AGRICULTURAL LAND-COVER AND FOREST CLASSIFICATION OF LANDSAT DATA IN THE PROVINCE OF BOLU

A.Nejat EVSAHIBIOĞLU
Assoc.Prof.Dr.
The University of Ankara
Agricultural Engineering
Department, TURKEY
ISPRS Commission No: IV

Mehmet TANKUT
Manager of Remote Sensing and GIS
AS-IŞLEM Remote Sensing
Engineering Co.Ltd.,
Ankara, TURKEY

ABSTRACT :

The main purpose of this study was to analyze LANDSAT TM data by computer processing for mapping of Agricultural and Forest cover types in the province of Bolu, Turkey. Primary LANDSAT 5 TM data used for evaluation was portion of scene Path 178, Row 32 (Center geographic coordinates N 40°20'00" E 31°43'00") acquired July 16, 1984. Reference data used to support the analysis was consisted of identification of ground observation and recording on large-scale map the crop type and land-use of necessary number of fields in the test areas. Supervised and Unsupervised approaches were utilized for classification. Digital analysis was performed using ERDAS image processing system. Classification schemes were evaluated using contingency tables and were ranked using KAPPA statistics. The study results indicate Supervised Classification technique was more accurate than Unsupervised Clustering over the study area. Maximum Likelihood classification method was superior to any digital classification approach for the area. The study results also display that classification accuracy is dependent on the selection of a particular classification scheme and that remote sensing techniques utilizing computer-aided image analysis methods can be used to identify the Agricultural Land-Cover and Forest pattern in the study area.

KEY WORDS: Classification, Image analysis, Image Processing, Landsat, Remote Sensing Application.

1. INTRODUCTION

Sound Land Management requires timely and accurate information about the type, amount, availability and condition of renewable resources being produced (DeGloria and Benson, 1987).

In all disciplines involving land management there exist a need for timely, reliable information on which to base resource management decisions. One of the most important types of resource information required for comprehensive planning is a current data base of the vegetation/land/water surface cover, subsequently referred to as "cover type" (Fleming, 1988).

Land-cover are used for many purposes. In land consolidation projects and in environmental and hydrological studies, accurate, up to data information about land-cover on a regional scale is often required. Knowledge of changes in land-cover is becoming increasingly important from both the ecological and economical point of view (Janssen et al., 1990).

Particularly in the last twenty years, remote sensing employing digital Landsat data has developed at a rapid pace. It has became a practical tool for monitoring the environment and assessing our natural resources in a number of application areas. Nevertheless, computer-generated land-cover classifications require significant improvement in both their accuracy and specificity in order to be used operationally in many applications. One faced to the solution of this problem is to improve the quality of the raw data. This has been initiated with the launch of the Landsat Thematic Mapper and SPOT satellites (Lo et al., 1986). Equally important is to decide the most proper methods to analyze and classify the data for a particular region.

On the other hand, timely and accurate knowledge of forest composition and condition can be an invaluable forest management tool. The ability of remote sensing analysis to augment traditional forest resource evaluation procedures has been demonstrated by researchers for a diverse range of applications, imagery types and analysis techniques. Satellite imagery is also recognized as a useful tool for forest mapping and inventory (Buchheim et al., 1985; Borry et al., 1990).

The objective of this study was to evaluate several classification schemes for identifying and mapping agricultural and forest cover types in the province of Bolu, Turkey from Landsat TM data.

2. LITERATURE REVIEW

The use of remote sensing for the investigation of terrestrial resources has become increasingly common since the launching of the first Landsat satellite in 1972. Conventional aerial photographs have of course been of great value for many years, and continue to provide important data. However, remotely sensed data from satellites have a number of important advantages leading to their increasingly wide adoption (Townshend, 1981).

Since Landsat MSS data became available, many efforts have been made to demonstrate the utility of remotely sensed data is developing land-use and crop area statistics on regional and national level (Holko and Sigman, 1984).

Gautam and Chennaiah (1985) have oriented their efforts toward studying the changes in land-use and land-cover in Tripura, India, using LANDSAT images of two different dates and to see how well data obtained help in the study of geographical phenomena with special reference to land-use and cover. Ninety percent accuracy of each land-use category has been achieved when compared with existing data compiled on ground surveys by the working plan of the Division of Forest Department.

Many studies for developing land-cover and forest information have shown that digital processing of Landsat MSS data requires some interpretation and/or integration using aerial photography and/or some type of ground observations (Hoffer and Staff, 1975). Most reports on forestry applications of TM imagery have used Thematic Mapper Simulator (TMS) data (Franklin, 1986; Vogelmann and Rock, 1986). In general, TM-type data have been found superior to MSS imagery (Horter and Ahern, 1986).

Latty and Hoffer (1981) studied the utility of TM spectral bands for a site in South Carolina using TMS data. They analyzed the statistical separability of spectral classes using various TMS spectral band combinations. Their results showed high separability between a number of forest classes.

Williams and Nelson (1986) report the results of a North Carolina study where substantial classification improvements were obtained with TMS data (relative to MSS data) for seven forest classes. The overall performance of TMS classification was 60 percent as compared to 39 percent for MSS data.

Benson and DeGloria (1985) present the results from the use of actual TM data in forestry application in Carolina. They stated that the best TM band combination was better than the MSS data and the best results indicated that TM data will provide higher classification.

The paper of Jones et al. (1988) describes the complementary use of digital terrain information and SPOT-1 HRV multispectral imagery for the study and mapping of semi-natural upland vegetation. A digital terrain model was derived for a study area in Snowdonia, Wales, and was used to generate slope and aspect images.

Borry et al. (1990) studied on the assessment of the value of monotemporal SPOT-1 Imagery for forestry applications under Flemish conditions. Research results revealed that the enhancement procedure for visual interpretation is of minor importance when compared to the acquisition data. The advantage of visual analysis is lost if detailed levels of forest information are required.

3. MATERIAL AND METHODS

3.1. Study Area

The study area chosen for analysis is located within the Abant forest and the Aladağ forest area in the eastern part of Bolu province. The study site is approximately 943 square kilometers (364 square miles) (Figure 1).

This area is situated on latitude N 40°42' and longitude E 31°28' with an altitude 738 meters above the sea level. The area has highly productive, level to rolling terrain and intensively vegetated.

The region has semiarid mesothermal climate with dry summers and cold winters. The coldest month is January with -4.6 °C mean and the warmest August with 39.4 °C . Annual average relative humidity is 74%. Mean annual precipitation is 529.2 milimeters, mostly taking place between late September and late June.

Soils of study area are generally within Brown Great Soil Group. Soil texture is mainly heavy, namely clay and loamy-clay.

Primary land-cover categories in the study area are Forests, Water bodies, Agricultural lands, Non-vegetated areas. Almost 70% of the total study area is covered by forest. Water bodies, namely lakes and reservoirs is 0.3 %. Agricultural lands, settlements and other areas constitute almost 29% of the study site.

3.2. Data Acquisition and Preparation

Cloud-free digital LANDSAT-5 TM data were acquired over the study area (Path 178 and Row 32) on 16 July 1984. Center geographic coordinates of the image is N 40°20' and E 31°43'. They became available on CCTs of BSQ format (radiometrically and geometrically calibrated). Reference data used to support the analysis was consisted of identification by ground observation and recording on maps with a scale of 1/25 000 the land-cover type of necessary number of fields in the test areas. Topographic land-cover maps with a scale of 1/100 000 were also used for geometric rectification of the image and the verification and accuracy assessment of the digital natural resources classification together with 1/25 000 scaled maps.

Applying special algorithms available on the ERDAS image processing system installed at the Remote Sensing Laboratory in the Ankara University, Agricultural Engineering Department, a subscene covering the study area was extracted from TM tapes and loaded on to floppies for further analysis.

The extracted subscene was then rectified to a state plane map projection. The implementation of the rectification process was based on the image and map coordinates of 14 control points uniformly scattered throughout the study area. The map coordinates were acquired by using a digitizing table and the appropriate routines of the available software. The set of the acquired coordinates were used to compute the coefficients of the transformation matrix using a least square algorithm. The nearest-neighbour interpolation was used to rectify the input image.

3.3. Field Work

Ground data provide indispensable support in the interpretation of remotely sensed data and are helpful in the verification obtained from actual site of study and in areas of particular interest such as agriculture, forestry, water bodies and other land-use categories (Gautam and Chennaiah, 1985). Therefore ground data collection was carried out in the study site. It involved the development of an overall systematic plan so that all the selected sites could be visited. Intensive study was carried out to collect data from the maximum number of representative areas of different land-cover type. Combining the information of land-cover gathered by field work with land-cover map on 1/100 000 scale, percentages and acreage of each land-cover category related to total area have been obtained.

3.4. Data Analysis

Preliminary visual interpretation for the delineation of land-cover classes was performed on 1/50 000 scale screen display of LANDSAT multispectral imagery with the help of a false color composite (Bands 4 red, 3 green and 2 blue).

In this study it was evaluated a number of biomass transformations, or vegetation indices, in order to emphasize the relative greenness of the land-cover classes of interest. This evaluation was done on a representative subset of test data, and it was determined that the Radiance Ratio TM 4 / TM 3 and the Normalized Difference Vegetation Index Transformation [(TM 4 - TM 3)/(TM 4 + TM 3)]

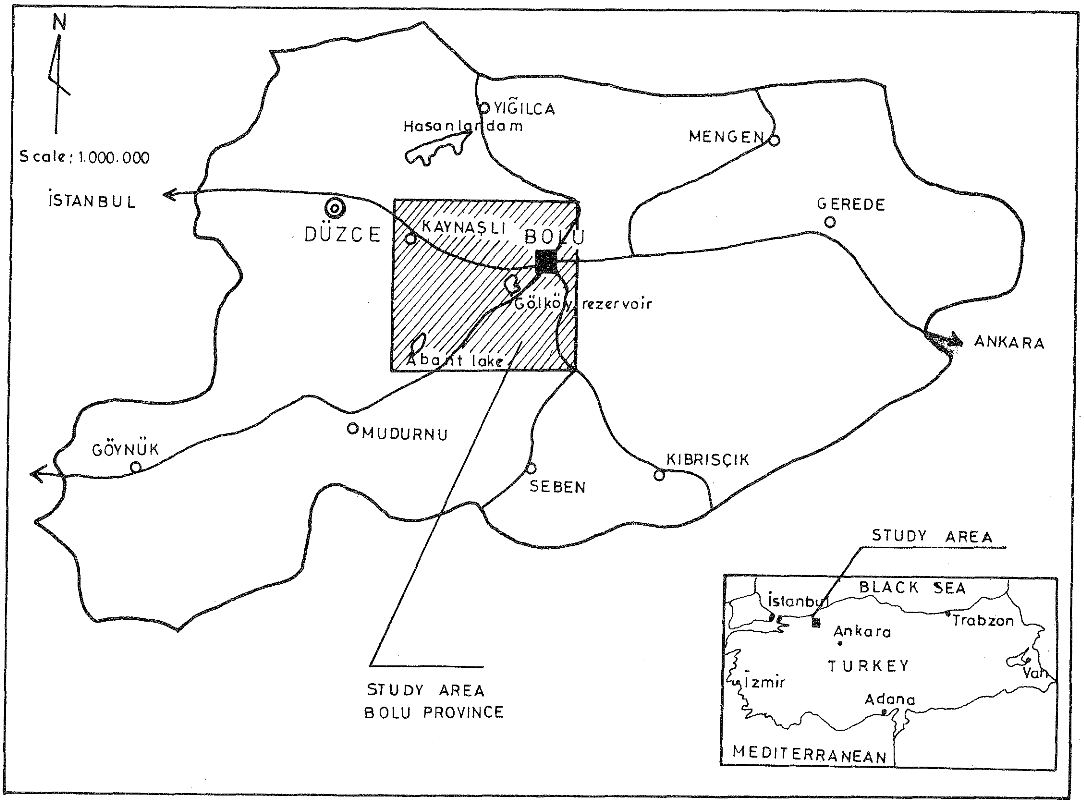

Figure 1. Map of Bolu province showing the location of the study area

provided the most useful temporal discrimination for the study area. By doing so it was aimed to dispose the vegetation cover in overall image in order to enhance the ability of visual interpretation.

Selection of multiple band combination was based on the decision to test TM data which are representative of the three major spectral regions, namely Near Infrared, Band 4; Red, Band 3; Green, Band 2 (Karteris, 1990).

Supervised and unsupervised classification techniques were performed for spectral pattern recognition of LANDSAT TM multispectral data by using ERDAS software package.

Supervised classification is based on the selection and use of various training data sets. The size and number of these data for each category are dependent mainly upon the spectral variability within that category throughout the study area and should be unique in terms of effectively defining the category. In this study 28 training sets were selected by displaying a composite of bands 4, 3 and 2. All these sets were located well within the boundaries of the corresponding categories. The training data were processed statistically and spectral signature files containing the means, the standard deviations and the variance-covariance matrices were generated for each category and used as input to the classifier for the classification of the whole study area (1.047.552 pixel in total). Classification stage was performed by using two different approaches; Gaussian maximum likelihood classifier and minimum-distance to means classifier. Output stage was presented in the form of tables of area statistics.

An alternative approach called unsupervised clustering was also carried out for the study area.

An accuracy assessment was performed to provide some degree of confidence to the classification results. Overall accuracy of each individual category were calculated for each set of analyzed data. The procedure was accomplished by overlaying and registering on the projected classification images the land-cover map and then collecting the required data. Data collection was done by stratifying the area into the classification categories and conducting a random sampling of points within each of the stratified land-cover categories.

Evaluation of spectral separability was provided by a confusion matrix for the purpose of large area accuracy analysis of study area data set that are different from, and considerably more extensive than, the training area (Lillesand and Kiefer, 1979). From this information, classification error of omission and commission was studied. Accuracy estimates based only upon diagonal elements of these tables may produce inflated accuracies. Therefore, the Kappa statistic was used as a coefficient of agreement since it corrects for chance agreements and accounts for errors for both omission and commission (Hudson, 1987).

4. RESULTS AND DISCUSSION

4.1. Land-Cover Pattern

According to the information subtracted from land-cover map of the study area, highlighting the land-cover pattern and its type existing at the time imagery was obtained, land-cover categories were determined for assistance to classification procedures. The area and percentages of each land-cover category is given in Table 1.

The study site covers an area of 94 280 hectares. As a matter of fact 73 % of the non-irrigated agricultural lands is devoted to Wheat and Barley production, which is 20 560 hectares. Since the ambient temperatures of 1984 plant growing season are generally higher than previous years, cereals namely wheat and barley were harvested before the LANDSAT image was acquired. For this reason, most of the agricultural lands of the study area on the imagery was occupied by fallow. On the other hand, 87 % of the irrigated lands over the study area is given for potato production. Remaining 13% of the irrigated area is mainly covered by Sugar beets (12 %) and 1 % is left for legumes. For this reason sugar beets and legumes were considered into the same category in order to improve the classification efficiency.

Nonvegetated areas was also considered in the same category with Fallow.

Consequently land-cover/use classes for machine-assisted image processing were obtained as shown in Table 1.

4.2. Classification Accuracy Assessment

Results from supervised classification schemes (maximum likelihood and minimum-distance to means approaches) and unsupervised clustering are summarized in contingency tables 2, 3 and 4. Accuracy assessments results, ranked according to Percent Correct, Commission Error and Kappa statistics are presented in Table 5.

With respect to Bolu study area maximum likelihood classifier provided the most accurate results. This method has given the highest Percent Correct value of 86.6 and the lowest Commission Error of 13.4. The greatest Kappa value of 0.809 is another indicator that reveals the maximum likelihood approach ranked the best classification scheme among the others for this study. Minimum-distance to means classifier has given a Percent Correct value of 82.4, Commission Error of 17.6 and Kappa value of 0.751. These accuracy symptoms ranked the minimum-distance to means classification technique second after maximum likelihood. Intuitively, it would seem that the more bands use in a maximum likelihood classifier, the better the classification results would be. Nevertheless Karteris (1990) reports that the basic remote sensing information tool may be considered to be the three-band combination (colour composites, other three-band sensors, less cost of acquiring and analyzing the data, good classification results etc.). Owing to the above, he considered that a comparison between the recorded accuracies of six-band and three-band combinations for each individual category would provide useful information. In most cases the difference in accuracy between them was negligible (a maximum of 3.9 percent). These findings encourage the use of three-band combinations in natural resources classifications and forest mapping projects. However he also stated that band 4 should be included in all the three-band combinations.

In this study it was thought that it may be reason of why supervised classification schemes used three band combinations (Band 4, Band 3 and Band 2) have given high classification accuracies.

Unsupervised clustering has displayed the lowest Percent Correct of 73.1, Commission Error of 26.9 and Kappa statistics of 0.617. According to these consequences unsupervised clustering was not so successful as being to other schemes to classify the study area of Bolu province. The reason for this circumstances may be resulted from the relatively large stand sizes and pure cover types provided sufficient training site statistics to characterize the existing study area cover adequately.

4.3. Area Estimation Accuracy

The classification performance indicated by 80 % correct recognition of test fields is believed to be adequate for satisfactorily estimating crop areas (Bauer et al., 1979). Therefore in this study Percent Correct value of 82.4 on the basis of maximum likelihood classifier for overall study area can be assumed acceptable percentage of correct recognition to estimate land-cover/use acreage. Table 6 presents the comparison of land-cover class area percentages with land-cover class area estimates based on classification of LANDSAT data for the study area of Bolu.

Table 1. The Area and Percentages of Each Land-Cover Category in The Study Area of Bolu, July 1984.

Category No	Category Name	Area (ha)	Percent of total area
1	Coniferous	38 800	41.15
2	Deciduous	26 800	28.43
3	Water	300	0.32
4	Potatoes	5 405	5.73
5	Sugar beets + Legumes	1 595	1.69
6	Fallow + Nonvegetated	21 380	22.68
	TOTAL	94 280	100.00

Table 2. LANDSAT TM Classification Performance, Maximum Likelihood Classifier, Bolu Province Test Site

Cover Type (Map)	LANDSAT Classification C.	D.	W.	P.	S.+L.	F.+N.	TOTAL	PERCENT CORRECT	OMISSION ERROR	COMMISSION ERROR
Coniferous	3414	183	0	120	14	66	3797	89.9	10.1	10.1
Deciduous	151	2221	0	78	53	90	2593	85.7	14.3	15.3
Water	0	0	50	2	2	1	55	90.9	9.1	9.1
Potatoes *	52	16	2	438	16	4	528	83.0	17.0	32.6
Sugar beets + Legumes	13	7	2	2	132	1	157	84.1	15.9	54.6
Fallow + Nonvegetative	166	115	1	10	74	1764	2130	82.8	17.2	8.4
TOTAL	3796	2542	55	650	291	1926	9260			
PERCENT CORRENT	89.9	87.4	90.9	67.4	45.4	91.6		86.6		13.4

Table 3. LANDSAT TM Classification Performance, Minimum-Distance To Means Classifier, Bolu Province Test Site

Cover Type (Map)	LANDSAT Classification C.	D.	W.	P.	S.+L.	F.+N.	TOTAL	PERCENT CORRECT	OMISSION ERROR	COMMISSION ERROR
Coniferous	3253	236	4	161	28	115	3797	85.7	14.3	12.2
Deciduous	183	2074	3	112	58	163	2593	80.0	20.0	16.7
Water	1	1	48	2	2	1	55	87.3	12.7	26.2
Potatoes	57	22	5	427	11	6	528	80.9	19.1	41.2
Sugar beets + Legumes	16	9	1	4	124	3	157	79.0	21.0	56.5
Fallow + Nonvegetative	193	147	4	20	62	1704	2130	80.0	20.0	14.5
TOTAL	3703	2489	65	726	285	1992	9260			
PERCENT CORRENT	87.8	83.3	73.8	58.8	43.5	85.5		82.4		17.6

Table 4. LANDSAT TM Classification Performance, Unsupervised Clustering, Bolu Province Test Side

Cover Type (Map)	LANDSAT Classification									
	C.	D.	W.	P.	S.+L.	F.+N.	TOTAL	PERCENT CORRECT	OMISSION ERROR	COMISSION ERROR
Coniferous	3029	357	6	179	112	114	3797	79.8	20.2	21.1
Deciduous	217	2018	4	96	90	168	2593	77.8	22.2	27.3
Water	4	3	42	3	2	1	55	76.4	23.6	37.3
Potatoes	76	44	6	352	17	33	528	66.7	33.3	46.6
Sugar beets + Legumes	22	8	3	5	109	10	157	69.4	30.6	71.2
Fallow + Nonvegetative	489	344	6	24	48	1219	2130	57.2	42.8	21.1
TOTAL	3837	2774	67	659	378	1545	9260			
PERCENT CORRENT	78.9	72.7	62.7	53.4	28.8	78.9		73.1		26.9

Table 5. Summary of LANDSAT TM Classification Performance, Bolu Province Test Site

Rank	Classification scheme	Percent Correct	Commission Error	Kappa
1	Maximum Likelihood	86.6	13.4	0.809
2	Minimum-Distance To Means	82.4	17.6	0.751
3	Unsupervised Clustering	73.1	26.9	0.617

Table 6. Comparison of Class Area Percentages with Area Estimates Based on Classification of LANDSAT TM Data for Bolu

Class No	Land-cover Class	Based on Map (Percent)	LANDSAT Classification Estimates (Percent)	Difference (Percent)
1	Coniferous	41.15	41.14	+ 0.01
2	Deciduous	28.43	27.87	+ 0.56
3	Water	0.32	0.32	0.00
4	Potatoes	5.73	7.05	- 1.32
5	Sugar beets + Legumes	1.69	3.11	- 1.42
6	Fallow + Nonvegetated	22.68	20.51	+ 2.17

As seen on the table, class area percentage values based on map are very close to those based on LANDSAT classification estimates. Differences between class values are changed from -1.42 % to +2.17 % . In fact, statistically F-Distribution test was applied to both group values in order to control the homogeneity of variances. By doing this, it was aimed to understand the significance level of difference between group values. For this purpose variances of groups were divided each other and resulting F value was compared to the pre-determined ones tabulated for the levels 5 % and 1 % (Düzgüneş et al., 1983). Results indicated that no significant difference between group values exists. In other words, both groups represent the same population. According to this result it can be concluded that LANDSAT TM classification estimates based on maximum likelihood classifier may be used to determine the land-cover class areas for the study area of Bolu Province.

It can also be deduced that the utility of TM data appears to be very high and successful for natural resources classification and should be more widely used in various applications. Furthermore, use of three band combination (4,3,2) yields information with an accuracy acceptable for agriculture and forestry purposes and at lower cost.

REFERENCES

Bauer,M.E., J.E.Cipra, P.E.Anuta and J.B.Etheridge. 1979. Identification and area estimation of agricultural crops by computer classification of LANDSAT MSS data. Remote Sens. Environ. 8:77-92.

Benson,A.S., and S.D.DeGloria, 1985. Interpretation of landsat-4 thematic mapper and multispectral scanner data for forest surveys. Photogrammetric Engineering and Remote Sensing, 51(9):1281-1289.

Borry,F.C., B.P. De Roover, R.R. De Wulf, and R.E.Goossens, 1990. Assessing the value of monotemporal SPOT-1 imagery for forestry applications under Flemish conditions. Photogrammetric Engineering and Remote Sensing 56(8):1147-1153.

Buchheim,M.P., A.L.Maclean, and T.M.Lillesand, 1985. Forest cover type mapping and spruce budworm defoliation detection using simulated SPOT imagery. Photogrammetric Engineering and Remote Sensing. 51(8):1115-1122.

DeGloria, S.D., and A.S.Benson, 1987. Interpretability of advanced SPOT film products for forest and agricultural survey. Photogrammetric Engineering and Remote Sensing. 53(1):37-44.

Düzgüneş,O., T.Kesici and F.Gürbüz, 1983. İstatistik Metodları. Ankara Üniversitesi Ziraat Fakültesi Yayınları:861, Ders Kitabı:229, Ankara Üniversitesi Basımevi, Ankara, 218 p.

Fleming, M.D., 1988. An integrated approach for automated cover-type mapping of large inaccessible areas in Alaska. Photogrammetric Engineering and Remote Sensing. 54(3):357-362.

Franklin, J., 1986. Thematic mapper analysis of coniferous forest structure and composition. International Journal of Remote Sensing. 7(10):1287-1301.

Gautam, N.C., and G.CH.Chennaiah, 1985. Land-use and land-cover mapping and change detection in Tripura using satellite LANDSAT data. International Journal of Remote Sensing. 6(3):517-528.

Holko,M.L., and R.S.Sigman, 1984. The role of landsat data in improving U.S. crop statistics. Proceedings of the 18 th International symposium on Romete Sensing of Environment held in Paris on 1-5 October 1984 (Ann Arbor:Environmental Research Institute of Michigan), pp:307-320.

Horler,D.N.H., and F.J.Ahern, 1986. Forestry information content of thematic mapper data. International Journal of Remote Sensing 7(3):405-428.

Hudson,W.D., 1987. Evaluation of several classification schemes for mapping forest cover types in Michigan. International Journal of Remote Sensing, 8(12):1785-1796.

Janssen,L.L.F., M.N.Jaarsma and E.T.M. van der Linden, 1990. Integrating topographic data with remote sensing for land-cover classification. Photogrammetric Engineering and Remote Sensing, 56(11):1503-1506.

Jones,A.R., J.J.Settle, and B.K.Wyatt, 1988. Use of digital terrain data in the interpretation of SPOT-1 HRV multispectral imagery. International Journal of Remote Sensing, 9(4):669-682.

Karteris,M.A., 1990. The utility of digital thematic mapper data for natural resources classification. International Journal of Remote Sensing, 2(9):1589-1598.

Latty,R.S., and R.M.Hoffer, 1981. Waveband evaluation of proposed thematic mapper in forest cover classification. LARS Technical Report 041581, Purdue University, W.Lafayette, Indiana, 12 p.

Lillesand, T.M. and R.W.Kiefer, 1979. Remote Sensing and Image Interpretation, John Wiley & Sons,Inc., New York, Chichester, Brisbane, Toronto, Singapore, 612 p.

Lo, T.H.C., F.L.Scarpace, and T.M. Lillesand, 1986. Use of multitemporal spectral profiles in agricultural land-cover classification. Photogrammetric Engineering and Remote Sensing, 52(4):535-544.

Townshend, J.R.G., 1981. Terrain Analysis and Remote Sensing. George Allen & Unwin, London, 232 p.

Vogelmann,J.E., and B.N.Rock, 1986. Assessing forest decline in coniferous forests of vermont using NS -001 thematic mapper simulator data. International Journal of Remote Sensing, 7(10):1303-1321.

Williams,D.L., and R.F.Nelson, 1986. Use of remotely sensed data for assessing forest stand conditions in the eastern United States, IEEE Transactions on Geoscience and Remote Sensing, GE 24(1):130-138.