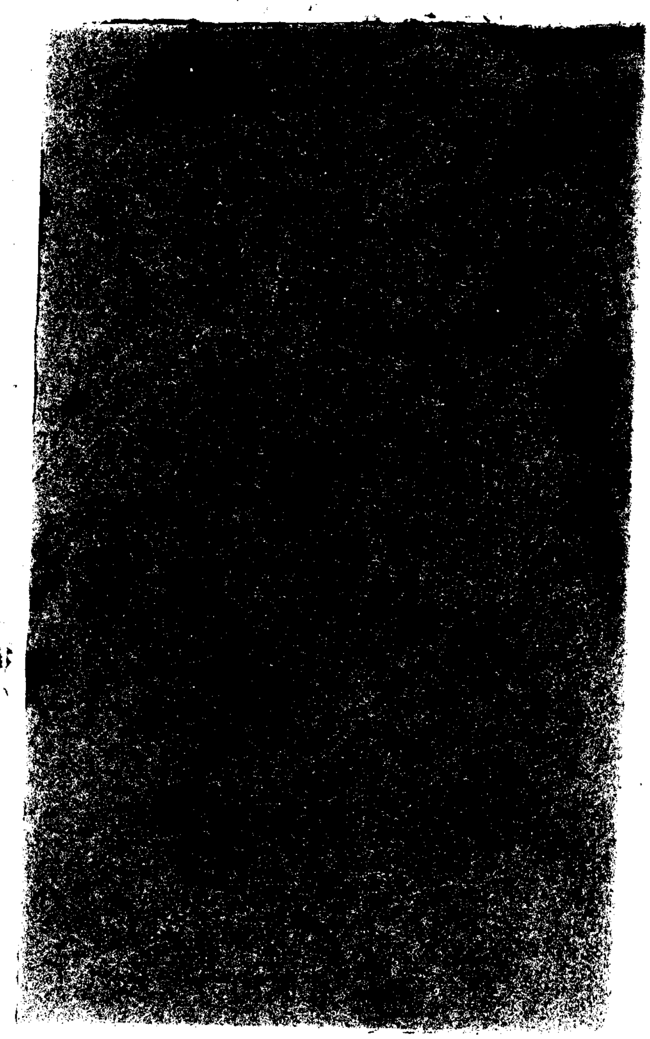
Using LANDSAT-2 Multispectral Data for Crop Area Estimation in Northern IRAN

by Ali Virasteh

The Laboratory for Applications of Remote Sensing
Purdue University W. Lafayette, Indiana
1977

This report outlines the training in remote sensing technology and its applications to agricultural statistics received by Ali Virasteh under the Andre Mayer Fellowship given by the F.A.O. at the Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, during the period April 1976 - March 1977.



SUMMARY OF TRAINING

2 April 1976	Arrived Lafayette, Indiana, and made initial contact with Laboratory for Applications of Remote Sensing (LARS).
5-9 April 1976	Participated in the LARS Short Course. This course provided an introduction to remote sensing technology and computer-aided data analysis.
12 April-- 31 May 1976	Completed self-study educational materials necessary to become more familiar with the LARSYS data analysis system and the rationale behind computer assisted remote sensing data analysis.
1 June-- 1 October 1976	Worked on an experiment entitled "Crop Identification and Area Estimation over Large Geographic Areas Using LANDSAT MSS Data", contracted to Dr. Marvin Bauer by NASA. This experience involved applying principles previously learned in a production mode. Because of this many new concepts and ideas were obtained and the practicalities of obtaining information from digital data were appreciated. The experiment was involved with determining the accuracy and precision with which area estimates for wheat in Kansas and corn and soybeans in Indiana could be obtained from LANDSAT data. This was valuable information as it provided the background for studying the same type of thing in Iran.
October 1976-- March 1977	Performed analysis of LANDSAT data taken over the Gilan Province in Iran. A complete report of that work is attached.

The period 29 June to 1 July was spent in attendance of the LARS Symposium on Machine Processing of Remotely Sensed Data.

Several days in December and January were spent with representatives of the Indiana branch of the U.S. Department of Agriculture, Statistical Reporting Service. This time was spent to become familiar with the system of data collection and formulation of agricultural statistics in the United States.

Further details of the first six months' training can be found in individual progress reports. The period 1 April to 15 May is covered in the first report, 16 May to 2 July in the second, 2 July to 1 September in the third, and 1 September to 1 December in the fourth.

ACKNOWLEDGEMENTS

I am thankful to the Senior Officer of the Statistics Division of the Statistical Development Service and the Officer-in-Charge of the Fellowships Group of the F.A.O. for granting the fellowship for this training in remote sensing technology and its applications obtained at the Laboratory for Applications of Remote Sensing (LARS) at Purdue University, West Lafayette, Indiana, U.S.A.

I am also thankful to the Ministry of Agriculture and Natural Resources, Government of Iran, for giving me the opportunity of receiving this training.

I am also thankful to the Director of LARS and his staff for arranging and participating in this training.

Using LANDSAT-2 Multispectral Data for Crop Area Estimation in Northern Iran

by

Ali Virasteh

Deputy General Director

Department of Agricultural Economy and Statistics

Ministry of Agriculture and Natural Resources

Teheran, Iran

ABSTRACT

Computer-assisted analyses of two LANDSAT frames, acquired in 1975, over the Gilan Province in Iran are reported. With only 1:250,000 color composites of the LANDSAT data and personal knowledge of the area to use as reference data, area estimates similar to Iranian government estimates were obtained for rice in both frames, and tea in one frame. The results of this study show that with a minimum of information, accurate crop area estimates can be obtained using remote sensing technology. In some cases, such as tea in this study, the size of the area being considered has a direct bearing on the success in obtaining good estimates.

INTRODUCTION

As world food demand increases the need for timely and accurate production information becomes more critical. Many countries and organizations make attempts to assess world food production. However, relatively few have the tools or technology in a production mode to completely and accurately accomplish such a task. In this report a method of survey and production estimation will be described that has potential for providing such information. The benefits to be realized from accurate estimates include increased price stability, more appropriate and realistic domestic and foreign policies, and more efficient use of resources.

A new technology that is evolving which holds some promise for quicker, more accurate and less costly crop production information is that of remote sensing.[2] Remote sensing, defined as the process of acquiring information about objects from a distance, is of several types. The type discussed here involves measuring the intensity of electromagnetic energy being returned from the ground in several wavelength bands and using that information to discern some qualities of the piece of ground being considered (crop type, yield potential, etc.)

Landsat 1 and 2, orbiting the earth at 940 kilometers are capable of making such energy measurements for areas 180 kilometers square anyplace on the globe every 18 days.[3] Each 180 kilometer square area is broken into 7.7 million points by the action of a line scanner. Each resulting point has four values associated with it; the level of energy being returned from the portion of ground it represents in the green (.5-.6μm), red (.6-.7μm), and two infrared (.7-.8, and .8-1.1 μm) portions of the spectrum. The spatial resolution of Landsat data is 80 meters and each data point represents 0.64 hectares.

In order to utilize these large volumes of data efficiently and to provide the type of information desired, computer aided analysis of the data is commonly performed. One widely accepted analysis concept known as LARSYS takes a maximum likelihood pattern recognition approach to classifying data such as that described above. This approach involves locating points representative of the cover types of interest, and using the spectral properties of those points to decide to which of those cover types an unknown point is most likely to belong. Other classification concepts include clustering, level slicing, minimum distance, and sample (or field) classification.

The topic of this report is computer aided analysis of Landsat 2 multispectral scanner data collected over northern Iran using LARSYS. Two frames of data were analyzed, both covering approximately the same area (Figure 1), but at two different times of the growing season (June 20, 1975 and August 31, 1975). The results of the analyses of these data will be discussed to illustrate the potential this technology has for providing timely, large area, crop production information.

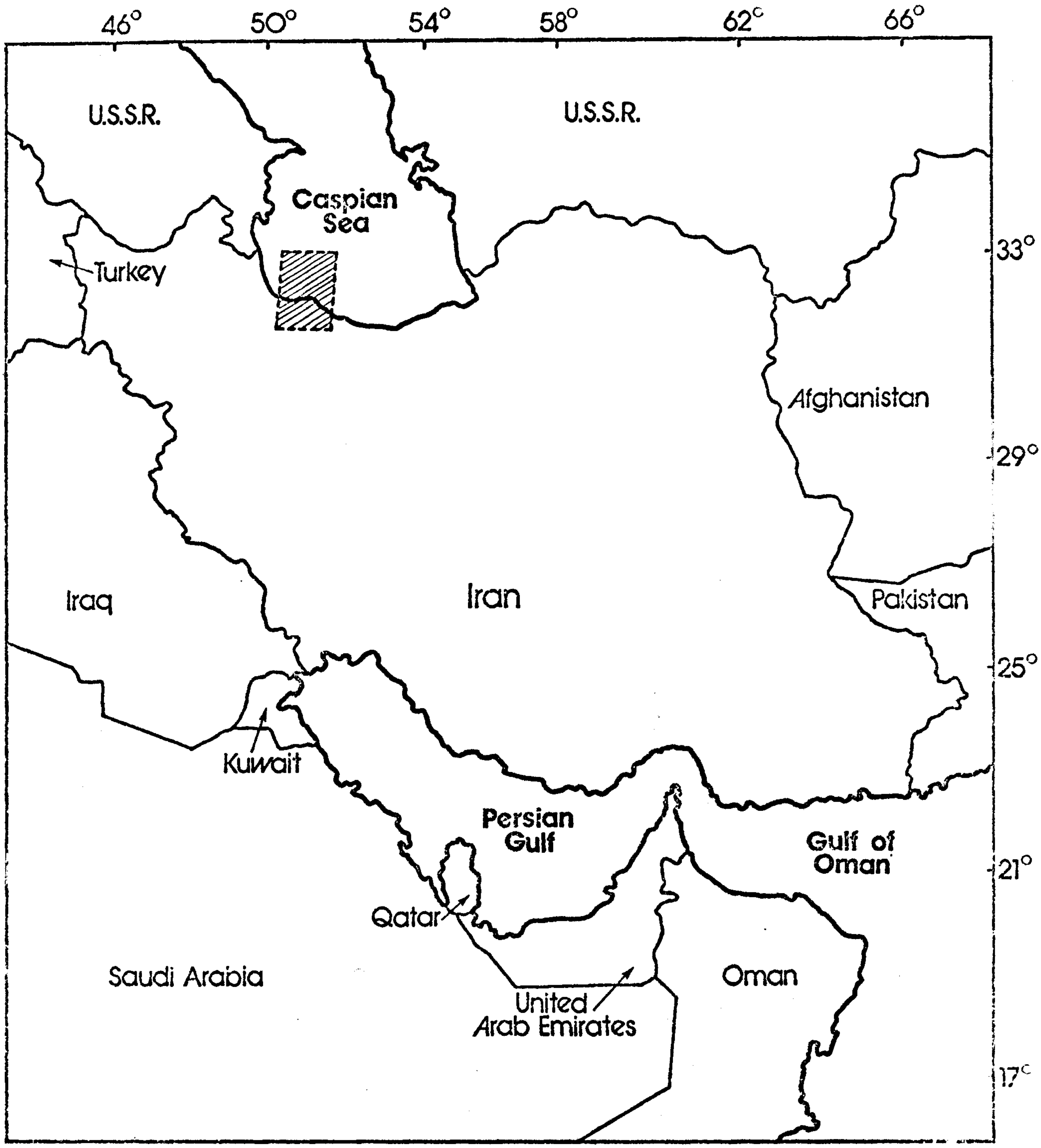


Figure 1. Map of Iran showing the approximate location of the LANDSAT frames used in this study.

BACKGROUND

Remote sensing as applied to identification of earth surface features is dependent upon measurable variations in electromagnetic field strength. Variations are of three main types. The first, spectral, is dependent on the physical conditions of the cover type being sensed. As the electromagnetic energy from the sun strikes the cover type it will either be reflected back into space, absorbed and then emitted at a different wavelength, transmitted through the material (and lost for measurement purposes) or scattered depending on the geometry and physiology or physical make-up of the cover type. In as much as different combinations of these things happen or in as much as they happen at different wavelengths, spectral variations will occur. The second, spatial, arises because the spatial location of cover type A is different than cover type B. Measurement of spatial variations is dependent on a device that can measure the electromagnetic energy being returned at many spatial locations, such as a line scanner. The third type of variation in electromagnetic field strength is temporal. Many cover types, and especially crops, change spectrally over time as they progress through growth and development stages.

Several investigators have reported the utility of using Landsat data to map crops.[2,4,5] In short the capability can be stated as this: any crop can be mapped using multispectral scanner data if the spectral values associated with the crop are detectably different from the values of the other features to be mapped.

Note that this statement of capability assumes that the person wishing to have the mapping done knows which spectral values belong to each cover type he is interested in. Therefore, the person must locate points representative of the cover types of interest and therefore must have some information about the location of each of those cover types. This is referred to as "ground truth" or reference data, and in general, the more ancillary information available, the more information that can be extracted from the data.

Several causes for a lack of detectable difference are 1) the feature to be mapped is smaller than the spatial resolution of the scanner, 2) the physical make-up of one of the features is very similar to that of other features such that spectral variation does not occur among them, 3) inappropriate wavelength bands for discrimination are available, and 4) uncertainty about the location of the features. An example of the third cause is snow and clouds, which are difficult to automatically distinguish between with Landsat data but are readily distinguishable with Skylab data which contains a middle infrared band (Landsat does not).[6] The fourth cause of problems in discrimination is the most common one, both due to inadequate reference data and improper use of it. However, there are also well defined exceptions to these rules. For example, water bodies one-fourth the spatial resolution of Landsat can often be mapped simply because water is so spectrally different from most other materials. Another is that even though spectral variation between cover types may not exist at one time of the year often it will at another. Although all the causes can be very significant, they can usually be remedied by intelligent selection

Table 1. Crop Calendar for Rice, for 8 subprovinces of the Gilan Province.

Name of Subprovinces	Date of Sowing	Date of Transplantation	Date of Maturity	Date of Yellowing	Date of Harvesting
Bandar Pahlavi	Apr. 5 to 15	May 5 to 15	July 5 to 15	Aug. 5 to 15	Aug. 15 to 25
Fuman	Apr. 5 to 15	May 1 to 10	July 1 to 10	Aug. 1 to 10	Aug. 10 to 20
Lahijan	Apr. 5 to 15	May 5 to 15	July 5 to 15	Aug. 5 to 15	Aug. 15 to 25
Langarud	Apr. 5 to 15	May 5 to 15	July 5 to 15	Aug. 5 to 15	Aug. 15 to 25
Rasht	Apr. 5 to 15	May 5 to 15	July 5 to 15	Aug. 5 to 15	Aug. 15 to 25
Rudbar	May 5 to 15	June 5 to 15	Aug. 5 to 15	Sep. 1 to 10	Sep. 10 to 20
Rudsar	Apr. 5 to 15	May 2 to 12	July 2 to 12	Aug. 2 to 12	Aug. 12 to 22
Sameh-Sara	Apr. 5 to 15	May 1 to 10	July 1 to 10	Aug. 1 to 10	Aug. 10 to 20

of the time data was taken and the sensor system used to take the data.

MATERIALS

Landsat data was obtained for an area in Northern Iran, next to the Caspian Sea. This area was selected because of the economic importance of the major crop grown there, rice. It is only along this strip bordered by the Caspian Sea and the Albourz Mountains that precipitation is plentiful enough (approx. 2000 mm/year) to raise this crop. For similar reasons this strip is the only area in Iran that tea can be grown. Two Landsat frames were selected that covered approximately the same geographic area but at different times in the growing season. The area included in these frames contained approximately 90% of the area under rice in the Gilan Province and more than 60% of the strip in which rice and tea can be grown. Different dates over the same area were selected so that the ability to distinguish rice and tea could be evaluated at different growth stages. For example on June 20, 1975 most of the rice would have been transplanted for about one month and would have been very green with a high percentage of ground cover. On the other hand, on August 31, 1975, most of the rice would have been harvested, leaving only a yellow stubble - quite spectrally different than other ground cover (Table 1). Tea would not have changed much between these dates but in order to establish the ease with which it could be differentiated from other crops at those times both frames were necessary. The corresponding Landsat frames immediately adjacent to the East would have provided a complete coverage of the rice and tea production area.

A 1:250,000 color composite image of the Landsat data was obtained for use as reference data. This image is constructed by successively exposing the same piece of film to three of the four Landsat bands, each band through a different color filter. The result of this process is a color print that simulates a color infrared photograph of the portion of ground covered by the Landsat frame. A 1:250,000 topographic sheet, and a 1:250,000 regional map of land resourses and potentialities were also available to use as reference data. The major problem with this reference data was its scale. Although all three items were at the same scale, the Landsat digital data itself was at a scale of 1:24,000. All efforts to obtain more detailed maps and aerial photography failed. Thus the reference data available was not of sufficient detail to pinpoint individual fields as being a specific cover type. Rather, only general locations were available and to as great an extent as was possible the color-composite was used to specify points representing the cover types of interest.

The Iranian Ministry of Agriculture and Natural Resources estimates of the area under cultivation of rice, and tea in the Gilan Province in 1975 were used as a basis of comparison for the results obtained in this study. Under the Crop Reporting System, these estimates are prepared on the basis of monthly interviews with sample farmers. During each interview the area under cultivation of each crop, their growth stages, and production forecasts before harvest or production amounts after harvest are recorded. This information is collected for all 4,000 sample villages in Iran, analyzed and processed, resulting in area estimates and crop condition reports for each province.

As mentioned before, the system used to perform the analysis of the data described above was LARSYS. LARSYS, made up of 18 integrated, interrelated processing functions installed on a computer, embodies an entire approach to conversion of remote sensing data to the information desired.[7] Central to this approach are the concepts "pattern recognition" - involving training the machine to recognize patterns, and "multispectral"- the data to be utilized has more than one dimension, the spectrum having been broken into many smaller regions.[8] At LARS, the LARSYS System is currently implemented on an IBM 360, Model 67 (time sharing) computer.

METHODS

The first part of any analysis is to specify analysis objectives. This normally includes a statement of the cover types to be mapped, at what geographic location, and with what accuracy. With these factors in mind an appropriate data set can be selected and ordered along with any appropriate and available reference data.

In this case a Landsat computer compatible tape (CCT) containing digital data and the color composite image of the data were obtained from EROS Data Center in Sioux Falls, South Dakota. Upon its arrival the tape was reformatted to make it compatible with the LARSYS System. LARSYS and many other data processing systems use a format different than that of the Landsat CCT to reduce processing costs associated with the CCT format.

During the time taken to get the data reformatted, the color composite was studied to identify 8-10 small areas (100 lines x 100 columns) that represented all the different conditions of the cover types present in the area to be classified. Once the data was reformatted a gray level map of each of these small areas was produced. That is, each point in each of these small areas was assigned one of ten symbols based on its level of response (brightness) in the channel specified. This was accomplished by the Pictureprint processor and resulted in an enhancement in the ability to visualize and delineate spatial boundaries present on the ground at the time the data was taken.

The boundaries could then be compared to those in the color composite of the data and the general locations of the cover types in each small area could be interpreted. With this information a smaller part (about one-fourth) of each small area was selected to be supplied to an algorithm for division into 6-15 spectral "classes". The processor involved in this step (Cluster) did not divide the points into the classes defined in the analysis objectives (information classes), rather it examined the spectral response values of each point in the four wavelength bands and established groups of points with similar response values (spectral classes). It performed this operation for each smaller or "training" area separately.

After the Cluster processor had defined these spectral classes for each training area it printed a map showing into which spectral class each point in each area had been assigned. These maps could then be compared to the color composite and reference maps to associate an informational name with each spectral class.

Another output of the cluster processor is punched cards that describe by line and column address the location of those points belonging to each spectral class in each training area. These addresses were then supplied to a processor (Statistics) that calculated a mean of the responses of the points in each class in each wavelength band and a covariance matrix for each class.

The means and covariance matrix calculated by statistics were then used to determine the probability of successfully distinguishing among the spectral classes. A distance measure, transformed divergence, calculated by the Separability processor using the means and covariance matrix of each class provided this information. When a low separability between two classes was reported, a decision as to whether to combine the classes or delete one of them had to be made. The usual basis for this decision was the assigned informational names of the spectral classes. Two spectral classes determined to represent the same cover type would be combined whereas if the spectral classes in question belonged to different informational classes the one less established as belonging to an informational class would normally be deleted. One other guideline would be that any spectral class showing low separability between many other classes of different informational identity would normally be deleted regardless of the firmness of its identity. Often the steps of combining and deleting of classes could be repeated many times before a completely separable and acceptable set of classes was obtained.

When an acceptable set was obtained sample classifications would be performed to evaluate the representativeness of the classes. This involved picking test areas (separate and different than the training areas) and supplying their addresses and the means and covariance matrices of the accepted classes (in some cases separate spectral classes and in others groups of spectral classes, each having an informational name, and referred to as a training class) to the Classifypoints algorithm which then compared the response values of each point in each test area to the means and covariance matrix of each class and assigned the point to the training class (any of the accepted classes) it had the highest probability of belonging to.

The results of these test classifications were then examined for correctness. If complete misclassifications of large areas occurred, the previous steps were examined for errors or questionable decisions. Whatever steps necessary were taken to try to represent the misclassified areas without causing new confusion among classes. In some cases this meant deciding the less important of two possible errors. When the results of test classifications were acceptable and considered optimum (for the purposes of the study) the entire area of interest was classified with the Classifypoints algorithm. The result of this classification was stored on tape.

Because of the lack of detailed reference data the percent of correct classification could not be estimated. Instead qualitative examination of its correctness was performed. In addition the printresults processor was used to obtain estimates of the acreage under cultivation of rice and tea in the Gilan Province which were then compared to Iranian Government estimates. These two methods together provide a good picture of the correctness of classification.

For the purpose of showing the data used, 35mm color composite slides of the data were generated using the Imagedisplay processing function. Similarly color coded 35mm slides of the classification results were obtained using the Photo processor. Prints of these slides were obtained and are used in this report to support the results and discussion section.

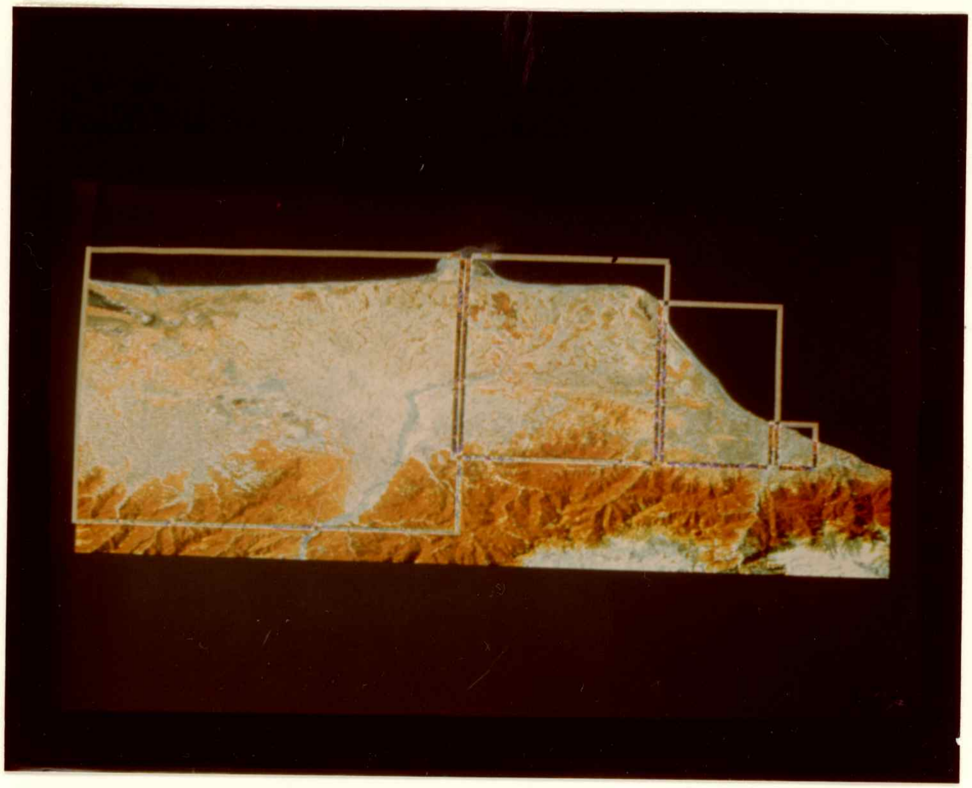


Figure 2. False Color Composite of Landsat-2 Data Acquired August 31, 1975 With the Four Subareas Classified Outlined.

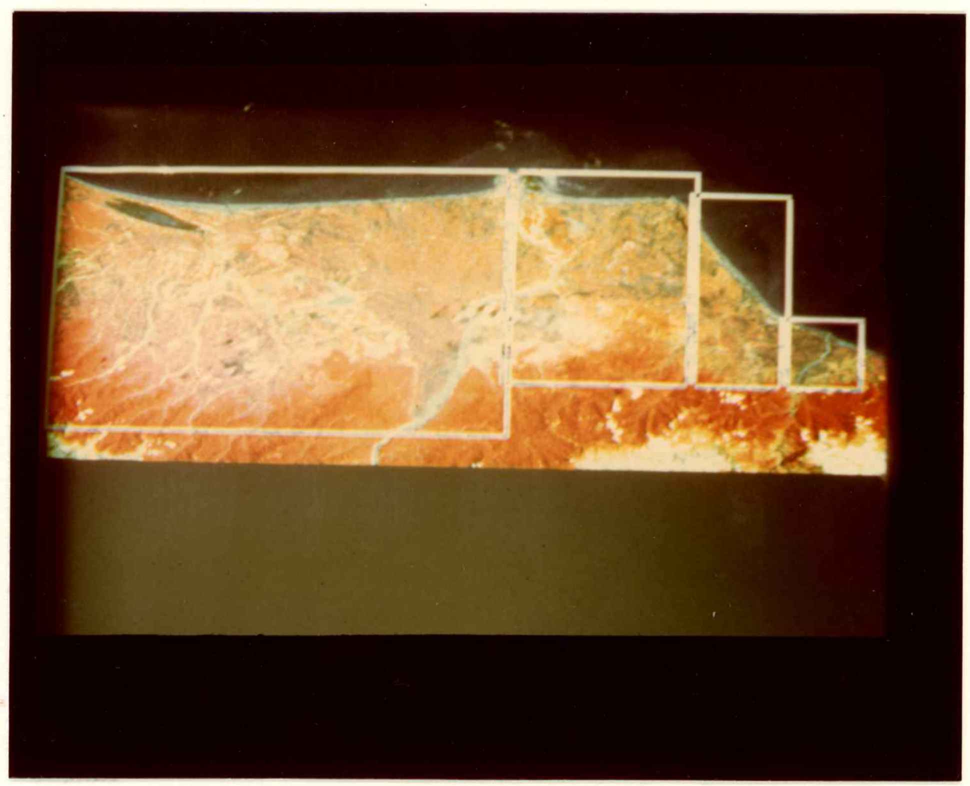


Figure 3. False Color Composite of Landsat-2 Data Acquired June 20, 1975 With the Four Subareas Classified Outlined.

Table 2. Identity of Spectral Classes August 31, 1975 Landsat Data.

Area/Classes	1	2	3	4	5	6	7	8
1	R	R	u.v	o.v	o.v	o.v		
2	o.v	R	R	o.v	o.v	o.v	o.v	o.v
3	o.v	o.v	o.v	R	o.v	o.v		
4	ur	ur	ur	ur	o.v	o.v		
5	B.S.	o.v	B.S.	o.v	R	W		
6	B.S.	R	R	B.S.	u.v	W		
7	B.S.	o.v	T	ur	T	T		
8	T	T	u.v	T	T	T		
9	R	R	o.v	o.v	o.v	o.v		
10	F	F	F	F	F	F		

R: Rice
u.v: unknown vegetation
o.v: other vegetation
ur: urban
B.S: Bare soil
W: Water
T: Tea
F: Forest

RESULTS AND DISCUSSION

At the outset the decision was made to obtain LANDSAT data over the Gilan Province and to analyze it to obtain the area under cultivation of rice and tea. LANDSAT data was chosen because of the large area for which information was desired. A computer search of available LANDSAT data over this area was requested from EROS Data Center, Sioux Falls, South Dakota. To facilitate use of the listings returned by EROS, LARS maintains a library of microfilm images of band 5 of all LANDSAT frames collected. Once the listing of data available over the Gilan Province had been received, the microfilm images were referred to and a specific frame selected. The first frame selected had been acquired August 31, 1975. A CCT and color composite of this data were then ordered. This date was selected because it was cloud free and close to the harvesting time of rice, when the rice would be a different color than any other vegetation (Figure 2).

Upon receipt of the color composite, eight small areas were selected to represent all of the cover types present. Out of these eight, ten smaller training areas were chosen. Six spectral classes were obtained for each of nine of the training areas and eight spectral classes were obtained for one of the training areas. Out of the resulting sixty-two spectral classes, ten were identified as rice, eight as tea, six as forest, five as urban, five as wet sand, two as water, and twenty-six as other (not specified) vegetation (Table 2).

When separability information had been obtained for these classes, it was found that tea and other vegetation and rice and other vegetation were being confused. Many other classes whose identity was not well established were deleted when it was found that they were being confused with classes of other identity. For the most part classes of the same identity tended to form groups, that is, they were confused mostly with each other. By pooling such classes together and by eliminating any of the group being confused with classes of another identity tentative sets of classes were obtained. By performing these operations many times, a set of classes (some composed of groups of classes), all acceptably separable from one another, were arrived at.

When the test classifications were performed, it was found that much of what had been previously identified as other vegetation was being classified as tea, and vice versa. At this point it was felt that this was due to incorrect definition of the spectral value's representative of tea. All subsequent efforts to more precisely define tea spectrally failed, primarily because of uncertainty about the exact location of tea. Consequently, it was decided to combine the tea classes with other vegetation and call them other vegetation. This resulted in three training classes of rice, four of other vegetation, four of urban-wet sand (another set of classes difficult to separate), four of forest, and two of water.

In order to reduce the amount of computer time necessary to classify the area of interest it was divided into four subareas (Figure 2), within which every second line and every second column were classified. This was done primarily to save the money allocated for computer usage for subsequent analyses. This 25% sample of the data was felt adequate to obtain accurate crop area estimates. Upon examination of the maps of the classification results, it was felt that the area had been accurately classified (Figures 4-7), except for tea. In addition, the estimate of the area of rice under cultivation (200,000 hectares Table 3-6) was close to the Iranian government estimate (220,000 hectares). The LANDSAT frame used to make this analysis covered only 95% of the area under cultivation of rice in the Gilan Province. 95% of the government estimate is 209,000 hectares which is certainly within the error of estimation associated with the government estimate.

To determine whether the difficulty encountered in correctly mapping tea and other vegetation could be resolved at a different time of the year, and because of a need to know the accuracy with which rice could be mapped on a different date using these methods, it was decided to select a second frame over the same area, but at an earlier time in the growing season. The earliest date available that was sufficiently cloud free was June 20, 1975. Therefore, a CCT and color composite image were ordered for this date.

As soon as the data was available, ten small areas were selected out of which ten smaller training areas were selected. Six spectral classes were obtained for each training area. Out of the resulting sixty spectral classes, twenty-three were identified as rice, two as tea, eight as other vegetation, six as forest, four as water, one as marsh, three as cloud, three as shadow, four as urban, three as bare soil, and three eliminated due to uncertainty about their identity (Table 7).

The problems encountered in the previous analysis were encountered in this one also. One class of tea was used in this

Cover Type	Color	Points	Acres	Hectares
Rice	orange	6924	8683	3514
Forest	light green	-	-	-
other vegetation	dark green	904	1134	459
urban/Bare Soil	red	1100	1379	558
Water	sky blue	1476	1851	749

Table 3. Distribution of Cover Types, Area 1, August 31, 1975 Data.

Figure 4. Color Coded Classification Result, Area 1, August 31, 1975 Data.

Cover Type	Color	Points	Acres	Hectares
Rice	orange	34528	43298	14523
Forest	light green	144	181	73
other vegetation	dark green	22824	28621	11583
urban/Bare Soil	red	6856	8597	3479
Water	sky blue	57052	71543	28953

Table 4. Distribution of Cover Types, Area 2, August 31, 1775 Data.

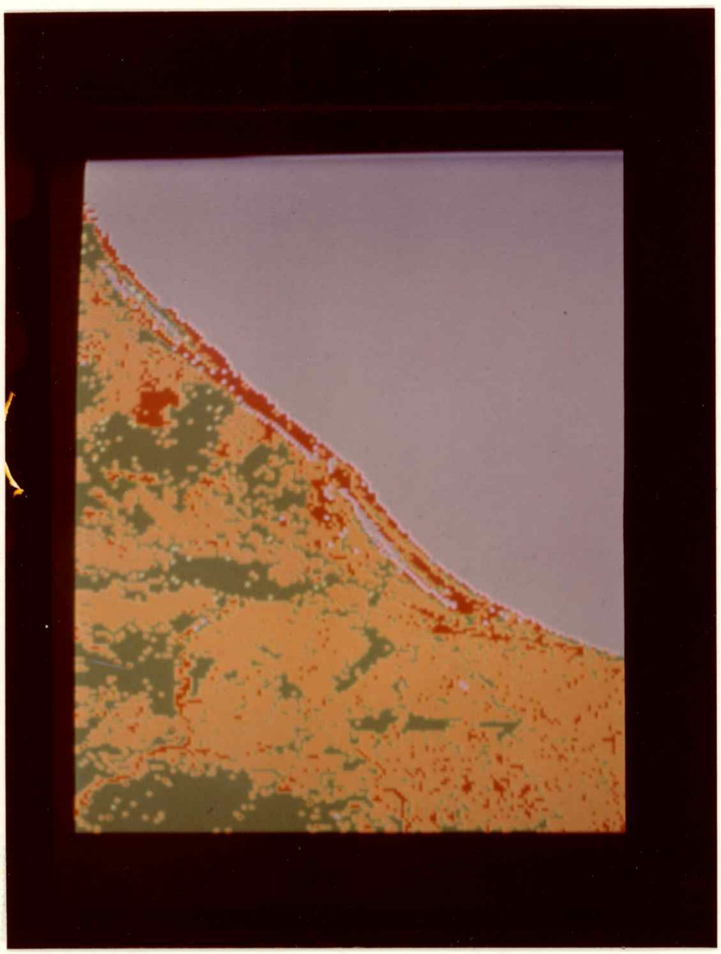


Figure 5. Color Coded Classification Result, Area 2, August 31, 1975 Data.

Cover Type	Color	Points	Acres	Hectares
Rice	orange	97848	122701	49657
Forest	light green	9788	12274	4967
other vegetation	dark green	103504	129794	52527
urban/Bare Soil	red	9968	12500	5059
Water	sky blue	30896	38744	15679

Table 5. Distribution of Cover Types, Area 3, August 31, 1975 Data.

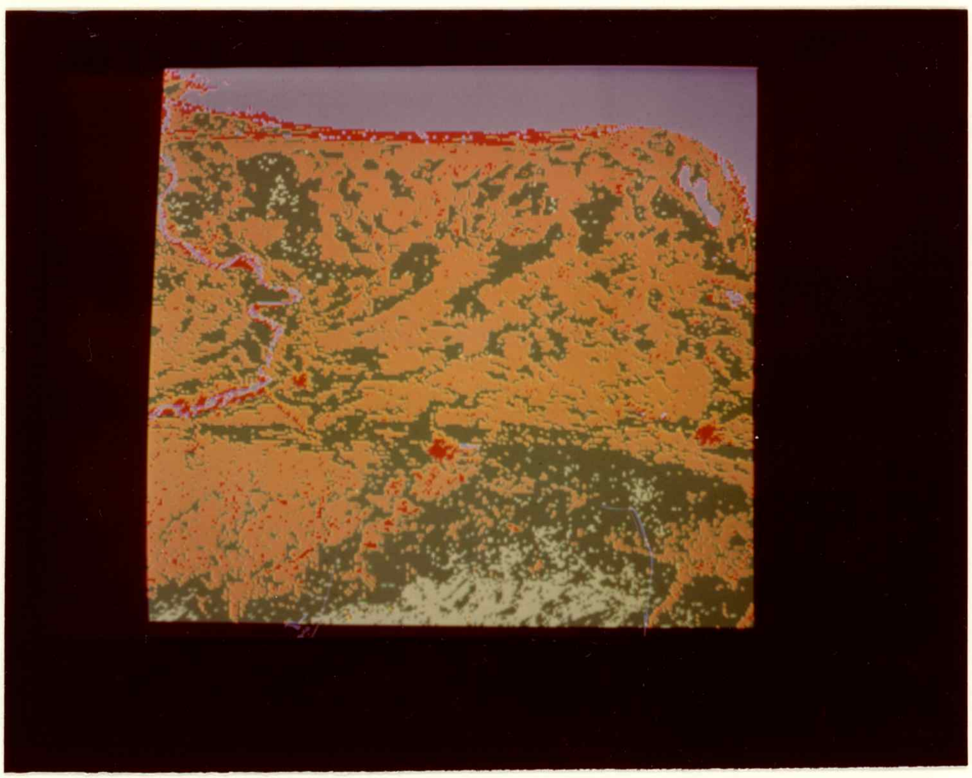


Figure 6. Color Coded Classification Result, Area 3, August 31, 1975 Data.

Cover Type	Color	Points	Acres	Hectares
Rice	orange	253804	318270	128802
Forest	light green	39648	49719	20121
other vegetation	dark green	227904	285792	115658
urban/Bare Soil	red	37188	46634	18872
Water	sky blue	69044	86581	35039

Table 6. Distribution of Cover Types, Area 4, August 31, 1975 Data.

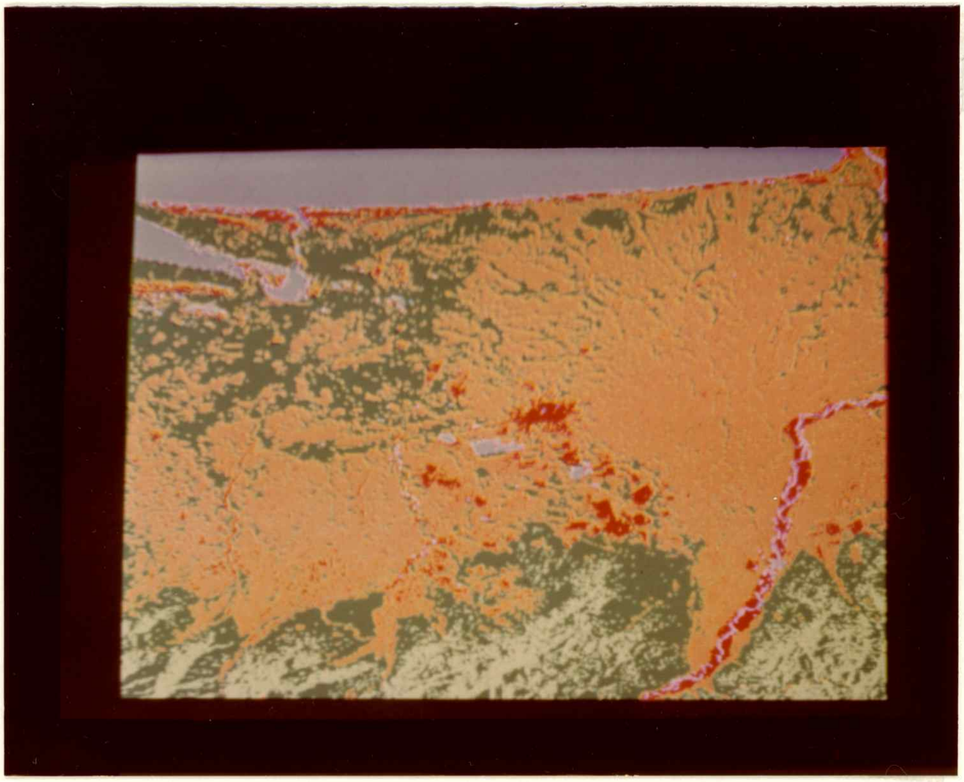


Figure 7. Color Coded Classification Result, Area 4, August 31, 1975 Data.

Table 7. Identity of spectral classes. First procedure, June 20, 1975 Landsat data.

Area/Classes	1	2	3	4	5	6
1	o.v	R	R	R	R	W
2	cl.	cl.	R	R	sh.	sh.
3	ur.	ur.	ur.	o.v	R	R
4	B.S.	B.S.	F	F	R	W
5	o.v	R	R	R	R	R
6	o.v	o.v	o.v	R	R	W
7	cl.	F	F	F	F	sh.
8	T	T	E	E	R	R
9	R	R	ur.	o.v	o.v	M
10	B.S.	R	R	R	E	W

o.v: other vegetation
R: Rice
W: Water
cl: cloud
sh: shadow
ur: urban
B.S: Bare Soil
F: Forest
T: Tea
E: Eliminate
M: Marsh

case, even after the test classification had been made. However, one class was not sufficient to represent tea and it was apparent that the area estimate for tea after the final classification had been made would be low. It was also apparent that rice would be overestimated. It was more difficult to determine when and where other vegetation was being classified as rice because both were green at that time of year and appeared as only very slightly different shades of red on the color composite (Figure 3). Therefore, instead of proceeding to make the final classification at this point, it was decided to take a slightly different approach in generating the spectral classes.

The problem with the spectral classes, as interpreted, was that they were too broad and lacked information class specifity. In other terms, it seemed that each spectral class in reality represented two or more conditions of different crops. This is especially likely to occur when cover types that are very similar spectrally are present in a training area. In addition, as mentioned before, it was quite difficult to distinguish rice from other vegetation and tea from other vegetation with only the LANDSAT color imagery as reference data. In order to improve the correspondence of spectral classes with information classes, a special and experimental version of the Cluster processor was used. This version computes a parameter for each number of clusters specified by the user. The number of clusters at which the parameter is smallest should provide the most distinct clusters. This version of Cluster was then used on nine new training areas. As a result, each training area was divided into 10 to 15 spectral classes, the number for each training area determined by the cluster distinctness parameter. Out of the resulting 122 spectral classes, 23 were identified as rice, three as tea, 29 as other vegetation, six as unknown (tea or other) vegetation, 13 as forest, 13 as urban, six as bare soil, four as marsh, three as water, one as cloud, one as shadow, two as haze, three as wet soil, and 15 unidentifiable (all from one training area and subsequently deleted (Table 8).

After the test classifications were made with the classes resulting from optimization of the separability of classes, it was determined that areas south of Rasht identified as fodder crops were being classified as tea, and haze was being classified as bare soil. To overcome this confusion, two additional training areas were selected and clustered into eight classes each. Four of the classes from the first area were identified as fodder crops, and four from the second area were identified as clouds or haze. These new spectral classes were then included with the original group and separability redetermined. After many iterations, 37 classes were arrived at. Of this 37, five were rice, seven other vegetation, three tea, six forest, five cloud, one shadow/rice, three water, one marsh, four urban/bare soil, one haze, and one shadow.

For final classification the Gilan Province area was divided into the same four subareas as for the first LANDSAT frame (Figure 3). Because of a prior knowledge about the first, second and fourth subareas, that little if any tea is grown in them, the three tea classes were deleted before classifying them. The tea classes were retained for classifying the third area - in and around Lahijan and Langarud - where most of the tea in Iran is produced. The estimate of area under cultivation for tea (14,700 hectares) compared favorably to the government estimate

Table 8. Identity of spectral classes. Second procedure, June 20, 1975 Landsat data.

Area/Classes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	o.v	o.v	o.v	R	R	R	R	R	R	R	R	R	R	R	R
2	o.v	o.v	o.v	o.v	o.v	o.v	o.v	o.v	o.v	o.v	o.v	o.v	R	R	R
3	F	F	F	F	u.v	u.v	u.v	u.v	R	R	R	R			
4	F	ur.	F	B.S.	B.S.	o.v/T	F	o.v	o.v/T	u.v	T	ur.	o.v	u.v	M
5	B.S.	B.S.	o.v	B.S.	o.v	B.S.	o.v	o.v	o.v	R	R	W			
6	cl.	H	F	F	F	H	F	F	F	sh.					
7	ur.	ur.	ur.	ur.	ur.	ur.	ur.	ur.	ur.	ur.	o.v	o.v	ur.	w.s.	M
8	o.v	o.v	o.v	o.v	R	o.v	w.s.	R	M	w.s.	W	W	M		
9				←	Not Identified		→								

o.v: other vegetation, R: Rice, F: Forest, u.v: unknown vegetation, ur: urban, B.S.: Bare Soil, T: Tea,
M: Marsh, W: Water, cl: cloud, H: Haze, sh: shadow, w.s: wet soil

Cover Type	Color	Points	Acres	Hectares
Rice	orange	18772	23540	9527
Forest	light green	1632	2047	828
other vegetation	dark green	4136	5187	2099
cloud	white	2044	2563	1037
Haze	gray	172	216	87
shadow	black	-	-	-
Marsh	dark blue	2252	2824	1143
Water	sky blue	5892	7389	2990
Bare Soil	red	652	818	331

Table 9. Distribution of Cover Types, Area 1, June 20, 1975 Data.

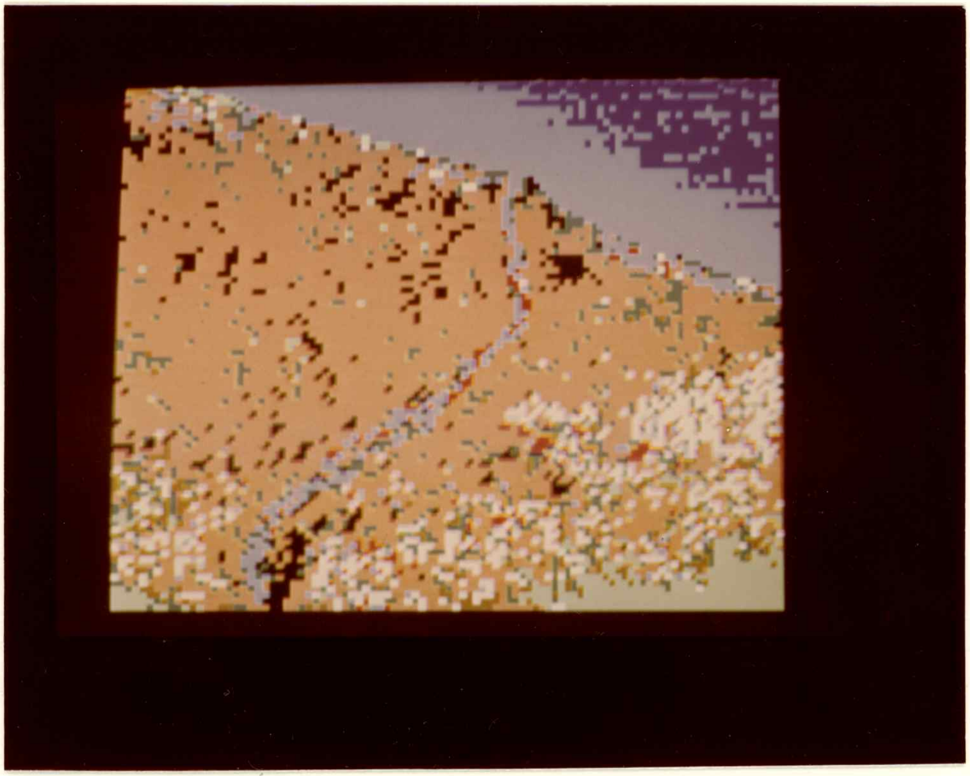


Figure 8. Color Coded Classification Result, Area 1, June 20, 1975 Data.

Cover Type	Color	Points	Acres	Hectares
Rice	orange	37828	47436	19197
Forest	light green	5248	6581	2663
other vegetation	dark green	23880	29946	12119
cloud	white	4976	6240	2525
Haze	gray	268	336	136
shadow	black	-	-	-
Marsh	dark blue	1420	1781	721
Water	sky blue	52268	65544	26525
urban/Bare Soil	red	616	772	313

Table 10. Distribution of Cover Types, Area 2, June 20, 1975 Data.

Figure 9. Color Coded Classification Result, Area 2, June 20, 1975 Data.

Cover Type	Color	Points	Acres	Hectares
Rice	orange	91944	115298	46660
Tea	brown	28964	36321	14699
Forest	light green	38136	47823	19354
other vegetation	dark green	55204	69226	28015
cloud	white	13344	16733	6772
Haze	gray	9888	12400	5018
shadow	black	4	5	2
Marsh	dark blue	732	918	371
Water	sky blue	30580	38347	15519
urban/Bare Soil	red	8308	10418	4216

Table 11. Distribution of Cover Types, Area 3, June 20, 1975 Data.

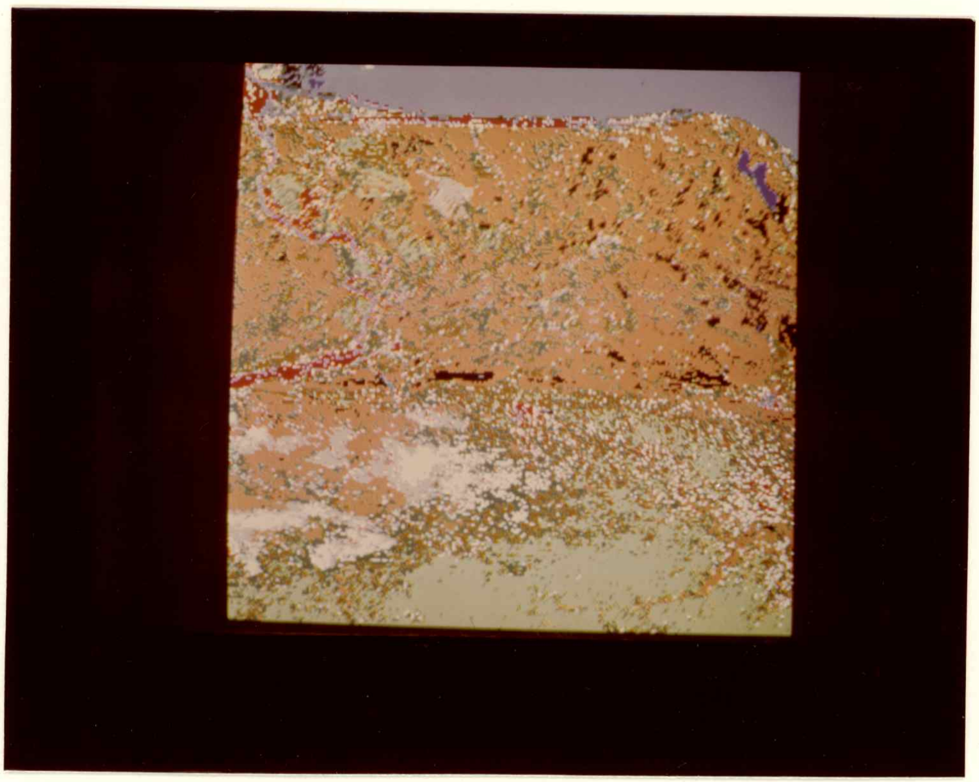


Figure 10. Color Coded Classification Result, Area 3, June 20, 1975 Data.

Cover Type	Color	Points	Acres	Hectares
Rice	orange	164012	205671	83234
Forest	light green	165508	207547	83993
other vegetation	dark green	307452	385545	156028
cloud	white	24452	30663	12409
Haze	gray	17620	22095	8942
shadow	black	104	130	53
Marsh	dark blue	21244	26640	10781
Water	sky blue	72116	90433	36598
urban/Bare Soil	red	30580	38347	15519

Table 12. Distribution of Cover Types, Area 4, June 20, 1975 Data.

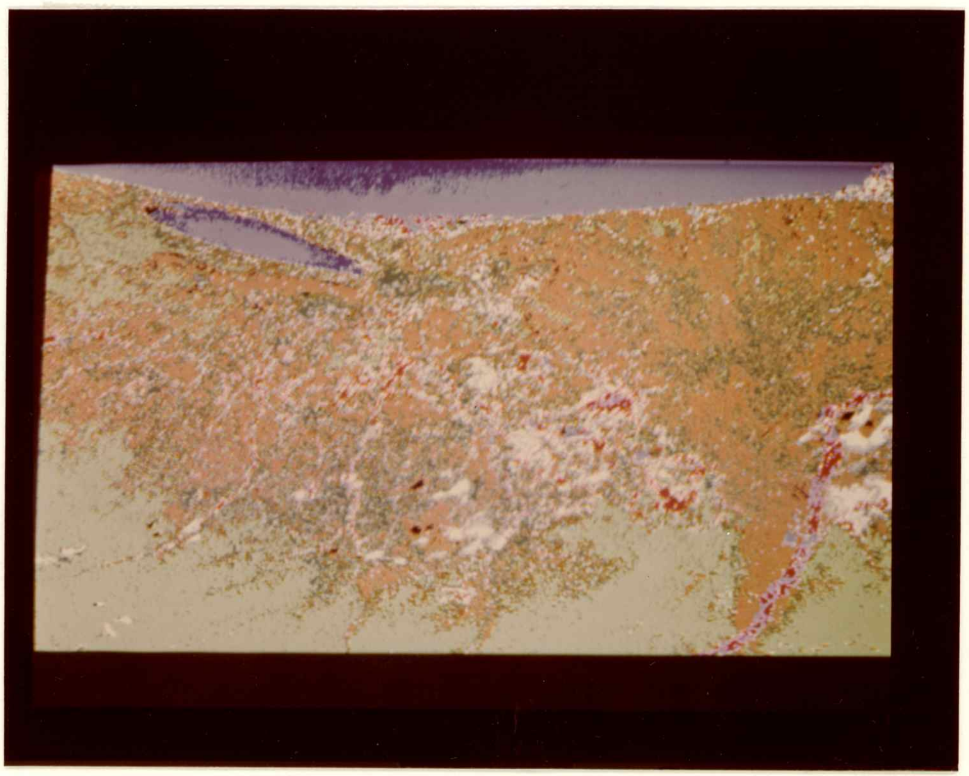


Figure 11. Color Coded Classification Result, Area 4, June 20, 1975 Data.

for this area (14,000 hectares). Most tea plantations occur on the slopes of hills. Therefore, in order to obtain a correct estimate of the area under cultivation of tea from satellite data a correction must be made for the slope. However, in this case because of a lack of information as to the slopes in the tea plantation areas, this was not done.

The estimate of area under cultivation of rice for this new classification requires some discussion and interpretation. The main problem is that in this LANDSAT frame, there are clouds, haze and shadows covering large parts of primarily rice producing areas. For example, immediately east of the Sefid Rud (the main river in the area), south of the point where it turns sharply to the east. With LANDSAT data the crops underneath the clouds cannot be mapped. In some cases when a cloud is very thin (haze) the shadow it casts does not completely mask the spectral response of the crop and the response of the two things combined can be mapped as that crop. In most cases though, the shadow is of such low reflectance that it can only be mapped as shadow. Because almost all the clouds, shadows and haze in this LANDSAT frame were over areas that had been mostly classified as rice in the previous analysis (August 31 data), all points classified into those categories in the present analysis were counted as being rice. Also there was one mixed class of rice/shadow that was counted as rice when the area estimate for rice was tallied. This gave an estimate of 196,000 hectares, 148,000 for rice, 37,000 for clouds, shadows, haze, and 11,000 for rice/shadow (Tables 9-12). This figure is quite close to the estimate obtained in the previous analysis.

Upon examination of the color coded classification maps of of the four subareas (Figures 8-11) several changes from the photos of the classification of the August 31, 1975 data are noticeable. First less spatial detail is apparent, although the same basic patterns appear. The homogeneity of the rice and other vegetation areas has been lost. This is to be expected when comparing the result from a date on which rice and other vegetation were similar in color to a date when they were quite different in color. When two cover types are similar in color only slight changes in the arrangement or percent ground cover of one of the cover types under one resolution element can cause the spectral values associated with that resolution element to be more like the spectral values normally associated with the other cover type. Certain classification algorithms that classify groups of points instead of a single point at a time can smooth out this heterogeneous result to look more like the homogeneous one shown earlier.

In both analyses, other ground features such as bare soil, urban, other vegetation, forest, marsh, and water were mapped in the areas classified. However, the areas classified were restricted to rice and tea-producing areas. Therefore, a good area estimate for the entire Province for forest, for example, is not available, because not all of the area under forest in the Province was classified. In addition, other vegetation was not broken into any more specific classes. There are two reasons for this. First, insufficient information was available to locate and identify such more specific classes; second, the objective of the analysis was to map rice and tea; other vegetation was of secondary importance. Although area estimates were obtained for other vegetation, because it is not broken down and because of the smaller amount of effort put toward insuring their correctness, they are not reported. These are also reasons for the apparent difference in the number and location of points classified as forest between the two dates (Figures 4-7 versus 8-11).

CONCLUSIONS

As stated earlier, the area estimates for rice were close to the government estimate. Possible reasons for the difference between the two would include 1) not all of the rice producing areas in the Gilan Province were covered by either LANDSAT frame, approximately 5% was missing; 2) insufficient reference information was available to completely and specifically represent the spectral characteristics of rice; 3) the area being considered is larger than that normally successfully mapped with one set of training classes. The main problem encountered when working with a large area is that Crop A at Location 1 is spectrally similar to Crop B at Location 2 because of differences in soil, precipitation or cultural practices. Again, lack of reference information at the sub-province level played an important role in forcing consideration of such a large area at one time. The same considerations apply to tea which was successfully mapped in the second analysis by limiting the area it was mapped in. By applying the experience gained here many it is felt many other crops could be mapped also. Recommendations for subsequent analyses would include working on one sub-province (or other smaller area) at a time. The success of mapping tea in the second analysis was mainly due to taking this course of action. In addition, it would be very helpful to have any type of current (same year) aerial photography. One type seen in use at LARS had been taken along several 4.5 x 50 kilometer flightlines equally spaced across the area to be mapped. The photography was photointerpreted and used as ground truth to locate training fields, fields of known cover type, in the satellite data. Had that type of data been available for this analysis, much more detail could have been extracted from it.

The technology used in this study has the potential for being very useful for mapping large areas of crops, such as rice, tea, wheat, cotton, and sugar beets. These are crops that normally occur in pure fields that are large enough to be discriminated by LANDSAT. Many other crops such as vegetables, pulses, pumkin and eggplant that occur in small irregularly shaped fields would be difficult to map with currently available technology. Other cover types such as mixed crops (vegetables, oil seeds, and temporary crops mixed with permanent crops) would also pose problems.

One other problem noted in this study is that extensive topographic relief usually results in shadows on the west side (LANDSAT collects data about 10:00 A.M.) of sufficiently high hills, and these shadows can cause difficulty in representing cover types. In Iran most unirrigated areas are in the valleys where they would be affected by this effect.

A consideration that should be elaborated is that this technology is most cost-effective for mapping large areas. Sufficiently small areas, 2000 hectares or less, can probably be done more efficiently with photointerpretive techniques.[6] In summary there are certain things that this technology does well and cost effectively - this analysis is an example of that. However, the technology is not a panacea and one must consider his problem carefully before he can decide how best to solve it.

REFERENCES

1. Bauer, M.E., 1977. "Crop Identification and Area Estimation Over Large Geographic Areas Using Landsat MSS Data." LARS Information Note 012477, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana.

2. Bauer, M.E., 1975. "The Role of Remote Sensing in Determining the Distribution and Yield of Crops." Advances in Agronomy, Vol. 27.

3. NASA Staff. "NASA Landsat Satellite Data User's Handbook." Goddard Space Flight Center, Greenbelt, Maryland.

4. Stockton, J.G., Bauer, M.E., Blair, B.O., and Baumgardner,M.F., 1975. The Use of ERTS-1 Multispectral Imagery for Crop Identification in A Semi-Arid Climate." Ph.D Thesis, LARS Information Note 040775, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana.

5. Draeger, W.C., 1973. "Agricultural Application of ERTS-1 Data." Symposium on Significant Results Obtained From the ERTS-1 Satellite, NASA, Washington, D.C. Vol. I. Section A: 197-204.

6. Hoffer, R.M., and Staff, 1975. "Computer-Aided Analysis of Skylab Multispectral Scanner Data in Mountainous Terrain for Land Use, Forestry, Water Resource, and Geologic Applications." LARS Information Note 121275, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana.

7. Phillips, T.L. (Editor), 1974. "LARSYS User's Manual." Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana.

8. Swain, P.H., 1972. "Pattern Recognition: A Basis for Remote Sensing Data Analysis." LARS Information Note 111572, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana.