

NASA Technical Memorandum 85995

NASA-TM-85995 19840022441

---

# The Effect of Spatial, Spectral, and Radiometric Factors on Classification Accuracy Using Thematic Mapper Data

---

Robert C. Wrigley, William Acevedo,  
Douglass A. Alexander, Joan S. Buis and Don H. Card

---

July 1984

**LIBRARY COPY**

JUL 21 1984

LANGLEY RESEARCH CENTER  
LIBRARY, NASA  
HAMPTON, VIRGINIA

**NASA**

National Aeronautics and  
Space Administration



NF00847

---

# The Effect of Spatial, Spectral, and Radiometric Factors on Classification Accuracy Using Thematic Mapper Data

---

Robert Wrigley

Don H. Card, Ames Research Center, Moffett Field, California

William Acevedo

Douglass A. Alexander

Joan S. Buis, Technicolor Government Services, Inc, Ames Research Center,  
Moffett Field, California



National Aeronautics and  
Space Administration

**Ames Research Center**  
Moffett Field California 94035

N84-30510 #

THE EFFECT OF SPATIAL, SPECTRAL AND RADIOMETRIC FACTORS  
ON CLASSIFICATION ACCURACY USING THEMATIC MAPPER DATA

R. Wrigley\*, W. Acevedo\*\*, D. Alexander\*\*, J. Buis\*\*, and D. Card\*

\*NASA/Ames Research Center, Moffett Field, CA 94035

\*\*Technicolor Government Services, Inc.  
Ames Research Center, Moffett Field, CA 94035

ABSTRACT

An experiment of a factorial design was conducted to test the effects on classification accuracy of land cover types due to the improved spatial, spectral and radiometric characteristics of the Thematic Mapper (TM) in comparison to the Multispectral Scanner (MSS). High altitude aircraft scanner data from the Airborne Thematic Mapper instrument was acquired over central California in August, 1983 and used to simulate Thematic Mapper data as well as all combinations of the three characteristics for eight data sets in all. Results for the training sites (field-center pixels) showed better classification accuracies for MSS spatial resolution, TM spectral bands and TM radiometry in order of importance

Keywords Thematic Mapper, factorial experiment, simulation, classification.

1. INTRODUCTION

The Multispectral Scanner (MSS) aboard the Landsat series of satellites has proven to be a very useful tool with which to categorize the land surface into various cover types using multispectral classification techniques. As a first generation instrument, it suffered from a number of disadvantages. Firstly, the eighty meter spatial resolution created serious limitations; cultural features often were blurred beyond recognition. Secondly, the six-bit radiometric resolution employed did not take full advantage of the sensitivity of the detectors. Modern detectors are engineered for even more sensitivity. Finally, the spectral characteristics of the MSS were not optimized for vegetative discrimination. Some bands were virtually redundant and one even spanned the vegetative signature reversal in the 700-750 nm region. Recent detector technology advances yielded an expanded spectral region to cover the middle infrared (1450-2350 nm), where leaf moisture content and clay mixtures play a significant role, as well as the thermal infrared.

The aforementioned limitations, in addition to potential enhancements, spurred development of the Thematic Mapper (TM)--a second generation land remote sensing satellite instrument designed to categorize land cover types. The TM incorporates thirty meter resolution, eight-bit radiometry with very stable detectors, and several spectral bands featuring better placement in the spectrum and including the middle and thermal infrared regions. Given these improvements over the MSS, a fair set of questions is: How well does TM categorize land cover types in comparison to MSS?, and How do the various improvements contribute to any differences in capability? This paper attempts to address these questions in a definitive way using high altitude aircraft scanner data acquired on August 12, 1983 in conjunction with ground reference data acquired at the same time. The aircraft scanner data has been used to simulate both TM and MSS spatial, spectral and radiometric characteristics in a factorial experiment designed to test the effects of each factor on classification accuracy, separately and in combination. TM and MSS data were obtained for the same day; in the future, they will be used to confirm results for the end members of the factorial experiment.

2. BACKGROUND

A number of studies have tried to characterize the effects of changing the spatial, spectral and radiometric parameters of satellite remote sensing instruments on the classification of land cover types. Several efforts concentrated on one parameter at a time and varied it over broad ranges. Among others, Morgenstern *et al.* (1) varied spatial resolution over 30 to 90 meters, simulating the proposed modulation transfer function (MTF) of the TM, and found that classification accuracy decreased as spatial resolution decreased except for field-center pixels which were unaffected. Such studies usually begin with 5-10 meter pixels in order to simulate the TM spatial characteristics. Very few pixels remain in the scan line at the coarsest resolution making it difficult to work effectively. As often as not, simple boxcar filters are used to degrade the spatial resolution (Ref. 2,3,4,5)

Morgenstern et al. (1) also degraded the radiometry by adding random noise in progressively larger amounts. Their results for field center pixels showed an immediate decrease in classification accuracy if the noise was increased above the nominal specification for TM. Landgrebe et al. (6) took a similar approach and obtained comparable results. Other efforts performed radiometric degradation by dropping the two least significant bits in eight bit data (Ref. 7,8) or proportionately reducing the dynamic range of eight bit data to an equivalent dynamic range in six bit data (Ref. 3,4). The noise addition approach is inappropriate for simulation of MSS data because the noise measured in MSS data is less than one count except for six-line striping which is an artifact often removed by histogram equalization techniques. In hindsight, both of the other approaches degrade the data excessively. In terms of digital counts, the difference in the dynamic range of comparable TM and MSS bands is of the order of a factor of 2-3. Consequently, a degradation scheme which focuses on 8 bits versus 6 bits (a factor of four) rather than the actual ranges will over-compensate.

Spectral simulation and degradation has often been a problem because aircraft scanners operate in fixed bands which may not correspond to either TM or MSS bands. Morgenstern et al. (1) were fortunate to be able to select bands from a 24 channel scanner. Even so, the bands were still not precise and some were not useable for extraneous reasons (noise, etc.). More recently, Thematic Mapper simulators have become available which accurately simulate the TM bands. However, they make it more difficult to simulate MSS. The NS-001 simulator omits the intermediate spectral regions between the TM bands entirely; the Daedalus Airborne Thematic Mapper (ATM) used in the present work includes them and makes a simulation of the MSS spectral regions feasible, if not exact.

The parametric studies (Ref. 1,6) do not investigate combined effects or specific differences between TM and MSS. A few studies (Ref. 3,4,7,8) examine only the end members of the degradation sequence: TM or MSS characteristics. Such studies do provide direct comparisons between the characteristics of TM and MSS. In addition, they used a factorial design so that the effects of a single characteristic could be isolated while still being able to study the combined effects. Sigman and Craig(7) used NS-001 data and found that the spectral and spatial characteristics of TM yielded some improvements in area mensuration of corn and soybeans. Williams et al. (8) used actual TM data acquired in November, 1982 and found that higher classification accuracies resulted with TM radiometric and spectral characteristics, but that the spatial resolution factor was not important. On the other hand, their classification accuracies were very low, perhaps due to the time of the year (most of the vegetation was dead or defoliated), and their results may not be representative. Previous studies by the present authors (Ref. 3,4) utilized data from a precursor to the ATM which could not simulate TM bands 5 or 6. Considering only field-center pixels from the training areas, it was found that lower spatial resolution (MSS) and higher radiometric resolution (TM) gave higher classification accuracies while the spectral factor showed little difference in classification accuracy. Unfortunately, lack of

sufficient ground reference data limited these studies to the training areas so that direct comparison with the other factorial experiments was not appropriate. Still, the conflicting nature of the results cited suggests that a definitive experiment should be conducted. Such an experiment should use a well vegetated data set and the most appropriate simulation methods in a factorial design. Since some of the differences in results may be attributable to the effects of boundary pixels, field-center pixels should be examined directly.

### 3. APPROACH

The present work seeks to improve on the earlier studies in a number of ways. First, a great deal more ground reference data was acquired so that analysis would not be limited to training sites and so that random sampling could be used for classification accuracy analysis. Second, a simultaneous data set of TM, MSS and ATM data was acquired at a time when vegetative cover was near maximum. The TM and MSS data will eventually serve as checks of the end members in the degradation sequence. Third, the spectral simulation of TM was more accurate with the use of ATM data although the simulation of MSS spectral characteristics suffered slightly. Fourth, the radiometric simulations used the actual radiometric gains of TM and MSS as opposed to a scaling factor related to the dynamic range of the ATM data. Fifth, the spatial degradation used the measured modulation transfer function of MSS to simulate MSS data. Finally, the analysis sequence was expanded to include not only field-center pixels for comparison with the earlier results (Ref. 3), but randomly selected pixels for improved statistical analysis within the factorial design.

On August 12, 1983, a test of the Tracking and Data Relay Satellite System for Landsat-4 data transmission occurred over the San Joaquin Valley and yielded the first TM data of that area during the growing season when agricultural crops were close to full canopy development. MSS data were acquired at the same time. Airborne Thematic Mapper data were acquired by the high altitude ER-2 aircraft with an instantaneous field of view of 25 meters within three hours of the Landsat-4 overpass. Ground reference data acquisition for time-dependent cover types, such as agricultural crops, began on August 12th and continued for two weeks. Sufficient ground reference data was collected in order to permit an assessment of area mensuration accuracy by random sampling from the data set for statistical analysis.

The factorial design of our experiment required analyzing a data set for each of the possible combinations of the factors. We had three factors (spatial, spectral and radiometric) at two levels (TM and MSS), requiring eight data sets to cover all combinations. Systematic degradations of the ATM data generated these eight data sets according to the scheme shown in Figure 1. The spectral and radiometric degradations began with the ATM data but the spatial degradation operated on those derived products. Each of these degradations will be described below.

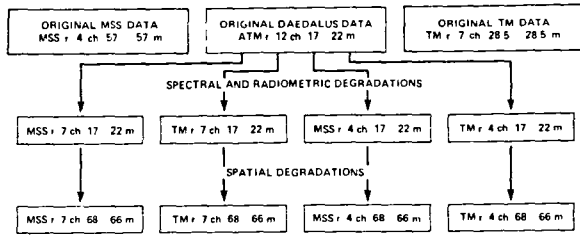


Figure 1. The sequence for systematically degrading the ATM data by varying the spatial, spectral, and radiometric characteristics to achieve approximate TM and MSS values. The "r" stands for radiometry, "ch" stands for channels and "m" stands for meters representing pixel size.

### 3.1 Study Areas

The Stockton area was selected as the prime study site for this experiment. A portion of the city of Stockton was selected for the urban segment. Agricultural classes from the Stockton area were supplemented with the addition of a rice-dominated agricultural area near the town of Colusa. Rangeland classes were found in the foothills east of Stockton. An area around Bucks Lake in the northern Sierras was selected for its forestry classes. Segments from six ATM flightlines covered the areas of interest.

The variation of the solar azimuth angle with respect to flightline directions created concern regarding scan angle dependent variation in the data. Two flightlines were identified as potential problems. Statistical models of the scan angle effects were calculated for both these flightlines along with a third selected as an example of data not expected to need correction. Only the Bucks Lake forestry segment required correction, it was flown with the flightline direction nearly perpendicular to the solar azimuth. The Bucks Lake segment was corrected using a second order polynomial model of column averages of the flightline. All six flightline segments were digitally mosaicked to minimize the number of data sets to be processed.

### 3.2 Degradations

Figure 2 shows the spectral bands of the ATM data and indicates the manner in which they were used to simulate TM and MSS spectral characteristics. ATM bands 2, 3, 5, 7, 9, 11 and 10 were direct analogs of TM bands 1-7, respectively, and were used without spectral modification. ATM band 3 (520-600 nm) was used to simulate MSS band 1 (500-600 nm), lack of coverage of the 500-520 nm region not being considered significant. ATM band 4 (600-630 nm) and ATM band 5 (630-690 nm) were combined to simulate MSS band 2 (600-690 nm). ATM band 6 (690-760 nm) was used to simulate MSS band 3 (700-800 nm). The extra coverage of the 690-700 nm region and the lack of coverage of the 760-800 nm region may tend to enhance the effects MSS band 3 creates in straddling the vegetative signature reversal in the 690-740 nm region. ATM bands 7 (760-900 nm) and 8 (900-1050 nm) were combined to simulate MSS band 4 (800-1100 nm). The differences in spectral coverage were not considered to be significant for this band. In cases where ATM

bands were combined, they were converted to radiance values using the ATM radiometric gains (Ref 9) and then weighted according to their proportional coverage of the simulated band.

In the earlier work (Ref. 3,4), the radiometric simulation of TM data used the aircraft scanner data as acquired. Typically, the dynamic range was much greater than is the case with actual TM data. The MSS characteristics were simulated by scaling the aircraft scanner data by the fourth root of its dynamic range. As discussed above, this procedure tended to exaggerate the simulation. In the present study, the actual radiometric gains of TM and MSS were used to avoid that problem and provide an accurate simulation of the digital count levels. The ATM digital counts for a given band were multiplied by the ATM radiometric gains (Ref. 9) to produce radiance values. These radiance values were divided by the appropriate radiometric gain for TM (Ref. 10) or MSS (Ref. 11) to produce a digital count. Table 1 lists the radiometric gains for TM, MSS and ATM. Note that the MSS gains are given in spectral radiance units for comparison with the TM and ATM gains. Normally, MSS gains are quoted for "radiance in the band". In some cases, ad hoc decisions were necessary to complete the data sets for the factorial experiment. For instance, the simulation of MSS radiometry of TM bands 5-7 had no basis in experience so the gains were set at exactly four times the TM gains. Also, the MSS gains were set as if they corresponded to 64 levels since resampling to 128 levels is done only to linearize the data. (This does not refer to spatial resampling.)

Spatial degradation by Alexander *et al.* was completed by averaging nine (3x3 pixel window) 25 meter (TM) pixels to simulate one 75 meter (MSS) pixel. The current analysis employed a point-spread function (PSF) that closely represented the MSS system. Several investigators developed a PSF for the MSS based on the nominal characteristics of the instrument, such as the instantaneous field of view, sampling rate, electronic filtering and the optical-blur circle (Ref. 12,13). Schowengerdt *et al.* (14) measured the complete optical transfer function, of which the modulation transfer function (MTF) is the real part, from operational MSS data. The Fourier transform of the MTF is the PSF. They measured the optical transfer function by scanning film products with a micro-densitometer, optimizing the match between scans of underflight imagery and

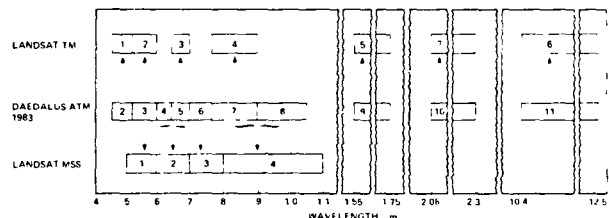


Figure 2. Spectral regions of the Airborne Thematic Mapper bands in comparison to Thematic Mapper and Multispectral Scanner bands. The arrows indicate how the ATM bands were used to simulate TM and MSS bands.

AIRBORNE THEMATIC MAPPER (ATM)	
ch 2	0.076
ch 3	0.072
ch 4	0.061
ch 5	0.076
ch 6	0.065
ch 7	0.079
ch 8	0.041
ch 9	0.012
ch 10	0.008
THEMATIC MAPPER (TM)	
ch 1	0.063
ch 2	0.124
ch 3	0.095
ch 4	0.092
ch 5	0.013
ch 7	0.007
MULTISPECTRAL SCANNER (MSS)	
ch 1	0.332
ch 2	0.294
ch 3	0.182
ch 4	0.281

Table 1. Radiometric gain factors used for the ATM, Landsat-4 TM and MSS data. The units are in terms of radiance/count.

either the 70mm film chips or film products derived specifically for their analysis from MSS digital data tapes, and forming the ratio of the two scans. The authors scaled the published MTF plots for the digital data products of Schowengerdt *et al.* (14) to obtain numerical results, and then fit a Gaussian curve to them. The Fourier transform of the Gaussian representation of the MTF was used as the PSF of the MSS in meters along the ground. The full-width-at-half-maximum of the PSF in the along-scan direction was 80.8 meters and in the along-track direction it was 72.2 meters.

In order to use the PSF in degrading the ATM data to approximate the spatial characteristics of MSS data, it was assumed that the MTF of the ATM was sufficiently high at the spatial frequencies important for MSS data that it could be ignored. A 7x7 convolution kernel was constructed to represent the PSF at the pixel spacings of the ATM data. The spacings were measured directly in the imagery because they depended on actual aircraft altitude above the terrain (along-scan) and ground speed (along-track), including effects due to headwinds or tailwinds. The spacings used were 17 meters along-track and 22 meters along-scan, even though one area in the data deviated from these values by 20% due to headwinds. The 7x7 convolution kernel (shown in Figure 3) was normalized so that the resultant image maintained the same digital data range as the original image. TM channel 6 required a different convolution kernel since it has a different spatial resolution. A similar methodology was used to calculate this kernel from pre-launch measurements of the line spread function.

0.0025	0.0068	0.0144	0.0151	0.0124	0.0068	0.0025
0.0054	0.0146	0.0267	0.0326	0.0267	0.0146	0.0054
0.0085	0.0237	0.0472	0.0515	0.0422	0.0237	0.0085
0.0099	0.0270	0.0492	0.0601	0.0492	0.0270	0.0099
0.0085	0.0237	0.0422	0.0515	0.0472	0.0237	0.0085
0.0054	0.0146	0.0267	0.0326	0.0267	0.0146	0.0054
0.0025	0.0068	0.0124	0.0151	0.0124	0.0068	0.0025

Figure 3. Convolution kernel representing the Point Spread Function of the MSS data for a pixel spacing of 17 meters along-track and 22 meters along-scan.

Convolution was performed at every pixel location in the original image so that the output image size was the same size as the input. The output image was resampled at integer pixel locations to approximate the MSS pixel spacing of 57m x 57m. Sampling every fourth line and every third pixel yielded a spacing of 68m x 66m for the 17m x 22m pixels of the ATM data. Thus, the ATM data was convolved with a PSF derived from actual MSS data and sampled to approximate the pixel spacing of the MSS.

This series of degradations provided a very close approximation to TM and MSS data. The other six data sets will provide information on how each factor contributes independently, and in combination, to classification accuracy.

### 3.3 Training Site Selection

Training sites were selected within polygonal areas of homogeneous urban, agricultural, forested, rangeland and water information classes using an interactive system to display the ATM data at full spatial resolution. Sites were selected to include the spectral variation in each cover type and to exclude boundaries between cover types. The training areas covered at least 5 percent of the total area occupied by that class

For the urban categories, photo-interpretation of high altitude color infrared (CIR) photography of Stockton collected in July, 1982 at a scale of 1:33,000 provided sufficient ground reference data. Nine classes were delineated.

Training sites for fifteen agricultural categories were delineated from crop type survey maps developed from the field work performed in San Joaquin County on August 12th and shortly thereafter. Field boundaries for a dozen different crops were delineated on the ATM data. Training sites for an additional crop, rice, were photo-interpreted from high altitude CIR photography flown over Colusa County in July, 1983 at a scale of 1:33,000.

The breakdown of forest classes differed from a standard classification scheme. Class definitions reflected structural (crown closure and size class) and taxonomic (plant community) information. Training sites were ground referenced from high altitude CIR photography flown over the Bucks Lake area of the Plumas National Forest in June, 1980. Forest stand maps, provided by the U.S. Forest Service, contained structural and taxonomic information for all areas identified as commercial forest, approximately 90% of the area. Eight information classes represented pole timber, saw timber, <40% crown closure, >40% crown closure, pines and firs, in all possible combinations.

Four rangeland categories represented grasslands photo-interpreted from the 1982 high altitude CIR photography of an area east of Stockton. The single water category was photo-interpreted from the Stockton and Bucks Lake CIR photography.

Training site polygons, delineated on the original high spatial resolution (17x22 meter) data set, were applied to the other three 17x22 meter data

sets directly. This ensured that exactly the same set of pixels would be used for training for all four data sets. The high resolution polygon vector file was converted to a strata mask and sampled to overlay the 68x66 meter data sets. If any high resolution pixels from outside the training site contributed to a low resolution pixel during the sampling process, the low resolution pixel was excluded from the low resolution training site. Thus, the low resolution training sites also contained only field-center pixels.

### 3.4 Clustering, Classification and Accuracy Assessment

Markham and Townshend(15) indicated that higher spatial resolution data contained greater spectral heterogeneity. Alexander *et al.* (3) attempted to account for this effect in earlier work by setting the number of spectral clusters representing an information class according to the spectral variability observed within the class. The same approach was employed in this study. Histograms were generated for all data sets for each information class and examined to determine the number of modes in each band for each information class. The maximum number of modes for a given class determined the number of clusters requested from the clustering algorithm for that class. Therefore, the number of clusters requested for each information class varied among classes and among data sets. Cluster statistics files were generated for each class using an EDITOR/LARSYS clustering algorithm implemented on a Cray XMP. Clusters representing less than 50 pixels in a class were deleted to avoid excessive variances. All remaining clusters for a given data set were combined to form a master cluster file without further editing for that data set. Maximum likelihood classification of each data set was performed on the Cray XMP using the appropriate cluster file on a per-pixel basis.

Three methods of analysis will be used eventually but results from only the first method will be presented in this paper. The other methods will be described briefly for completeness. In the first method, classification accuracy of the training sites was computed and compared among the data sets to identify trends in a manner similar to the previous work (Ref. 3). This was done to provide a direct comparison at three levels of information, as before: Level II/III Overall, Level I, and Level I Overall. The overall accuracies were analyzed in a 2x2x2 factorial experiment by first averaging the classification accuracies over each of the three factors, one factor at a time, and then averaging them over two factors, one at a time. This scheme allows easy comparison among the unaveraged factor or factors.

Analysis of the classification accuracy trends for the training sites involves a complete enumeration of the pixels within them and, therefore, does not permit inferences about other areas not included in the tested population. However, the training site pixels represent, by definition, data that are not contaminated by the boundaries between cover types. Such pixels may be called "pure pixels" or field-center pixels. Their importance is enhanced in the present study because higher spatial resolution data has a greater proportion of pure pixels and a lower proportion of boundary pixels than does the

lower spatial resolution data. Thus, this first mode of analysis provides important results representing the pure-pixel case.

The second mode of analysis will examine the classification accuracy trends for a random sample of pixels drawn from all areas where ground reference data is available. This phase of the analysis will permit inferences to be drawn for a larger area than the training sites, and allow statistical tests to be performed to calculate the significance of the results. (Since, in the previous case, the training site pixels were completely enumerated, the results were exact and without variance, so that statistical tests had no meaning.) To support this mode of analysis, the mosaics will be gridded into 40x40 pixel blocks and a random sample of blocks will be drawn from each portion of the mosaic. This process, sometimes referred to as "cluster sampling", will permit a more efficient use of analyst effort than if single pixels were selected. The ground reference data will be transferred to the randomly selected blocks by drawing the boundaries between the cover types within each block on the raw data under cursor control using an interactive display system. When the boundary lines in each block have been converted into solid cover type polygons, contingency tables will be constructed between each pair of classified data/ground reference data blocks and the results combined for each data set. At this point, the analysis will proceed as in the case of the training site pixels. Results will be compared across the data sets at three levels of information, and the overall accuracies will be subjected to the 2x2x2 factorial experiment. In the latter case, either analysis of variance or Tukey's methods will be used to determine the statistical significance of the results, depending on the level of interaction between the factors. These results will represent the more usual situation where both pure and boundary pixels are included in analysis.

The final mode of analysis will use the previously selected random blocks to derive accuracies for field-center pixels once again. The boundary lines in each block will be expanded to be 2-4 pixels wide so that the non-boundary areas can be considered to be pure pixels. A set of classified data which includes only the pure pixels will be generated from the original classified data by logical operations using the expanded boundaries. The analysis sequence used in the first and second modes of analysis will be repeated for this field-center pixel case. Comparison of the results between the second and third modes of analysis will provide an assessment of the effects of the boundary pixels based on random sampling procedures.

## 4. RESULTS

The classification accuracies for the training sites are reported for three levels of information (Level II/III Overall, five Level I classes, and Level I Overall) in Table 2 as the percentage of correctly classified pixels. The spatial, radiometric and spectral characteristics of each of the eight data sets in the factorial experiment are listed beside the corresponding results. Six spectral bands were used to simulate the TM

DATA SET CHARACTERISTICS	LEVEL II/III	LEVEL I CLASSES				LEVEL I	LEVEL I	
	OVERALL	AGRICULTURE	FOREST	WATER	URBAN	RANGE	OVERALL*	
17x22 m, TM rad, 6 bands	71.9	98.4	100.0	81.4	90.0	98.9	93.7	96.8
17x22 m, TM rad, 4 bands	65.0	97.0	99.9	98.4	75.7	95.7	93.4	92.1
17x22 m, MSS rad, 6 bands	71.9	97.3	100.0	39.5	86.4	98.0	84.2	95.4
17x22 m, MSS rad, 4 bands	57.1	93.9	99.5	39.3	72.5	93.9	79.8	89.9
68x66 m, TM rad, 6 bands	89.9	97.4	100.0	100.0	99.0	99.4	99.2	98.9
68x66 m, TM rad, 4 bands	79.4	97.8	100.0	34.0	94.2	98.4	84.9	97.6
68x66 m, MSS rad, 6 bands	84.3	97.9	95.9	34.0	97.9	99.4	85.0	97.8
68x66 m, MSS rad, 4 bands	69.6	94.2	100.0	34.0	88.7	93.0	82.0	94.0

Table 2. Classification accuracies for three levels of information for each of the eight data set configurations representing all combinations of the three factors (spatial, radiometric and spectral) in two states (TM and MSS). The accuracies are given as the percentage of correctly classified pixels. \* The Level I class Water was not included in the overall average in this column.

CLASSIFICATION ACCURACIES AVERAGED OVER SPECTRAL FACTOR				CLASSIFICATION ACCURACIES AVERAGED OVER SPATIAL FACTOR				CLASSIFICATION ACCURACIES AVERAGED OVER RADIOMETRIC FACTOR			
		RADIOMETRIC				SPECTRAL				SPATIAL	
		TM	MSS			6 BAND	4 BAND			17X22 M	68X66 M
SPATIAL	17X22 M	68.4	64.5	RADIOMETRIC	TM	80.9	72.2	SPECTRAL	6 BANDS	71.9	87.1
	68X66 M	84.7	76.9		MSS	78.1	63.4		4 BANDS	61.0	74.5

CLASSIFICATION ACCURACIES AVERAGED OVER SPECTRAL & RADIOMETRIC FACTORS				CLASSIFICATION ACCURACIES AVERAGED OVER SPATIAL & SPECTRAL FACTORS				CLASSIFICATION ACCURACIES AVERAGED OVER RADIOMETRIC & SPATIAL FACTORS			
		RADIOMETRIC				SPECTRAL				SPATIAL	
		TM	MSS			6 BAND	4 BAND			17X22 M	68X66 M
SPATIAL	17X22 M	66.6		RADIOMETRIC	TM	76.5		SPECTRAL	6 BANDS	79.5	
	68X66 M	80.8			MSS	70.7			4 BANDS	67.8	
DIFFERENCE		-14.2		DIFFERENCE		5.8		DIFFERENCE		11.7	

Table 3. Classification accuracies from the Level II/III Overall column of Table 2 are averaged over one factor (spectral, spatial or radiometric) at a time in the upper portion of the table to display the interactions of the other two factors. The accuracies are averaged over two factors in the lower portion to display the effect of the remaining factor.

CLASSIFICATION ACCURACIES AVERAGED OVER SPECTRAL FACTOR				CLASSIFICATION ACCURACIES AVERAGED OVER SPATIAL FACTOR				CLASSIFICATION ACCURACIES AVERAGED OVER RADIOMETRIC FACTOR			
		RADIOMETRIC				SPECTRAL				SPATIAL	
		TM	MSS			6 BAND	4 BAND			17X22 M	68X66 M
SPATIAL	17X22 M	94.5	92.7	RADIOMETRIC	TM	97.9	94.9	SPECTRAL	6 BANDS	96.1	98.4
	68X66 M	98.3	95.9		MSS	96.6	92.0		4 BANDS	91.0	92.1

CLASSIFICATION ACCURACIES AVERAGED OVER SPECTRAL & RADIOMETRIC FACTORS				CLASSIFICATION ACCURACIES AVERAGED OVER SPATIAL & SPECTRAL FACTORS				CLASSIFICATION ACCURACIES AVERAGED OVER RADIOMETRIC & SPATIAL FACTORS			
		RADIOMETRIC				SPECTRAL				SPATIAL	
		TM	MSS			6 BAND	4 BAND			17X22 M	68X66 M
SPATIAL	17X22 M	93.6		RADIOMETRIC	TM	96.4		SPECTRAL	6 BANDS	97.2	
	68X66 M	97.1			MSS	94.3			4 BANDS	93.4	
DIFFERENCE		-4.5		DIFFERENCE		2.1		DIFFERENCE		3.8	

Table 4. Classification accuracies from the Level I Overall\* column of Table 2 treated in the same manner as Table 3.



spectral characteristic. The thermal band was not used at this time because of its lower spatial resolution. The Level II/III Overall and Level I Overall accuracies were calculated by averaging the percentage of correctly classified pixels within each class. This procedure treated each class equally and did not include weighting by the number of pixels in the training site.

The results for the Level I class Water in Table 2 show that five of the eight data sets had very low classification accuracies. Investigations showed that these low accuracies were due to an artifact of the clustering process. By asking for multiple clusters for each category and then deleting clusters with too few pixels, it is possible to delete too many clusters so that those remaining fail to adequately represent the spectral variation in the data since the experimental design precluded further editing of the clusters. Classification accuracy would then suffer. Although this effect was noticed in a few other isolated instances, its impact on the results in Table 2 was limited to the Water class primarily because it was not averaged with anything else. When Water was removed from the Level II/III Overall averages, the results were affected very little because it was only one of 36 classes. Furthermore, when Water was removed from the Level I Overall averages, the results (see below) were far more consistent with those derived from the Level II/III Overall averages. It was decided to delete the Water class from further consideration because of its anomalous results and the improved consistency of the results derived below. The Level I Overall averages calculated without the Water class are given in the last column in Table 2. The Water class is not included in the Level II/III Overall averages either.

An examination of the error matrices for each of the eight data sets revealed some characteristics of these classifications. For urban classes in the high spatial resolution data sets, the attempt to distinguish levels of information beyond Level II (residential, industrial and commercial) was not very successful. Generalizing the classification scheme to Level II should provide acceptable accuracies at the higher resolution. The lower spatial resolution data sets provided acceptable accuracies for urban distinctions beyond Level II if the TM spectral characteristics were included. For Level II/III forest categories, significant confusion occurred in the high spatial resolution data sets although the Level I accuracies were uniformly high. Analysis showed that accuracies improved the most when the size variable (pole timber/saw timber) was averaged out. Again, low spatial resolution coupled with the TM spectral characteristics provided good accuracies for all forest Level II/III classes.

Careful inspection of Table 2 shows that when the spatial and radiometric factors are held constant, the TM spectral characteristic (six bands) yields higher classification accuracies in 22 of 24 comparisons with the MSS spectral characteristic (four bands). Similarly, when the spatial and spectral characteristics are held constant, the TM radiometric characteristic yields higher classification accuracies than the MSS radiometric characteristic in 23 of 24 cases. When the spectral and radiometric characteristics are held

constant, it is the MSS spatial characteristic (lower resolution) that yields higher classification accuracies in 21 of 24 cases. These results are quite similar to those of Alexander et al. (3) except that the earlier spectral comparisons were not conclusive. The factorial design of the experiment permits making these observations more explicit by averaging the results over one or two factors, as shown in Tables 3 and 4 for the two Overall levels.

The Level II/III Overall results have been averaged over one factor at a time in the upper part of Table 3 to eliminate that particular factor's effect. When the spectral factor is averaged out, the TM radiometry gives higher accuracies for both spatial resolutions and the lower spatial resolution gives better results for both radiometries. When the spatial factor is averaged out, TM radiometry gives better results with either spectral configuration and the TM spectral bands yield higher accuracies with either radiometry. With the radiometry averaged out, the MSS spatial resolution and the TM spectral bands always gave higher accuracies. Thus, the results of all of these comparisons are consistent.

The lower part of Table 3 averages the Level II/III Overall results over two of the factors and displays the difference in accuracies obtained for the TM and MSS configurations of the remaining factor. After this averaging, MSS spatial resolution, TM radiometry and TM spectral bands gave higher accuracies. The largest effect was due to the spatial factor with the spectral factor close behind. Since the training site pixels were completely enumerated, the results can be considered to be exact for the training sites used but cannot be applied to other areas. When the differences for the three factors at the bottom of Table 3 are taken together, there is a net difference in favor of the TM characteristics. In fact, Table 2 shows the full TM configuration yields higher accuracies than the full MSS configuration.

Table 4 presents a similar averaging of the three factors for the case of the Level I Overall results. Except for the fact that the classification accuracies are higher at this more general level of detail, all the relationships found in Table 3 hold in Table 4 as well. When averaged over two factors at a time, the differences in accuracy for the two configurations of the remaining factor are smaller than in Table 3 due to the compression in the range of accuracy values. However, the direction of the difference and the relative magnitudes are similar: the spatial effect is the largest, followed closely by the spectral effect, with the net difference favoring the TM configuration. Similar analyses of the individual Level I classes tend to yield more erratic results, however Agriculture, Urban and Range did show better accuracies for TM radiometry, TM spectral characteristics and MSS spatial characteristics although the difference for the latter was small for Agriculture and Range. All differences were small for Level I Forest due to its very high accuracies.

Alexander et al. (3) found very similar differences in accuracy for the spatial and radiometric factors when the other two factors were

averaged out. They had discounted the spectral results due to an imperfect simulation. Thus, the results of the factorial experiment with the training site data show that for three levels of information the MSS spatial resolution, TM spectral bands and TM radiometry contribute to higher classification accuracy. Moreover, the results are quite consistent with our earlier results (Ref. 3) using less exacting methods in a different area.

## 5. DISCUSSION AND CONCLUSIONS

The result that the Thematic Mapper's improved spectral and radiometric characteristics should yield improved classification accuracies is welcome. The result that the improved spatial characteristic yields lower accuracies is not. The effect is apparently due to increased spectral and radiometric heterogeneity of the higher spatial resolution data (Ref. 15). The use of multiple clusters in each training site did not overcome this effect. These results apply only to the training sites which consist of field-center pixels by definition. Since the proportion of field-center pixels will increase and the proportion of boundary pixels will decrease with higher spatial resolution, the effect on the classification accuracy of the whole area is not clear. The next phase of the analysis for the present data sets is designed to resolve this problem (as discussed in Section 3.4).

The spatial resolution results make clear that optimal use of high spatial resolution data will have to move beyond the per-pixel spectral classifier used here and incorporate spatial information so clearly present in the data in some form to maintain the classification accuracies the remote sensing community has come to expect from MSS data. Suggestions have included the use of texture, per-field classifiers and contextual classifiers. In terms of an operational, land remote sensing satellite system, the suggestion by Alexander *et al.* (3) to include high radiometric sensitivity on any such system can be re-emphasized and augmented: the spectral characteristics of TM could also be incorporated because they appear to provide a larger improvement in accuracy. Selection of a minimal set of bands would have to be considered for an operational system to control data volume.

## 5. REFERENCES

1. Morgenstern, J.P., R.F. Nalepka and J.D. Erickson 1977, Investigation of Thematic Mapper spatial, radiometric and spectral resolution, Proc. 11th International Symposium on Remote Sensing of Environment, Ann Arbor 25-29 April 1977, 693-701.
2. Clark, J. and N.A. Bryant 1977, Landsat-D Thematic Mapper simulation using aircraft multispectral scanner data, Proc. 11th International Symposium on Remote Sensing of Environment, Ann Arbor 25-29 April 1977.
3. Alexander, D.A., J.S. Buis, W. Acevedo, and R.C. Wrigley 1983, Spectral, spatial and radiometric factors in cover type discrimination, Proc International Geoscience and Remote Sensing Symp., San Francisco August 31-September 2, 1983.
4. Buis, J.S., W. Acevedo, D.A. Alexander and R.C. Wrigley 1983, The role of spectral, spatial and radiometric resolution on information content, Proc Machine Processing of Remotely Sensed Data Symp., W. Lafayette 21-23 June 1983.
5. Sadowski, F.G. and J.E. Sarno 1976, Forest classification accuracy as influenced by multispectral scanner spatial resolution, Final Report to NASA/Johnson Space Center, Houston, TX, Contract No. NAS9-14123.
6. Landgrebe, D.A., L. Beihl and W. Simmons 1977, An empirical study of scanner system parameters, IEEE Trans. Geosci. Electr., Vol GE-15, 120-130.
7. Sigman, R. and M. Craig 1981, Potential utility of Thematic Mapper data in estimating crop areas, Proc 15th International Symposium on Remote Sensing of Environment, Ann Arbor 11-15 May 1981.
8. Williams, D.L., J.R. Irons, B.L. Markham, R.F. Nelson D.L. Toll, R.S. Latty and M.L. Stauffer 1983, Impact of Thematic Mapper sensor characteristics on classification accuracy, Proc International Geoscience and Remote Sensing Symp., San Francisco August 31-September 2, 1983.
9. NASA Ames Research Center Airborne Missions and Applications Division Flight Summary Report, Flight No. 83-155, August 12, 1983.
10. Barker, John, 1983, Personal communication.
11. Bartolucci, L.A. and S.M. Davis 1983, The calibration of Landsat MSS data as an analysis tool, Proc Machine Processing of Remotely Sensed Data Symp., W. Lafayette 21-23 June 1983.
12. Dye, R.H. 1975, Restoration of Landsat images by discrete two-dimensional deconvolution, Proc. 10th International Symposium on Remote Sensing of Environment, Ann Arbor 6-10 October 1975.
13. Pont, W.F. Jr. 1982, Spatial and spectral simulation of Landsat images of agricultural areas, Proc Machine Processing of Remotely Sensed Data Symp., W. Lafayette 22-24 June 1982.
14. Schowengerdt, R.A., R.L. Antos and P.N. Slater 1974, Measurement of the Earth Resources Technology Satellite Multispectral Scanner OTF from operational imagery, Proc Society of Photo-Optical Instrumentation Engineers, Vol. 46, 247-257.
15. Markham, B.L. and J.R.G. Townshend 1981, Land cover classification accuracy as a function of sensor spatial resolution, Proc 15th International Symposium on Remote Sensing of Environment, Ann Arbor 11-15 May 1981.

1 Report No NASA TM-85995		2 Government Accession No		3 Recipient's Catalog No	
4 Title and Subtitle THE EFFECT OF SPATIAL, SPECTRAL AND RADIOMETRIC FACTORS ON CLASSIFICATION ACCURACY USING THEMATIC MAPPER DATA				5 Report Date July 1984	
				6 Performing Organization Code	
7 Author(s) R. Wrigley (Ames Research Center), and W Acevedo, D. Alexander, J. Buis (Technicolor Government Services, Inc , Ames Research Center, Moffett Field, CA 94035), and D. Card (Ames Research Center)				8 Performing Organization Report No A-9836	
				10 Work Unit No T-5819	
9 Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035				11 Contract or Grant No	
				13 Type of Report and Period Covered Technical Memorandum	
12 Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546				14 Sponsoring Agency Code 677-21-26	
15 Supplementary Notes Point of Contact: R. Wrigley, Ames Research Center, MS 242-4, Moffett Field, CA 94035 (415) 965-6060 or FTS 448-6060					
16 Abstract  An experiment of a factorial design was conducted to test the effects on classification accuracy of land cover types due to the improved spatial, spectral and radiometric characteristics of the Thematic Mapper (TM) in comparison to the Multispectral Scanner (MSS). High altitude aircraft scanner data from the Airborne Thematic Mapper instrument was acquired over central California in August, 1983 and used to simulate Thematic Mapper data as well as all combinations of the three characteristics for eight data sets in all. Results for the training sites (field-center pixels) showed better classification accuracies for MSS spatial resolution, TM spectral bands and TM radiometry in order of importance.					
17 Key Words (Suggested by Author(s)) Thematic Mapper Factorial experiment Simulation Classification			18 Distribution Statement  Unlimited  Subject Category - 43		
19 Security Classif (of this report) Unclassified		20 Security Classif (of this page) Unclassified		21 No of Pages 11	22 Price* A02

**End of Document**