NAG5-898
GODDARD
GRANT
IN-61-CR
177194
398

Cloud Classification from Satellite Data Using a Fuzzy Sets Algorithm: a Polar Example

J.R. Key, J.A. Maslanik, and R.G. Barry

Cooperative Institute for Research in Environmental Sciences
and Department of Geography
University of Colorado, Boulder
Boulder, Colorado
80309-0449

Abstract. Where spatial boundaries between phenomena are diffuse, classification methods which construct mutually exclusive clusters seem inappropriate. The Fuzzy c-means (FCM) algorithm assigns each observation to all clusters, with membership values as a function of distance to the cluster center. The FCM algorithm is applied to AVHRR data for the purpose of classifying polar clouds and surfaces. Careful analysis of the fuzzy sets can provide information on which spectral channels are best suited to the classification of particular features, and can help determine likely areas of misclassification. General agreement in the resulting classes and cloud fraction was found between the FCM algorithm, a manual classification, and an unsupervised maximum likelihood classifier.

(NASA-CR-184568) CLOUD CLASSIFICATION FROM SATELLITE DATA USING A FUZZY SETS ALGORITHM: A POLAR EXAMPLE Semiannual Progress Report (Colorado Univ.) 39 p CSCL 09B

N89-14708

Unclas
G3/61 0177194

Cloud Classification from Satellite Data Using a Fuzzy Sets Algorithm: a Polar Example

1. Introduction

Cloud detection and classification from satellite remote sensing data has received considerable attention in view of the significance of cloud cover for global climate. Various techniques are reported in the literature based on threshold, bispectral, 2-d or 3-d histograms, and split-window methods. Smith (1981) and Crane and Barry (1984) summarize these procedures. From a classification standpoint, most current approaches seek to designate mutually exclusive classes with well defined boundaries; these are termed "hard" classifications. Clustering algorithms used in such classifications are commonly based on either the Euclidean distance measure (e.g., Parikh 1977; Desbois et al. 1982) or the maximum likelihood classifier (e.g., Bolle 1985; Pairman and Kittler 1986; Ebert 1987). Areas where cloud identification is uncertain are usually treated by forcing them into existing classes, or leaving them unclassified.

Our particular interest in cloud conditions in polar regions indicates that this approach is especially undesirable where the spectral characteristics of the clouds and the underlying surface frequently overlap. Where cloud categories are poorly defined and the spatial boundaries between them are diffuse, it seems appropriate to represent this uncertainty in the taxonomic strategy.

The purpose of this study is to examine the applicability of the fuzzy sets approach to the classification of clouds from satellite data. In contrast to hard classifiers, the fuzzy sets approach assigns each observation to every class, with the strength of the membership being a function of the its similarity to the class mean. Fuzzy clustering was introduced by Ruspini (1969) and was later developed into the fuzzy c-means algorithm by Dunn (1974) and generalized by Bezdek (1975). Previous applications of the procedure to climatic data are limited to McBratney and Moore (1985) where the fuzzy c-means algorithm was applied to temperature and precipitation data, and Leung (1987) who took a linguistic approach to describing the imprecision of regional boundaries. There has been an increasing use of fuzzy set theory and fuzzy algorithms with digital images (e.g., Huntsberger et al. 1985, Pal and King 1983), but these procedures have not yet found their way into satellite data processing applications. We do not intend to present new information on cloud characteristics, but rather to provide an alternative method of dealing with the poorly defined boundaries of clouds and surfaces in satellite data.

2. Data

The AVHRR (Advanced Very High Resolution Radiometer) on board the NOAA-7 polar orbiting satellite is a scanning radiometer that senses in the visible, reflected infrared, and thermal (emitted) infrared portions of the electromagnetic spectrum with a nadir resolution of 1.1 km (IFOV of 1.4 milliradians) at a satellite

altitude of 833 km. Global Area Coverage (GAC) data provide a reduced-resolution product created through on-board satellite processing. GAC pixel resolution is used, with each pixel representing a 3 x 5 km field of view. Of the five channels available (0.58-0.68 μm, 0.725-1.00, 3.55-3.93, 10.3-11.30, 11.5-12.50) channels 1, 3 and 5 are employed here.

First-order calibration of the AVHRR GAC data was performed following the methods described in the NOAA Polar Orbiter Users Guide (NOAA 1984) and Lauritsen, et al. (1970). Channel 1 was converted to albedo and corrected for solar zenith angle; channels 3 and 5 were converted to radiance in mW m^{-2} sr^{-1} cm.

3. Example of Polar Clouds and Surfaces

Determination of the amount of cloud cover is the principal objective of cloud classification for the study of ice-atmosphere interactions in the polar regions. Secondarily, breakdown of the cloud cover into different types, e.g. stratus, cirrus, cumulus provides useful information on cloud radiative properties, availability of moisture, and source of the cloudiness. To determine the amount of cloud requires that the classifier discriminate between clouds and underlying surfaces of snow, ice, water, and land. Distinguishing between cloud type may require information on cloud height (estimated from cloud-top temperature) and cloud morphology (related to large-scale patterns or local texture).

The study area is shown in Figure 1 (channels 1, 3, and 5).

This is a 250x250 pixel or (1250 km)^2 area centered over Novaya Zemlya and the Kara and Barents Seas on July 1, 1984. Open water, snow-covered and snow-free land, sea ice (various concentrations), and high, middle, and low cloud over different surfaces are present in the image. For computational efficiency, means of 2x2 pixel cells were used in the classification process, reducing the number of pixels from 62,500 to 15,625. A manual interpretation of this area is given in Figure 2.

The problem of distinguishing discrete cloud and surface categories is illustrated by Figure 3, which shows scatter plots of visible vs. near-infrared and visible vs. thermal data for a (1250 km)^2 segment of the study area. Based on training area statistics, the spectral responses of four surface types (snow-free and snow-covered land, sea ice, and open water) and three general cloud categories (high, middle, and low) are identified in the plots by their mean plus and minus two standard deviations in each of the two channels. The principal sources of confusion are likely to occur between snow/ice and cloud due to their similar responses in AVHRR Channel 1 and, to a lesser extent, Channel 2. In the thermal channels, similarities exist between the physical temperatures of low or thin clouds, ocean, and melting sea ice. The data in Figure 1 present several examples of cloud of varying optical depth overlying different concentrations of sea ice. In addition, the surface conditions of the sea ice (as estimated by reflectance and passive microwave emissivity differences) are not constant throughout the image. It is clear that the spectral

properties of the clouds and ice are not likely to form compact and distinct clusters in multispectral space. Hard classifiers are required to force these indistinct areas into spectrally similar, but perhaps unsuitable, classes. Otherwise, large areas of the image will remain unclassified.

4. Classification using Fuzzy Sets

In the fuzzy sets approach, points do not belong to only one class but instead are given membership values for each of the classes being constructed. Membership values are between zero and one and all the membership values for a given point must sum to unity. Memberships close to one signify a high degree of similarity between the sample point and a cluster while memberships close to zero imply little similarity.

In this respect, memberships are similar to probabilities. However, no assumption of distribution type is made in fuzzy c-means (FCM) clustering, and calculations of memberships are not based on probability density functions. Therefore, this methodology bears little theoretical relationship to probability-based techniques such as maximum likelihood which assumes multivariate normal distributions, or discriminant analysis which is based on the general linear model.

The fuzzy c-means algorithm is neither a "lumper" (conjunctive or clustering procedure), which operates by combining small clusters into larger clusters, or a "splitter" (disjunctive or divisive classification procedure) which begins with all pixels

belonging to the same class then subdividing. Instead, in the FCM algorithm all pixels begin and end with memberships in each of the specified number of clusters; each iteration adjusts these memberships to minimize an error function.

A brief explanation of the FCM procedure is provided below; for a more complete description, see Bezdek (1981) and Kandel (1982). Following Bezdek et al. (1984) and McBratney and Moore (1985), the fuzzy c-partition space is

$$M = \{U:u_{ik} \in [0,1];\ \sum_{k=1}^{n} u_{ik}>0,\ i=1..c;\ \sum_{i=1}^{c} u_{ik}=1,\ k=1..n\}$$

where U is a fuzzy c-partition of a sample of n observations and c clusters. Each element of U, u_{ik}, represents the membership of a particular observation x_k in the ith fuzzy group. Each x_k is a vector of length p where p is the number of features (e.g. spectral channels, texture measures, etc.). These membership coefficients are values between 0 and 1 and for each observation sum to one. Also, the sum of the membership values for each cluster is greater than zero, otherwise the group does not exist.

Optimal fuzzy c-partitions may be identified with the generalized least-squared errors functional

$$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m (d_{ik})^2$$

where U is the fuzzy c-partition of the data, x_k, which is a c by n matrix with elements u_{ik}; V is a c by p matrix where each element v_{ik} represents the mean of the kth of p attributes in the ith of c

groups; n is the number of observations; m is a weighting component, $1<m<\infty$, which controls the degree of fuzziness; d_{ik} is the distance between each observation x_k and a fuzzy centroid v_i, a measure of dissimilarity given as

$$(d_{ik})^2 = (x_k - v_i)^T A(x_k - v_i)$$

where A is the inner product norm metric, discussed below. An optimal fuzzy c-partition is obtained when J_m is minimized. This is achieved by the Fuzzy c-Means algorithm, which is given in the appendix.

4.1 FCM Parameters

A number of options are available in the FCM algorithm so that the results may be tailored to the problem at hand. These are the weighting exponent, initial matrix, A-norm, and computational considerations.

Weighting exponent. According to Bezdek et al. (1984), no computational or theoretical evidence distinguishes an optimal weighting exponent. The range of useful values seems to be [1, 30] while for most data, $1.5 \leq m \leq 3.0$ gives good results. In choosing values for m, it is important to remember that as m approaches unity the partitions become increasingly hard and as m approaches infinity the optimal membership for each data point approaches 1/c. Therefore increasing m tends to increase "fuzziness".

McBratney and Moore (1985), applied the fuzzy c-means method

to temperature and precipitation data from stations in Australia. They tested a range of values for m and found that m=100 yielded memberships almost constant at 0.5 for each of two classes indicating that clustering was so fuzzy that no clusters would be distinguished. They also attempted to identify optimal combinations of c, the number of classes, and m by plotting the change in the error functional, J_m, with m for each number of clusters, c. In general, J_m decreases with increasing c and m, but its rate of change with changing m is not constant. Their work showed that, at least empirically, m of approximately 2 is optimal, though for a large number of groups m should be less than for a smaller number of groups to obtain similar balance between structure and continuity.

Initial matrix. The initial U matrix also provides a number of options: a random start, a random nonfuzzy start, or an almost uniform start. Alternatively, the results from another clustering method can be used as the initial matrix. In the random start, each membership coefficient is given a random value between zero and one. The random nonfuzzy start assigns a membership coefficient of one to a randomly chosen class and zero to the remaining sets. An almost uniform start is obtained by setting each membership to 1/c plus or minus a small random component. The algorithm presented by Bezdek et al. (1984) employs a random start, while McBratney and Moore (1985) found that an almost uniform start yielded faster convergence and similar results from different runs.

Starting with results from another cluster procedure has not previously been tested; in our experiments the number of iterations needed for convergence was usually reduced by 10-20%.

It is suggested by Bezdek that the FCM be run for several different starting membership matrices since the iteration method used, like all descent methods, is susceptible to local stagnations. If different starting matrices result in different final memberships, further analysis should be made.

A-norm. A detailed discussion of the geometric and statistical implications of the choices of the A-norm is given in Bezdek (1981). Three of these norms, Euclidean, diagonal, and Mahalanobis, are of interest in FCM. When the Euclidean norm is used, J_m identifies hyperspherical clusters, but for any other norm, the clusters are essentially hyperellipsoidal. A Euclidean metric can be used for uncorrelated variables on the same scale, a diagonal metric for uncorrelated variables on different scales, and Mahalonobis' for correlated variables on the same or different scales.

Computational considerations. The fuzzy sets program was not originally designed for application to very large data sets such as satellite images. The number of computations necessary is a function of the number of data items (pixels), the number of features, and the number of clusters. The number of data items being processed at any one time can be reduced by using a random

sample of the entire image, hopefully obtaining a representative subset. Clustering local areas of the image with the ultimate goal of global description is another possibility.

No alternative method of calculating cluster centers or updating the membership matrix is evident. However, an alternative method of error calculation - which controls termination of the algorithm - is to compare elements of each cluster center matrix from two successive iterations rather than comparing successive membership matrices. The cluster center matrix is of dimensions c by c rather than n by c for the membership matrix. If n is much larger than c, the savings in CPU time are significant. Additionally, computer memory would be reduced by approximately 40%. If this method is chosen, however, data should be on the same scale - either originally or standardized to a zero mean and unit variance - so that cluster centers can be compared with the same error criteria. Of course, relaxing the convergence criterion (maximum allowable error; see Appendix, step 4) will reduce the number of iterations required. If the channels employed are statistically independent, then the number of computations may be further reduced by eliminating those involving the A-norm metric, which for uncorrelated variables on the same scale would be the identity matrix.

Without these modifications for image processing, computational resources are certainly not trivial, as the execution of the FCM algorithm on a 125 x 125 pixel area requires approximately one hour of CPU time on a DEC VAX 8550 and up to ten

hours on a DEC MicroVAX configured for "typical" user loads. Adjustment of system parameters such as working set size can significantly reduce disk paging, which will in turn reduce total CPU time. With adjustments for large images, computation time can be reduced by a factor of ten.

4.2 Validity functionals

It is possible to obtain data sets where the error functional is globally minimal but where the resulting classes are visually unappealing. To aid in the resolution of this problem, two validity functionals are used to evaluate the effect of varying the number of clusters: the partition coefficient, F, and the entropy, H:

$$F = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^2/n$$

and

$$H = - \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik} \log_{\alpha} u_{ik})/n, \qquad 0 < \alpha < \infty$$

F will take values between 1/c and one, while H has a range of zero to $\log_{\alpha}c$. When F is unity or H is zero, clustering is hard, while an F value of 1/c or an H value of $\log_{\alpha}c$ implies that memberships are approximately 1/c. A plot of F or H by the number of groups may be examined for local maxima of F or minima of H, which will give some indication of optimal c.

5. Results

The FCM program was applied to the study area in Figure 1 as represented by AVHRR channels 1, 3, and 5. A variety of fuzziness index values were tested as well as a range in the number of clusters. The partition coefficient, F, and entropy, H, for each run is listed in Table 1. Run #5 represents an essentially hard classification (m=1.25) where F is large and H is small. Conversely, the fuzziness index of 2.6 in run #2 resulted in a small F and large H. Run #6 produced the least visually appealing and least realistic results of all runs. This is supported statistically by the minimal F and maximal H. Figure 4 illustrates the change in F and H for a varying number of clusters. For these tests, m=2.0. A local maximum for F and minimum for H occur at c=6, with c=10 also being acceptable.

A visual examination of the results from each of these tests revealed that the 10-cluster solution best identified the cloud and surface types present in the scene, therefore an interpretation of this solution is given. Figures 5a-5j (hereafter Sets A-J) illustrate each of the ten classes where grey level represents membership of each pixel in the class, lighter grey shades indicating larger membership values. The most distinct classifications are shown by the bright areas (high probabilities) assigned primarily to land in Set C, sea ice under clear skies (Set H), and open water (Set I). The varying gradation of cloud conditions are represented in several of the other sets. Sets E

and G describe high stratocumulus, Sets A, D, and J show high memberships for lower stratus. Thin stratus over ice is represented in Set F. Large memberships in Set B occur for thin cloud over water, but also for mixed pixels at land, cloud, and ice edges. Areas that are not distinctly classified in a single set appear as intermediate gray-tones in several of the sets in Fig. 6. In particular, the ice cap on Novaya Zemlya is confused with thin cloud over ice (Set F), and thicker, higher clouds in Sets A, D, E, G, and J. These are areas that - at least for this particular algorithm - require additional information to be distinguished from other classes.

The distribution of memberships between the fuzzy sets described above presents a convenient graphical tool for interpreting the physical properties of clouds and surfaces, and thus the potential sources of confusion in multispectral classifiers. For example, the similarity between clouds and the Novaya Zemlya ice cap in several of the fuzzy sets is apparently due to similar albedos and temperatures yielding similar responses in AVHRR channels 1 and 5. Interestingly, the ice cap has the largest membership in Set F, with memberships similar to the thin cloud over ice in the upper-left portion of the image. A physical interpretation of the memberships in Set F suggests that the combination of thin cloud with an underlying, high-albedo surface yields a combined spectral return with physical temperature and albedo similar to the Novaya Zemlya ice cap under clear skies.

If desired, a hard classification can be obtained from the

fuzzy sets results where, for each pixel, the largest membership value is replaced with a membership of one, while membership values for the other classes are set to zero. In this manner, the same basic classes will result, but the fuzziness is eliminated.

5.1 Statistical Properties

The previous discussion pointing out the ability of the fuzzy sets to combine multispectral information into individual probability sets is suggestive of artificial orthogonal features created through principal components analysis. The fuzzy sets are, however, simpler to interpret in physical terms since their development is not restricted by the objective of creating uncorrelated components and maximizing the amount of variance accounted for by each component. No attempt is made to include as much information as possible in the first few sets created. Unlike principal components, the information content of each successive fuzzy set does not necessarily decrease. In fact, Sets H and I represent two of the most spectrally-distinct classes in the AVHRR data.

These differences between the fuzzy sets classifier and principal components is demonstrated by examining the cross-correlations between the individual probability sets. The maximum correlation (37%) occurs between Set A and Set J. Sets H and I are not positively correlated with any of the other sets. Sets A and J both predominantly represent slightly different conditions of stratus cloud. The lack of a requirement that the

two sets be uncorrelated allows the gradation of cloud height and thickness to be clearly represented in these two sets. On the other hand, the ability of the fuzzy sets classifier to identify basically uncorrelated classes such as open water and sea ice is demonstrated in Sets H and I.

Application of principal components analysis with fuzzy sets as variables and individual pixels as observations allows us to identify similarities among the sets more quantitatively. Using unrotated components, eight components are required to account for 94% of the variance present in the sets, while the first five components describe 69% of the variance. The large number of components required to represent the information content of the fuzzy sets confirms that each set provides a considerable amount of unique information. Comparison of the factor loadings in each set suggests that Principal Component 1 discriminates between different conditions of stratus cloud and open water (high loadings for Sets A and J, and negative loading for Set I. A similar type of interpretation can be made for Component 2, which appears to represent high cloud, with the greatest positive loadings for Fuzzy Sets E and G). With the exception of Components 1 and 2, no loadings exceed 50%. The relationships between the fuzzy sets as variables is perhaps slightly masked by the potential confusion of unique and common variances inherent in principal components. However, the component-derived groupings agree well with the correlations in the cross-correlation matrix. As a final confirmation of the uniqueness of each fuzzy set, a Varimax

rotation was applied to the principal components. Results of the rotation approach the desired ideal of simple structure, with a loading of nearly 1.0 on one set per component, again suggesting that large correlations are not found between groups of fuzzy sets.

5.2 Supervised Approach

A supervised approach may be taken if class means are known. In this case, the algorithm may be modified to simply calculate the memberships for each pixel in each of the known classes. The memberships are still a function of the weighted distance to the class means, but the class means are no longer determined by the algorithm. These are instead supplied in a manner analagous to using training sites to provide spectral statistics for a supervised classification. This approach is very fast (30-40 times faster than unsupervised) as it requires only one pass through the data.

We have found that class means must be very carefully selected, and that some experimentation may be necessary to reach a realistic solution. For example, Run #7 in Table 1 was a supervised classification where a seven-cluster solution was specified and class means were provided for snow-covered and snow-free land, sea ice, open water, high cloud, middle cloud, and low cloud in AVHRR channels 1, 3, and 5. Snow-covered land did not uniquely define any fuzzy set, but was instead grouped with low cloud because of similar albedos and brightness temperatures. While this problem may be solved by adjusting the class means,

perhaps a better solution would be to add a weighting function to the algorithm so that features which better define a particular class will be more influential in the calculation of membership coefficients.

5.3 Maximum Likelihood Classification

To provide a source of comparison to the fuzzy sets approach, the data shown in Figure 1 were classified using an unsupervised maximum likelihood (ML) procedure. The results are shown in Figure 6. The unsupervised clustering approach (with all image pixels taking part in the definition of spectral signatures) yielded 21 clusters, with four clusters accounting for 67% of the area. None of the remaining 17 clusters represented more than one percent of the image. Sixteen percent of the scene remained unclassified, and an additional 12% of the image pixels fell in more than one cluster. Misclassifications are noted for indistinct classes, specifically low concentration ice (grouped with low clouds), optically thin clouds, and for boundary pixels between different classes.

The restrictive effects of the hard classifier vs. fuzzy sets are apparent in the large number of unclassified pixels. Most high and middle cloud layers were left unclassified, as was the ice cap on Novaya Zemlya. For indistinct classes common in polar cloud analyses, the fuzzy sets approach avoids errors of commission and ommission that occur when such indistinct values are forced into the nearest class in spectral space.

6. Discussion

Sets A, B, D, E, F, G, and J of the fuzzy sets classification each represent a separate cloud class, although other surface/cloud mixtures sometimes had large membership values in these classes. A map of cloud classes constructed from the maximum membership value for each pixel is shown in Figure 7. These results generally agree with the manual classification in Figure 2 and the maximum likelihood classification shown in Figure 6. Discrepancies occur with middle and high clouds (unclassified in the ML method), and with cumulus which, in the ML procedure, is grouped with an optically thin stratus deck over sea ice.

While there were some obvious differences in number of cloud classes and the cloud types that each class represented in the three methods, the total cloud amount computed for each procedure was similar. For the manual and ML classifications, cloud fraction is simply the proportion of cloud pixels in the image. In the ML results, this was computed for only those areas labeled as cloud in Figure 6, and again with the unclassified areas included. For the fuzzy sets results, two methods of computing cloud fraction were examined. In both cases, the membership values of each pixel in each of the cloud classes were summed. This may be considered an estimate of a pixel's "cloudiness". Then, in the first case, for each threshold from 0.4 to 0.9 in increments of 0.1, a pixel was considered cloud-filled if its cloudiness exceeded the threshold. Cloud fraction was expressed as the proportion of

cloudy pixels in the image. In the second case, the pixels were considered partially cloud-filled, with cloud fraction being the sum of all cloudiness values that exceeded the threshold, as a proportion of the total number of pixels in the image. Cloud fractions computed for the manual classification, ML method, and the fuzzy sets are given in Table 2. Best agreement between the methods occurs when the threshold is high (0.7) if pixels are considered completely cloud-filled, or in the midrange (0.4-0.6) is pixels are treated as partially cloud-filled.

7. Conclusion

The fuzzy sets method of classification was successfully adapted to the analysis of multispectral satellite imagery. The ability of the fuzzy sets approach to address indistinct spectral classes by calculating class memberships as opposed to the "in-or-out" decision required of hard classifiers is particularly well suited to the range of albedos and physical temperatures encountered in the analysis of ice and cloud conditions in the polar regions.

Application of the fuzzy sets classifier to an AVHRR image containing sea ice and cloud of varying condition and opacity yielded ten membership sets containing contextually and statistically unique information. Interpretation of intensities in images of these sets demonstrates the ability of the fuzzy sets to describe well-defined classes (such as open water and land) as well as classes that fall in intermediate spectral space (e.g.,

ice cap, thin stratus over water, or sea ice of varying concentration). Identification of such fuzzy areas in taxonomic space provides information on where data in additional spectral regions are required for accurate classification. Future work will use the fuzzy sets approach as a tool to help "tune" hard classifiers such as unsupervised clustering and bispectral threshold methods for cloud and ice mapping in the polar regions.

Acknowledgements

This work was supported under NASA grant NAG-5-898 and a DOD University Research Instrumentation Program grant N00014-85-C-0039. Thanks are due to W. Rossow and E. Raschke for providing AVHRR GAC data.

Appendix

Following Bezdek et al. (1984), the Fuzzy c-Means (FCM) algorithm is:

(1) Fix: c, $2<c<n-1$
m, $1<m<\infty$; the larger the m, the fuzzier the solution; many practitioners use m=2.
A, the inner product norm metric for R^p, where p is the number of attributes
U^o, the initial fuzzy c-partition
ϵ, the value for the stopping criterion (ϵ=0.01 gives reasonable convergence.)

Repeat until convergence (step 4):

(2) Calculate the c fuzzy group centroids, v_i

$$v_i = \sum_{k=1}^{n} (u_{ik})^m x_k \Big/ \sum_{k=1}^{n} (u_{ik})^m \quad \text{for all i}$$

(3) Update U^l using

$$u_{ik} = 1/\left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}\right]$$

which may be rewritten in the more computationally efficient form:

$$u_{ik} = [1/d_{ik}^2)^{1(m-1)}][1/\sum_{j=1}^{c} (1/d_{ik}^2)^{1/(m-1)}]$$

The measure of dissimilarity, d_{ik}^2, is given as

$$(d_{ik})^2 = (x_k - v_i)^T A(x_k - v_i)$$

where A is the inner product norm metric.

(4) Compare U^{l+1} to U^l. If the difference between all corresponding elements is less than or equal to ϵ, then stop. Otherwise, set $U^l = U^{l+1}$ and return to step (2).

References

Bezdek, J.C., 1981, Pattern Recognition with Fuzzy Objective Function algorithms. (New York: Plenum Press)

Bezdek, J.C., Ehrlick, R., and Full, W., 1984, FCM: the fuzzy c-means clustering algorithm. **Computers & Geosciences**, 10, 191-203.

Bezdek, J.C., 1975, Mathematical models for systematics and taxonomy. In Proc. Eighth Int. Conf. on Numerical Taxonomy, G. Estabrook (ed.), Freeman, San Francisco, pp. 143-164.

Bolle, H.J., 1985, Assessment of thin cirrus and low cloud over snow by means ofthe maximum likelihood method. **Advances in Space Research**, 5(6), 169-175.

Crane, R.G. and Barry, R.G., 1984, The influence of clouds on climate with a focus on high latitude interactions. **Journal of Climatology**, 4, 71-93.

Desbois, M., Seze, G., and Szejwach, G., 1982, Automatic classification of clouds on METEOSAT imagery: application to high level clouds. **Journal of Applied Meteorology**, 21, 401-412.

Dunn, J.C., 1974, A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters. **Journal of Cybernetics**, 3, 22-57.

Ebert, Elizabeth, 1987, A pattern recognition technique for distinguishing surface and cloud types in the polar regions. **Journal of Climate and Applied Meteorology**, 26, 1412-1427.

Huntsberger, T.L., Jacobs, C.L., and Cannon, R.L., 1985, Iterative fuzzy image segmentation. **Pattern Recognition**, 18 (2), 131-138.

Kandel, A., 1982, Fuzzy Techniques in Pattern Recognition. (New York: Wiley)

Lauritson, L., Nelson, G.G., and Port, R.W., 1979, Data extraction and calibration of TIROS-N/NOAA A-G radiometer. **NOAA Tech. Memor., NESS 107**, Natl. Oceanic and Atmos. Admin., Boulder, Colorado.

Leung, Yee, 1987, On the imprecision of boundaries. **Geographical Analysis**, 19 (2), 125-151.

McBratney, A.B. and Moore, A.W., 1985, Application of fuzzy sets to climatic classification. **Agricultural and Forest**

Meteorology, 35, 165-185.

NOAA, 1984, NOAA polar orbiter data user's guide. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, NESDIS, February, 1984.

Pairman, D. and Kittler, J., 1986, Clustering algorithms for use with images of clouds. **International Journal of Remote Sensing**, 7(7), 855-866.

Pal, S.K. and King, R.A., 1983, On edge detection of X-ray images using fuzzy sets. **IEEE Transactions**, PAMI-5, 69-77.

Parikh, J.A., 1977, A comparative study of cloud classification techniques. **Remote Sensing of Environment**, 6, 67-81.

Ruspini, E.H., 1969, A new approach to clustering. **Information Control**, 15, 22-32.

Smith, E.A., 1981, Review of cloud climatologies. In Clouds and Climate; Modeling and Satellite Observational Studies, NASA, Goddard Institute for Space Studies (New York), 113.

Table 1

Results of FCM tests for varying m and c.
Scaling norm is diagonal in all cases.

Run #	Number of Clusters	Fuzziness Index, m	Partition Coefficient[1]	Entropy[1]	Comments
1	8	2.00	0.46	1.23	
2	8	2.60	0.25	1.73	
3	7	2.30	0.37	1.40	
4	10	1.25	0.90	0.19	Hard
5	8	1.80	0.12	2.08	Poor
6	10	2.00	0.50	1.20	Study run
7	7	2.00	0.53	0.98	Supervised

1. See text.

Table 2

Cloud fraction computed for three classification methods.

Manual Interp.	ML[1]	ML[2]	Fuzzy Sets (Threshold:) 0.4	0.5	0.6	0.7	0.8	0.9
0.53	0.40	0.55	0.69	0.65	0.62	0.57	0.46	0.19[3]
			0.56	0.54	0.52	0.49	0.40	0.18[4]

1. Classes labeled as cloud only.
2. Including unclassified areas as cloud.
3. Pixels treated as completely cloud-filled.
4. Pixels partially cloud-filled.

Figure 1. Study area centered on Novaya Zemlya (approximately 75°N, 60°E) and containing the Kara and Barents Seas. The area covers $(1250\ km)^2$. AVHRR channels 1, 3, and 5.

Figure 2. Manual interpretation of the area shown in Figure 1. Cloud classes: LCLI - low cloud over sea ice; LCLW - low cloud over water; MCL - middle cloud; HCL - high cloud, Cu - cumulus.

Figure 3. Bispectral plots of AVHRR data for the arctic. Class means ± two standard deviations are shown as rectangles.
a) visible vs. near-infrared; b) visible vs. thermal.

Figure 4. Plot of the partition coefficient, F (solid line), and entropy, H (broken line), as a function of the number of classes. In all cases, m=2.0.

Figure 5. Ten classes produced by the FCM algorithm from the the study area data. See text for interpretation of classes.

Figure 6. Study area as classified by a unsupervised maximum likelihood procedure. Cloud classes are as defined for Figure 2. Additional class codes: U - unclassified, M - mixed classes, low cloud is defined by two classes: LCL1 and LCL2.

Figure 7. Fuzzy sets classification of the study area. The class to which a pixel belongs is the one with the largest membership value. Cloud classes: LCLI - low cloud over sea ice; LCLW - low cloud over water; MCL - middle cloud; HCL - high cloud, Cu - cumulus.

*** NOTE TO REVIEWERS ***

Due to the high cost of producing prints, only one set of Figures 1 and 5 has been sent to the editor. These have been photocopied to the best of our ability and included here, but some grey levels have been lost. We feel that these are adequate for conveying the point of "partial memberships" in each class, but we ask that you keep in mind the unavoidable reduction in the quality of these copies.

ORIGINAL PAGE IS
OF POOR QUALITY

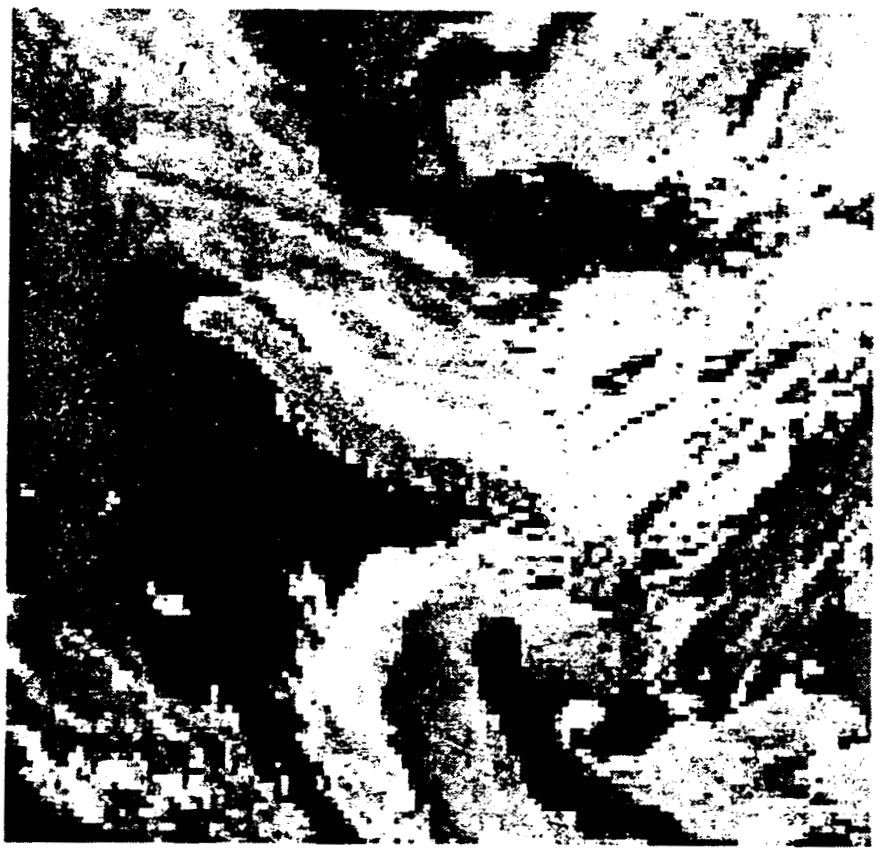


Fig. 1

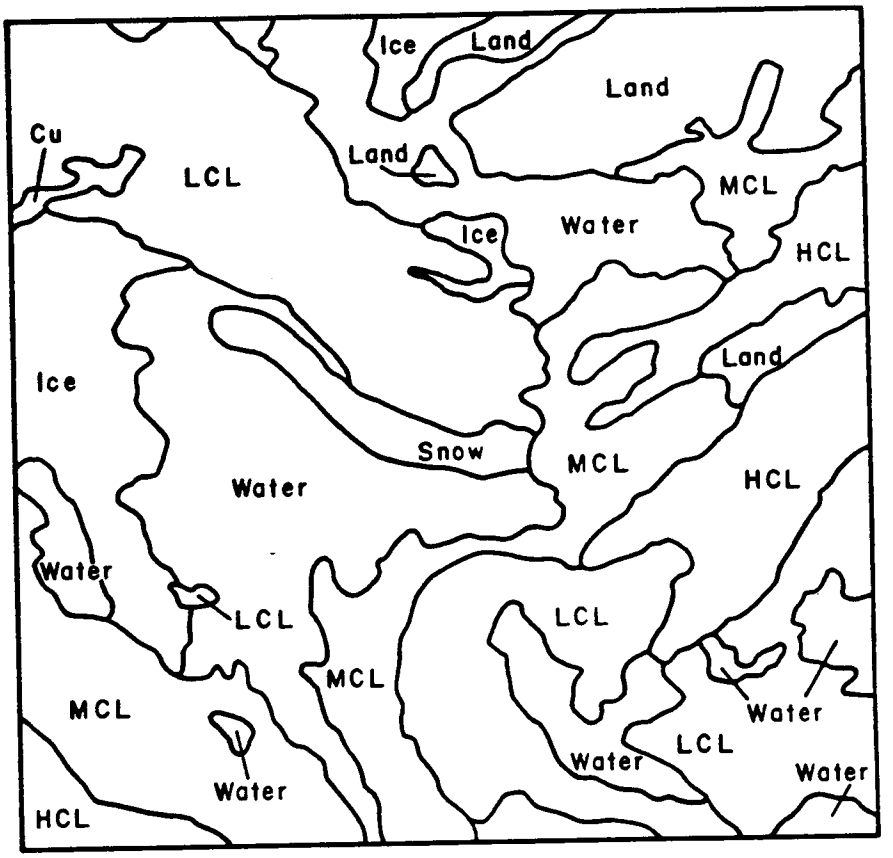

Ice
Land
Land
Cu
LCL
Land
MCL
Ice
Water
HCL
Land
Ice
Snow
MCL
HCL
Water
Water
LCL
LCL
MCL
MCL
Water
Water
LCL
Water
Water
HCL

Fig. 2

AVHRR Channels 1 and 3

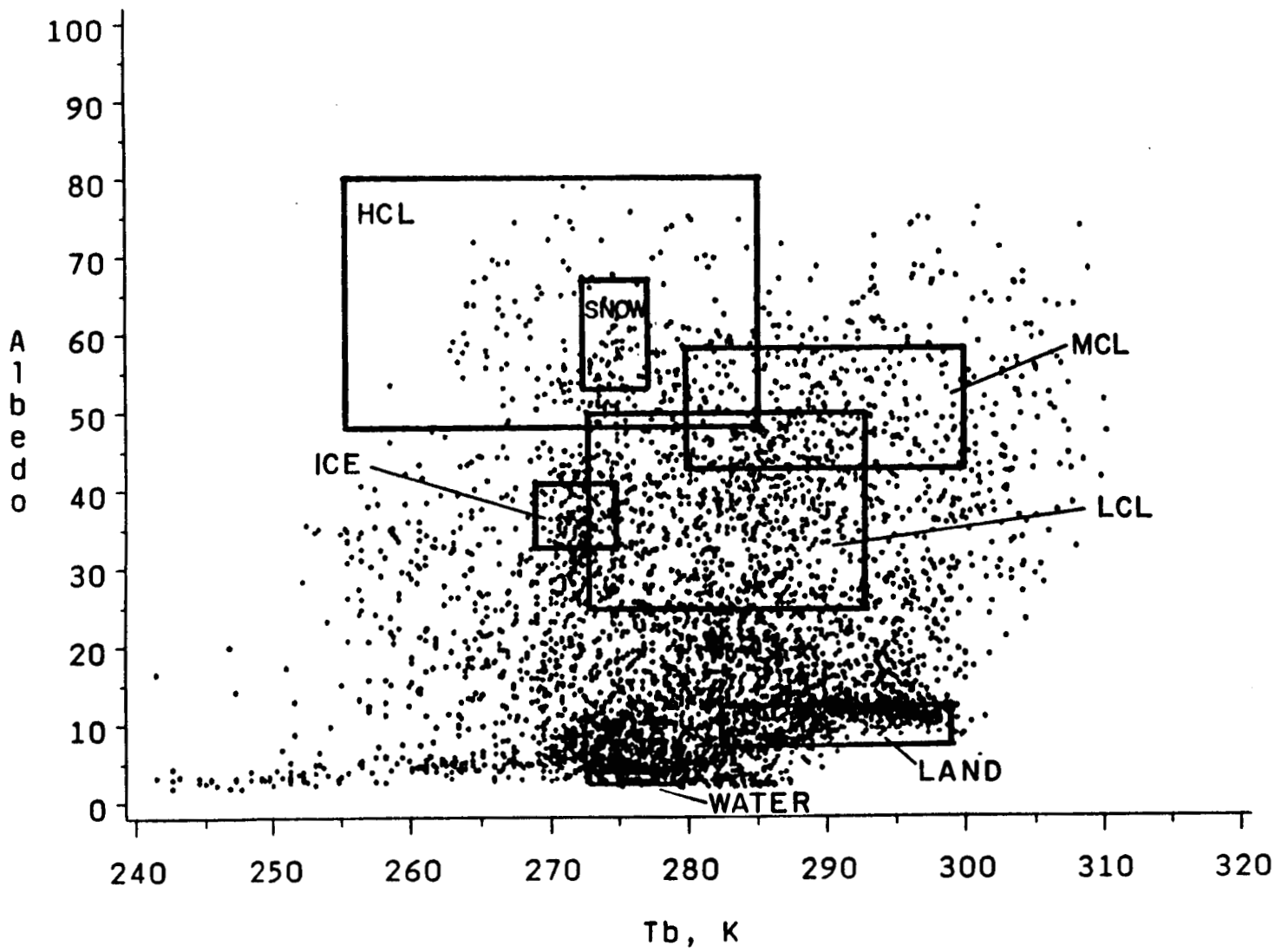


Fig. 3

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

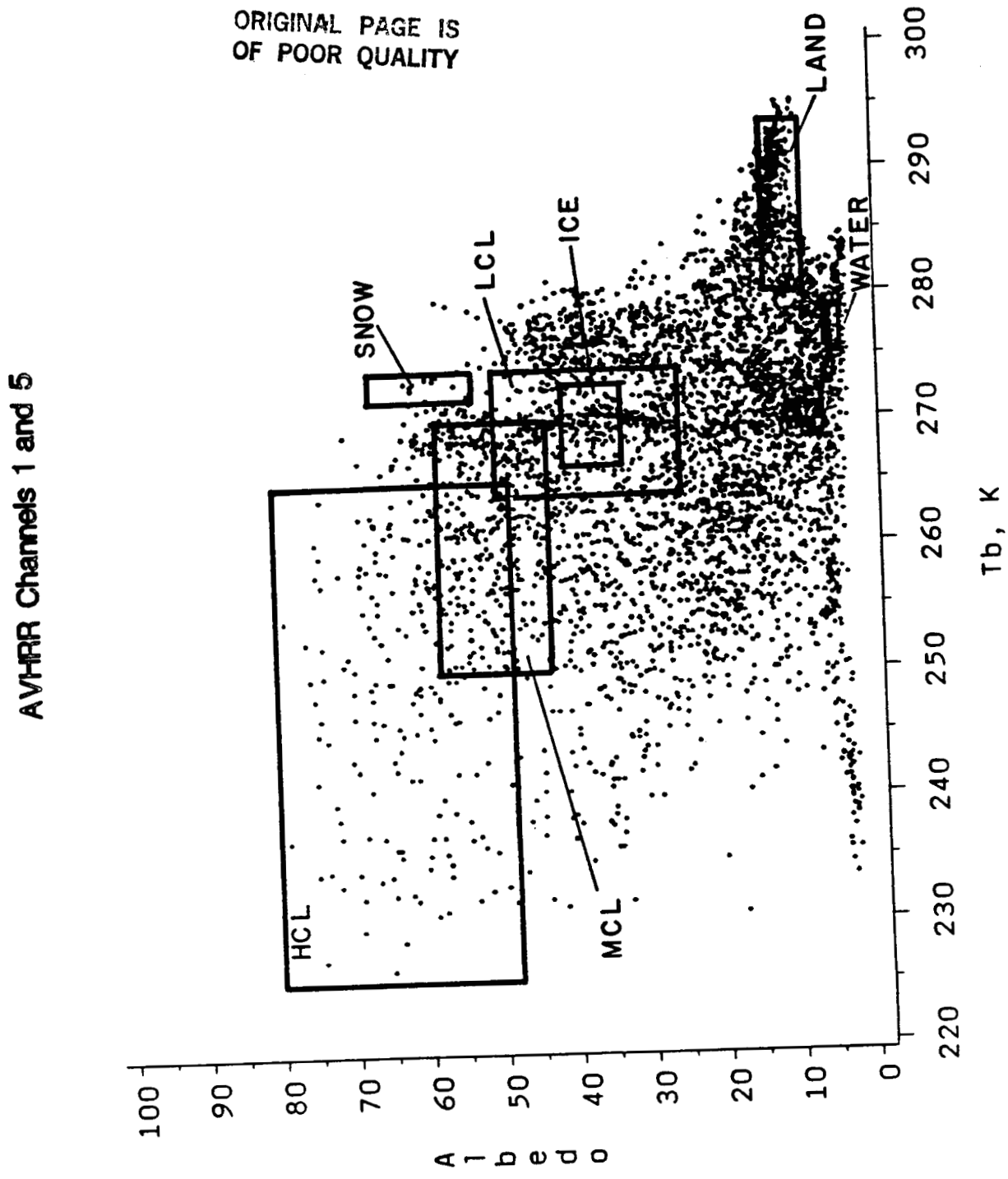
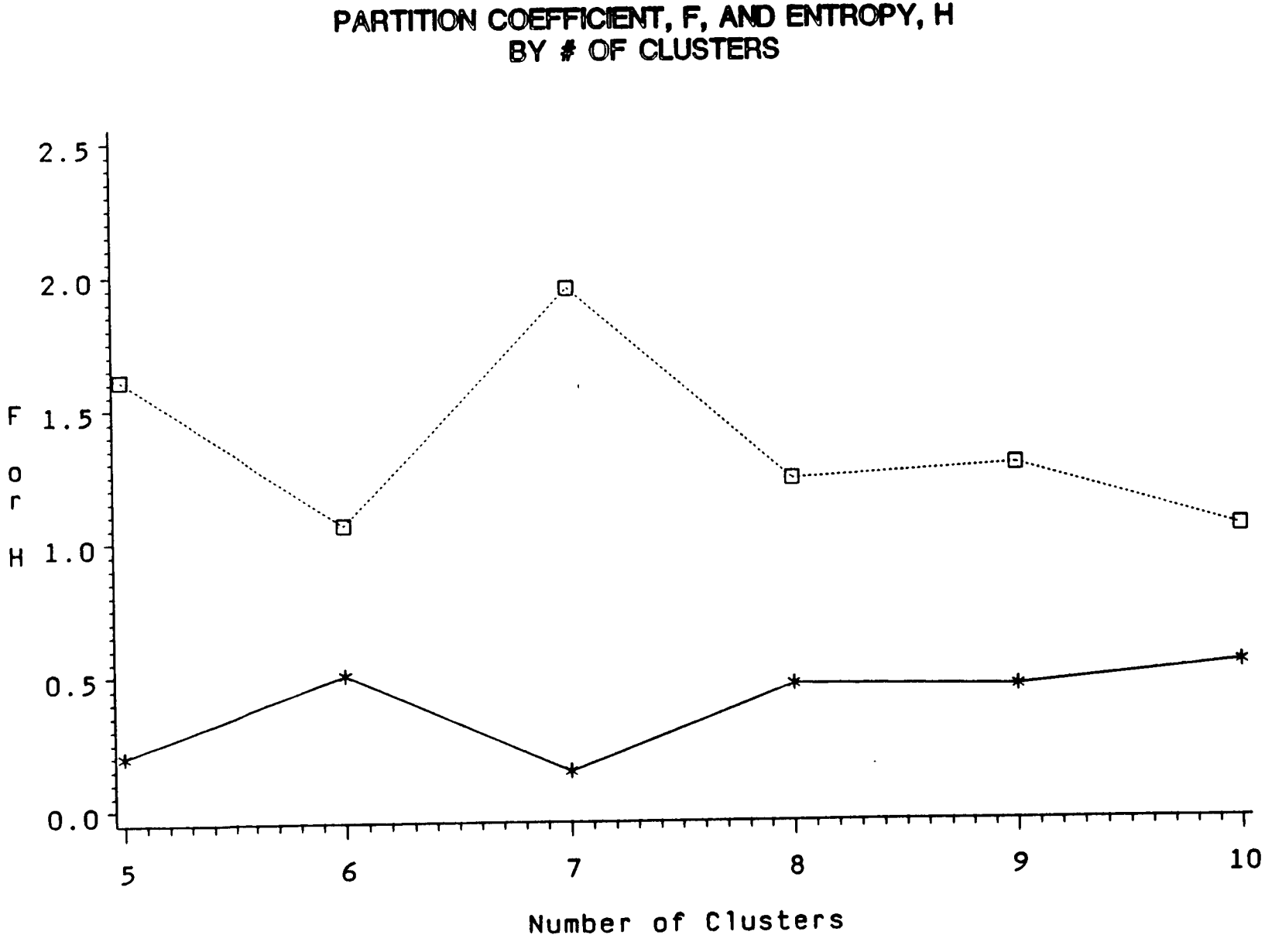
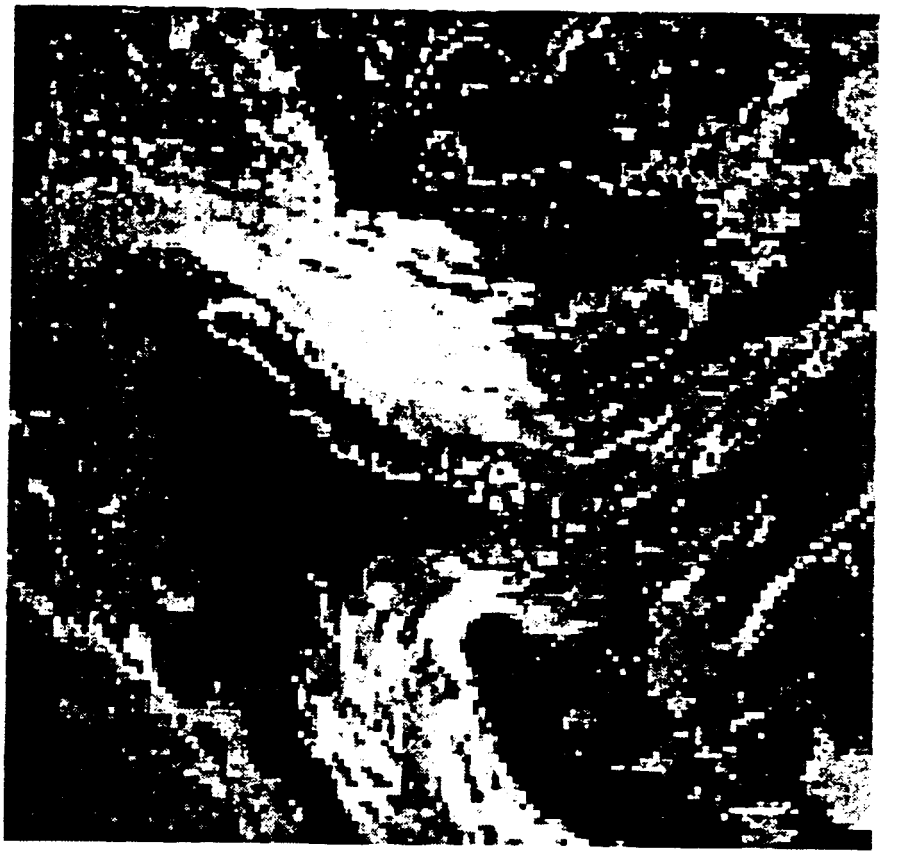


Fig. 3, cont.

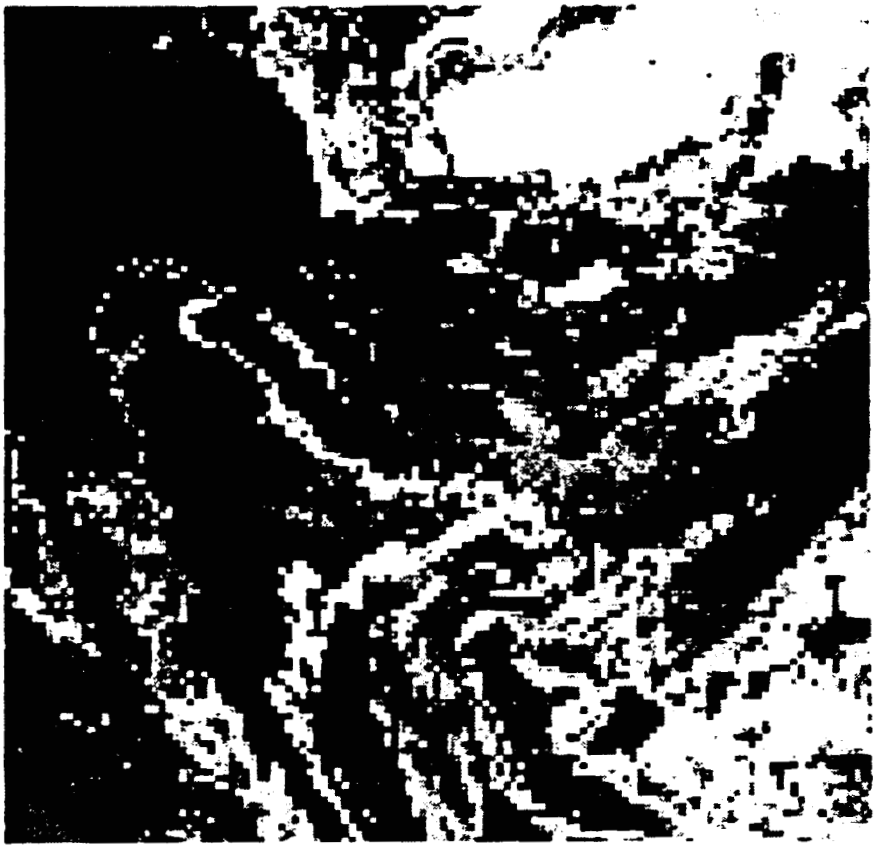
Fig. 4


PARTITION COEFFICIENT, F, AND ENTROPY, H
BY # OF CLUSTERS
2.5
2.0
1.5
1.0
0.5
0.0
F
or
H
5
6
7
8
9
10
Number of Clusters



ORIGINAL PAGE IS
OF POOR QUALITY

Fig. 5



ORIGINAL PAGE IS
OF POOR QUALITY

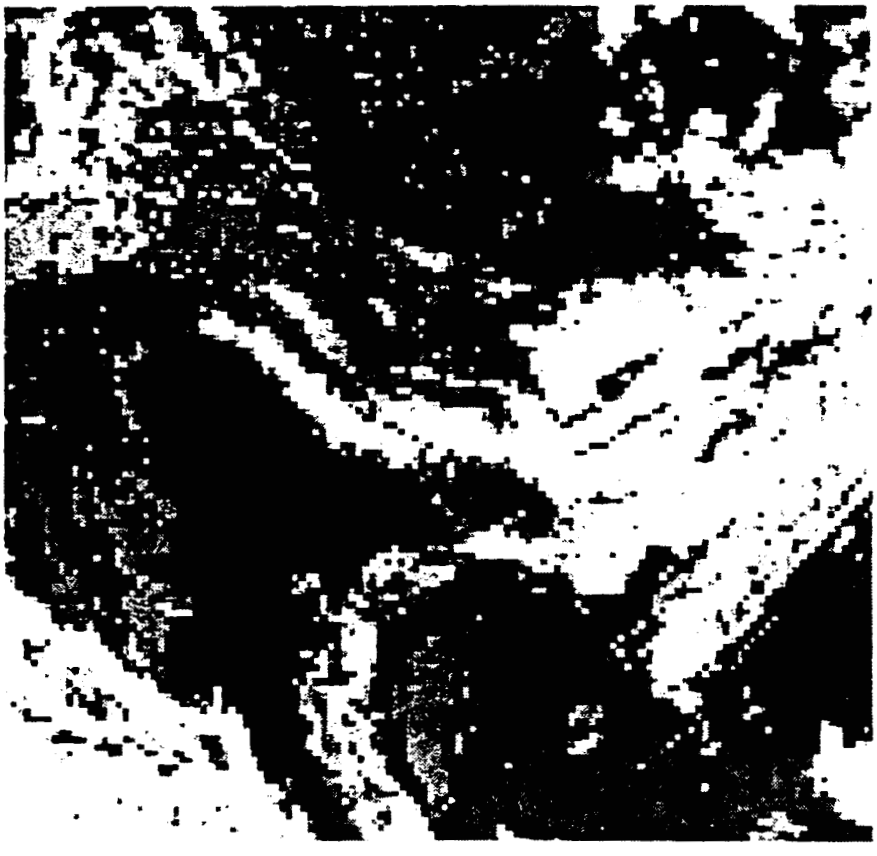
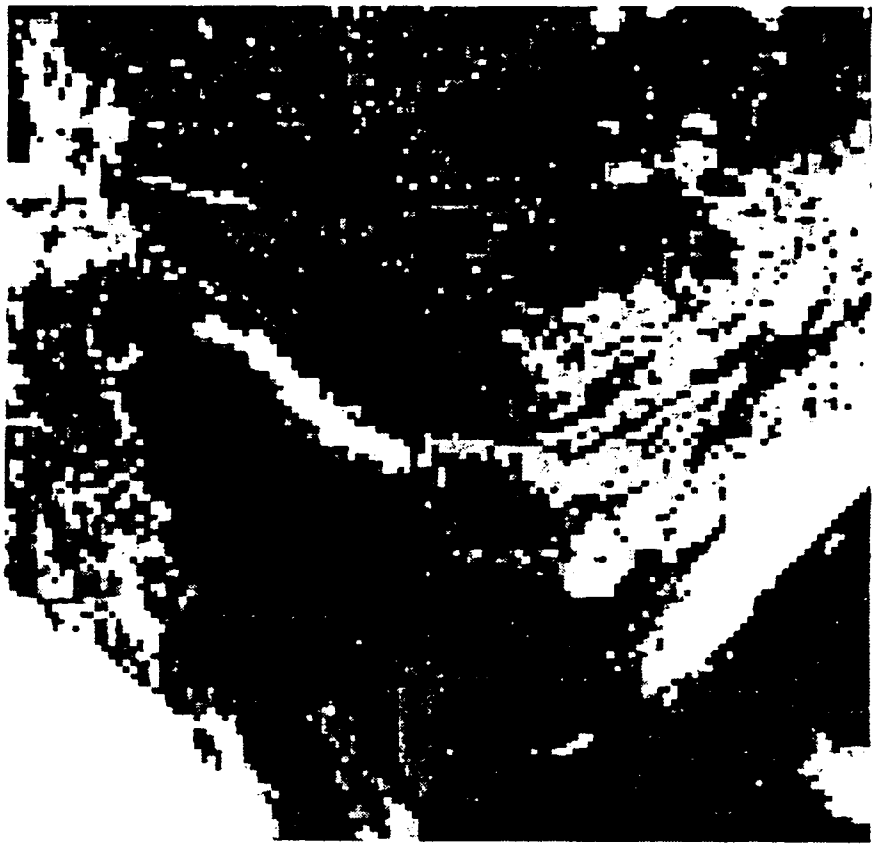


Fig 5, Conv.



ORIGINAL PAGE IS
OF POOR QUALITY

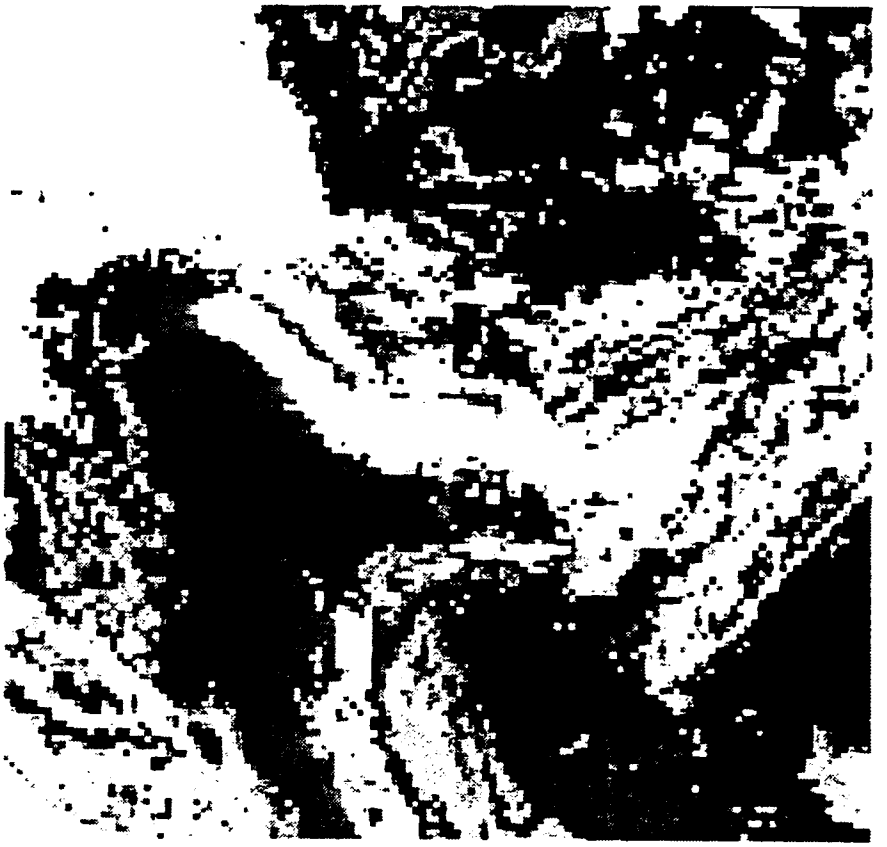
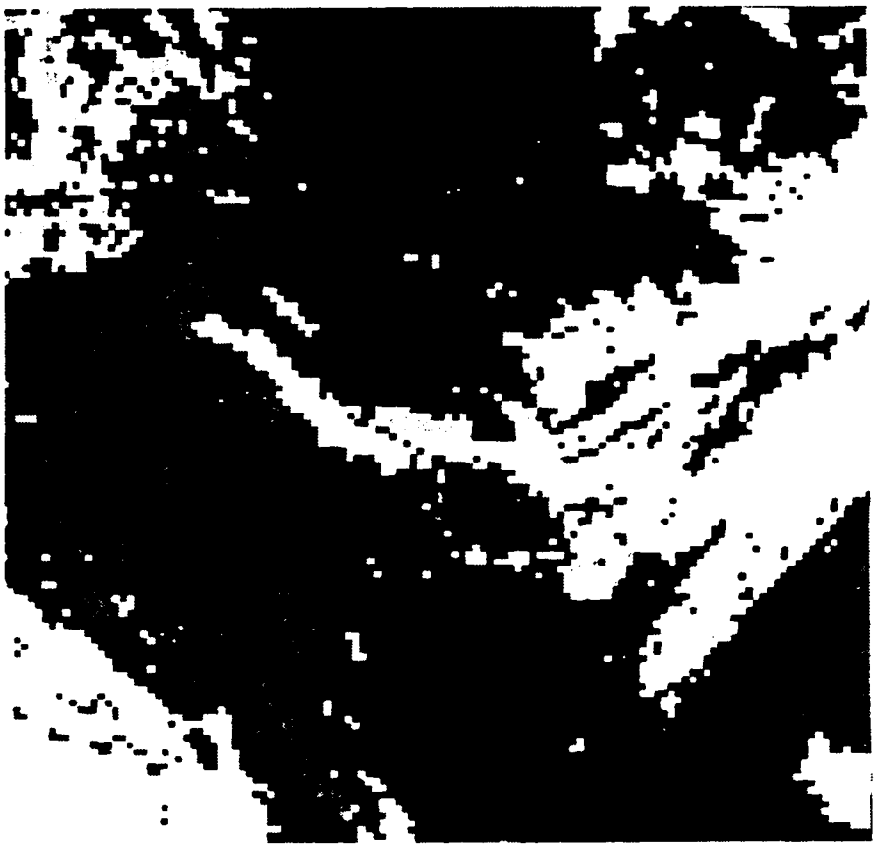


Fig. 5, cont.



ORIGINAL PAGE IS
OF POOR QUALITY

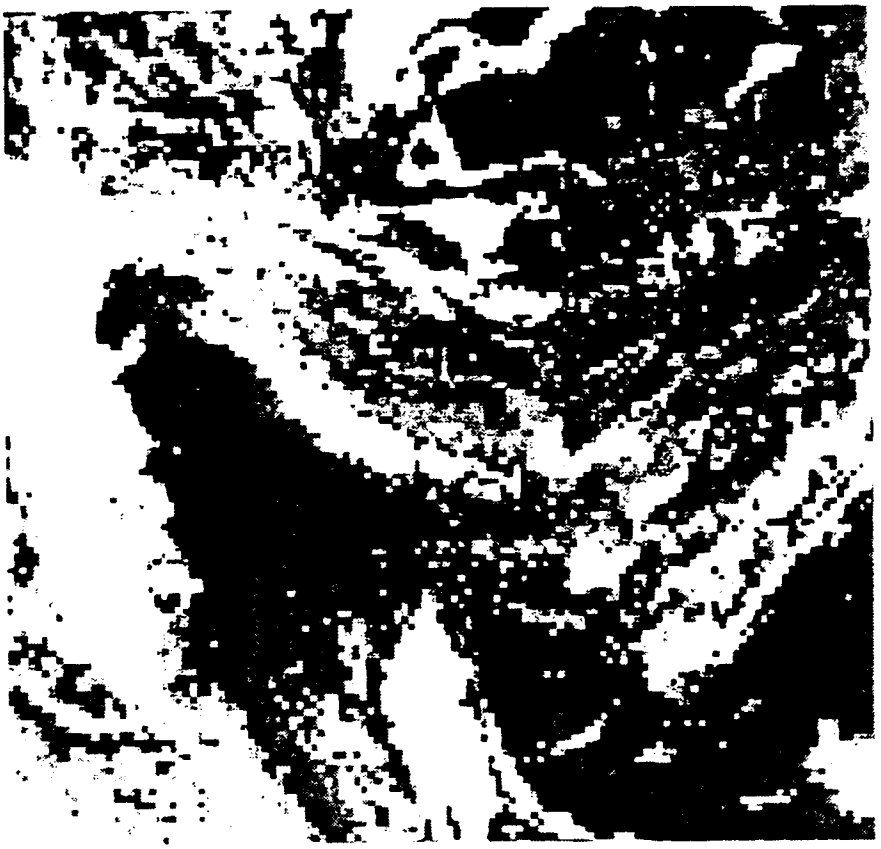
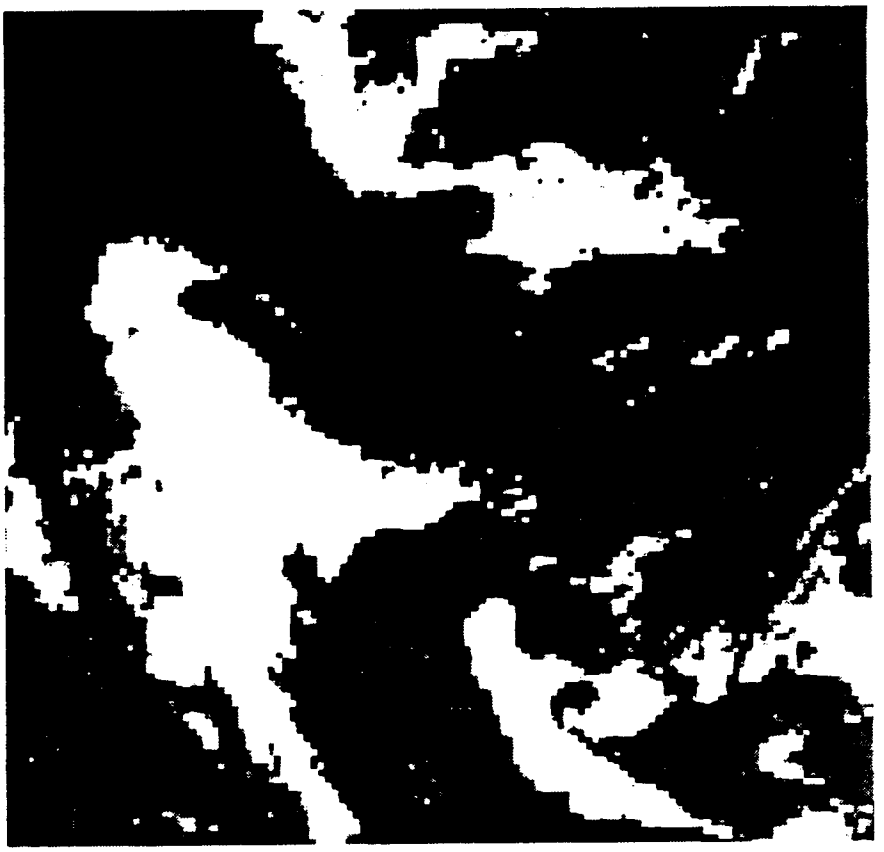


Fig. 5, cont'd.



ORIGINAL PAGE IS
OF POOR QUALITY

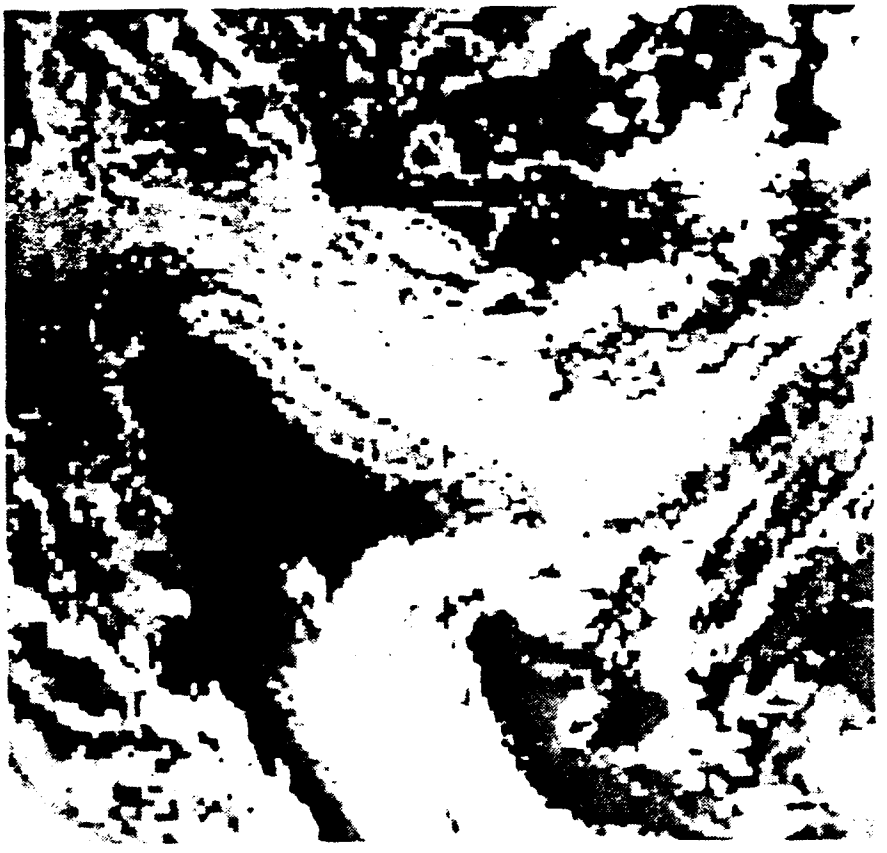


Fig. 5, cont.

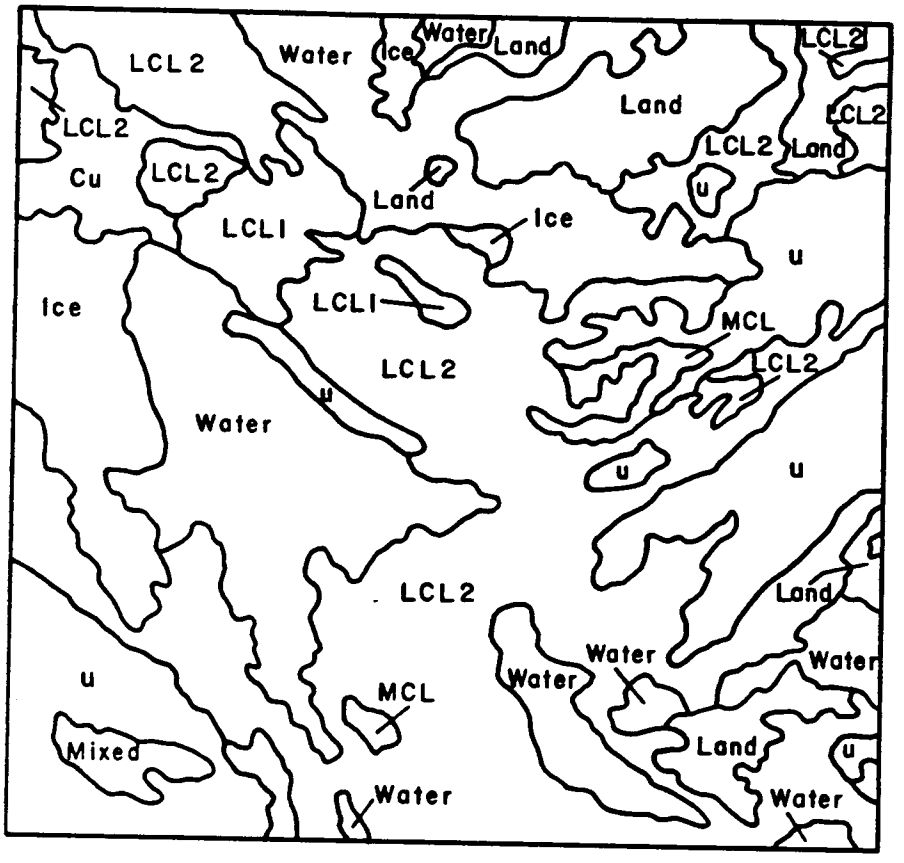

LCL 2
Water
Ice
Water
Land
LCL2
LCL2
Land
LCL2
LCL2
Cu
LCL2
Land
LCL2
Land
u
LCLI
Ice
u
Ice
LCLI
MCL
LCL2
LCL2
u
Water
u
u
LCL2
Land
Water
Water
Water
u
MCL
Mixed
Land
u
Water
Water

Fig. 6

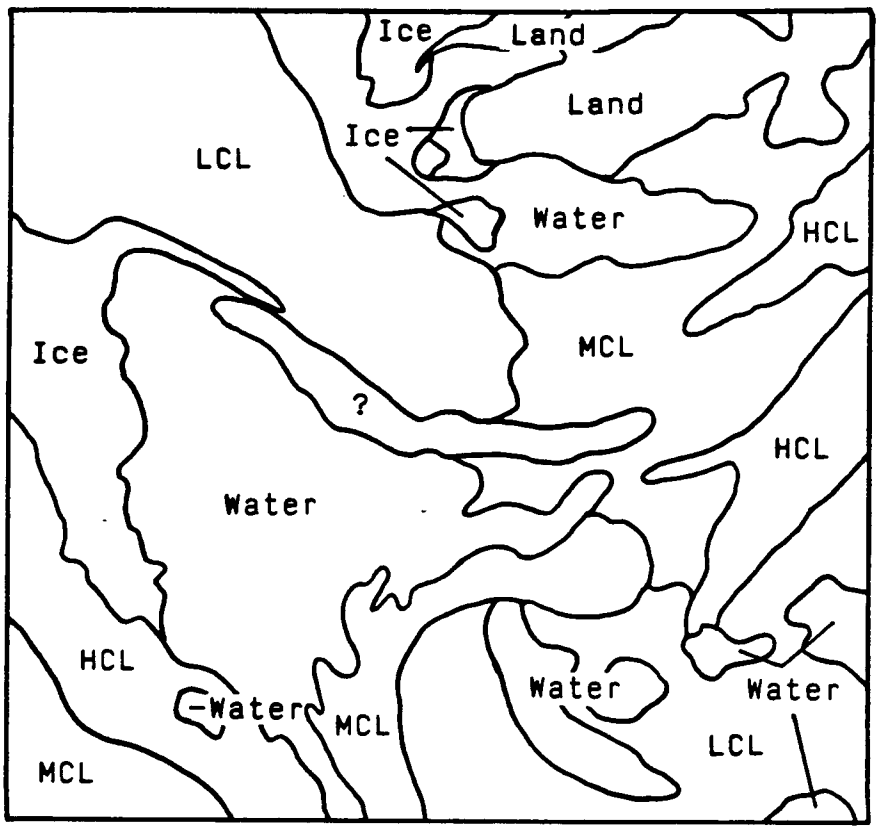

Ice
Land
Land
Ice
LCL
Water
HCL
Ice
MCL
?
HCL
Water
HCL
Water
Water
Water
MCL
LCL
MCL

Fig. 7