sets for mine 1, (b) result of geochemical data sets for mine 1, (c) result of airborne geophysical data sets for mine 3, (d) result of geochemical data sets for mine 3, (e) result of airborne geophysical data sets for mine 7, (f) result of geochemical data sets for mine 7.

results, range of each class can be referred from Table 1.

In the case of mine 1 (class 4 in geology, Fig. 6 (a), (b)), spatial trends in airborne data sets appeared to be the same class as in airborne data sets; class 5 of airborne magnetic anomaly. However, geochemical data sets showed relative variations; class 4 to 5 of Cd, class 3 to 5 of Cu, class 3 to 4 of Zn, except class 2 of Ag, and class 6 of Pb. In the case of mine 3 (class 3 in geology, Fig. 6 (c), (d)), some different distributed patterns were visible. Airborne data sets varied relatively more than geochemical data sets; class 3 to 5 of airborne magnetic anomaly, class 3 to 6 of gamma ray and class 3 to 5 of airborne uranium. It was thought that these patterns resulted from class 3 in geology; airborne radiometric data usually showed a large response in coal seams or black shale.

Table 2. Entropy measures, main classes of all input layers corresponding to selected zones over 95% by CF estimation and the Yule coefficient and odds ratio with respect to integrated layer

Input layers		Entropy measures	Main class	Yule coefficient	Odds ratio
Airborne geophysical data sets	Gamma ray	0.0159	7	0.350	4.313
	K	0.0557	2	0.260	2.903
	Th	0.0634	5	0.393	5.269
	U	0.0243	6	0.399	5.414
	Mag.	0.066	7	0.607	16.739
Geochemical data sets	Ag	0.081	2	0.822	105.295
	Cd	0.070	5	0.572	13.500
	Cu	0.151	3	0.335	4.036
	Pb	0.152	6	0.618	17.972
	U	0.096	2	0.488	8.460
	Zn	0.135	4	0.551	11.938
Geological map		0.145	Granite	0.518	9.909

Although class 3 was the representative value, the heterogeneity of the geology in the buffered zone resulted in variation of airborne radiometric data.

As shown at Fig. 6 (e), (f), in mine 8 (class 3 in geology), in spite of a point source of Pb- and Zn-rich ore deposit, the classified pattern of Zn was invariant; however, there was a class change of 3 to 5 for Pb. Furthermore, the clumped trend on most layers was slightly changed, class 5 to 3 for the airborne magnetic anomaly, class 6 to 4 for the airborne U, and class 2 to 4 for K, except for gamma ray and U.

Actually, this changed pattern and the circular clumped trend extracted by these example case studies in this area may not be explained fully by traditional geological investigation. However, it is regarded as informative supporting evidence for the spatial integrated and newly generated layer such as in Fig. 4, with regard to GIS's application in prediction-related thematic mapping of mineral occurrences. Some advantages over traditional interpretation in the non-GIS environment are as follows: provision of supporting information to build a knowledge base around targets for a geo-based expert system ore deposit model for mineral exploration and the delineation of a generalized model for given source features.

Quantitative assessment of spatial interrelationship

To analyze quantitatively the spatial integration result, firstly a contingency table of each input layer was constructed, and then, entropy measures were computed to reveal the relationship between original input layers and the resultant layer. In addition, the Yule coefficient and odds ratio were computed to reveal the most dominant class value within classified zones over highest rank zone (95%) of the CF estimation layer. To calculate these statistics, the main classes of all input layers corresponding to selected zones over 95% by CF were masked and then these masked zones were reproduced as a binary map.

These results (Table 2) could be regarded as a kind of post-processing or assessment of modeling results of the spatial integration task with respect to the spatially integrated layer for mineral potential mapping.

As for entropy measures, the highest value represents the strongest association. The overall tendency of each layer was to show the highest rank of geochemical data sets (Pb, Cu, Zn) and geology. As for airborne geophysical data sets, the residual magnetic intensity data showed higher values than radiometric data sets. However, the integrated result was

relatively more affected by geochemical data sets and the underlying geology than airborne geophysical data. It is thought that these results are influenced by a major element of the ore deposit and the host rock. Considering the mineral occurrence, the integrated result is more highly affected by the surface catchment data, which are directly related to ore deposits and residual magnetic intensity, because of the iron, a regionally important and dominant element. Thus, in the scope of exploratory data analysis, if there exists an anomaly in the geochemical data set, then that can severely affect the integration result. Among the geochemical data sets, Ag and U elements show a relatively smaller effect than other elements. However, the effects of these elements are larger than those of the airborne geophysical data sets. The high effect of Ag and U, which generally show more response with respect to Precambrian metasedimentary rocks than in the sedimentary rocks, can be attributed to a non-diagonal deviation.

The fact that airborne and geochemical U data show different responses can be explained as follows. Although both surveys measure the common element U, airborne geophysical survey measures the response with respect to the surface or subsurface. On the other hand, the geochemical survey was based on a ground collection of stream rock samples. Thus, the ground survey does not directly nor necessarily reflect the underlying geological rock type.

In terms of the Yule coefficient and odds ratio, the statistical result reveals that the dominant class of each data set correlates to a high mineral potential zone. As a result, airborne magnetic intensity, Ag, Pb and Zn all show relatively high values. This means that the spatial pattern for these four data sets is strongly related to the newly integrated layer for prediction of mineral occurrences. As for the airborne geophysical data sets, most elements, except potassium (K), show high class values. The meaning of these values is that the granite zone and the black shale containing coal seams from the Pyongan Supergroup correspond to a high potential zone. It is reasonable that the high class value of the residual magnetic intensity is due to the Fe component of the iron ore deposits. In spite of the high response of potassium with respect to granite, there are two cases where K shows a low value; that low association in the granite zone might result from the large non-diagonal deviation which probably has an effect in computing a binary case. As for the geochemical data sets, the highest class of Pb is related to the integrated layer. In particular, the extremely large value for Ag might reflect a large effect from supplementary classes containing anomalies. Therefore, it is thought that K, Ag, and U elements are needed to assess field data by exploratory data analysis, although we shall not discuss this further in this study. In addition, further research using this approach would be useful to extend the GIS-based exploratory data analysis.

In comparison with the quantitative assessment result with respect to the integrated layer, the main classes of input layers, which have an influence, were not fully fitted to the spatially clumped class (Fig. 6). The main class of Ag, Cd, U and Zn in the geochemical data sets was fitted to the spatially clumped class. However, other data sets were not directly fitted because this combined result might not be directly related to the richest element. The reason is that areas of mineral deposit, modeled with points in GIS, do not normally appear in a wide spatial pattern and a spatial resolution of the cell is not appropriate for the detailed ore deposit model.

Discussion and Conclusions

GIS-based data integration for geoscientific applications can be realized as a spatial analytical functionality. In this case study, we dealt with pre- and post-processing as well as the interpretation of the spatially integrated layer based on specific methodologies for data integration.

The main purpose of this case study was to

extract quantitative information between the integrated layer and input layers, as well as the mineral potential distribution through multi-buffer zone analysis and statistical analysis based on a contingency table. As a result, input layers, which closely affect the integrated layer, and the spatial pattern were revealed. The overall assessment of this integrated result can be considered as a decision-supporting layer from the viewpoint of GIS with additional quantitative evidence, not represented in the integrated layer of Fig. 4. However, in the computed statistical coefficients related to association, those results may not perfectly reflect the relationship between each class, but rather the relationship between non-diagonal components. Notwithstanding this problem, based on the results of this case study, we think that this statistical/quantitative information is helpful to outline the relationship between input layers. In addition, there is a relationship to the integrated layer and to each input layer and even data quality assessment by spatial pattern, which might be overlooked by simple visual interpretation of the resultant layer.

Although multi-buffer zone analysis is not a development of new technology, but a practical scheme using basic GIS function of application level, this proposed scheme can be utilized to reveal the interrelationship of spatial pattern between multiple data sources and developed as supplementary information for interpretation. However, several aspects still need to be addressed; the selection problems of the representative value of each data layer within buffered zones, the uncertainty or accuracy level caused by limitation of cell-based processing and problems of misunderstanding resulting from intrinsically complex interactions between data used. Nevertheless, it is expected that this approach provides possibilities for further extended studies, particularly the handling of other types of geo-based features. It is helpful to analyze the spatially clumped pattern of multiple data oriented to given sources whatever type of primitives is applied, although only one case study related to thematic mapping for mineral occurrences is presented.

Within a spatial data integration framework, analysis of the characteristics of input data should generally precede execution of the spatial integration. Geoscience data and its interpretation are naturally uncertain due to finite, discrete and other limiting characteristics. Anomalous values of data having some physical meaning of natural phenomena and outliers showing extremely large or small value are often included in the data. In particular, outliers unclassified from anomalous values can severely affect the interpretation of a decision-supporting layer generated by spatial data integration. As for data representation and the subsequent interpretation stage, finding an interrelationship among data sets should be considered significant, particularly when the spatial integration task is performed.

From this point of view, spatial integration with a pre- or post-processing step of multi-buffer zone analysis and a post-processing step of statistical analysis in this case study can provide significant information for assessment of the integrated layer and detailed qualitative interpretation, regardless of its result.

In addition, it is expected that these methodologies can be utilized effectively for exploratory analysis to interpret quantitatively an appropriate integration model, particularly when these approaches are applied to data sets within the geoscientific and other fields simultaneously. Conversely, in the scope of exploratory data analysis, these approaches can be utilized to assess multiple geoscience data sets, if provided with sufficient geological evidence in the actual field.

Finally, this case study covers the geoscientific application of GIS in spatial integration perspectives. However, we expect it to be applicable not only for predictive mineral mapping, but also for general or specific GIS schemes handling hybrid geoscience data models.

References

An, P., Moon, W.M., and Rencz, A., 1991, Application of fuzzy set theory to integrated mineral exploration. Canadian Journal of Exploration Geophysics, 27, 1-11.

Bonham-Carter, G.F. and Agterberg, F.P., 1988, Integration of geological data sets for gold exploration in Nova Scotia. Photogrammetric Engineering & Remote Sensing, 54, 1585-1592.

Bonham-Carter, G.F., 1994, Geographic information systems for geoscientists: modeling with GIS. Pergamon, New York, 398 p.

Chung, C.F. and Fabbri, A.G., 1993, The representation of geoscience information for data integration. Nonrenewable Resources, 2, 122-139.

Chung, C.F. and Fabbri, A.G., 1999, Probabilistic prediction models for landslide hazard mapping. Photogrammetric Engineering & Remote Sensing, 65, 1389-1399.

Heckerman, D., 1986, Probabilistic interpretations for MYCIN's certainty factors. In: Kanal, L.N. and Lemmer, J.F. (eds.), Uncertainty in artificial intelligence. Elsevier, New York, 167-196.

Lee, K. and Kwon, B.-D., 1995, Application of artificial neural networks to spatial integration of raster-based geological data. Journal of the Korean Earth Science Society, 16, 358-367.

Moon, W.M., 1990, Integration of geophysical and geological data using evidential belief function. IEEE Transactions on Geoscience and Remote Sensing, 28, 711-720.

Park, N.-W., Chi, K.-H., Chung, C.F., and Kwon, B.-D., 2003, Application of spatial data integration based on the likelihood ratio function and Bayesian rule for landslide hazard mapping. Journal of the Korean Earth Science Society, 24, 428-439.

Shortliffe, E.H. and Buchanan, B.G., 1975, A model of inexact reasoning in medicine. Mathematical Biosciences, 23, 351-379.

Manuscript received, November 15, 2003
Revised manuscript received, January 12, 2004
Manuscript accepted, January 30, 2004