NASA

Technical Memorandum 79546

(NASA-TM-79546) KOREAN COASTAL WATER DEPTH/SEDIMENT AND LAND COVER MAPPING (1:25,000) BY COMPUTER ANALYSIS OF LANDSAT IMAGERY (NASA) 23 p HC A02/MF A01 CSCL 08B

N78-23532

Unclas
G3/43 17686

Korean Coastal Water Depth/Sediment and Land Cover Mapping (1:25,000) by Computer Analysis of Landsat Imagery

K. Y. Park and Lee D. Miller

MAY 1978

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

PREPRINT of a paper to appear in:
Remote Sensing of Earth Resources, Vol. 7. 1978. F. Shahrokhi, ed.
Univ. of Tenn. Space Inst. Tullahoma, Tenn.

KOREAN COASTAL WATER DEPTH/SEDIMENT AND LAND COVER MAPPING (1:25,000) BY COMPUTER ANALYSIS OF LANDSAT IMAGERY

K. Y. Park* and L. D. Miller+

Department of Civil Engineering
Colorado State University
Fort Collins, Colorado 80523

Abstract

Computer analysis was applied to single date LANDSAT MSS imagery of a sample coastal area near Seoul, Korea equivalent to a 1:50,000 topographic map. Supervised image processing yielded a test classification map from this sample image containing 12 classes: 5 water depth/sediment classes, 2 shoreline/tidal classes, and 5 coastal land cover classes at a scale of 1:25,000 and with a training set accuracy of 76%. Unsupervised image classification was applied to a subportion of the site analyzed and produced classification maps comparable in results in a spatial sense. The results of this test indicated that it is feasible to produce such quantitative maps for detailed study of dynamic coastal processes given a LANDSAT image data base at sufficiently frequent time intervals.

Introduction

Extensive earlier LANDSAT coverage exists in the "public domain" for large areas of the world. At present new image acquisitions of geographic areas are impeded by the lack of tracking stations, on-board tape capacity, and current recognition of nationalist sentiments. Computer classification of this earlier public imagery must, of necessity, be completed with a minimum or absence of ground control data. It is also the nature of large scale processes, such as desertification, coastal dynamics, tropical forest exploitation, etc., that they do not lend themselves to characterization or represen-

---

* Ph.D. Candidate, Department of Civil Engineering.

+ Senior Postdoctoral Scientist, NASA/Goddard Space Flight Center, Greenbelt, Maryland.

tation by ground control even when attempted simultaneously with image acquisition.

The western coast of Korea consists of an extensive network of bays, estuaries, offshore islands, extensive tidal flats, and other complex features. Very large dynamic sediment transport processes are clearly evident in the LANDSAT mosaic of this area but are very difficult to characterize by "ground control". Detailed study of the dynamics of this coastal area will require computer classification of time sequences of LANDSAT images. Efforts to develop the initial procedures required for such a program were undertaken to determine what classification patterns of the coastal areas could be extracted in the absence of ground control. The results of these initial efforts should enable those responsible for the analysis of coastal dynamics to determine whether they wish to promote the needed LANDSAT coverage of the area and implement the image analysis techniques in Korea.

This demonstration proceeded in several steps. Supervised image classification (maximum likelihood ratioing) was employed with training sets selected by direct examination of the digitally displayed imagery by one familiar with the selected study site. Stepwise discriminant analysis was employed as a mechanism to refine these training sets and determine those image variables making the most significant contributions to classification accuracy. A complete classification and display was next prepared at a scale of 1:25,000 for an area equivalent to a 1:50,000 topographic map. Finally, an unsupervised (ISOCLAS) clustering analysis was completed for a portion of the same area to assess the relative value of this approach to the image classification in areas of sparse or absent ground control.

Test Site

Korea consists of generally mountainous lands with small valleys and narrow coastal plains. Most of its agricultural activity occurs on tiny plots which are either irrigated rice paddies or nonirrigated, dry, upland fields. Due to the high population density, rural settlements are found in almost every portion of arable lands, while numerous small clustered fishing villages occur along the coastline. Recent population increases have also caused rapid expansion of major urban areas and attendant changes in urban and suburban landscapes.

The site selected for initial LANDSAT analysis is 20 kilometers from the margin of Seoul (population of seven million) and contains a combination of suburban and rural features distributed along a 30 kilometer coastline (Fig. 1). The area of the site represents one quadrangle of the 1:50,000 Korean topographic map series and is approximately equally divided between the shallow water area of Inchun Bay and the coastal land area. Approximately a dozen usable LANDSAT images of this site were available for the intervening years since the initial 1972 launch. All illustrate the complex, dynamic offshore sediment patterns and processes which occur along the western coast of Korea and, more particularly, in the Inchun Bay. One LANDSAT frame of 31 October 1972 was selected for this preliminary test application of computer image classification techniques to extract information of use for the study of coastal sediment processes and land use (Fig. 2).

The four band, digital multispectral scanner (MSS) data available for the selected LANDSAT frame was preprocessed using portions of the LANDSAT Mapping System (LMS) package developed at Colorado State University.[1] A picture element recorded on the LANDSAT computer compatible tape (CCT) represents a ground area as a 57 meter (E-W) by 79 meter (N-S) parallelogram inclined about 12 degrees east of north. The LMS geometric rectification module was applied to the digital imagery to remove systematic distortions and resample the inclined picture elements based on the nearest neighbor approach. The data element resulting from this process now represents a ground area of 63.5 meters (E-W) and 79.4 meters (N-S) (approximately 0.5 hectare) and of known geographic position. The resulting reformatted data was displayed on an eight line per inch computer printer as a map overlay at a scale of 1:25,000 (Fig. 3). Photo-like displays of the selected LANDSAT MSS coverage of the test site were prepared from this resampled data using a microfilm graymap routine (Figs. 4 and 5).

Additional preprocessing with the LMS system consisted of forming the six interband ratios between the four original MSS spectral bands (the six inverse ratios were omitted). The resulting composite of ten bands (four original and six derived) were tested to determine their relative contribution to the image analysis procedures.

Selection of Training Sets

The map classification algorithm in the LMS package is based upon the Gaussian likelihood method. It is a supervised classification algorithm which requires ground control

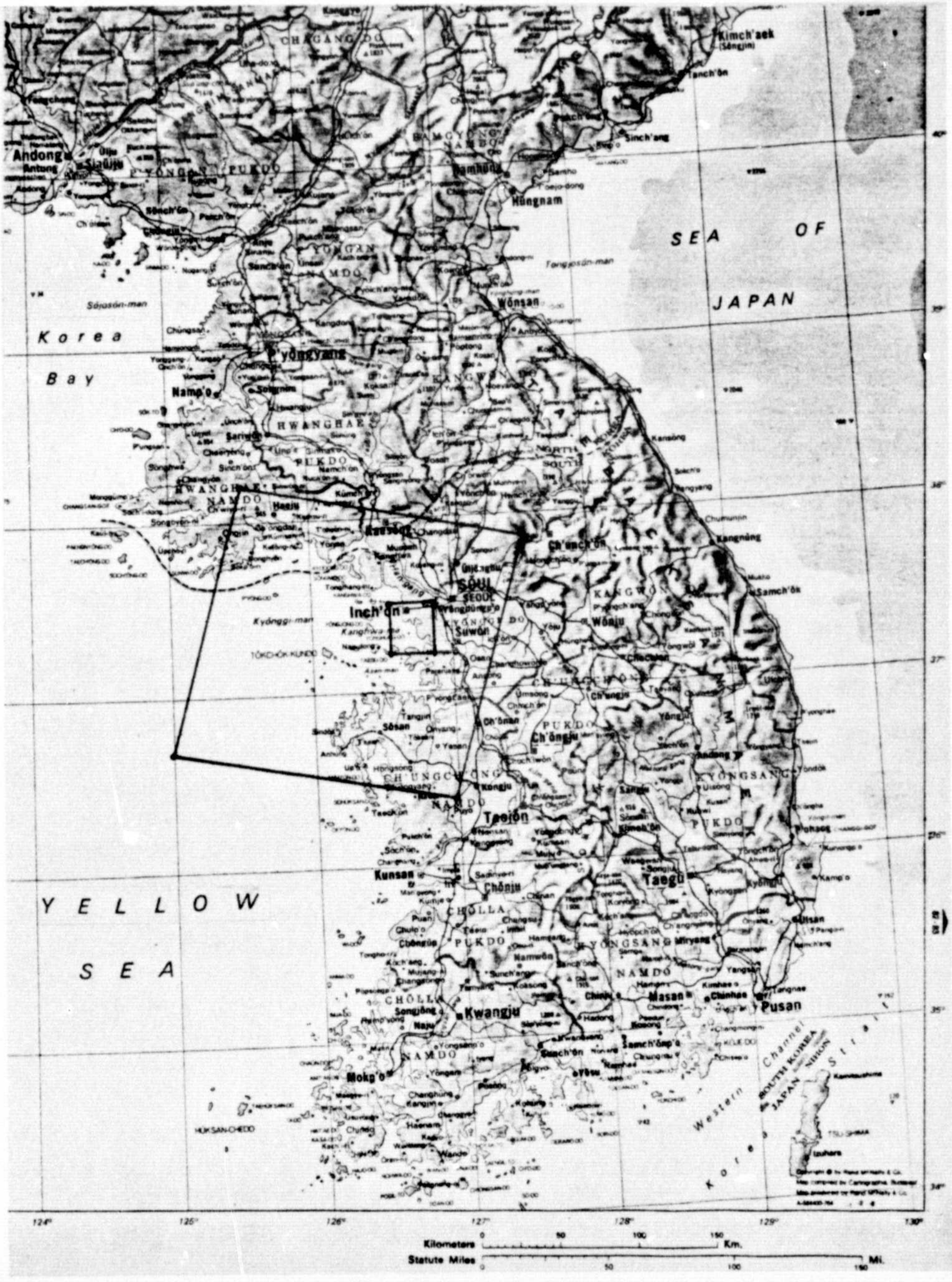

Figure 1. Geographic location of the study site. Smaller rectangle represents approximate location of site consisting of the area of one of the 1:50,000 Korean topographic maps. Larger parallelogram represents approximate area of the series of LANDSAT images available for this site.

Figure 2a. Photographic format LANDSAT MSS band 5 image of the study site. Rectangle represents approximate location of the site consisting of the area of one of the 1:50,000 Korean topographic maps.

Figure 2b. Photographic format LANDSAT MSS band 7 image of the study site. Rectangle represents approximate location of the site consisting of the area of one of the 1:50,000 Korean topographic maps.

Figure 3. Lineprinter display of the LANDSAT MSS band 7 of a portion of the study site. Scale 1:25,000, 31 October 1972.

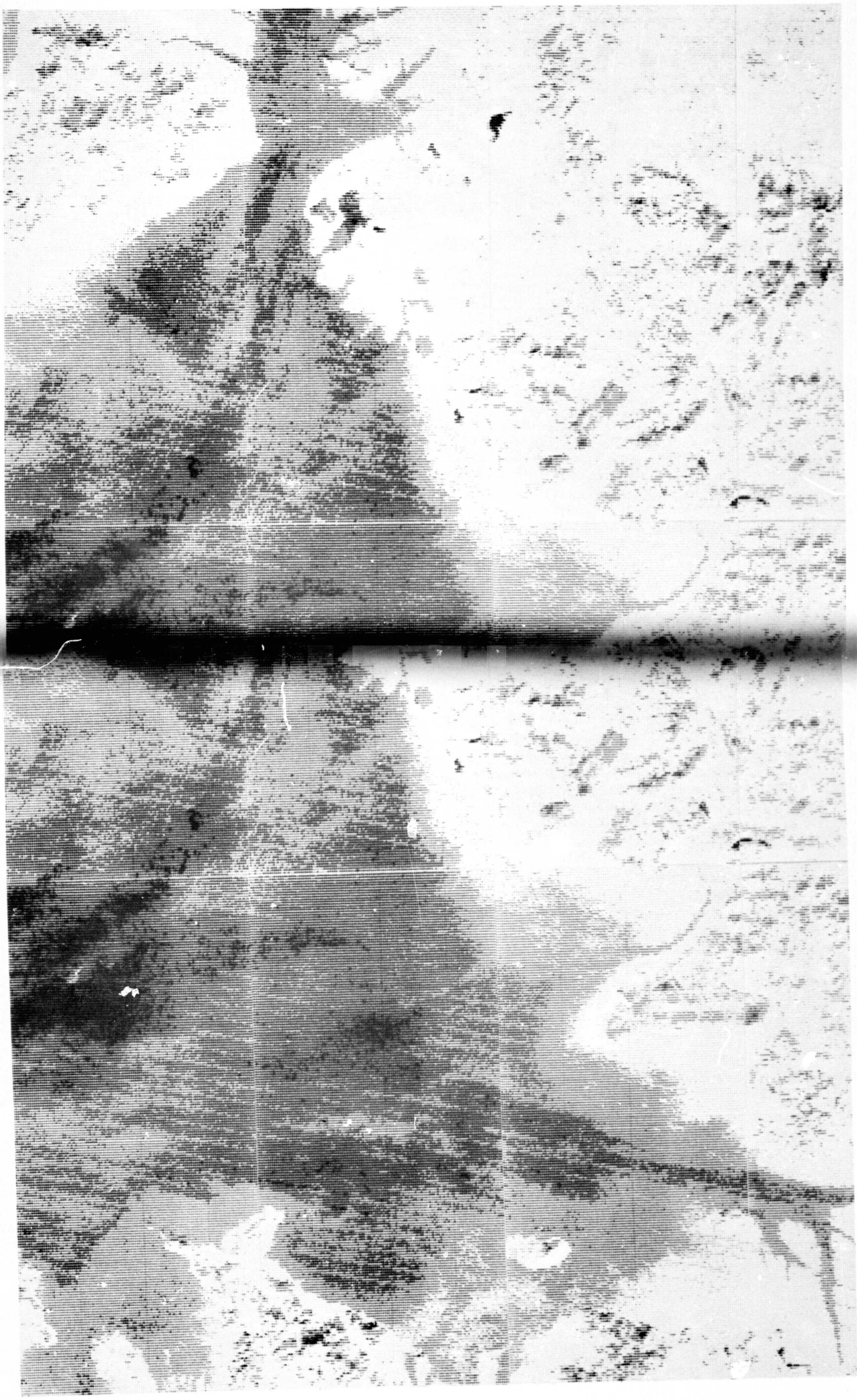

Figure 4. A microfilm graymap of LANDSAT MSS band 5. Geometrically rectified and then displayed on a computer microfilm plotter. Imaged 31 October 1972.

FOLDOUT FRAME

Figure 5. A microfilm graymap of LANDSAT MSS band 7. Geometrically rectified and then displayed on a computer microfilm plotter. Imaged 31 October 1972.

information for the development of training classes to represent each land cover or water depth/sediment type. This training procedure is used to provide estimates for statistical parameterization of each map class sought. These parameters, the mean vector and covariance matrix, are normally computed from the imaged values of the training set areas identified by ground control data. No meaningful ground control data was available for this study. Hence, unidentified training sets were obtained by selecting several rectangular groups of cells exhibiting similar graytones or radiance levels in both LANDSAT MSS bands 5 and 7. These groups of cells in aggregate constituted the training set selected to represent each class to be identified subsequently on the ground using the initial classification map. This training set selection procedure may be interpreted as a simple spatial clustering and is essentially equivalent to photo interpretation without field checking. The photo interpreter often recognizes specific objects in an airphoto by earlier experience with the area and uses the synoptic coverage of the photo to expand or extrapolate this knowledge of the area. Similarly, the image analyst in this case was generally familiar with the land uses in the area of the site and selected representative, but not positively identified, groups of training sets. Supervised image classification was then used to objectively extend their location to other portions of the site. Nine land cover classes and five water depth/sediment classes were initially specified in this fashion. Three of the initial land cover classes were shoreline beaches and/or tidal areas of exposed mud. The five water classes represented varying gradations of water depth and sediment. All eight of these classes are of particular interest when examining dynamic coastal erosion/deposition processes.

Refinement of Training Sets

Initial classification performance for these 14 pseudo-supervised classes using the 10 bands or variables available was examined by a stepwise linear discriminant analysis. Best performance of this classifier is expected when each class has the distinctive unimodal normal distribution in the available multidimensional feature space. Tabulations of training set classification accuracy were obtained by the stepwise discriminant analysis module in LMS. Refinement of training sets was carried out by examining the distribution characteristics of the MSS data selected as a training set for each class and the classification accuracy tableaus displayed as each new variable is added in the stepwise fashion (Table 1). Earlier attempts at further refinement of such

Table 1. Classification matrix for the 14 initial training sets using stepwise linear discriminant analysis on the four MSS bands of the 31 October 1972 image. Values in percent.

CLASS	*symbolic representation* ↓→	PREDICTION: Land Cover ⊞	X	ǂ	I	≠	=	Shoreline −	+	.	Water *	/	0	$	■	Individual Class Accuracy
Land Cover	⊞	60	31	2	2	2	2									60
	X	23	67			2	2	5	2							67
	ǂ	3	14	78			4		1	①						78
	I	1	①		74	4	20									74
	≠	10	7	1	8	25	19	12	3	1	1	7	3			25
	=				16		52		32							52
Shoreline	−					1		86	11	2						86
	+	9	2			2		2	32	45	4	3				32
	.							1	32	66						66
Water	*										95	5				95
	/										8	77	15			77
	0											7	76	18		76
	$											1	21	60	19	60
	■											1	13	29	57	57
Commission Accuracy ②		34	61	99	92	51	25	86	32	49	71	86	53	60	73	73

① Represents percentages less than 1.

② Commission Accuracy = number of correctly assigned cells expressed as a percentage of total number assigned to a class.

training sets by the deletion of individual cells of that set which were found to be most dissimilar were not repeated here. Removal of such individual outlying cells from the population of cells initially selected to represent a given class purifies the training set and increases training set accuracy but has been shown to have no impact on the accuracy of the final map product.[2]

Overall training set accuracy is used as a measure of success for each new variable added. It is computed as a percent representing the total number of training set cells correctly assigned by the classifier back into their initial classes. The 14 training sets selected yielded an overall training set classification accuracy of 72.7% using the four original MSS bands (Fig. 6 and Table 2). Examination of the classification matrix and histograms of the data values for each training set clearly indicated confusion between two of the 14 training sets. Elimination of the confusing classes by regrouping the 14 classes into 12 by combining two land cover

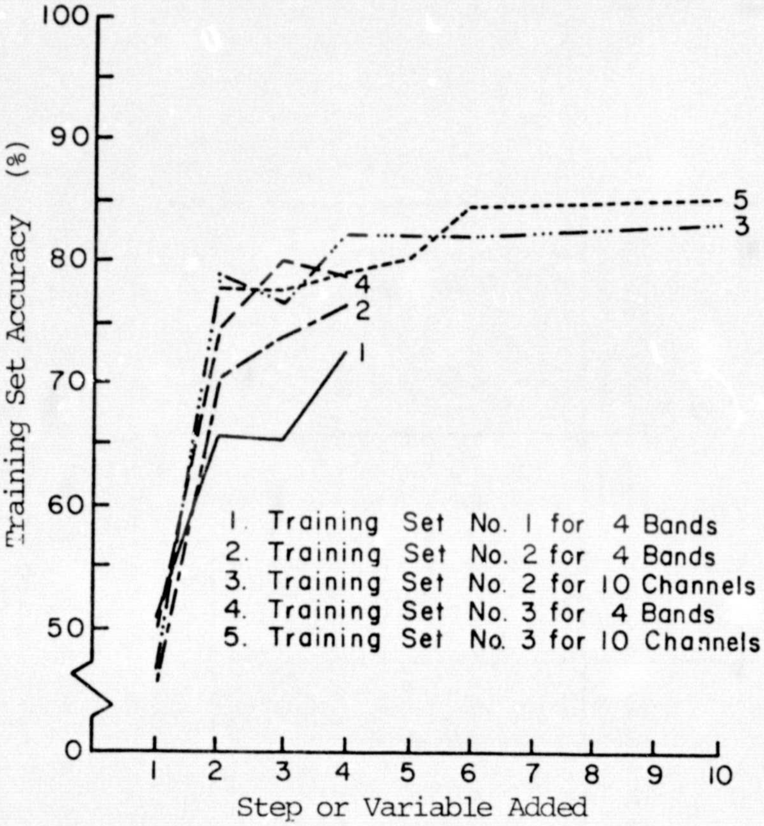

Figure 6. Overall training set classification accuracy achieved by stepwise linear discriminant analysis.

and two shoreline classes, respectively, yielded an apparent increase in training set accuracy from 72.7% to 76.6%. Re-siting some of the original rectangular training set selections yielded an additional increase in accuracy for the four MSS bands from 76.6% to 78.9%. Much of the apparent increase between 72.7% and 76.6% may well be due to the associated reduction of the number of choices for classes from 14 to 12. The change between 76.6% and 78.9% represents a small, but real, refinement in training set characteristics. When the six MSS band ratios were added to the original four MSS bands a further increase of 6.4% in overall training set accuracy was achieved with the refined training sets and yielded an acceptable training set classification matrix (Table 3). The classes of water depth/sediment content have good separability with confusion rates of less than 27%. Land cover classes have higher confusion rates ranging from 17% to 35%. Almost no confusion occurs between the desired shoreline and land cover or water classes. It should also be clearly noted that

Table 2. Overall training set accuracy as a function of the variable added. The three iterations represent successive refinement of classes and training sets.

Set No.		1		2				3			
No. of Classes		14		12				12			
No. of Cells		1,524		1,451				1,339			
	Step	Band Added	Percent Accuracy	Band Added	Percent Accuracy	Band Added	Percent Accuracy	Band Added	Percent Accuracy	Band Added	Percent Accuracy
	1	6	51.0	7	45.7	6/4	46.5	7	50.1	6/4	46.6
	2	4	65.9	4	70.3	5	79.0	4	74.6	4	77.9
	3	7	65.6	6	73.9	7	76.8	5	80.1	7	77.7
	4	5	72.7	5	76.6	7/5	82.4	6	78.9	7/6	79.1
	5					7/6	82.1			7/4	80.2
	6					6/5	82.1			5/4	84.7
	7					5/4	.			6/5	.
	8					4	.			5	.
	9					6	.			7/5	.
	10					7/4	83.3			6	85.3
Final		4 bands	72.7	4 bands	76.6	10 bands	83.3	4 bands	78.9	10 bands	85.3

Table 3. Classification matrix for the 12 final training sets using stepwise linear discriminant analysis on the foui MSS bands and six ratios of the 31 October 1972 image. Values in percent.

CLASS	*symbolic representation* ↓	Land Cover ⊞	Land Cover ⧗	Land Cover ǂ	Land Cover I	Land Cover =	Shoreline –	Shoreline •	Water *	Water /	Water 0	Water $	Water ▣	Individual Class Accuracy
Land Cover	⊞	65	33	2										65
	⧗	20	74			2	4							74
	ǂ	3	13	83		1		①						83
	I	5	①		78	16	①							78
	=	7		3	10	80								80
Shoreline	–						96	4						96
	•						①	99						99
Water	*								100					100
	/								4	95	1			95
	0									1	98	1		98
	$										4	79	17	79
	▣											27	73	73
Commission Accuracy ②		40	68	99	98	48	95	96	88	99	94	76	79	85

① Represents percentages less than 1.

② Commission Accuracy = number of correctly assigned cells as a percentage of total number assigned to a class.

the training set classification accuracy examined here is always less than final map or verification accuracy.

Stepwise discriminant analysis also provides insight into the relative significance of each variable employed in the classification. The first step or variable selected was chosen to yield the best overall separability of each class from all others in the multimodal space specified. A large additional improvement in training set accuracy is achieved when a second step is taken with the addition of a second variable. The total amount of improvement was generally reduced as additional variables are added although the rate of change in accuracy may increase or decrease (Fig. 5). Stepwise discriminant analysis is also a linear process which most certainly does not provide the absolute optimal combination of variables less than the total number available. It is useful, however, as one method for quickly determining the relative contribution of a collection of available variables.

Map Classification

The LMS map classification module employs the Gaussian likelihood ratio (maximum likelihood) classifier, in which mean vector and covariance matrix of each class have to be provided for the classification. When the ratios of the MSS bands were used unusually high correlations occurred between selected ratio bands and their original bands for some classes like water. The Gaussian likelihood algorithm fails to perform adequately in these conditions. Any high dependency between variables or bands has to be reduced under such circumstances so that the covariance matrices of the classes can be inverted. This was not an impediment in the stepwise discriminant analysis since it uses a single common covariance matrix for all the classes. The five water level classes are very important to the mapping and study of coastal zone processes and could not be deleted. Hence, only the original four MSS bands were utilized in the final mapping of the entire area. Restricting the classification to these four variables negated the small possible increase in accuracy detected when using band ratios in training set development. This accuracy loss was compensated for by the significantly reduced cost of classifying with four variables in lieu of ten.

After classification the resulting 1:25,000 scale map of the site was displayed as a symbol map on the lineprinter (Fig. 7). This symbol map portrayed approximately 170,000 cells of 0.5 hectare whose water type and land cover have been identified with an accuracy yet to be determined. Reasonably clear land use patterns become even more apparent when the classification results are displayed as graytone theme maps on the microfilm plotter (Figs. 8 and 9).

Test of an Unsupervised Classification

The initial map classifications were prepared by a supervised approach employed without known ground data. A comparison was made of this crude approach to that which might be achieved with an unsupervised or clustering algorithm. ISOCLAS was employed for this comparison and consists of a modified version of the clustering algorithm ISODATA developed at the Stanford Research Institute.[3] Preparation of input data for ISOCLAS from the LMS data format was carried out with the assistance of personnel of the U. S. Forest Service.[4] Thus the same data input to the earlier supervised classification was input to this unsupervised classification. The images had been subjected to the same geometric rectification and resampling noted earlier. The original data of MSS bands 4, 5, and

Figure 7. Supervised classification map of the 12 water level/sediment and land cover types. Computed using Gaussian likelihood ratioing of four LANDSAT MSS bands. Scale 1:25,000, 31 October 1972. One cell represents approximately 0.5 hectares. Symbolism corresponds with that of Table 3.

Figure 8. Theme map of water level/sediment and coastal classes. Computed using Gaussian likelihood ratioing on four LANDSAT MSS bands. Scale approximately 1:150,000. Image taken 31 October 1972.

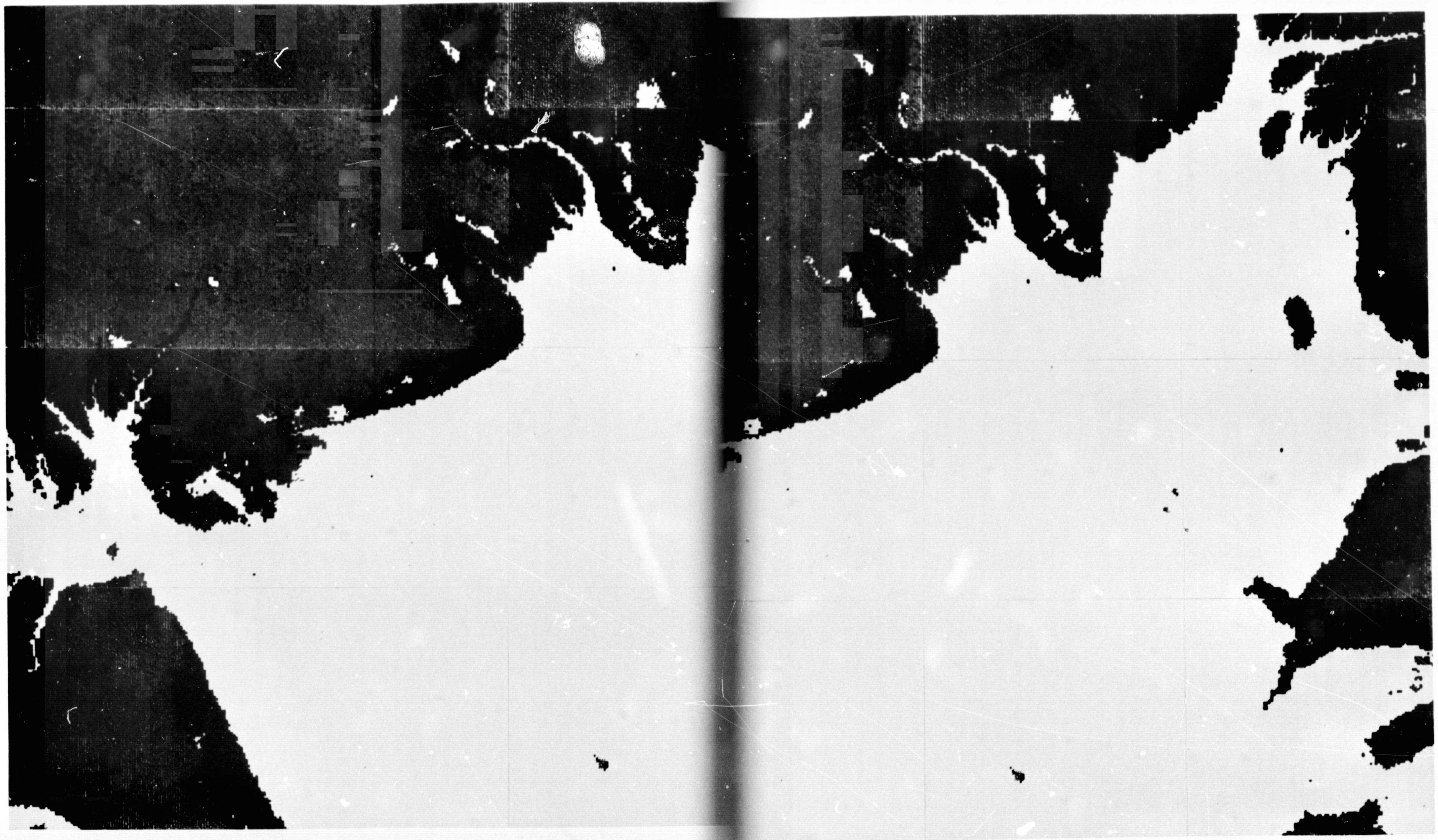

Figure 9. Theme map of water level/sediment and coastal classes. Computed using Gaussian likelihood ratioing on four LANDSAT MSS bands. Scale approximately 1:150,000. Image taken 31 October 1972.

6 were doubled from the original data magnitude and MSS band 7 was quadrupled. This differential range adjustment did not affect the earlier classification results but doubled the relative weight of the band 7 data when computing the similarity measures of Euclidean distance with the ISOCLAS algorithm. As a result, similar magnitudes of standard deviations were obtained for all clusters.

Classification maps were prepared for a portion of the study site using this unsupervised approach and were displayed for visual comparison with those produced earlier (Fig. 10). Gross spatial patterns shown in the supervised and unsupervised approach were generally comparable at the 12 class levels. However, scattered small spatial features tend to form different distribution configurations. These inconsistent patterns were most prevalent in the land cover classes. This scattered dissimilarity of clusters reflects the variety of small scale agricultural activities currently practiced in Korea. The unsupervised clustering algorithm provided greater flexibility in adjusting to the variations of typical clusters. The classes having inter-class distances in the spectral space which were closer than specified were able to be chained and displayed as spectrally similar on the classification map (Fig. 10b). A similar supervised (Fig. 10a) and unsupervised pattern (Fig. 10c) was observed when classes closer than 4.5 units in distance were chained. Only two distinct classes were left when the threshold distance of chaining was 5.0 (Fig. 10d).

Conclusions

This effort was carried out to provide a first quick, crude attempt to apply computer analysis of digital LANDSAT imagery to Korea. It was performed without access to the ground area being surveyed but does provide a clear indication that large scale, timely maps could be prepared for the study of coastal zone features, especially for the erosion and deposition processes along the western coastline. Although the procedure for the use of the supervised approach (LMS) has not been described in detail the possibility of its employment in the total absence of ground control has been illustrated. Clearly, adequate known ground control would increase the applicability and meaningfulness of the results. However, future large scale studies of coastal dynamics where very large and dynamic processes are at work may never be supported with extensive ground control procedures. It has been shown that shoreline and water depth/sediment classes could be classified in a supervised procedure for later easy

(a) (b)

(c) (d)

Figure 10. Comparison of supervised and unsupervised classification maps of the 12 water level/sediment and land cover types. Image taken 31 October 1972. One cell represents approximately 0.5 hectare. (a) Supervised using Gaussian likelihood ratioing. (b) Unsupervised using ISOCLAS and 90 iterations without chaining for 12 classes. (c) ISOCLAS with chaining and d $\leq$ 4.5, where d = interclass distance. (d) ISOCLAS with chaining and d $\leq$ 5.0.

identification as necessary. Comparison of the results of the supervised method without ground control and the unsupervised or cluster algorithm indicated a clearer need for ground control data when dealing with the more spatially heterogeneous land cover features.

References

[1] Maxwell, E. L., T. C. Hart, R. L. Riggs and L. D. Miller, "Land Use Classification for Six Rocky Mountain States Using LANDSAT Multispectral - Multitemporal Data," Final Report prepared under contract to the Federation of Rocky Mountain States under NASA contract NAS-5-22338, June 1977, Earth Resources Dept. and Civil Engineering Dept., Colo. State Univ., Ft. Collins, CO, 173p.

[2] Sung, Q. C. and L. D. Miller, "Land Use/Land Cover Mapping (1:25,000) of Taiwan, Republic of China by Automated Multispectral Interpretation of LANDSAT Imagery," NASA/Goddard Space Flight Center, Doc. X-923-77-210, August 1977, Greenbelt, MD, 180p.

[3] Senkus, W. M., "ISOCLAS - User's Guide, Version 1.1," January 1976, Remote Sensing Research Program, University of Calif., Berkeley, CA.

[4] Merritt, N., Personal communication in 1977, Program PREPIT and a CDC version of ISOCLAS were provided the investigator by N. Merritt for use in this study, Rocky Mountain Forest and Range Experiment Station, Ft. Collins, CO.

BIBLIOGRAPHIC DATA SHEET

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle KOREAN COASTAL WATER DEPTH/SEDIMENT AND LAND COVER MAPPING (1:25,000) BY COMPUTER ANALYSIS OF LANDSAT IMAGERY		5. Report Date 6. Performing Organization Code
7. Author(s) K. Y. Park and L. D. Miller		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil Engineering Colorado State University Ft. Collins, CO 80523		10. Work Unit No. 11. Contract or Grant No. 13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address National Aero. Space Admin. Goddard Space Flight Center Greenbelt, MD 20771		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Computer analysis was applied to single date LANDSAT MSS imagery of a sample coastal area near Seoul, Korea equivalent to a 1:50,000 topographic map. Supervised image processing yielded a test classification map from this sample image containing 12 classes: 5 water depth/sediment classes, 2 shoreline/tidal classes, and 5 coastal land cover classes at a scale of 1:25,000 and with a training set accuracy of 76%. Unsupervised image classification was applied to a subportion of the site analyzed and produced classification maps comparable in results in a spatial sense. The results of this test indicated that it is feasible to produce such quantitative maps for detailed study of dynamic coastal processes given a LANDSAT image data base at sufficiently frequent time intervals.

17. Key Words (Selected by Author(s)) LANDSAT, Korea, Inchun Bay, Supervised Classification, Unsupervised Classification, Stepwise Discriminant Analysis		18. Distribution Statement	
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages 20	22. Price