General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

Thematic Mapper Design Parameter Investigation

NASA CR-156756

(NASA-CR-156756) THEMATIC MAPPER DESIGN PARAMETER INVESTIGATION Final Report, Jun. 1976 - Jan. 1978 (IBM Federal Systems Div.) 147 p HC A07/MF A01 CSCL 08B

N78-23505

Unclas

G3/43 16643

MAY 1978
RECEIVED
NASA STI FACILITY

THEMATIC MAPPER DESIGN PARAMETER INVESTIGATION

Final Report for Period June 1976 to January 1978

January 1978

Prepared for

Goddard Space Flight Center
Greenbelt, Maryland 20771

by

C. P. Colby, Jr.
S. G. Wheeler

Federal Systems Division
INTERNATIONAL BUSINESS MACHINES CORPORATION
Gaithersburg, Maryland 20760

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Thematic Mapper Design Parameter Investigation		5. Report Date January, 1978 6. Performing Organization Code
7. Author(s) C. P. Colby, S. G. Wheeler, IBM R. Kauth, R. Nalepka, ERIM		8. Performing Organization Report No. FSD 780001
9. Performing Organization Name and Address IBM Federal Systems Division Gaithersburg, Maryland 20760		10. Work Unit No. 11. Contract or Grant No. NAS5-23585
12. Sponsoring Agency Name and Address Goddard Space Flight Center Greenbelt, Maryland 20771 Warner Miller, Technical Officer		13. Type of Report and Period Covered Final Report June, 1976-Jan. 1978 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This study simulated the multispectral data sets to be expected from three different Thematic Mapper configurations, and the ground processing of these data sets by three different resampling techniques. The simulated data sets were then evaluated by processing them for multispectral classification, and the Thematic Mapper configuration and resampling technique which provided the best classification accuracy were identified.

17. Key Words (Selected by Author(s))	18. Distribution Statement
System Simulation Thematic Mapper Multispectral Classification	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
		141	

*For sale by the Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.

ORIGINAL PAGE IS
OF POOR QUALITY

APPROVAL

This final report for the Thematic Mapper Design Parameter Investigation has been reviewed and approved by the undersigned.

Charles P. Colby, Jr.

Charles P. Colby, Jr.
Project Engineer
IBM Federal Systems Division
Gaithersburg, Maryland

Ralph Bernstein

Ralph Bernstein
Senior Engineer
IBM Federal Systems Division
Gaithersburg, Maryland

ACKNOWLEDGEMENT

The technical group which performed the Thematic Mapper Design Parameter Investigation wishes to acknowledge with gratitude the financial and technical support provided by NASA-GSFC personnel for this study. Worth of particular comment is the interest and encouragement provided by M. Maxwell, and the technical assistance provided by the Technical Officer for this contract, W. Miller, and by Dr. J. Barker, M. Ritter, K. Duck, L. Thompson, and L. Linstrom, in defining the parameters used to model the various Thematic Mapper configurations.

PREFACE

The objective of this study was to obtain a quantitative evaluation of the relative performance of several potential multispectral scanner designs for the Landsat-D Thematic Mapper. The study produced simulated multispectral Thematic Mapper data sets by processing aircraft resolution digital multispectral data through models of the various Thematic Mapper configurations and then through three sampling techniques as a simulation of ground processing. The resulting data sets were subjected to multispectral classification, and the effectiveness of the various systems was compared on the basis of the classification results obtained.

Two data sets were used in this study, an aircraft data set with nominally a 3-meter resolution and a synthetic data set of the same nominal resolution. This latter data set was generated using psuedo-random numbers based on the class statistics of the aircraft data set, and it was constructed to provide a more satisfactory distribution of field size and shape than was available in the aircraft data set. Three Thematic Mapper configurations, differing in the type of filtering and sampling were considered. One configuration modeled an "integrate and dump" sampler, while the other two employed a conventional presampling filter with a "sample and hold" circuit, and differed only in the along-scan sampling rate. The resampling techniques considered were nearest neighbor assignment, cubic convolution resampling, and point-spread-function-compensation resampling. Classification of the simulated Thematic Mapper data sets was performed using a maximum-likelihood classifier.

While the classification results exhibited differing trends depending on such variables as class, on the basis of an average over classes, and with respect to both classification accuracy and proportion estimation, the configuration with a conventional presampling filter and 1.4 samples/IFOV along-scan provided the best overall performance for the image data derived from the synthetic data set. There was little difference among the resampling techniques. On the same basis, although with a more limited data base available for statistical analysis, the configuration with a conventional presampling filter, 1.0 samples/IFOV and point-spread-function-compensation resampling provided the best overall performance for the image data derived from the aircraft data set.

CONTENTS

Section		Page
1	INTRODUCTION	1-1
2	DATA SET DESCRIPTION	2-1
2.1	Aircraft Data Set	2-1
2.2	Synthetic Data Set	2-1
3	PREPROCESSING OF DATA SETS	3-1
3.1	Aircraft Data Set Preprocessing	3-1
3.2	Synthetic Data Set Preprocessing	3-5
4	THEMATIC MAPPER DATA SET SIMULATION	4-1
4.1	Simulation Processing	4-1
4.1.1	Aperture Convolution	4-1
4.1.2	Noise Generation and Insertion	4-7
4.1.3	Filtering	4-10
4.1.4	Sampling	4-12
4.2	Resampling	4-12
4.2.1	Nearest Neighbor Assignment	4-12
4.2.2	Cubic Convolution	4-12
4.2.3	Point Spread Function Compensation	4-12
4.3	Simulated Thematic Mapper Data Sets	4-16
5	CLASSIFICATION PROCESSING AND ANALYSIS OF RESULTS	5-1
5.1	Classification Processing	5-1
5.1.1	Aircraft Image and Related Thematic Mapper Images	5-1
5.1.2	Synthetic Image and Associated Thematic Mapper Images	5-2
5.2	Analysis of Percent Correctly Classified	5-5
5.2.1	Synthetic Image Analysis	5-5
5.2.2	Aircraft Image Analysis	5-8
5.3	Analysis of Proportion Errors	5-36
5.3.1	Synthetic Image Analysis	5-39
5.3.2	Aircraft Image Analysis	5-48
6	CONCLUSIONS AND SUMMARY	6-1
7	REFERENCES	7-1
Appendix		
A	ERIM FINAL REPORT ON THE THEMATIC MAPPER DESIGN PARAMETER INVESTIGATION	A-1

ILLUSTRATIONS

Figure		Page
2.1-1	Aircraft Data Set	2-3
3.1-1	Aircraft Data Set Preprocessing	3-2
3.1-2	Geometry of Panoramic Distortion Compensation	3-4
3.2-1	Synthetic Data Set Preprocessing	3-6
3.2-2	Rotated Synthetic Data Set	3-9
4.1-1	Simulation Processing Flow	4-2
4.1-2	Simulation of Thematic Mapper Scan Processing	4-5
4.1-3	Aperture Function Cross Section Thematic Mapper Band 1	4-6
4.1-4	Goldberg Filter	4-11
4.2-1	Nearest Neighbor Assignment	4-13
4.2-2	Cubic Convolution Resampling	4-14
4.2-3	Point Spread Function Compensation Resampling	4-15
4.2-4	Central Along-Scan Cross-Section of Impulse Response Configuration 1, Band 1	4-18
4.2-5	Central Along-Scan Cross-Section of Impulse Response Configuration 3, Band 1	4-19
4.3-1	Data Key to Figures 4.3-2 through 4.3-7	4-20
4.3-2	Simulated Thematic Mapper Data Derived from Aircraft Data Set -- Configuration 1	4-21
4.3-3	Simulated Thematic Mapper Data Derived from Aircraft Data Set -- Configuration 2	4-22
4.3-4	Simulated Thematic Mapper Data Derived from Aircraft Data Set -- Configuration 3	4-23
4.3-5	Simulated Thematic Mapper Data Derived from Synthetic Data Set -- Configuration 1	4-24

Figure		Page
4.3-6	Simulated Thematic Mapper Data Derived from Synthetic Data Set -- Configuration 2	4-25
4.3-7	Simulated Thematic Mapper Data Derived from Synthetic Data Set -- Configuration 3	4-26
5.2-1	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Resampling and Configuration for Entire Fields in TM Images Derived from Synthetic Data Set	5-13
5.2-2	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Field Size and Field Shape for Entire Fields in TM Images Derived from Synthetic Data Set	5-14
5.2-3	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Resampling and Configuration for Field Center Fields in TM Images Derived from Synthetic Data Set	5-15
5.2-4	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Field Size and Field Shape for Field Center Fields in TM Images Derived from Synthetic Data Set	5-16
5.2-5	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Resampling and Configuration for Entire Training Fields in TM Images Derived from Aircraft Data Set	5-31
5.2-6	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Resampling and Configuration for Entire Test Fields in TM Images Derived from Aircraft Data Set	5-32
5.2-7	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Resampling and Configuration for Field Center Training Fields in TM Images Derived from Aircraft Data Set	5-33
5.2-8	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Resampling and Configuration for Field Center Test Fields in TM Images Derived from Aircraft Data Set	5-34
5.2-9	Probabilities of Correct Classification, Averaged Over Classes, Showing Effects of Field Size and Field Shape for Field Center Training Fields in TM Images Derived from Aircraft Data Set and in Aircraft Data Set	5-35

Figure		Page
5.3-1	Estimated Root Mean Square Proportion Errors for 2.5 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)	5-44
5.3-2	Estimated Root mean Square Proportion Errors for 5.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)	5-45
5.3-3	Estimated Root Mean Square Proportion Errors for 10.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)	5-46
5.3-4	Estimated Root Mean Square Proportion Errors for 20.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)	5-47
5.3-5	Estimated Root Mean Square Proportion Errors for 2.5 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50 Percent Corn Scenario)	5-53
5.3-6	Estimated Root Mean Square Proportion Errors for 5.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50 Percent Corn Scenario)	5-54
5.3-7	Estimated Root Mean Square Proportion Errors for 10.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50 Percent Corn Scenario)	5-55
5.3-8	Estimated Root Mean Square Proportion Errors for 20.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50 Percent Corn Scenario)	5-56
5.3-9	Estimated Root Mean Square Proportion Errors for Smaller Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (Equally Likely Classes)	5-60
5.3-10	Estimated Room Mean Square Proportion Errors for Intermediate Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (Equally Likely Classes)	5-61

Figure		Page
5.3-11	Estimated Root Mean Square Proportion Errors for Larger Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (Equally Likely Classes)	5-62
5.3-12	Estimated Root Mean Square Proportion Errors for Smaller Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images from Aircraft Data Set (50 Percent Corn Scenario)	5-66
5.3-13	Estimated Root Mean Square Proportion Errors for Intermediate Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (50 Percent Corn Scenario)	5-67
5.3-14	Estimated Root Mean Square Proportion Errors for Larger Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (50 Percent Corn Scenario)	5-68

TABLES

Table		Page
2.1-1	M-7 Scanner Characteristics	2-2
3.1-1	Aircraft Data Set Quantization Conversion and Atmospheric Transformation	3-3
3.2-1	Synthetic Data Set Atmospheric Transformation	3-7
4.1-1	Thematic Mapper Configurations Simulated	4-3
4.1-2	Power Spectrum Weighting Constants	4-8
4.1-3	Signal Dependent Noise Parameters	4-9
4.2-1	Gaussian Approximation to Point Spread Function	4-17
5.1-1	Identified Fields in Aircraft and Related Images	5-3
5.1-2	Definitions and Values of Cluster Parameters for Clustering Aircraft-derived Thematic Mapper Images	5-4
5.1-3	Fields in Synthetic and Related Images	5-5
5.2-1	Analysis of Logits of Correct vs. Incorrect Classification for Entire Fields of Synthetic Image	5-6
5.2-2	Analysis of Logits of Correct vs. Incorrect Classification for Field Center Fields of Synthetic Image	5-7
5.2-3	Probabilities of Correct Classification Showing Effects of (class, resampling, configuration) for Entire Fields of Synthetic Image Data	5-9
5.2-4	Probabilities of Correct Classification Showing Effects of (class, size, shape) for Entire Fields of Synthetic Image Data	5-10
5.2-5	Probabilities of Correct Classification Showing Effects of (class, resampling, configuration) for Field Center Fields of Synthetic Image Data	5-11
5.2-6	Probabilities of Correct Classification Showing Effects of (class, size, shape) for Field Center Fields of Synthetic Image Data	5-12
5.2-7	Analysis of Variance from Stepwise Regression of Percent Correctly Classified for Thematic Mapper Data Derived from Aircraft Data Set	5-18

ORIGINAL PAGE IS
OF POOR QUALITY

Table		Page
5.2-8	Probabilities of Correct Classification Showing Effects of (class, resampling, configuration) for Field Center Training Fields of TM Data Derived from Aircraft Image	5-19
5.2-9	Probabilities of Correct Classification Showing Effects f (class, resampling, configuration) for Field Center Test Fields of TM Data Derived from Aircraft Image	5-20
5.2-10	Probabilities of Correct Classification Showing Effects of (class, resampling, configuration) for Entire Training Fields of TM Data Derived from Aircraft Image	5-21
5.2-11	Probabilities of Correct Classification Showing Effects of (class, resampling, configuration) for Entire Test Fields of TM Data Derived from Aircraft Image	5-22
5.2-12	Percentage Correctly Classified Showing Effects of (class, size, shape) for Field Center Training Fields of TM Data Derived from Aircraft Image	5-23
5.2-13	Percentage Correctly Classified Showing Effects of (class, size, shape) for Field Center Training Fields of Aircraft Image	5-24
5.2-14	Percentage Correctly Classified Showing Effects of (class, size, shape) for Entire Training Fields of TM Data Derived from Aircraft Image	5-25
5.2-15	Percentage Correctly Classified Showing Effects of (class, size, shape) for Entire Training Fields of Aircraft Image	5-26
5.2-16	Percentage Correctly Classified Showing Effects of (class, size, shape) for Field Center Test Fields of TM Data Derived from Aircraft Image	5-27
5.2-17	Percentage Correctly Classified Showing Effects of (class, size, shape) for Field Center Test Fields of Aircraft Image	5-28
5.2-18	Percentage Correctly Classified Showing Effects of (class, size, shape) for Entire Test Fields of TM Data Derived from Aircraft Image	5-29
5.2-19	Percentage Correctly Classified Showing Effects of (class, size, shape) for Entire Test Fields of Aircraft Image	5-30
5.3-1	Estimated Average Proportion Errors for 2.5 Acre Fields of Simulated Image (Equally Likely Classes)	5-40

Table		Page
5.3-2	Estimated Average Proportion Errors for 5.0 Acre Fields of Simulated Image (Equally Likely Classes)	5-41
5.3-3	Estimated Average Proportion Errors for 10.0 Acre Fields of Simulated Image (Equally Likely Classes)	5-42
5.3-4	Estimated Average Proportion Errors for 20.0 Acre Fields of Simulated Image (Equally Likely Classes)	5-43
5.3-5	Estimated Average Proportion Errors for 2.5 Acre Fields of Simulated Image (50 Percent Corn Scenario)	5-49
5.3-6	Estimated Average Proportion Errors for 5.0 Acre Fields of Simulated Image (50 Percent Corn Scenario)	5-50
5.3-7	Estimated Average Proportion Errors for 10.0 Acre Fields of Simulated Image (50 Percent Corn Scenario)	5-51
5.3-8	Estimated Average Proportion Errors for 20.0 Acre Fields of Simulated Image (50 Percent Corn Scenario)	5-52
5.3-9	Estimated Average Proportion Errors for Smaller Fields in Aircraft Image Derived TM Images (Equally Likely Classes)	5-57
5.3-10	Estimated Average Proportion Errors for Intermediate Fields in Aircraft Image Derived TM Images (Equally Likely Classes)	5-58
5.3-11	Estimated Average Proportion Errors for Larger Fields in Aircraft Image Derived TM Images (Equally Likely Classes)	5-59
5.3-12	Estimated Average Proportion Errors for Smaller Fields in Aircraft Image Derived TM Images (50 Percent Corn Scenario)	5-63
5.3-13	Estimated Average Proportion Errors for Intermediate Fields in Aircraft Image Derived TM Images (50 Percent Corn Scenario)	5-64
5.3-14	Estimated Average Proportion Errors for Larger Fields in Aircraft Image Derived TM Images (50 Percent Corn Scenario)	5-65

ORIGINAL PAGE IS
OF POOR QUALITY

Section 1

INTRODUCTION

The Thematic Mapper is the multispectral scanner which will be flown on Landsat-D. One of the primary applications for the digital imagery obtained from this sensor will be the classification and discrimination of agricultural crops in order to provide accurate and timely crop inventory information. The accuracy of the inventory estimates obtained by this method depend on many factors, but of these only the sensor system characteristics and the ground data processing performed prior to classification are under control of the system designer. There is therefore great interest on the part of these designers in the effects their design choices will have on the accuracy of the classification results to be expected from the operational system.

The Thematic Mapper Design Parameter Investigation was undertaken in order to obtain comparative measures of the effect of alternative choices of sensor and ground processing parameters on the utility of Thematic Mapper data. The approach used was to start with high-resolution multispectral aircraft scanner data, process this data through models of several Thematic Mapper system configurations, machine classify agricultural areas in the resulting simulated Thematic Mapper data, and use the classification and mensuration accuracies obtained to quantitatively compare the performance of the configurations. Participants in this study were IBM's Advanced Digital Image Processing Group and its Earth Resources Laboratory (ERL), the Environmental Research Institute of Michigan (ERIM), and NASA/GSFC. The Digital Image Processing Group assumed overall study responsibility and provided design and execution of the data processing required to generate the simulated Thematic Mapper data sets; the Earth Resources Laboratory performed the multispectral classification of data sets. ERIM provided consultation on the overall study design. Both ERIM and ERL independently contributed their assessment of the performance of the several simulated Thematic Mapper configurations, based on analysis of the classification results obtained. NASA/GSFC provided review, guidance, and the electronic and optical specifications for the Thematic Mapper configurations to be simulated.

At the start of this effort, meetings were held to review the study plan. All parties involved found the plan acceptable and the study proceeded to address the problem of data set selection.

It was originally contemplated that three multispectral aircraft data sets would be employed, with the ERIM M-7 Scanner identified as a likely source of data because of its spectral responses, resolution (2-3 mrad), and channel-to-channel registration. During the early discussions, it was discovered that the M-7 data, as normally provided, is on a sampling lattice of approximately 10 meters along-scan and along-track. This sampling interval corresponds to the nominal resolution (5 mrad) of the M-7 thermal band for an aircraft altitude of 5,000 feet. The study participants recognized that data sampled this coarsely was inadequate for use in simulation of a system whose resolution was 30 meters. Therefore it was necessary to investigate other sources of the required aircraft data sets, seeking an available data set with a smaller sampling

interval which still provided a good spectral match with the bands of the Thematic Mapper and registered spectral channels. The various sources of aircraft scanner data were contacted (e.g., LARS, NASA/JSC) in an effort to locate a satisfactory data set. However, the combined resolution and spectral requirements could not be met by any available multispectral data set. For this reason, arrangements were made for ERIM to reprocess an analog M-7 scanner data tape to provide a data set sampled on a 3-meter sampling lattice for use in this study. This sampling interval equals the resolution of the M-7 visible bands from a 5,000 foot altitude.

Furthermore, in addition to the problems encountered in obtaining a data set with the required spectral and resolution characteristics, it proved impossible to locate data sets with a satisfactory distribution of crops, field sizes, and field shapes, and also with reliable, consistent ground truth. It was therefore decided to generate a synthetic multispectral data set for use in this study. This data set was constructed using six single-subclass classes whose statistics (means and covariance matrices) were selected from the statistical characterizations of the classes and subclasses in the aircraft image. Selection of these well defined classes, with the resulting idealized statistical class characterization, was consistent with the study objective of comparative sensor and ground processing system evaluation, rather than with an objective of classifier performance evaluation. Within this data set, the distribution of classes, field sizes, and field shape factors (ratio of length to width) were specified to provide an extensive data base which would support a thorough statistical analysis of the dependence of classification accuracy on sensor and processing system design choices. The comparative multispectral classification performance on simulated Thematic Mapper data derived from this synthetic data set, validated where possible against the performance on simulated Thematic Mapper data derived from the aircraft data set, would then provide a complete and convincing presentation of the effects of these design choices.

Details of the simulation processing are given in Sections 2 through 4 of this report. The classification results and the evaluation by ERL of the implications of results for the various system configurations are given in Section 5, and the conclusions derived from the study are summarized in Section 6. ERIM's evaluation of the study results is presented in Appendix A.

Section 2

DATA SET DESCRIPTION

2.1 AIRCRAFT DATA SET

The Aircraft Data Set employed in this study contained data gathered as part of the Corn Blight Watch Experiment, specifically data segment 204, 13 August 1971, 10:30 EST. The entire data set is approximately 1.6 km by 19.2 km. The data was taken with the 12 band ERIM M-7 multispectral scanner whose spectral response is given in Table 2.1-1, along with the RFOV associated with these channels for a 5,000 foot aircraft altitude.

Because this study required simulation of the 30-meter resolution of the Thematic Sensor, modeling the aperture function, it was necessary to have the data set sampled at an interval considerably smaller than 30 meters so that a precise model of the aperture function could be employed. Therefore, the aircraft data set was digitized at sampling intervals of 3.0 meters along scan and 2.74 meters along track. Each line of the data set contained 416 samples, and 2,300 lines were used for simulation processing. The FOV of the data set extended approximately $+15^{\circ}$ to -45° from nadir, with nadir located at roughly pixel 323. The data set had been scan angle corrected by means of multiplicative coefficients derived from a subsample of this data. This provided a radiometric equalization along the scan. The channels were in channel-to-channel registration. Absolute radiometric calibration was provided for all channels except 10 and 11.

Ground truth data, in the form of annotated photography of the data segment, together with tables relating to the annotation to specific crops and planting information, was obtained. Crops represented in this segment include corn, soybeans, wheat, pasture, woods, and oats. Portions of the segment are also idle farmland, set aside and non-farm areas. This ground truth contained some apparent inconsistencies with respect to visual inspection of the aircraft data set, and considerable effort was required to deal with this problem.

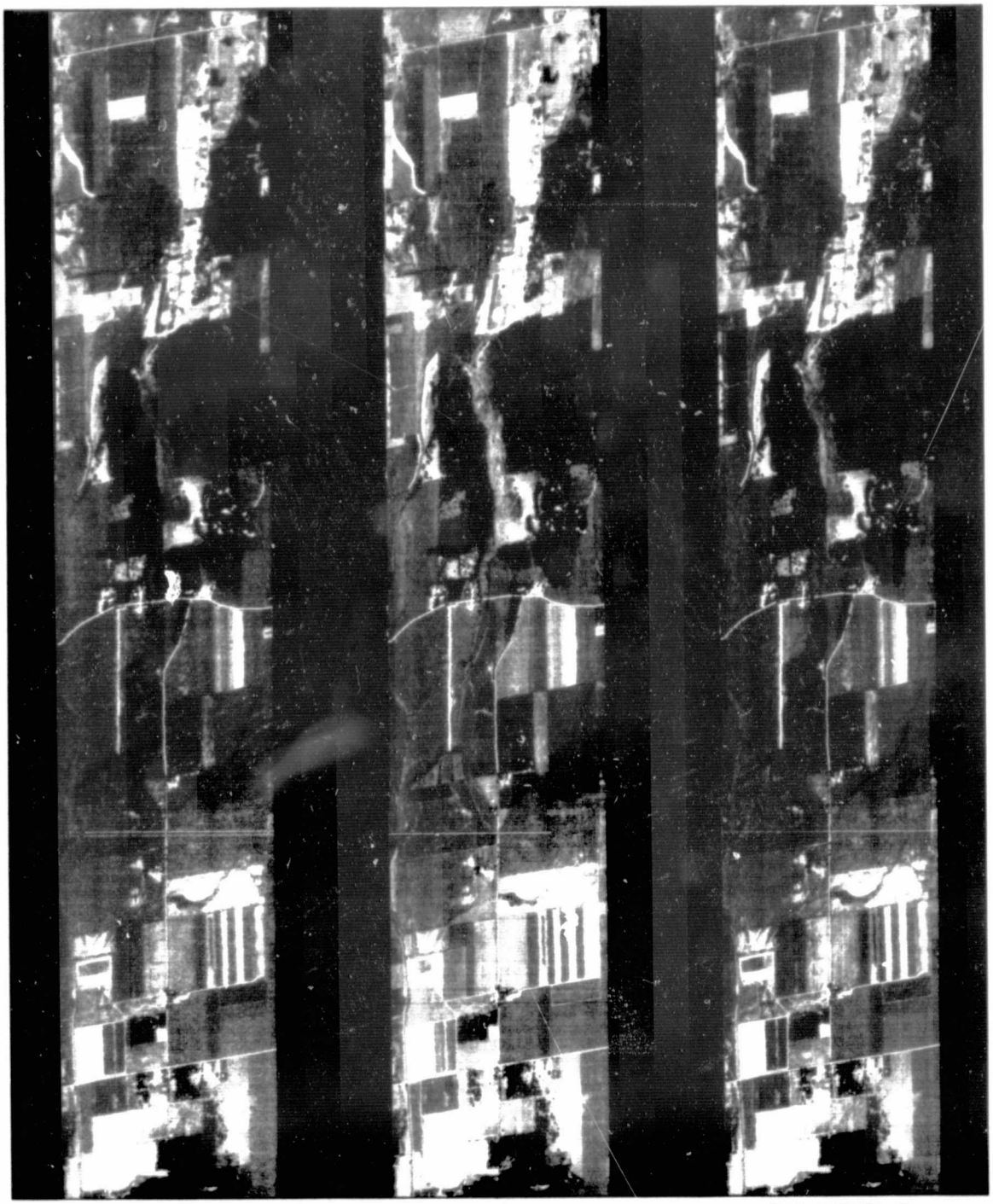
A photographic presentation of the region employed is provided in Figure 2.1-1.

2.2 SYNTHETIC DATA SET

A synthetic 6-channel multispectral data set was created especially for use in this study, constructed in a manner which provided a desirable distribution of field sizes and shapes. The purpose of using this data set was to augment the data available for the subsequent parametric analysis of multispectral classification results by providing a data set with a more extensive distribution of field shapes and sizes than was available in any of the aircraft data sets considered and which was supported by a reliable and consistent set of ground truth data. Since this study was concerned with the relative performance of sensor and ground processing systems, and not with evaluating classification techniques, the improved absolute classification accuracy

Table 2.1-1. M-7 Scanner Characteristics

	Data Channel	Waveband μm	Footprint Along Scan	Footprint Along Track
Visible	3	0.46 - 0.49	3.8m	3.1m
	4	0.48 - 0.52	3.8m	3.1m
	5	0.50 - 0.54	3.8m	3.1m
	6	0.53 - 0.57	3.8m	3.1m
	12	0.55 - 0.60	3.8m	3.1m
	8	0.58 - 0.64	3.8m	3.1m
	2	0.62 - 0.70	3.8m	3.1m
	1	0.67 - 0.94	3.8m	3.1m
Near IR	7	1.0 - 1.4	3.8m	7.6m
	9	1.5 - 1.8	3.8m	7.6m
	11	2.0 - 2.6	3.8m	7.6m
Thermal	10	9.3 - 11.7	5.0m	5.0m



TM BAND 1 TM BAND 2 TM BAND 3

Figure 2.1-1. Aircraft Data Set (Sheet 1 of 2)

ORIGINAL PAGE IS
OF POOR QUALITY

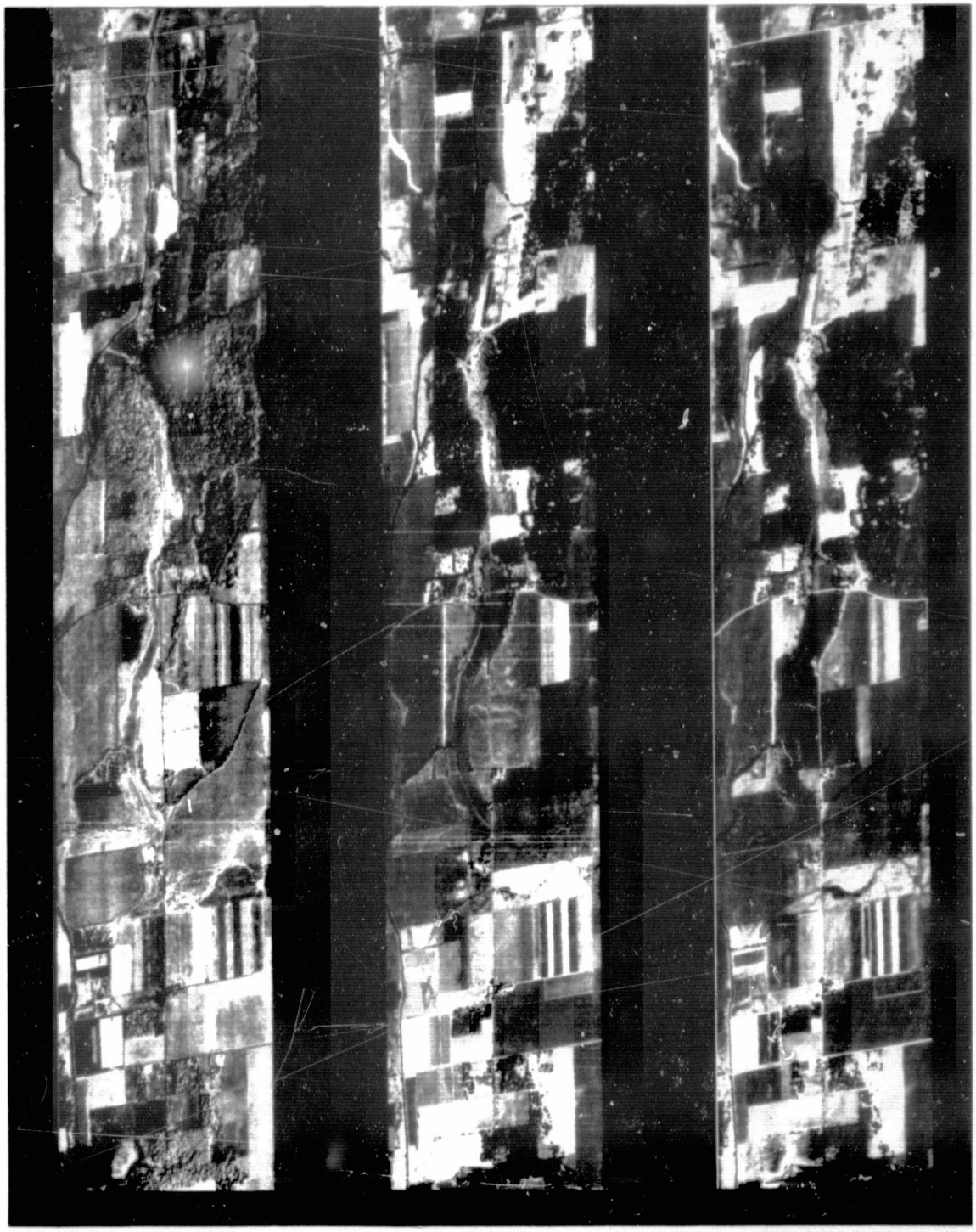


Figure 2.1-1. Aircraft Data Set (Sheet 2 of 2)

which might result from the idealized spectral crop characterization implicit in a synthetic data set was judged to be of less importance than the need to have an adequate data base for a thorough statistical analysis of the relative performance of the sensor and ground processing systems as measured by classification results. This data set was generated by a computer program (SERID) which had been previously created to provide test data for use in the development and testing of classification software at the Johnson Space Center. It consisted of 1,650 lines of 442 samples, on the same 3.0 meter by 2.74 meter sampling lattice as the aircraft data set. This pattern provided field sizes of 2.5, 5, 10, and 20 acres, with field form relations of 1 x 1, 1 x 2, and 1 x 4. Six classes were provided: trees, corn, pasture, winter wheat, soybeans, and soil. The spectral signatures for these classes were obtained from the covariance matrix for these classes derived from the aircraft data set.

A photographic presentation of this data set is provided in Figure 3.2-2.

Section 3

PREPROCESSING OF DATA SETS

The differing physical characteristics of the synthetic and aircraft data sets employed necessitated separate preprocessing to provide a common input format for the TM simulation processing. This preprocessing included application of a radiometric transformation which modeled the atmospheric effects introduced in transit from the aircraft altitude to the spacecraft altitude, and in the case of the aircraft data set, compensation for panoramic distortion.

3.1 AIRCRAFT DATA SET PREPROCESSING

The scan angle dependent radiometric effects in this data were compensated as part of the original preparation of the data set. However, the panoramic distortion associated with wide-angle, aircraft altitude scanners, remained. The preprocessing applied to this data set is shown schematically in Figure 3.1-1.

The first step in preprocessing was decommutation of the data in the aircraft data set for the six spectral channels to be used to simulate Thematic Mapper data. This produced six complete spectral images of 2,300 lines, with 416 samples per line, and with the 9-bit data field for each sample. Each of these images was then histogrammed to determine its dynamic range, so that an appropriate linear radiometric transformation could be specified to transform the 9-bit samples into an 8-bit field. (Since the data from the Thematic Mapper is to be quantized to 8-bits, and since 8-bit data was more suitable for use in the simulation and classification software, it was decided to perform this conversion as part of the preprocessing.)

Table 3.1-1 provides a tabulation of the dynamic range of the 9-bit data and the resulting 8-bit data, together with the radiometric transformation which was used to model atmospheric effects. This latter will be discussed subsequently. After conversion to 8-bit samples, the resulting six spectral images of the aircraft data set were processed into the LARS-II format for use as a control data set in the classification evaluation of the simulated Thematic Mapper data sets.

For this aircraft data, the next step in preprocessing was the application of an along-scan geometric compensation for panoramic distortion. This compensation, whose geometry is diagrammed in Figure 3.1-2, consisted of an along-scan resampling using the nearest neighbor technique, to produce a spectral image whose along-scan sampling interval, in linear units, was equal to the nadir sampling interval of the aircraft data, 3 meters. While the original intent had been to use only a portion of the aircraft data whose centerline was the nadir track, the asymmetry of the aircraft data set with respect to the nadir track made it necessary to employ portions of the data set with more severe panoramic distortion than anticipated, and required compensation of the data before simulation processing could be performed. In conjunction with this resampling, the samples in each line were inverted through

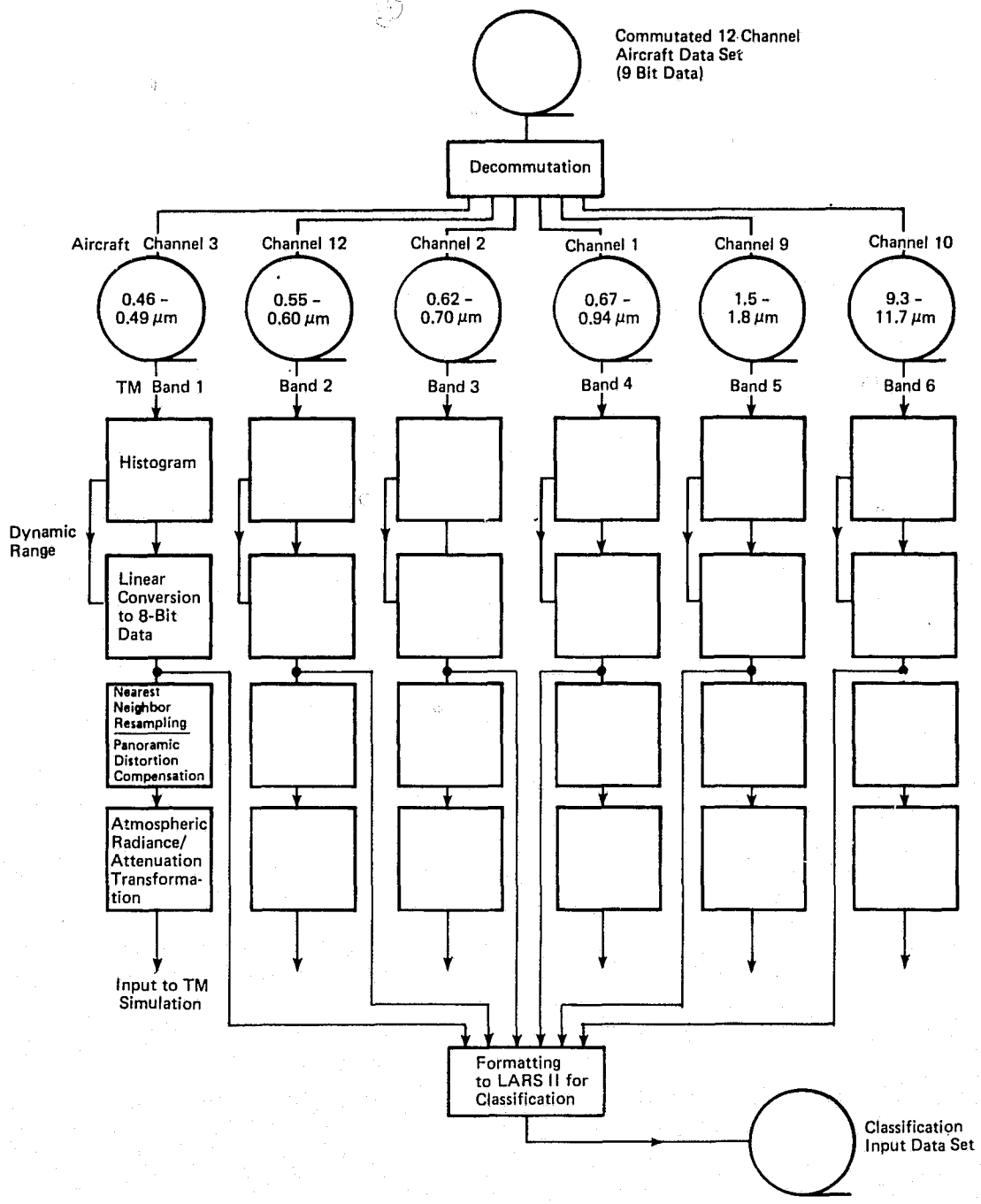


Figure 3.1-1. Aircraft Data Set Preprocessing

Table 3.1-1. Aircraft Data Set Quantization Conversion and Atmospheric Transformation

M7	TM	9 Bit M7 Output (C_9)		M7 Data Scaled to Bytes (C_8')		$C_8 = A_i + B_i C_9$		$C_8 = A_i' + B_i' C_8'$			
Band	Band	Min	Max	Min	Max	A_i	B_i	A_i'	B_i'	TM Response	M-7 Response
3	1	0	277	20	240	6.0246	0.6584	-10.5551	0.8290	0.45-0.52μm	0.46-0.49μm
12	2	38	269	20	240	13.2721	0.4536	20.9833	0.4763	0.52-0.60μm	0.55-0.60μm
2	3	25	269	20	240	13.0146	0.4387	14.2509	0.4866	0.63-0.69μm	0.62-0.70μm
1	4	0	241	20	240	9.0401	1.0194	-13.2940	1.1167	0.76-0.90μm	0.67-0.94μm
9	5	0	241	20	240	14.3860	0.8549	-5.1213	0.9754	1.55-1.75μm	1.5-1.8μm
10	6	0	198	20	240	59.1102	0.4029	51.8580	0.3626	10.40-12.50μm	9.3-11.7μm

ORIGINAL PAGE IS
OF POOR QUALITY

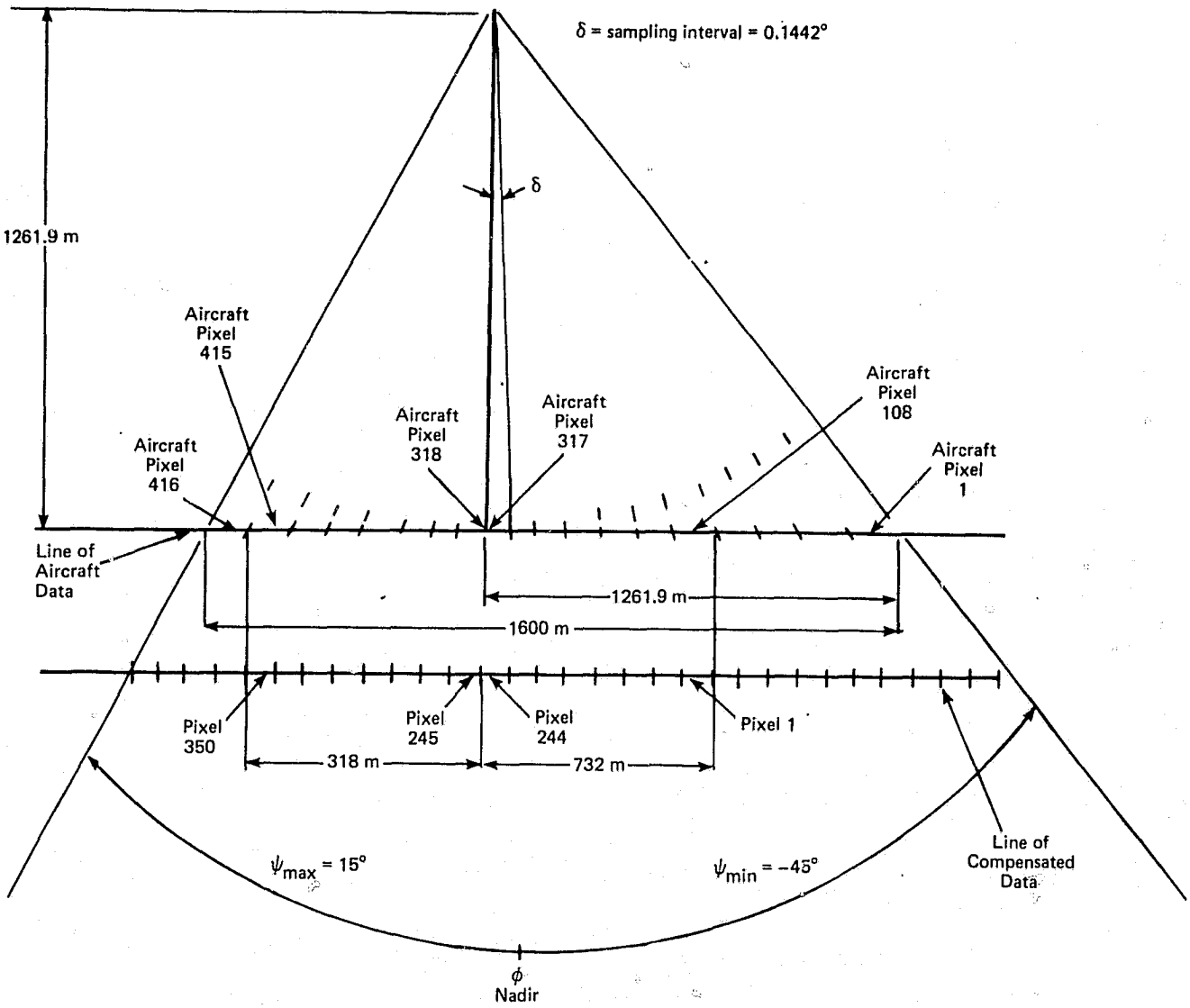


Figure 3.1-2. Geometry of Panoramic Distortion Compensation

the nadir so that the digital data would have the conventional relationship with the ground truth imagery, with the first sample of the first line corresponding to the upper left-hand corner of the imagery.

To model the radiometric effects of the atmosphere which will modify the radiance sensed at the spacecraft altitude, a radiometric transformation was applied to each band of the compensated aircraft data. For TM bands 1 through 5, the transformations were developed by using the calibration data provided for the aircraft data set to determine the radiance sensed by the aircraft in absolute units, and using an atmospheric simulation program, (RESET, developed at JSC) to establish the relationship between absolute radiance at aircraft altitude and the corresponding absolute radiance at spacecraft altitude. This latter relationship was discovered to be essentially linear. Clear atmospheric conditions interpreted as visibility of approximately 23 km, were assumed throughout, since the aircraft data set was taken under these conditons. This information, together with specification of the Thematic Mapper's dynamic range, was sufficient to define the transformation between the 9-bit aircraft data samples and the 8-bit samples at the altitude of the Thematic Mapper. The coefficients of this transformation, modified to include the effect of the initial 9-bit to 8-bit conversion, are presented in Table 3.1-1.

For the emissive channel of the aircraft data set, which was used to simulate band 6 of the Thematic Mapper, no calibration data was available, since this channel of the M7-scanner is not calibrated to absolute units. It was therefore necessary to make some reasonable assumptions regarding the temperature of sensed ground area for the known scene content (mainly corn) at this time of year (mid-August), and to assume a reasonable range of temperatures (260°K to 320°K) corresponding to the dynamic range of the digital data in this channel. By this means, a rough calibration relationship was established for the emissive channel, and a linear radiometric transformation was determined between the 9-bit aircraft samples and the 8-bit samples at the Thematic Mapper altitude. This transformation, again including the effect of the initial 9-bit to 8-bit conversion, is presented in Table 3.1-1.

A detailed discussion of the development of these atmospheric transformations is contained in Reference 1.

3.2 SYNTHETIC DATA SET PREPROCESSING

Preprocessing for the synthetic aircraft data set was simplified by the facts that it consisted of 8-bit samples and, by construction, did not suffer from panoramic distortion. As part of the preprocessing, the data in this set was rotated 5° counterclockwise with respect to its sampling lattice, so that the field boundaries, in the set, which were required by the synthesizing program to be parallel and perpendicular to the sampling lattice, would no longer be so. The preprocessing applied to this data set is shown schematically in Figure 3.2-1.

For the synthetic data set, the first step in preprocessing was a reformatting into separate spectral images. Since the data set produced by the SERID program was in a modified LARS-II format which was incompatible with the classification software, these separate spectral images were then reformatted into a conventional LARS-II format for use as a control data set in the multispectral classification.

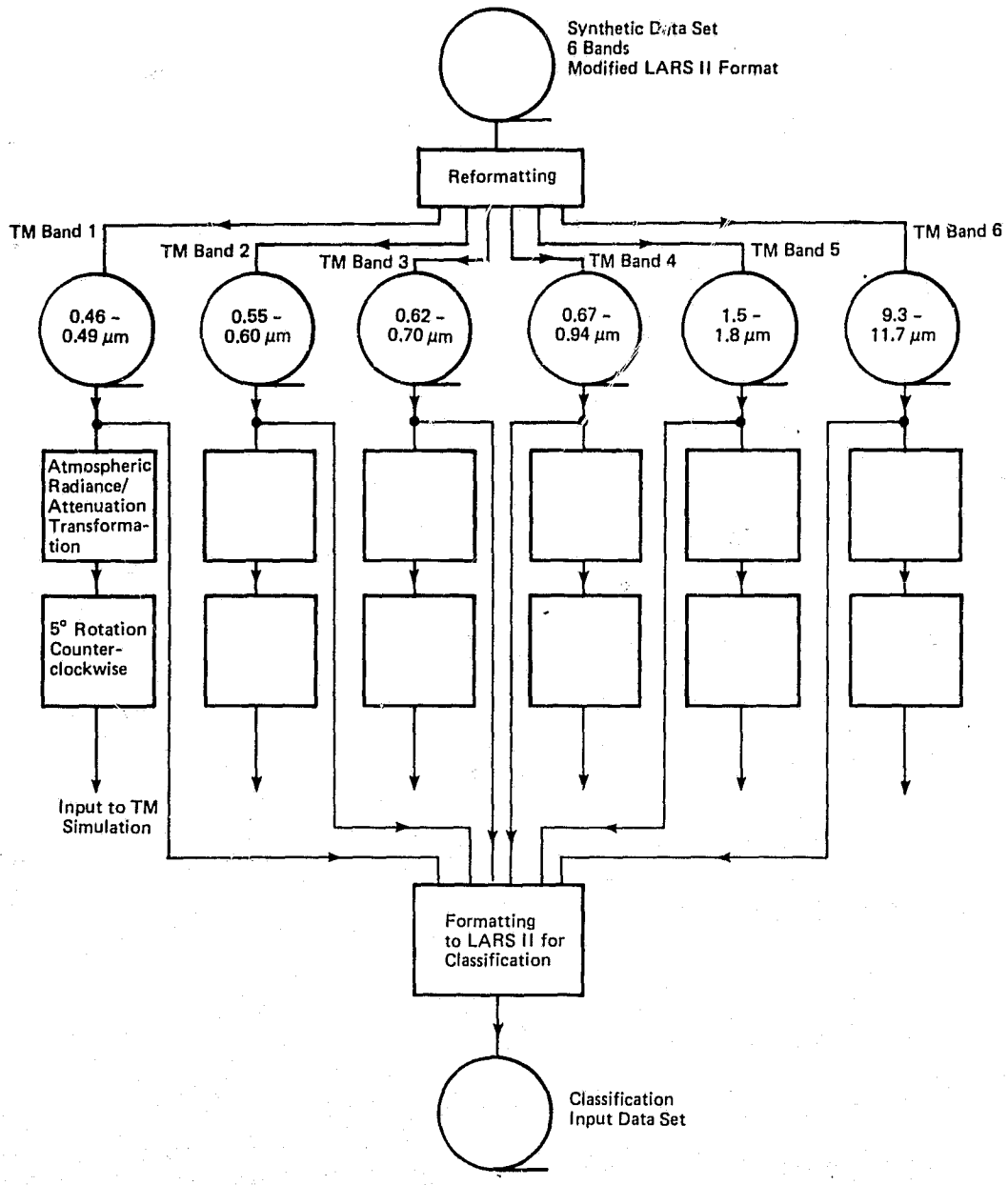


Figure 3.2-1. Synthetic Data Set Preprocessing

Because the statistical characterization used to generate the synthetic data set had been derived from the aircraft data set, the atmospheric transformation to be applied to it was in principal identical to that used for the aircraft data set. However, in order to produce 8-bit samples for the synthetic data set, the synthesized samples had been reduced in magnitude by a factor of two before being recorded in the data set. The atmospheric transformation was therefore modified to compensate for this scaling. The transformations employed for the spectral bands are given in Table 3.2-1.

The final preprocessing step was the counterclockwise rotation of the data set by 5° to destroy the parallelism of the field boundaries with the sample and line directions of the data set. The rotation was performed by nearest neighbor resampling. The radiometrically transformed and rotated data set is shown in Figure 3.2-2.

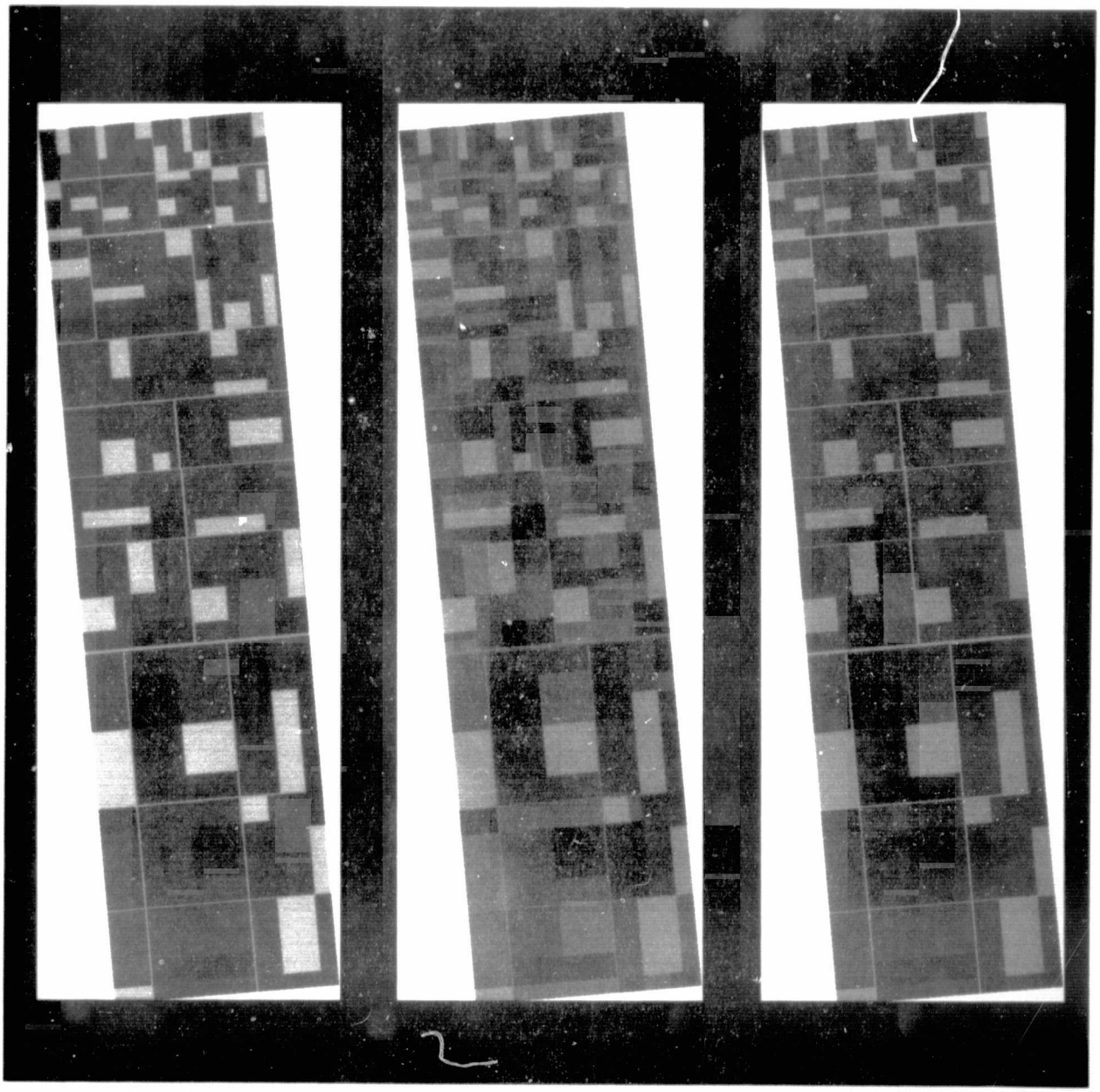
ORIGINAL PAGE IS
OF POOR QUALITY

Table 3.2-1. Synthetic Data Set Atmospheric Transformation

Synthetic Data Set Channel	TM Band	Limits of C_8' Min	Limits of C_8' Max	$C_8 = A_i + B_i C_9$ A_i	$C_8 = A_i + B_i C_9$ B_i	$C_8 = A_i' + B_i' C_8'$ A_i'	$C_8 = A_i' + B_i' C_8'$ B_i'
1	1	45	129	6.0246	0.6584	6.0246	1.3168
2	2	49	143	13.2721	0.4536	13.2721	9.9720
3	3	22	117	13.0146	0.4387	13.0146	0.8774
4	4	63	251	9.0401	1.0194	-62.06664	1.2238
5	5	16	155	14.3860	0.8549	-17.3568	1.7098
6	6	111	205	59.1120	0.4029	59.1102	0.8058

(N. B. Data in synthetic data set is reduced by factor of 2 with respect to ERIM radiances. Therefore, B_i correction coefficients should be multiplied by 2.)

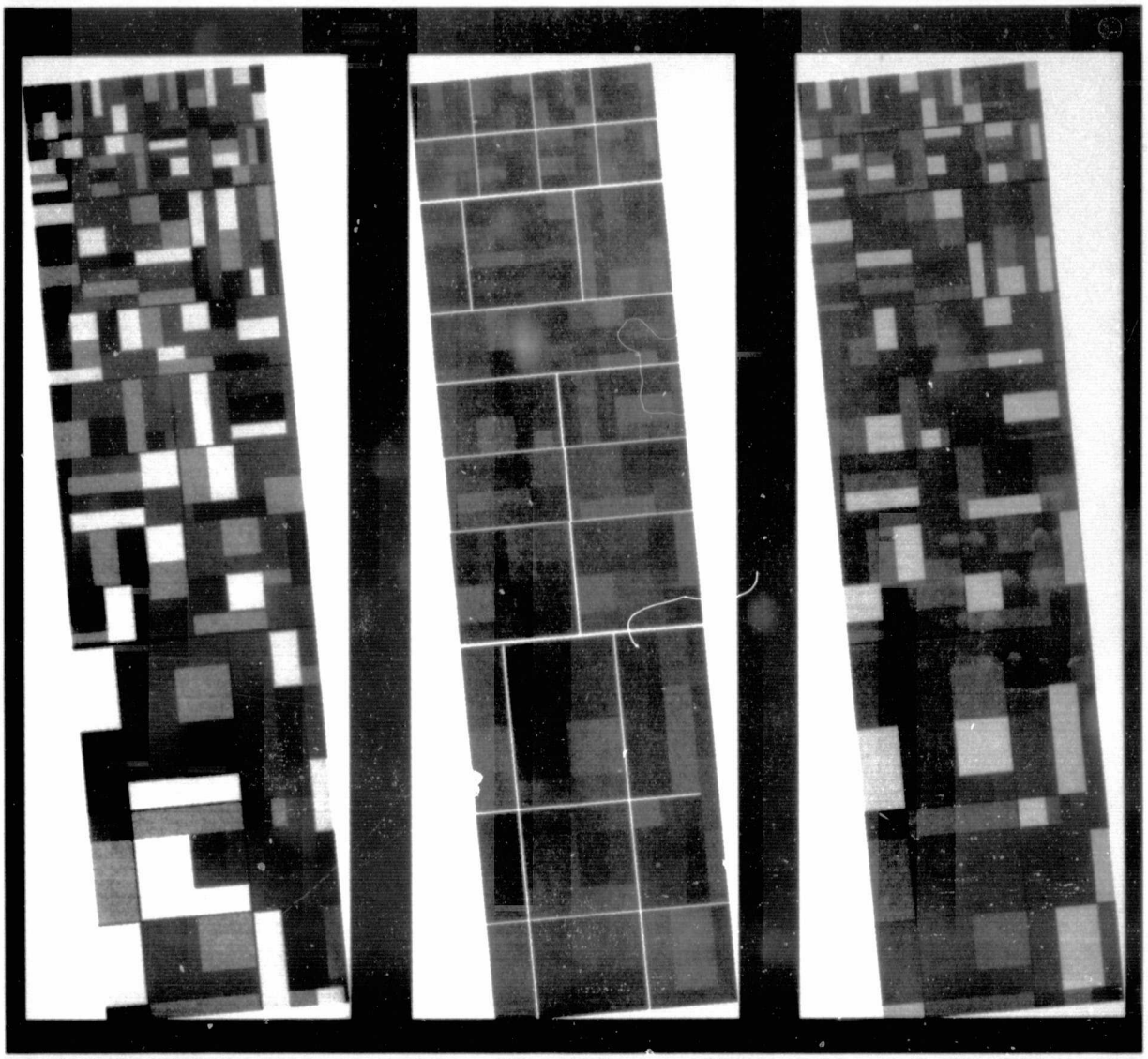
ORIGINAL PAGE IS
OF POOR QUALITY



TM BAND 1 TM BAND 2 TM BAND 3

Figure 3.2-2. Rotated Synthetic Data Set

ORIGINAL PAGE IS
OF POOR QUALITY



TM BAND 4 TM BAND 5 TM BAND 6

Figure 3.2-2. Rotated Synthetic Data Set (Cont.)

Section 4

THEMATIC MAPPER DATA SET SIMULATION

4.1 SIMULATION PROCESSING

The data sets produced as a result of the preprocessing described in Section 3 provided the input for the simulation of Thematic Mapper data. The processing flow for a single spectral image is shown in Figure 4.1-1. In this flow, the input and output data sets contain 8-bit image samples; the intermediate data sets and the simulation programs employ floating point sample representation to maintain the precision of the overall simulation. Three configurations of the Thematic Mapper were simulated in this study. Unique aperture function models were used for each spectral band, but were the same set for each configuration. The noise models were also unique for each band but identical across the three configurations. The configurations differed in the presampling filter and sampling characteristics modeled. These are summarized in Table 4.1-1. Details of the simulation are discussed subsequently.

4.1.1 Aperture Convolution

Processing to simulate the scanning performed by the Thematic Mapper aperture was performed in the Fourier domain. The need to model signal dependent noise in the system made it desirable to perform the rest of the simulation processing in the spatial domain.

To simulate the optical effects of the sensor aperture, the aperture model accepted an operator of linear dimension three times that of the nominal 30 meter (120 meters in the case of the thermal band) instantaneous field of view. This permitted the use of weighting coefficients to describe an imperfect IFOV. The use of such an extended IFOV representation, with a non-uniform spatial characterization, and the overlap of such representations vertically down the image, made Fourier domain processing desirable. For each scan of the Thematic Mapper aperature, the processing which implemented simulated scanning convolved along-scan cross-sections of the aperture function with the corresponding lines of the aircraft scanner data, and summed the corresponding samples of each of these convolved aircraft scanner lines to produce an aperture convolved Thematic Mapper line, sampled along-scan at the same interval as the original aircraft data (see Figure 4.1-2). This computation was performed by Fourier transforming the required aircraft scanner lines, multiplying each of these by the Fourier transform of the corresponding "slice" of the aperture function, applying an operator to combine the corresponding cross-scan samples into samples of the Fourier transform of the aperture convolved Thematic Mapper line, and taking the inverse transform to produce the spatial representation of the line.

Along-scan and cross-scan cross-sections of the aperture function employed for Thematic Mapper band 1 are given in Figure 4.1-3. Only half of the function is

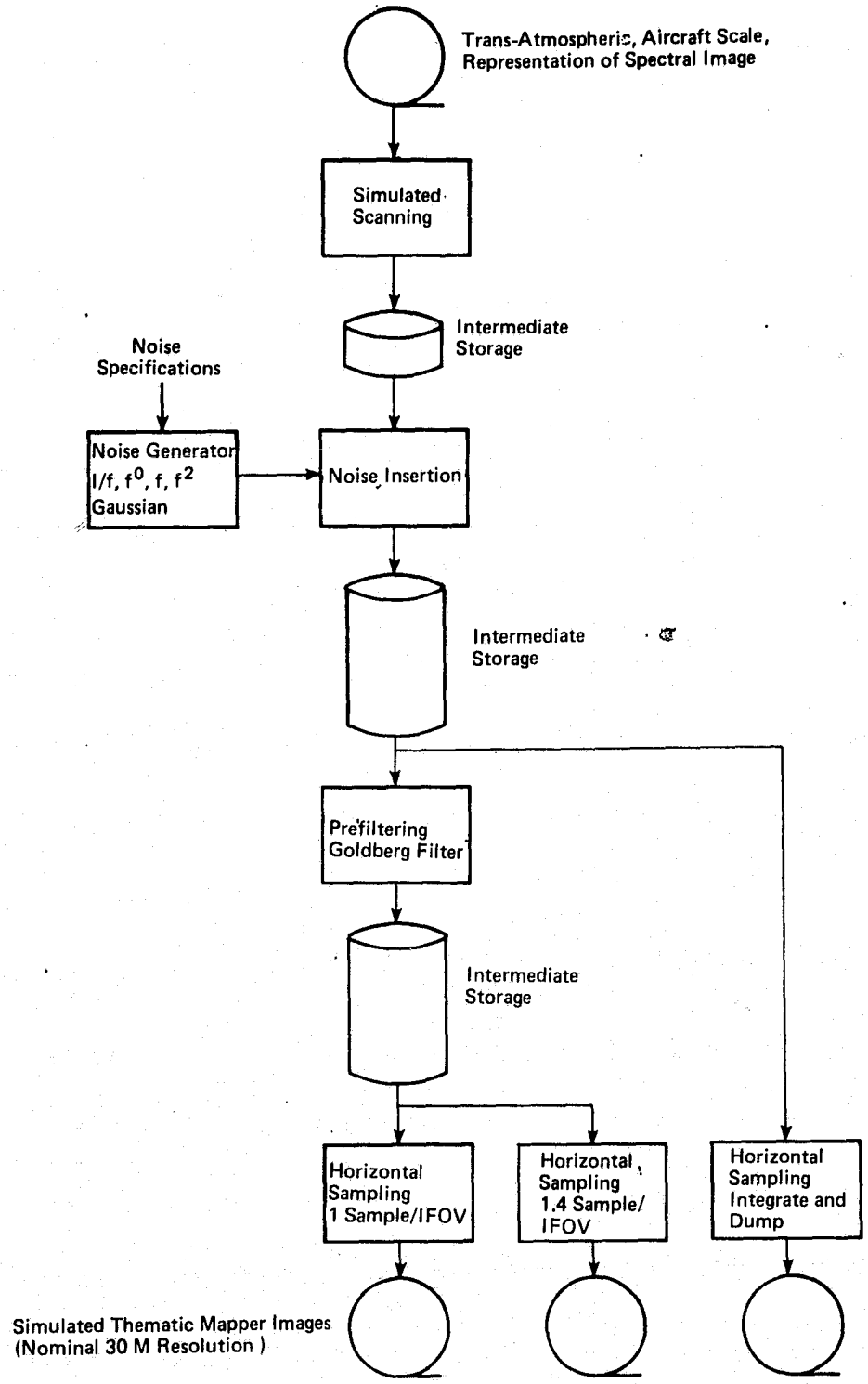


Figure 4.1-1. Simulation Processing Flow

ORIGINAL PAGE IS OF POOR QUALITY

Table 4.1-1. Thematic Mapper Configurations Simulated

Configuration	TM Band	Nominal IFOV	Dwell Time	Filter and Cutoff	Samples
1	1	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.0 Samples/IFOV
	2	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.0 Samples/IFOV
	3	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.0 Samples/IFOV
	4	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.0 Samples/IFOV
	5	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.0 Samples/IFOV
	6	120 m	39.49μs	Goldberg 12.66 khz	Sample-and-Hold 1.0 Samples/IFOV
2	1	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.4 Samples/IFOV
	2	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.4 Samples/IFOV
	3	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.4 Samples/IFOV
	4	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.4 Samples/IFOV
	5	30 m	9.226μs	Goldberg 54.19 khz	Sample-and-Hold 1.4 Samples/IFOV
	6	120 m	39.49μs	Goldberg 12.66 khz	Sample-and-Hold 1.4 Samples/IFOV

Table 4.1-1. Thematic Mapper Configurations Simulated (Cont.)

Configuration	TM Band	Nominal IFOV	Dwell Time	Filter and Cutoff	Samples
3	1	30 m	17.29 μs	None	Integrate-and-Dump 1.0 Samples/IFOV Integration Time = 13.5 μs
	2	30 m	17.29 μs	None	Integrate-and-dump 1.0 Samples/IFOV Integration Time = 13.5 μs
	3	30 m	17.29 μs	None	Integrate-and-Dump, 1.0 Samples/IFOV Integration Time = 13.5 μs
	4	30 m	17.29 μs	None	Integrate-and-Dump 1.0 Samples/IFOV Integration Time = 13.5 μs
	5	30 m	17.29 μs	None	Integrate-and-Dump, 1.0 Samples/IFOV Integration Time = 13.5 μs
	6	120 m	69.07 μs	Goldberg 7.24 khz	Sample-and-Hold 1.0 Samples/IFOV

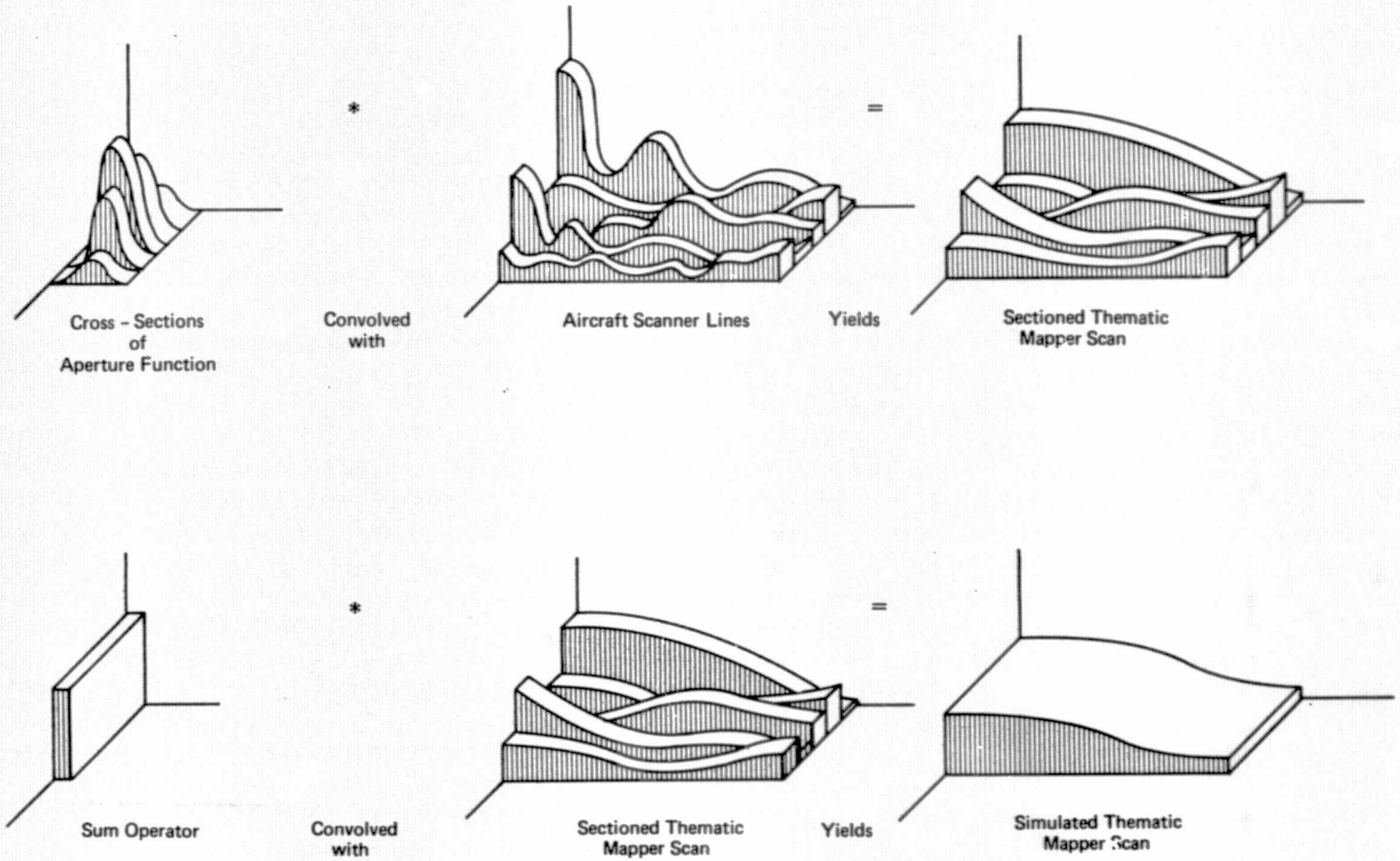
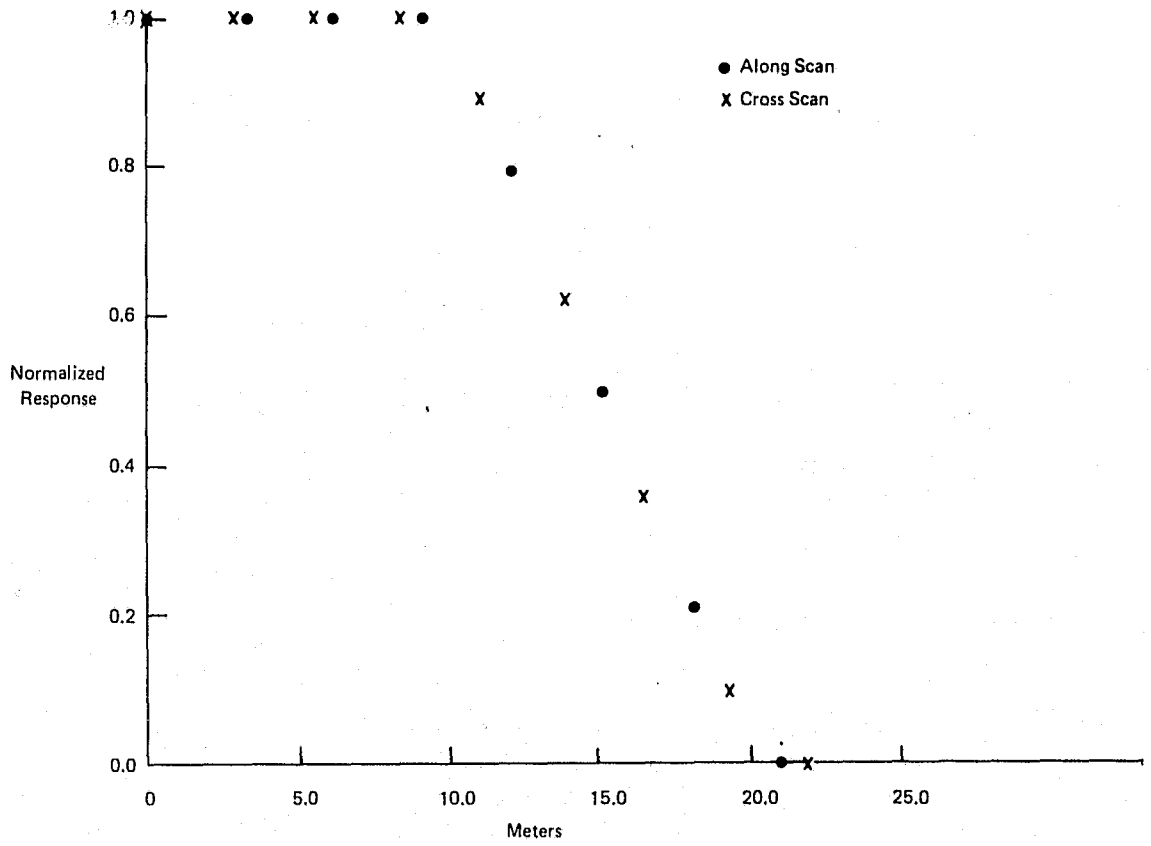


Figure 4.1-2. Simulation of Thematic Mapper Scan Processing

ORIGINAL PAGE IS OF POOR QUALITY



Function 4.1-3. Aperture Function Cross Sections Thematic Mapper Band 1

displayed, since the aperture function possesses inversion symmetry through its centroid and reflection symmetry about the along-scan and cross-scan directions.

4.1.2 Noise Generation and Insertion

The noise models appropriate to this sensor simulation can conveniently be divided into two types: signal dependent noise and signal independent noise. Signal dependent noise (photon limited noise) is encountered in photomultiplier tubes and the advanced solid state detectors such as will be employed in the Thematic Mapper and can be modeled as a function of the square root of scene radiance. Signal independent noise, which arises from various sources, is characterized by a power spectrum which typically exhibits three regions:

a. a low-frequency region, where the power spectrum exhibits 1/f variation with frequency

b. a mid-frequency range where the power spectrum is essentially flat

c. a high-frequency region, where the noise power spectrum rapidly increases in magnitude, typically as a linear and quadratic function of frequency.

To minimize the cost of simulating noise for the processing, a signal independent noise population, tailored to the noise characteristics and levels specified for the Thematic Mapper, was generated and stored for use in the simulation. Samples drawn from this population in a random manner were the immediate source of noise values which were added to the signal produced by convolving the sensor aperture with the trans-atmospheric, aircraft scale representation of the image being processed.

The noise population was generated by constructing a set of unit magnitude, random phase samples in the Fourier domain, shaping the power spectrum of these samples with a weighting function of the form, in bits^2/hertz,

$$G(f) = K_1(1/f) + K_2 + K_3 f + K_4 f^2,$$

and then inverse Fourier transforming the samples into the time domain. The values of the K_i for the various Thematic Mapper bands are given in Table 4.1-2.

Signal dependent noise was modeled by adding to each sensor convolved sample a noise sample obtained by weighting a noise sample drawn from a Gaussian population by the square root of the signal sample value, expressed as a percentage of the peak-to-peak signal. For this noise generation procedure a pre-stored population of Gaussian noise samples was used.

Table 4.1-3 presents the parameters employed to model this signal-dependent noise.

Table 4.1-2. Power Spectrum Weighting Constants

Thematic Mapper Band	K_1	K_2	K_3	K_4
1	0.0	2.080×10^{-5}	3.289×10^{-10}	2.603×10^{-14}
2	0.0	1.423×10^{-6}	2.247×10^{-11}	1.779×10^{-15}
3	0.0	3.107×10^{-6}	4.906×10^{-11}	3.884×10^{-15}
4	0.0	4.234×10^{-7}	6.687×10^{-12}	5.293×10^{-16}
5	0.0	1.770×10^{-5}	1.845×10^{-10}	3.430×10^{-14}
6*	2.032×10^{-2}	4.064×10^{-5}	0.0	0.0

$$G(f) = \frac{K_1}{f} + K_2 + K_3 f + K_4 f^2 \qquad \text{bits}^2/\text{hertz}$$

*Noise = 0 for $f < 16$ hertz

Table 4.1-3. Signal Dependent Noise Parameters

Thematic Mapper Band	σ (Bits)	ψ_o	$\Delta\psi$
1	1.95	20	220
2	0.99	20	220
3	1.20	20	220
4	0.75	20	220
5	1.26	20	220
6	0.00	--	--

ψ = Signal sample value

ψ_o = Minimum signal value

$\Delta\psi$ = Peak-to-peak signal

n = Noise sample value

$G(\sigma)$ = Sample of Gaussian population with variance σ^2

$$n = G(\sigma)\left[\frac{\psi - \psi_o}{\Delta\psi}\right]^{\frac{1}{2}} \text{ bits}$$

4.1.3 Filtering

Two of the Thematic Mapper configurations simulated in this study employed a pre-sampling filter. This was a Goldberg filter, whose poles are at

$$s = -0.8198,\ -0600 \pm j1.028$$

The analog transfer function for this filter

$$H(s) = \frac{1.1615}{(s + 0.8198)\ (s + 0.600 + j1.028)\ (s + 0.600 - j1.028)}$$

was transformed, using the bilinear z transform, into a digital filter

$$H(z) = (1 + Z^{-1}) \sum_{j=1}^{2} H_j(Z)$$

where

$$H_j(Z) = \frac{A_{1j}(1+Z^{-1}) + A_{2j}(1-Z^{-1})}{A_{3j}Z^{-2} + A_{4j}Z^{-1} + 1}$$

For this parallel decomposition of the filter transfer function, the simulation performed a calculation of the form:

$$O_j'(k) = A_{1j}\left[i(k) + i(k-1)\right] + A_{2j}\left[i(k) - i(k-1)\right] - A_{3j}O_j'(k-2) - A_{4j}O_j'(k-1)$$

$$O(k) = \sum_{j=1}^{N} O_j'(k) + O_j'(k-1)$$

where

$O(k)$ is the kth output sample

$O_j'(k)$ is the kth output sample from the jth filter section

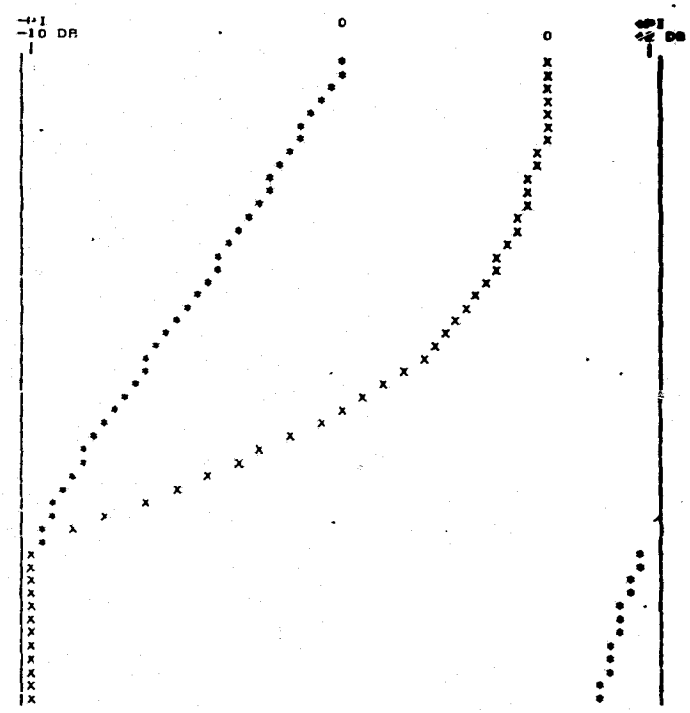
$i(h)$ is the kth input sample

A_{1j} are the filter coefficients

N is the number of filter sections in the parallel decomposition (2).

A frequency magnitude plot of the digital filter employed for Thematic Mapper bands 1 through 5 for configurations 1 and 2 is given in Figure 4.1-4.

FREQUENCY (HERTZ)	MAGNITUDE (DECIBELS)	PHASE (DEGREES)	DELAY (SECS)
0.0	0.0	0.0	0.0
0.22120E+04	-0.5736[illegible]E-02	-4.63	0.6069E-05
0.44241E+04	-0.22612E-01	-9.66	0.6064E-05
0.66361E+04	-0.50884E-01	-14.46	0.6053E-05
0.88481E+04	-0.89396E-01	-19.29	0.6039E-05
0.11060E+05	-0.13767E+00	-24.06	0.6023E-05
0.13272E+05	-0.19484E+00	-28.87	0.6006E-05
0.15484E+05	-0.26009E+00	-33.64	0.5992E-05
0.17696E+05	-0.33256E+00	-38.40	0.5982E-05
0.19908E+05	-0.41151E+00	-43.17	0.5979E-05
0.22120E+05	-0.49657E+00	-47.93	0.5984E-05
0.24332E+05	-0.58731E+00	-52.71	0.6002E-05
0.26544E+05	-0.68377E+00	-57.51	0.6031E-05
0.28756E+05	-0.78644E+00	-62.35	0.6075E-05
0.30968E+05	-0.89616E+00	-67.23	0.6133E-05
0.33180E+05	-0.10143E+01	-72.18	0.6205E-05
0.35392E+05	-0.11428E+01	-77.19	0.6290E-05
0.37604E+05	-0.12840E+01	-82.27	0.6388E-05
0.39816E+05	-0.14408E+01	-87.44	0.6488E-05
0.42029E+05	-0.16166E+01	-92.69	0.6591E-05
0.44241E+05	-0.18150E+01	-98.02	0.6692E-05
0.46453E+05	-0.20391E+01	-103.42	0.6781E-05
0.48665E+05	-0.22945E+01	-108.87	0.6851E-05
0.50877E+05	-0.25825E+01	-114.36	0.6894E-05
0.53089E+05	-0.29063E+01	-119.86	0.6902E-05
0.55301E+05	-0.32678E+01	-125.33	0.6875E-05
0.57513E+05	-0.36675E+01	-130.75	0.6799E-05
0.59725E+05	-0.41050E+01	-136.07	0.6663E-05
0.61937E+05	-0.45768E+01	-141.26	0.6524E-05
0.64149E+05	-0.50656E+01	-146.26	0.6329E-05
0.66361E+05	-0.55226E+01	-151.16	0.6102E-05
0.68573E+05	-0.61151E+01	-155.82	0.5852E-05
0.70785E+05	-0.67708E+01	-160.27	0.5586E-05
0.72997E+05	-0.72734E+01	-164.56	0.5310E-05
0.75209E+05	-0.76598E+01	-168.51	0.5033E-05
0.77421E+05	-0.86161E+01	-172.30	0.4759E-05
0.79633E+05	-0.92469E+01	-175.87	0.4490E-05
0.81845E+05	-0.98652E+01	-179.24	0.4232E-05
0.84057E+05	-0.10523E+02	177.56	-.4481E-05
0.86269E+05	-0.11159E+02	174.59	0.3754E-05
0.88481E+05	-0.11793E+02	171.78	0.3536E-05
0.90693E+05	-0.12422E+02	169.12	0.3331E-05
0.92905E+05	-0.13047E+02	166.62	0.3141E-05
0.95117E+05	-0.13665E+02	164.26	0.2964E-05
0.97329E+05	-0.14276E+02	162.05	0.2798E-05
0.99541E+05	-0.14865E+02	159.95	0.2640E-05
0.10175E+06	-0.15470E+02	157.93	0.2504E-05
0.10397E+06	-0.16065E+02	156.04	0.2370E-05
0.10618E+06	-0.16645E+02	154.25	0.2250E-05
0.10839E+06	-0.17218E+02	152.55	0.2137E-05



$$H(Z) = (1+Z^{-1}) \sum_{j=1}^{2} H_j(Z)$$

$$H_j(Z) = [A_{1j}(1+Z^{-1}) + A_{2j}(1-Z^{-1})] / [A_{3j}Z^{-2} + A_{4j}Z^{-1} + 1]$$

Section j	A_{1j}	A_{2j}	A_{3j}	A_{4j}
1	0.731242×10^{-1}	0.731242×10^{-1}	0.0	-0.771852×10^{0}
2	-0.805765×10^{-2}	-0.134932×10^{0}	0.691884×10^{0}	-0.157761×10^{1}

Figure 4.1-4. Goldberg Filter

ORIGINAL PAGE IS
OF POOR QUALITY

4.1.4 Sampling

To produce simulated Thematic Mapper imagery with the desired along-scan sampling characteristics (see Table 4.1-1), the simulated scan lines were processed through a sampling routine. For the "sample and hold," this sampling was done by selecting samples at the appropriate interval from the output signal of the filter. For the "integrate and dump," the sampling was performed by averaging over the integration time the oversampled discrete values at the output of the noise simulation step.

4.2 RESAMPLING

In order to evaluate the effects which different resampling procedures for digital imagery may have on remotely sensed data, three possible resampling procedures were employed to transform the simulated Thematic Mapper data onto a sampling lattice which differed from the original sampling lattice by a small non-integral translation (15 meters in each axis), and a small clockwise rotation of 1.0°. The three resampling procedures employed were nearest neighbor assignment, cubic convolution and point spread function compensation. The resampled imagery was produced at 1 sample/IFOV both along-scan and cross-scan, except for band 6, where the resampling produced resampled imagery at 4 samples/ IFOV in each direction in order to provide band 6 samples on a lattice congruent to that of the other 5 bands. After resampling, the six bands of each of the nine resampled simulated Thematic Mapper datas sets derived from each input data set, aircraft and synthetic, were formatted into LARS-II for evaluation by multispectral classification.

4.2.1 Nearest Neighbor Assignment

As the name implies, this resampling technique involves establishing the geometric relationship between the input and output sampling lattices, and then assigning as values for the samples of the output lattice the values of the closest input samples. This procedure is diagrammed in Figure 4.2-1.

4.2.2 Cubic Convolution

Cubic convolution is a four-point interpolating function developed to approximate the ideal resampling function, sinc(x). This function, together with an outline of its use for interpolating digital imagery, is presented in Figure 4.2-2.

For each output value, the one-dimensional algorithm is used to interpolate along the four nearest horizontal input lines to points defined by a vertical line through the output point (the points marked "x" in Figure 4.2-2). The one-dimensional algorithm is applied again along this vertical line to produce the required output value.

4.2.3 Point Spread Function Compensation

Point spread function compensation is implemented using a bivariate, anisotropic Gaussian approximation to characterize the point spread function of the Thematic Mapper sensor system and a polynomial approximation to the Thematic Mapper data in a neighborhood of the desired resampled point. The development of this resampling procedure is presented in Reference 2. The process is outlined in Figure 4.2-3.

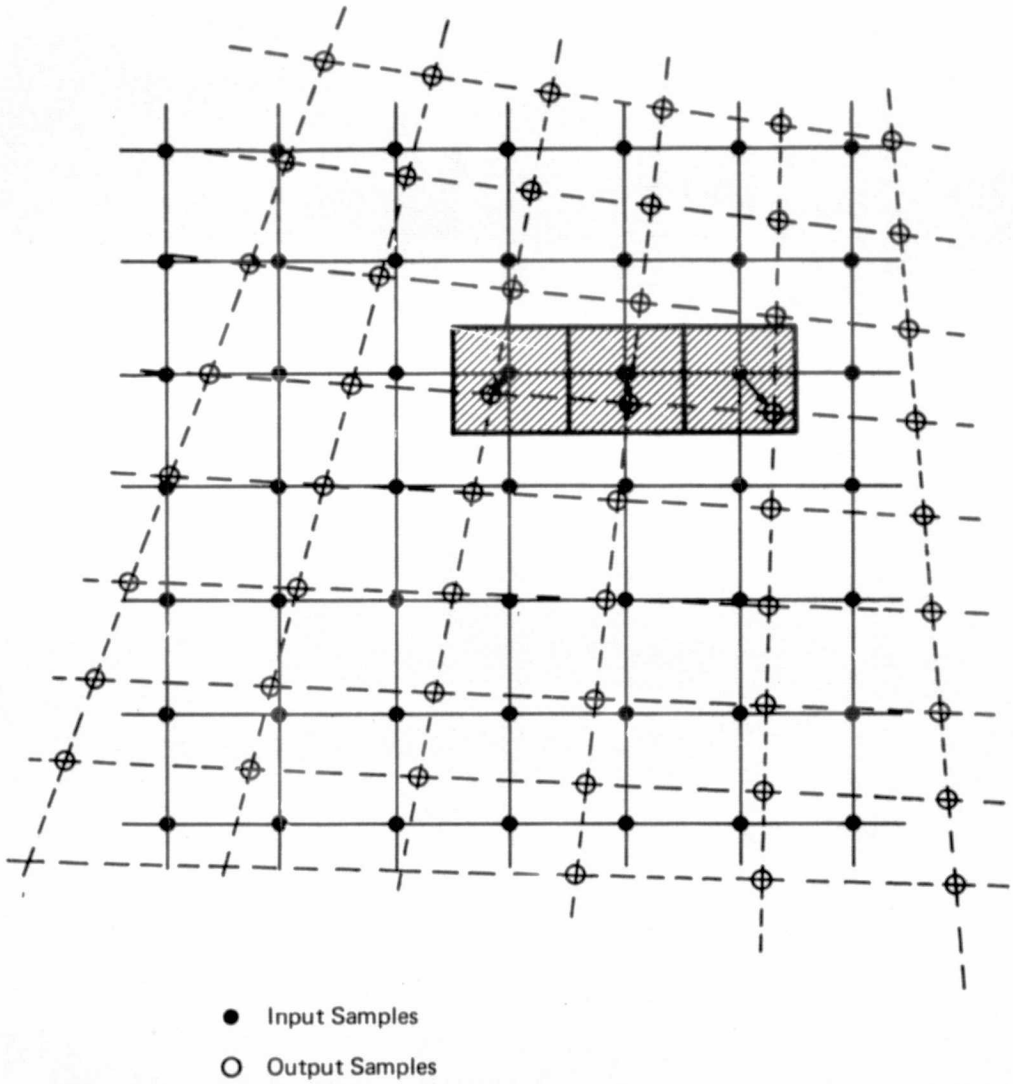
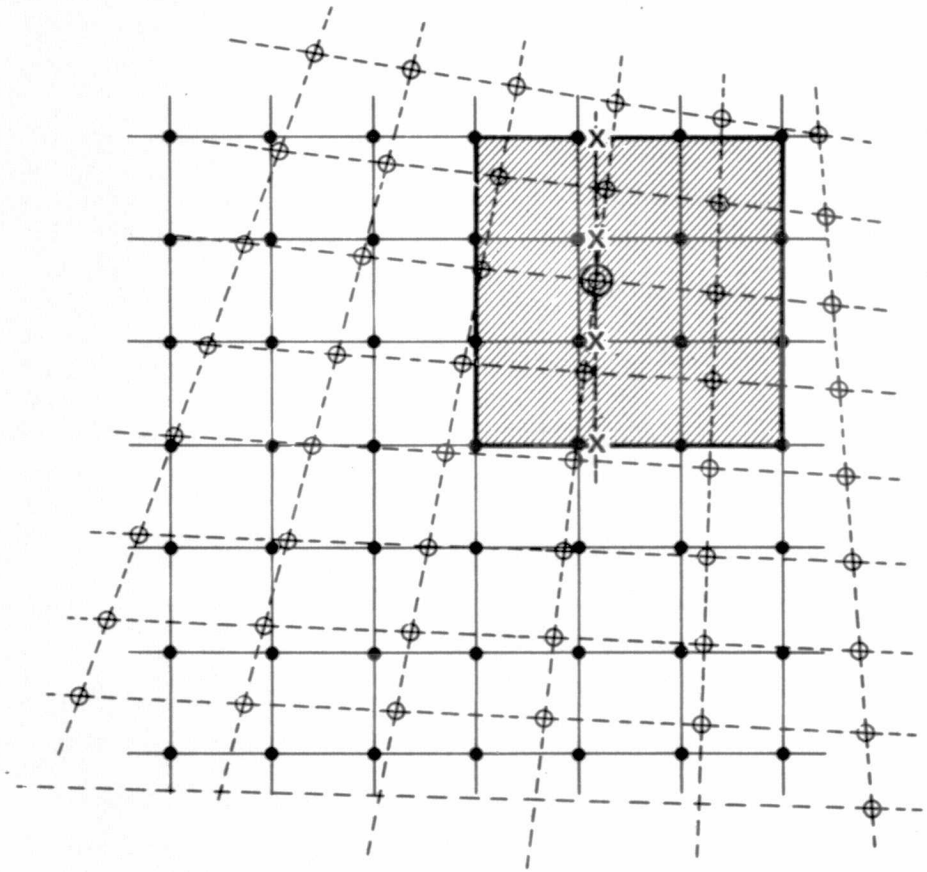
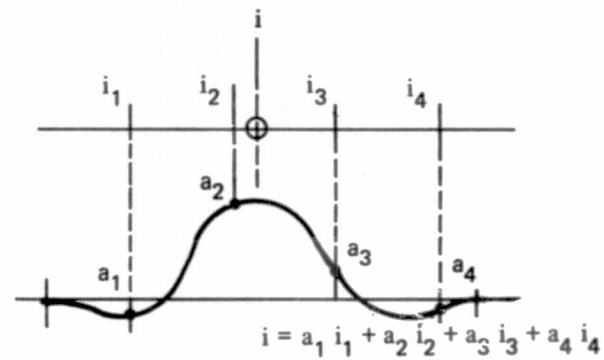


Figure 4.2-1. Nearest Neighbor Assignment





Cubic Convolution Polynomial

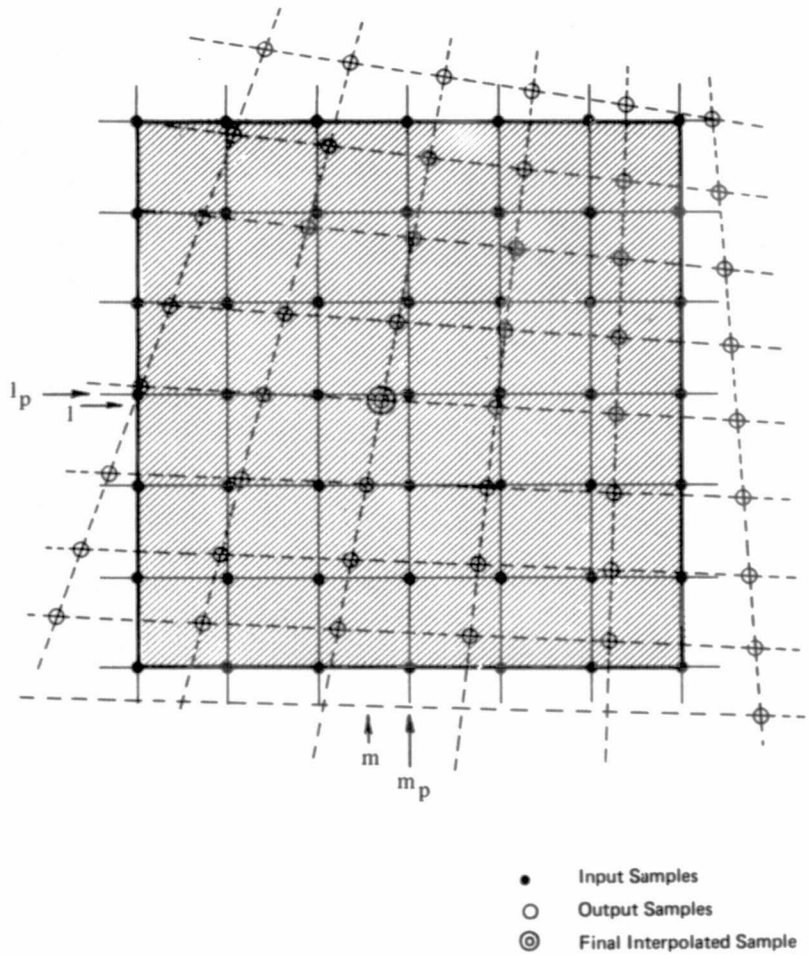
$$f(x) = 1 - 2|x|^2 + |x|^3 \qquad 0 \leq |x| \leq 1$$

$$= 4 - 8|x| + |x|^2 - |x|^3 \qquad 1 < |x| \leq 2$$

$$= 0 \qquad < |x|$$

- ● Input Samples
- ○ Output Samples
- X Intermediate Interpolated Samples
- ◎ Final Interpolated Sample

Figure 4.2-2. Cubic Convolution Resampling



$$i = \hat{i}(l, m) - \frac{\alpha\sigma^2}{2} \nabla_\alpha^2 \hat{i}(l, m) + \left(\frac{\alpha\sigma^2}{2}\right)^2 \left(\frac{1}{2}\right) \nabla_\alpha^4 \hat{i}(l, m)$$

where $\hat{i}(l, m) = \sum_{i=0}^{4} \sum_{j=0}^{4} \hat{a}_{ij} (l - l_p)^i (m - m_p)^j$

$(i + j) = 0, 1, 2, 3, 4$

$\hat{a} = M F$ = coefficient vector

F = vector of 7x7 input sample array

M = constant matrix

Figure 4.2-3. Point Spread Function Compensation Resampling

ORIGINAL PAGE IS
OF POOR QUALITY

In this resampling procedure, the compensated, resampled radiance value, i, is given by

$$i = \hat{i}(l,m) - \left(\frac{\alpha\sigma^2}{2}\right)\Delta\alpha^2\,\hat{i}(l,m) + \left(\frac{\alpha\sigma^2}{2}\right)^2\left(\frac{1}{2}\right)\Delta_\alpha^4\, i(l,m)$$

where $\alpha\sigma$ and σ are the widths of the point spread function (cross-track and along-track, respectively) and $\hat{i}(l,m)$ is a noise-smoothed polynomial approximation to the original Thematic Mapper data in the neighborhood of (l,m). The function $\hat{i}(l,m)$ is given by

$$\hat{i}(l,m) = \sum_{i=0}^{4}\sum_{j=0}^{4} \hat{a}_{ij}(l-l_p)^i\,(m-m_p)^j$$

where (l_p,m_p) is the center of the approximation neighborhood and is obtained by a least squares fit to a 7x7 array of Thematic Mapper samples centered on (l_p,m_p). The Laplacian $(\Delta\alpha^2)$ and biharmonic (Δ_α^4) operators are simple functions of the a_{ij}.

For this resampling procedure, it was necessary to generate the required Gaussian approximation to the Thematic Mapper point spread function. In the configurations employed in this study, there were twelve unique point spread functions, corresponding to the six aperture functions of the six spectral bands and the two different sampling schemes. The parameters of the twelve Gaussians were obtained by performing a least-squares fit of a Gaussian function to the impulse response of the Thematic Mapper configurations. The impulse response was obtained by processing through the simulation software a unit impulse in a field of zeros constructed on the sampling lattice of the aircraft data set. The scanning and sampling specifications in the simulation were set to produce an output image sampled on the sampling lattice of the aircraft data set, thus producing a finely sampled representation of the impulse response. The parameters of the resulting Gaussians are presented in Table 4.2-1. Typical impulse response cross-sections (unnormalized) are given in Figures 4.2-4 and 4.2-5.

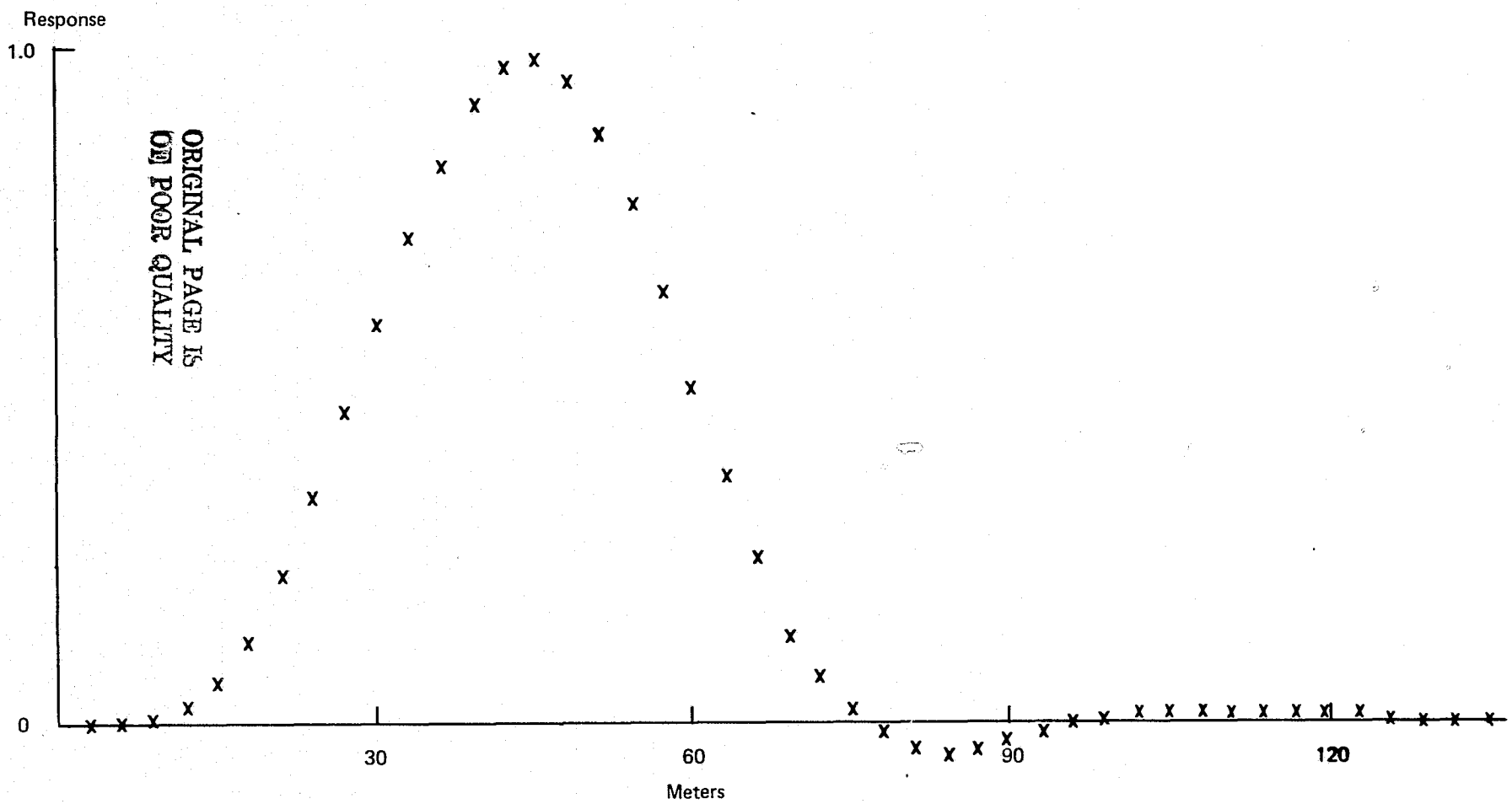
4.3 SIMULATED THEMATIC MAPPER DATA SETS

Figures 4.3-2 through 4.3-7 are photographic recordings of the 18 simulated Thematic Mapper data sets which were produced and submitted for multispectral classification. Figure 4.3-1 provides identification of the data presented in these figures. Only five of the six Thematic Mapper bands are included here, since the 120-meter resolution of band 6, combined with the narrowness of aircraft image employed in the simulation (approximately 1 km), resulted in a simulated image only 8 pixels wide before resampling. After resampling, this width effectively ranged from 2 to 8 120 meter pixels, depending on which resampling technique was employed. For such a narrow image, field definition was impossible, and it was decided that including this data in the classification evaluation would only contribute confusion to the classification results. The classification evaluation thus was performed using only bands 1 through 5.

It should be noted that each of the individual Thematic Mapper images had to be separately adjusted radiometrically in order to produce Figures 4.3-2 through 4.3-7, so that visual comparison of the images in these figures can be misleading. They are provided here only to give an indication of the nature of the data employed in the classifications.

Table 4.2-1. Gaussian Approximation to Point Spread Function

Band	Configuration	Gaussian Parameters α	σ
1	1	0.4536	13.15 meters
	3	0.7481	12.91 meters
2	1	0.4613	13.39 meters
	3	0.7617	13.15 meters
3	1	0.4690	13.65 meters
	3	0.7710	13.37 meters
4	1	0.4905	14.34 meters
	3	0.8013	13.98 meters
5	1	0.4517	12.85 meters
	3	0.6955	12.51 meters
6	1	0.8776	52.64 meters
	3	0.8776	52.64 meters



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 4.2-4. Central Along-Scan Cross-Section of Impulse Response (Normalized) Configuration 1, Band 1

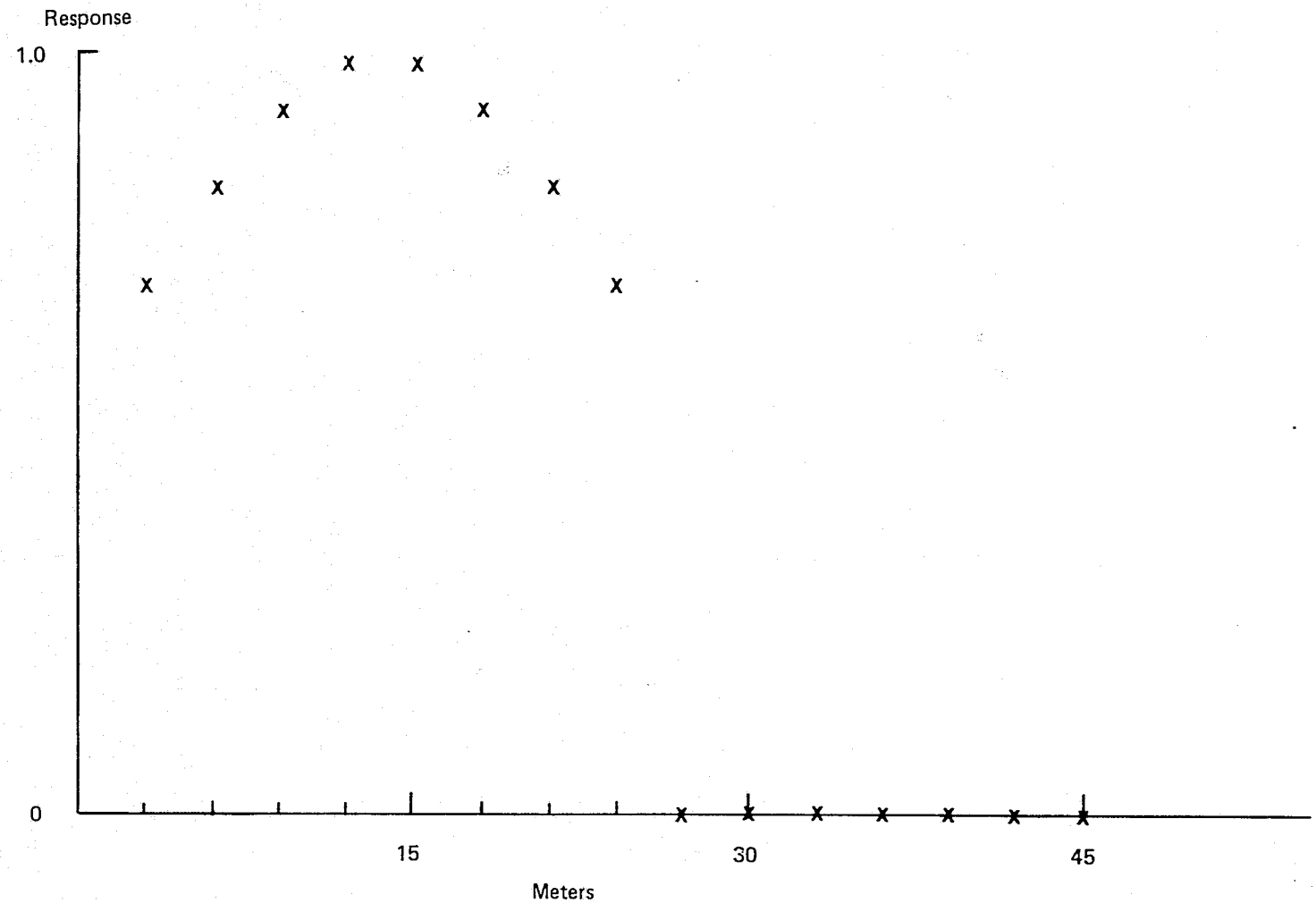
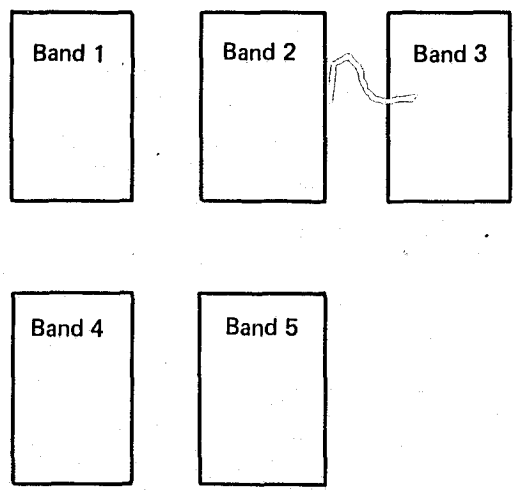


Figure 4.2-5. Central Along-Scan Cross-Section of Impulse Response Configuration 3, Band 1



A = Nearest Neighbor Resampling
B = Cubic Convolution Resampling
C = Point Spread Function Compensation Resampling

Figure 4.3-1. Data Key to Figures 4.3-2 through 4.3-7

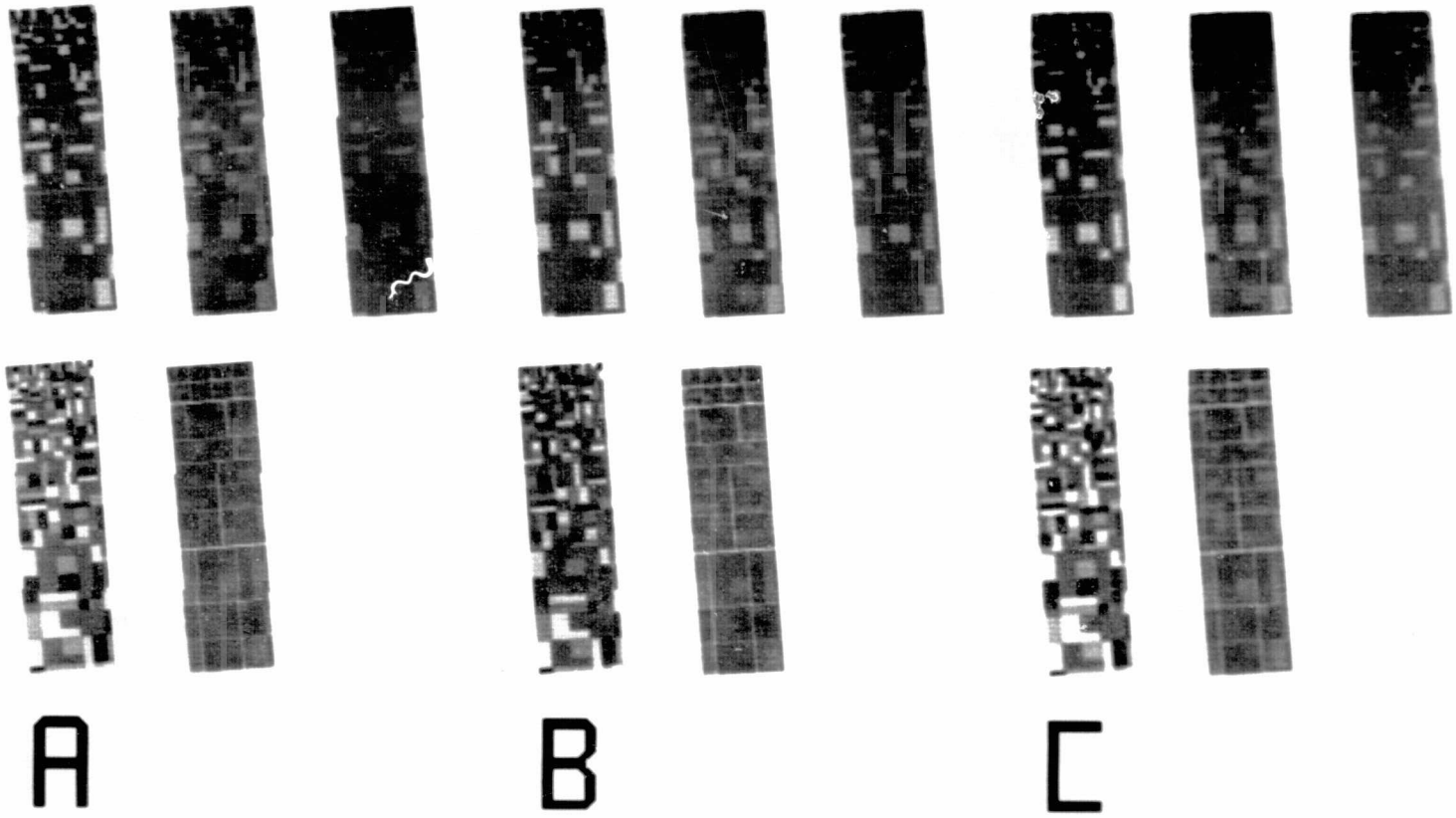


Figure 4.3-2. Simulated Thematic Mapper Data Derived from Aircraft Data Set, Configuration 1

ORIGINAL PAGE IS
OF POOR QUALITY

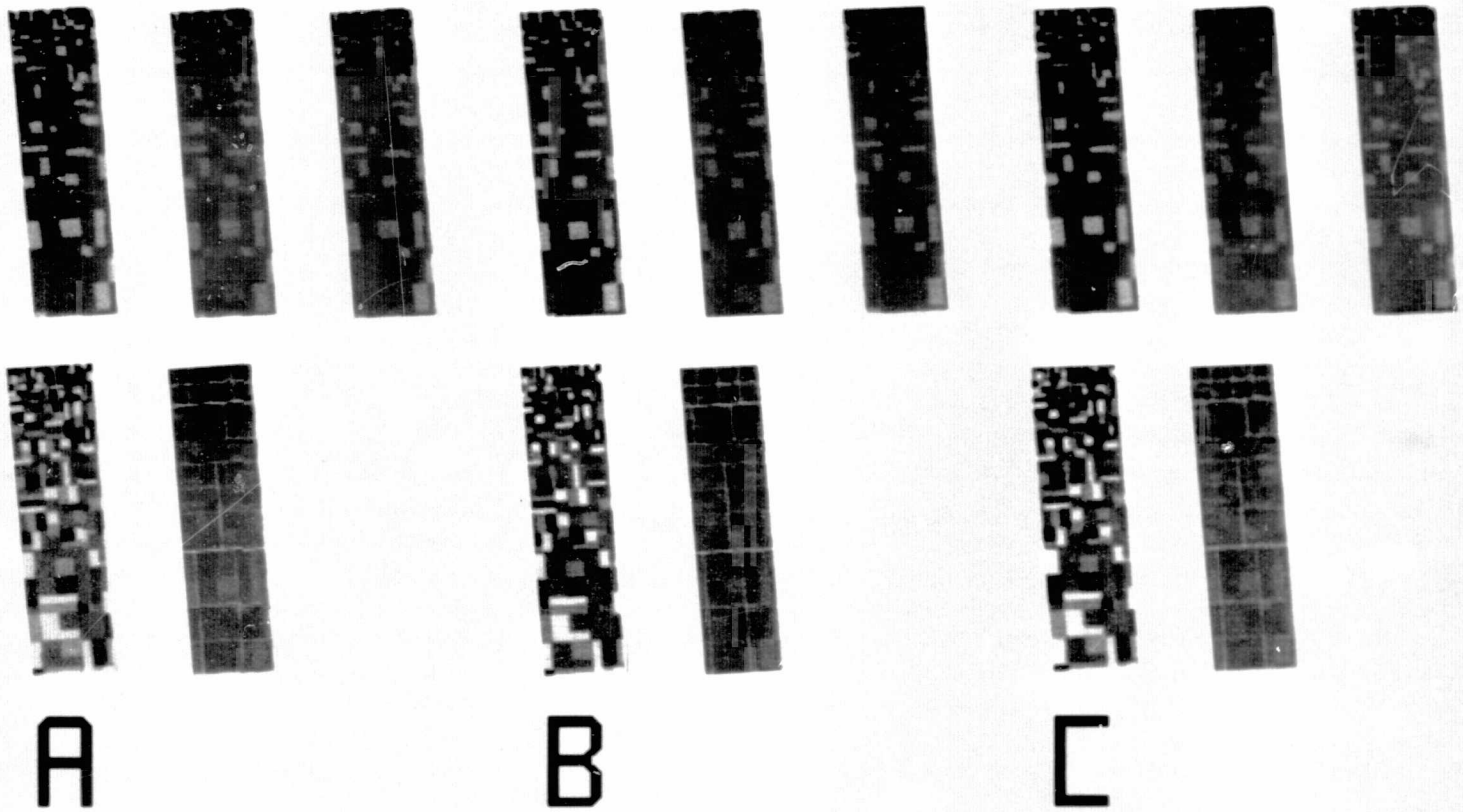


Figure 4.3-3. Simulated Thematic Mapper Data Derived from Aircraft Data Set -- Configuration 2

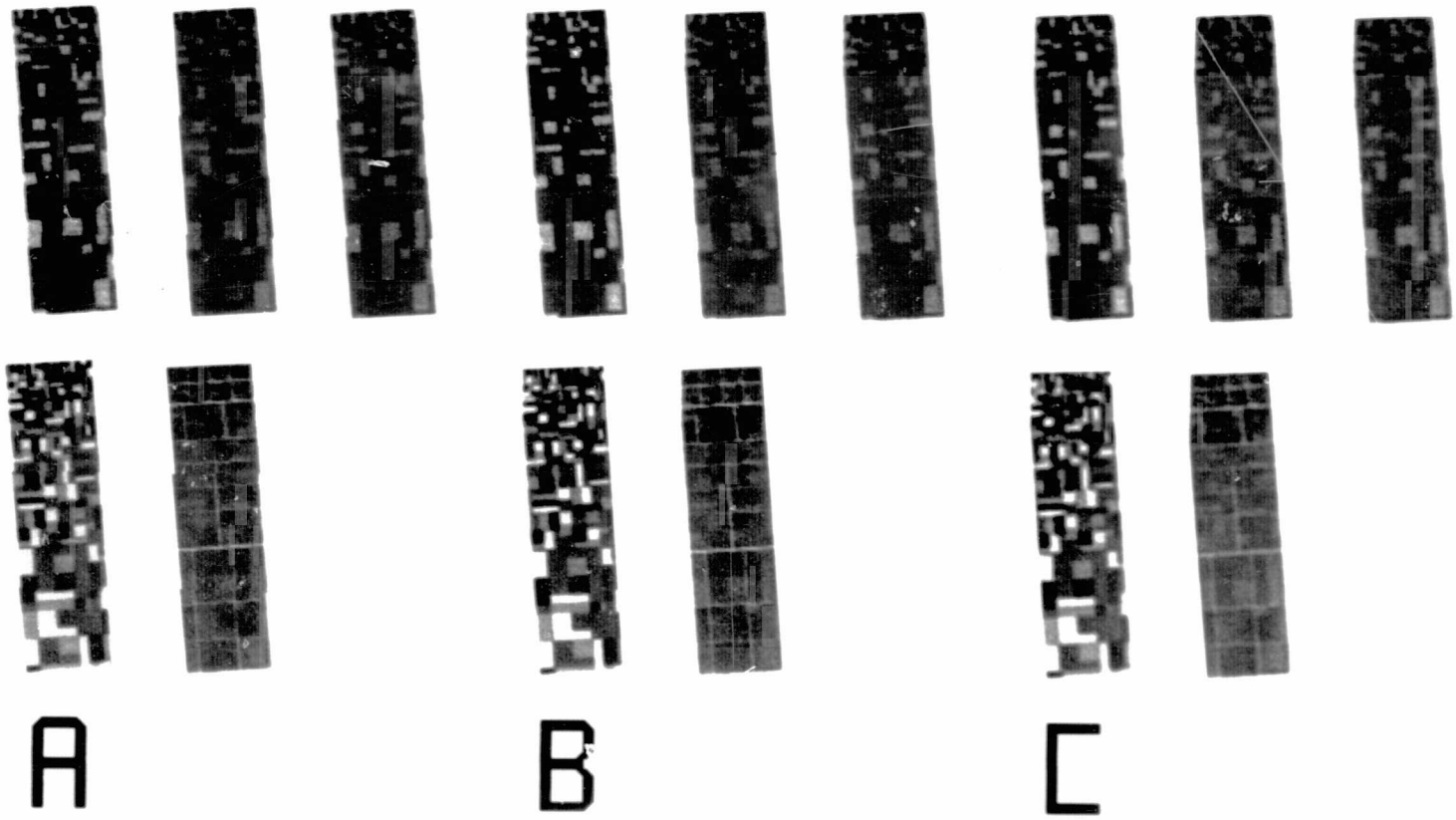


Figure 4.3-4. Simulated Thematic Mapper Data Derived from Aircraft Data Set -- Configuration 3

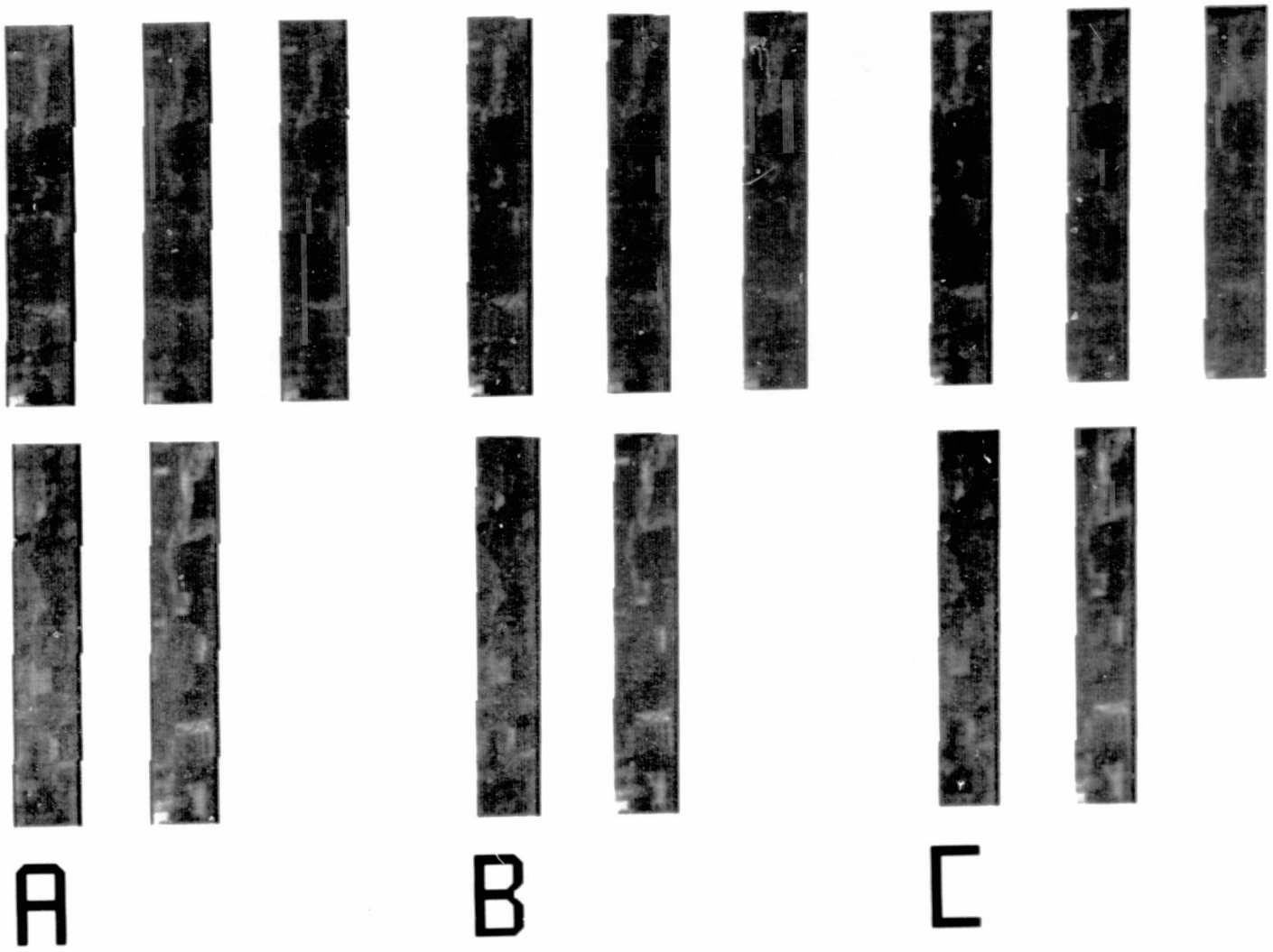


Figure 4.3-5. Simulated Thematic Mapper Data Derived from Synthetic Data Set -- Configuration 1

ORIGINAL PAGE IS
OF POOR QUALITY

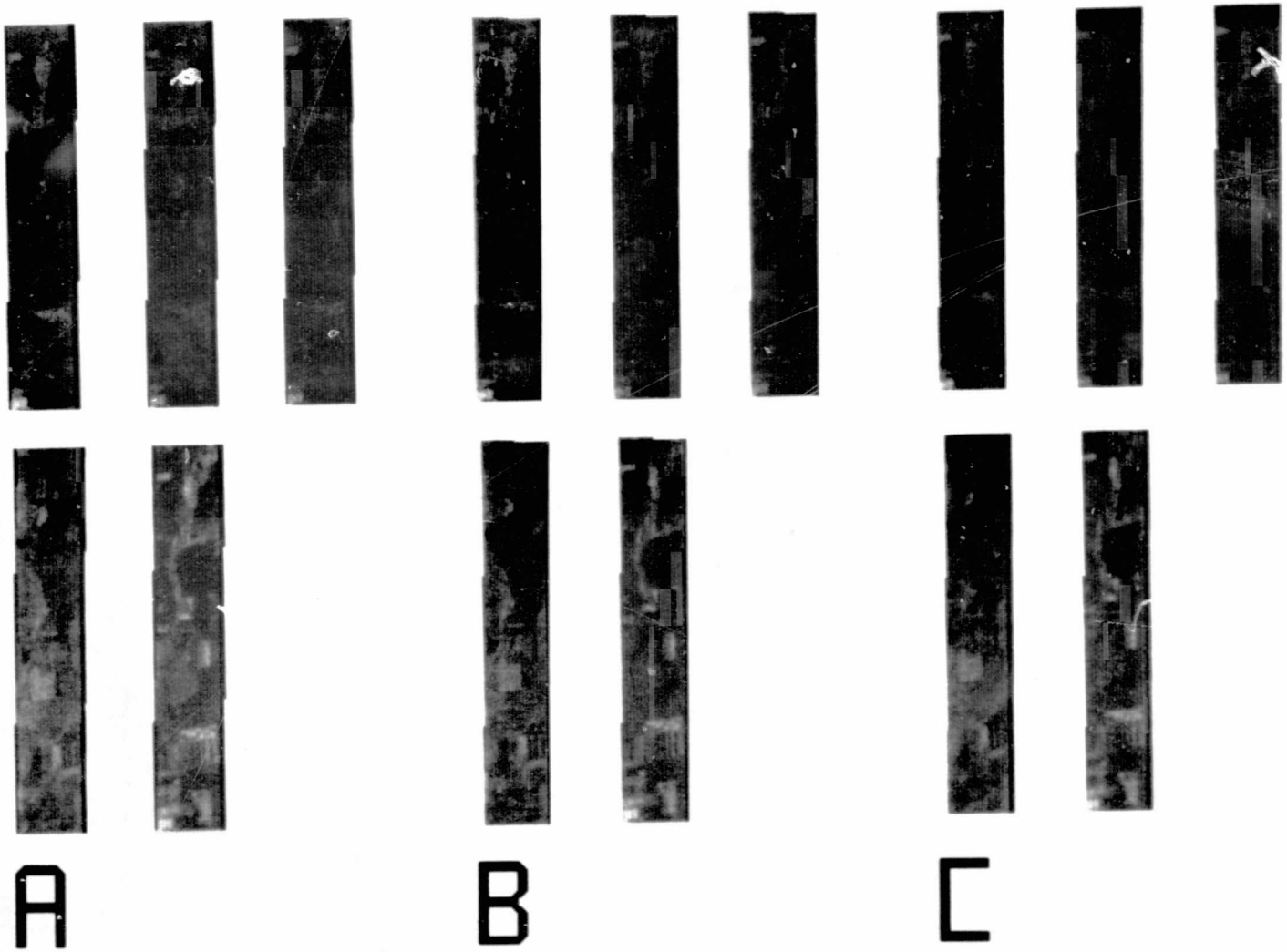


Figure 4.3-6. Simulated Thematic Mapper Data Derived from Synthetic Data Set -- Configuration 2

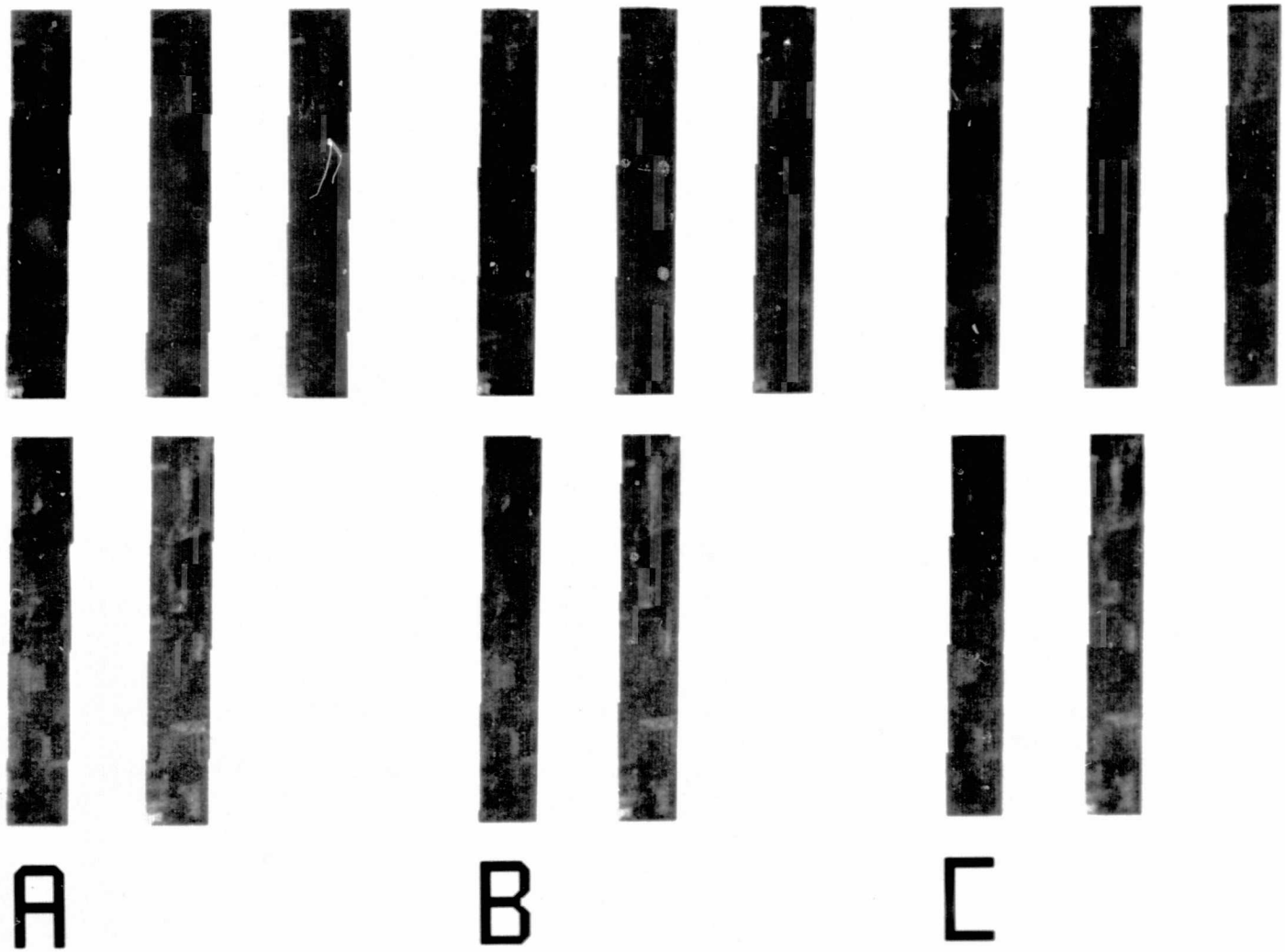


Figure 4.3-7. Simulated Thematic Mapper Data Derived from Synthetic Data Set -- Configuration 3

Section 5

CLASSIFICATION PROCESSING AND ANALYSIS OF RESULTS

5.1 CLASSIFICATION PROCESSING

This section describes the procedures and steps used in classifying the various image data sets. The following terminology is used in this discussion:

a. Aircraft image will refer to the high resolution aircraft image.

b. Synthetic image will refer to the constructed image containing pseudo-gaussian data.

c. Thematic Mapper (TM) image will refer to the simulated data derived from either the aircraft or synthetic image.

d. Configuration 1 is the Thematic Mapper design incorporating a presample filter and sampled at 1.0 samples/IFOV along the scan.

e. Configuration 2 is the Thematic Mapper design incorporating a presample filter and sampled at 1.4 samples/IFOV along the scan.

f. Configuration 3 is the Thematic Mapper design employing an "integrate and dump" sampler.

g. The tables and figures employ the notation NN, CC, and PS for the nearest neighbor, cubic convolution, and point-spread-function-compensation resampling techniques, respectively.

h. The field shape factor is the length to width radio for a field.

5.1.1 Aircraft Image and Related Thematic Mapper Images

The first step was to identify all the fields which could be found in the displayed aircraft image, the aerial photograph of the ground area, and the ground truth listing which accompanied the imagery. At this time various discrepancies among these three sources were noted and resolved.

The vertices of the fields defined in the aircraft image were transformed into field vertices in the various Thematic Mapper images. These became the vertices of the Entire fields in these images. Then, vertices were found which identified field center fields in all images. (Here a field center field is defined as a field containing only field center pixels, where a field center pixel is a pixel which lies completely inside the boundary of the defined field.) These vertices became the vertices of the Field Center fields.

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.1-1 shows the number of fields which were identified for each of the ground truth classes. Except for class corn and possibly trees, there was an insufficient number of fields to conduct an entirely adequate classification experiment. However, these were the only fields available in this image.

Approximately half of the field center fields in each class were chosen for training the classifier, with remaining fields being held as independent test fields. The training fields for each class containing more than one training field were subjected to a cluster analysis to find subclasses of more or less homogeneous pixels. The clustering was done using a version of Johnson Space Center's Iterative Clustering algorithm, a form of Ball and Hall's ISOCLASS algorithm. Reference 4 contains an extensive discussion of this algorithm. The cluster parameters are defined and their values listed in Table 5.1-2. These parameters were chosen to give a "nice" cluster map for each class in the aircraft image. The prime justifications for using them for the Thematic Mapper images were (1) the cluster results are fairly insensitive to the cluster parameters, and (2) there was no acceptable unbiased method for choosing parameters specifically for the Thematic Mapper images. Cluster maps were constructed and displayed for some of the Thematic Mapper images and found to be acceptable. The training fields in the classes containing a single training field were not clustered but were used to define a single subclass for the class.

Statistics (i.e., mean vectors and dispersion matrices) were calculated for each subclass and the image data within the Entire and Field Center portion of the training and test fields were classified using a standard quadratic classifier. (The classifier is the maximum likelihood classifier for gaussian populations with assumed known parameters.) The complete statistics and classification results are available in the computer listings of this processing. NASA/GSFC, IBM, and ERIM have copies of these listings.

5.1.2 Synthetic Image and Associated Thematic Mapper Images

The synthetic image, together with the associated Thematic Mapper images, was handled in much the same way as the aircraft image, with the following exceptions:

a. Two field center fields were chosen for the training for each class.

b. One test field of each size and shape was chosen for each class.

c. The training fields were used to calculate statistics without clustering. This was done because of the basic synthetic image classified virtually without error. Thus, there was no reasonable way to choose clustering parameters.

Table 5.1-3 shows the number of fields in each of the ground truth classes which were present in the synthetic data set.

Table 5.1-1. Identified Fields in Aircraft and Related Images

Class	Number of Fields in Image	Number of Training Fields	Number of Test Fields
Corn	17	8	9
Trees	10	5	5
Pasture	3	2	1
Oats	1	1	0
Soybeans	1	1	0
Mixed Grains	1	1	0
Wheat	3	2	1
Set Aside	6	3	3
Non-farm	4	2	2
Total	46	25	21

Table 5.1-2. Definitions and Values of Cluster Parameters for Clustering Aircraft-derived Thematic Mapper Images

PERCENT	Percent of clusters which must be "stable" in all channels to terminate the initial sequence of splitting iterations (between 0.0 and 1.0).	0.8
SEP	Factor multiplied by standard deviation to give quantity to be added to and subtracted from mean as cluster is split in a certain channel.	1.0
STDMAX	Threshold standard deviation for determining cluster stability in each channel.	8.0
DLMIN	Distance threshold for combining of clusters during a combine iteration (not L1).	8.0
NMIN2	Cluster size criteria used during possible cluster eliminations prior to splitting or combining clusters.	10
ITMAX	Maximum number of iterations allowed in the preliminary sequence of splitting iterations. If stability is not achieved in PERCENT of the clusters by the ITMAX initial split iteration, the input split-combine sequence is invoked anyway.	3
SPLIT/ COMBINE SEQUENCE	Sequence of S and C characters which indicate the desired types of iterations following the initial splitting iterations.	CC

Table 5.1-3. Fields in Synthetic and Related Images

Class	Number of Fields in Image
Corn	42
Trees	44
Pasture	45
Soybeans	40
Wheat	40
Soil	24
Default (pasture)	18

5.2 ANALYSIS OF PERCENT CORRECTLY CLASSIFIED

5.2.1 Synthetic Image Analysis

Only a few (less than 50) pixels were misclassified within the synthetic image, so no comparison was attempted between this and the associated simulated Thematic Mapper images.

A logistic analysis was used to evaluate the effects of the TM configuration, resampling technique, field size, field shape, and class on the percent correctly classified. (A logistic analysis expresses the logarithm of the probability of correct classification for a given set of levels of the factors as a linear function of the levels. A maximum-likelihood chi square test is then used to determine the significant factors. See Reference 3 for a more extensive discussion of this.) Tables 5.2-1 and 5.2-2 provide, for Entire and Field Center fields respectively, a logistic analysis of correct versus incorrect classification, showing the degrees-of-freedom (df) and the chi-squared test statistic (X^2) resulting from the analysis. From these tables we deduce that:

a. The variations among percent correctly classified can be adequately described by a 1-factor, 2-factor, and 3-factor interactions. This means that it is not necessary to look for effects of four or more factors (e.g., class, TM configuration, resampling, shape) on the percent correctly classified.

b. The effects (class, size, shape) and (class, TM configuration, resampling) with all 1-factor and 2-factor subsets of these are sufficient to adequately describe variations among entire fields.

c. Additional 3-factor effects are required to describe variations among percent correctly classified in Field Center fields. These could not be fully investigated due to the small number of pixels in the field center fields.

Thus sets of three-way tables showing percent correctly classified as a function of (class, TM configuration, resampling) and (class, field-size, field-shape) describe all effects of interest.

Table 5.2-1 Analysis of Logits of Correct vs. Incorrect Classification for Entire Fields of Synthetic Image

Source of variation	df	X^2
1. Total variation of the logits	540	2941.38 *
2. Effect of 1-factor and 2-factor interactions (assuming that 3-factor and higher effects are nil)	79	2200.95 *
3. Effect of 3-factor interactions (assuming that 4-factor and 5-factor effects are nil)	164	481.58 *
4. Effect of 4-factor interactions (assuming that 5-factor effects are nil)	200	196.52
5. Effect of 5-factor interactions	96	62.33
4a. Residual 4-factor and 5-factor interactions	297	258.85
Partition of 2 and 3 by forward selection		
2.1 Effect of (class,size,shape)	35	2004.48 *
2.2 Effect or (class,configuration,resampling) given (class,size,shape)	46	467.38 *
2.3 Effect of (class,configuration,size) given (class,configuration,resampling) and (class,size,shape)	30	85.06 *
2.4 Residual from above effects	428	384.46
2.4a Effect of (class,configuration,shape) given above effects	20	34.32
2.4b Residual	408	350.14

* denotes significance at the 0.05 level

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.2-2 Analysis of Logits of Correct vs. Incorrect Classification for Field Center Fields of Synthetic Image

Source of variation	df	X^2
1. Total variation of the logits	245	1212.67 *
2. Effect of 1-factor and 2-factor interactions (assuming that 3-factor and higher effects are nil)	79	849.42 *
3. Effect of 3-factor interactions (assuming that 4-factor and 5-factor interactions are nil)	164	291.20 *
4. Effect of 4-factor and 5-factor interactions	297	71.77
Partition of 3		
3.1 Effect of (class,size,shape), (class,configuration,resampling), (class,configuration,size), and (class,configuration,shape)	131	193.08 *
3.2 Residual 3-factor interactions	33	98.12 *

* denotes significance at the 0.05 level

Tables 5.2-3 to 5.2-6 contain such sets of three way tables. Figures 5.2-1 to 5.2-4 contain plots of the data, averaged over classes. From a study of these it is seen that:

a. Field Center fields are classified more accurately than Entire fields (average of 73.0 percent versus 94.8 percent correct). This is not surprising, since the Entire fields contain border pixels.

b. The effects of field size and shape vary over the classes. In general, larger fields are more correctly classified than smaller fields. Effects of field shape are not predictable.

c. After averaging over classes, effects of field size are seen very clearly. Field shapes, however, do not significantly affect these results as is evident in Figure 5.2-2 where the reduction in classification accuracy which one would expect as fields become longer and narrower is only minimally present, and then only for fields smaller than 10 acres. The effect of field shape is implicitly observed in Figure 5.2-4, where the absence of data points for the smaller, narrower fields reflects the operational impossibility of defining Field Center fields for these fields. It is difficult to explain these classification results intuitively, for the process of multispectral classification is a data transformation which depends not only on field shape and field size, but also on the interactions among the class signatures, and such statistical characteristics as the degree of normality of the class statistics and the amount of pixel-to-pixel stochastic dependence. A satisfactory explanation for this unexpected behavior would require an extensive investigation of the individual pixel classifications which combine to produce it.

d. The effects of the TM configuration and resampling technique differ from class to class. After averaging over classes, there is a clear advantage in using TM configuration 2. For this configuration the cubic convolution and point spread resampling techniques are slightly better than the nearest neighbor technique in terms of the average percent classified correctly. There does not appear to be a significant difference between the cubic convolution and point spread resampling techniques when using TM configuration 2.

From this analysis of the percent correctly classified in the simulated TM images, one would recommend TM configuration 2 with either the cubic convolution or point spread resampling. This recommendation might change if a system were being designed for monitoring a particular class or set of classes.

5.2.2 Aircraft Image Analysis

For these data sets, the effects of TM configuration, resampling technique, field size, field shape, and crop type of classification accuracy were evaluated by means of a stepwise regression analysis. The more appropriate logistic analysis which was used to evaluate the classifications of the synthetic image data sets could not be used because the available computer programs were unable to deal with the incomplete complement of field size/shape combinations encountered in these data sets.

Table 5.2-3 Probabilities of Correct Classification Showing Effects of (class,resampling,configuration) for Entire Fields of Synthetic Image Data

a. Class Trees

Config.	Resampling NN	CC	PS	Avg
1	83.8	86.4	72.3	80.9
2	81.1	79.5	80.6	80.4
3	80.6	82.0	72.3	78.3
Avg	81.8	82.7	75.1	79.9

b. Class Pasture

Config.	Resampling NN	CC	PS	Avg
1	78.3	75.3	77.4	77.0
2	82.1	83.2	83.7	83.0
3	69.0	68.8	68.8	68.9
Avg	76.5	75.8	76.7	76.3

c. Class Corn

Config.	Resampling NN	CC	PS	Avg
1	64.9	48.4	62.1	57.7
2	67.5	62.8	58.7	63.0
3	60.4	52.9	58.6	57.3
Avg	63.5	54.7	59.8	59.3

d. Class Wheat

Config.	Resampling NN	CC	PS	Avg
1	73.0	80.7	68.6	74.1
2	74.7	81.4	76.6	77.5
3	70.0	71.3	58.3	66.6
Avg	72.5	77.7	67.8	72.6

e. Class Soybeans

Config.	Resampling NN	CC	PS	Avg
1	71.1	77.6	83.3	77.3
2	66.2	85.2	80.8	77.4
3	80.1	79.2	83.1	80.8
Avg	72.5	80.7	82.4	78.5

f. Average over Classes

Config.	Resampling NN	CC	PS	Avg
1	73.6	73.3	72.6	73.0
2	74.2	78.2	76.1	76.1
3	71.8	70.5	67.9	70.1
Avg	73.2	73.8	72.1	73.0

Table 5.2-4 Probabilities of Correct Classification Showing Effects of (class,size,shape) for Entire Fields of Synthetic Image Data

a. Class Trees

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	57.1	82.8	76.8	85.7	80.4
1x2	63.4	67.0	73.0	90.5	80.2
1x4	36.7	77.1	74.6	90.5	78.9
Avg	52.5	75.4	74.8	88.9	79.9

b. Class Pasture

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	81.3	67.9	72.6	83.4	78.4
1x2	53.6	62.6	72.5	76.4	71.6
1x4	81.5	66.7	79.5	84.4	80.3
Avg	71.5	65.5	74.6	81.0	76.3

c. Class Corn

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	41.5	58.1	72.0	61.5	62.0
1x2	55.3	52.2	57.3	64.2	59.8
1x4	45.6	43.0	50.0	65.0	56.2
Avg	48.0	51.0	59.2	63.6	59.3

d. Class Wheat

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	45.3	55.2	70.2	77.8	69.7
1x2	63.9	64.6	79.5	78.0	75.2
1x4	44.4	62.1	81.6	76.2	72.8
Avg	51.3	60.8	77.0	77.3	72.6

e. Class Soybeans

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	57.6	65.9	69.1	92.3	79.2
1x2	55.6	69.1	79.1	77.7	75.0
1x4	54.3	74.1	81.9	89.8	82.3
Avg	55.9	69.4	76.5	85.8	78.5

f. Average over Classes

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	56.9	66.1	72.2	79.6	73.4
1x2	58.3	62.8	71.9	77.2	72.2
1x4	52.0	63.3	72.9	80.3	73.4
Avg	55.5	64.0	72.3	78.9	73.0

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.2-5 Probabilities of Correct Classification Showing Effects of (class,resampling,configuration) for Field Center Fields of Synthetic Image Data

a. Class Trees

Config.	Resampling NN	CC	PS	Avg
1	99.1	100.0	92.9	97.3
2	98.6	98.6	96.2	97.8
3	94.0	96.6	86.8	92.5
Avg	97.2	98.4	91.8	95.8

b. Class Pasture

Config.	Resampling NN	CC	PS	Avg
1	100.0	98.7	100.0	99.6
2	99.6	99.6	99.6	99.6
3	91.5	92.4	93.6	92.5
Avg	97.0	96.9	97.7	97.2

c. Class Corn

Config.	Resampling NN	CC	PS	Avg
1	92.3	83.8	88.0	88.0
2	96.8	96.4	94.7	96.0
3	85.9	80.9	81.3	82.7
Avg	91.5	86.8	87.9	88.8

d. Class Wheat

Config.	Resampling NN	CC	PS	Avg
1	95.4	100.0	93.8	96.4
2	98.7	98.7	98.7	98.7
3	92.0	93.1	84.3	89.8
Avg	95.2	97.6	92.3	95.0

e. Class Soybeans

Config.	Resampling NN	CC	PS	Avg
1	95.1	98.1	100.0	97.7
2	93.9	100.0	100.0	98.8
3	96.9	97.4	97.8	97.4
Avg	96.0	98.5	99.3	98.0

f. Average over Classes

Config.	Resampling NN	CC	PS	Avg
1	96.2	96.0	94.9	95.7
2	97.9	98.9	98.1	98.3
3	91.9	91.8	88.5	90.7
Avg	95.3	95.5	93.8	94.8

Table 5.2-6 Probabilities of Correct Classification Showing Effects of (class,size,shape) for Field Center Fields of Synthetic Image Data

a. Class Trees

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	88.9	99.0	98.5	98.1	97.9
1x2	----	82.8	94.9	96.9	94.9
1x4	----	----	77.5	100.0	93.9
Avg	88.9	91.1	92.1	98.1	95.8

b. Class Pasture

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	100.0	92.5	96.2	98.7	97.6
1x2	----	93.3	94.2	97.1	96.0
1x4	----	----	97.3	99.4	98.9
Avg	100.0	92.9	95.6	98.2	97.2

c. Class Corn

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	72.7	93.5	97.8	91.7	92.5
1x2	----	81.0	84.0	90.3	87.6
1x4	----	----	89.9	83.4	85.0
Avg	72.7	86.9	90.2	88.8	88.8

d. Class Wheat

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	77.8	89.6	96.9	95.3	94.3
1x2	----	93.1	52.9	95.3	84.1
1x4	----	----	96.5	95.6	95.9
Avg	77.8	91.3	80.1	95.4	95.0

e. Class Soybeans

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	100.0	94.6	98.4	99.2	98.5
1x2	----	100.0	96.7	97.1	97.2
1x4	----	----	96.5	99.7	98.6
Avg	100.0	97.1	97.2	98.4	98.0

f. Average over Classes

Shape	Size (acres) 2.5	5.0	10.0	20.0	Avg
1x1	86.5	93.9	97.6	96.5	96.0
1x2	----	89.5	92.7	95.4	94.2
1x4	----	----	92.0	94.9	94.1
Avg	86.5	91.7	94.2	95.6	94.8

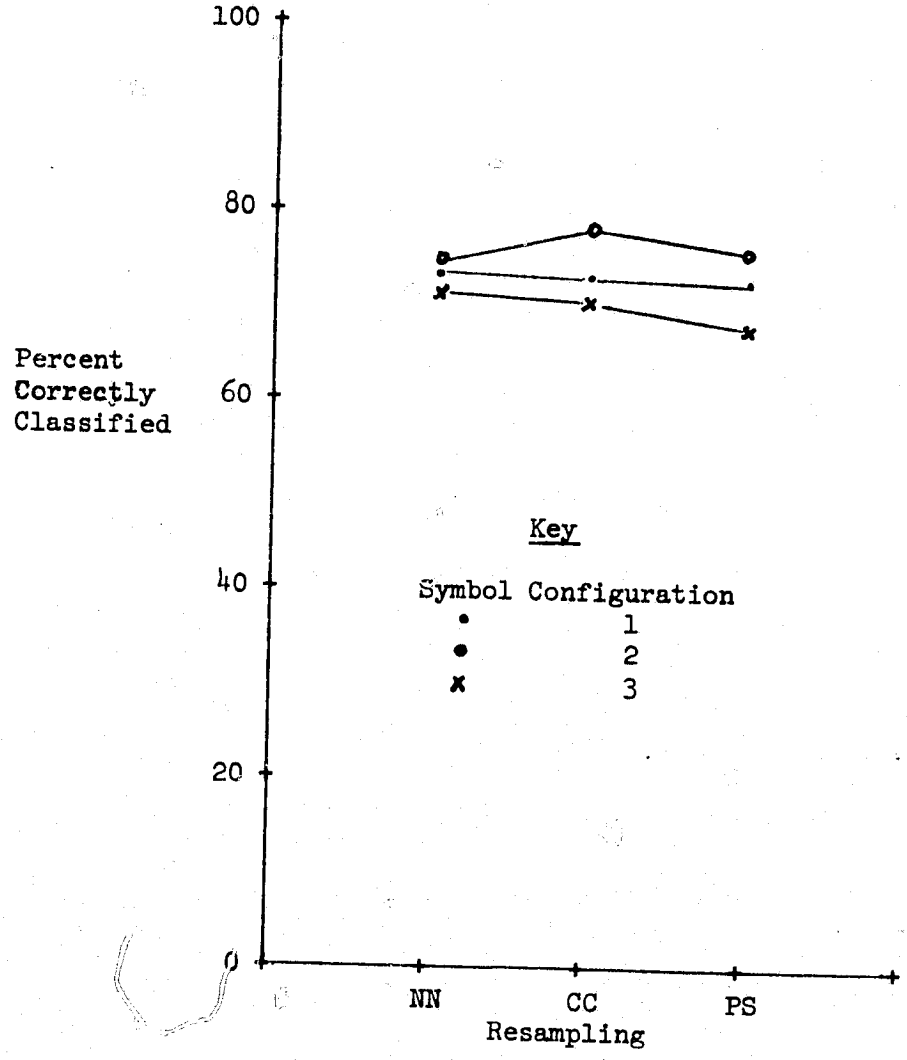
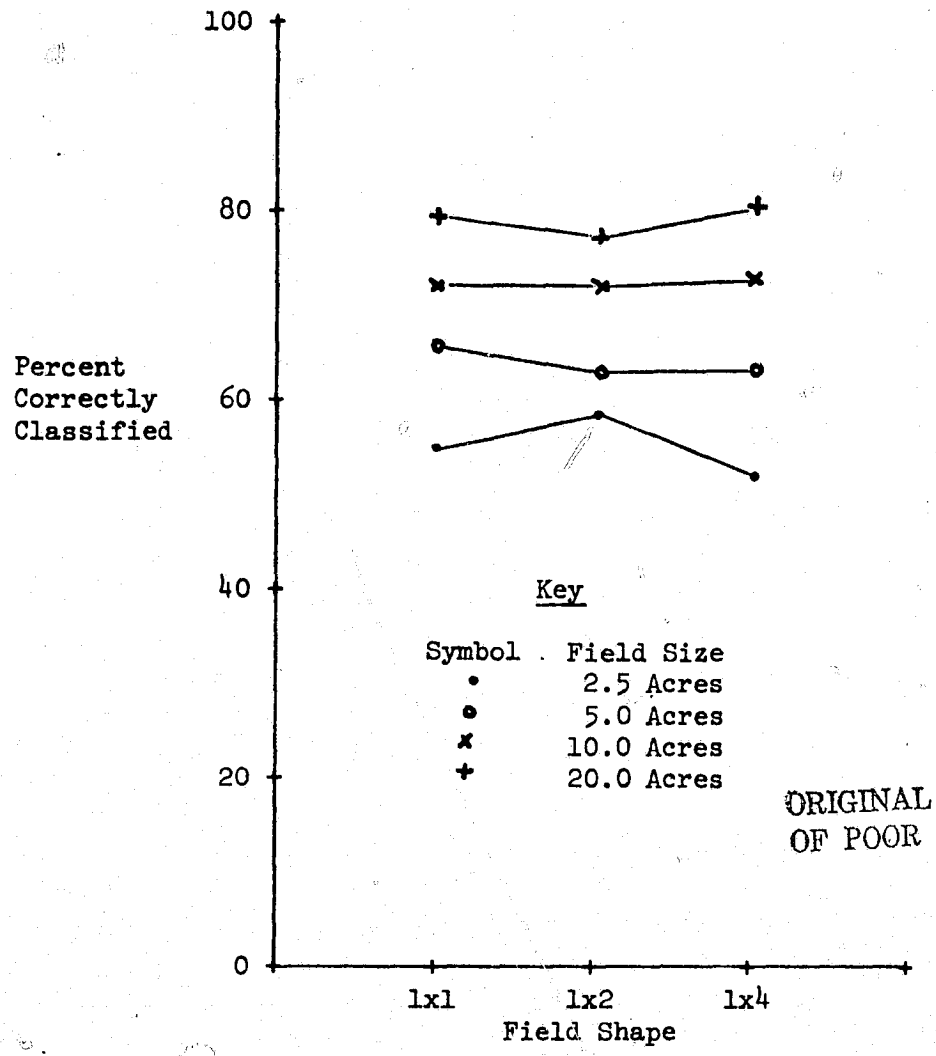


Figure 5.2-1 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Resampling and Configuration for Entire Fields in TM Images Derived from Synthetic Data Set

ORIGINAL PAGE IS
OF POOR QUALITY



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 5.2-2 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Field Size and Field Shape for Entire Fields in TM Images Derived from Synthetic Data Set

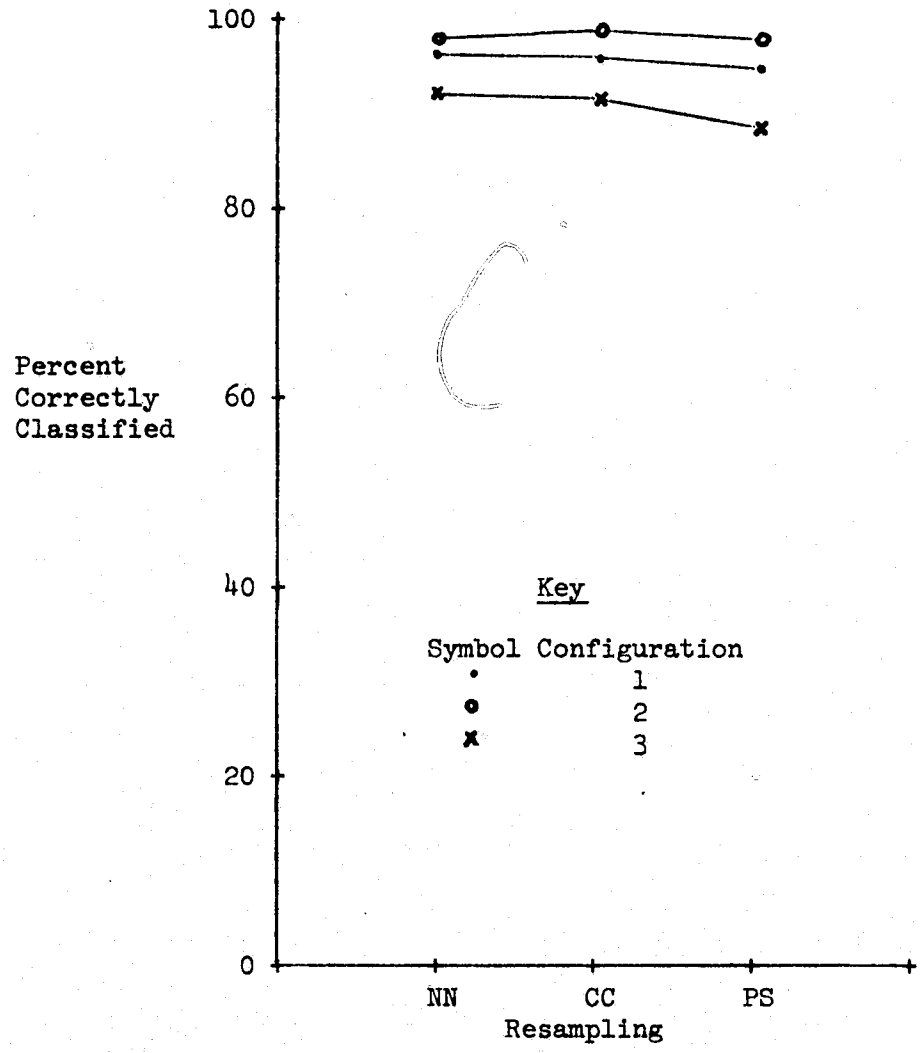


Figure 5.2-3 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Resampling and Configuration for Field Center Fields in TM Images Derived from Synthetic Data Set

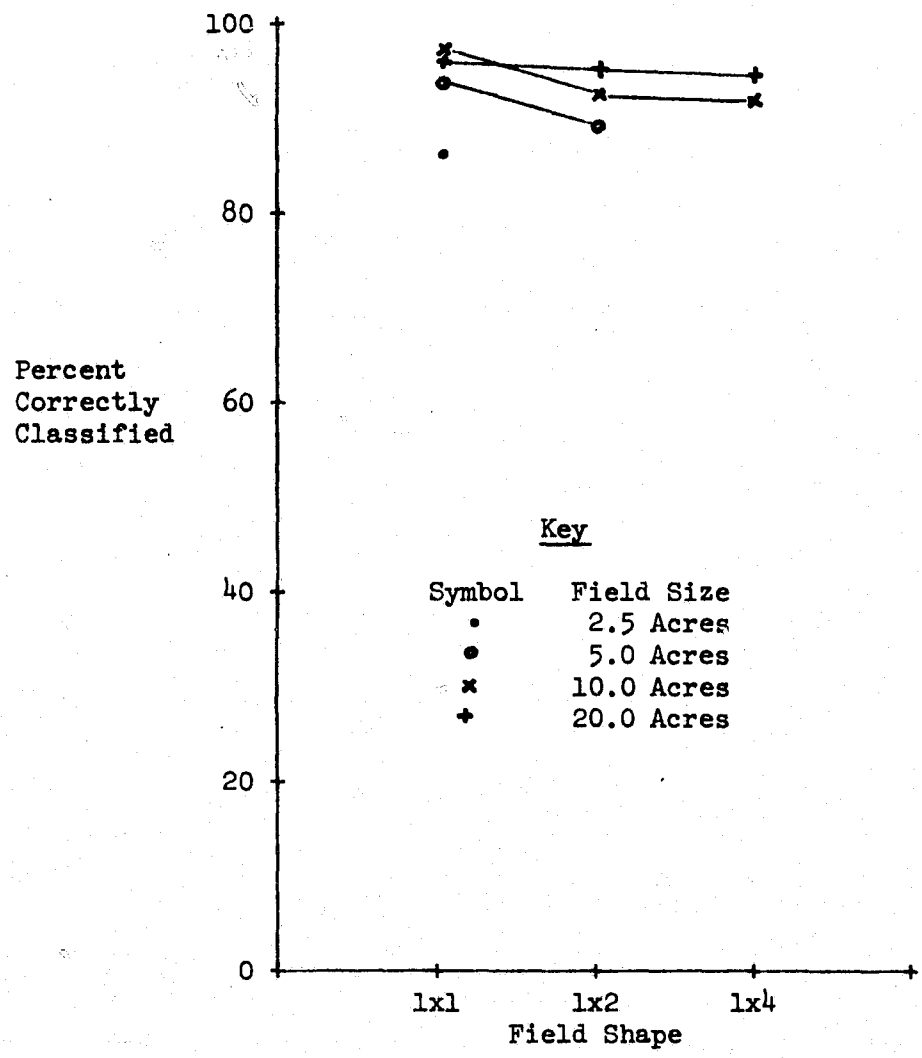


Figure 5.2-4 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Field Size and Field Shape for Field Center Fields in TM Images Derived from Synthetic Data Set

ORIGINAL PAGE IS
OF POOR QUALITY

To perform this analysis, each field's percent correctly classified was expressed as a linear combination of terms associated with the factors mentioned above. A weighted regression analysis, in which each percent correctly classified observation was weighted according to the number of classified pixels, was used to determine the combinations of factors which provided the best explanation of the observed variations. The term best is to be interpreted in a standard stepwise regression sense for which factors are entered in the order of their contributation to the variability. The resulting analysis-of-variance table is given in Table 5.2-7 which gives the degrees-of-freedom (df), sum-of-squares of the variation, mean-square variation, and the corresponding F statistic value for each source of variation in the analysis. In general, the greater the value of the F statistic, the greater the confidence level that this variation is not the result of a random variation. The sources of variation given in the table are significant at the 95 percent confidence level. Note that all of the factors are identified as important sources of interaction; some of them interact in pairs or triplets. Tables 5.2-8 through 5.2-11 show averaged percent correctly classified for different crop types, field types (i.e., training or test fields), TM configurations, and resampling techniques. Note that all averages are weighted averages of the percent correctly classified, with the number of classified pixels serving as the weight.

Tables 5.2-12 through 5.2-19 show the averaged correct classification for combinations of crop types, field sizes, and field shapes. The field sizes and shapes were quantized to form these tables, with the cut points of the quantization chosen after study of histograms of the available sizes and shapes. The field shape factor in these tables is the length to width ratio for a field. Figures 5.2-5 through 5.2-9 contain graphs which present the same information found in the tables for the data averaged over classes.

The following observations can be made, based on the data presented in the tables and graphs:

a. In many cases, the pixels in the Thematic Mapper images were classified more correctly than those in the original aircraft image. This is true for both the training and test field data. A possible explanation for this might be the smoothing imposed on pixel values by convolution with the Thematic Mapper aperture, although accuracy improvement resulting from this process would be diminished by the concomitant contamination of field boundary pixels by the smoothing.

b. The fact that many field size/shape combinations were not available for experimentation precludes any definite conclusions regarding the effect of field size and shape. Even after averaging over classes, there is no indication that larger or more square fields are classified more accurately than other fields.

c. Classification accuracies were very poor in test fields of all crops except, possibly, corn, trees, and pasture. This was probably caused by the fact that only a few fields of the other crop types were available for training, with a resulting inadequate representation of the image data for these crops.

Table 5.2-7 Analysis of Variance from Stepwise Regression of Percent Correctly Classified for Thematic Mapper Data Derived from Aircraft Data Set

Source of Variation	df	Sums of Squares	Mean Square	F
Ground truth class	3	437489.1	145829.7	432.8
Training field vs. test field	1	14400.6	14400.6	47.7
Shape	1	7998.4	7998.4	23.7
Size x filter (linear)	1	3963.1	3963.1	11.8
Shape x filter (linear)	1	12184.6	12184.6	36.2
Shape x filter (linear) x resampling (linear)	1	2056.1	2056.1	6.1
Shape x filter (quadratic)	1	2032.6	2032.6	6.0
Size x filter (quadratic) x resampling (linear)	1	1483.3	1483.3	4.4
Residual	781	263129.1	336.9	
Total	791	744746.0		

Estimated Standard Deviation - 18.36%

Table 5.2-8 Probabilities of Correct Classification Showing Effects of (class,resampling,configuration) for Field Center Training Fields of TM Data derived from Aircraft Image

a. Class Corn

Config.	Resampling NN	CC	PS	Avg
1	94.2	92.2	99.0	95.1
2	77.4	78.9	90.1	82.1
3	79.9	83.7	54.9	72.8
Avg	83.8	84.9	81.3	83.3

Aircraft Image - 87.6

b. Class Trees and Pasture

Config.	Resampling NN	CC	PS	Avg
1	85.4	87.0	96.0	89.5
2	77.4	76.4	93.2	82.3
3	80.1	83.0	89.7	84.3
Avg	81.0	82.1	92.9	85.3

Aircraft Image - 71.1

c. Class Soybeans and Small Grains

Config.	Resampling NN	CC	PS	Avg
1	74.3	80.2	92.8	82.4
2	80.3	82.4	89.5	84.1
3	65.0	72.2	87.3	74.8
Avg	73.2	78.3	89.9	80.5

Aircraft Image - 67.0

d. Class Set-aside and Non-farm

Config.	Resampling NN	CC	PS	Avg
1	92.1	92.1	100.0	94.8
2	85.7	92.3	96.7	91.6
3	78.2	72.4	87.4	79.3
Avg	85.4	85.8	94.8	88.6

Aircraft Image - 68.0

e. Average over Classes

Config.	Resampling NN	CC	PS	Avg
1	86.7	87.8	96.7	90.4
2	78.7	79.9	91.6	83.4
3	76.6	80.1	76.4	77.7
Avg	80.7	82.6	88.2	83.8

Aircraft Image - 78.2

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.2-9 Probabilities of Correct Classification Showing Effects of (class,resampling,configuration) for Field Center Test Fields of TM Data Derived from Aircraft Image

a. Class Corn

Config.	Resampling NN	CC	PS	Avg
1	80.2	82.5	86.8	83.1
2	70.5	69.8	82.8	74.4
3	65.6	67.8	30.4	54.6
Avg	72.3	73.5	68.4	71.4

Aircraft Image - 76.2

b. Class Trees and Pasture

Config.	Resampling NN	CC	PS	Avg
1	77.4	76.1	87.2	80.2
2	66.3	63.3	75.5	68.4
3	82.8	80.3	79.4	80.9
Avg	78.6	77.6	83.7	80.0

Aircraft Image - 81.5

c. Class Soybeans and Small Grains

Config.	Resampling NN	CC	PS	Avg
1	0.0	0.0	35.9	12.0
2	12.8	5.1	2.6	6.8
3	0.0	0.0	0.0	0.0
Avg	4.3	1.7	12.8	33.3

Aircraft Image - 11.0

d. Class Set-aside and Non-farm

Config.	Resampling NN	CC	PS	Avg
1	19.3	29.4	14.3	21.0
2	15.8	15.8	5.0	12.2
3	18.4	21.9	43.0	27.8
Avg	17.8	22.4	20.4	20.2

Aircraft Image - 19.5

e. Average over Classes

Config.	Resampling NN	CC	PS	Avg
1	63.2	64.9	72.8	67.0
2	60.2	57.9	67.4	61.8
3	63.4	62.3	59.3	61.7
Avg	62.3	61.7	66.6	63.5

Aircraft Image - 66.5

Table 5.2-10 Probabilities of Correct Classification Showing Effects of (class,resampling,configuration) for Entire Training Fields of TM Data Derived from Aircraft Image

a. Class Corn

Config.	Resampling NN	CC	PS	Avg
1	86.2	85.8	89.6	87.2
2	72.4	73.4	86.6	77.5
3	73.8	73.5	49.2	65.5
Avg	77.4	77.6	75.2	76.7

Aircraft Image - 78.6

b. Class Trees and Pasture

Config.	Resampling NN	CC	PS	Avg
1	72.6	73.8	85.4	77.3
2	65.2	64.9	83.0	71.0
3	68.0	69.6	80.1	72.6
Avg	68.6	69.4	82.9	73.6

Aircraft Image - 62.7

c. Class Soybeans and Small Grains

Config.	Resampling NN	CC	PS	Avg
1	74.6	64.4	81.4	73.4
2	70.2	67.7	67.7	68.5
3	50.4	50.4	63.2	54.6
Avg	64.5	60.5	70.4	65.2

Aircraft Image - 54.4

d. Class Set-aside and Non-farm

Config.	Resampling NN	CC	PS	Avg
1	19.3	29.4	14.3	21.0
2	15.8	15.8	5.0	12.2
3	18.4	21.9	43.0	27.8
Avg	17.8	22.4	20.4	20.2

Aircraft Image - 19.5

e. Average over Classes

Config.	Resampling NN	CC	PS	Avg
1	78.5	77.4	83.9	79.9
2	68.2	68.3	81.3	72.6
3	67.7	67.3	61.9	65.6
Avg	71.4	71.0	75.7	72.7

Aircraft Image - 68.9

Table 5.2-11 Probabilities of Correct Classification Showing Effects of (class,resampling,configuration) for Entire Test Fields of TM Data Derived from Aircraft Image

a. Class Corn

Config.	Resampling NN	CC	PS	Avg
1	73.2	75.5	76.9	75.2
2	60.3	62.1	73.3	65.2
3	60.3	61.1	32.8	51.4
Avg	64.5	66.1	60.8	63.8

Aircraft Image - 70.5

b. Class Trees and Pasture

Config.	Resampling NN	CC	PS	Avg
1	67.8	68.0	79.4	71.7
2	66.3	63.3	75.5	68.4
3	73.3	70.4	74.4	72.7
Avg	69.1	67.2	76.4	70.9

Aircraft Image - 72.2

c. Class Soybeans and Small Grains

Config.	Resampling NN	CC	PS	Avg
1	11.7	7.4	27.7	15.6
2	12.2	7.1	5.1	8.2
3	5.5	4.4	5.5	5.1
Avg	9.9	6.4	12.7	9.7

Aircraft Image - 10.5

d. Class Set-aside and Non-farm

Config.	Resampling NN	CC	PS	Avg
1	16.0	26.0	14.3	18.8
2	10.0	13.0	6.5	9.8
3	15.1	27.6	38.7	27.1
Avg	13.7	22.1	19.7	18.5

Aircraft Image - 18.6

e. Average over Classes

Config.	Resampling NN	CC	PS	Avg
1	58.5	60.5	65.8	61.6
2	52.7	52.0	60.2	55.0
3	56.5	57.2	51.2	54.9
Avg	55.9	56.5	59.1	57.2

Aircraft Image - 56.0

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.2-12 Percentage Correctly Classified Showing Effects of (class,size,shape) for Field Center Training Fields of TM Data Derived from Aircraft Image

a. Class Corn

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	72.2	88.2	81.4	83.1
1.5-4.0	----	----	----	----
4.0	----	----	86.4	86.4
Avg	72.2	88.2	82.4	83.3

b. Class Trees and Pasture

Shape	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	79.6	84.7	----	83.7
1.5-4.0	62.8	----	85.3	82.7
4.0	----	80.2	95.0	92.3
Avg	68.1	83.5	88.2	85.3

c. Class Soybeans and Small Grains

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	93.2	----	----	93.2
1.5-4.0	----	----	79.1	79.1
4.0	----	----	----	----
Avg	93.2	----	79.1	80.5

d. Class Set-aside and Non-farm

Shape	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	100.0	81.1	----	88.9
1.5-4.0	----	----	----	----
4.0	88.0	----	----	88.0
Avg	94.2	----	----	94.2

e. Average over Classes

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	91.0	85.8	81.4	84.4
1.5-4.0	62.8	----	82.1	80.9
4.0	88.0	80.2	91.4	89.8
Avg	84.3	85.3	83.5	83.9

Table 5.2-13 Percentage Correctly Classified Showing Effects of (class,size,shape) for Field Center Training Fields of Aircraft Image

a. Class Corn

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	83.1	89.6	87.2	87.8
1.5-4.0	----	----	----	----
4.0	----	----	86.7	86.7
Avg	83.1	89.6	87.1	87.6

b. Class Trees and Pasture

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	43.0	64.7	----	61.7
1.5-4.0	42.0	----	70.2	67.8
4.0	----	57.9	89.8	84.4
Avg	42.3	63.0	75.6	71.1

c. Class Soybeans and Small Grains

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	84.5	----	----	84.5
1.5-4.0	----	----	65.4	65.4
4.0	----	----	----	----
Avg	84.5	----	65.4	67.0

d. Class Set-aside and Non-farm

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	93.7	43.3	----	61.4
1.5-4.0	----	----	----	----
4.0	87.3	----	----	87.3
Avg	90.6	43.3	----	68.0

e. Average over Classes

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	81.3	74.1	87.2	81.8
1.5-4.0	42.0	----	67.7	66.6
4.0	56.6	57.9	88.6	83.9
Avg	69.7	72.8	77.2	75.9

Table 5.2-14 Percentage Correctly Classified Showing Effects of (class,size,shape) for Entire Training Fields of TM Data Derived from Aircraft Image

a. Class Corn

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	63.2	82.1	77.6	78.5
1.5-4.0	----	----	----	----
4.0	----	----	69.4	69.4
Avg	63.2	82.1	75.3	76.7

b. Class Trees and Pasture

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	68.5	69.9	----	69.7
1.5-4.0	53.0	----	80.0	73.6
4.0	----	42.9	93.4	76.0
Avg	55.9	59.2	84.1	73.6

c. Class Soybeans and Small Grains

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	91.8	----	----	91.8
1.5-4.0	----	----	64.3	64.3
4.0	----	----	----	----
Avg	91.8	----	64.3	71.5

d. Class Set-aside and Non-farm

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	74.0	54.1	----	60.9
1.5-4.0	----	----	----	----
4.0	56.6	----	----	56.6
Avg	65.5	54.1	----	59.8

e. Average over Classes

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	69.3	72.0	77.6	74.0
1.5-4.0	53.0	----	76.0	71.7
4.0	56.6	42.9	79.5	70.6
Avg	62.2	68.4	77.6	72.7

Table 5.2-15 Percentage Correctly Classified Showing Effects of (class,size,shape) for Entire Training Fields of Aircraft Image

a. Class Corn

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	55.6	83.7	79.6	80.2
1.5-4.0	----	----	----	----
4.0	----	----	71.5	71.5
Avg	55.6	83.7	77.5	78.6

b. Class Trees and Pasture

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	37.2	59.4	----	56.0
1.5-4.0	38.2	----	62.0	57.1
4.0	----	48.8	89.8	76.8
Avg	38.0	55.4	71.0	62.7

c. Class Soybeans and Small Grains

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	67.5	----	----	67.5
1.5-4.0	----	----	48.5	48.5
4.0	----	----	----	----
Avg	67.5	----	48.5	54.4

d. Class Set-aside and Non-farm

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	83.8	43.1	----	54.1
1.5-4.0	----	----	----	----
4.0	60.2	----	----	60.2
Avg	71.2	----	----	55.5

e. Average over Classes

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	66.9	67.9	79.6	73.0
1.5-4.0	38.2	----	58.5	55.3
4.0	60.9	48.8	79.3	72.5
Avg	57.9	65.8	72.8	68.9

Table 5.2-16 Percentage Correctly Classified Showing Effects of (class,size,shape) for Field Center Test Fields of TM Data Derived from Aircraft Image

a. Class Corn

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	76.1	74.5	64.0	70.8
1.5-4.0	86.3	----	----	86.3
4.0	65.1	----	----	65.1
Avg	75.1	74.5	64.0	71.4

b. Class Trees and Pasture

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	19.4	----	----	19.4
1.5-4.0	----	----	86.6	86.6
4.0	----	----	84.5	84.5
Avg	19.4	----	86.0	80.0

c. Class Soybeans and Small Grains

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	----	----	----	----
1.5-4.0	----	----	6.3	6.3
4.0	----	----	----	----
Avg	----	----	6.3	6.3

d. Class Set-aside and Non-farm

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	10.0	22.0	----	21.2
1.5-4.0	16.9	----	----	16.9
4.0	----	----	----	----
Avg	15.5	22.0	----	20.2

e. Average over Classes

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	34.9	45.2	64.0	45.2
1.5-4.0	47.1	----	49.0	48.9
4.0	65.1	----	46.0	48.4
Avg	41.4	45.2	49.7	47.6

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.2-17 Percentage Correctly Classified Showing Effects of (class, size, shape) for Field Center Test Fields of Aircraft Image

a. Class Corn

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	87.5	77.0	67.6	74.7
1.5-4.0	83.4	----	----	83.4
4.0	85.1	----	----	85.1
Avg	86.0	77.0	67.6	76.2

b. Class Trees and Pasture

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	23.4	----	----	23.4
1.5-4.0	----	----	86.0	86.0
4.0	----	----	81.8	81.8
Avg	23.4	----	85.0	81.2

c. Class Soybeans and Small Grains

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	----	----	----	----
1.5-4.0	----	----	11.0	11.0
4.0	----	----	----	----
Avg	----	----	11.0	11.0

d. Class Set-aside and Non-farm

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	0.0	23.1	----	21.9
1.5-4.0	11.0	----	----	11.0
4.0	----	----	----	----
Avg	9.2	23.1	----	19.5

e. Average over Classes

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	58.3	47.2	67.6	56.2
1.5-4.0	41.8	----	78.8	75.2
4.0	85.1	----	81.8	82.3
Avg	57.3	47.2	77.5	69.2

Table 5.2-18 Percentage Correctly Classified Showing Effects of (class,size,shape) for Entire Test Fields of TM Data Derived from Aircraft Image

a. Class Corn

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	68.6	64.3	58.8	63.7
1.5-4.0	78.4	----	----	78.4
4.0	55.1	----	----	55.1
Avg	66.0	64.3	58.8	63.8

b. Class Trees and Pasture

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	18.7	----	----	18.7
1.5-4.0	----	----	78.0	78.0
4.0	----	----	77.2	77.2
Avg	18.7	----	77.8	70.9

c. Class Soybeans and Small Grains

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	----	----	----	----
1.5-4.0	----	----	9.7	9.7
4.0	----	----	----	----
Avg	----	----	9.7	9.7

d. Class Set-aside and Non-farm

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	16.7	19.1	----	18.8
1.5-4.0	17.3	----	----	17.3
4.0	----	----	----	----
Avg	17.1	19.1	----	18.5

e. Average over Classes

Shape Factor	Size (aircraft pixels) less than 4000	4000 to 8000	more than 8000	Avg
1.5	20.0	16.5	7.1	12.8
1.5-4.0	26.8	----	65.9	55.4
4.0	29.6	----	77.2	52.7
Avg	23.0	16.5	32.7	26.8

Table 5.2-19 Percentage Correctly Classified Showing Effects of (class,size,shape) for Entire Test Fields of Aircraft Image

a. Class Corn

Size (aircraft pixels)

Shape Factor	less than 4000	4000 to 8000	more than 8000	Avg
1.5	76.0	65.2	62.6	66.9
1.5-4.0	82.8	----	----	82.8
4.0	79.2	----	----	79.2
Avg	78.7	65.2	62.6	70.5

b. Class Trees and Pasture

Size (aircraft pixels)

Shape Factor	less than 4000	4000 to 8000	more than 8000	Avg
1.5	25.5	----	----	25.5
1.5-4.0	----	----	79.9	79.9
4.0	----	----	69.6	69.6
Avg	25.5	----	76.8	72.2

c. Class Soybeans and Small Grains

Size (aircraft pixels)

Shape Factor	less than 4000	4000 to 8000	more than 8000	Avg
1.5	----	----	----	----
1.5-4.0	----	----	10.5	10.5
4.0	----	----	----	----
Avg	----	----	10.5	10.5

d. Class Set-aside and Non-farm

Size (aircraft pixels)

Shape Factor	less than 4000	4000 to 8000	more than 8000	Avg
1.5	0.7	21.7	----	19.7
1.5-4.0	14.4	----	----	14.4
4.0	----	----	----	----
Avg	10.9	21.7	----	18.6

e. Average over Classes

Size (aircraft pixels)

Shape Factor	less than 4000	4000 to 8000	more than 8000	Avg
1.5	49.1	41.5	62.6	48.6
1.5-4.0	48.3	----	68.1	65.5
4.0	79.2	----	69.6	72.2
Avg	55.7	41.5	67.6	60.2

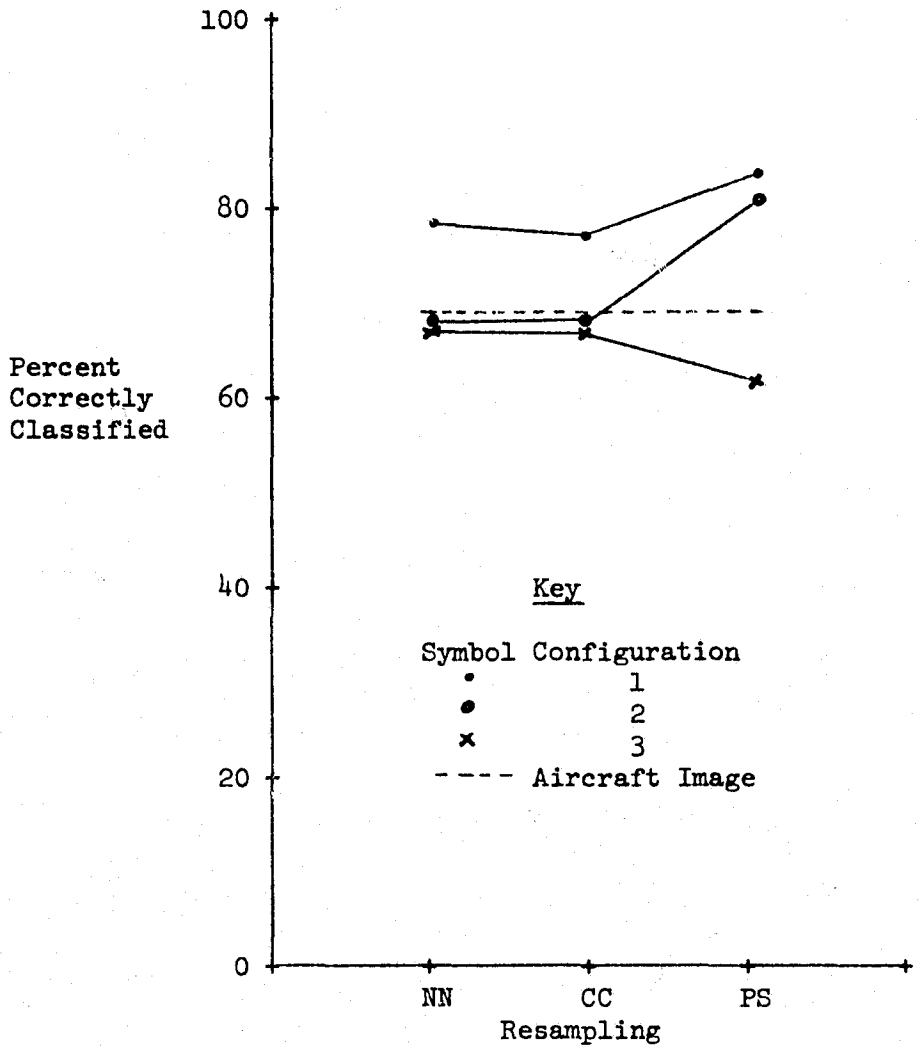


Figure 5.2-5 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Resampling and Configuration for Entire Training Fields in TM Images Derived from Aircraft Data Set

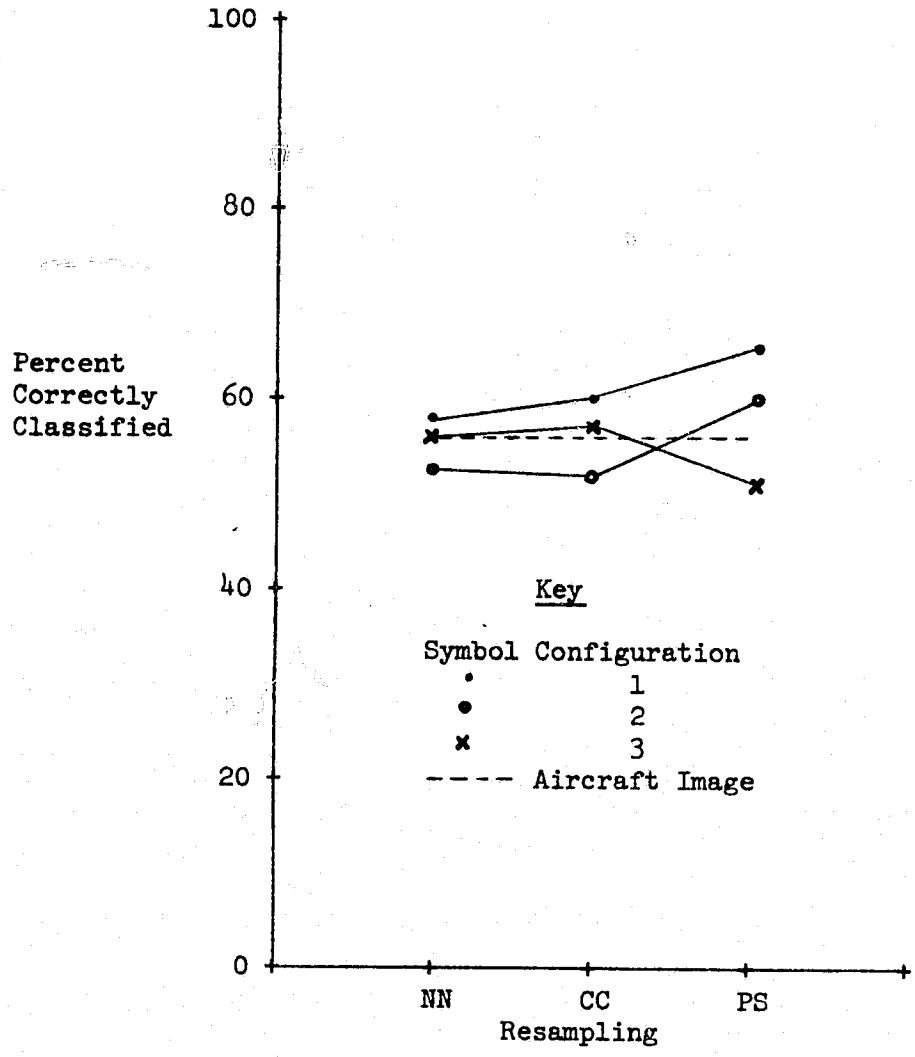


Figure 5.2-6 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Resampling and Configuration for Entire Test Fields in TM Images Derived from Aircraft Data Set

ORIGINAL PAGE IS
OF POOR QUALITY

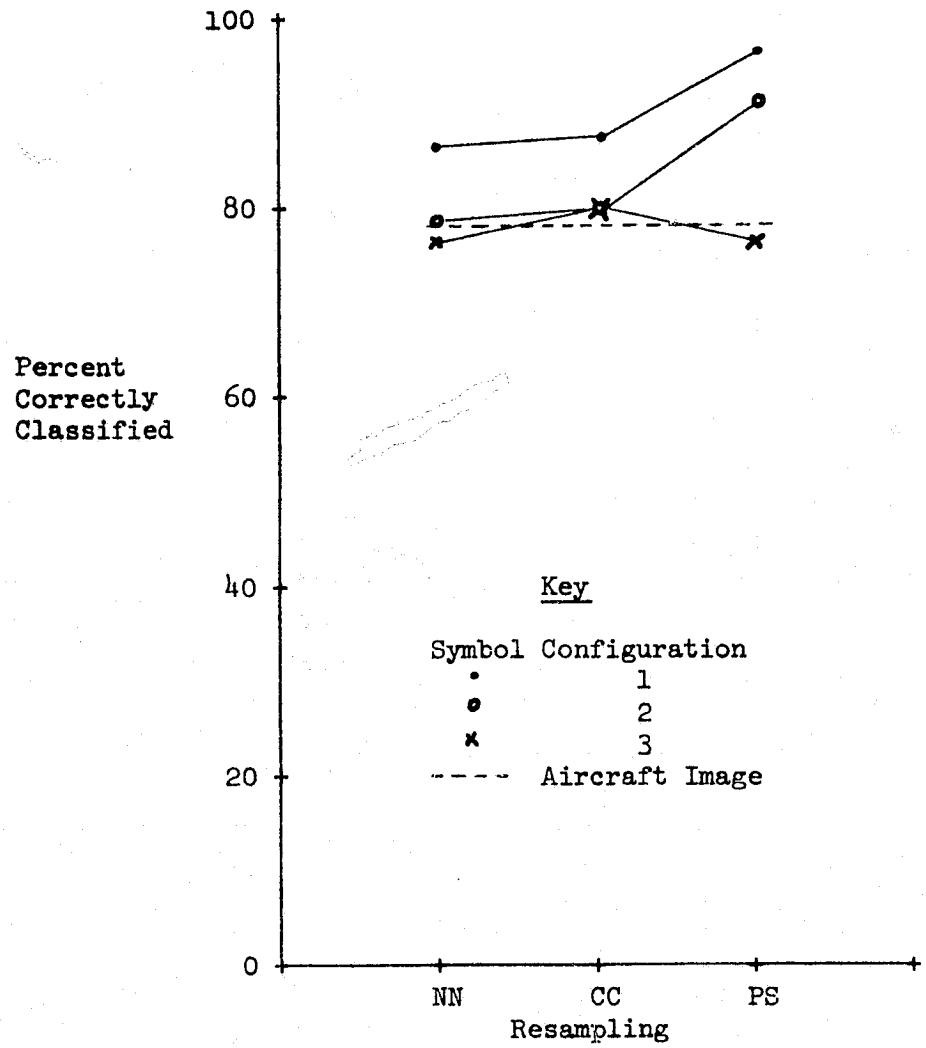


Figure 5.2-7 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Resampling and Configuration for Field Center Training Fields in TM Images Derived from Aircraft Data Set

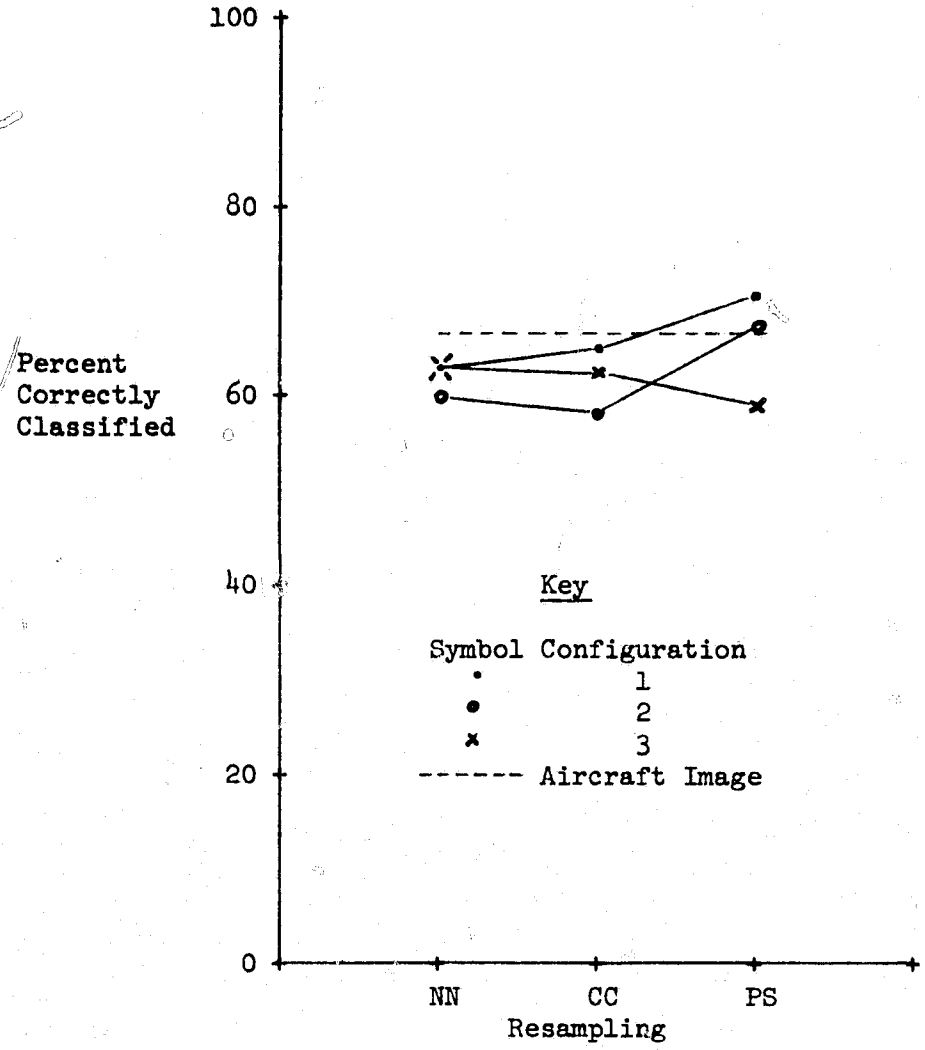


Figure 5.2-8 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Resampling and Configuration for Field Center Test Fields in TM Images Derived from Aircraft Data Set

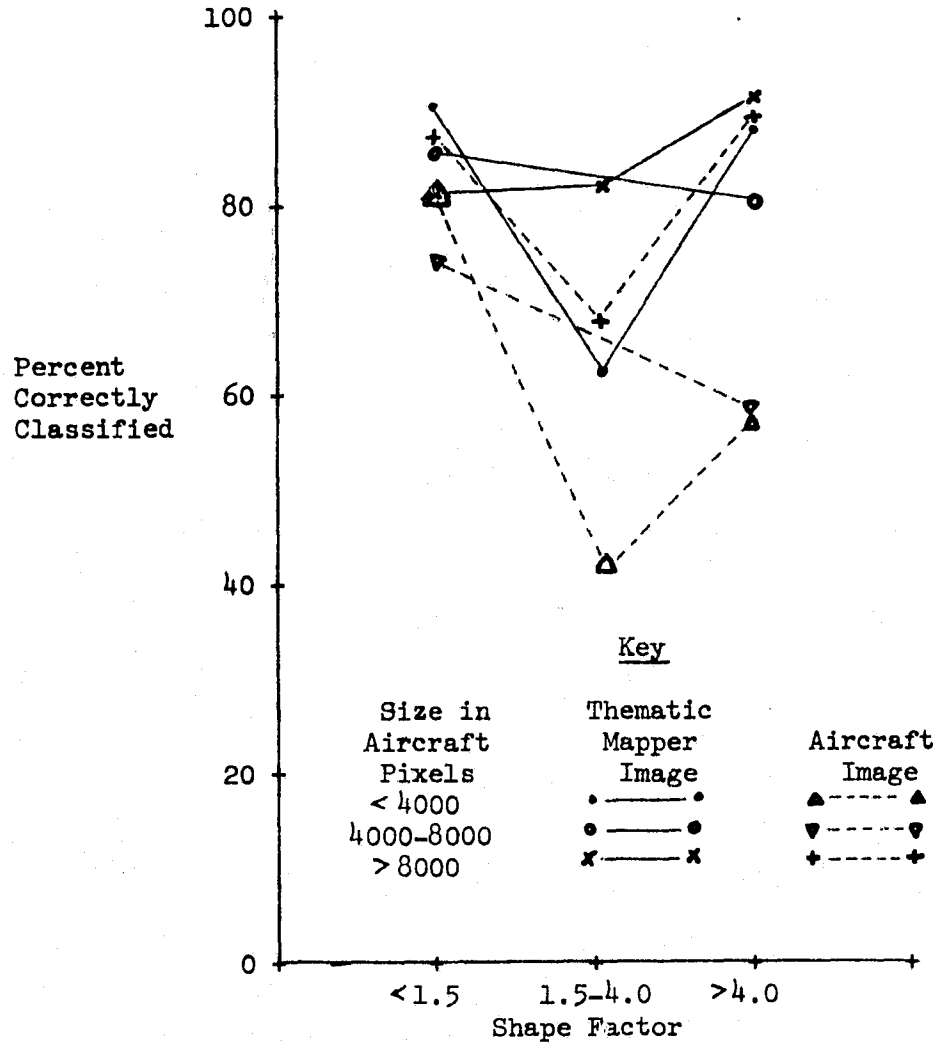


Figure 5.2-9 Probabilities of Correct Classification, Averaged over Classes, Showing Effects of Field Size and Field Shape for Field Center Training Fields in TM Images Derived from Aircraft Data Set and in Aircraft Data Set

ORIGINAL PAGE IS
OF POOR QUALITY

d. Even the field center training fields do not exhibit the very high classification accuracies usually associated with field centers of training fields. This is particularly surprising since supervised clustering was used to develop subclass statistics for classification. Note, in particular, the poor result for class corn when using TM configuration 3 with point spread resampling. This is caused by the fact that only one corn cluster was found when clustering these data. All other factor combinations resulted in three or more corn clusters. Apparently, statistics from this one cluster did not adequately represent the corn image data; even some small training fields were badly misclassified. It was, unfortunately impossible to pursue a data analysis of this strange clustering behavior, or other anomalies, within the available study resources.

e. There is an indication that TM configuration 1 gives better classification results than either of the other two configurations. There is considerable overlap of the results obtained with configurations 2 and 3, with configuration 2 slightly preferred over configuration 3.

f. With respect to classification accuracy, the point spread function compensation resampling technique provided better results than cubic convolution, which in turn provided better results than nearest neighbor assignment.

Based on an analysis of the percent correctly classified for these data sets, the best configuration for the Thematic Mapper is configuration 1, and any resampling performed as part of the ground processing should employ point spread resampling. The obvious limitation associated with this recommendation is that it is based on a single acquisition of a single aircraft image with a limited number of fields.

The TM configuration identified as best on the basis of classification of the aircraft image derived TM data sets differs from that identified on the basis of the classification of the synthetic image derived TM data sets, while the resampling technique recommendation is consistent with the previous result. It is worthy of note that two TM configurations differ only with respect to the along-scan sampling rate, this being 1.0 samples/ IFOV for configuration 1 and 1.4 samples/IFOV for configuration 2.

5.3 ANALYSIS OF PROPORTION ERRORS

For the purpose of this analysis, it was assumed that proportions should be estimated in the standard manner, using a simple classify and count procedure. The results may not be indicative of those which would be obtained with other proportion estimators. For Example, if accurate labels can be attached to some points in the image, the bias correction procedure used in LACIE is known to produce unbiased proportion errors. In that case, the variance of the proportion errors is of primary importance. However, the only adequate way to estimate proportion error variances is to classify multiple images, and the results of such multiple classifications could not be obtained from the data available in this study.

The evaluation of proportion errors is based on a well known mathematical model of the classification process. To establish this model, assume that there are s classes of interest in an image and that the image contains t field types (field size and shape combinations). The following notation is used:

$c_{ji}^{(k)}$ = probability that a pixel from field type k and class i is classified into class j.

$N_i^{(k)}$ = the number of pixels in field type k and class i.

$n_j^{(k)}$ = the number of pixels from field type k classified into class j.

$$T^{(k)} = \sum_i N_i^{(k)}$$

$$T = \sum_k T^{(k)}$$

$$n^{(k)} = \left(n_1^{(k)}, n_2^{(k)}, \ldots\ldots, n_s^{(k)}\right)'$$

$$N^{(k)} = \left(N_1^{(k)}, N_2^{(k)}, \ldots\ldots, N_s^{(k)}\right)'$$

$$C^{(k)} = \left\{c_{ij}^{(k)}\right\}$$

where ' denotes the transpose of the matrix.

The vector of estimated class proportions, P, is defined as

$$P = \frac{1}{T} \sum_k n^{(k)} = \sum_k \frac{T^{(k)}}{T} \frac{n^{(k)}}{T^{(k)}}$$

Similarly, the vector of actual class proportions, $\underline{P}$, is defined as

$$\underline{P} = \sum_k \frac{T^{(k)}}{T} \frac{N^{(k)}}{T^{(k)}}$$

and the errors of the proportion estimates are contained in the vector e defined as

$$e = P - \underline{P}$$

Each term in the sum has a physical interpretation. That is

$\frac{T^{(k)}}{T}$ = the proportion of image pixels present in fields of type k.

$\frac{n^{(k)}}{T^{(k)}}$ = the estimated proportion vector for fields of type k.

$\frac{N^{(k)}}{T^{(k)}}$ = the true proportion vector for fields of type k.

To provide a more compact formulation, the following additional notation will be used:

$$a^{(k)} = \frac{T^{(k)}}{T}, \quad p^{(k)} = \frac{n^{(k)}}{T^{(k)}}, \quad \underline{p}^{(k)} = \frac{N^{(k)}}{T^{(k)}}, \text{ and } e^{(k)} = p^{(k)} - \underline{p}^{(k)}$$

The vector of proportion errors can then be written as

$$e = \sum_k a^{(k)} (p^{(k)} - \underline{p}^{(k)}) = \sum_k a^{(k)} e^{(k)} .$$

Since classification is essentially a random process, the statistics of e may be considered. In particular, the average proportion error, E(e), is easily seen to be

$$E(e) = \sum_k a^{(k)} E(e^{(k)}) = \sum_k a^{(k)} (C^{(k)} - I)\underline{p}^{(k)}$$

where I is the identity matrix of order s. Other statistics (e.g., variance) are more complicated since they depend upon the distributional properties of the pixel classifications and are better studied by replicating the experiment (with additional images) than by mathematical modeling. These are not considered in the following analysis. The vector of expected (average) proportion errors depends upon:

a. The class confusion matrices, $C^{(k)}$, for the different field types.

b. The true class proportion vectors, $\underline{p}^{(k)}$, for each field type, and

c. The proportion of pixels, $a^{(k)}$, in each field type.

It is very likely that the optimal choice of filter and resampling technique depends upon the anticipated values of the $C^{(k)}$, $a^{(k)}$, and $\underline{p}^{(k)}$.

The effect of Thematic Mapper configuration and resampling technique were studied by

a. Estimating the confusion matrices $C^{(k)}$ based upon the observed classification results,

b. Establishing a senario of class proportions $P^{(k)}$, and

c. Estimating the proportion errors $e^{(k)}$ for each field type.

Since the average proportion error is zero (the total proportion must add to 100 percent) two summary statistics were computed for each k. These are the average of the absolute values of the errors

$$\frac{1}{t} = \sum_j \left| e_j^{(k)} \right|$$

and the root mean square error defined by

$$\frac{1}{t} \sum_j \left\{ \left[e_j^{(k)} \right]^2 \right\}^{1/2}$$

Two scenarios were used for the images. The first assumed that the classes were equally likely; that is, that all the elements of $\underline{P}^{(k)}$ equal (1/t). The second scenario assumed an image with 50 percent corn. In each class and field type combinations which were missing in the imagery were assumed to be missing in this analysis. This was done because there was no information available to estimate the appropriate terms in the $C^{(k)}$.

The following sections contain plots, tables, and discussion of the results derived from evaluating the proportion errors in the synthetic and aircraft derived data sets. Only the plots of root mean square errors have been included in this report, in order to limit the size of the presentation.

5.3.1 Synthetic Image Analysis

Tables 5.3-1 through 5.3-4 gives values of the estimated average proportion errors for each class, field size, field shape, TM configuration, and resampling technique combination for the equally-likely-classes scenario. The root mean square errors are shown in Figures 5.3-1 through 5.3-4. In these tables and figures, the field shape factor is the length-to-width ratio for a field. In general, the smaller fields have larger root mean square (and other) errors than do the larger fields. The optumum choice of a TM configuration and resampling technique depends on the error criterion, field size, and field shape. TM configuration 2 with cubic convolution resampling is generally the best for the root mean square criterion. As shown in the following table, this combination is also best for the root mean square criterion averaged over all field sizes and shapes.

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.3-1 Estimated Average Proportion Errors for 2.5 Acre Fields of Simulated Image (Equally Likely Classes)

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Trees	6.7	21.8	-1.8	1.7	-1.0	-1.9	6.0	12.7	0.1
Pasture	11.3	2.2	8.9	19.1	9.6	18.1	-2.4	1.3	-4.6
Corn	-2.3	-14.4	-3.0	-2.8	-3.7	-6.1	-9.5	-12.3	-2.2
Wheat	-5.7	-3.2	-15.0	-7.5	-4.4	-8.5	-8.4	-8.4	-13.9
Soybeans	-10.0	-6.4	10.9	-10.4	-0.6	-1.7	14.3	6.8	20.7
Avg Abs Error	7.2	9.6	7.9	8.3	3.8	7.2	8.1	8.3	8.3
RMS Error	7.9	12.1	9.3	10.4	5.0	9.4	9.0	9.3	11.4

a. Field Shape Factor 1

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Trees	17.7	13.6	-3.7	8.0	3.2	2.1	11.2	14.4	4.8
Pasture	4.1	-8.1	1.5	9.4	3.4	12.2	-3.6	-10.8	-14.1
Corn	-2.5	-3.7	2.1	-2.3	-0.8	1.3	0.7	0.2	2.8
Wheat	-9.3	5.7	-4.5	-1.3	-1.8	-6.8	-8.8	0.2	-4.2
Soybeans	-10.0	-7.5	4.6	-13.8	-4.0	-8.8	0.5	-4.0	10.7
Avg Abs Error	8.7	7.7	3.3	6.9	2.6	6.2	5.0	5.9	7.3
RMS Error	10.2	8.4	3.5	8.3	2.9	7.5	6.6	8.3	8.5

b. Field Shape Factor 2

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Trees	6.3	4.9	-13.6	-5.5	-8.4	-6.8	7.9	-0.3	-5.3
Pasture	11.2	7.7	4.7	24.3	23.3	18.7	0.2	2.2	-0.9
Corn	-2.1	-2.2	8.5	1.0	0.8	-0.8	-2.9	5.4	-1.4
Wheat	-12.0	-8.4	-20.0	-10.9	-12.0	-10.6	-9.3	-8.0	-1.3
Soybeans	-3.4	-2.0	20.4	-9.0	-3.7	-0.5	4.0	0.6	9.0
Avg Abs Error	7.0	5.1	13.4	10.1	9.7	7.5	4.9	3.3	3.6
RMS Error	8.1	5.7	14.8	12.8	12.4	10.1	5.9	4.4	4.8

c. Field Shape Factor 4

Table 5.3-2 Estimated Average Proportion Errors for 5.0 Acre Fields of Simulated Image (Equally Likely Classes)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	9.1	11.8	3.8	0.0	3.2	1.1	8.3	8.2	4.7
Pasture	9.7	0.8	1.2	12.9	6.3	9.1	-3.1	-3.1	-2.4
Corn	-2.8	-3.8	-1.6	-3.1	-5.3	-7.4	-4.5	-3.7	-3.7
Wheat	-7.4	-0.2	-9.5	-3.3	-0.5	-1.2	-3.3	-1.1	-10.2
Soybeans	-8.6	-8.6	6.1	-6.5	-3.7	-1.6	2.5	-0.2	11.6
Avg Abs Error	7.5	5.0	4.4	5.2	3.8	4.1	4.3	3.3	6.5
RMS Error	7.9	6.7	5.4	6.8	4.3	5.3	4.8	4.3	7.5

a. Field Shape Factor 1

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	4.7	9.3	-2.6	-1.0	0.3	0.3	1.7	1.7	-2.1
Pasture	9.1	-2.4	-0.0	10.5	5.0	5.3	-2.4	-2.4	-2.4
Corn	-5.8	-8.1	0.9	-2.2	-3.4	-5.1	-6.3	-3.9	-0.7
Wheat	-1.7	-0.5	-2.4	-1.7	-0.5	-0.4	1.2	1.7	-5.4
Soybeans	-6.4	1.7	4.1	-5.6	-1.4	-0.1	5.8	2.9	10.5
Avg Abs Error	5.5	4.4	2.0	4.2	2.1	2.3	3.5	2.5	4.2
RMS Error	6.0	5.7	2.5	5.5	2.8	3.3	4.1	2.6	5.5

b. Field Shape Factor 2

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	12.3	15.2	-6.1	6.8	-1.3	2.0	10.4	17.7	-1.5
Pasture	7.8	-7.0	-1.9	11.7	0.7	8.2	-4.0	-8.2	-8.0
Corn	-7.5	-4.5	3.1	-5.5	-1.5	-8.8	-7.1	-6.2	3.6
Wheat	-2.7	-0.0	-4.5	-3.7	5.0	2.3	-6.9	-9.4	-13.1
Soybeans	-10.0	-3.7	9.4	-9.3	-2.8	-3.7	7.7	6.0	19.0
Avg Abs Error	8.1	6.1	5.0	7.4	2.3	5.0	7.2	9.5	9.0
RMS Error	8.7	7.9	5.6	7.9	2.7	5.8	7.5	10.4	11.1

c. Field Shape Factor 4

Table 5.3-3 Estimated Average Proportion Errors for 10.0 Acre Fields of Simulated Image (Equally Likely Classes)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	8.7	13.0	0.9	0.9	-2.1	4.1	7.0	9.5	-0.3
Pasture	5.9	1.6	1.8	12.4	8.3	7.9	-0.1	-1.1	-2.9
Corn	-5.7	-7.2	0.1	-3.2	-5.6	-5.2	-5.1	-2.8	2.3
Wheat	-4.1	-3.7	-5.0	-3.2	-3.5	-3.7	-3.7	-2.4	-6.0
Soybeans	-4.8	-3.7	2.1	-6.9	2.8	-3.1	1.8	-3.1	6.8
Avg Abs Error	5.8	5.8	2.0	5.3	4.5	4.8	3.5	3.8	3.7
RMS Error	6.0	7.1	2.6	6.6	5.0	5.1	4.3	4.8	4.4

a. Field Shape Factor 1

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	1.1	2.1	-3.9	-2.0	-1.6	-3.9	-4.2	2.0	-4.4
Pasture	5.1	1.0	3.1	8.3	5.1	6.8	17.1	-2.6	-2.2
Corn	-7.2	-9.6	-5.6	-5.9	-7.5	-5.8	-3.7	-6.4	-1.8
Wheat	2.8	4.8	-0.1	3.4	3.1	0.9	1.9	3.7	-1.5
Soybeans	-1.8	1.7	6.6	-3.8	0.9	2.1	-11.0	3.3	9.9
Avg Abs Error	3.6	3.8	3.9	4.7	3.6	3.9	7.6	3.6	4.0
RMS Error	4.2	5.0	4.5	5.2	4.4	4.5	9.5	3.9	5.0

b. Field Shape Factor 2

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	1.5	9.4	-5.1	1.1	-0.2	1.7	2.8	6.4	-3.9
Pasture	8.2	3.1	3.8	12.4	7.5	9.8	-3.7	-6.0	-7.9
Corn	-7.0	-12.2	-5.3	-6.2	-9.6	-10.0	-5.7	-6.1	-0.5
Wheat	-3.8	-0.6	-2.4	-3.7	-0.6	-2.9	-1.7	-0.1	-2.5
Soybeans	1.1	0.3	9.0	-3.5	2.9	1.5	8.3	5.8	14.8
Avg Abs Error	4.3	5.1	5.1	5.4	4.2	5.2	4.5	4.9	5.9
RMS Error	5.2	7.0	5.6	6.6	5.6	6.5	5.0	5.4	7.8

c. Field Shape Factor 4

Table 5.3-4 Estimated Average Proportion Errors for 20.0 Acre Fields of Simulated Image (Equally Likely Classes)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	6.7	7.5	0.3	1.4	0.3	1.2	7.4	6.9	2.1
Pasture	0.8	-2.9	-1.8	5.4	0.0	2.3	-2.7	-2.6	-2.2
Corn	-4.9	-7.2	-2.3	-3.6	-3.5	-3.8	-4.9	-5.9	-2.3
Wheat	-3.7	-0.9	-3.9	-3.3	-0.6	-2.5	-3.3	-3.3	-6.4
Sotbeans	1.1	3.6	7.7	0.1	3.8	2.8	3.6	4.9	8.8
Avg Abs Error	3.4	4.4	3.2	2.8	1.6	2.5	4.4	4.7	4.4
RMS Error	4.1	5.1	4.1	3.3	2.3	2.7	4.7	5.0	5.2

a. Field Shape Factor 1

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	5.2	13.3	4.8	4.6	1.0	7.3	2.0	7.2	3.2
Pasture	-13.8	-2.3	-1.8	5.8	-1.4	0.1	-3.3	-3.7	-3.1
Corn	-4.3	-9.1	-4.4	-3.4	-5.6	-5.2	-3.9	-7.0	-6.7
Wheat	-2.0	1.3	-3.1	-1.1	1.4	-0.3	-2.5	-2.3	-4.7
Soybeans	-5.1	-3.2	4.5	-5.9	4.5	-1.9	7.6	5.8	11.2
Avg Abs Error	6.1	5.9	3.7	4.2	2.8	3.0	3.9	5.2	5.8
RMS Error	7.3	7.5	3.9	4.5	3.4	4.1	4.4	5.6	6.5

b. Field Shape Factor 2

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	6.0	8.8	-0.6	4.4	-0.9	1.0	5.5	9.9	1.3
Pasture	2.4	-1.8	0.1	7.7	2.4	4.3	-3.7	-3.2	-2.9
Corn	-5.9	-9.3	-2.5	-4.5	-4.3	-3.5	-18.8	-8.8	-3.4
Wheat	-0.0	2.7	-1.7	-2.9	1.0	-1.4	-4.2	-0.7	-6.0
Soybeans	-2.5	-0.4	4.7	-4.8	1.8	-0.4	21.3	2.8	11.0
Avg Abs Error	3.4	4.6	1.9	4.8	2.1	2.1	10.7	5.1	4.9
RMS Error	4.1	5.9	2.5	5.1	2.4	2.6	13.2	6.3	6.0

c. Field Shape Factor 4

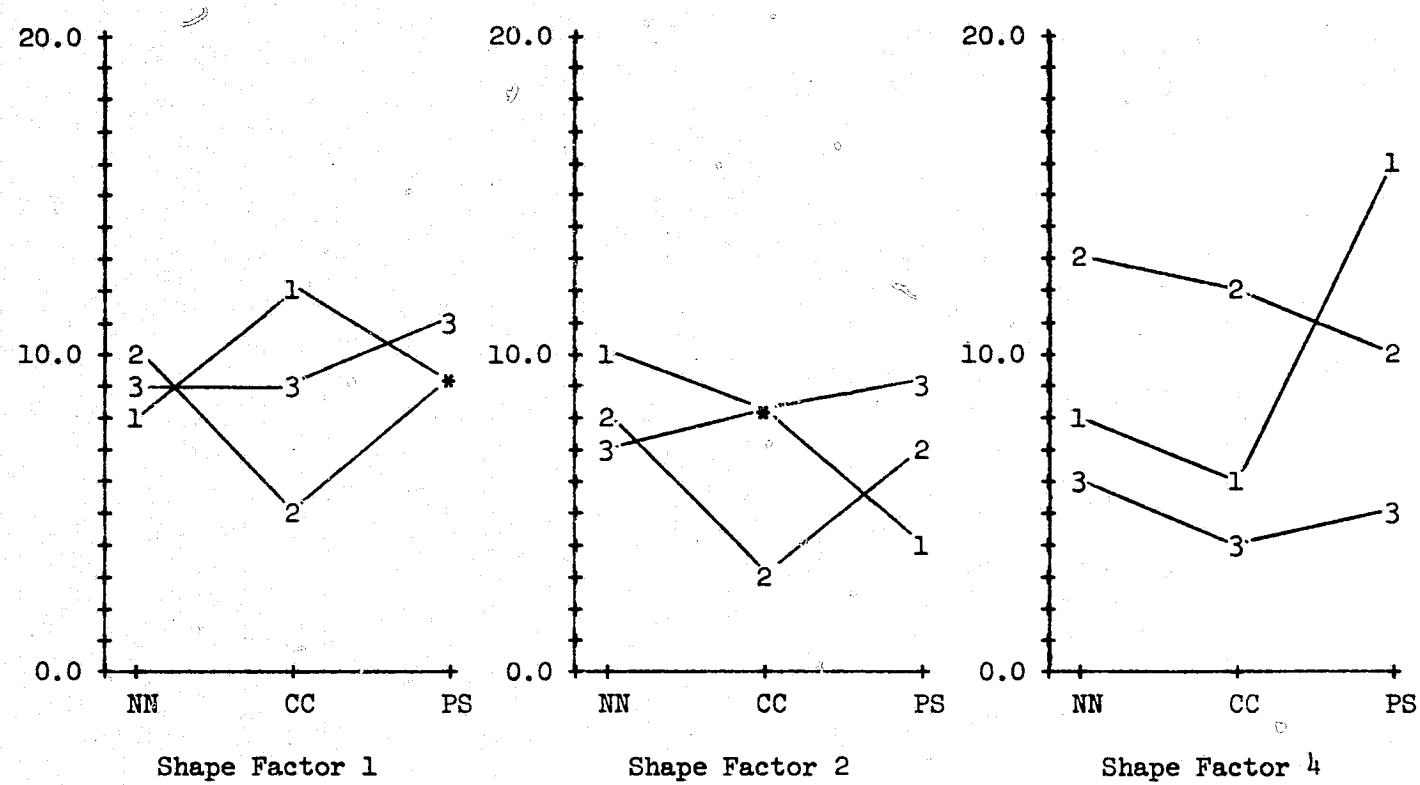


Figure 5.3-1 Estimated Root Mean Square Proportion Errors for 2.5 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)

ORIGINAL PAGE IS
OF POOR QUALITY

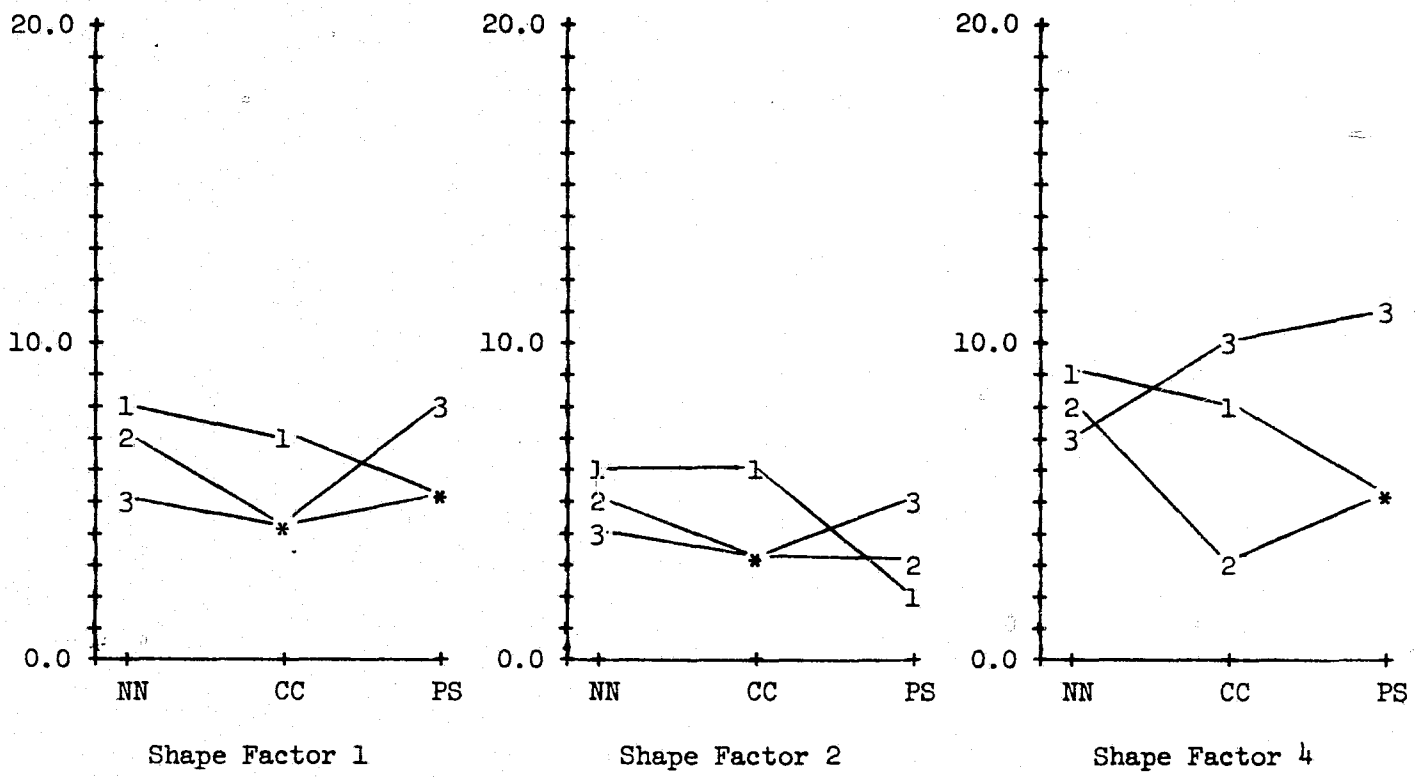


Figure 5.3-2 Estimated Root Mean Square Proportion Errors for 5.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)

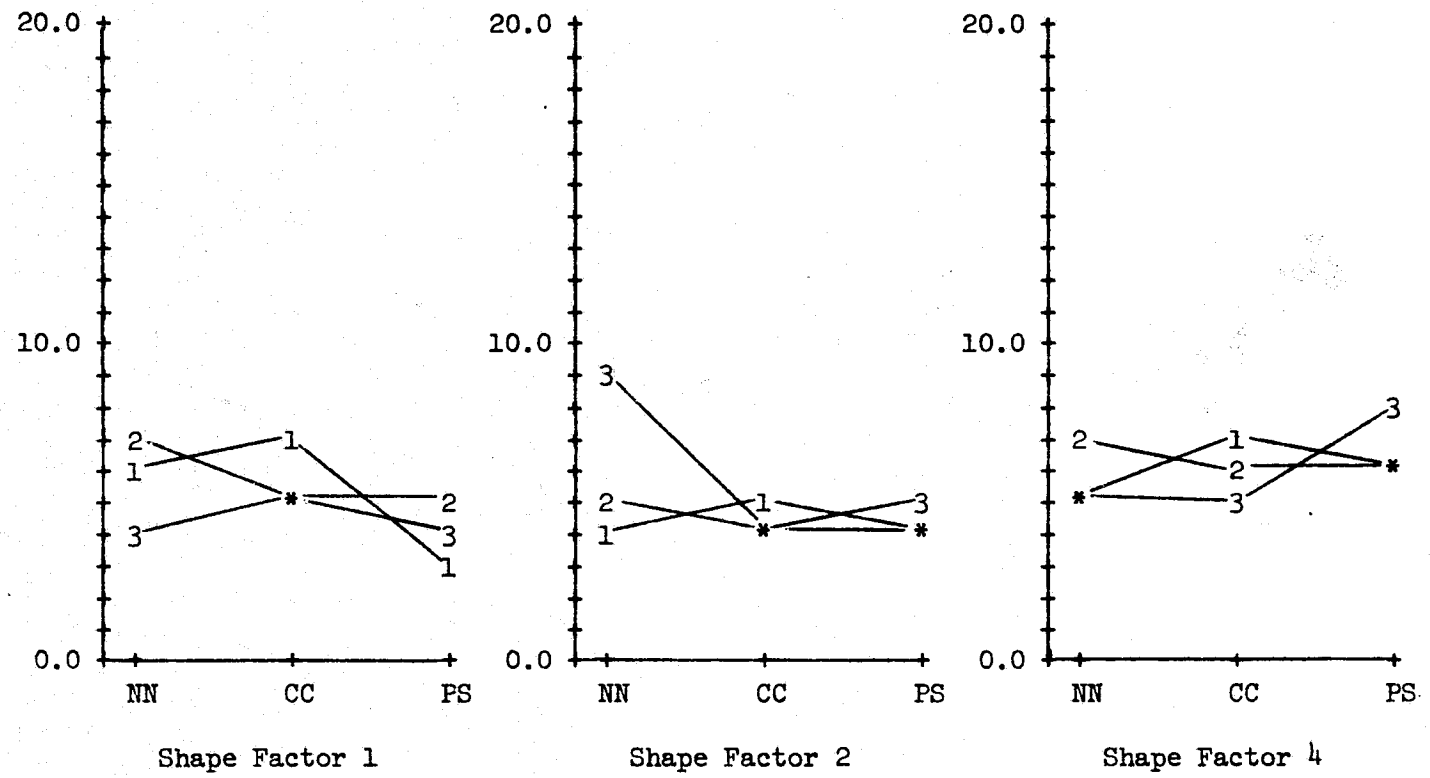


Figure 5.3-3 Estimated Root Mean Square Proportion Errors for 10.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)

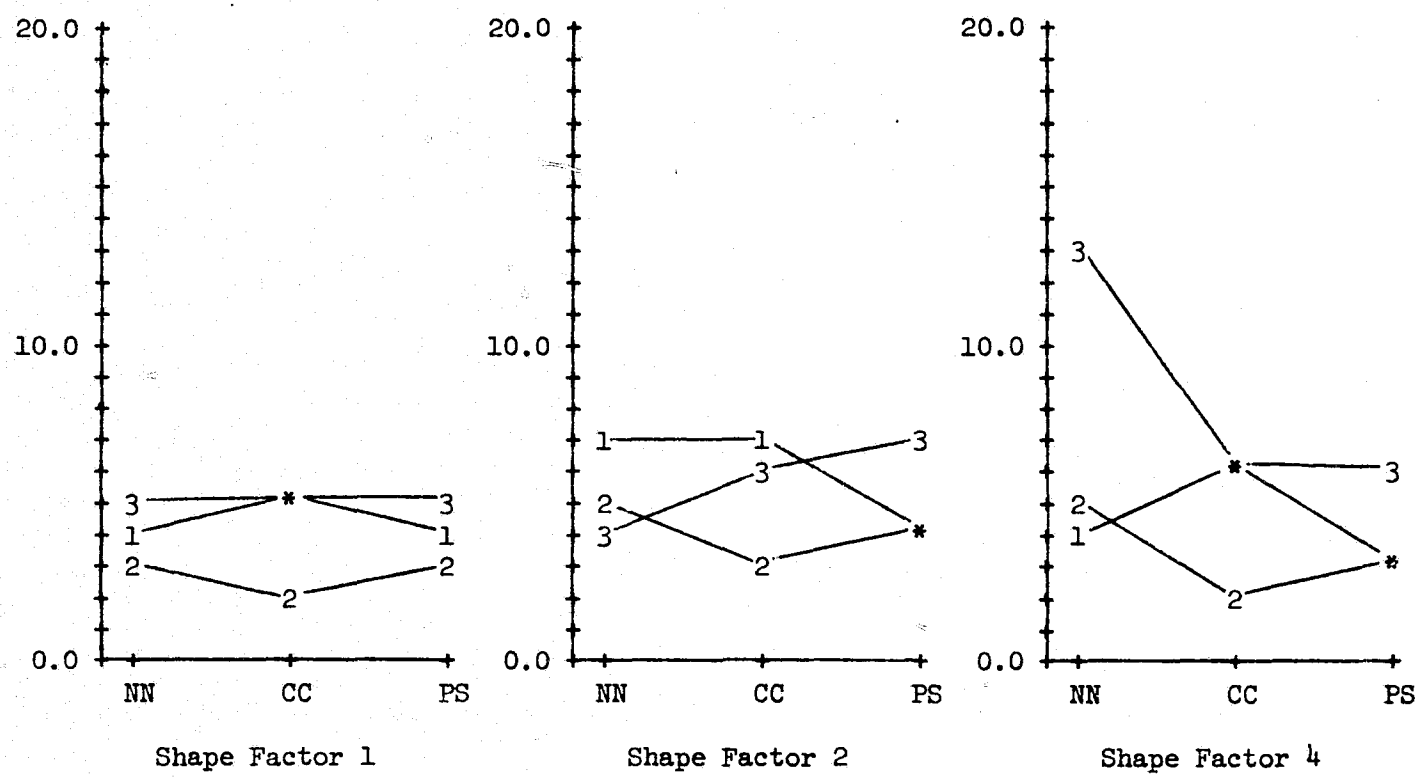


Figure 5.3-4 Estimated Root Mean Square Proportion Errors for 20.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (Equally Likely Classes)

ORIGINAL PAGE IS OF POOR QUALITY

TM Configuration	Root Mean Square Error Averaged over Field Sizes and Shapes NN	CC	PS
1	6.6	7.0	5.4
2	6.9	4.4	5.6
3	6.6	5.9	7.0

Tables 5.3-5 through 5.3-8 give values of the estimated proportion errors for an image containing 50 percent corn rather than equally likely classes. The root mean squares of the estimated proportion errors are found in Figures 5.3-5 through 5.3-8. The dependence of the distribution of proportion on the true class proportions is seen readily by comparing these figures and tables with those for the image containing equally large classes. Errors are generally larger in the 50 percent corn image and, more importantly, there is no clearly superior TM configuration and resampling technique combination. The following table shows the average of the root mean square errors, over all field sizes and shapes, for a 50 percent corn image.

TM Configuration	Root Mean Square Errors Averaged over Field Sizes and Shapes NN	CC	PS
1	11.5	14.1	10.6
2	11.1	10.8	13.2
3	14.0	13.0	12.4

TM configuration 1 with point-spread resampling and TM configuration 2 with cubic convolution resampling are virtually identical according to this criterion. Other combinations are quite similar to these.

A recommendation of the best TM configuration and resampling technique based upon the proportion estimation criterion would depend upon the exact purpose of the sensor system. For a corn monitoring system one might use a combination quite different from that for a wheat monitoring system. However, based on this analysis of the results for a single image, the proportion error criterion points weakly to a choice of TM configuration 2 with cubic convolution resampling. This choice is strengthened by the similar conclusion from the classification analysis evaluation.

5.3.2 Aircraft Image Analysis

Tables 5.3-9 through 5.3-11 give values of the estimated average proportion errors for an aircraft-type image with equally likely classes. The values of the root mean square criterion are given in Figures 5.3-9 through 5.3-11. Similar values for an aircraft-type image in the 50 percent corn scenario are given in Tables 5.3-12 through 5.3-14 and Figures 5.3-12 through 5.3-14. In these tables and figures, the same field shape and field size quantization is employed as was used in Section 5.2.2. Shape factor set A consists of fields whose shape factor is 1.5 or less, set B consists of fields whose shape factors are in the range from 1.5 to 4.0, and set C consists of those fields whose shape factor is greater than 4.0. The smaller fields are those containing less than 4000 aircraft pixels, the intermediate fields are those containing 4000 to 8000 aircraft pixels, and

Table 5.3-5 Estimated Average Proportion Errors for 2.5 Acre Fields of Simulated Image (50% Corn Scenario)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	15.1	41.1	11.2	9.5	5.1	6.4	15.9	32.6	11.8
Pasture	10.6	1.1	13.0	20.6	8.5	25.1	1.5	0.6	-2.3
Corn	-19.6	-41.1	-22.5	-16.8	-14.1	-25.8	-28.8	-40.8	-23.4
Wheat	-1.1	1.4	-15.0	-8.1	-0.4	-6.2	-4.4	-4.4	-10.5
Soybeans	-5.0	-2.6	13.3	-5.2	1.0	0.4	15.8	12.1	24.4
Avg Abs Error	10.3	17.5	15.0	12.1	5.8	12.8	13.3	18.1	14.5
RMS Error	12.2	26.0	15.5	13.4	7.7	16.6	16.4	24.1	16.7

a. Field Shape Factor 1

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	25.0	16.2	5.2	11.0	11.7	3.4	19.9	16.2	8.7
Pasture	9.4	-1.1	7.4	10.0	11.8	25.2	2.0	-3.3	-7.1
Corn	-20.1	-15.9	-19.5	-12.9	-17.9	-16.8	-18.0	-15.7	-18.0
Wheat	-9.3	4.5	-6.2	-1.3	-3.6	-7.4	-9.4	-0.6	-6.5
Soybeans	-5.0	-3.8	13.0	-6.9	-2.0	-4.4	5.6	3.3	22.9
Avg Abs Error	13.8	8.3	10.3	8.4	9.4	11.5	11.0	7.8	12.6
RMS Error	15.7	10.5	11.6	9.4	11.1	14.2	13.0	10.3	14.2

b. Field Shape Factor 2

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	7.6	13.1	-6.1	-2.1	-2.9	-3.4	14.7	4.6	-2.0
Pasture	20.6	14.5	14.8	38.5	38.7	36.4	16.6	15.9	13.6
Corn	-18.5	-21.1	-10.8	-23.2	-23.9	-24.8	-29.0	-18.5	-19.5
Wheat	-12.0	-9.6	-20.0	-11.4	-12.0	-11.3	-10.7	-9.3	-1.3
Soybeans	2.3	3.0	22.0	-1.8	0.2	3.1	8.4	7.3	9.2
Avg Abs Error	12.2	12.3	14.8	15.4	15.5	15.8	15.9	11.1	9.1
RMS Error	14.0	13.6	15.9	20.8	21.1	20.4	17.4	12.3	11.5

c. Field Shape Factor 4

Table 5.3-6 Estimated Average Proportion Errors for 5.0 Acre Fields of Simulated Image (50% Corn Scenario)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	14.9	22.3	9.1	5.7	8.7	3.4	16.2	17.4	11.7
Pasture	11.5	5.5	6.2	17.8	15.5	22.0	3.8	5.1	5.5
Corn	-14.3	-20.5	-15.1	-15.9	-19.8	-22.2	-19.6	-20.5	-20.5
Wheat	-7.8	-3.1	-11.3	-5.1	-3.7	-4.4	-6.1	-5.0	-11.7
Soybeans	-4.3	-4.3	11.1	-2.5	-0.8	1.2	5.7	3.0	15.0
Avg Abs Error	10.6	11.1	10.6	9.4	9.7	10.6	10.3	10.2	12.9
RMS Error	11.3	14.0	10.9	11.2	12.0	14.2	12.1	12.5	13.8

a. Field Shape Factor 1

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	7.1	13.9	0.9	5.1	4.6	4.6	3.1	3.4	2.6
Pasture	13.7	1.3	3.3	13.0	9.4	11.4	1.0	1.0	1.0
Corn	-21.8	-25.2	-16.2	-17.8	-19.5	-22.6	-23.2	-20.8	-22.6
Wheat	1.0	3.0	1.4	1.3	3.0	2.9	6.7	8.5	-2.1
Soybeans	-0.0	7.0	10.6	-1.6	2.5	3.7	12.4	7.9	21.0
Avg Abs Error	8.7	10.1	6.5	7.7	7.8	9.0	9.3	8.3	9.9
RMS Error	11.9	13.3	8.8	10.1	10.0	11.7	12.2	10.8	13.9

b. Field Shape Factor 2

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	17.8	16.6	-1.7	8.7	2.3	4.0	20.7	28.3	3.2
Pasture	8.6	-0.8	9.0	17.8	8.3	21.4	-0.2	-0.5	-1.6
Corn	-24.8	-14.9	-16.4	-21.8	-18.4	-28.1	-27.5	-28.8	-16.1
Wheat	3.3	0.3	-1.9	-0.5	8.5	3.8	-4.2	-9.4	-13.1
Soybeans	-5.0	-1.2	11.0	-4.3	-0.8	-1.2	11.2	10.3	27.6
Avg Abs Error	11.9	6.8	8.0	10.6	7.7	11.7	12.7	15.5	12.3
RMS Error	14.4	10.0	9.8	13.3	9.9	16.0	16.3	19.1	15.5

c. Field Shape Factor 4

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.3-7 Estimated Average Proportion Errors for 10.0 Acre Fields for Simulated Image (50% Corn Scenario)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	16.2	23.2	9.6	5.9	6.2	10.8	15.0	16.8	8.3
Pasture	8.1	4.5	3.5	15.1	12.9	12.4	5.3	3.2	1.7
Corn	-17.9	-22.3	-11.4	-13.5	-20.9	-19.6	-17.6	-15.8	-10.0
Wheat	-4.6	-4.1	-4.6	-4.3	-3.9	-3.5	-4.9	-3.3	-6.6
Soybeans	-1.9	-1.2	2.8	-3.3	5.8	-0.1	2.2	-0.8	6.6
Avg Abs Error	9.7	11.1	6.4	8.4	9.9	9.3	9.0	8.0	6.6
RMS Error	11.6	14.7	7.3	9.7	11.8	11.5	10.9	10.5	7.2

a. Field Shape Factor 1

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	6.1	7.9	-2.0	-0.3	-0.1	-1.8	1.7	7.0	-0.7
Pasture	8.6	7.5	10.8	14.1	12.4	16.4	12.3	3.0	2.9
Corn	-20.2	-24.8	-18.8	-20.3	-21.1	-21.6	-16.4	-20.6	-16.2
Wheat	5.7	7.7	2.5	8.4	7.8	5.2	4.6	7.2	0.2
Soybeans	-0.1	1.7	7.5	-1.9	1.0	1.8	-2.3	3.4	13.8
Avg Abs Error	8.1	9.9	8.3	9.0	8.5	9.3	7.5	8.3	6.8
RMS Error	10.5	12.6	10.4	11.7	11.5	12.4	9.5	10.5	9.6

b. Field Shape Factor 2

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	7.3	20.4	0.6	2.5	5.0	8.7	9.7	15.5	2.1
Pasture	15.5	9.2	14.7	21.1	16.6	23.3	-0.5	-2.1	-4.0
Corn	-22.0	-30.6	-22.2	-20.7	-24.8	-32.2	-19.0	-19.0	-16.7
Wheat	-4.4	-1.4	-2.8	-4.1	-1.4	-3.7	-2.7	-1.6	-3.5
Soybeans	3.5	2.4	9.6	1.2	4.7	3.9	12.5	7.2	22.1
Avg Abs Error	10.6	12.8	10.0	9.9	10.5	14.4	8.9	9.1	9.7
RMS Error	12.7	17.0	12.7	13.4	13.7	18.4	11.1	11.5	12.6

c. Field Shape Factor 4

Table 5.3-8 Estimated Average Proportion Errors for 20.0 Acre Fields for Simulated Image (50% Corn Scenario)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	10.4	13.9	5.1	3.0	3.6	4.5	10.6	11.5	6.4
Pasture	3.7	-0.5	0.7	9.4	3.0	6.1	0.2	0.3	0.7
Corn	-14.4	-19.1	-13.3	-11.7	-12.9	-13.6	-13.8	-16.0	-12.6
Wheat	-2.6	1.2	-2.6	-2.5	1.3	-1.6	-3.3	-2.0	-6.2
Soybeans	2.9	4.5	10.2	1.9	5.0	4.6	6.3	6.2	11.7
Avg Abs Error	6.8	7.8	6.4	5.7	5.1	6.1	6.8	7.2	7.5
RMS Error	8.3	10.7	7.9	7.0	6.5	7.3	8.4	9.3	8.7

a. Field Shape Factor 1

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	6.8	19.5	4.5	4.2	6.3	6.9	5.5	14.5	9.2
Pasture	-0.5	1.0	2.8	8.7	1.1	7.4	-0.2	-0.8	-0.3
Corn	-13.8	-24.9	-15.7	-10.3	-15.3	-15.1	-14.2	-22.0	-22.1
Wheat	-0.2	5.4	-1.3	0.3	5.2	1.7	-0.6	0.5	-4.1
Soybeans	-2.3	-1.0	9.8	-2.9	2.6	-0.9	9.5	7.8	17.4
Avg Abs Error	4.7	10.4	6.8	5.3	6.1	6.4	6.0	9.1	10.6
RMS Error	7.0	14.4	8.6	6.4	7.9	8.2	8.1	12.3	13.4

b. Field Shape Factor 2

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Trees	6.6	14.8	0.9	3.4	0.8	1.3	3.6	15.2	2.1
Pasture	4.2	-0.5	3.2	10.5	4.5	8.4	-1.9	-1.6	-1.4
Corn	-15.3	-23.2	-14.6	-12.0	-13.0	-13.0	-48.8	-24.3	-16.4
Wheat	3.3	6.4	0.5	-0.3	4.1	0.7	-5.2	4.2	-3.7
Soybeans	1.2	2.5	10.0	-1.6	3.5	2.7	52.2	6.5	19.4
Avg Abs Error	6.1	9.5	5.8	5.6	5.2	5.2	22.3	10.4	8.6
RMS Error	7.8	12.7	8.1	7.3	6.6	7.1	32.1	13.3	11.5

c. Field Shape Factor 4

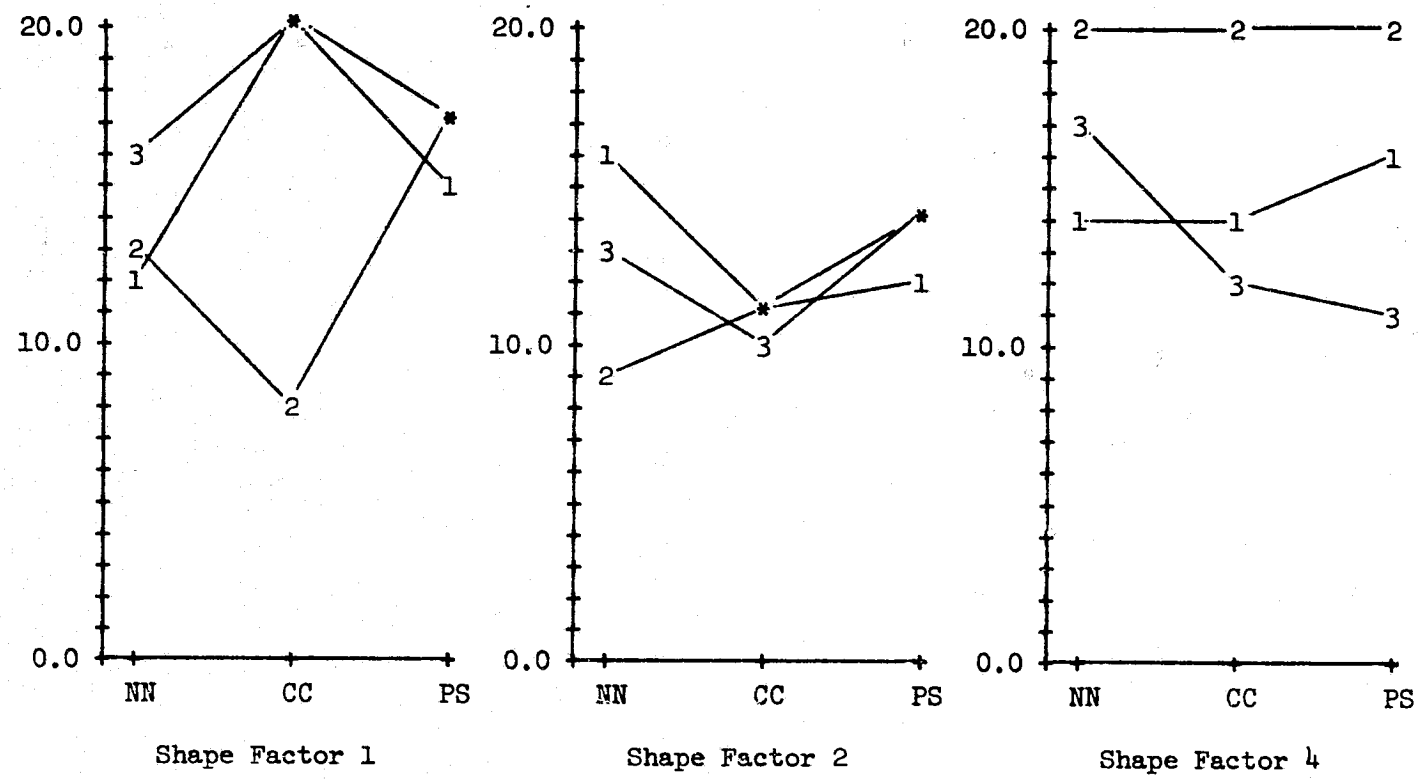


Figure 5.3-5 Estimated Root Mean Square Proportion Errors for 2.5 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50% Corn Scenario)

ORIGINAL PAGE IS
OF POOR QUALITY

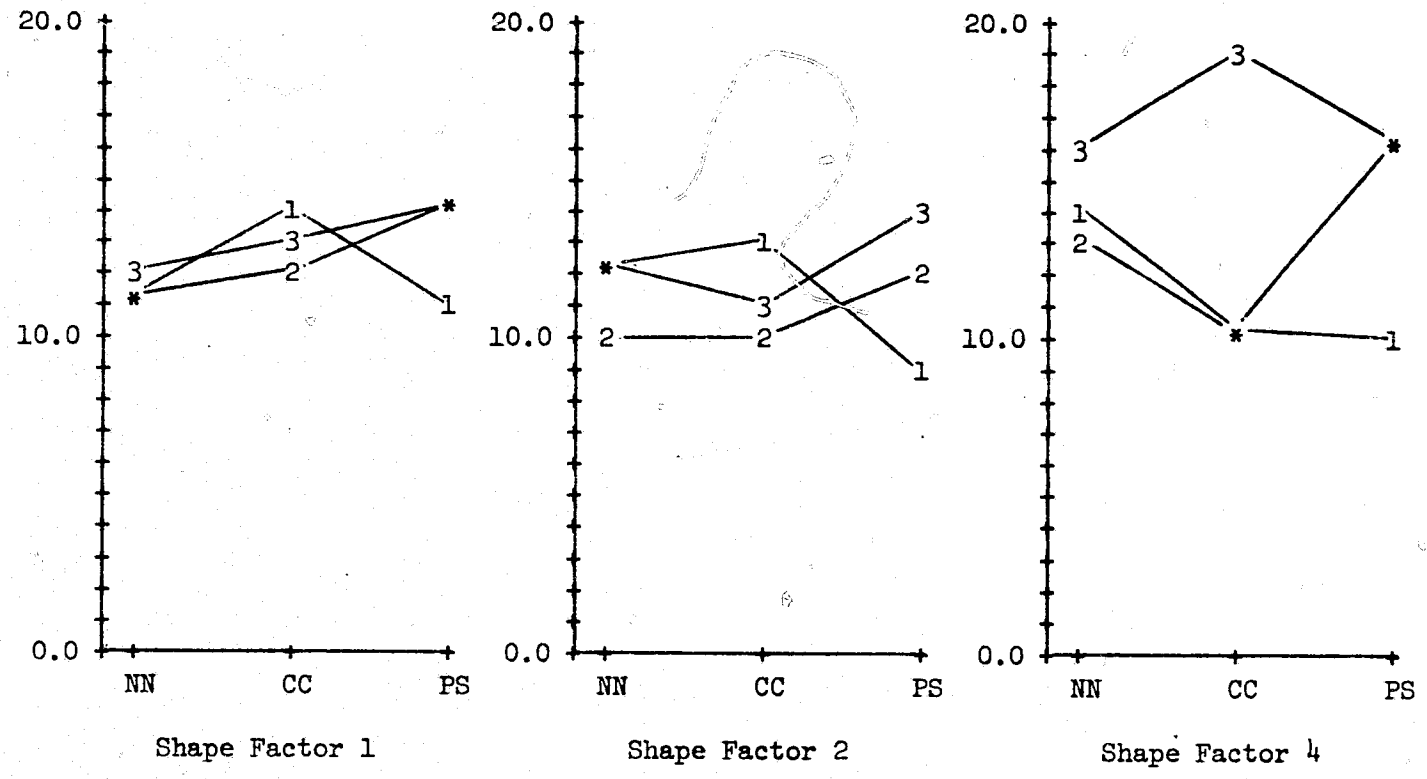


Figure 5.3-6 Estimated Root Mean Square Proportion Errors for 5.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50% Corn Scenario)

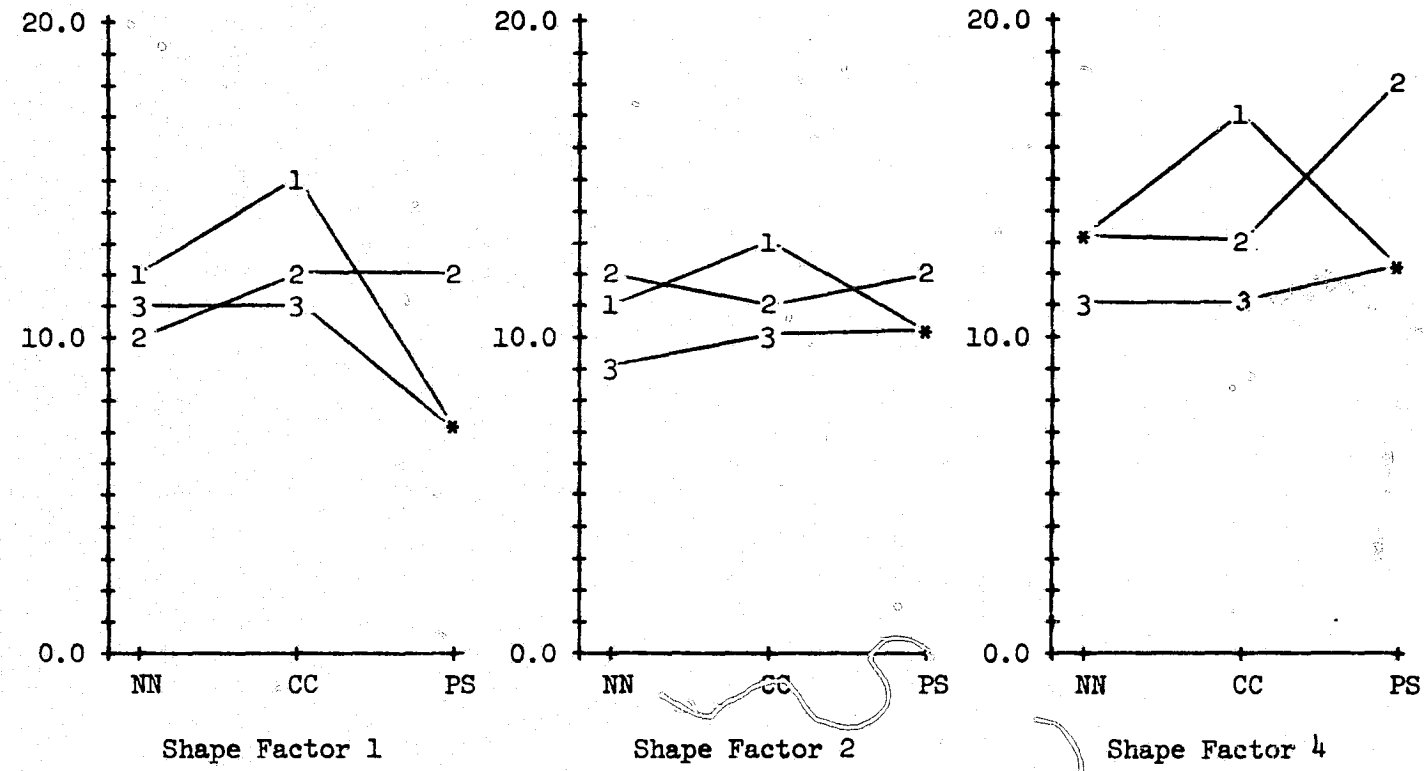


Figure 5.3-7 Estimated Root Mean Square Proportion Errors for 10.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50% Corn Scenario)

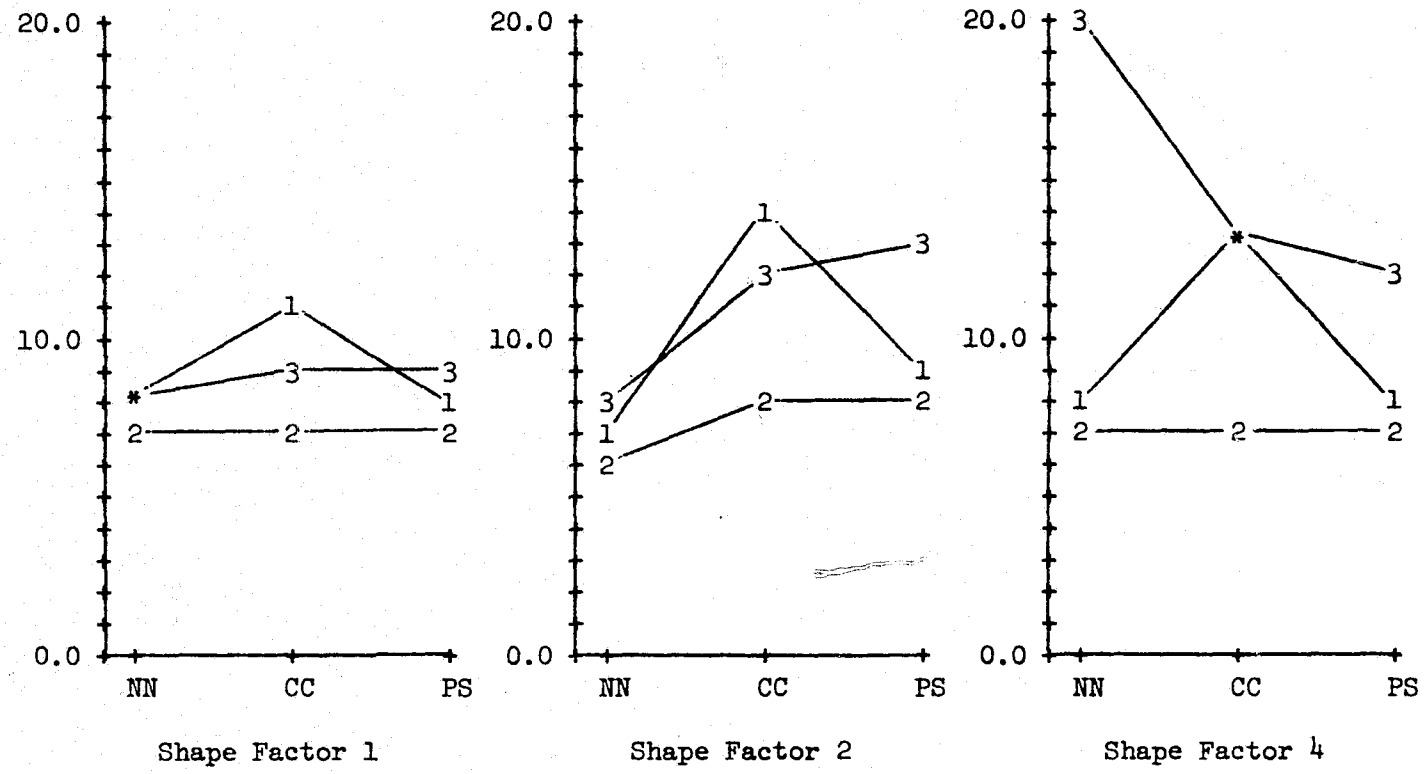


Figure 5.3-8 Estimated Root Mean Square Proportion Errors for 20.0 Acre Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Synthetic Data Set (50% Corn Scenario)

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.3-9 Estimated Average Proportion Errors for Smaller Fields in Aircraft Image Derived TM Images (Equally Likely Classes)

Class	Configuration 1			Configuration 2			Configuration 3		
	NN	CC	PS	NN	CC	PS	NN	CC	PS
Corn	2.2	3.0	2.5	0.7	0.2	1.8	-2.9	-2.5	-16.8
Trees/Pasture	-9.8	-11.0	-5.0	-8.8	-2.7	-3.6	-2.6	-12.9	5.5
Small Grains/Soy	4.8	-3.7	6.9	8.7	1.6	2.9	-0.2	-0.8	1.6
Set Aside/Nonfarm	2.8	11.6	-4.4	-0.7	0.9	-1.1	5.7	16.3	9.8
Avg Abs Error	4.9	7.3	4.7	4.7	1.3	2.4	2.8	8.1	8.4
RMS Error	5.8	8.3	5.0	6.2	1.6	2.5	3.4	10.5	10.1

a. Field Shape Factor Set A

Class	Configuration 1			Configuration 2			Configuration 3		
	NN	CC	PS	NN	CC	PS	NN	CC	PS
Corn	-1.8	-2.0	-2.0	-4.1	-6.3	-1.2	-4.8	-6.9	-8.8
Trees/Pasture	-7.2	-5.7	1.8	-7.3	-8.4	0.4	-1.7	-5.3	6.7
Small Grains/Soy	14.9	10.6	16.4	19.4	19.9	17.3	12.0	11.7	11.5
Set Aside/Nonfarm	-5.8	-3.0	-16.1	-8.1	-5.3	-16.6	-5.5	0.6	-9.4
Avg Abs Error	7.4	5.3	9.1	9.7	10.0	8.9	6.0	6.1	9.1
RMS Error	8.8	6.3	11.6	11.3	11.5	12.0	7.1	7.3	9.3

b. Field Shape Factor Set B

Class	Configuration 1			Configuration 2			Configuration 3		
	NN	CC	PS	NN	CC	PS	NN	CC	PS
Corn	-5.1	-2.5	-6.1	-20.3	-21.1	-17.5	-5.2	-8.9	-15.0
Trees/Pasture	2.4	1.6	3.3	1.4	1.5	3.5	2.0	3.3	18.7
Small Grains/Soy	7.1	7.2	10.8	24.2	25.8	18.1	8.0	9.4	6.9
Set Aside/Nonfarm	-4.4	-6.3	-8.0	-5.3	-6.2	-4.1	-4.8	-3.9	-10.6
Avg Abs Error	4.7	4.4	7.0	12.8	13.6	10.8	5.0	6.4	12.8
RMS Error	5.0	5.0	7.5	16.0	17.0	12.9	5.4	6.9	13.5

c. Field Shape Factor Set C

Table 5.3-10 Estimated Average Proportion Errors for Intermediate Fields in Aircraft Image Derived TM Images (Equally Likely Classes)

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Corn	-0.5	0.1	-0.5	-3.4	-3.8	-0.2	-3.5	-1.9	-13.2
Trees/Pasture	-3.2	-2.0	3.3	-4.2	-2.4	4.8	-2.8	-4.1	8.5
Small Grains/Soy	11.0	8.2	10.3	12.6	12.6	7.8	11.7	10.1	7.3
Set Aside/Nonfarm	-7.4	-6.3	-13.1	-5.0	-6.4	-12.4	-5.4	-4.1	-2.6
Avg Abs Error	5.5	4.2	6.8	6.3	6.3	6.3	5.9	5.1	7.9
RMS Error	6.8	5.3	8.5	7.3	7.4	7.7	6.8	5.9	8.8

a. Field Shape Factor Set A

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Corn	0.3	2.2	0.0	1.2	1.5	2.7	0.8	0.0	0.0
Trees/Pasture	3.8	2.6	2.6	2.7	3.3	5.4	8.1	5.8	5.0
Small Grains/Soy	-4.2	-6.1	-2.9	-5.1	-5.1	-8.0	-9.7	-7.2	-5.0
Set Aside/Nonfarm	0.0	1.3	0.3	1.2	0.3	0.0	0.8	1.4	0.0
Avg Abs Error	2.1	3.0	1.4	2.5	2.5	4.0	4.9	3.6	2.5
RMS Error	2.8	3.5	1.9	3.0	3.1	5.0	6.3	4.7	3.5

b. Field Shape Factor Set B

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Corn	-9.7	-6.4	-8.5	-8.9	-7.1	-7.1	-10.0	-5.9	-15.7
Trees/Pasture	-9.7	-10.3	-7.7	-12.3	-11.4	-10.0	-5.3	-7.2	0.1
Small Grains/Soy	16.2	11.5	13.8	18.2	15.6	15.0	10.8	7.3	13.7
Set Aside/Nonfarm	3.2	5.2	2.4	3.0	3.0	2.1	4.5	5.8	1.9
Avg Abs Error	9.7	8.3	8.1	10.6	9.3	8.5	7.7	6.6	7.8
RMS Error	10.7	8.8	9.0	11.9	10.4	9.7	8.2	6.6	10.5

c. Field Shape Factor Set C

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.3-11 Estimated Average Proportion Errors for Larger Fields in Aircraft Image Derived TM Images (Equally Likely Classes)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	-3.6	-4.2	-3.0	-8.7	-7.6	-4.2	-7.5	-7.5	-14.0
Trees/Pasture	0.6	1.1	0.4	1.5	0.8	0.7	1.6	1.7	7.9
Small Grains/Soy	2.3	1.7	2.3	5.7	5.0	2.4	5.3	4.4	4.3
Set Aside/Nonfarm	0.6	1.4	0.4	1.5	1.8	1.1	0.6	1.5	1.7
Avg Abs Error	1.8	2.1	1.5	4.3	3.8	2.1	3.8	3.8	7.0
RMS Error	2.2	2.5	1.9	5.3	4.7	2.5	4.7	4.5	8.4

a. Field Shape Factor Set A

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	1.2	1.7	2.6	1.4	1.7	4.8	2.1	1.2	5.0
Trees/Pasture	-0.1	-3.9	2.3	5.5	3.6	7.6	-0.3	-0.8	1.6
Small Grains/Soy	-19.3	-20.9	-15.5	-16.1	-18.7	-21.5	-19.0	-20.1	-20.6
Set Aside/Nonfarm	18.1	23.1	10.5	9.1	13.3	9.1	17.2	19.7	14.1
Avg Abs Error	9.7	12.4	7.7	8.0	9.3	10.7	9.6	10.5	10.3
RMS Error	13.3	15.7	9.5	9.7	11.6	12.5	12.8	14.1	12.8

b. Field Shape Factor Set B

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	-7.2	-6.4	-6.1	-7.1	-7.3	-4.0	-7.2	-7.2	-12.5
Trees/Pasture	-1.8	-0.3	0.2	-3.9	-5.9	-1.9	0.6	-1.0	5.6
Small Grains/Soy	4.7	2.4	2.8	4.4	5.6	1.6	3.0	3.3	2.9
Set Aside/Nonfarm	4.2	4.3	3.1	6.6	7.6	4.4	3.6	5.0	4.0
Avg Abs Error	4.5	3.3	3.0	5.5	6.6	3.0	3.6	4.1	6.3
RMS Error	4.9	4.0	3.7	5.7	6.7	3.2	4.3	4.7	7.3

c. Field Shape Factor Set C

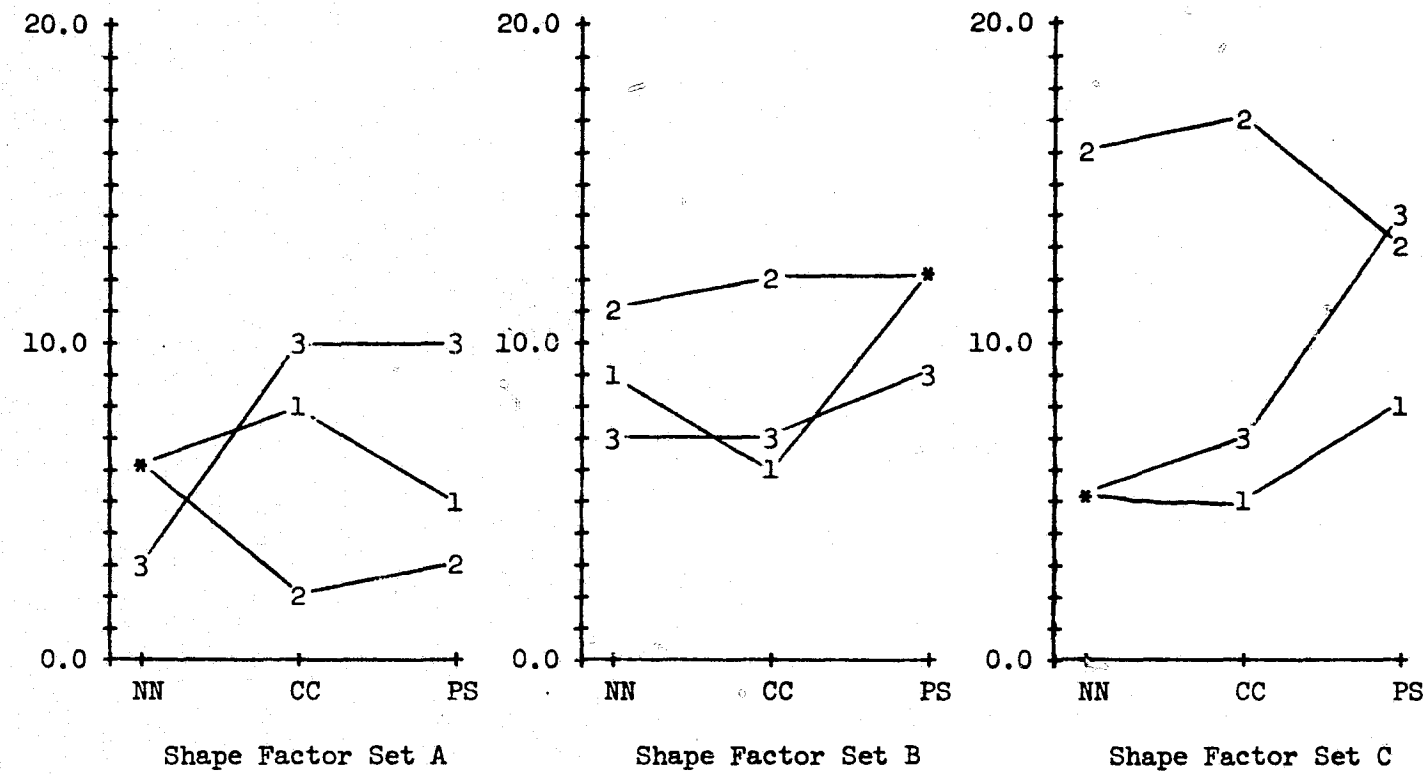


Figure 5.3-9 Estimated Root Mean Square Proportion Errors for Smaller Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (Equally Likely Classes)

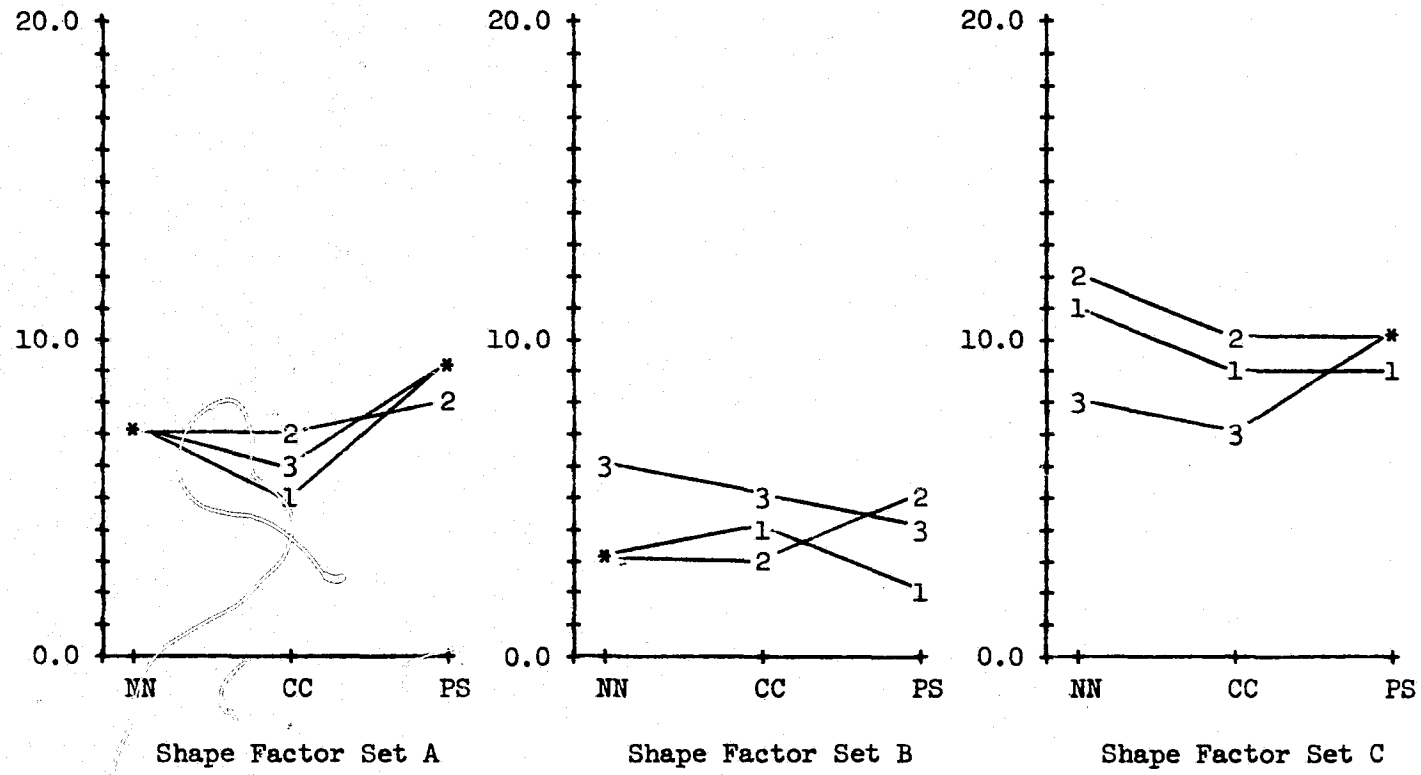


Figure 5.3-10 Estimated Root Mean Square Proportion Errors for Intermediate Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (Equally Likely Classes)

ORIGINAL PAGE IS OF POOR QUALITY

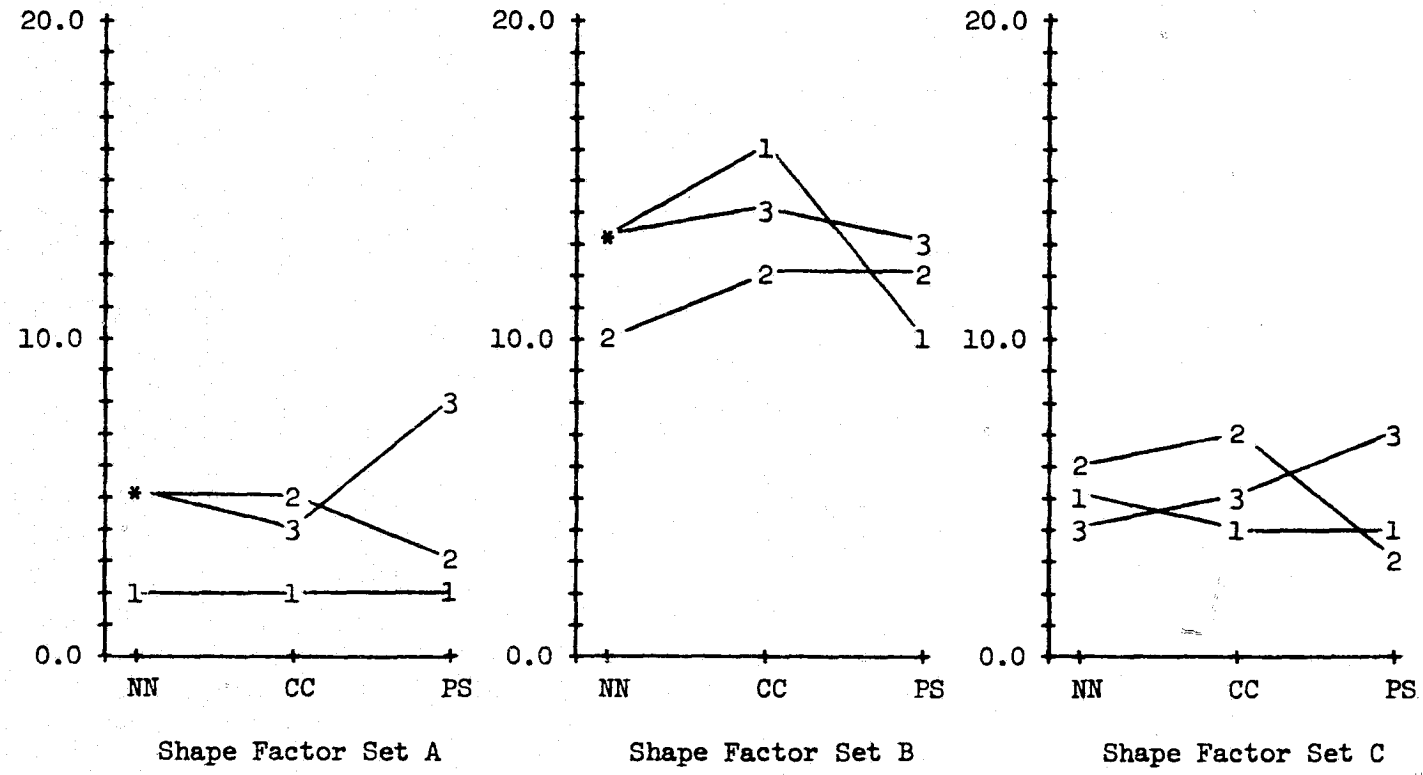


Figure 5.3-11 Estimated Root Mean Square Proportion Errors for Larger Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (Equally Likely Classes)

Table 5.3-12 Estimated Average Proportion Errors for Smaller Fields in Aircraft Image Derived TM Images (50% Corn Scenario)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	-5.0	-5.5	-2.8	-7.2	-8.0	-5.1	-13.9	-13.2	-37.2
Trees/Pasture	-9.0	-8.2	-3.5	-7.6	-2.4	-2.4	-0.3	-8.9	14.1
Small Grains/Soy	4.2	-3.3	4.1	6.7	2.3	3.0	1.1	-0.5	3.1
Set Aside/Nonfarm	9.7	17.0	2.1	8.1	8.1	4.4	13.1	22.6	20.0
Avg Abs Error	7.0	8.5	3.1	7.4	5.2	3.7	7.1	11.3	18.6
RMS Error	7.4	10.0	3.2	7.4	5.9	3.9	9.6	13.8	22.3

a. Field Shape Factor Set A

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	-4.9	-4.5	-4.5	-9.4	-13.3	-3.2	-12.2	-16.3	-18.3
Trees/Pasture	-6.9	-3.7	1.7	-8.0	-8.1	-1.4	2.9	-3.9	12.9
Small Grains/Soy	7.9	5.2	7.8	10.3	11.3	8.0	7.5	6.1	5.7
Set Aside/Nonfarm	3.8	3.0	-5.0	7.0	10.2	-3.5	1.8	14.7	-0.3
Avg Abs Error	5.9	4.1	4.7	8.7	10.7	4.0	6.1	10.4	9.3
RMS Error	6.1	4.2	5.2	8.8	10.9	4.7	7.3	11.7	11.6

b. Field Shape Factor Set B

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	-11.0	-10.8	-16.4	-42.7	-44.4	-36.4	-14.0	-20.6	-30.7
Trees/Pasture	4.0	1.6	2.3	2.9	3.1	6.3	3.4	3.0	24.5
Small Grains/Soy	6.8	8.7	17.3	34.3	34.0	23.6	11.7	13.1	9.6
Set Aside/Nonfarm	0.3	0.4	-3.2	5.6	7.3	6.5	-1.1	4.4	-3.4
Avg Abs Error	5.5	5.4	9.8	21.4	22.2	18.2	7.5	10.3	17.0
RMS Error	6.8	7.0	12.1	27.6	28.2	22.1	9.3	12.5	20.3

c. Field Shape Factor Set C

Table 5.3-13 Estimated Average Proportion Errors for Intermediate Fields in Aircraft Image Derived TM Images (50% Corn Scenario)

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	-5.8	-4.2	-4.2	-10.1	-11.2	-4.7	-12.4	-10.6	-27.9
Trees/Pasture	-3.6	-0.6	1.2	-5.9	-3.9	0.6	-1.3	-2.8	17.4
Small Grains/Soy	7.8	4.5	5.7	9.6	10.6	4.9	9.9	8.2	6.5
Set Aside/Nonfarm	1.6	0.4	-2.7	6.4	4.5	-0.7	3.8	5.3	4.0
Avg Abs Error	4.7	2.4	3.5	8.0	7.5	2.7	6.9	6.7	14.0
RMS Error	5.2	3.1	3.8	8.2	8.2	3.4	8.2	7.3	16.9

a. Field Shape Factor Set A

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	0.3	1.8	0.0	1.0	1.2	2.1	0.7	0.0	0.0
Trees/Pasture	3.1	2.1	2.1	2.1	2.6	4.3	6.4	4.7	4.0
Small Grains/Soy	-3.3	-4.9	-2.3	-4.0	-4.0	-6.4	-7.8	-5.8	-4.0
Set Aside/Nonfarm	0.0	1.0	0.3	1.0	0.2	0.0	0.7	1.1	0.0
Avg Abs Error	1.7	2.4	1.2	2.0	2.0	3.2	3.9	2.9	2.0
RMS Error	2.3	2.8	1.5	2.4	2.5	4.0	5.1	3.8	2.8

b. Field Shape Factor Set B

Class	Configuration 1 NN	CC	PS	Configuration 2 NN	CC	PS	Configuration 3 NN	CC	PS
Corn	-19.9	-12.8	-17.0	-18.3	-14.8	-14.8	-20.6	-12.4	-31.4
Trees/Pasture	-6.4	-6.3	-2.3	-9.8	-9.2	-5.2	1.3	-3.0	10.5
Small Grains/Soy	23.8	15.0	17.4	25.7	21.6	18.3	15.6	8.7	18.6
Set Aside/Nonfarm	2.5	4.1	1.9	2.4	2.4	1.7	3.6	6.8	2.3
Avg Abs Error	13.2	9.5	9.7	14.0	12.0	10.0	10.3	7.7	15.7
RMS Error	15.9	10.5	12.3	16.6	13.9	12.1	13.1	8.4	19.0

c. Field Shape Factor Set C

ORIGINAL PAGE IS
OF POOR QUALITY

Table 5.3-14 Estimated Average Proportion Errors for Larger Fields in Aircraft Image Derived TM Images (50% Corn Scenario)

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Corn	-7.1	-8.5	-6.1	-17.3	-15.2	-8.4	-15.0	-15.0	-28.0
Trees/Pasture	1.2	2.3	0.8	2.9	1.7	1.5	3.2	3.3	15.9
Small Grains/Soy	4.7	3.4	4.6	11.4	10.0	4.7	10.6	8.8	8.7
Set Aside/Nonfarm	1.2	2.8	0.8	3.0	3.5	2.2	1.2	3.0	3.5
Avg Abs Error	3.6	4.2	3.0	8.7	7.6	4.2	7.5	7.5	14.0
RMS Error	4.3	4.9	3.8	10.6	9.3	5.0	9.4	9.0	16.8

a. Field Shape Factor Set A

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Corn	1.0	1.3	2.1	1.1	1.4	3.8	1.7	1.0	4.0
Trees/Pasture	-0.0	-3.1	1.9	4.4	2.9	6.1	-0.2	-0.7	1.3
Small Grains/Soy	-15.4	-16.7	-12.4	-12.8	-14.9	-17.2	-15.2	-16.1	-16.5
Set Aside/Nonfarm	14.5	18.5	8.4	7.3	10.7	7.3	13.7	15.8	11.3
Avg Abs Error	7.7	9.9	6.2	6.4	7.5	8.6	7.7	8.4	8.2
RMS Error	10.6	12.6	7.6	7.7	9.3	10.0	10.3	11.3	10.2

b. Field Shape Factor Set B

Class	Configuration 1 NN	Configuration 1 CC	Configuration 1 PS	Configuration 2 NN	Configuration 2 CC	Configuration 2 PS	Configuration 3 NN	Configuration 3 CC	Configuration 3 PS
Corn	-14.9	-13.3	-12.1	-14.6	-15.1	-8.1	-15.4	-15.6	-25.7
Trees/Pasture	1.3	4.9	2.6	-1.7	-4.6	-1.2	5.5	2.8	15.9
Small Grains/Soy	8.3	3.8	5.1	6.2	8.3	2.4	5.4	5.0	5.0
Set Aside/Nonfarm	5.3	4.6	4.4	10.1	11.3	6.8	4.5	7.8	4.8
Avg Abs Error	7.5	6.7	6.1	8.2	9.8	4.6	7.7	7.8	12.9
RMS Error	9.0	7.7	7.0	9.4	10.6	5.4	8.9	9.2	15.5

c. Field Shape Factor Set C

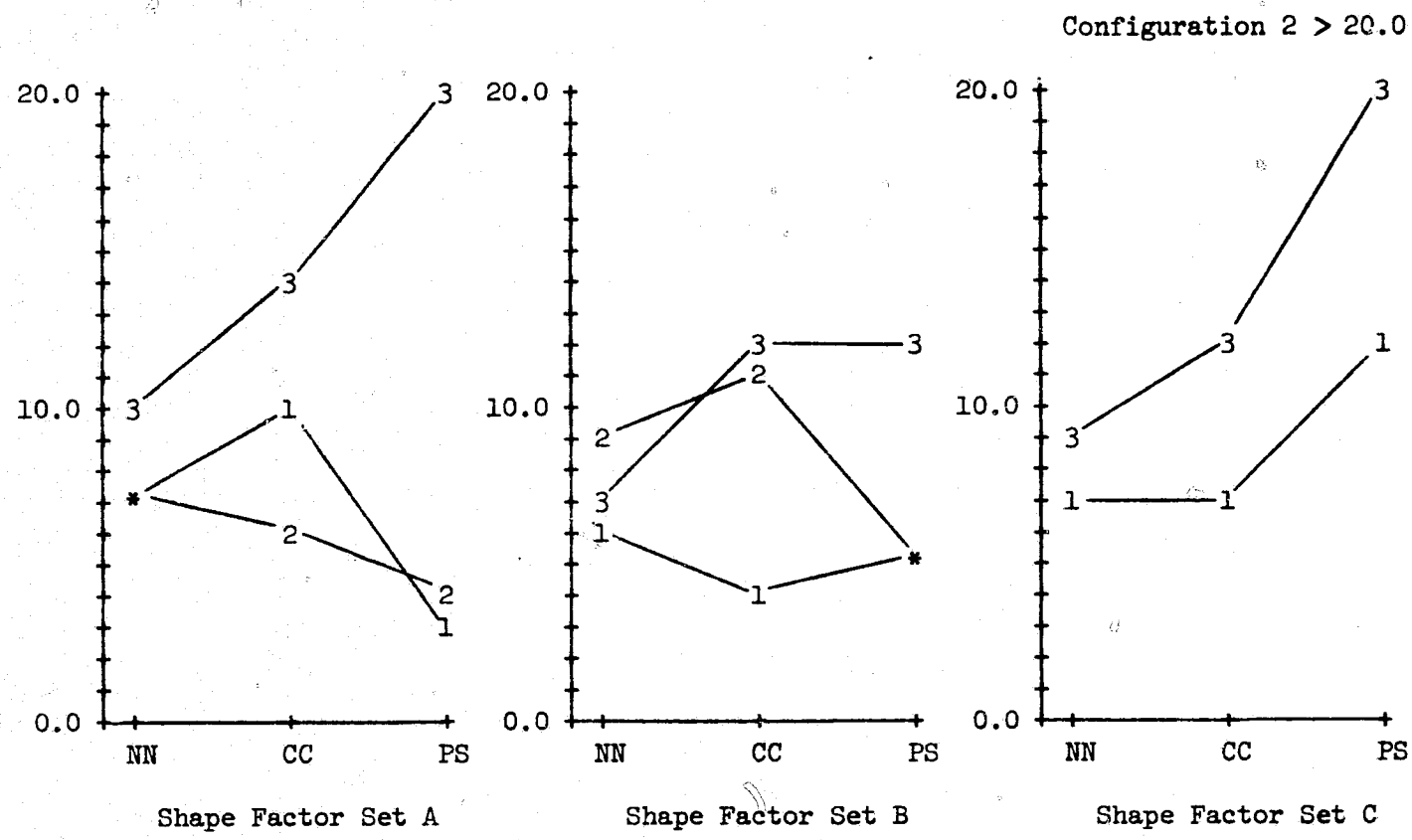


Figure 5.3-12 Estimated Root Mean Square Proportion Errors for Smaller Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (50% Corn Scenario)

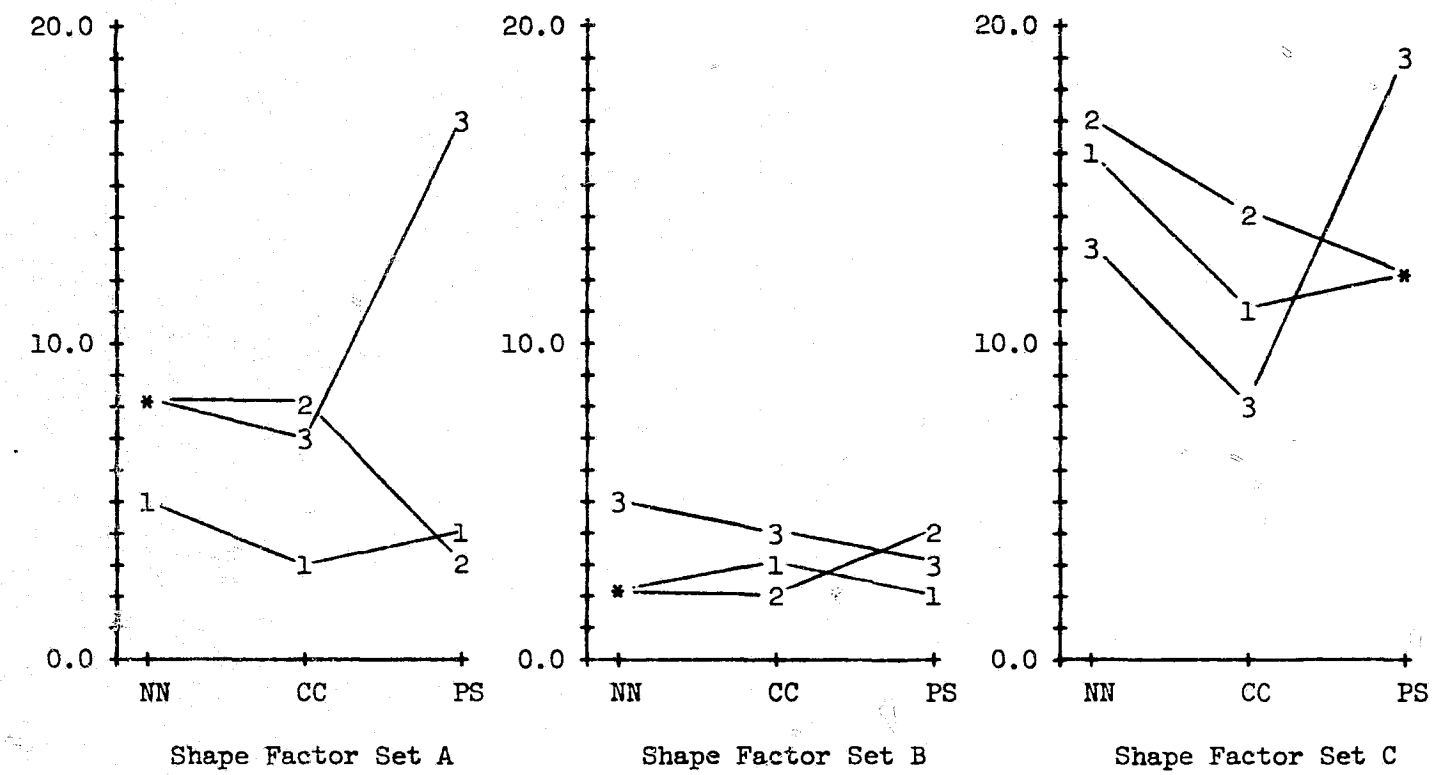


Figure 5.3-13 Estimated Root Mean Square Proportion Errors for Intermediate Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (50% Corn Scenario)

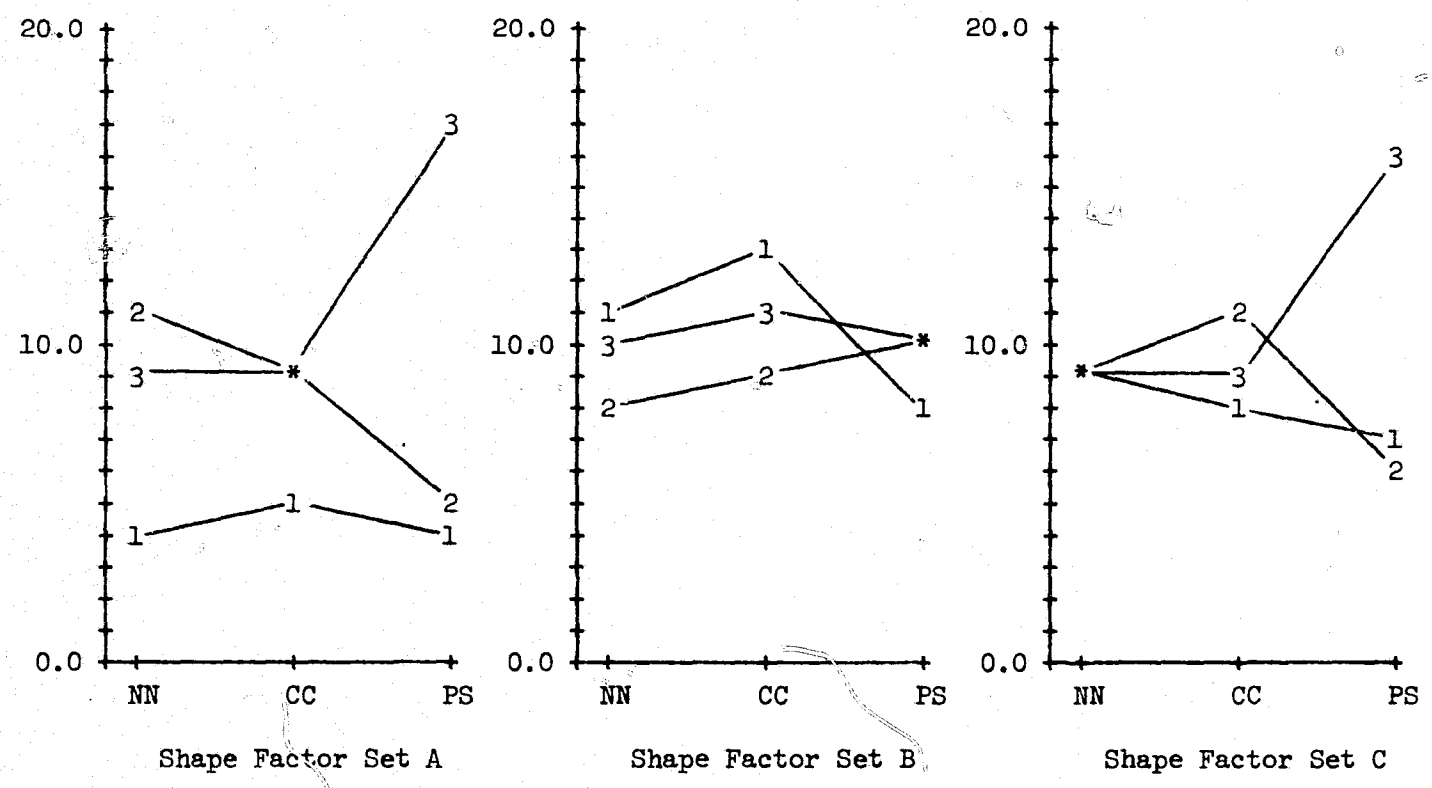


Figure 5.3-14 Estimated Root Mean Square Proportion Errors for Larger Fields Showing Effects of Field Shape Factor and TM Configuration for TM Images Derived from Aircraft Data Set (50% Corn Scenario)

ORIGINAL PAGE IS
OF POOR QUALITY

the larger fields are those containing more than 8000 aircraft pixels. Many of the values in these tables are not as reliable as those given in the corresponding tables for the synthetic image because of the missing field size and field shape combinations in the aircraft image. The following conclusions should therefore be relied upon cautiously.

The size of the proportion errors for the equally likely aircraft and synthetic images are roughtly the same. The proportion errors for the 50 percent corn scenario in the aircraft image are approximately the same as those for the equally likely case. This differs from the result obtained for the synthetic image and these two differing results provide a demonstration of the sensitivity of proportion estimation to the nature of the underlying class confusion matrices.

Again, there is a great deal of variation in the results obtained with the different TM configuration and resampling technique combinations, depending on class, field size, and field shape. In general, TM configuration 1 appears to be best (at least under a root mean square error criterion), but there is little real difference among the resampling techniques.

The following table shows the averages of the root mean square criterion from the aircraft image.

TM Configuration	Root Mean Square Errors Averaged Over Field Size and Shape: Equally Likely Classes NN	CC	PS	50 Percent Corn Image NN	CC	PS
1	6.70	6.60	6.51	4.91	5.16	4.32
2	8.49	8.22	7.56	7.94	8.24	5.68
3	6.56	7.24	9.36	6.09	7.50	10.74

As in the case with the evaluation of classification errors, it appears that TM configuration 1 is superior. There is little to choose among the resampling techniques, but because of the classification accuracy results, it would seem that point spread resampling should be chosen.

Section 6

CONCLUSIONS AND SUMMARY

The previous sections of this report have discussed the simulation processing and subsequent multispectral classification and analysis of the simulated data sets. This section provides an overall summary and discussion of the study results.

For the synthetic data set, the various resampling techniques were found to be approximately equivalent while for the real aircraft data set, in terms of classification accuracy, point spread of function-compensation resampling was superior to cubic convolution, which in turn was superior to nearest neighbor assignment. For the real aircraft data set, TM configuration 1 (Goldberg prefilter, 1.0 samples/IFOV) was preferred, while the synthetic data set, TM configuration 2 (Goldberg prefilter, 1.4 samples/IFOV) produced better classification results. In both cases, there was significant variation among results for the parameters considered, and considerable overlap among the results obtained with all of the TM configuration and resampling technique combinations.

Since the simulated data sets which were used in the classification had all been resampled onto a lattice of 1.0 sample/IFOV in both directions, their geometry was the same whether they were produced using TM configuration 1 or TM configuration 2. The different results obtained for the real aircraft data set and the synthetic data set, other things being equal, then depends only on this single parameter. ERIM, which acted as a consultant on this study has raised some questions on the issue of whether other things really are equal with respect to the synthetic data set. They have observed that the statistical description used in the creation of the synthetic data set was much simplified over that derived from the actual aircraft data set, and that this data could be expected to enhance the apparent classification performance obtained for the modeled Thematic Mapper configurations. These considerations are discussed in Appendix A, which presents the ERIM report on this study. In any case, they suggest that this characteristic of the synthetic data set could cause spurious differences among the nine TM configuration-resampling combinations.

In its report, ERIM also has questioned the method which was used to define field boundaries in the various related images, stating the belief that a non-integer method should be used for this. While this is a reasonable point to raise and is worthy of consideration in any continuation of this work, the constraints involved in performing multispectral classification on ten geometrically-related data sets within a facility not designed to deal with the difficulties of non-integer boundary specification in related digital images prevented the issue from being directly addressed in this study.

While this study has provided an indication of the superiority of higher order resampling techniques for processing digital imagery destined for multispectral classification, the absence of any great effect of resampling on the classification results obtained with the Thematic Mapper data derived from the synthetic data set suggests that it is not simply the case that higher order resamplers provide a superior

reconstruction of the data, but that they "shape" the data so that it is a better (in the case of configurations 1 and 2) or worse (in the case of configuration 3) match to the models on which the classifier is based. Since the synthetic data set was constructed on the basis of these models, it was already a good match to the classifier. With respect to the design of the Thematic Mapper sensor itself, the results obtained for the Thematic Mapper data derived from the synthetic data set indicate that a conventional prefilter/sampler design, oversampled at 1.4 samples/IFOV, yields better classification results. This choice of a sampling rate is in agreement with the intuitive idea that a higher sampling rate provides a better characterization of an analog signal. The results obtained for the Thematic Mapper data derived from the aircraft data set, while supporting the choice of a conventional prefilter/sampler design, indicate that an along-scan sampling rate of 1.0 samples/IFOV is preferred. However, in the light of the above comments on the apparent "shaping" effect of the higher order resamplers, and the similarity of the classification results for the two sampling rates when the point-spread-function compensation resampler is employed, the disagreement between the results for the two data sets is less surprising. Because of the thorough statistical analysis which was possible with the synthetic data set, the comparative results obtained with this data set are persuasive, and the choice of a higher sampling rate seems justified. The results obtained in examination of proportion errors for the Thematic Mapper data derived from both the synthetic and aircraft data sets are consistent with this choice for the design of the Thematic Mapper sensor.

It should also be observed that, in the opinion of IBM-Earth Resources Laboratory classification specialists, further consideration of the simulated data sets produced in this study should provide considerable illumination of the effects of the sensor system on data to be used for multispectral classification. These data sets provide the opportunity to investigate in detail the statistical characteristics of various classes whose initial structure can be well known, and to follow the transformation of these characteristics as the image is processed through the sensor system.

The results obtained in this study are derived from work with only two data sets. Because of the notorious scene-to-scene variation in results from classification studies, these results should not be interpreted as a valid indication of the performance to be expected from the Landsat-D Thematic Mapper. They represent a extensive comparison of the variation of classification performance against typical multispectral data sets when the variables are the parameters of the sensor system and the ground processing.

REFERENCES

1. Duck, Kenneth I., GSFC Memorandum, "Transformation of ERIM Multispectral Scanner Data to Emulate Thematic Mapper Data"; February 8, 1977.

2. Point Spread Function Compensation Resampling, Thematic Mapper Design Parameter Investigation proposal, IBM Corporation, April 1976.

3. Goodman, L. A., The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications, Technometrics, Vol. 13, No. 1, February 1971, pages 33-61.

4. Ball, G. H., and Hall, D. J., A Clustering Technique for Summarizing Multivariate Data, Behavioral Science, Vol. 12, pp. 153-155, March 1967.

Appendix A

ERIM FINAL REPORT ON THE THEMATIC MAPPER DESIGN PARAMETER INVESTIGATION

Note: This appendix does not contain Appendix I of the ERIM report regarding thermal band radiance.

ENGINEERING SUPPORT OF TMDPI

123300-2-L

December 1977

Prepared by

R. Kauth

R. Nalepka

ENGINEERING SUPPORT OF TMPDI

1.0 INTRODUCTION

This report is written in final fulfillment of the terms of ERIM's subcontract from IBM to assist in the experiment design, data selection and results interpretation of IBM's Thematic Mapper Design Parameter Investigation (TMDPI). The period covered is May 1976 - December 1977.

1.1 ERIM'S ROLE

ERIM's role in this effort is basically that of a consultant. The ERIM effort in actually carrying out the experiment and conducting analysis of the results has been minimal.

1.2 THE BROAD TECHNICAL ISSUES

The technical problem at issue is as follows:

The design parameters of the Thematic Mapper include: the filter which follows the detectors; the sampling scheme which is used to digitize the filtered detector output on board the satellite; and the resampling scheme which is employed at the ground station to reconstruct a rectified image to a prescribed scale, suitable for general use by interpreters and for machine classification. NASA has requested an evaluation of:

a five-pole Bessel filter combined with a sample and hold sampling scheme,

versus

a five-pole Butterworth filter combined with an integrate and dump sampling scheme

and

a sample spacing nominally the same as the instantaneous field of view (IFOV),

versus

a sample spacing 1/1.4 times as large as the IFOV

and

nearest neighbor reconstruction,

versus

cubic convolution reconstruction,

versus

point spread function reconstruction.

The above combinations result in 12 treatment combinations. Subsequent negotiations with NASA have presumably led to the deletion of the Butterworth filter with the higher sampling rate, leaving a total of nine system configurations to be evaluated.

The array of conditions over which these system configurations are to be evaluated include a variety of field sizes and field shapes, for several agricultural crop classes.

The performance measure to be used in evaluation is the probability of correct classification. The final figure of merit is to be the average probability of correct classification over the conditions of observation.

1.3 REPORT ORGANIZATION

Section 2, following, is a brief summary of ERIM's activities in support of IBM in this effort. Section 3 consists of technical commentary on the synthetic data aspects of the program. Section 4 concerns the TM simulated data based on aircraft (A/C) imagery. Section 5 contains conclusions and recommendations.

2.0 STATEMENT OF ERIM ACTIVITIES

ERIM assisted IBM in three task areas; experiment design, data set selection; and interpretation of results. In addition ERIM responded to questions and consulted with GSFC on numerous occasions, regarding the interpretation of M-7 scanner data, and the question of atmospheric transmission and emission in the thermal band. Under a separate contract with Goddard Space Flight Center (GSFC), ERIM prepared M-7 scanner data for IBM.

2.1 EXPERIMENT DESIGN AND DATA SELECTION

IBM and ERIM personnel worked together to establish an approximate experiment plan which would distinguish between the parameter sets, if any differences exist.

Of available data sets ERIM recommended that Corn Blight Watch Experiment (CBWE) data be used but pointed out that the data, in its immediately available form had a resolution of 8-10 m and would be of marginal value for the purposes of the experiment. Subsequently it was resolved that the data should be reprocessed from analog tape at a higher (3.5 m) resolution so as to better meet the needs of the experiment.

2.2 INTERPRETATION OF RESULTS

ERIM has viewed raw data classification results provided by IBM and has noted certain discrepancies which have been reported to IBM, mainly in the simulated data sets. ERIM has not been provided with composite summaries, which would make it easy to view the results by class. ERIM has viewed preliminary hand drawn plots of composite summaries but has not been provided with copies.

ERIM has reviewed IBM's immediate future plans for producing results and drawing conclusions. Based on that review ERIM agrees that

ORIGINAL PAGE IS
OF POOR QUALITY

IBM is proceeding along a reasonable route and will be able to draw a conclusion in one of the three broad classes of possible outcomes of the experiment, namely, there are differences between treatments; there are no differences between treatments; or the experiment does not reveal whether there are differences. If it should turn out that the experiment does not reveal whether there are differences, but is almost significant then ERIM recommends that consideration be given to reprocessing the data with a floating point rather than integer boundary decision algorithm, as described in section 3.3 following.

2.3 GSFC SUPPORT

ERIM provided direct support to GSFC by a) delivering a signature set which Goddard personnel could use in creating a synthetic simulated data set, b) answering questions concerning the calibration of M-7 scanner data, and c) answering questions concerning the modelling of the thermal band radiance. The thermal band radiance work is incorporated as Appendix I.

3.0 TECHNICAL DISCUSSION

At the time of this writing a number of unanswered questions remain regarding the experiment. In the following we attempt to state these questions as clearly as possible.

3.1 SYNTHETIC DATA SIGNATURES: ASSUMPTIONS

The signature set which ERIM delivered to GSFC contained several spectral subclasses for each major class in order to more closely approximate the broad spectral spread of each class. In producing the synthetic data set, however, GSFC selected single modes of each class to represent the entire class. Furthermore, GSFC used the variances based on 8 meter resolution data and applied them to creating a synthetic scene with 3 meter resolution elements. These were then smoothed to create 30 meter elements. We believe the net effect of these procedures will be to substantially improve predicted performance compared to real TM performance, when tested on field center pixels far removed from any boundary. Further we believe that, when tested on field center pixels which are nearer boundaries, (i.e., the great majority of field center pixels), the effect is likely to be to create substantial apparent differences among the treatments where only small ones exist. We attempt to describe the reasons we believe this in the following paragraphs.

There are two major sources of within field variance in 30 meter resolution synthetic pixels; within field variance in 3 meter resolution pixels contained in the 30 meter pixels; and influence from 3 meter resolution pixels outside of the field, i.e., belonging to other classes. This latter component of variance will differ depending on the particular filtering/sampling/resampling treatment. Influence due to this component will be primarily dependent on the mean of the adjacent field type signature and since the number of possible combinations

of adjacent types is large and the available number of field samples is modest, (even in the synthetic data) it is doubtful whether this source of variation is sampled adequately.

The net effect of the assumptions made by GSFC is to reduce the first variance contribution unrealistically without reducing the differences between signature means in a comparable way, thus increasing the relative importance of the second variance contribution; and it is this second variance contribution which could in principle, cause differences between the treatments.

In a more realistic simulation the boundary effects on signatures will be diluted by the within field variance.

3.2 SYNTHETIC DATA SIGNATURE VARIATIONS

We have examined, in a preliminary manner, the synthetic data signatures drawn both from the 3 meter resolution data and from the nine treatments. The signature means vary among treatments by what amounts to several sigma of the signature variances. These signatures are drawn from the larger fields available--yet they appear to be dominated by the accidents of neighboring fields and the treatment used in an unrealistic way. We suspect, but cannot be certain without extensive analysis, that these effects may be explained by the GSFC model assumptions described in the previous section.

3.3 SCENE BOUNDARY EFFECTS IN SYNTHETIC DATA

The classification results at the edge of the scene appear to consist of long strings of repetitions of class decisions, regardless of true class. We do not know whether this is a computer bug which may affect all the results, or whether it is another manifestation of "edge effects" in the sense that the scene continues across the boundary as a series of zeros. If the latter is the case it would explain why

so many of the boundary pixels seem to be classified as "trees", since trees have lower signal values than most other classes.

4.0 COMMENTS REGARDING SIMULATED TM DATA BASED ON AIRCRAFT IMAGERY

In general the A/C Simulated TM Data appears free of the major objections above which apply to the synthetic data. There are, however, additional problems which apply to both data sets, but which have extra force with respect to the simulated data set because of its smaller size. The problems primarily have to do with the definition of field boundaries in the various image products.

To begin with we should state that we believe the existence of the problems we will describe in the next section do not invalidate statistically significant conclusions which might be drawn by IBM personnel regarding the simulated TM data. However the existence of these problems does dilute the distinctions between treatments and makes an experiment of any given size less statistically significant than it could be with a different treatment of field boundaries.

4.1 THE BOUNDARY DEFINITION PROBLEM

Ideally the field boundaries seen in A/C data should be projected onto the various TM scenes to provide field boundaries in those scenes. Inset boundaries, used to define field center pixels, should also be drawn initially in the A/C data scene and then projected onto various TM scenes.

The decision as to whether a pixel is in or out of a given boundary would be then determined by an analytical relationship between the projected boundary and the center of the TM pixel in question. In the present experiment this ideal case is compromised in two specific ways.

a. The coordinates of boundary vertices projected onto TM scenes are truncated or rounded to the nearest integer in the TM pixel coordinates. Thus the vertices and the boundaries connecting them do not represent the best available estimate as to the true position of the boundaries, but are misplaced by as much as several aircraft resolution pixels.

ORIGINAL PAGE IS
OF POOR QUALITY

b. These effects are not constant from one TM image to another. The reason for this is that the various TM images are not resampled to identical grids, but rather are resampled to grids which are typically 0.1 TM pixels different from each other. Comparing two TM images then there may be perhaps 1 out of 5 vertices which will be differently placed in either line number or point number or both. This suggests that in a comparison of two TM images a large proportion of the fields will be found to have one or more mismatched vertices.

Since the vertices can be mismatched in any direction they will, on the average, tend to average out. Hence in a large enough experiment the difference between TM images will be retained. However, as mentioned above, the clarity of distinctions that can be made in a smaller experiment is unnecessarily diluted.

The cure for this difficulty is to leave the boundary vertex definitions in floating point form and decide the "in" or "out" status of pixels relative to this floating point boundary.

5.0 CONCLUSIONS AND RECOMMENDATIONS

5.1 SYNTHETIC DATA

ERIM believes that the assumptions made in producing the synthetic data set specifically and directly influence the results in the direction of creating artificial differences between the treatments.

ERIM does not recommend that the synthetic data approach should be scrapped, however. This approach probably offers the only reasonable way of resolving the present issue as well as others which will arise in the future. ERIM recommends that the assumptions be modified and experience be gained with more and more realistic approximations.

5.2 A/C IMAGE BASED TM SIMULATED DATA

ERIM sees no present reason to quarrel with the approach being taken by IBM to analyze the A/C based data. It may be that the conclusions that can be drawn will be found to be only marginally conclusive. If this turns out to be the case, we believe that a sharper distinction may be made by incorporating a floating point boundary description into the processing.

These same remarks apply to the synthetic data, (if the more severe criticisms of that experiment are first resolved).

5.3 GENERAL RECOMMENDATIONS

We believe that additional activities in addressing the Thematic Mapper Design Parameters are required to more conclusively establish the technical needs and cost-effective solutions. Data more specifically gathered to satisfy the needs of such a design study are certainly called for. These data would include more basic class types at various stages of growth and provide more samples of the range of field sizes desired.

Because of the critical effect of the class signatures on the final conclusions, additional care should be taken in initially defining the signature set to be used throughout the investigation for both the A/C and synthetic data sets. In addition, since many users of satellite multispectral scanner data are primarily interested in large area inventories, more emphasis needs to be placed on performance measures such as proportion estimation of the various class types over a region.