02

77-10005 II
CR-148974

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
6		

4. Title and Subtitle	Analysis of Data for LANDSAT (ERTS) Follow-on
5. Report Date	9/15/76
6. Performing Organization Code	
7. Author(s)	Joseph E. Sizer
8. Performing Organization Report No.	
9. Performing Organization Name and Address	Minnesota State Planning Agency, 550 Cedar Street, St. Paul, Minnesota 55101
10. Work Unit No.	
11. Contract or Grant No.	NAS5 - 20985
12. Sponsoring Agency Name and Address	
13. Type of Report and Period Covered	Type II Progress Report 3/25/76 - 6/25/76
14. Sponsoring Agency Code	
15. Supplementary Notes	

16. Abstract

The reporting period of this document is from 3/25/76 - 6/25/76. During this time effort continued to be expended in the development of 1) computer systems for analysis of the digitally recorded LANDSAT data for forest vegetation, land use and water quality classification and 2) the visual analysis of iron mining features and surficial geology. Work was begun for the training of resource managers and planners in the application of LANDSAT data

17. Key Words (Selected by Author(s))	18. Distribution Statement
Resource Information, Land Use classes, feasibility, LANDSAT Imagery, Quaternary geology	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
unclassified	unclassified	18	

*For sale by the Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.

(E77-10005) ERTS-B APPLICATIONS TO MINNESOTA RESOURCE MANAGEMENT Progress Report, 25 Mar. - 25 Jun. 1976 (Minnesota State Planning Agency, St. Paul.) 21 p HC A02/MF A01 CSCL 05B G3/43 N77-11485 Unclas 00005

Figure 2. Technical Report Standard Title Page

TABLE OF CONTENTS

PREFACE

The objectives of the contract are to develop a method of using LANDSAT data that will be reliable enough to be used by resource managers and planners. This method must be economical, easily used and readily available to the users. Secondly, to set up training sessions to acquaint potential users with LANDSAT data. To find out their needs and involve them in the continuing research. This involvement will result in the data collection techniques that will be accepted by the potential users.

The reporting period of this document is from 3/25/76 to 6/25/76. During this time effort continued to be expended in the development of 1) computer systems for analysis of the digitally recorded LANDSAT data for forest vegetation, land use and water quality classification and 2) the visual analysis of iron mining features and surficial geology. Work was begun for the training of resource managers and planners in the application of LANDSAT data.

INTRODUCTION

The objectives of this contract are first to develop a methodology for using LANDSAT data for resource managers or planners and secondly to conduct training sessions to acquaint them with LANDSAT data and applications. Three research projects are currently underway, each of which demonstrate an application of LANDSAT data. These are (1) a classification of forest vegetation and other land uses from digital tapes, (2) the compilation of a Quaternary geology map of Minnesota based on previous works and assisted by visual interpretation of LANDSAT bulk imagery and (3) projects of regional interest such as visual interpretation of land use/land cover change for the Mesabi Range iron mining district and regional water quality analysis using digital tapes.

A. ACCOMPLISHMENTS

Forest Vegetation and Land Use Classification

In this quarter, all of the planned preliminary work was completed - including software development, training set selection, and performance of classifications. Although the classification accuracies have not been determined yet, general results can be reported at this time. Also, the relative value of two algorithms, maximum likelihood and parallelepiped, for classifying the vegetation in northern Minnesota can be reported. Two new techniques for analyzing LANDSAT data were developed and will be briefly described. The feasibility of detecting lake water quality in the study area was considered and the results are included. The value of temporally registered data has not been fully evaluated, but some initial results were obtained. The ability to combine training sets for compatability with various classification systems was tested with very interesting results, and the implications may be quite significant.

The categories of natural resources which will be feasible to map from LANDSAT data have not yet been determined. However, during this period classification of the following categories was attempted:

Water
Lowland conifer
Upland conifer
Mixed forest
Brush and shrub
Grassland
Agriculture
Mined land
Uncategorized

After the accuracy of mapping these categories is determined some changes to the above classification scheme may be necessary.

1. Water - At this time only one water category seems feasible since the variability in water signatures is large within the lakes, as well as between lakes. This may be due to such factors as turbidity, depth, algal growth, tannins, etc., which may have complex interactions.

2. Lowland Conifers - This category includes black spruce, northern white cedar and tamarack. Possibly the tamarack, due to its deciduous nature, may be separated if imagery at dates other than May 29 and July 17 were used. The cedar stands are small and generally occur on the fringe of the spruce stands and, hence, from the user's point of view, do not require a separate category.

3. Upland Conifers - The upland conifer category includes red pine, jack pine, and white pine. The white pine does not commonly occur in stands but as single, typically large, trees scattered throughout the forests in Itasca County. Jack pine, red pine and spruce - the latter from the lowland conifer category - are found in relatively pure stands and the signatures are obtainable (see Table 1). As a consequence, when signatures for jack pine are developed and the classifications are performed, there are cross-overs between jack pine, red pine and spruce. When several jack pine training sets were selected it was found that the "overlap" could be pushed down - i.e., most of the crossover could be forced between the red pine and jack pine, leaving the spruce generally separate. Since the jack pine cannot be reliably broken out as a separate category, efforts were made to select training sets which caused the least overlap with the black spruce types. Thus, the upland and lowland coniferous forest types can be separated but not without some errors.

The above discussion illustrates a very important point when considering machine processing as a means of interpreting LANDSAT data. The selection of training sets has a tremendous effect on the final classification results and is subject to bias by the analyst. Also, the importance of the investigator's field experience and knowledge of the vegetation types can significantly affect the classification - the case of jack pine being a good example. If the investigator had not had ground truth for large areas in addition to the small training area, the first classification might have been accepted. The final evaluation, using test sets, would undoubtedly show the effects of the spruce-pine crossovers, but only after the fact. In this case, adequate ground truth allowed the investigator to see the problem early in the process, consider the possible alternatives, and adjust the classification system to the limitations set by the spectral signatures. In short, why define a category such as jack pine if there will be a high percentage of confusion with the spectrally-adjacent categories?

4. Mixed Forest - The mixed forest category consists of northern hardwood types (mainly oaks and maples), aspen and birch. These forest types were grouped for two reasons. First, the aspen and birch trees only rarely occur as distinct pure stands. In fact, it was nearly impossible to locate stands of aspen and birch large enough for use as training sets. The northern hardwood stands (although having a distinct signature) are not extensive and consequently, were grouped with the aspen/birch (see Table 2).

5. Brush and Shrub - The brush and shrub category includes both upland and lowland types such as alder, hazel, willow, leatherleaf and young trees. It includes transition zones between upland and lowland forest

types as well as between lake shores and the forest. While this category is general, including several species, it might be adequate for the forest manager since it is all nonforested. However, the level of detail would hardly be adequate for detailed wetland inventories, although forested wetlands (i.e., spruce bogs) can be mapped. More discussion of the problems related to mapping wetlands will be given when considering the advantages of various algorithms for classification.

6. Grassland - As with the shrub/brush category, grassland includes both lowland and upland vegetation. In fact, sedges, cattails and canary grass (all wetland species) comprise a large percentage of the area classified as grassland rather than agriculture.

7. Agriculture - Agriculture is possibly the most difficult category to classify from the LANDSAT data. The crops grown in Itasca County include mainly alfalfa, timothy, potatoes and sunflowers with the remaining land in pasture. The problem results from two factors. First, the classification is very dependent upon the date of the imagery. For example, freshly plowed fields have considerably different signatures than crops later in the season. On a given date, agricultural lands have very different signatures, but from the forest land manager's perspective, the land is all agriculture - bare ground or mature crop.

8. Mined Land - No evaluation of this category has been done.

The vegetation signatures vary with factors such as crown closure, precipitation, phenology and, perhaps most importantly in this area, intermixing of the plant communities. In Itasca County there are very few "pure" stands of any vegetation type, so considerable care must go into the selection of training sets. If a training set is too "pure",

Table 1. Mean Spectral Response from Training Sets of Five Coniferous Forest Types.

Forest Type	LANDSAT Band			
	4	5	6	7
Red pine	21.70	13.91	34.70	21.30
Jack pine	21.67	14.38	29.63	17.33
Black spruce	22.60	14.00	31.48	17.75
White cedar	23.50	16.17	37.67	21.33
Tamarack	23.38	14.25	37.63	21.13

Table 2. Mean Spectral Response from Training Sets of Two Forest Types Which Comprise the Mixed Forest Category.

Forest Types	LANDSAT Band			
	4	5	6	7
Aspen/birch	22.04	11.88	30.93	17.42
Northern hardwood	25.97	15.12	61.55	36.15

it will not be useful in classifying that particular forest type as it typically occurs. Furthermore, there are zones of transition between the types which have intermediate signatures. In areas such as northern Minnesota where the vegetation patterns are complex, particular care must also be given to selection of the best algorithm for performing the classifications. Furthermore, the optimum strategy used in selecting the training sets is dependent upon which classification algorithm is used.

Experience in this investigation suggests that the selection of the algorithm depends upon the type of material to be classified (i.e., the intended application) and the variability in the signatures of the types to be classified. Some specific examples will be given later in this report to illustrate these points.

The applicability of the LANDSAT data for detecting patterns within water which may relate to water quality was considered. Training sets of equal size (each 252 pixels) were selected in open water of five lakes: Lake Winnibigoshish, Ball Club Lake, Deer Lake, Round Lake, and Jessie Lake in Itasca County. These training sets were used to generate statistics (i.e., mean variance, and the range) upon which to classify the entire scene using a level slice algorithm. The means for each of these classes are given in Table 3 and are presently being compared and related to ground truth.

When the level slice algorithm was used to perform the classification, 90.3% (30,963 of 34,263) of the pixels classified as water were overlapping between the categories defined by the training sets taken from the five lakes. This suggests that the variability within each sample used as a training set was large by comparison to the variability between the lakes. The overlapping pixels were subsequently classified using a maximum likelihood algorithm. Thus, in the final classification, all of the water areas were forced into one of the five categories of water (see Table 3). The results of this study show a low feasibility to obtain a coefficient indicative of the over-all water quality for an entire lake. Imagery obtained at dates other than those used in this study might be more useful for this application. Various preprocessing transformations such as band ratioing might give different results. Research might also be done to evaluate this application in other parts of Minnesota.

Although the above conclusions were reached, it was of interest to consider the patterns of reflectance within lakes. A level slice was performed with the Image 100 in each band with various thresholds. Noise in the data was very apparent in Band 7 when level slicing the May 29, 1973, imagery. The noise was found to be mainly in a very narrow range of radiance values. When all of the areas with a radiance value of 3 were color coded most of the noise in the data was enhanced. Noise may be a significant factor confounding the classification over the whole scene, but it is most apparent in the water. A nearest neighbor or some type of filtering technique might be used before analyzing the data to remove the noise, but this was not done in this study. Slicing the same data (Band 7, for May 29, 1973) for the range 1 through 2 resulted in alarming most of Lake Winnibigoshish, Sand Lake, and Bowstring Lake, but only the deepest center part of Ball Club Lake and the shallow edge of Deer Lake. When all the pixels with values of 4 were alarmed none of Lake Winnibigoshish was colored except near Tamarack Point. At this level, radiance values of 4, the northeast half of Ball Club Lake was colored but none of Bowstring Lake. These patterns will be compared to maps showing lake depths. While this has not been completed, the results initially suggest that there is no relationship between either depth or turbidity with the reflectance values in Band 7 recorded on May 29,

Table 3. Mean Spectral Response from Training Sets of Five Lakes for Two Dates.

Lake name	Band 4		Band 5	
	May	July	May	July
Winnibigoshish	18.5	19.6	9.2	10.6
Ball Club	18.4	19.1	9.3	10.3
Deer	19.6	25.6	9.0	11.1
Round	18.4	19.3	9.8	10.4
Jessie	17.1	19.2	8.9	10.3

1973. This is suggested since it is known that Lake Winnibigoshish is shallow and turbid and Ball Club Lake is cleaner and deeper. If a correlation would have been found between these factors (depth and turbidity) and the radiance measured by the MSS, it may be masked by a third interacting factor - algal growth.

One application of LANDSAT data for inventorying water resources, however, is unquestionably feasible. The shape, pattern, and extent of surface water can be easily mapped. Comparing detailed maps of several lakes in Itasca County with the areas classified as water on the LANDSAT data show that some lakes have changed considerably since they were mapped. Due to the drought this year, the water level in most lakes has dropped. It seems feasible with the LANDSAT data to estimate this change in water level.

Two algorithms were used to classify the data into the nine major categories discussed earlier. These include a parallelepiped approach and a maximum likelihood classifier, as implemented on the General Electric Image 100 and the Bendix Corporation M-DAS systems, respectively. The same data were classified with both systems using the same categories and nearly identical training sets. It was found that the resulting classifications were considerably different when the two systems were used.

When a parallelepiped technique was used, less than half of the scene was classified. The remaining pixels did not fit into any of the areas defined in four dimensional feature space, as established by the training sets. These results imply that the training sets were "atypical" examples of the types they were selected to represent. Since all the training sets were validated in the field and carefully located on the graymaps, the results were at first not understandable. However, further examination of the photography and more field checking gave a possible explanation. Pure "classic" stands of each type had been selected for use as training sets and were not representative of the types in general. The natural vegetation patterns of the region are highly variable. Much of the total area is covered by mixes of the various types with complex transition zones. These patterns result in intermediate signatures which were not within the range of radiance values for any of the classes. Using the maximum likelihood classifier, however, permitted classification of more than 98% of the scene. With this algorithm the pixels of intermediate value were forced into one of the categories.

There are advantages and disadvantages to both classification techniques. With the first (parallelepiped), it is nearly impossible to classify the entire scene and with the second (maximum likelihood), pixels are apt to be classified incorrectly. However, there seems to be a logical approach to the dilemma considering the tradeoffs. First, the scene should be classified using a parallelepiped technique. The results of this classification should be quite accurate. The signatures classified with this technique would have to be quite similar to those for the "classic" training sets and should not include the mixtures and transition zones. Once these classifications are performed they could be printed with various symbols all in one color. Then the remaining pixels can be classified using the maximum likelihood algorithm. The results of this second classification could be printed with the same symbols as used in the parallelepiped technique but this time in a different color. The results of such a two step procedure would permit mapping nearly the entire scene at two levels of confidence (the accuracy could be determined for areas classified with each technique individually). The user of such a map would thus know for which areas photography and field checking are most needed.

Some general statements can be made concerning the relative merits of the maximum likelihood and the parallelepiped techniques. Classification of large distinct natural resource types can be done using the Image 100 system. However, when small units of vegetation are to be mapped, or for transition zones, the maximum likelihood classifier is far superior. In Itasca County, water is very distinct and occurs in large fairly homogeneous blocks permitting very adequate mapping with the parallelepiped technique. When the same algorithm was used for typing wetland vegetation, which often occurs in narrow bands, the results were very poor. In this case pixels along the shores of lakes were nearly all classified as wetland vegetation. Using the maximum likelihood technique the shoreline pixels were usually mapped correctly (either as water or the same as the adjacent vegetation type - on the land). That the selection of training sets seems to be more critical with the parallelepiped technique is understandable since training sets including the whole range for each type must be included (i.e., variable density, age, and understory) when developing the themes. Thus, several training sets may be required in defining the signature for each vegetation class while, with the maximum likelihood classifier, only one training set representing the center of the range of signatures for each type is needed. The selection of training sets is the most important step in classifying LANDSAT data by any supervised machine processing technique. Human bias in selecting the training sets can significantly influence the classification results.

The value of using temporally registered data is difficult to determine. Two LANDSAT scenes for Itasca County (May 29, 1973 and July 17, 1974) were registered to each other by techniques used at the Jet Propulsion Laboratory, Pasadena, California. Training sets were defined and three classifications were performed using the Image 100 system. A portion of the scene was mapped first using only bands 5 and 7 from the May imagery, then using the same bands on the July imagery. Finally, using bands 5 and 7 from both dates simultaneously, a third classification was performed. The accuracies have not yet been determined but the three color-coded maps are quite different although exactly the same training sets were used in each case. Thus it can be concluded that the degree of crossovers and types of crossovers change with the phenology of the vegetation. After the accuracy for each of the three classifications is determined more comments concerning this problem may be possible.

Bendix Corportion recently completed a project with the Bureau of Outdoor Recreation, U.S. Department of the Interior, for mapping several land use types along the Mississippi River and includes all of Itasca County. Considerable time was spent evaluating the maps produced. Many of the categories seemed ill-defined and very inadequate for forestry applications. However, the personnel at the Bureau of Outdoor Recreation found the maps to be well suited to their needs. In this project, Bendix developed 35 original categories of land cover and subsequently grouped the categories to form the classes desired by the Bureau of Outdoor Recreation. Since the original 35 categories were stored in the computer it was possible to simply regroup the categories by changing the color coding. This was tried and hard copy prints of these classifications are now being produced. When these products are obtained the accuracy will be evaluated and compared to the results from the Image 100.

Two new techniques were developed for analysis of LANDSAT 70mm imagery using a density level slicer. First, a means of determining the distribution of the gray levels in a scene (i.e., histogram of radiance) was tested and found to be accurate and repeatable. Second, and unsupervised "quasi-clustering" technique for classification using graph paper and a pocket calculator was found to be useful in analysis of the LANDSAT data. A manuscript describing both of these techniques in detail is being prepared, and should be finished within the next quarter.

A. ACCOMPLISHMENTS - cont.

Regional Development Commission Projects

During the reporting period the interpretation of all 28 Mesabi Iron Range Quadrangles was completed. Final drafting and editing was also completed. Twelve 1:24,000 scale transparent quad overlays are currently being evaluated by ARDC personnel. The remaining 16 quadrangles are now in the process of photo reproduction. Statistics on the types of land cover changes are nearly complete. Total changes from 1969 through 1976 are expected to amount to over 18,000 acres, about half of which are expansion of mine features. The total changes represent roughly two percent of the 28 quadrangle are in 6 years.

Most of the problems associated with the use of computer LANDSAT tapes for water quality analysis have been over come and digital tapes covering the Douglas county area in Region 4 Regional Development Commission have been ordered.

Quaternary Geology Project

Progress on the compilation of the Quaternary Geologic map of the State of Minnesota has permitted the location and classification of bog sites throughout the state. Sufficient data has been collected to proceed with the identification and location of glacial features.

The bog areas of the state were generally located with the assistance of the Minnesota Soils Atlas. A black and white (positive) print of May imagery at a scale of 1:500,000 from band seven was used to locate the bogs.

They were identified visually by darker tones associated with wet areas. The grassy bogs reflected more light and did not have the darker tones associated with wetness. In these cases the shape and location as well as the apparent flatness of the area was used to define the bog. The bogs were classified by their association to the surficial drainage pattern.

Technology Transfer Sessions

Work continues at two levels for the training of resource managers and planners in the applications of LANDSAT data.

Assistance and information was given on a one to one basis to the staff of Region 5 Regional Development Commission and Region 7 a Regional Development Commission. As a result of these meetings, the Region 5 Regional Development Commission is writing a contract with the Bendix Corporation for the purchase of regional land use maps generated from LANDSAT data.

Secondly, work was begun on the text and 35mm slides for a multi-media (slide-tape) show introducing the various resource inventory programs, including LANDSAT, in Minnesota.

B. Problems

There were no significant problems during this reporting period.

C. Significant Results

As mentioned previously in this report, the application of LANDSAT data in digital tape format is very feasible. The shape, pattern and extent of surface water (eg lakes) can be readily mapped. Comparing detailed maps of several lakes in Itasca County with the areas classified as water by the LANDSAT data show that some lakes have changed considerably since they were mapped.

Due to the several droughts this year (1976), the water level in most lakes has dropped. At this time it seems feasible that LANDSAT digital tape data to estimate lake water level change, due to the 1976 drought conditions.

D. Publications

There were no publications during this period.

E. Recommendations

There are no recommendations at this time.

H. Aircraft Data

None at this time.

I. Work Planned for next quarter

Forest Vegetation and Land Use Classification Project

The products of the Image 100 work will be enlarged and the classification accuracies will be estimated by sampling several areas and comparing them with the ground truth. The Bendix products will also be evaluated. Once these statistics are known, the products will be duplicated and delivered to the user-cooperators for their evaluation. Line printer classification maps will be produced and converted to transparent overlays for the areas included in several 7½ USGS topographic quadrangles. These materials will also be given to the field cooperators for their evaluation. Plans have been made for working with UNIVAC Corporation to determine if various preprocessing operations will enhance the imagery. Detailed descriptions of the two new techniques developed for analysis of LANDSAT imagery will be written. Finally, the feasibility of using a film recorder to classify vegetation types in the study area will be evaluated.

Regional Development Commission Projects

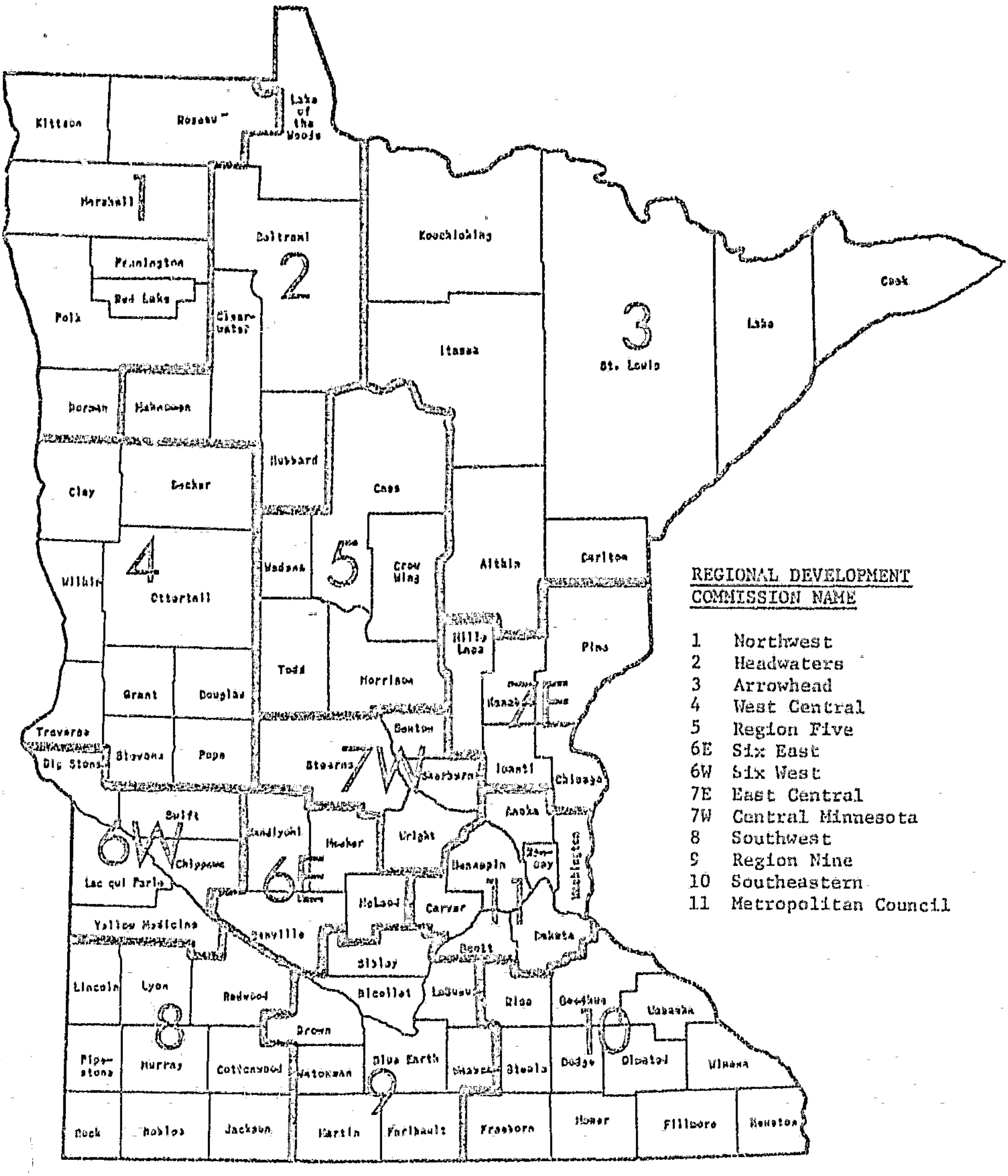
In the next reporting period the lake classification procedures will be developed and attempts will be made to apply these procedures to lakes in the Douglas County area. Arrowhead Regional Development Commission personnel will also complete their evaluation of the Iron Range land cover change maps. In addition Department of Natural Resources personnel in the Hibbing office have expressed strong interest in the quads. A set will be provided to them for evaluation in their work, by the end of July.

Quaternary Geology Project

The next step in the compilation of the Quaternary Geology map will be to visually recognize glacial features (moraines, lake plains, outwash plains, etc.) on the LANDSAT imagery that were identified in previously published books and maps. As the features are recognized on the LANDSAT imagery they will be located and delineated on the draft map.

During the next reporting period it is anticipated that the characteristics of LANDSAT imagery which facilitate the recognition of glacial features will be understood. This will be the final step in compiling the Quaternary Geologic Map of Minnesota. The maps will then be prepared for the printing process. Later the techniques used in recognizing glacial features will be presented to people and agencies who wish to use remote sensing in evaluating their resources.

Minnesota Development Regions



REGIONAL DEVELOPMENT COMMISSION NAME

1 Northwest
2 Headwaters
3 Arrowhead
4 West Central
5 Region Five
6E Six East
6W Six West
7E East Central
7W Central Minnesota
8 Southwest
9 Region Nine
10 Southeastern
11 Metropolitan Council

State Planning Agency
Office of Local & Urban Affairs
FEBRUARY 1976