DC-Y3-04439
NSTL/ERL-224

AgRISTARS

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Domestic Crops and Land Cover

SEPTEMBER 1983

Technical Report

STRATIFICATION OF SAMPLED LAND COVER BY SOILS FOR LANDSAT-BASED AREA ESTIMATION AND MAPPING

E. R. Stoner

National Aeronautics and Space Administration
National Space Technology Laboratories
Earth Resources Laboratory
NSTL, MS 39529

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO. DCLC-Y3-04439; NASA/NSTL/ERL-224	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
4. TITLE AND SUBTITLE Stratification of Sampled Land Cover by Soils for Landsat-Based Estimation and Mapping		5. REPORT DATE September 1983
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) E. R. Stoner		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS NASA/NSTL Earth Resources Laboratory NSTL, Mississippi 39529		10. WORK UNIT NO.
		11. CONTRACT OR GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS National Aeronautics and Space Administration		13. TYPE OF REPORT & PERIOD COVERED AgRISTARS Technical Report
		14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Analysis of Landsat MSS data within the framework of land areas sampled using June Enumerative Survey (JES) procedures has proven beneficial for land cover area estimation and mapping. One of the possible factors leading to misidentification of crops is the variable soil background on which the crops are grown; therefore, an analysis procedure which minimizes spectral confusion resulting from soil differences should be of value in improving classification accuracy.

Multidate Landsat MSS data and JES land cover information from 175 sampled segments were available from Robeson County, North Carolina, for the 1980 growing season. The county soil survey was used to interpret and assign soil class values to each 5-acre, or larger, field, serving as the basis for stratification of the sampled land areas into units which were relatively homogeneous in soil drainage, topography, and land use. For the four principal cover types--forest, soybeans, corn, and tobacco--soil stratification resulted in improved classification accuracy over the analysis of single-date, unstratified Landsat MSS data. Multidate Landsat analysis, however, resulted in similar or more accurate classification than stratified, single-date analysis. Tradeoffs between the cost of stratifying by soils and the expense of additional Landsat scenes would have to be considered over the useful life of a set of sample segments.

17. KEY WORDS Soils; Land Cover; Landsat MSS; Area Estimation and Mapping; Sampling; Remote Sensing; June Enumerative Survey	18. DISTRIBUTION STATEMENT

19. SECURITY CLASSIF. (of this report) Unclassified	20. SECURITY CLASSIF. (of this page) Unclassified	21. NO. OF PAGES 19	22. PRICE *

NSTL FORM 13 (JAN 1975)

*For sale by NTIS, Springfield, VA 22151

STRATIFICATION OF SAMPLED LAND COVER BY SOILS FOR LANDSAT-BASED AREA ESTIMATION AND MAPPING

September 1983

by E. R. Stoner

National Aeronautics and Space Administration
National Space Technology Laboratories
Earth Resources Laboratory

NASA/NSTL/ERL Report No. 224

AgRISTARS Technical
Report DCLC-Y3-04439

TABLE OF CONTENTS

LIST OF TABLES

I. SUMMARY

Analysis of Landsat MSS data within the framework of land areas sampled using June Enumerative Survey (JES) procedures has proven beneficial for land cover area estimation and mapping. One of the possible factors leading to misidentification of crops is the variable soil background on which the crops are grown; therefore, an analysis procedure which minimizes spectral confusion resulting from soil differences should be of value in improving classification accuracy.

Multidate Landsat MSS data and JES land cover information from 175 sampled segments were available from Robeson County, North Carolina, for the 1980 growing season. The county soil survey was used to interpret and assign soil class values to each 5-acre, or larger, field, serving as the basis for stratification of the sampled land areas into units which were relatively homogeneous in soil drainage, topography, and land use. For the four principal cover types—forest, soybeans, corn, and tobacco—soil stratification resulted in improved classification accuracy over the analysis of single-date, unstratified Landsat MSS data. Multidate Landsat analysis, however, resulted in similar or more accurate classification than stratified, single-date analysis. Tradeoffs between the cost of stratifying by soils and the expense of additional Landsat scenes would have to be considered over the useful life of a set of sample segments.

II. INTRODUCTION

Stratification of remotely sensed data into broad land areas which are homogeneous in cropping intensity has been a successful approach to sampling and analyzing Landsat MSS data for crop area estimation (Hanuschak, Allen, and

Wigton, 1982). For specialty crops, sample frame construction techniques and stratification procedures using soil information were found to produce more precise estimates of planted acreage than was possible with estimates derived from a more general land use stratification (Fecso et al., 1982). The use of stratification by soils within the framework of the Statistical Reporting Service (SRS/USDA) June Enumerative Survey (JES) sampling procedures is discussed in this paper as a technique for improving the accuracy of Landsat-derived land cover classifications.

Soil characteristics exert a marked effect on spectral response from vegetated surface features. Soils differing in drainage, depth, moisture-holding capacity, inherent fertility, and topography affect the morphological and phenological characteristics of cultivated crops as well as forested land and rangeland. The spectral properties of surface soils themselves contribute substantially to the overall spectral response of sparse vegetative canopies typical of developing cultivated crops. The subtle spectral differences among cultivated crops suggest that any extraneous influence attributable to soil differences should be minimized, if possible. A convenient way to reduce the confounding effects of site factors is to use soil survey information at a level of cartographic and categoric detail sufficient to allow the remotely sensed data to be stratified by selected soil characteristics.

Stratification of Landsat MSS data along soil parent material boundaries defined by soil association maps for an area in Lincoln County, South Dakota, resulted in improved accuracy in corn classification (Dalsted, Worchester, and Devries, 1979). Land areas stratified based on historically uniform soil productivity with the Prairie Provinces of Canada correlated highly with productivity detraction features recognizable on enhanced Landsat data (Schubert et al., 1980). Soils stratified by physiographic position based on detailed county level soil survey information for Gentry County, Missouri provided increased overall classification accuracy for the predominantly agricultural land cover types using a multidate Landsat MSS data set (Stoner, 1982).

III. OBJECTIVE

The overall objective of this study was to evaluate the effect of stratifying Landsat MSS data by soil information on land cover classification accuracy. This stratification was made on a field-by-field basis by interpreting detailed soil survey information for sampled land areas, using procedures employed in the SRS/USDA June Enumerative Survey. The 42 soil-mapping units were aggregated into six soil classes differing mainly in drainage and topographic position, for the sake of simplifying analysis of the Landsat data. Four dates of Landsat MSS data were available for analysis.

IV. STUDY AREA DESCRIPTION AND DATA SOURCES

The study area consisted of sampled land areas scattered among the 607,104 acres (245,686 ha) of Robeson County, situated on the Coastal Plain in the Southeastern part of North Carolina. Of the total acreage, 58 percent of the county is in farm tracts, 30% of which is in forest and about 45% of which is in harvested crops (McCachren et al., 1978). About 70 percent of the harvested crop acreage is in corn and soybeans, with smaller amounts in cotton, tobacco, and truck crops.

The topography of Robeson County is nearly level to gently sloping; and the soil, although generally acid and low in natural fertility and organic matter content, is well-suited to farming on the better-drained sites. The well-drained soils occur on the interstream divides, while the more poorly-drained soils are located in the flood plains of streams and in depressional Carolina bays.

A. Land Cover Sampling

Crop information for 1870 segments covering the entirety of Robeson County was collected in a specially conducted June Enumerative Survey during the 1980 growing season. Typical JES segments from this county, consisting of about 0.5 mi^2 of land area, are divided into individual farm fields or management tracts of forested land. Complete enumeration of crop types was performed in a ground survey, while the identity of additional tracts of noncropland was interpreted from color IR aerial photography. About 10 percent, or 175 of the available segments, were selected for detailed analysis with Landsat MSS data.

Segment boundaries were digitized and registered to the coordinate system of the September 9, 1980, Landsat scene. Digital segment data files contained information pertaining to segment number, field number within segment, cover type code, and a boundary pixel identifier. Data analysis was directed to the field interior, nonboundary pixels, so as to exclude boundary pixels with mixed land use from the analysis.

B. Landsat-Crop Calendar Considerations

Landsat scenes from four 1980 growing season dates were available for analysis. The September 9 Landsat scene was used as a base to which all other scenes were registered. Available scenes and data quality were June 11 (scattered clouds, good quality), July 17 (hazy, fair quality), August 4 (cloud free, excellent quality), and September 9 (cloud free, excellent quality).

Evaluation of the 1980 crop calendars (Table 1) revealed that the September 9 scene, although of excellent quality, fell too late in the growing

Table 1. Status of Crop Calendars for Principal Cover Types, Robeson Co. NC, 1980 Growing Season (North Carolina Crop and Livestock Reporting Service, 1980).

	DATES OF LANDSAT ACQUISITION			
COVER TYPE	JUNE 11	JULY 17	AUGUST 4	SEPTEMBER 9
Soybeans	82% planted	12% blooming	38% blooming	83% pods set
Corn	98% planted by May 11	81% silked	94% silked	45% harvested
Tobacco	97% transplanted	8% harvested	22% harvested	89% harvested
Forest	leafed-out	leafed-out	leafed-out	leafed-out

season for discrimination between corn and tobacco, a good portion of which had already been harvested. The July 17 date was only fair in quality and was not yet at the peak vegetative growth of soybeans. The August 4 date was chosen for single-date analysis because of its excellent quality and optimal timing for a majority of the cover types, in spite of some harvesting of tobacco. For multidate analysis, the August 4/June 11 combination was chosen. Landsat MSS data from a date earlier than June 11 would have been desirable, but was not available.

C. Stratification by Soils

Soil survey maps at 1:20,000 were overlaid with plots of segment and field boundaries at the same scale in order to locate individual fields for assignment of soil class number on a field-by-field basis. The 46 soil mapping units of the Robeson County Soil Survey (McCachren et al., 1978) were aggregated into a simplified classification of soils along the lines of drainage and topographic position (Table 2). In the case of fields with more than one soil class, the soil class covering the largest proportion of the field was chosen as the single soil class identifier for that field. Soil class numbers were coded into the existing segment data file for all fields five acres or larger. Altogether, 153 segments containing more than 1500 fields were coded for soil information.

The soil classes selected and their attributes are described in Table 2. Soil class 1 covers 43 percent of the county and consists of well-drained soils on upland and terraces. Almost 80 percent of the area of soil class 1 is planted in row crops, while 15 percent is forested. Soil class 2 consists of moderately well-drained to somewhat poorly drained soils, of which about half is in row crops. Soil class 1 would be expected to have a light-colored,

Table 2. Characteristics of Soil Class Strata Defined from Mapping Units of the Robeson County, NC Soil Survey (McCachren, et al., 1978).

SOIL CLASS	AREAL EXTENT (% OF COUNTY TOTAL)	SOIL SUBORDERS	MAJOR SOIL CHARACTERISTICS	PREDOMINANT LAND USE
1	42.9	Udults, Psamments	Well drained, light, low organic matter content, on uplands and terraces	80% row crops 15% forested
2	10.4	Aquults, Udults, Psamments	Moderately well to somewhat poorly drained, dark, medium organic matter content, on uplands and bays	50% row crops 45% forested
3	18.0	Aquods, Aquults	Poorly drained, dark, medium organic matter content, on uplands and terraces	15% row crops 80% forested
4	10.2	Aquepts, Aquults, Saprists	Poorly to very poorly drained, black high organic matter content, on uplands and depressional bays	15% row crops 85% forested
5	17.8	Aquents, Aquepts, Aqualfs, Aquults	Poorly to very poorly drained, black, high organic matter content, on drains and low terraces	almost 100% forested
6	0.1	Orthents	Soils altered by man, rendering original relief and profile unrecognizable	100% urban built-up land

highly reflecting surface, while soil class 2 would have a dark surface. Soil classes 3, 4, and 5 are predominantly forested, exhibiting increasingly poorer drainage and resultant restrictions for cultivation. Soil class 6 consists of man-altered land and exists only in urban built-up land situations.

A pixel-by-pixel tally of each cover type by soil class on which it occurs is given in Table 3. Not all these cover types were analyzed, but their tally by soil class is instructive to convey the fact that the occurrence of certain cover types is restricted to certain land areas. For example, 95 percent of all tobacco fields occur on soil class 1. Hayfields and orchards are also restricted to these well-drained sites. A total of 85 percent of all forested tracts occur on the poorly drained sites of soil classes 3, 4, and 5. Urban land areas occur mainly on the better-drained sites of soil classes 1 and 2, as well as on the man-altered sites of soil class 6.

It is clear that, as an analysis tool, the stratification of Landsat data into land areas with a predominance of a given cover type should reduce confusion with other less frequently occurring cover types. As an example, soil class 5, which represents 18 percent of the county land area is 98 percent forested, thus reducing the possibility of misidentification of classified pixels with cover types such as corn or soybeans.

V. LAND COVER CLASSIFICATION

The 153-segment land cover data set was divided into one group of 76 segments used for training and one group of 77 segments used for testing. Spectral class development was directed to those training pixels identified as belonging to an individual soil class, or in the unstratified case, to the entire set of training pixels. This was accomplished by using the program

Table 3. Cover Type/Soil Class Tally of Field Interior Pixels from 153 Robeson County, NC June Enumerative Survey Segments.

COVER TYPE	SOIL CLASS (STRATUM)						COVER TYPE TOTALS
	1	2	3	4	5	6	
Forest	2,031	1,165	6,167	4,054	8,259	—	21,676
Corn	4,061	359	362	209	5	—	4,996
Soybeans	5,446	814	703	515	67	—	7,545
Tobacco	911	32	14	—	—	—	957
Cotton	525	31	4	61	—	—	621
Grassland	109	32	422	14	55	—	632
Hay	162	6	—	—	—	—	168
Orchards	6	—	—	—	—	—	6
Urban Land	282	28	54	7	—	103	474
Stratum Totals	13,533	2,467	7,726	4,860	8,386	103	37,075

WCCL, or within-class cluster, part of the Earth Resources Laboratory Applications Software (ELAS, Junkin et al., 1981). A companion program, WMAX, assigned each pixel within the selected data set to one of the point-cluster-derived statistics using a maximum likelihood ratio algorithm. Spectral class labeling was done using a pixel-by-pixel tally of the classified data with the training data. Soil class 6 was not analyzed because, by definition, it represented disturbed, urban land.

A two-channel, single-date classification was made using Landsat MSS bands 5 and 7 from the August 5 acquisition. A multidate, four-channel classification was also run using Landsat MSS bands 5 and 7 from the June 11 and August 4 acquisitions.

VI. RESULTS AND DISCUSSION

Accuracy of identification of pixels in the 77-segment test data set was evaluated for classifications of individual soil strata and on an overall basis for single-date and multidate Landsat MSS data, both stratified by soil and unstratified. For both the single-date (Table 4) and multidate (Table 5) data sets, overall classification accuracy summarized by soil stratum showed significant differences among all soil strata. In all cases, accuracy improved for each successive stratum from soil class 1 to soil class 5. In general, this outcome can be explained by the decreasing number of row crop pixels in proportion to forest pixels in the higher numbered strata. The heavily forested strata 3, 4, and 5 have by far the highest classification accuracy, while there is still considerable misidentification among all cover types of stratum 1.

Comparison of results for single-date analysis (Table 6) shows improvement in the stratified approach to be significant for forest and soybeans, as well

Table 4. Percent Correct Identification of Test Pixels for Major Cover Types within Individual Soil Strata, Using August 4 Landsat MSS bands 5 and 7.

	SOIL STRATUM				
COVER TYPE	1	2	3	4	5
Forest	1111/1434 = 77.5	490/557 = 88.0	3029/3320 = 91.2	2107/2157 = 97.7	4169/4223 = 98.7
Soybeans	1903/2845 = 66.9 abc*	396/466 = 85.0 cde	238/297 = 80.1 bef	114/185 = 61.6 adf	0/1 = 0
Corn	1409/1854 = 76.0	62/133 = 46.6	42/161 = 26.1	6/74 = 8.1	—
Tobacco	198/448 = 44.2	0/24 = 0_a	0/14 = 0_a	—	—
Weighted Average	4621/6581 = 70.2	948/1180 = 80.3	3309/3792 = 87.3	2227/2416 = 92.2	4169/4224 = 98.7

*Values within a row not followed by the same letter are significantly different at the 0.05 level as determined by arcsin transformation and the Newman-Keuls test.

Table 5. Percent Correct Identification of Test Pixels for Major Cover Types Within Individual Soil Strata, Using June 11 and August 4 Landsat MSS Bands 5 and 7.

	SOIL STRATUM				
COVER TYPE	1	2	3	4	5
Forest	$\frac{1173}{1434}$ = 81.8	$\frac{434}{557}$ = 77.9	$\frac{3217}{3320}$ = 96.9	$\frac{2117}{2157}$ = 98.1	$\frac{4223}{4223}$ = 100.0
Soybeans	$\frac{1934}{2845}$ = 68.0 abc*	$\frac{376}{466}$ = 80.7 cde	$\frac{242}{297}$ = 81.5 bef	$\frac{176}{185}$ = 95.1 adf	$\frac{0}{1}$ = 0
Corn	$\frac{1432}{1854}$ = 77.2	$\frac{113}{133}$ = 85.0	$\frac{76}{161}$ = 47.2	$\frac{1}{74}$ = 1.4	—
Tobacco	$\frac{176}{448}$ = 39.3	$\frac{0}{24}$ = 0_a	$\frac{0}{14}$ = 0_a	—	—
Weighted Average	$\frac{4715}{6581}$ = 71.6	$\frac{923}{1180}$ = 78.2	$\frac{3535}{3792}$ = 93.2	$\frac{2294}{2416}$ = 95.0	$\frac{4223}{4224}$ = 100.0

*Values within a row not followed by the same letter are significantly different at the 0.05 level as determined by arcsin transformation and the Newman-Keuls test.

Table 6. Percent Correct Identification of Test Pixels for Major Cover Types for Single and Multidate Landsat MSS Data Sets, Both Stratified by Soils and Unstratified.

	AUGUST 4 MSS 5, 7		JUNE 11 MSS 5, 7/ AUGUST 4 MSS 5/7	
COVER TYPE	UNSTRATIFIED	STRATIFIED	UNSTRATIFIED	STRATIFIED
Forest	10316/11691 = 88.2	10906/11691 = 93.3	10816/11691 = 92.5	11164/11691 = 95.5
Soybeans	2509/3794 = 66.1	2651/3794 = 69.9 a*	2824/3794 = 74.4	2728/3794 = 71.9 a
Corn	1472/2222 = 66.2 a	1519/2222 = 68.4 a	1600/2222 = 72.0 b	1622/2222 = 73.0 b
Tobacco	175/486 = 36.0 ab	198/486 = 40.7 acd	226/486 = 46.6 c	176/486 = 36.2 bd
Weighted Average	14472/18193 = 79.5	15274/18193 = 84.0	15466/18193 = 85.0	15690/18193 = 86.2

*Values within a row not followed by the same letter are significantly different at the 0.05 level as determined by arcsin transformation and the Newman-Keuls test.

as on an overall basis. With a multidate data set, forest and overall classification improved considerably for the stratified approach, although the accuracy of soybean and tobacco classification actually decreased significantly.

An analysis of all four classification approaches reveals the stratified, multidate procedure to have the highest overall accuracy. The addition of the June 11 Landsat MSS data to the August 4 data set had about the same effect on overall classification accuracy as the stratification of the August 4 data set by soils; although the unstratified, multidate approach was clearly better for crop discrimination.

VII. SUMMARY AND CONCLUSIONS

As a means of analyzing Landsat MSS data from sampled land areas, stratification of individual fields by soil characteristics is an improvement over the analysis of the best available Landsat data set without the benefit of soil information. Improved results are clearly related to the fact that certain soil strata are predominantly forested, presenting a less complicated situation for the classification of these strata. Classification accuracy was poorest for soil strata with a predominance of cultivated land. There is little evidence to suggest that the outcome of crop classification was related in any way to the attempt to stratify soils along light versus dark soil backgrounds. The fact that the detailed soil map information was summarized by field rather than being digitized for pixel-by-pixel matching with Landsat data, resulted in a soil stratifiction resembling more a localized land use scheme than a true soil classification.

An evaluation of the costs involved would have to be made to determine whether the effort of performing a soil stratification would be justified over the expense of acquiring an additional Landsat scene. While the analysis of multidate Landsat data would require the purchase of additional scenes each year, a permanent soil stratification of sampled land areas, unchanged for yearly revisits, would be a one-time expense. The eventual availability of soil survey information in digital format as a routine product of cooperating soil mapping agencies would facilitate soil stratification and merit its use as an operational procedure for sensor-derived land cover estimation and mapping.

VIII. REFERENCES

1. Dalsted, K. J., B. K. Worcester, and M. E. Devries. 1979. Influence of soils on Landsat spectral signatures of corn. Eighth Annual Remote Sensing of Earth Resources Conference, March 27-29, 1979. University of Tenn. Space Inst., Tullahoma.

2. Fecso, R., J. Geuder, B. Hale, and S. Pavlasek. 1982. Estimating dry bean acreage in Michigan. SRS-USDA. Staff Report No. AGES 820225. U.S. Government Printing Office, Washington, D.C.

3. Hanuschak, G. A., R. D. Allen, and W. H. Wigton. 1982. Integration of Landsat data into crop estimation program of USDA's Statistical Reporting Service (1972-1982). Proceedings of the Eighth International Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, IN. pp. 45-55.

4. Junkin, B. G., R. W. Pearson, B. R. Seyfarth, M. T. Kalcic, and M. H. Graham. 1981. ELAS, Earth Resources Laboratory Applications Software. Report No. 183. Revised Edition. National Aeronautics and Space Administration. National Space Technology Laboratories. Earth Resources Laboratory. NSTL Station, MS.

5. McCachren, C. M., R. E. Horton, A. Weaver, and D. G. Spangler. 1978. Soil survey of Robeson County, North Carolina. SCS-USDA. U.S. Government Printing Office, Washington, D.C.

6. North Carolina Crop and Livestock Reporting Service. 1980. Weekly crop summaries. NCCLRS-USDA. Raleigh, NC.

7. Schubert, J., P. Chagarlamudi, J. A. Shields, and A. R. Mack. 1980. Stratification of Landsat data by uniformity productivity of soils. Proceedings of the 6th International Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, IN. pp. 186-193.

8. Stoner, E. R. 1982. Agricultural land cover mapping with the aid of digital soil survey data. Proceedings of the Eighth International Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, IN. pp. 495-502.