NASA Technical Memorandum 84244

Final Report: Landsat Land Cover Analysis Completed for CIRSS/San Bernardino County Project

William Likens, Keith Maw,
and David Sinnott

OCTOBER 1982

NASA Technical Memorandum 84244

Final Report: Landsat Land Cover Analysis Completed for CIRSS/San Bernardino County Project

William Likens
Ames Research Center
Moffett Field, California

Keith Maw
Technicolor Graphic Service
Mountain View, California

David Sinnott
Ames Research Center
Moffett Field, California

National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1982

TABLE OF CONTENTS

CALIFORNIA INTEGRATED REMOTE SENSING SYSTEM: LANDSAT LAND-COVER ANALYSIS FOR THE SAN BERNARDINO COUNTY PROJECT

William Likens, Keith Maw,* and David Sinnott

Ames Research Center

The Landsat analysis carried out as part of Ames Research Center's San Bernardino County Project, one of four projects sponsored by NASA as part of the California Integrated Remote Sensing System (CIRSS) effort for generating and utilizing digital geographic data bases, is described. Topics explored in this Landsat analysis include use of data-base modeling with spectral cluster data to improve Landsat data classification, and quantitative evaluations of several change techniques. Both 1976 and 1979 Landsat data were used in the project. The Landsat analyses took place between April 1980 and September 1981.

INTRODUCTION

The San Bernardino County Project, sponsored by Ames Research Center, is one of four California Integrated Remote Sensing System (CIRSS) projects undertaken to investigate the concept of vertical data integration, or the utilization of a data set at a number of jurisdictional levels (ref. 1). A major focus is the use of Landsat data in a digital Geographic Information System (GIS) environment. The use of GIS data as an integral part of the Landsat image classification process was examined before undertaking data base applications for San Bernardino County and the San Bernardino National Forest. Also investigated was the means by which Landsat can be used as a mechanism to periodically update data bases.

A data base containing more than a dozen data layers, including elevation, slope, aspect, soils, land use, environment hazard, and growth management data, was constructed by integrating a number of special purpose data bases. Through hierarchical modeling, many of these data elements were used to guide a 1976 Landsat land-cover classification (ref. 2). A similar model, incorporating 1976-1979 Landsat spectral change data, in addition to other data base elements, was used in the classification of a 1979 Landsat image. The resultant Landsat products were integrated, as additional layers, into the data base.

In the San Bernardino project, the institutional and technical issues involved with an "industry-assisted" approach to vertical data integration (ref. 1) are also being examined. The industry participant, Environmental Systems Research Institute (ESRI), has performed a major portion of the data-base construction, integration, and applications development tasks (ref. 3). The San Bernardino County Planning Department and the San Bernardino National Forest, the two governmental participants, have provided several data sets, as well as acting as the "clients" for the applications studies. The Southern California Edison Company provided land-use data sets for the valley portion of the project area.

Ames Research Center (ARC), through its Technology Applications Branch, provided Landsat image processing and technical assistance during the data-base integration phase. This report describes the Landsat image processing carried out at Ames to support the San Bernardino Project.

The project study area (fig. 1) consists of the southwest corner of San Bernardino County, California. The study area comprises 750,000 acres (1170 miles) and consists of the urban-agricultural San Bernardino Valley, the brush and forest covered San Bernardino Mountains, and a portion of the Mojave Desert.

Two Landsat scenes were utilized by the project. The baseline scene was an August 6, 1976 image originally used by the Jet Propulsion Laboratory (JPL) for an earlier cooperative project between NASA and the California Department of Forestry (CDF). The choice of this image, part of a statewide mosaicked data set including Landsat multispectral scanner (MSS) raw and classified data, as well as digital terrain, was an essential part of the overall vertical data integration theme. A near-anniversary date (July 22, 1979) Landsat update image was also chosen to evaluate changes and data base updating potential. The original unmanipulated August 1976 computer compatible tape (CCT) used by JPL for the California Department of Forestry was obtained for use in the change detection portion of the San Bernardino project. The processed JPL raw data were not used for change analysis (but were used during classification) because of the extensive spectral and spatial manipulation applied to the data. The 1976 and 1979 images differed in

*Staff Analyst, Technicolor Graphic Service, Mountain View, CA.

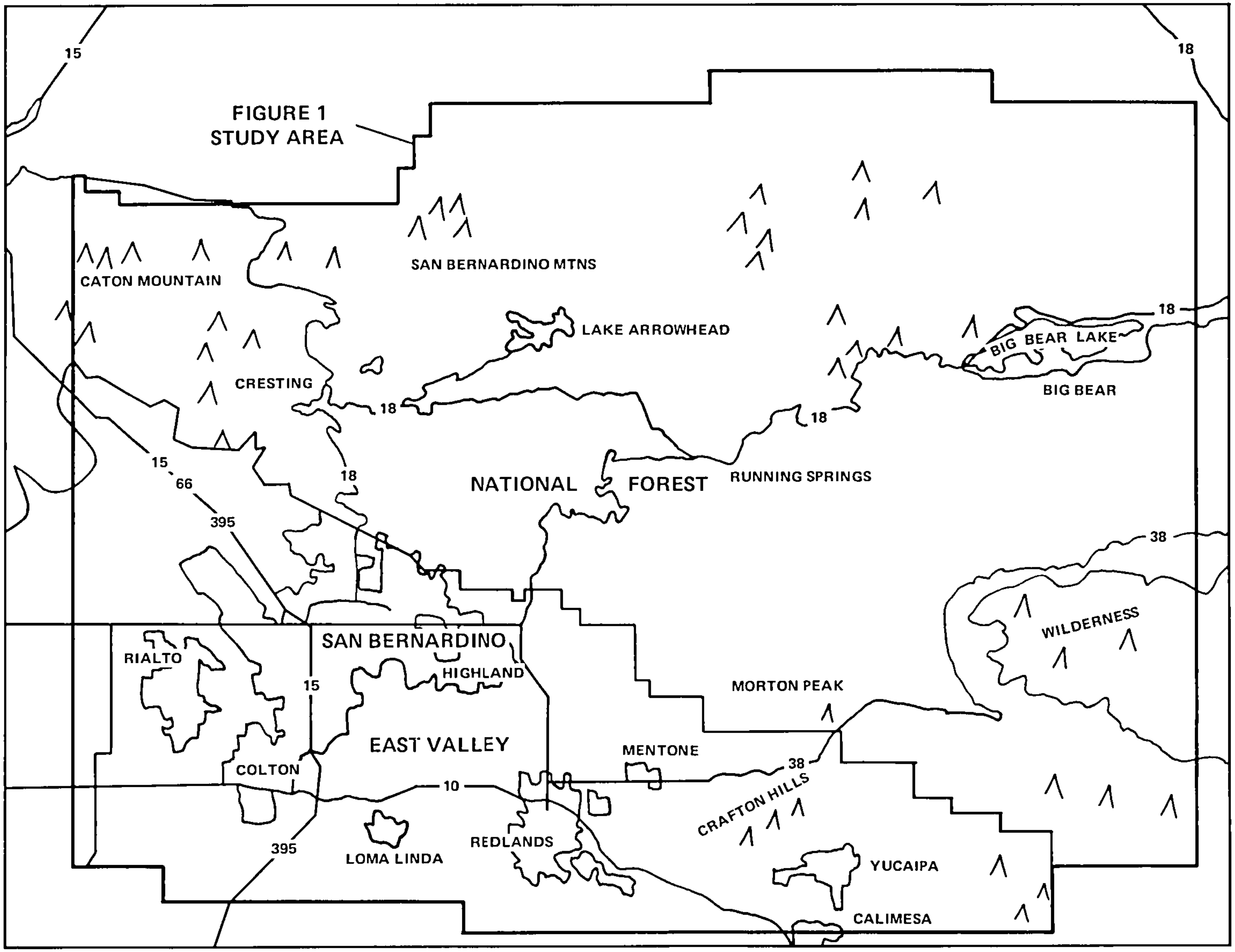

Figure 1.– Project study area.

several significant ways. First, the 1976 data had not been subjected to radiometric and geometric preprocessing by EROS; the 1979 data were in the EDIPS format, with haze correction and contrast enhancement. Secondly, the August 1976 image followed an unusually dry winter in southern California, while the July, 1979 image followed a wet winter with extensive mountain snow coverage.

Several other data sets for various portions of the study area were integrated by ESRI to produce a prototype integrated environmental planning data base. Data sets included 1974 and 1979 land-use, photo-interpreted from low-altitude black-and-white photographs; integrated terrain unit (ITUM) data for the San Bernardino National Forest (SBNF) area, containing several geologic, cover, soil, and physiographic data items; a partial extension of the terrain unit mapping to the valley floor area performed by ESRI; county general plan support data; and census geography from the local GBF/DIME program (ref. 4). Many of these various data sets were registered to an intermediate 1-acre grid used during much of the Landsat processing.

The 1976 Landsat data analysis is described in the next section; it is followed by a discussion of the 1979 Landsat analysis and the 1976-1979 change-detection analysis. Landsat land-cover codes and information about acreages and accuracies are presented in the appendixes.

We wish to thank the following people for their support during the conduct of this project: Jerrold Christenson and Russel Michel, ESRI; Craig Gooch, Ron Maytias, and Kenneth Topping, San Bernardino County Planning Department; Gay Almquist, James Bridges, Jo Bridges, and Jeanine Derby, San Bernardino National Forest; Len Gaydos, USGS/Ames Research Center; Ethel Bauer, Don Card, and Robert Wrigley, Ames Research Center; and Douglas Alexander, Frank Croft, and Kenneth Weinstock, Technicolor Graphic Services.

1976 LANDSAT DATA ANALYSIS

Data Set Description

The August, 1976 scene consisted of two mosaicked 1° X 1° quadrangles from the CDF project statewide data set (ref. 5). This data set was prepared by JPL through geometric and radiometric mosaicking of 32 Landsat MSS scenes, most of which were acquired during August 1976. During mosaicking, the data were registered to a rotated Lambert conformal conic projection grid, with a nominal cell size of 80 meters square. The full data set was then subsectioned into fifty-four 1° X 1° sections to facilitate file-handling operations during subsequent classification. Defense Mapping Agency digital terrain files, obtained from the USGS National Cartographic Information Center, were used by JPL to generate registered elevation (40-ft quantization), slope, and aspect files for use in the classification process. The Landsat MSS data had been classified using an unsupervised classification process with extensive pre-classification stratification (ref. 6). This process was executed by (1) dividing the state of California into 32 ecological zones (or "ecozones"), (2) unsupervised clustering of spectral data within each ecozone, (3) maximum likelihood classification of each quadrangle once spectral classes had been developed for all ecozones in that quadrangle, creating an ungrouped classified image, (4) identifying cover types for each of the classes developed, and (5) creating a grouped classified image consisting of the aggregation of spectral classes into a land-cover type categorization developed by CDF. The bulk of the clustering and classification was performed on the EDITOR image processing system at ARC.

The San Bernardino study area includes portions of four CDF ecozones, each with about 40 ungrouped spectral classes. Additionally, elevation data had been used in one of the ecozones to split roughly a dozen classes at the 5000-ft elevation point into a "high-mountain" zone to better differentiate conifer types, and a low-mountain component. A total of 178 spectral classes had been developed for the study area. Resource labels were assigned to each spectral class after extensive review of each quad using the interactive color display. Photo-interpretation and label assignments were performed by both NASA and CDF staff. The final grouped classification included 16 land-cover types, with emphasis on forest covers.

The classification produced over the study area during the CDF project lacked sufficient detail for the SBC project. This was particularly true in the urbanized valley area, primarily because of a lack of emphasis on stratifying and developing detailed classes for urban areas. Illustrative of the problems with the CDF data set was the complete omission of agricultural land in the valley, which was a part of the Los Angeles basin ecozone. Since agriculture was a very minor constituent of the ecozone, no agricultural classes were identified. Brush areas in the valley were similarly overgeneralized as "urban."

Because of the "vertical integration" objective of the project, it was decided to proceed with the utilization of the CDF data set despite these difficulties. The relabeling of spectral classes present in the ungrouped classification was chosen as the best means of generating an improved classification within the available time and resources. Formulating new "spectral class-to-land cover" assignments would allow the development of classes that would be both more accurate and specifically tailored to the needs of the County and the Forest Service.

Spectral Cluster Label Modification Methods

A previous pilot vertical-integration project completed by ESRI and ARC in 1978 suggested a rationale for cluster relabeling (ref. 7). This project, conducted in the Redlands, California 7.5′ quadrangle in the east-central portion of the study area, generated a small test data set consisting of the CDF 1976 Landsat classified data, 1974 and 1977 photo-interpreted land use, and several other data sets. Contingency tables were generated between the CDF 1976 ungrouped Landsat classification and both the 1974 and 1977 land-use data for use in relabeling Landsat spectral classes. Spectral classes were labeled according to their predominant land-use category. For those few classes without a clear majority, compound labels (i.e., grass/baresoil) were developed. This approach was tentatively chosen for Landsat class relabeling for the SBC project.

Because of two factors, the task of relabeling CDF Landsat spectral classes was split into valley floor and mountain area: (1) CDF ecozone boundaries treat the mountain and valley areas in two separate ecozones (the two other ecozones present have limited coverage within the study area, and (2) the major supplemental data sources were photo-interpreted land-use mappings for the valley floor (1974) and a separate vegetation mapping (1977) as part of the Integrated Terrain Unit Map (ITUM) file for the San Bernardino National Forest. Initial relabeling efforts concentrated on analysis in the valley area.

Attempts to analyze the original CDF spectral classes in light of the contingency analysis revealed a much lower apparent correlation between Landsat classes and photo-interpreted land-use categories than anticipated, based on the earlier Redlands results. Although some spectral classes, generally with relatively few pixels, fit well with the land-use data, other caused substantial problems. Five spectral classes, together accounting for well over 50% of the intersection between the classified Landsat and land-use data sets, each correlated equally well with each of the following four photo-interpreted land uses: single-family residential, undeveloped-improved, orchards and vineyards, and brush. Two distinct situations were noted: (1) the area in question consisted largely of old, apparently abandoned vineyards and small farms invaded by low-density residential, presenting an extremely high-frequency spatial pattern; and (2) minimum-mapping-unit conventions used in developing the land-use map (nominal 3-acre land-use minimum mapping unit versus 1.3-acre pixel size for the Landsat data) had apparently resulted in identical land-cover types being placed in quite different land-use classes, depending on their situation.

It was initially thought that these problems could be overcome by partitioning the valley area into several large "block-strata," corresponding generally to 7.5′ quadrangles, in order to reduce overall diversity. This approach was tested, but it too produced unsatisfactory results, owing to high-frequency variations in spatial patterns of land use in the valley. The satisfactory operation of these techniques in the Redlands test site was probably attributable more to high scene-homogeneity than to small test-site size.

Development of Relabeling Model

In an attempt to solve the cluster-labeling problems resulting from low homogeneity, a new approach was developed. It used the 1974 photo-interpreted land-use data to develop 13 land-use-based strata for use as masks. A separate cluster relabeling table could then be developed for each stratum. This "microstratification" is conceptually similar to the pre-classification stratification approach used in the original CDF project.

The strata developed from the 1974 land-use data, including such broad groups as residential, orchard/vineyard, and brush, were used to mask the Landsat classified data such that each stratum could be produced as an image on an Interactive Digital Image Manipulation System (IDIMS) display (ref. 8). The residential image, for example, showed the Landsat spectral class for each pixel classed as residential in the 1974 land-use image. Both spatial orientation and class identification were greatly improved over previous approaches in which entire spectral classes were examined without reference to the land-use data. Several types of residential cover could be differentiated in the residential stratum, including some significantly sized grass and brush areas between developments that had been classed as residential by the land-use data. Differentiation between orchards and vineyards was possible within the orchard/vineyard stratum, and many areas cleared for development were also detected in the Landsat data. Several distinct, fairly narrowly defined orchard classes, including a mature-to-declining citrus class, were identified, as were several classes of declining or apparently abandoned vineyards.

A similar technique-development process was followed in the mountainous SBNF area. Contingency analysis between CDF grouped data and 1977 photo-interpreted vegetation revealed a high degree of disagreement and potential classification error. These differences were believed to be related to several factors: (1) mislocation (resulting from mis-registration within the original CDF data set) of the 5000-ft elevation strata use to separate certain conifer classes in the Landsat data; (2) relatively large minimum mapping units (nominally 40 acres but in practice much larger) and species-level classification in the photo-interpreted vegetation data, resulting in large amounts of "misclassified" intermixture within vegetation polygons; and (3) substantial unresolved spectral similarity between various chapparal and conifer types.

Six broad strata, described by combinations of normally geographically adjacent vegetation categories, were developed for use in postclassification stratification. Separate cluster-labeling decision tables were developed for each stratum, and in many cases for specific elevation and aspect substrata. Figure 2 describes the general model and structure and strata employed for both valley (land use) and mountain (vegetation) areas.

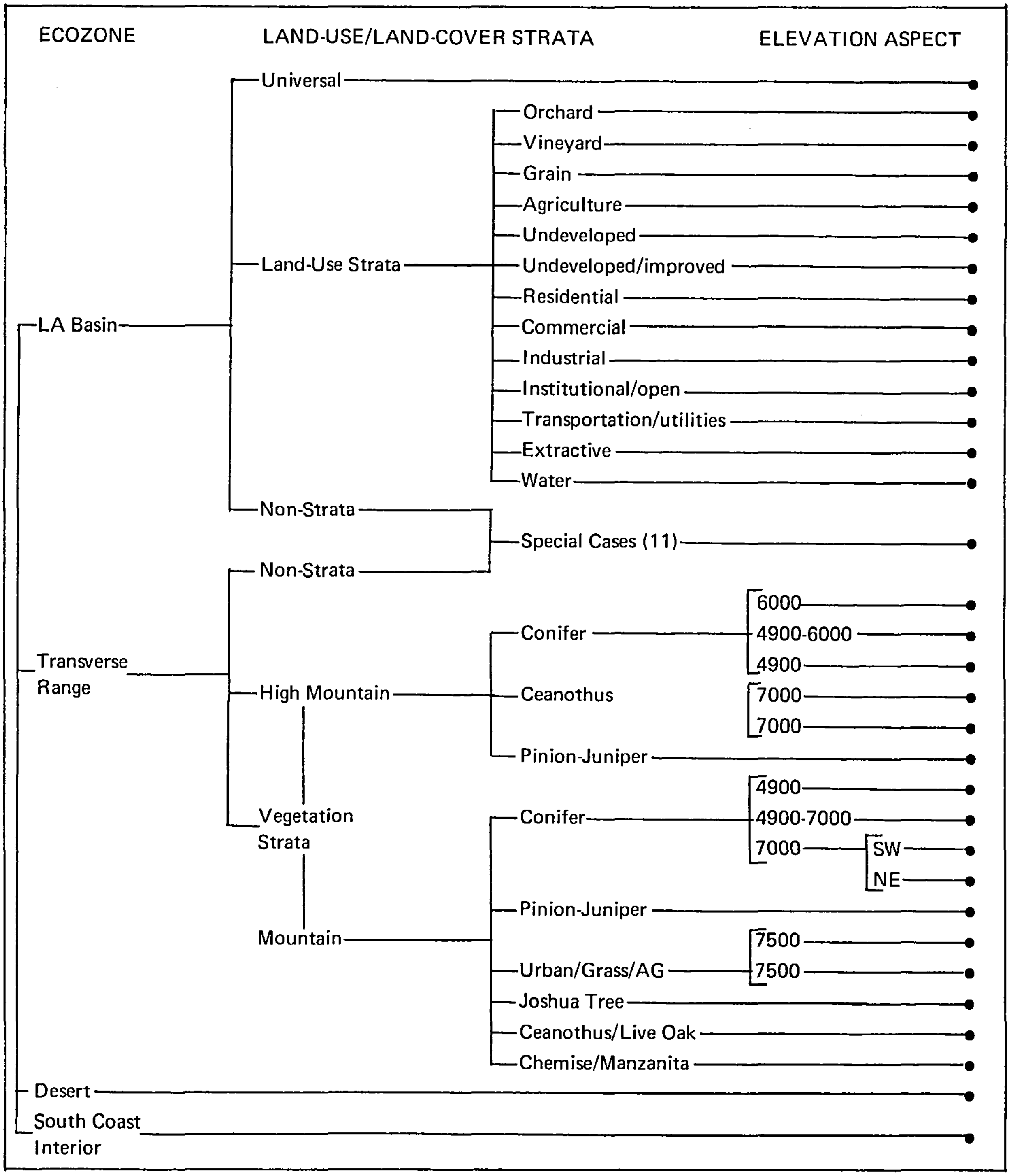

Figure 2.– Generalized model structure, showing 13 land use and 6 vegetation strata. Each filled dot at the right represents a cluster relabeling table for up to 40 clusters. Several of the "special case" tables also contain slope/aspect decisions.

Model Structure

A total of 49 land-cover classes (Appendix A) were developed for the composite valley/mountain area relabeling scheme. These relabeling decisions were first constructed as a large matrix that related ungrouped 1976 Landsat spectral classes to a total of 19 primary microstrata defined by land use and vegetation. Eleven special case microstrata blocks were delimited by line-sample location. Several strata were further subdivided by elevation and aspect, particularly in the mountain areas.

None of the conventional image-processing techniques available at ARC provided a workable method for implementing these decision rules without the generation of an extremely large number of intermediate images and logical masks. Instead, the decisions were incorporated into a model, using the ESRI GRID MAP/MODEL software. The several data layers required were registered and merged into a multivariable file (MVF), similar in structure to a pixel-interleaved-by-band image file.

A typical model statement consists of a range check that conditions the addition to or multiplication of an accumulator. "IF" statements can be concatenated with "AND" statements to generate logical forms equivalent to "IF Landsat class within rangel AND vegetation class within range2, then add x (in this case, the new class number) to the accumulator and skip to the designated instruction (usually the end of the model)." A full expansion of the decision matrix would yield approximately 175 Landsat classes x 19 microstrata x 6 elevation strata x 8 aspect strata or more than 160,000 pairs of model statements. This fully expanded decision matrix was not directly implemented in the model. By selectively processing certain cells in the matrix, it was possible to collapse remaining rows and columns with similar decisions, greatly reducing the required number of statements. Certain spectral classes, such as those representing water, were treated as "universals" and routed to the same decision table regardless of containing strata. The model evaluated remaining decisions through decision-tree structures representing five smaller matrices: (1) land-use strata areas; (2) vegetation strata areas with "mountain" Landsat classes; (3) vegetation strata area with "high-mountain" Landsat classes; (4) desert ecozone; and (5) special cases, primarily gaps and overlaps among the various masks. The final model contained about 400 statements. Although some thought was given to optimizing the model by arranging most frequent cases to the top, no formal optimization was performed. The data layers used by the model included (1) 1976 ungrouped Landsat spectral class data; (2) 1974 landuse data; (3) SBNF 1977 vegetation data; (4) elevation; (5) topographic aspect; (6) line coordinate; and (7) column coordinate. Because of the complexity of the model and the size of the data sets, several iterations were required on both test and complete data sets before a satisfactory analysis was completed. Generation of the final classification required nearly 7000 CPU seconds on ARC's IBM 360/67, including generation of a complete line-printer map. Much of this time may be attributable to somewhat inefficient FORTRAN I/O under the TSS operating system.

Evaluation and Secondary Stratification

A brief review of the output classification was made by Ames, SBC, and SBNF personnel. No noticeable deficiencies were found in the valley, but several errors needing correction were noted in the mountains. These errors seemed to be of a systematic nature and appeared resolvable through the incorporation of rainfall information and revision of some terrain-based decisions. A second model was constructed to effect improvements, using as input the output from the 1976 classification model, 1976 classified Landsat, average annual rainfall, elevation, and aspect to create a final 1976 modeled classification.

The final version of the 1976 land-cover product was evaluated for accuracy. A stratified random sampling was used to select 8 x 8 pixel blocks within 11 land-cover strata. On average, seven blocks were selected per stratum, with each stratum consisting of a unique subgroup of the 49 land-covers mapped. The study area was gridded into a series of 8 x 8 pixel blocks, with blocks being assigned to strata based on the land-cover evaluation group having plurality in a given block. A total of 74 blocks were evaluated (109 plots originally selected), with accuracies being determined for most of the 49 land covers. Accuracies could not be determined for some land covers because of an insufficient number of pixels with those land covers being selected for evaluation. The 1976 Landsat-derived land cover for each block was compared to photo-interpreted land cover obtained from examination of 1976 aerial photographs. The evaluation compared each pixel in the Landsat data to the grid cell in the corresponding location in the 8 x 8 photo-interpreted block. The overall accuracy for the 1976 classification was computed by summing the accuracies after weighting by the number of pixels in each individual class.

The Landsat classification was evaluated at several levels of thematic detail (see appendix B). The original land-cover mapping was at a high level of thematic detail, approximately equal to a modified information Level III, as defined by the U.S. Geological Survey (ref. 9). The U.S. Geological Survey's Level I and II specifications were modified to allow the Landsat land-cover data to be evaluated at more generalized informational Levels as well. Overall classification accuracies at Levels I, II, and III were 76%, 69%, and 52%, respectively. There were several possible factors that adversely affected the accuracy of the 1976 classification: (1) poor spatial integrity of the Landsat data (the data were resampled once during the CDF project and once more in the San Bernardino project, with potentially detrimental effects; (2) the unsupervised classification technique used as part of the CDF project, and subsequently used in the San Bernardino project, may have been inadequate for this area and application; (3) emphasis on mapping Level III features may have resulted in lowered Level I and II accuracies; (4) analysts had difficulty in photo-interpreting the land-cover classes developed on a per-pixel basis, and (5) the classification modeling technique may not be as suitable as hoped, or may have been less than optimally executed because of inexperience. Level III land-cover classifications are rarely attained, so the 52% accuracy obtained at this level is not necessarily bad.

CONCLUSION

The advantages of integrating Landsat classified data with other data sets in the context of a geographic information system can include (1) potential improvements in the spatial resolution of the data base; (2) identification of land-cover features not previously included on other data layers; (3) the use of Landsat on a periodic basis to provide updates; and (4) potentials for geographic area expansion of the data base, using signature-extension techniques. The combination of spectrally based Landsat MSS classification with a more spatially based aerial photograph classification may provide useful information not readily available from independent analysis of either data set. In this project, only one potential avenue for combining two such data sets was investigated. The specific technique examined here involves using several data base layers including photo-interpreted land use, to develop a postclassification microstratification for a hierarchical-cluster-labeling model for use with an existing unsupervised classification. This technique significantly reduced confusion between spectral classes and allowed the development of more sophisticated thematic information than would have been possible with a conventional unsupervised classification. The Landsat classification produced is more compatible with other data layers, and contains fewer of the classification irregularities that sometimes result from mis-registration. Yet to be examined are possibilities for obtaining better results through (1) using hierarchical modeling techniques in conjunction with supervised classification data; (2) semiautomated training site selection; or (3) extending the modeling technique to include pre-classification stratification. Using some or all of the layers of the data base to model the classification of Landsat data to be incorporated can result in thematic classes more in tune with data-base layers, higher thematic detail, and better spatial registration of the product to other data base layers.

1979 LANDSAT ANALYSIS AND 1976-79 CHANGE-DETECTION ANALYSIS

Introduction

The San Bernardino County Planning Department had expressed interest in detecting urban changes and monitoring growth. In one possible method of detecting change with Landsat, a direct pixel-by-pixel comparison between independent image classifications from each of the two dates would be used. This classification comparison approach would yield both information on whether change had occurred, as well as the combinations of "from-to" changes taking place. Previous studies, including a parallel CIRSS project being conducted by researchers at the University of California-Santa Barbara (ref. 10), have found several problems with this approach. Given a random distribution of errors within each classification, owing to both classification and minor registration errors, the accuracy of the detection of correct "from-to" change combinations would tend toward the product of the accuracies of the two classifications. Given two independent classifications, each with an accuracy of 80%, their ability to correctly portray accurate information for both dates would be only 0.8 x 0.8 = 0.64, or 64%. As the assumption of random error among classes is not likely to be met, actual errors might vary widely. Two independently conducted classifications may result in either dependent or independent data. Reported results in the literature do not yet describe the full relationship between errors during classification comparisons. The multiplicative error hypothesis was taken as being descriptive of the problem, and procedures were examined which would result in less error than the multiplicative case. Whether errors are multiplicative, or conform to some other relationship, any errors directly attributable to the comparison process are bound to be undesirably larger than the actual change in a given area, as typical high-growth areas experience only about one or two percent change per year. Provided that image registration and radiometric corrections are satisfactory, lower errors might be expected from either (1) a single classification of an image containing spectral information from two dates or (2) a classification of only those pixels in the second date identified as "change" pixels through an independent analysis. The second of these two options was chosen for demonstrating change analysis and land-use updating; that is, a two-step process, a detection of change and determination of "from-to" combinations. As a further means of minimizing multiplicative comparison errors, a modeling process was used in the classification of the update image within change areas to constrain changes to likely combinations (table 1).

TABLE 1.– MATRIX OF ALLOWED CHANGE POSSIBILITIES

1977 land-cover types	1979 Land-cover types[a]					
	Urban	Agr.	Range	Water	Forest	Barren
Urban	x	–	–	–	–	–
Agriculture	x	x	x	x	–	x
Range	x	x	x	x	x	x
Water	– –	–	–	x	–	x
Forest	x	–	x	–	x	x
Barren	x	x	x	x	–	x

[a]Symbol x indicates an allowed change; "–" indicates prohibited change.

A mask of possible change areas between the 1976 baseline image and the 1979 update Landsat image was created, using several change-detection techniques. The "change image" thus developed was used as a mask to limit analysis of the 1979 update Landsat image. The use of a change mask to limit areas classified in an update image has previously been examined by Angelici as a means of reducing computer computation time and minimizing erroneous changes from occurring when two independent classifications are compared (ref. 11). A similar approach was also undertaken by Todd (ref. 12).

Data Preparation

The first step in the multidate analysis was construction of a composite 1976-1979 image. Because of the geometric problems with the 1976 JPL-processed MSS data, it was necessary to revert to the original CCT. After de-skewing (oblique correction) and de-stripping the 1976 data, common ground-control points were chosen in each image. A second-order polynomial transformation, along with a cubic-convolution resampling algorithm, were used to register the 1976 image to the 1979 image. Histograms of the two images revealed considerable differences in both means and variances in each channel. Most of these radiometric differences were not related to per-pixel change, but rather to atmospheric differences, complex cover-variant reflection changes owing to sun angle (ref. 13), satellite performance parameters (Landsat I versus Landsat III), substantial ground-moisture differences, and ground-processing system differences. Rather than invest the considerable time that might be involved in an attempt to analytically remove each of these factors, a histogram normalization procedure, similar to that used in destriping, was developed and applied band-by-band to the 1979 image. The procedure developed used a contingency analysis between 1976 and 1979 spectral values. The contingency table was used to generate a histogram of 1979 spectral values for each 1976 digital spectral value. These histograms were then examined for modes (local maximums) in order to determine for a given 1976 digital number the corresponding 1979 digital number. For example:

1976 digital number	1979 digital number
0	0
1	2
2	4
3	5
.	.
.	.
.	.
127	119

In some instances, several 1979 digital number modes corresponded to a single 1976 value. These were resolved through interpolation between the 1979 digital number modes for the 1976 digital numbers that are centered about the 1976 value in question in order to insure that the 1979 values increased monotomically. The resulting 1976 to 1979 digital value correspondence table was applied as a piecewise linear mapping function with 128 steps.

Change Analysis

Following the normalization process, four potential change-detection methods were developed: (1) unsupervised classification of a 5-7-5-7 multidate image (a four-band image containing 1976 bands 5 and 7, and 1979 bands 5 and 7; (2) band-by-band thresholding of a four-band 1976-1979 spectral arithmetic difference image; (3) unsupervised classification of the band-difference image; and (4) a multiband chi-square analysis of the band-difference image. Upon subjective visual analysis, the results of methods 1, 3, and 4 were found to produce potentially adequate delineations of change. Rather than choosing a "best" approach, these three images were combined and spatially filtered, as explained below, to form a composite "change mask." This change mask was used to limit the 1979 unsupervised classification to those areas with high likelihood of change. The final 1979 land-cover update classification was created by a model that grouped 1979 image unsupervised spectral classes for change areas into cover types; the cover types were based on spectral characteristics, the 1976 baseline classification, elevation aspect, SBC General Plan designation, and in special cases, distance of a spectral class to water. For nonchange areas, the 1976 baseline classification, values were retained. By combining the new classification for the change-mask areas with the baseline classification for nonchange-mask areas, multiplicative errors, when comparing the baseline and update images for change detection, were limited to areas within the change mask. The effect of these errors was further reduced by using a "change possibilities" matrix in the 1979 classification model to prohibit the accuracy of unlikely "from-to" change combinations (table 1–. Change from agriculture to urban is "likely" to be a real change, and was therefore allowed; apparent change from urban to agriculture probably represents classification error and is an examply of those types of changes not allowed in the 1979 classification.

Unsupervised classification of multidate image– The four band 5-7-5-7 multidate image underwent unsupervised clustering for 63 clusters (number arbitrarily chosen), followed by classification. Seven clusters were determined to represent change, by examining the classification on a color video display and comparing the classification with land-use and cluster spectral plots. An image of change areas was generated and incorporated along with the products from clustering and chi-square analysis of the difference image. Subsequent comparison of the multidate change classification with the photo-interpreted land-use change for the Redlands area indicated that one of the clusters selected as representing change was not a significant change indicator.

Thresholding 7% tails of the distribution in difference image– The spectral values of the difference image as originally created ranged from -127 to +127, with no change centered on zero. The final difference image had the no-change mean shifted to 127, resulting in a data range from 0 through 255. Histograms were generated of each band and were used to define breakpoints in the distribution tails that were useful in delineating change. The 7% tails of the spectral distribution in each of the four bands were selected as best delineating change based on evaluation of the histograms and of the image, using a color video monitor. The areas delineated in each band were summed to yield the change image. Visual comparison of this product with those from the other change detection approaches showed the thresholded product to be much noisier. Comparison of the distribution threshold product with aerial photography showed that most of areas delineated did not in fact represent useful thematic change. The thresholding of the spectral distribution of the difference image was not further investigated, because of noise and excessive commission errors. Additionally, this approach required more work than the other change detection approaches examined. A detailed evaluation to determine whether the selection of different breakpoints would yield better results was not undertaken.

Unsupervised classification of difference image– an unsupervised clustering followed by classification was carried out on the difference image. The classification was evaluated using a color monitor in conjunction with land-use and cluster spectral plots. Ten clusters of the 30 developed were identified as indicating change and were grouped to create an image depicting change areas. This change image was then combined with the change products created by multidate image classification and difference-image chi-square analysis. Later, detailed comparison of the difference classification with the 1976-79 photo-interpreted land-use for the Redlands area indicated that three of the clusters originally selected as change may not be significant change indicators.

Chi-square analysis of difference image– Using EDITOR image analysis software at Ames Research Center, it was possible to cluster and classify the four-banded difference image for just one class (the mean of which corresponded to the mean difference between the two images) and, as a by-product, obtain a threshold value for each pixel classified. The resulting threshold values generated by the algorithm were chi-square values. The chi-square values define how well each pixel has fitted the general population; in this case change pixels should fit poorly and thus have high values. The image of chi-square values was compared with aerial photographs and photo-interpreted land-use data in order to select a chi-square breakpoint score that was significant in detecting change. Chi-square scores of 7.779 and over (chi-square probability score less than 0.1) were determined to indicate pixels with significant change. An image delineating these pixels was created and incorporated into the composite map of change areas. Later, detailed comparison of the chi-square image to the edited land-use change for the Redlands area confirmed that in this case chi-square scores above 7.779 were significant indicators of change and that lower scores were not.

Development and Use of Composite Change Mask

A change image resulted from each approach just discussed. These were assessed by the project participants through subjective comparison with land-use data and through participant knowledge of the study area. As a result of this assessment, the thresholding of the distribution tails of the band-by-band difference image was discarded as a useful change indicator owing to excessive noise. The other three change detection indicators were merged to create a composite map of change areas. Change pixels in the composite image consisted of all those pixels flagged as change by two or more of the input images. The composite image was spatially smoothed using a 3 x 3 filter to minimize noise through removal of isolated change pixels and isolated nonchange pixels within change areas. Todd had previously cited noise removal algorithms as having potential for improving change-detection products (ref. 12). The smoothed composite change image was then used as a mask to limit classification of the 1979 Landsat update image to those areas that were determined to be possible change areas. A matrix of change possibilities and a variety of other ancillary data were used to constrain the classification to likely "from-to" combinations, thereby minimizing to some extent the introduction of multiplicative classification errors when the 1976 and 1979 classifications are compared.

Evaluation of Change Products

During the preparation of the multidate and difference image products to be incorporated in the composite change mask, the emphasis had been minimizing omission-of-change features. The three change-image inputs to the composite were developed as the result of rather cursory analysis and without detailed evaluation. The compositing of several change products was undertaken as a means of reducing commission errors without the necessity for detailed change-product evaluation. Detailed evaluation of the various change products was only carried out after all processing was completed. After-the-fact evaluation had the advantage of testing whether the process of compositing several minimal-effort change-detection products was an effective means of generating a product as accurate as one resulting from a more thorough analysis for change, using a single change-detection technique.

The various Landsat change data were compared with previously created (as part of a Southern California Edison effort) 1974 and 1979 land-use generated by a combination of aerial photographic interpretation and examination of zoning and other collateral planning data. Initial comparisons of the photo-interpreted land use and the Landsat-based change data showed several irregularities which had to be removed. The photo-interpreted 1974-79 land-use data revealed unlikely changes resulting from two factors: (1) slight mis-registrations between the 1974 and 1979 land-use data sets, and (2) coding errors that were present in the 1974 land use which were not duplicated in the 1979 land-use data, thus resulting in apparent changes where no changes had occurred. Another problem was that many of the changes detected by comparing the photo-interpreted 1974 and 1979 land-use data sets occurred before 1976, and thus were not changes in the 1976-1979 timeframe period of the Landsat data sets. Manual editing on a pixel-by-pixel basis of the photo-interpreted land-use change data through comparison with 1976 and 1979 aerial photographs was used to minimize these extraneous irregularities and maximize the comparability of the photo-interpreted land-use and Landsat data. An additional factor introducing irregularities was not compensated for: namely, that the photo-interpreted land-use information mapped a different class of features than the land-cover-oriented information extracted by Landsat.

Data editing and subsequent evaluations were limited to a portion of the Redlands 7.5′ quadrangle in order to minimize the scope of editing required. All Landsat change products (both the composited and uncomposited products) were evaluated against the edited land-use data. Changes in the Redlands quadrangle were considered representative of types of changes in the rest of the urbanized valley floor, but not those in the San Bernardino Mountains. The Redlands site was also attractive because 1976 and 1979 photographs were available, and because previous work had emphasized examining the Redlands area (1979). Limiting evaluations to the Redlands quadrangle does not provide definitive data on the accuracy of various Landsat change-products across the whole study area, but does allow inferences to be made of the performance of the Landsat data in the urban areas.

Table 2 summarizes the accuracies of the various change products when compared with land-use change data within the Redlands portion of the study area. The multidate image classification and the chi-square analysis of the difference image both yielded similar results. The chi-square approach is notable as requiring the least analyst evaluation of the four techniques explored. The multidate and difference classifications, and chi-square data were reevaluated using the following relationship to determine whether the change products developed from these approaches were the optimum possible: category (spectral class or chi-square interval) was determined to be a significant detector of change if $K_c/T_c > K_{nc}/T_{nc}$, where K_c is the number of pixels of the category in question that were identified as change by land-use ground-truth; T_c is the total number of pixels of change as identified by land-use ground truth; K_{nc} is the number of pixels of the category in question that were identified as nonchange by land-use ground-truth; and T_{nc} is the total number of pixels identified as nonchange by ground-truth.

The chi-square analysis was determined to have been optimally executed while both the multidate and difference classification change products incorporated in the composite change-mask were not the optimum possible. The assessment indicates that an optimized multidate classification would have an accuracy similar to the chi-square analysis, and that the optimized difference image still would have more commission error than the other techniques (table 2). The cause may be either the low number of cluster developed (30 for the difference image classification versus 63 for the multidate classification) or the possibility that difference-image clustering and classification is inherently less accurate. The relatively high commission errors in the difference classification have had only a small effect on the composite product, indicating that products may not need to be optimized, if a number of products are merged. In areas without preexisting change data it may in fact not be possible to create optimized change products using any one technique. Smoothing the composite change mask through

TABLE 2.– LANDSAT PRODUCT-CHANGE DETECTION ACCURACY COMPARED WITH LAND-USE GROUND-TRUTH FOR REDLANDS AREA[a]

Landsat change product	Total Landsat-identified change pixels	Landsat change pixels correct	Landsat change pixels, percent correct	Percent of real land-use change detected	Percent change omitted by Landsat	Percent changes committed by Landsat
Difference image 7% distribution tails	not evaluated	–	–	–	–	–
5-7-5-7 Multidate image classification	1273	258	20	47	53	80
Difference-image classification	2552	312	12	57	43	88
Difference image chi-square analysis	907	240	26	44	56	74
Change mask: unsmoothed	1213	253	21	46	54	79
Change mask: smoothed	528	252	48	46	54	52
Change detected by 1976 baseline and 1979 update classification comparison (irrespective of whether changes detected were properly identified as to "from-to" land-cover combination)	396	240	61	44	56	39
Optimized 5-7-5-7 multidate image classification	896	250	28	46	54	72
Optimized difference image classification	1622	293	18	54	46	82

[a]547 pixels were identified by photo-interpreted land-use as change. The area evaluated comprised 7161 pixels.

use of a 3 x 3 spatial filter greatly lowers commission errors without affecting the degree of capture of real changes, and appears to have potential for improving the per-pixel accuracy of any change-detection technique.

Despite the utilization of several techniques that have made change detection more accurate than it would otherwise have been, the actual land-use change-detection accuracy remains low. Many of the change features identified by Landsat but determined to be in error based on comparison with the photo-interpreted land-use data in the Redlands area, were agricultural and water-level changes and changes in characteristics of the Santa Ana river wash. These changes were not land-use changes but were counted as errors even though real spectral change had occurred. The land-use changes undetected by Landsat were, in order of occurrence, from (1) orchards and vineyards to vacant, (2) orchards and vineyards to residential, and (3) vacant to residential. The detailed evaluation of the Landsat products in the Redlands area indicate that no amount of spectral data manipulation would have made detection of about 50% of the omitted land-use change. Examination of aerial photographs revealed that some land-use changes around Redlands actually involved no land-cover change, particularly those changes from orchards and vineyards to vacant. About 75% of the commission errors appeared to be actual land-cover changes that were not land-use changes.

Changes from fallow to planted agricultural fields, small changes in the course of the Santa Ana wash, and the filling of water basins and gravel pits with water in 1979 composed most of the land-cover changes not present in the land-use data. No amount of spectral manipulation could have resulted in the removal of these features from the Landsat change data. Better change-detection accuracies therefore appear to be possible for applications stressing detection of land-cover rather than land-use, changes.

The comparison of various Landsat data with land-use change has allowed the generation of readily comparable accuracy figures for technique evaluation. These accuracies do not state the actual Landsat accuracies with respect to detecting land-cover change. The 1976 and 1979 classifications describe land cover, and not land use, and must ultimately be evaluated from the land-cover perspective.

A brief land-cover accuracy evaluation of the 1976 and 1979 land-cover classifications when grouped and evaluated at a generalized information level (urban, range, forest, agriculture, barren) was conducted by comparing a number of randomly selected change-mask features with photo-interpreted 1976 and 1979 ground-truth. Land-cover change detection accuracy was also evaluated, and overall results are summarized in table 3. As had appeared probable,

TABLE 3.– LANDSAT ACCURACIES WHEN EVALUATED AGAINST PHOTO-INTERPRETED LAND COVER

Pixels evaluated	77
Photo-interpreted land-cover pixels	
Change	29
Nonchange	48
Landsat land-cover classification comparison	
Change	29
Nonchange	48
Landsat land-cover pixels correct	
1976 baseline classification	57(74%)
1979 update classification	17(22%)
Landsat change pixels correct	35(73%)
Landsat pixels with correct from-to land cover (includes pixels correctly identified as having same cover on both dates)	12(16%)

the accuracy of Landsat-detected land-cover change is higher than that of land-use change (73% versus 61%). The 1979 land-cover accuracy is very low, possibly because the analysts did not have sufficient contiguous areas in the classification with which to make their evaluations, as a result of the small size (15 pixels average) and separated locations of the change-mask polygons. During analysis, the 1979 clustered data were superimposed over the 1976 classification to enhance analyst-perceived spatial relationships. Spectral plots were also heavily used, but these techniques did not appear to fully overcome the problem. Artificial expansion of change-mask polygons to yield more contiguous units or reliance upon supervised classification techniques that do not require post-clustering cluster identification may be needed. The from-to accuracy of 16% within the change mask is identical to that which would be predicted by multiplying the 1976 and 1979 land-cover accuracies. The 16% accuracy is also much lower than the 73% accuracy developed for the detection of 1976-1979 land-cover changes, without respect to the amount of correct from-to cover assignments. Applications that require the ability to detect change, without respect to type, appear to have much greater chances for success than those that require determination of the types of change. Use of Landsat change data as a detector of change areas to be further evaluated by photo-interpretation and ground-truthing may be desirable, particularly if the number of change areas is small.

By using tables 2 and 3, it is possible to determine if the overall technique used in the project, the chaining of baseline and update image classifications through use of change-detection and masking techniques, has successfully reduced multiplicative comparison errors when two classifications are compared for changes. If we are charitable and assume that an independently completed 1979 classification would have the same accuracy as the 1976 classification, the predicted accuracy for detecting changes when comparing the two classifications would be 0.74 x 0.74 = 0.55, or 55%. The accuracy actually achieved can be predicted as:

$$A = A_i F_i + A_o F_o$$

where

A = overall accuracy
A_i = accuracy within change-mask boundaries
F_i = fraction of image covered by change mask
A_o = accuracy outside change mask (after omission error is considered)
F_o = fraction of image covered

or $(0.73 \times 0.9856) + (0.16 \times 0.0144) = 72\%$.

Conclusions

Land-cover updating using change-detection and masking techniques requires (1) the creation of an initial baseline classification, (2) use of spectral change analysis techniques to generate a map of possible change areas, (3) the classification of the possible change areas within the update image data, and (4) creation of a final update classification by combining the classification of the baseline image for the nonchange areas with the update classification of the change areas. Generating an update classification only within areas defined as having changed, results in fewer errors than would result from comparing two independently generated, full-image, land-cover classifications. Minimization of comparison errors between co-registered products is of particular importance to geographic information system users who interact with a multitude of co-registered data sets. The technique also requires less effort than the generation of two separate classifications, for the area to be evaluated in the update image has been greatly reduced. Yet to be examined is the issue of how well the technique can be repeated, with change-detection and masking techniques being used between new Landsat images as they are obtained, and what may be only the latest of a series of chained classifications.

Several other techniques were also shown to improve Landsat change-detection accuracies over some standard methods. Spatial smoothing of Landsat-derived change-detection products greatly minimizes commission error, with little or no increase in omission error. Conducting a single-date land-cover classification of the update image-change within areas delineated as change by conventional techniques, followed by comparison with a baseline classification, also reduced commission error, with small effect on omission error. Using several different change products in combination required less ground-truthing and change-product evaluation than would have been required to generate a comparable quality product using a single change-analysis approach.

Mapping conducted based on the concept of extracting land use was found to yield unresolvable differences in results from mappings based on the extraction of land-cover data. It appears that spectral change, inherently an indicator of land-cover change, accounts for only about half of the real change in land use. Therefore, applications emphasizing the detection of land-cover rather than land-use changes will have greater chances of success.

Ames Research Center
National Aeronautics and Space Administration
Moffett Field, California 94035, May 12, 1982

APPENDIX A

LANDSAT LAND-COVER CODES

The Landsat data have been evaluated at several levels of thematic detail. The original land-cover mapping was at a high level of thematic detail, approximately equal to a modified information Level III, as defined by the U.S. Geological Survey (ref. 10). The data can also be used at more generalized information levels by grouping land-cover types. The U.S. Geological Survey's Level I and II specifications were modified to allow the Landsat land-cover data developed in this project to be interpreted at more generalized informational levels (generalization has been conceptual only—there was actual modification of digital values). Use of the data at generalized, rather than detailed, levels of information may be desirable either because detailed information is not required, or because the detailed information is insufficiently accurate.

The following land-cover legends describe both the 1976 and 1979 Landsat land cover classifications.

LANDSAT LAND-COVER DIGITAL CODES

Information Level I

Codes	Land Cover
18, 19, 23, 24, 27-33	Urban
6-12, 17, 34	Agriculture
3-5, 15, 16, 44-49	Rangeland
25	Water
13, 35-43	Forest
1, 2, 20-22, 26	Barren

Information Level II

Codes	Land Cover
28-33	Residential
23, 24, 27	Commercial
18, 19	Other urban

Information Level II (Continued)

Codes	Land Cover
17	Crops
6-12, 34	Orchards and vineyards
15, 16	Grass
3-5, 44-49	Brush
25	Water
13, 41	Deciduous forest
35-40, 42-43	Coniferous forest
1, 2, 21, 22, 26	Mixed bare
20	Extractive

Information Level III

Code	Cover
28	Residential: with trees
29	Residential: newer irrigated
30	Residential: sparse cluster
31	Residential: large lot unirrigated
32	Residential: rural strip
33	Mobile homes/high-density residential
23	Structures
24	Structures: strip commercial
27	Structures: with brush
19	Concrete
18	Asphalt

Information Level III (Continued)

Code	Cover
17	Crops
6	Orchards: young
7	Orchards: moderate vigor
8	Orchards: mature
9	Orchards: declining
34	Vineyards: moderate vigor
10	Vinyards: moderate vigor
11	Vineyards: high vigor
12	Vineyards: declining
15	Lush grass
16	Dry grass
3	Sparse brush
4	Moderate density brush
5	Thick brush
46	Chamise
47	Chamise/ceanothus mix
45	Ceanothus/scrub oak
44	Bracken fern/ceanothus
48	Coastal sage
49	Sagebrush
25	Water
13	Woodland
41	Riparian mixed hardwoods
35	Big cone Douglas fir
36	White fir
37	Jeffrey pine: closed understory

Information Level III (Continued)

Code	Cover
38	Jeffrey pine: open understory
42	Jeffrey pine: ceanothus
43	Coulter pine
39	Lodgepole/limber pine
40	Pinyon/juniper
21	Cinder
1	Cleared
2	Bare
22	Slag
20	Extractive
26	Snow

LEVEL III LANDSAT CLASSIFICATION LAND-COVER LEGEND

(The digital data value representative of each land-cover class is noted in parentheses.)

WATER (25): Lakes, ponds, other standing water

CLEARED (1): Vegetation has recently been removed and area is now essentially bare

BARE (2): Exposed soil, rock, and snow

CINDER (21): Railroad roadbed rock and other blackened rock

SLAG (22): Industrial waste tailings

EXTRACTIVE (20): Gravel pits and other mining resulting in exposed and disturbed soil

LUSH GRASS (15): Lush grasses and forbs

DRY GRASS (16): Grasses and forbs that have dried and yellowed

SPARSE BRUSH (3): Less than 30% brush closure, with grass or exposed soil understory

MODERATE DENSITY BRUSH (4): 30%-70% brush cover with grass or soil understory

THICK BRUSH (5): Over 70% brush cover

WOODLAND (13): Trees undifferentiated as to type, with over 50% crown closure

HIGH VIGOR VINEYARD (11): Vineyards with bright infrared reflectance

MODERATE VIGOR VINEYARD (10): Vineyards with moderate infrared reflectance with some soil and grass, and a few dead vines

LOW VIGOR VINEYARD (34): Vineyards with low chlorophyll content, often not presently cultivated

DECLINING VINEYARD (12): Poor condition, possibly uncultivated with numerous dead vines

MATURE ORCHARD (8): Orchards with over 70% crown closure

MODERATE VIGOR ORCHARD (7): Orchards of intermediate age and vigor, with between 45% and 70% crown closure

YOUNG ORCHARD (6): New orchards characterized by young, small trees with less than 45% crown closure

DECLINING ORCHARD (9): Orchards with less than 70% crown closure with large intermittently spaced bare patches indicative of dead trees

CROPS (17): Agricultural plantings other than orchards and vineyards

RESIDENTIAL WITH TREES (28): Wooded residential lots, generally older neighborhoods with established trees

RESIDENTIAL–NEWER IRRIGATED (29): New residential neighborhoods without developed large trees, but with well-watered lawns

RESIDENTIAL–SPARSE CLUSTER (30): Sparse, generally small, residences occurring in clusters amidst largely undeveloped lands

RESIDENTIAL–LARGE LOT UNIRRIGATED (31): Sparsely spaced residences on large lots of dry grass or brush

RESIDENTIAL–RURAL STRIP (32): Sparsely spaced residences facing onto roadways, with large vacant brush or grass covered areas to their rear

MOBILE HOMES/HIGH-DENSITY RESIDENTIAL (33): Trailer parks and apartment complexes with a large amount of roof area

STRUCTURES (23): Buildings, predominantly nonresidential

STRUCTURES-STRIP COMMERCIAL (24): Nonresidential buildings sparsely located along highway corridors

STRUCTURES–WITH BRUSH (27): Predominantly nonresidential buildings surrounded by brush-covered lots

CONCRETE (19): Concrete parking lots, roofs, some drainage surfaces, and road surfaces

ASPHALT (18): Asphalt-covered parking lots, roofs, and road surfaces

COASTAL SAGE (48): Over 30% vegetative cover of buckwheat, and other coastal sage types

CHAMISE (46): Over 30% crown closure of chamise-dominated chaparral

CHAMISE/CEANOTHUS (47): Over 30% crown closure of chaparral co-dominated by chamise and ceanothus

CEANOTHUS/SCRUB OAK (45): Over 30% crown closure chaparral dominated by either ceanothus, scrub oak, or a mix thereof. Some manzanita or black oak may also be present

BRACKEN FERN/CEANOTHUS (44): Over 70% vegetative cover, dominated by a mix of bracken and ceanothus

RIPARIAN MIXED HARDWOODS (41): Over 30% crown closure of riparian woodland

BIG CONE DOUGLAS FIR (35): Over 30% crown closure of big cone Douglas fir

WHITE FIR (36): Over 30% forest crown closure, dominated by white fir

JEFFREY PINE MIXED COMMUNITY–CLOSED UNDERSTORY (37): Over 50% forest crown closure dominated by Jeffrey pine, and mixed with incense cedar, sugar pine, and black oak with ceanothus understory

JEFFREY PINE/CEANOTHUS (42): 30%-50% forest crown closure dominated by Jeffrey pine. Over 50% ceanothus-dominated chaparral in understory

JEFFREY PINE-OPEN UNDERSTORY (38): Over 30% Jeffrey pine crown closure, with young ponderosa and grasses in the understory

COULTIER PINE MIXED FOREST (43): Over 30% forest crown closure, dominated by Coultier pine, and containing varying mixes of incense cedar, sugar pine, black oak, pinyon, or juniper

LODGEPOLE/LIMBER PINE (39): Over 30% lodgepole or limber pine crown closure

PINYON/JUNIPER (40): Over 30% crown closure of either pinyon pine, juniper, or a mix thereof

GREAT BASIN SAGE (49): Over 30% great basin sage vegetative cover

BACKGROUND (0): Some areas south of the San Bernardino county line for which no Landsat classification was carried out (principally areas covered by the CDF South Coast Interior Ecozone

APPENDIX B

1976 LANDSAT LAND-COVER ACREAGES AND ACCURACIES

Level I
(Overall accuracy of Level I data = 76.5%.)

Class name: Urban
Accuracy = 74.19%
Commission errors = 25.81%
Omission errors = 19.68%
Acres identified = 47,149
Estimator = 0.924
Modified acreage = 43,552.5
Digital codes: 18, 19, 23, 24, 27-33

Class name: Agriculture
Accuracy = 75.76%
Commission errors = 24.24%
Omission errors = 30.50%
Acres identified = 20,601
Estimator = 1.091
Modified acreage = 22,473.8
Digital codes: 6-12, 17, 34

Class name: Rangeland
Accuracy = 76.78%
Commission errors = 23.22%
Omission errors = 33.84%
Acres identified = 464,046
Estimator = 1.160
Modified acreage = 538,530.6
Digital codes: 3-5, 15, 16, 44-49

Class name: Water
Accuracy: Not evaluated
Acres identified = 3,442
Digital code: 25

Class name: Forest
Accuracy = 79.19%
Commission errors = 20.81%
Omission errors = 13.10%
Acres identified = 182,532
Estimator = 0.911
Modified acreage = 166,332.3
Digital codes: 13, 35-43

Class name: Barren land
Accuracy = 59.74%
Commission errors = 40.26%
Omission errors = 36.11%
Acres identified = 31,021
Estimator = 0.935
Modified acreage = 29,006.6
Digital codes 1, 2, 20-22, 26

Level II
(Overall accuracy of Level II data = 69.4%)

Class name: Residential
Per-point accuracy
Accuracy = 71.48%
Commission errors = 28.52%
Omission errors = 20.29%
Acres identified = 40,155
Estimator = 0.897
Modified acreage = 36,007.8
Digital codes 28-33
Per block accuracy: 82.95%

Class name: Commercial
Accuracy = 49.56%
Commission errors = 50.44%
Omission errors = 32.53%
Acres identified = 6,194
Estimator = 0.734
Modified acreage = 4,549.6
Digital codes 23, 24, 27

Class name: Other urban
Accuracy: Not evaluated
Acres identified = 800
Digital codes: 18,19

Class name: Crops
Accuracy = 81.08%
Commission errors = 18.92%
Omission errors = 36.17%
Acres identified = 3,060
Estimator = 1,270
Modified acreage = 3,887.0
Digital code: 17

Class name: Orchards and vineyards
Accuracy = 70.33%
Commission errors = 29.67%
Omission errors = 25.58%
Acres identified = 17,541
Estimator = 0.945
Modified acreage = 16,577.2
Digital codes: 6-12, 34

Class name: Grasses
Accuracy = 36.11%
Commission errors = 63.89%
Omission errors = 63.07
Acres identified = 97,251
Estimator = 0.978
Modified acreage = 95,089.9
Digital codes: 15,16

Class name: Brush
Accuracy = 74.71%
Commission errors = 25.29%
Omission errors = 38.92%
Acres identified = 366,795
Estimator = 1.223
Modified acreage = 448,693.9
Digital codes: 3-5, 44-49

Class name: Water
Accuracy = Not evaluated
Acres identified = 3.442
Digital code: 25

Class name: Deciduous
Accuracy = 35.34%
Commission errors = 64.66%
Omission errors = 64.39%
Acres identified = 10,318
Estimator = 0.992
Modified acreage = 10,240.4
Digital codes: 13, 14, 41

Class name: Conifers
Accuracy = 81.19%
Commission errors = 18.81%
Omission errors = 10.18%
Acres identified = 172,214
Estimator = 0.904
Modified acreage = 155,661.7
Digital codes: 35-40, 42

Class name: Mixed/Barren
Accuracy = 56.47%
Commission errors = 43.53%
Omission errors = 38.25%
Acres identified = 30,864
Estimator = 0.915
Modified acreage = 28,228.2
Digital codes: 1, 2, 21, 22

Class name: Extractive
Accuracy = Not evaluate
Acres identified = 157
Digital code: 20

Level III
(Overall accuracy of Level III data = 51.5%)

Class name: Cleared
Accuracy = 16.39%
Commission errors = 83.61%
Omission errors = 77.27%
Acres identified = 10,489
Estimator = 0.721
Modified acreage = 7,565.8
Digital code: 1

Class name: Bare
Accuracy = 47.83%
Commission errors = 52.17%
Omission errors = 48.44%
Acres identified = 19,623
Estimator = 0.927
Modified acreage = 18,201.0
Digital code: 2

Class name: Sparse brush
Accuracy = 28.27%
Commission errors = 71.22%
Omission errors = 75.60%
Acres identified = 25,939
Estimator = 1.159
Modified acreage = 30,053.5
Digital code: 3

Class name: Medium density
Accuracy = 10.29%
Commission errors = 89.70%
Omission errors = 90.0%
Acres identified = 17,673
Estimator = 1.03
Modified acreage = 18,192.8
Digital code: 4

Class name: Thick brush
Accuracy: Not evaluated
Acres identified = 9,792
Digital code: 5

Class name: Young orchard
Accuracy: Not evaluated
Acres identified = 1,012
Digital code: 6

Class name: Moderate vigor orchard
Accuracy: Not evaluated
Acres identified = 6,203
Digital code: 7

Class name: Mature orchard
Accuracy: Not evaluated
Acres identified = 4,580
Digital code: 8

Class name: Declining orchard
Accuracy: Not evaluated
Acres identified = 1,216
Digital code: 9

Class name: Moderate-vigor vineyard
Accuracy: Not evaluated
Acres identified = 1,790
Digital code: 10

Class name: High vigor vineyard
Accuracy: Not evaluated
Digital code: 11

Class name: Declining vineyard
Accuracy: Not evaluated
Digital code: 12

Class name: Woodland
Accuracy: Not evaluated
Digital code: 13

Class name: Lush grass
Accuracy = 20.78%
Commission errors = 79.22%
Omission errors = 56.76%
Acres identified = 32,524
Estimator = 0.480
Modified acreage = 15,628.4
Digital code: 15

Class name: Dry grass
Accuracy = 38.35%
Commission errors = 61.65%
Omission errors = 71.58%
Acres identified = 64,727
Estimator = 1.350
Modified acreage = 87,350.0
Digital code: 16

Class name: Agriculture
Accuracy = 81.08%
Commission errors = 18.92%
Omission errors = 36.17%
Acres identified = 3,060
Estimator = 1.270
Modified acreage = 3,887.0
Digital code: 17

Class name: Asphalt
Accuracy: Not evaluated
Digital code: 18

Class name: Concrete
Accuracy: Not evaluated
Digital code: 19

Class name: Extractive
Accuracy: Not evaluated
Acres identified = 157
Digital code: 20

Class name: Cinder
Accuracy = 86.63%
Commission errors = 13.37%
Omission errors = 17.22%
Acres identified = 572
Estimator = 1.046
Modified acreage = 598.6
Digital code: 21

Class name: Slag
Accuracy: Not evaluated
Acres identified = 180
Digital code: 22

Class name: Structures
Accuracy = 54.01%
Commission errors = 45.99%
Omission errors = 41.28%
Acres identified = 3,795
Estimator = 0.920
Modified acreage = 3,490.6
Digital code: 23

Class name: Structures strip
Accuracy = 15.18%
Commission errors = 84.82%
Omission errors = 54.05%
Acres identified = 1,477
Estimator = 0.330
Modified acreage = 487.9
Digital code: 24

Class name: Water
Accuracy: Not evaluated
Acres identified = 3,442
Digital code: 25

Class name: Snow
Accuracy: Not evaluated
Digital code: 26

Class name: Structures with brush
Accuracy: Not evaluated
Acres identified = 877
Digital code: 27

Class name: Residential with trees
Accuracy = 49.06%
Commission errors = 50.94%
Omission errors = 47.30%
Acres identified = 11,803
Estimator = 0.931
Modified acreage = 10,986.4
Digital code: 28

Class name: Irrigated newer residential
Accuracy = 34.72%
Commission errors = 65.28%
Omission errors = 79.51%
Acres identified = 5,754
Estimator = 1.694
Modified acreage = 9,749.8
Digital code: 29

Class name: Cluster
Accuracy = 28.12%
Commission errors = 71.88%
Omission errors = 43.75%
Acres identified = 3,021
Estimator = 0.500
Modified acreage 1,510.5
Digital code: 30

Class name: Large low unirrigated
Accuracy = 34.11%
Commission errors = 65.89%
Omission errors = 58.09%
Acres identified = 14,901
Estimator = 0.814
Modified acreage = 12,128.7
Digital code: 31

Class name: Rural/strip
Accuracy = 15.62%
Commission errors = 84.38%
Omission errors = 75.61%
Acres identified = 3,682
Estimator = 0.641
Modified acreage = 2,358.8
Digital code: 32

Class name: Mobil home/high density
Accuracy: Not evaluated
Acres identified = 994
Digital code: 33

Class name: Low-vigor vineyard
Accuracy: Not evaluated
Acres identified = 1,091
Digital code: 34

Class name: Big cone Douglas fir
Accuracy: Not evaluated
Acres identified = 9,138
Digital code: 35

Class name: White fir
Accuracy = 32.97%
Commission errors = 67.02%
Omission errors = 35.21%
Acres identified = 12,620
Estimator = 0.509
Modified acreage = 6,423.1
Digital code: 36

Class name: Jeffrey Pine–closed understory
Accuracy = 31.91%
Commission errors = 68.08%
Omission errors = 61.37%
Acres identified = 33,397
Estimator = 0.826
Modified acreage = 27,594.0
Digital code: 37

Class name: Jeffrey Pine - open understory
Accuracy = 29.41%
Commission errors = 70.59%
Omission errors = 69.59%
Acres identified = 35.008
Estimator = 0.967
Modified acreage = 33.863.9
Digital code: 38

Class name: Lodgepole/limber pine
Accuracy = 85.71%
Commission errors = 14.28%
Omission errors = 49.65%
Acres identified = 9,737
Estimator = 1.702
Modified acreage = 16,576.1
Digital code: 39

Class name: Pinyon/juniper
Accuracy = 81.00%
Commission errors = 19.00%
Omission errors = 20.59%
Acres identified = 34,809
Estimator = 1.02
Modified acreage = 35,505.2
Digital code: 40

Class name: Riparian hardwood
Accuracy = 42.04%
Commission errors = 57.95%
Omission errors = 39.34%
Acres identified = 8,656
Estimator = 0.693
Modified acreage = 6,000.2
Digital code: 41

Class name: Jeffrey pine/ceanothus
Accuracy = 40.00%
Commission errors = 60.00%
Omission errors = 68.91%
Acres identified = 18,553
Estimator = 1.287
Modified acreage = 23,871.5
Digital code: 42

Class name: Coultier Pine mixed forest
Accuracy = 42.62%
Commission errors = 57.38%
Omission errors = 29.44%
Acres identified = 18,952
Estimator = 0.604
Modified acreage = 11,447.5
Digital code: 43

Class name: Bracken/ceanothus
Accuracy: Not evaluated
Acres identified = 1,260
Digital code: 44

Class name: Ceanothus/scrub oak
Accuracy = 72.57%
Commission errors = 27.43%
Omission errors = 57.74%
Acres identified = 56,066
Estimator = 1.717
Modified acreage = 96,282.1
Digital code: 45

Class name: Chamise
Accuracy = 51.26%
Commission errors = 48.73%
Omission errors = 48.08%
Acres identified = 87,827
Estimator = 0.987
Modified acreage = 86,715.3
Digital code: 46
Per block accuracy: 58.86

Class name: Chamise/ceanothus
Accuracy = 62.39%
Commission errors = 37.61%
Omission errors = 61.17%
Acres identified = 57,214
Estimator = 1.607
Modified acreage = 91,933.6
Digital code: 47
Per block accuracy: 77.78

Class name: Coastal sage
Accuracy: Not evaluated
Acres identified = 14,736
Digital code: 48

Class name: Sagebrush
Accuracy = 84.03%
Commission errors = 15.97%
Omission errors = 0.99%
Acres identified = 96,288
Estimator = 0.849
Modified acreage = 81,723.4
Digital code: 49

APPENDIX C

1976 – 1979 LANDSAT LAND COVER CHANGE ACREAGES AND ACCURACIES

1976 – 1979 CHANGE DETECTION ACCURACIES

1. 1976-1979 per-point correct assignment accuracy of Level I from-to land-cover combinations for all those pixels identified as change in the San Bernardino Valley floor: 12%.

2. 1976-1979 per-point accuracy of Level I from-to land-cover combinations for the full study area (both change and nonchange areas included): 72%.

3. 1979 land-cover accuracy for full study area: 74%.

4. 1976-1979 per-point accuracy for detecting Level I land-cover change (irrespective of whether or not the type of change identified was correct): 73%.

5. Inventory of acres of Level I changes over full study area: 5,400 acres.

6. Acreage inventory detected for valley changes (irrespective of type of specific change) was determined to 100% correlation to that obtained by ground-truth.

7. Accuracies for Level II and III 1976-1979 information were not examined.

8. Conclusions. The 1979 land-cover accuracy may be suitable for single-date 1979 applications work. The number and locations of Level I 1976-1979 changes detected may also be suitably accurate for use in applications. The detection of specific pixels with correct 1976-1979 from-to land-cover combinations is too inaccurate to be useful. However the inventory of total from-to change combinations (ignoring whether or not these changes were detected in the correct locations) is more accurate and may be useful information.

LAND-COVER CHANGES BETWEEN 1976 AND 1979

Information Level I

Acres	Cover
+3410	Urban
-1885	Agriculture
-2203	Range
+1990	Water[1]
-1310	Forest
0	Barren

Information Level II

Acres	Cover
+3189	Residential
+ 221	Commercial
- 531	Crops
- 1354	Orchards and vineyards
-3218	Grass
+1015	Brush
+1990	Water
- 132	Deciduous forest
-1178	Coniferous forest

[1]Mostly due to increased rainfall in 1979 in high water levels in Lakes and reservoirs; 1976 was a drought year.

Information Level III

Acres	Cover
+ 338	Residential: newer irrigated
+2064	Residential: sparse cluster
+ 723	Residential: large lot unirrigated
+ 64	Mobile homes/high-density residential
+ 184	Structures
+ 37	Structures: with brush
- 531	Crops
- 60	Orchards: young
- 491	Orchards: moderate vigor
- 426	Orchards: mature
- 81	Orchards: declining
- 21	Vineyards: poor vigor
- 36	Vineyards: moderate vigor
- 209	Vineyards: high vigor
- 30	Vineyards: declining
- 707	Lush grass
-2511	Dry grass
+ 120	Sparse brush
+1207	Moderate-density brush
+1097	Thick brush
- 137	Chamise
+ 150	Chamise/ceanothus mix
+ 52	Ceanothus/scrub oak
- 41	Bracken/ceanothus
- 169	Coastal sage
-1264	Sagebrush
+1990	Water
- 85	Woodland
- 47	Riparian mixed hardwoods
- 47	Big cone Douglas fir
- 38	White fir
- 189	Jeffrey pine: closed understory
- 191	Jeffrey pine: open understory
- 172	Coulter pine
- 528	Lodgepole/limber pine
- 111	Pinyon/juniper
- 197	Cleared
- 53	Bare
+ 436	Snow

REFERENCES

1. Wilson, P. M.; and Gialdini, M. J.: Operational Alternatives for Landsat in California, Annual Report. Geogroup Corporation, Berkeley, Calif., NASA CR-166353, 1981.

2. Likens, W.; and Maw, K.: Hierarchical Modeling for Image Classification. Proceedings of the Pecora VII Symposium, Sioux Falls, S.D., 1981.

3. Vertical Data Integration–San Bernardino Study, Phases I-A, I-B. Environmental System Research Institute, Redlands, Calif., NASA CR-166355, 1981.

4. Inventory of Available Data Elements for the San Bernardino, California Region. Environmental Systems Research Institute, Redlands, Calif., NASA CR-166354, 1981.

5. Peterson, D. L.; Tosta-Miller, N.; Norman, S. D.; Wierman, D.; and Newland, W. L.: Land Cover Classification of California Using Mosaicking and High-Speed Processing, Proceedings of the 14th International Symposium on Remote Sensing of Environment, Vol. 2, San Jose, Costa Rica, 1980.

6. Tosta-Miller, N.: Ecological Stratification for Vegetation Classification Using Remotely Sensed Data. Paper presented at In-Place Resource Inventories. Principles and Practices–A National Workshop, Orono, Maine, Aug., 1981.

7. Michel, R.; and Maw, K.: Explorations into Technical Procedures for Vertical Integration. Technical working paper, Environmental Systems Research Institute, Redlands, Calif., NASA CR-166352, 1979.

8. IDIM Users Guide. TM-705, ESL, Inc., Sunnyvale, Calif., 1976.

9. Anderson, J. R.; Handy, E. E.; Roach, J. T.; and Witmer, R. E.: A Land Use and Land Cover Classification System for Use With Remote Sensor Data. U.S. Geological Survey Professional Paper 964, 1976.

10. Stow, D. A.; Tinney, L. R.; and Estes, J. E.: Deriving Land Use/Land Cover Change Statistics from Landsat: A Study of Prime Agricultural Land. Proceedings of the 14th International Symposium on Remote Sensing of Environment, Vol. 2, San Jose, Costa Rica, 1980.

11. Angelici, G. L.; Bryant, N. A.; and Friedman, S. Z.: Techniques for Land Use Change Detection Using Landsat Imagery. Proceedings of Fall ASP-ASCM Convention, Little Rock, Ark., Oct., 1977.

12. Todd, W. J.: Urban and Regional Land Use Change Detected by Using Landsat Data. Journal of Research, U.S. Geological Survey, Vol. 5, No. 5, 1977, pp. 1533-1539.

13. Miller, L. D.; Nualchawee, K.; and Tom, C.: Analysis of the Dynamics of Shifting Cultivation in the Tropical Forests of Northern Thailand Using Landsat Modeling and Classification of Landsat Imagery. Proceedings of the 12th International Symposium on Remote Sensing of Environment, Vol. 2, Ann Arbor, Mich., 1978.

1. Report No. NASA TM-84244	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle FINAL REPORT: LANDSAT LAND COVER ANALYSIS COMPLETED FOR CIRSS/SAN BERNARDINO COUNTY PROJECT		5. Report Date OCTOBER 1982
		6. Performing Organization Code
7. Author(s) William Likens, Keith Maw, and David Sinnott		8. Performing Organization Report No. A-8920
9. Performing Organization Name and Address NASA Ames Research Center Moffett Field, CA 94035		10. Work Unit No. T 5882
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		13. Type of Report and Period Covered Technical Memorandum
		14. Sponsoring Agency Code 677-16-15

15. Supplementary Notes

Point of Contact: William Likens, Ames Research Center, M/S 242-4, Moffett Field, CA 94035
(415)965-5596 or FTS 448-5596

16. Abstract

The Landsat analysis carried out as part of Ames Research Center's San Bernardino County Project, one of four projects sponsored by NASA as part of the California Integrated Remote Sensing System (CIRSS) effort for generating and utilizing digital geographic data bases, is described. Topics explored in this Landsat analysis include use of data-base modeling with spectral cluster data to improve Landsat data classification, and quantitative evaluation of several change techniques. Both 1976 and 1979 Landsat data were used in the project. The Landsat analyses took place between April 1980 and September 1981.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Landsat Geographic information systems GIS Modeling Data Bases	Unclassified - Unlimited Subject Category – 43

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	28	A02

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1982

National Aeronautics and
Space Administration

Washington, D.C.
20546

Official Business
Penalty for Private Use, $300

THIRD-CLASS BULK RATE

Postage and Fees Paid
National Aeronautics and
Space Administration
NASA-451

U.S.MAIL

NASA

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return