Reprinted from

Eleventh International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Quantifying Global Process:
Models, Sensor Systems, and Analytical Methods

June 25 - 27, 1985

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

THE STUDY OF THE NATURAL GEOGRAPHIC DIFFERENCES IN THE COASTAL AREAS OF WATER COVERED PARTS OF MARMARA REGION IN TURKEY WITH THE HELP OF LANDSAT-4 MSS DATA USING AN UNSUPERVISED CLASSIFICATION ALGORITHM WITH EUCLIDEAN DISTANCE

DERYA MAKTAV

Department of Geodesy and Photogrammetry
Istanbul Teknik Universitesi
Ayazaga, Istanbul, TURKEY

ABSTRACT

In this paper the natural geographic differences in the coastal areas of big lakes in Marmara Region in Turkey are investigated. In order to obtain the different classes at the coasts an unsupervised classification algorithm has been developed where the Euclidean distance between pixels is chosen as the distance measure. This algorithm is applied to the data obtained from the Earth Resource Technology Satellite, LANDSAT-4.

Original as well as ratio data are used when applying the classification algorithm.

I. AN INTRODUCTION TO THE REGION INVESTIGATED

Marmara Region which is one of the seven major geographic regions in Turkey is located on north-west part of Turkey between geographic coordinates 26^{o}- 31^{o} longitudes east and 39^{o} - 42^{o} latitudes north and includes Sea of Marmara which connects Black Sea to the Aegean Sea. In this region some big lakes like Lake Ulubat (formerly Lake Apolyont: 156 km^2), Lake Kuş (formerly Lake Manyas: 178 km^2), Iznik Lake (300 km^2), Lake Ömerli, Lakes Büyük Çekmece and Küçük Çekmece and several small lakes are located. Besides these lakes, two rivers: Mustafa Kemal Paşa River and Gönen River, and two straights, Bosphorus and Dardanelles are located in this region. Bosphorus connects Sea of Marmara to Black Sea and Dardanelles connects Sea of Marmara to Aegean Sea.

II. AIM OF THE INVESTIGATION

In the coastal areas of the above mentioned lakes and Sea of Marmara there are natural geographic differences such as sedimentation areas, sandy areas, swamps or rushes, etc. The goal of this investigation is to present the natural geographic differences in these coastal areas in Marmara Region with the help of digital methods using an unsupervised classification algorithm (Maktav,D., 1985). In order to accomplish this goal two imageries recorded with LANDSAT-4 MSS obtained on December 8, 1982 (Marmara-2) and August 5, 1983 (Marmara-1) covering the areas with geographic coordinates 41^{o} 20' - 39^{o} 50' latitudes north and 29^{o} 35' - 27^{o} 30' longitudes east are used (Figure 1).

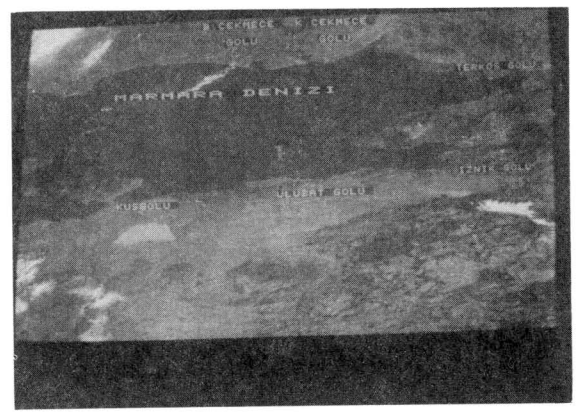


Figure 1. LANDSAT-4 Image of Marmara Region in Turkey (Track Nr: 180, Frame Nr: 32, Date: 8 December,1982).

III. IMAGE PROCESSING SYSTEM AND THE COMPUTER USED

Classifications were carried out with an unsupervised algorithm implemented in the digital image processing system DIDAK, developed at the Institute for Photogrammetry and Topography of the University of Karlsruhe. DIDAK-functions can be grouped in the following classes: Input/Output, preprocessing, geometric rectification, multispectral classification, contrast stretching, display of image data,etc. (Wiesel, J., 1977, pp. 219 - 221).

The computer used is Prime 500 and a Grinnel GMR-27 which is controlled by this computer. The GMR-27 generates displays on a dot-matrix grid consisting of up to 512 horizontal television raster lines, with up to 512 picture elements (pixels) spaced along each line. Image data is transferred to the GMR-27 as picture elements up to 12-bits in depth.

IV. BASIC FEATURES OF THE ALGORITHM PROPOSED

The algorithm developed for this purpose is based on unsupervised classification technique in which the clustering algorithm determines the classes based on the clustering techniques in the remotely sensed data. The algorithm has been coded in FORTRAN-IV (Maktav, D., 1985). This kind of classification requires no training areas which mean without a priori knowledge about the features in an area it is also possible to classify the pixels according to the algorithm developed.

All the pixels in the image are defined as vectors with elements as digital values of spectral reflectances of the pixels in different channels like $x^T = (x_1, x_2, x_3, x_4)^T$ where x_1, x_2, x_3, x_4 are spectral reflectances in four different channels for one pixel.

In the literature several different distance measures between clusters are proposed such as Swain-Fu distance which is the distance between probability density functions and Jeffries-Matusita distance which compares the separation of cluster centers to the dispersion of the data in clusters; and the Euclidean distance which is the point to point distance measure (Swain, P.H., 1972). In this work Euclidean distance is used which is defined for two n-dimensional pixels (or vectors) Y and Z as follows:

$$E = \left(\sum_{i=1}^{n} (Y_i - Z_i)^2\right)^{1/2}$$

In the algorithm the values chosen as parameters are:

i. Maximum number of pixels used to derive fixed cluster centers from temporary cluster centers: MAXPIX.

ii. Maximum number of clusters: MAXSIN.

iii. Euclidean distance between pixels:E.

These parameters can be explained as follows:

i. The first pixel read forms the first temporary center (A) of the first temporary class. The arithmetic mean of the second pixel belonging to the first temporary class and (A) will become the current temporary class center (B) of the first temporary class, that means (A) is moved to (B). The arithmetic mean of (A), (B) and the third pixel belonging to the first temporary class becomes the current temporary center (C) of the first temporary class, that means (B) is moved to (C). Thus the temporary centers are moved until the parameter MAXPIX reaches an a priori given value in the algorithm. If the parameter reaches this a priori given value in the algorithm, the center of the class is "fixed" which means the class will also be fixed and it will not be moved any more. This movement of the temporary cluster centers is done in order to achieve fixibility of all classes (Figure 2).

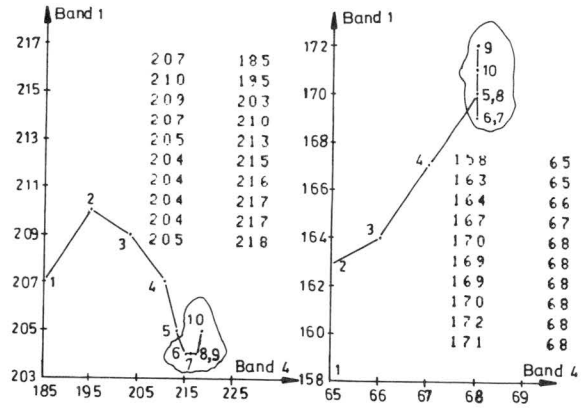


Figure 2. Movement of the temporary class centers for MAXPIX=10.

ii. The maximum number of classes is also chosen as a parameter to avoid the unnecessary spectral information which also increases the CPU-time.

iii. The Euclidean distance is the most important parameter in the algorithm. It controls which pixel belongs to which class. The best way to check the appropriate choice of this parameter is to try to interpret the classified image on the monitor by help of measurements from field work. If the result doesn't satisfy the observer the parameter can be increased or decreased.

In order to find out which pixel belongs to which cluster there are two methods used in the algorithm:

i. Distances between any pixel X and all cluster centers are calculated and the pixel X is assigned to the closest center.

ii. If the measured distance from any pixel X to a cluster center is equal to or smaller than a given parameter value, pixel X is decided to be a member of this cluster. Thus it is not necessary anymore to calculate the distances from X to all cluster centers.

There is not a big difference between the results of the above mentioned two methods, but second method is preferred mostly due to the smaller CPU-time (Maktav, D., 1985).

If the pixel is farther than an a priori defined distance from all cluster centers it would be interpreted as "unclassified pixel". It can be seen that as MAXSIN increases, i.e. as the number of cluster centers increases, the number of unclassified pixels will decrease.

V. USE OF RATIO DATA

In unsupervised classification algorithm not only original but also ratio data are used to avoid haze, illumination and topographic effects. To prepare the ratio images the following relationships are calculated (Odegaard, H.A., Ottesen, R., 1980, pp. 723 - 733; Fasler, F., Itten, K.I., Staenz, K., 1982, pp. TP-9/1.1 - 1.5)

- The digital number in one band is divided by the corresponding digital number in another band for each pixel:z.

- The difference between two bands is divided by their sum:z.

To obtain the values of the ratio data in the range 0-256, the following formulea are used in the algorithm:

$$\text{if } z \geq 1 \longrightarrow z' = 256 - 128 / z \geq 128$$

$$\text{if } z < 1 \longrightarrow z' = 128 \cdot z < 128$$

VI. APPLICATION OF THE ALGORITHM TO THE COASTAL AREAS OF THE MARMARA REGION

7,461,504 pixels (3264 columns, 2286 rows)are used in the whole area (for each band). Four histograms are given below which show the spectral distribution of the spectral data of the Marmara Region all covered in one tape. Here especially the distributions with two apexes in bands 3 and 4 show the spectral separation of land/water. Besides, in bands 1,2 and 3 (Marmara-2) there are small groups with high reflectances towards the right tails of the histograms. These groups represent the clouds in the image and cover 20% of the whole digital data (Figure 3).

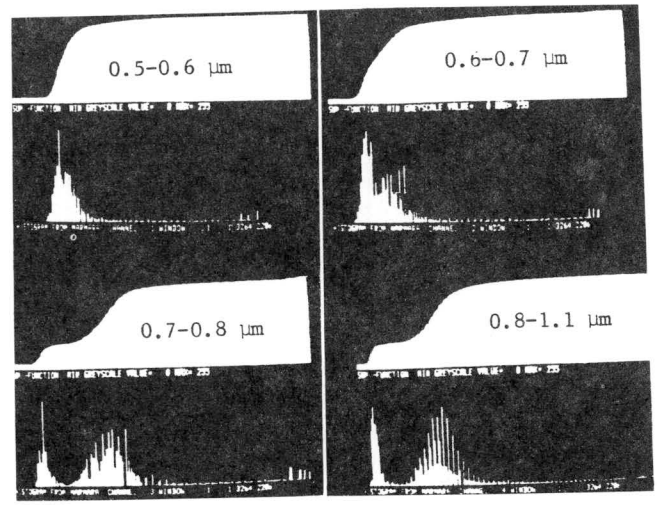


Figure 3. Histograms of four bands for 7,461,504 pixel area in Marmara Region (photographed from monitor screen).

A. COASTS OF LAKE ULUBAT

A 10-day field research at the northern coasts of Lake Ulubat showed that the class "water" can be clearly distinguished from land. The same result was obtained with the LANDSAT data using the proposed unsupervised classification algorithm.

But at the southern coasts of Lake Ulubat some classes can be distinguished such as muddy water, alluvion covered areas and swamps. These different classes, their potential distribution, number of pixels in each class and the values of the parameters MAXPIX, MAXSIN, ExE are presented below in Figure 4 and Table 1.

Table 1. Classes at the southern coasts of Lake Ulubat (Bands 1,3,4).

(Only for water covered areas)	Number of classified pixels	Percentage of the pixels (%)	Interpretation
black	41050	55.2	Sea of Marmara (Salty water)
light grey	28092	37.8	muddy water
white	5192	7.0	swamps
	74334	100.0	
MAXPIX=10,MAXSIN=30,ExE=200			

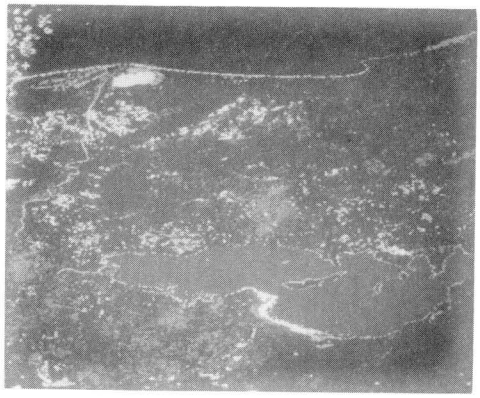


Figure 4. Unsupervised classification at the southern coasts of Lake Ulubat.

If MAXSIN is increased to 40 and ExE decreased to 100the classified image (2-times magnified) of the southwestern coasts of the Lake shows the distinction between sand and swamp. This is given in Figure 5. Dark grey represents the sand and light grey the swamp.

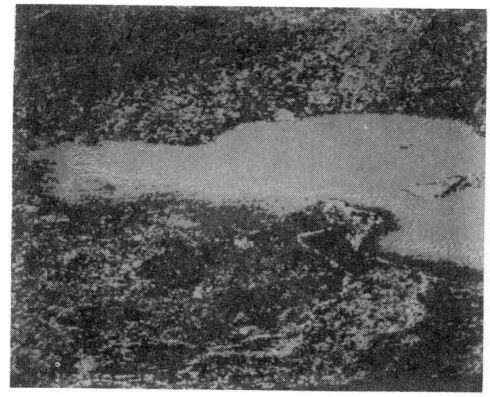


Figure 5. Unsupervised classification results at the southern coasts of Lake Ulubat.

B. COASTS OF LAKE IZNIK

A great portion but not the whole of Lake Iznik is included in the tape used for this research. The results of the classification compared to the field research show that at the northern coasts of Lake Iznik there are about 2 meters high rushes which cover 3% of the image of the portion of Iznik Lake available on this tape. A photograph of these rushy areas (photographed in April 1984) and the classified image of these areas are presented in Figure 7 within rectangles. The chosen parameters are ExE=100 and MAXSIN=40.

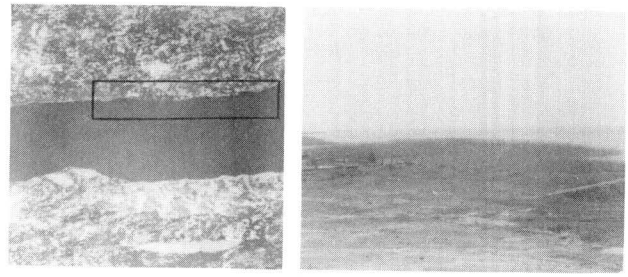
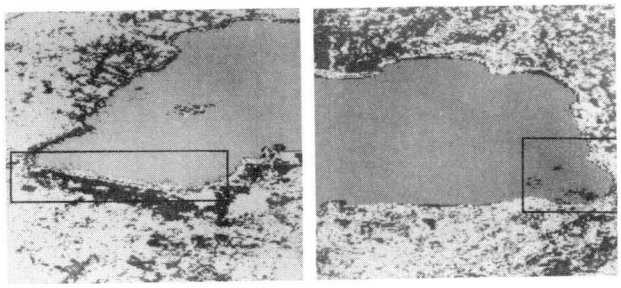


Figure 7. Unsupervised classification at Lake Iznik and the photograph of the same area.

C. COASTS OF LAKE KUŞ

The third biggest lake of this area is the Lake Kuş, which lies about 36 km east of Lake Ulubat. At the southeastern coasts between Dereboğazı and Kocagöl sandy areas and at the eastern coasts of the Lake between Yılgınburnu and Ergili swamps are to

be seen in Figures 8(a) and (b) within rectangles respectively.



(a) (b)

Figure 8. Unsupervised classification at the coasts of the Lake Kuş. (a): ExE=100, MAXSIN=40, bands=1,2,4; (b):ExE=75, MAXSIN=50, bands=1,2,4).

D. BAY OF ERDEK

Gönen River comes from Edremit Bay and joins the Sea of Marmara in Çayağzı at the Erdek Bay. The hell grey color shows the sedimentatiton areas of River Gönen in east and west direction (ExE=100, MAXSIN=40) in Figure 9.

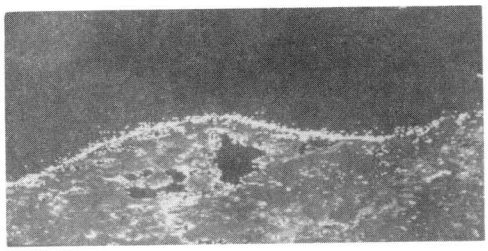


Figure 9. Unsupervised classification at Bay of Erdek.

VII. COMPUTATIONAL ASPECTS OF THE ALGORITHM

During the application of the classification algorithm a relationship between parameters and CPU-time has been observed.

The effect of the number of bands on CPU-time is presented in Table 2. According to this Table the CPU-time increases by 66% when the number of bands increases from 2 to 3, where MAXSIN is kept constant at 10.

Table 2. The effect of number of bands on CPU-time.

Number of bands	ExE	MAXSIN	Unclass. pixels	CPU-t (sec)	Increase of CPU-t (%)
2	50	10	87012	114.96	66
3	50	10	104082	190.84	
2	100	10	25920	104.98	67
3	100	10	40983	175.36	
2	150	10	7502	83,59	65
3	150	10	19687	137.92	

The effect of Euclidean distance on CPU-time is given in Table 3. According to it CPU-time increases by~10% when ExE decreases by increments of 50.

Table 3. The effect of Euclidean distance on CPU-time.

Number of bands	ExE	MAXSIN	Unclass. pixels	CPU-t (sec)	Increase of CPU-t
3	150	10	20010	158.14	10.9
	100	10	40983	175.36	8.8
	50	10	104082	190.84	
	200	30	12703	153.01	8.9
	150	30	16072	166.63	11.8
	100	30	30263	186.27	
	200	40	9984	150.16	12.0
	150	40	11092	168.13	6.0
	100	40	24984	178.14	

When choosing the parameters the interpretation of the image is the most important criterion but it is also important not to increase the CPU-time by obtaining unnecessary information.

VII. RESULTS

The results obtained by the unsupervised classification algorithm are compared to the field study in the same region and to a map at scale of 1:200,000. In northern coasts of Lake Iznik it is possible to interpret areas covered with about 2 meters high rushes. In the eastern part of Lake Kuş swamps and in the southwestern coasts of the Lake Kuş sandy areas are obtained. Because of the alluvium thrown down by rivers into the southern coasts of the salty Sea of Marmara (in Erdek Bay) and into the southwestern coasts of Lake Ulubat, sedimentation areas are interpreted. In addition to this, the course of the propagation of this alluvium areas are also obtained by the use of the classified image. In the coastal areas of Lake Ömerli, on the northern part of Marmara Region, water/land separation is very clear.

At the outset of this research there were some doubts whether LANDSAT MSS data, which actually are used for earth resource research purposes could be employed for investigating natural geographic differences in coastal areas. But as has been reported above satisfactory results have been obtained. It is expected though that better results are to be expected when the proposed unsupervised classification algorithm is applied to Seaset or Nimbus CZCS data. Such a comparison can be an interesting extension of this research.

Classification results obtained from the tests with the ratio images don't add any further information to the above results obtained from original data and thus are not presented here in detail.

REFERENCES

Fasler,F., Itten, K.I., Staenz, K.: Improved Landuse Classification Through Principal Component Analysis Based On Category Statistics And Synthetic Variables,IEEE Digest Volume 1, 1982 International Geoscience and Remote Sensing Symposium (IGARSS '82), June 1-4, 1982 Munich), TP-9/1.1-1.5.

Maktav, D.: Study Of The Coastal Areas In Marmara Region With The Help Of LANDSAT-4 Original And Ratio Data (in Turkish), Unpublished Ph.D. Thesis, Technical University of Istanbul, Turkey 1985, 100 p.

Odegaard, H.A., Ottesen, R.: Mapping Polar Vegetation At Spitsbergen Bye Use of LANDSAT, Proc. of the ISP, Com. VII, Vol. XXIII, Part B8, Hamburg, 1980, pp. 723 - 733.

Swain, P.H.: Pattern Recognition: A Basis For Remote Sensing Analysis, LARS information Note 111572, Purdue University, Lafayette 1972.

Wiesel, J.: DIDAK, A Digital Image Processing, Proc. of the Int.Symp. on Image Processing, Interactions with Photogrammetry and Remote Sensing, Graz 3-5 October 1977, pp. 219 - 221.

Derya Maktav. The author received the Degree of B.Sc. from the Technical University of İstanbul (İ.T.Ü.), Faculty of Civil Engineering, Department of Geodesy and Photogrammetry in 1975. Then he spent a year at the University College London, Department of Photogrammetry and Surveying as student. He received the M.Sc. degree from the University of Karlsruhe in W.Germany in 1979. Later he received his Doctor's degree from İ.T.Ü. in 1985. He is 34 years old and has served as instructor at İ.T.Ü. since 1979.