On Knowledge Based Approach of Integrating Geophysical and Geological Data Sets

P. An and W. M. Moon
Geophysics, Department of Geological Sciences
The University of Manitoba, Winnipeg, Canada R3T 2N2
Tel. (204)-474-8395, Fax. (204)-261-7581
geop04@ccu.umanitoba.ca

Abstract

An object-oriented and map-based prototype expert system has been developed for integrating geophysical, geological data for base metal exploration. The object-oriented knowledge representation structure and uncertainty propagation mechanisms used work well for this integrated resource exploration problem. Evidential belief function theory is utilized to manage the uncertainties in the system. It appears a more natural and adequate theoretical basis for representing spatially unbalanced geophysical and geological information. The problem of dependent information can be dealt with in a knowledge-based system of this type by explicitly introducing important uncertainties and by organizing the relation network properly. The prototype system has been tested using real exploration data sets.

Introduction

An integrated approach is one of the most distinguishing characteristics of today's non-renewable resource exploration. With the introduction of new efficient data acquisition techniques and subsequent rapid accumulation of exploration data, the number of data sets available are often large in many exploration areas. The tasks of integrated interpretation are also becoming more complicated with rapidly increasing volume of data. Use of the conventional intuitive approach is becoming less efficient and less effective. A number of statistical and mathematical approaches have recently been developed. The Bayesian, regression, and weights of evidence models (Bonham-Carter et al, 1988, 1989) derive prediction maps based on known mineral occurrences. These new approaches provide us with more objectively generated prediction maps based on data. New approaches are needed because most of the exploration areas are underexplored and there are too few known mineral occurrences. Evidential belief function (Moon, 1990) and fuzzy set (An et al, 1991) approaches can be used in these cases. The difficulty in representing exploration data into an evidential space or into a fuzzy space has impeded their widespread application. An automated expert system can be a useful tool to overcome this difficulty.

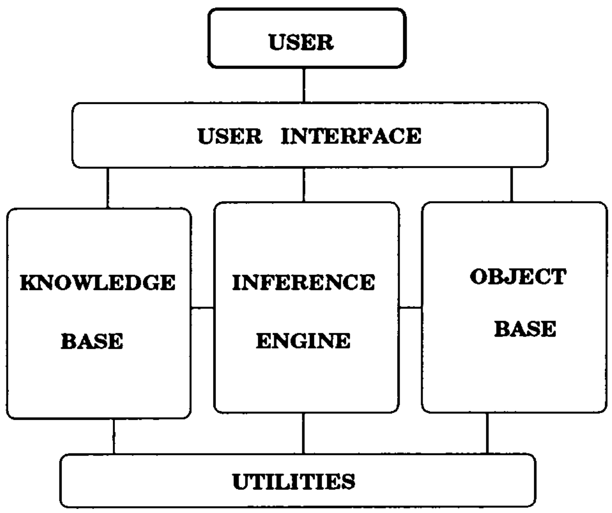

Figure 1: Block Diagram of the System

The current prototype system consists of a knowledge base, an inference engine and an object base (Figure 1). A set of utilities has been developed to define and provide the necessary programming environment because a relatively low level programming language C^{++} is used to build the system. It currently has 23 nodes (objects) and about 150 rules. The study is focused on the knowledge representation structure, its corresponding inference mechanisms and object simulation of the data and related concepts, so that the user interface is very simple.

Integrated Mineral Exploration

Non-renewable resource exploration activities can be divided into two categories in the set-theoretic representation of exploration data integration (An, 1992). The first includes tasks of acquiring and improving signal/noice ratio of data sets. The task in this category can be divided into many sub-tasks such as airborne EM data acquisition, surfacial geological mapping, seismic data processing and etc. There is often no interaction between different subtasks. Classical set theory is essentially the backbone of data management in this category. The second category includes interpretation of the data sets to find a certain exploration target. Although the data sets can be interpreted individually, the final decision has to be based on all the data sets. The tasks in this category are highly multidisciplinary and yet less properly defined than those in the first.

The tasks in the second category can not be represented adequately in the framework of classical set theory. Most of the exploration data are intrinsically imprecise although they are established in a framework of the classical set theory. During the interpretation of a data set or data sets, one can rarely simply answer "yes" or "no" to a proposition or exploration target before it has been observed derictly or drilled, even though all the data sets are of a good quality. At this moment, one can only say "to some degree yes" or "to some degree no". Usually no clear cut solution or answer is available. This situation reflects the fact that the information we gather can not provide us with sharp or absolute evidences. Representation of imprecise information and inabsolute knowledge, uncertainty processing and inference mechanisms thus have fundamental importance for expert system application. Clearly, the expert system applications in this category are more difficult than those in the first.

The tasks in the second category are less adequately defined and less well investigated. The whole integrated interpretation can be regarded as a specific mapping from input to output. The input includes different data sets which are relevant to the exploration targets. The output is a set of spatially distributed exploration targets. Unfortunately, there is generally no physical and/or mathematical models for such a mapping which takes data sets from different survey as input and outputs expected exploration targets. In this case, conventional computation which rely on mathematical models can not work very effectively. Intelligent approach can be efficient and effective. The consequent intelligent answer can be very useful in isolating the potential exploration targets.

Object-Oriented Knowledge Representation

Many different geological disciplines are relatively independent of each other. In any exploration strategy, data sets are collected independently and first interpreted for different parameters. One can often interpret one data set without knowing how to interpret other data sets. For example, one can interpret an airborne magnetic data set while he/she does not have to know how to interpret a gravity data set. The rules that govern processing and interpretation of one specific data set usually are not applicable for interpretation of another data set. Further more, certain data sets should be interpreted separately in order to keep the results of interpretation independent for the subsequent processing. If a data set is interpreted in reference to another data set, result of the interpretation will most likely depend on the referred data set and its interpretation. This may cause serious difficulty in subsequent processing. Equally important are the propositions (concepts) and intermediate results which are necessary in a later integrated analysis. The propositions receive evidences from data sets or other propositions. These evidences are then processed and the results are kept as new evidences for other propositions.

The interpretations can then be combined for a specific exploration target. The way in which the evidences are combined is determined by the relationship relevant to the exploration target among the data and propositions. An object-oriented knowledge representation structure for integrated exploration can be designed based on the characteristics described above. Intelligent objects can be built to handle the data sets and propositions, and a relation network can be established to represent the relationships among the objects and to control the combination of evidences from the objects.

The Relation Network

The basic idea of representing relations between objects as a network is formulated from the recent theories on problem reduction (Slagle and Gini, 1987) and frame theory (Maida, 1987). Even though the network (Figure 2) appears very similar to an AND/OR tree, a special case of the AND/OR graph, the ways of interpreting relations between the nodes are quite different from that of an AND/OR graph (Pearl, 1987). But one can still see considerable similarity of problem reduction reflected in the network. This network is specifically designed to satisfy requirements of the integrated exploration problem and associated inferencing steps.

The arc between two nodes indicates an evidential relationship and it connects lower level nodes to higher level node. The lower level nodes provide evidence to a higher level node. The higher level node represent a possible source of lower level nodes if the lower level node does not represent a possible source of noise. The AND and OR relationships are used for the cases where more than one lower level nodes are connected to a higher level node. The OR relationship represents a case where either one of the lower level nodes can provide evidence for the higher level node. The AND relationship means that only all lower level nodes together can provide evidence to a higher level node. There is only one such case in the current system.

The Intelligent Objects

An intelligent object referred in this research represents a relatively independent block of a computer program which has all the fundamental capabilities of an object (storage, processing, and communication). This object can process a data set or propositions intelligently. According to their relative locations in a network, the objects can be refered as terminal objects (nodes), middle objects, or the top object.

The terminal objects are those which have no child node. A terminal node can accept a data set and acquires necessary information about the data set from users. Each data set must have a corresponding terminal object so that it can be integrated by the system. The primary function of a terminal node includes receiving of messages from its upper level node, searching for a corresponding data set, acquiring pertinant information about the data, interpretation of the data and then, submission of a belief function about the proposition represented by its upper node. The complete procedure can either be very simple or fairly complex as long as a reasonable belief function can be provided to its upper proposition. The middle nodes are those which have both upper and lower level nodes. The functional requirement for a middle node includes receiving of a message from upper nodes, sending of messages to all its lower nodes, combination of belief functions from all lower level nodes, and then, calculation and submission of the belief function about the proposition represented by its upper node. Clearly, the middle nodes correspond to the propositions (concepts) or intermediate results. The top object is the one which has no higher node. It corresponds to the exploration target in this study. There is only one target "base metal deposit" in the present system.

Each terminal node or each middle node has a set of rules which is incorporated into the corresponding object. Rules built in the terminal node control the interpretation and processing of the data, and the rules of each middle node govern propagation of belief functions through it. Incorporating the rules into the corresponding objects instead of using a global rule base improves efficiency of computation. The interpreter of an object only need to search the rules in its own rule base. Another advantage includes easier development and updating of the rule base.

Uncertainty Management

It is important to correctly interpret certainty factors estimated by an expert system. The results are less useful if they are difficult to interpret. If the output of a map-based system is a map, it will show different degrees of spatial favorabilities towards a specific exploration target. These maps show spatial distributions of specific information for the prospecting area. Even when certain pixel values appear to be meaningless, the overall distribution usually provides a relative favorability distribution towards the chosen exploration target.

In most cases of non-renewable resource exploration, the spatial data coverages are incomplete, and data efficiency and effectiveness are spatially unbalanced. Some data sets only cover a very small part of the exploration area. In this situation, it is desireable to distinquish between lack of information and information providing negative evidences. Dempster-Shafer evidential belief function appears to provide a more suitable theoretic basis and it is used in this research to manage the uncertainties associated with the rules and the data sets.

The rules in an object have a format

If *premise*

Then *conclusion* (Bel_r),

where Bel_r is a belief function. The propositions of Bel_r are R_1, R_2, and R_3. R_1 presents "The rule is true"; $R_2 = \overline{R}_1$; and $R_3 = \{R_1, R_2\}$. Their basic probability numbers are $m(R_1)$ which is lower probability or degree to which the rule is true, $m(R_2)$ which is the probability of the negation of R_1 or degree to which the rule is false, and $m(R_3)$ which represents the uncertainty or degree to which one does not know whether the rule is true or false. Only two independent numbers, however, are required because a belief function must satisfy $m(R_1)+m(R_2)+m(R_3) = 1$. An example of a rule in an object is shown below.

If there exists *mafic-intermidiate-volcanic* rocks,

then there is a *favorable geological* condition for a base metal deposit (0.5 0.1 0.4).

The Inference

The inference engine is divided into two parts corresponding to the knowledge representation structure. The first part is a relatively simple interpreter and can be incorporated into the objects. When it receives a message or is activated, it searches applicable rules in its sub-rulebase and apply the rules to carry out the tasks which may include processing, interpretation, and evidence combination. During this process it communicates with other objects. Results of the tasks are belief functions to be submitted to its upper node. The second part involves systematically searching over the relational network backward to find applicable objects and activate them, and then, chaining forward to propagate the belief functions to the top node.

The Search Algorithm

The search task in an AI system is to find a solution path from an initial node to the goal node (Winston, 1990). The algorithm being developed here searches systematically through the relation network (Figure 2) and finds all existing evidences relevant to the top proposition, activates pertinat objects, and propagates the belief functions to the top proposition. The systematic search method used in this system is depth-first and it is given as follow:

1 . Put the top node on a list, called OPEN, of unvisited nodes. If the top node has no lower node, exit with failure.

2 . If OPEN is empty, exit successfully.

3 . If the last node, n, of OPEN is a terminal node or has no unvisited successor, do the following:

(1) activate n;

(2) set n visited;

(3) move n from OPEN.

4 . If n has unvisited successor, get a successor of n and place it to the last of OPEN.

5 . Go to step 2.

The Inference Mechanisms

The inference mechanisms are referred to as the ways in which belief functions are propagated through the network. In this research, Dempster's rule of combination (Dempster, 1967; Shafer, 1976) is used as the basis for propagating belief functions through the relation network. Three basic types of inference mechanisms which correspond to the knowledge representation structure are used in this research.

Suppose A_i $(i = 1,2,3)$ and B_j $(j = 1,2,3)$ are propositions of belief functions Bel_1 and Bel_2. A_1="The hypothesis is true from evidence E_1"; $A_2 = \overline{A}_1$; and $A_3 = \{A_1, A_2\}$. B_1= "The hypothesis is true from evidence E_2"; $B_2 = \overline{B}_1$; and $B_3 = \{B_1, B_2\}$. Bel_1 has probability measures $m_1(A_i)$ $(i = 1,2,3)$ and Bel_2 has probability measures $m_2(B_j)$ $(j = 1,2,3)$. Probability measures $m(H_l)$ of a new belief function Bel can be obtained by

$$m(H_l) = \frac{1}{1-k} \sum_{A_i \cap B_j = H_l} m_1(A_i) m_2(B_j)$$

where

$$k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j) < 1.0.$$

An OR operator is given by defining H_1="The hypothesis is true from E_1 or E_2", $H_2 = \overline{H}_1$, and $H_3 = \{H_1, H_2\}$. An AND operator is obtained by defining H_1="The hypothesis is true from both E_1 and E_2", $H_2 = \overline{H}_1$, and $H_3 = \{H_1, H_2\}$. In the case that a rule is involved and represented by Bel_r, an operator, PASS-RULE can be obtained by defining H_1="The conclusion of the rule is true from E_1", $H_2 = \overline{H}_1$, and $H_3 = \{H_1, H_2\}$. These three operators are utilized to propagate belief function through a relation network (An, 1992).

Problem with Dependent Evidence

Dependent evidence and concept management have always been problematic in information integration and uncertainty processing. In a frame work of the Bayesian probability theory, assumptions such as conditional independence (Chung and Moon, 1990) have to be made when the theory is applied to real world problems. In certain situations, assumptions such as conditional independence or local independence can lead to unacceptable conclusions (Konolige, 1979).

Since the belief function theory includes Bayesian models as a special case (Shafer, 1976). pieces of evidences which are independent in belief function framework are not necessarily statistically independent in a Bayesian probability model. Evidential independence (Moon, 1990) is used in this research to differentiate the concept of independence in a belief function system from the concept of statistical independence in the Bayesian probability models.

Belief function theory emphasizes the use of information as evidence for a given hypothesis or proposition. The fact that pieces of evidences are evidentially independent or not relies often on the hypothesis to be proven. For example, let us consider a case where two airborne magnetic surveys were carried out over the same location at different times. If the two measurements are used to prove a hypothesis that there is a base metal deposit at a specific location, they are evidentially dependent. If the same measurements are used to prove a hypothesis that there is an airborne magnetic anomaly at the same location, they can be interpreted as evidentially independent. This means that evidential independence is relative to a frame of discernment. Using this property of the knowledge-based approach, a knowledge representation structure can be organized in a way in which the lower level nodes are evidentially independent relative to their higher level nodes. Actually, this method is related to the fundamental concepts of reframing (Shafer, 1984) and partitioning (Shafer and Logan, 1985).

Shafer (1984) remarked that independence is always relative to the frames of discernment. By explicitly introducing certain important common uncertainties into the frame of discernment, remaining uncertainties may be treated as independent with respect to that frame of discernment. In another paper, Shafer and Logan (1985) introduced an idea of partitioning a frame of discernment. Such a partition can itself be regarded as a frame of discernment. By partitioning the frame of discernment, complexity of Dempster's rule can be greatly reduced. These ideas are employed in the knowledge representation structure and in construction of new belief functions. If a complete target-oriented exploration process can be represented as a frame of discernment, it includes all the evidences (data sets), propositions, and rules. In this frame of discernment, many propositions are brought explicitly into the frame, such as "there is a high conductivity", "there is a high magnetic anomaly", and etc. The rules in the frame connect evidences and propositions together and functionally control the transform of belief functions from evidence to a proposition and from a proposition to another proposition. If a frame only has one proposition that represents a specific exploration target, and if it has only OR relation, it becomes the same as the case described in Moon (1990) and Moon et al. (1991).

Test of the Prototype System

The prototype system is tested using data sets from Farley Lake area, Manitoba, Canada. The geophysical and geological data sets used in the test are the same as those used for testing an application study using the fuzzy set method (An et al., 1991). A total of 10 data sets, most of which have only a partial coverage, are used in this test.

The final outputs from the system are belief function maps showing different degree of favorability towards the chosen exploration target. Figure 3 shows a spatial distribution of support (lower probability) for the exploration target, "a base metal deposit". The higher level of support is found in the middle west of the test area. Figure 4 shows the spatial distribution of uncertainty. The high level of uncertainties are found in the areas where there is less data coverage and/or the data collected using a specific survey method are less efficient for the exploration target. Low uncertainties are found in the west central area where there is better data coverage and/or more efficient data sets.

Discussion and Conclusion

Use of evidential belief functions in knowledge representation provides a theoretical basis for a spatial reasoning system. It also allows uncertain information and knowledge to be represented in a more natural manner. Object-oriented knowledge representation partitions a complicated task of integrated resource exploration into sub-tasks which are more manageable by using artificial intelligence techniques. The dependent evidences can be in general better dealt with by introducing important uncertainties explicitly and by organizing the relation network properly. The test example successfully outlines the favorable exploration target areas. The knowledge representation structure described above appears to be a promising form of knowledge representation for integrated resource exploration. More comparative tests are needed.

References

An, P., 1992. Spatial reasoning technique and integration of geophysical and geological information for resource exploration. Ph.D. thesis, The University of Manitoba, Canada. Submitted.

An, P., Moon, W. M., and Rencz, A., 1991. Integration of geological, geophysical, and remote sensing data using fuzzy set theory, Canadian Journal of Exploration Geophysics, Vol. 27, No. 1., pp. 1-11.

Bonham-Carter, G. F., Agterberg, F. P., and Wright, D. F., 1989. Weights of evidenced modelling: a new approach to mapping mineral potential. Statistical Applications in the Earth Sciences, in Agterberg, F. P. and Bonham-Carter, G. F. (eds), Geological Survey of Canada, Paper 89-9, pp. 171-183.

Bonham-Carter, G. F., Agterberg, F. P., and Wright, D. F., 1988. Integration of geological database for gold exploration in Nova Ascotia. Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 11, pp. 1585-1592.

Chung, C. F., and Moon, W. M., 1990. Combination rules of spatial geoscience data for mineral exploration. Extended abstract, ISME-AI'90 (Tokyo), pp. 131-141.

Dempster, A. P., 1967. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics, Vol. 38, pp. 325-339.

Konolige K., 1979. Bayesian methods for updating probabilities. In Duda, R. O. et al. (1979), *A Computer-based Consultant for Mineral Exploration* (Final Report, SRI Project 6415). Menlo Park, CA: SRI International, Artificial Intelligence Center.

Maida, A. S., 1987. Frame theory. In Shapiro, S. C. et al, eds., Encyclopedia of artificial intelligence. Wiley-Interscience Publication.

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

Moon, W. M., Chung, C. F., and An, P., 1991. Representation and integration of geological, geophysical and remote sensing data. Geoinformatics. Vol. 2, No. 2, pp. 177-188.

Pearl, J., 1987. AND/OR graph. In Shapiro, S. C. et al, eds., Encyclopedia of artificial intelligence. Wiley-Interscience Publication.

Shafer, G., 1976. A mathematical theory of evidence. Princeton, N.J.: Princeton University Press.

Shafer, G., 1984. The problem of dependent evidence. working paper No. 164, School of Business, University of Kansas, U.S.A.

Shafer, G., and Logan, R., 1985. Implementing Dempster's rule for hierarchical evidence. working paper No. 174, School of Business, University of Kansas, U.S.A.

Slagle, J., and Gini, M., 1987. Problem reduction. In Shapiro, S. C., et al, eds., Encyclopedia of artificial intelligence. Wiley-Interscience Publication.

Waterman D. A., 1986. A guide to expert systems. Addison-Wesley Publishing Company.

Winston P. H., 1984. Artificial Intelligence (second edition). Addison-Wesley Publishing Company.

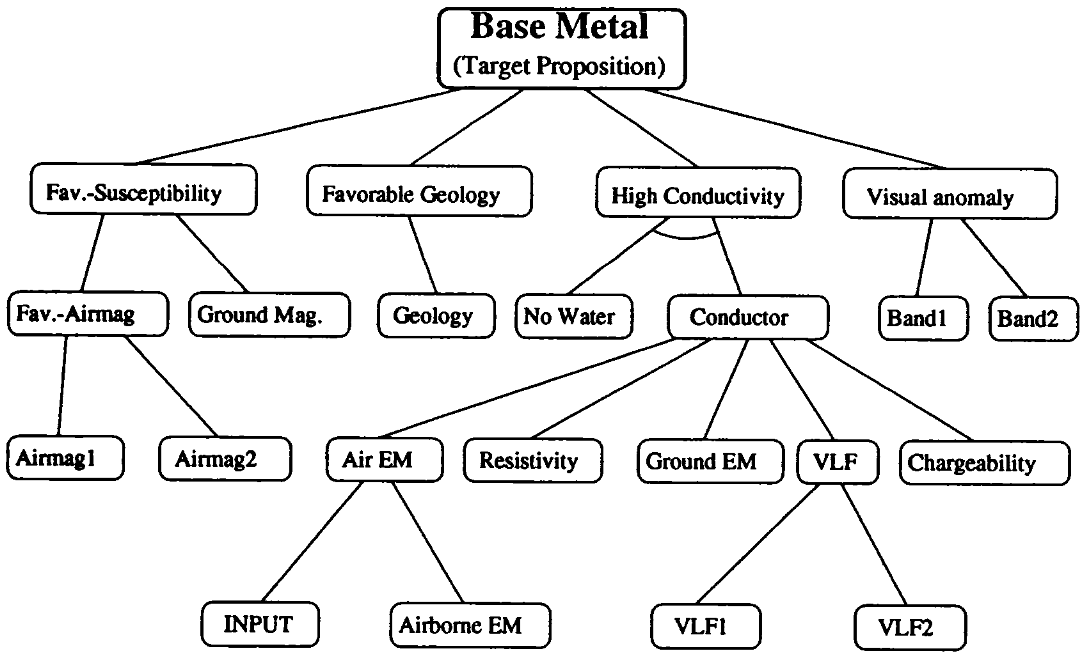

Figure 2. A Relation Network

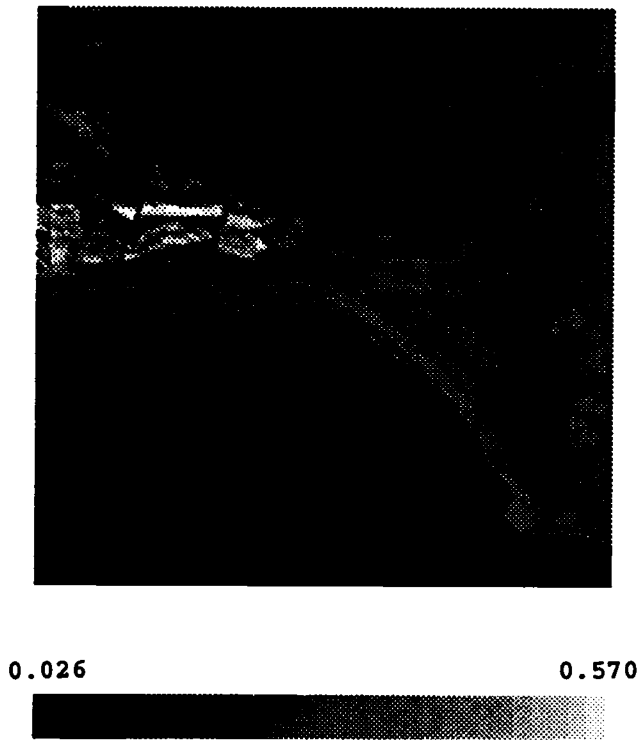

Figure 3. Support for Base Metal

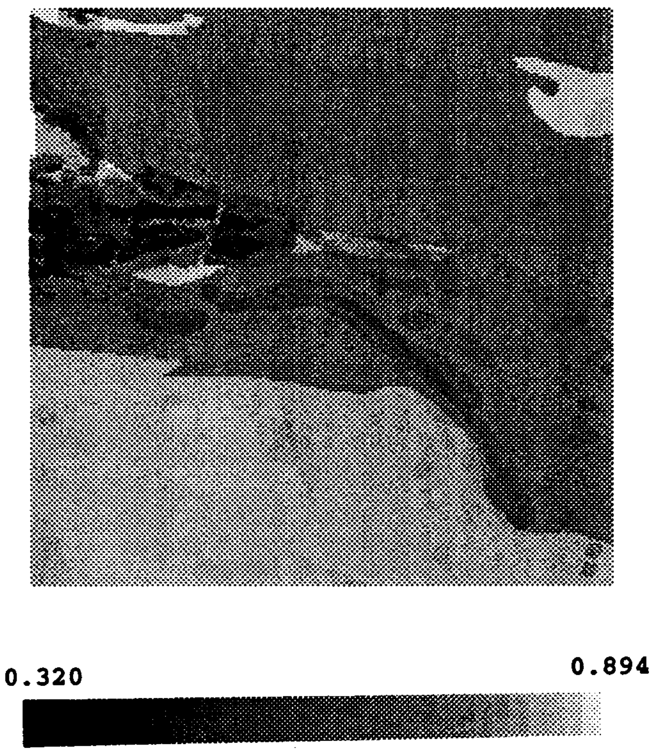

Figure 4. Uncertainty for Base Metal