NASA

Technical Memorandum 79710

Spatial Land-Use Inventory, Modeling, and Projection/ Denver Metropolitan Area, with Inputs from Existing Maps, Airphotos, and Landsat Imagery

Craig Tom, Lee D. Miller and Jerrold W. Christenson

(NASA-TM-79710) SPATIAL LAND-USE INVENTORY, MODELING, AND PROJECTION/DENVER METROPOLITAN AREA, WITH INPUTS FROM EXISTING MAPS, AIRPHOTOS, AND LANDSAT IMAGERY (NASA) 215 p HC A10/MF A01 CSCL 05B G3/43

N79-18422

Unclas 14343

AUGUST 1978

National Aeronautics and Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

SPATIAL LAND-USE INVENTORY, MODELING, AND PROJECTION/DENVER METROPOLITAN AREA, WITH INPUTS FROM EXISTING MAPS, AIRPHOTOS, AND LANDSAT IMAGERY

C. Tom

Colorado State University
Fort Collins, Colorado 80523

L. D. Miller

Texas A & M University
College Station, Texas 77840

and

J. W. Christenson

Goddard Space Flight Center
Greenbelt, Maryland 20771

August 1978

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

This document makes use of international metric units according to the Systeme International d'Unites (SI). In certain cases, utility requires the retention of other systems of units in addition to the SI units. The conventional units stated in parentheses following the computer SI equivalents are the basis of the measurements and calculations reported.

Throughout this document, conventional units of measure, which are customary for land-use planning, are used in the figures and tables and their captions.

SPATIAL LAND-USE INVENTORY, MODELING, AND PROJECTION/DENVER METROPOLITAN AREA, WITH INPUTS FROM EXISTING MAPS, AIRPHOTOS, AND LANDSAT IMAGERY

C. Tom; L. D. Miller*
and
J. W. Christenson

ABSTRACT

The research objectives were to overlay ancillary map data onto a Landsat multispectral data base, create and test land-use change prediction models, and map land uses in the Denver Metropolitan Area. Linear-discriminant analysis was used to model spatially projected land-use changes and to classify the single-date Landsat-1 image with ancillary data inputs.

A landscape model was constructed with 34 land-use, physiographic, socioeconomic, and transportation maps. A simple Markov land-use trend model was constructed from observed rates of change and non-change from photointerpreted 1963 and 1970 airphotos. Seven multivariate land-use projection models predicting 1970 spatial land-use changes achieved accuracies from 42 to 57 percent. A final modeling strategy was designed, which combines both Markov trend and multivariate spatial projection processes.

Landsat-1 image preprocessing included geometric rectification/resampling, spectral-band, and band/insolation ratioing operations. Rectangular training-set selection gave biased and unreproducible results. A new, systematic grid-sampled point training-set approach proved to be useful when tested on the four original MSS bands, ten image bands and ratios, and all 48 image and map variables (less land use). Ten-variable accuracy was raised over 15 percentage points from 38.4 to 53.9 percent, with the use of the 31 ancillary variables. Only three optimal ancillary variables added almost 12 of these 15 percentage points of improvement. The maximum-likelihood ratio classifier was also established to be inferior to linear-discriminant analysis for land-use mapping.

A land-use classification map was produced with an optimal ten-channel subset of four image bands and six ancillary map variables. Point-by-point verification of 331,776 points against a 1972/1973 U.S. Geological Survey (USGS) land-use map prepared with airphotos and the same classification scheme showed average first-, second-, and third-order accuracies of 76.3, 58.4, and 33.0 percent, respectively.

The average direct development cost for this 24-class land-use mapping effort with the 48-variable overlayed Landsat spectral/spatial map data was $0.0468 per hectare ($0.0189 per acre), or $704.19 per 1:24,000-scale USGS quadrangle, with the direct production cost using only the optimal ten-channel spectral/spatial data set estimated at an average of $0.0317 per hectare ($0.0128 per acre), or $477.17 per 7.5 minute USGS quadrangle.

Recommendations for further evaluation study include densification of socioeconomic data, improved digitization of topographic elevation, input of collateral soils data, multidate land-use change detection, greater use of ancillary map data, and further information systems development.

*A portion of the work was performed while serving as a Senior Post-doctoral Research Associate, Goddard Space Flight Center.

ACKNOWLEDGMENTS

This research has benefited from the cooperation, interaction, and counsel of many people over its span of 6 years. Throughout the long days and equally long nights, many people contributed their dedication, encouragement, ideas, expertise, labor, understanding, and just plain enthusiasm and friendship to the multitude of tasks both mundane and monumental that led to the eventual fruition of this effort. Any attempt to name them all would be incomplete and would do injustice to someone. Therefore, it is hoped that those who participated will recognize their contributions to the success of this project and that we are indebted to each and every one of them.

All the work was completed at Colorado State University and was supported chronologically as follows:

- U.S. Department of Interior, EROS Program, Bureau of Outdoor Recreation, Denver, Colorado
- Colorado State University, University Computer Center, Ft. Collins, Colorado
- Colorado State University, Department of Forest and Wood Science and Department of Civil Engineering, Ft. Collins, Colorado
- NASA/Goddard Space Flight Center, Landsat Follow-On Support to the Federation of Rocky Mountain States, Denver, Colorado
- NASA/Goddard Space Flight Center, Earth Resources Application Branch, Census Project, Greenbelt, Maryland

IV PAGE INTENTIONALLY BLANK

CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGMENTS	v
CHAPTER 1 - INTRODUCTION	3
GENERAL	3
Planning Process	4
Needed Technology Transfer	4
STUDY OBJECTIVES	5
SCOPE OF THE RESEARCH	6
GENERAL APPROACH	6
ANALYSIS PROCEDURES	8
Data Resources	8
Analysis Procedures	11
Utilization of Computer Program Modules	12
DETAILED REPORT ORGANIZATION	13
CHAPTER 2 - LANDSCAPE MODEL CONSTRUCTION	17
INTRODUCTION	17
CONSTRUCTION OF A SPATIAL DATA PLANE	19
Map Sampling Procedure	19
Compilation of the Data Plane	22
CONSTRUCTION OF LAND-USE SUBMODEL	27
CONSTRUCTION OF PHYSIOGRAPHIC SUBMODEL	35
CONSTRUCTION OF TRANSPORTATION SUBMODEL	46
CONSTRUCTION OF SOCIOECONOMIC SUBMODEL	54
SUMMARY	56
CHAPTER 3 - SPATIAL LAND-USE PROJECTION	69
INTRODUCTION	69
DETERMINATION OF RECENT LAND-USE DYNAMICS	69
MARKOV LAND-USE TREND MODEL	70
MULTIVARIATE LAND-USE SPATIAL PREDICTION MODEL	77
Relationship of Land-Use Change to Landscape Variables	79
Spatial Modeling Concepts	79

vi

CONTENTS (Continued)

	Page
Initial Spatial Modeling Test	81
Tests of Other Generic Spatial Models	83
Proposed Advanced Spatial Model	96
CHAPTER 4 - LANDSAT LAND-USE CLASSIFICATION	101
INTRODUCTION	101
OVERLAYING LANDSAT IMAGES ON THE LANDSCAPE MODEL	104
Review of Image Processing	104
Preprocessing for Geometric Rectification	105
Preprocessing to Form MSS-Band Ratios	111
Merging of Landsat and Landscape Data Planes	111
Preprocessing to Remove Terrain Effects	111
MULTIVARIATE CLASSIFICATION WITH RECTANGULAR TRAINING SETS	114
MULTIVARIATE CLASSIFICATION WITH GRID-SAMPLED TRAINING SETS	120
Verification	123
MAXIMUM-LIKELIHOOD CLASSIFICATION WITH GRID-SAMPLED TRAINING SETS	125
PRODUCTION OF LAND-USE MAPS	128
Selection of Optimal Mapping Variables	132
First-Order Theme Map	137
Second- and Third-Order Theme Maps	143
Verification	145
Direct Cost/Time Analysis	164
SUMMARY	164
CHAPTER 5 - CONCLUSIONS	171
SUMMARY	171
RELATED APPLICATIONS AREAS	172
RECOMMENDATIONS	173
Socioeconomic Census Data Densification	173
Topographic Elevation Data Digitization	173
Collateral Soils Data Inputs	173

CONTENTS (Continued)

	Page
Multidate Land-Use Change Detection	174
Symbiotic Use of Ancillary Map Data	174
Future Information Systems Development	174
REFERENCES	177
APPENDIX A - MULTIPLE DISCRIMINANT ANALYSIS	183
APPENDIX B - LANDSAT MAPPING SYSTEM	189
APPENDIX C - MACHINE CLASSIFICATION ERROR-RATE ESTIMATION	197
APPENDIX D - MAXIMUM-LIKELIHOOD RATIO	203
APPENDIX E - CORRELATION MATRICES FOR RECTANGULAR/ POINT-SAMPLED TRAINING SETS	207

ILLUSTRATIONS

Figure		*Page*
1	Simple Schematic Representation of the Landscape Modeling Concept	7
2	Land-Use Change Analysis, Mapping, Modeling, and Change-Prediction Flowchart	10
3	Conceptual Diagram of the Landscape Model of the Denver Study Area	18
4	Topographic Map of the Denver Study Area	21
5	Topographic Map Index of the Denver Study Area	23
6	Geographic Reference Overlay for Data Forms	24
7	Overlay Data Form Used for Sampling a 1:24,000-Scale Map	25
8	Complete Overlay Data Form Ready for Keypunching	26
9	A Block of Punched Cards Representing One Overlay Data Form	27
10	Display of the Data on One Keypunched Overlay Data Form	28
11	A Portion of the 1970 Denver Land-Use Data Plane	29
12	A Selective "Open-Space" Display of a Portion of the 1970 Denver Land-Use Data Plane	30
13	Display of the 1970 Land-Use Data Plane Emphasizing All of the Urban and Built-Up Areas (in black)	31
14	Display of the 1970 Land-Use Data Plane Emphasizing All Remaining Agricultural Lands (in black)	32
15	Display of the 1970 Land-Use Data Plane Emphasizing Single- and Multiple-Unit Residential Areas (in black)	33
16	Display of the 1970 Land-Use Data Plane Emphasizing the Location of Strip and Clustered Developments (in black) Relative to Other Urban and Built-Up Areas (in gray)	34
17	Display of the 1963 Land-Use Data Plane Emphasizing Single- and Multiple-Unit Residential Areas (in black)	36
18	Display of the 1963 Land-Use Data Plane Emphasizing the Location of Strip and Clustered Developments (in black) Relative to Other Urban and Built-Up Areas (in gray)	37
19	Display of the original 1972-1973 USGS 1:100,000-Scale Land-Use Source Map	38
20	Display of the 1972-1973 USGS Land-Use Data Plane Emphasizing Single- and Multiple-Unit Residential Areas (in black)	39
21	Display of the Topographic Elevation Data Plane Emphasizing the Lowest Areas (in black)	40
22	Display of the Topographic Slope Data Plane Emphasizing Shallow Slopes (in black)	41
23	Display of the Topographic Slope Data Plane Emphasizing Steep Slopes (in black)	42

ILLUSTRATIONS (Continued)

Figure		*Page*
24	Display of the Topographic Aspect Data Plane Emphasizing Northwest-Facing Areas (in black)	43
25	Three-Dimensional Perspecitve Display of the Topographic Elevation Data Plane Looking from the Southeast	44
26	Contour Map of the Topographic Elevation Data Plane	45
27	Three-Dimensional Perspective Display of the Topographic Elevation Data Plane	46
28	Three-Dimensional Perspective Display of the Topographic Elevation Data Plane Emphasizing All Urban Built-Up Lands (in black)	47
29	Display of the Insolation Data Plane Emphasizing Areas of Lowest Incoming Solar Energy (in black)	48
30	Display of the Surficial Geologic Data Plane Emphasizing Aeolian Deposits Deposits (in black)	49
31	Display of the 1971 Road Transportation Data Plane Showing Freeway Interchanges	50
32	Display of the Minimum Distance to Freeway Interchange Data Plane Emphasizing Minimum Distance (in black)	51
33	Display of the 1971 Freeway and Major-Road Transportation Data Planes and Their Associated Transformed Planes That Emphasize Minimum Distances (in black)	52
34	Display of the 1971 Minor-Road and Built-Up Urban Area Transportation Data Planes and Their Associated Transformed Planes That Emphasize Minimum Distance (in black)	53
35	Census Tract Overlay of the Denver Metropolitan Area	55
36	Display of the 1970 Socioeconomic Data Emphasizing the Highest Total Number of Families Per Census Tract (in black)	57
37	Display of the 1970 Socioeconomic Data Emphasizing Highest 1969 Mean Family Income, Median Housing-Unit Value, and Median Housing-Unit Rent (in black)	59
38	Display of the 1970 Socioeconomic Data Emphasizing Highest Population Totals and Averages (in black)	61
39	Display of the 1970 Socioeconomic Data Emphasizing Highest One-, Two-, Three-, and Three-Plus-Car Family Ownership and Average Number of Cars Per Family (in black)	63
40	Display of the Loss in Open Space Betewen 1963 and 1970 Emphasizing the Open Space Embedded in the Urban Area That was Converted Into Other Land Uses (in black)	71
41	Display of the Loss in Embedded Open Space Between 1963 and 1970 Emphasizing the Agricultural Land That was Converted Into Other Land Uses (in black)	72
42	Prediction of Future Trends in the Amount of Open Space and Competing Land Use in the Denver Metropolitan Area	77

ILLUSTRATIONS (Continued)

Figure		*Page*
43	Correlation of the Change in Agricultural Land Use with Elevation	80
44	Simple Linear Discriminant-Function Diagram	82
45	Comparative Displays of the Spatial Predictions of Three Land-Use Models Emphasizing All Changes From Agricultural Lands (in black)	91
46	Proposed Combination Markov and Linear Discriminant Models for Improved Spatial-Change Prediction	97
47	Geometric Relationship of Original and Nearest-Neighbor Transformed Landsat Image Cells	106
48	Resampling Efficiencies of the Image Geometric Transformation	107
49	Comparative Display of the Original Unrectified and the Transformed (Rectified/Resampled) Multispectral Landsat Imagery of the Denver Metropolitan Area	108
50	Landsat-1 MSS Band 5 Visible Red Image of the Denver Metropolitan Area	109
51	Landsat-1 MSS Band 7 Solar Infrared Image of the Denver Metropolitan Area	110
52	Display of the Effect of the Expansion of the Ancillary Map Data Planes in the Landscape Model for Overlay on Landsat Image Data Planes	112
53	Display Illustrating the Overlay of Selected Physiographic Map Data on the Landsat Multispectral Imagery of the Denver Metropolitan Area	113
54	Training-Set Accuracies and Costs of Processing Single-Date Landsat Imagery of the Denver Metropolitan Area with 30 Ancillary Landscape Variables	118
55	Verification of the Map Accuracy for the Denver Metropolitan Area for All 1972-1973 USGS Land-Use Classes (August 15, 1973 Landsat-1 image)	121
56	Low Verified Land-Use Accuracy for Automated Interpretation Using Rectangular Training-Set Statistics	122
57	Verification Accuracies and Costs of Processing Single-Date (August 15, 1973) Landsat Imagery of the Denver Metropolitan Area with Four MSS Bands	124
58	Verification of Accuracies and Costs of Processing Single-Date (August 15, 1973) Landsat Imagery of the Denver Metropolitan Area with the Four Primary and Six MSS-Band Ratios	126
59	Verification Accuracies and Costs of Processing Single-Date (August 15, 1973) Landsat Imagery of the Denver Metropolitan Area with Ten MSS-Band/Ratio Variables and 31 Ancillary Landscape Variables	129
60	Map-Verification Accuracies and Costs of Processing Single-Date Landsat Imagery of the Denver Metropolitan Area with Ten MSS-Band/Ratio Variables and 31 Ancillary Landscape Variables	131
61	Step-By-Step Classification of the Land Use of the Denver Metropolitan Area Verified at the First Order (August 15, 1973, Landsat-1 image)	133
62	Step-By-Step Classification of the Land Use of the Denver Metropolitan Area Verified at the Second and Third Orders (August 15, 1973, Landsat-1 image)	135

ILLUSTRATIONS (Continued)

Figure		*Page*
63	Verified Map Accuracy for the Denver Metropolitan Area for All USGS Land-Use Classes at the 41st Step (August 15, 1973, Landsat-1 image)	137
64	Maximum-Likelihood Ratio (or Gaussian) Image Classification of the Denver Metropolitan Area Verified at the First Order (August 15, 1973, Landsat-1 image)	139
65	Maximum-Likelihood Ratio (or Gaussian) Image Classification of the Denver Metropolitan Area Verified at the Second and Third Orders (August 15, 1973, Landsat-1 image)	141
66	Test of Final Ten-Channel Combination of Landsat and Landscape Variables for the Full Image Classification	144
67	Comparative Displays of the Discriminant Classification, a Verification of That Classification, and the USGS Map of the Six *First-Order Land-Uses* of the Denver Area	147
68	Comparative Displays of the Discriminant Classification, a Verification of That Classification, and the USGS Map of the *Urban Land Uses* of the Denver Area	149
69	Comparative Displays of the Discriminant Classification, a Verification of That Classification, and the USGS Map of the *Agricultural Land-Uses* of the Denver Area	151
70	Comparative Displays of the Discriminant Classification, a Verification of That Classification, and the USGS Map of the *Water-type Classes* of the Denver Area	153
71	Comparative Displays of the Discriminant Classification, a Verification of That Classification, and the USGS Map of the *Rangelands* of the Denver Area	155
72	Comparative Displays of the Discriminant Classification, a Verification of That Classification, and the USGS Map of the *Barrenlands* of the Denver Area	157
73	Comparative Displays of the Discriminant Classification, a Verification of That Classification, and the USGS Map of the *Forestlands* of the Denver Afea	159
B-1	Step 1: Image Preparation/Map Overlay	190
B-2	Step 2: Interleaves Images from Various Dates	191
B-3	Step 3: Computes Statistical "Signatures" of Materials to be Mapped	192
B-4	Step 4: Maps Distribution of Each Material	193

TABLES

Table		*Page*
1	Hierarchical Land-Use/Land-Cover Classification Scheme Used for the Denver Metropolitan Area	9
2	List of Spatial Landscape Modeling Variables	20
3	Net Changes in 1963 Denver Land Use Relative to 1970	73
4	Changes in Denver Land Use, 1963 to 1970	74
5	Denver Land-Use Probability Transition Matrix, 1963 to 1970	76
6	Denver Aggregate Land-Use Projections by a Markov Trend Model, 1963 to 2068	78
7	Initial 38-Class 1963 to 1970 Land-Use Change Model-Classification Matrix	85
8	Accuracy of Prediction by the Initial Model of Future Changes in Land Use on a Cell-to-Cell or Spatial Basis for the Denver Metropolitan Area	87
9	Landscape Variable Entry Order of the Initial Spatial Land-Use Projection Model	88
10	Comparison of Seven Test Land-Use Models of Spatial Change in the Denver Metropolitan Area, 1963 to 1970	89
11	Second Composite 38-Class 1963 to 1970 Land-Use Change Model Classification Matrix	93
12	Accuracy of Prediction of Future Changes in Land Use on a Cell-to-Cell or Spatial Basis for the Denver Metropolitan Area	95
13	Training-Set Accuracy of Automated Interpretation of the Four Original MSS Bands of a Landsat Image (August 15, 1973) of the Denver Metropolitan Area	115
14	Training-Set Accuracy of Automated Interpretation of the Four Original Landsat MSS Bands and Six Ratios of a Single-Date Landsat Image (August 15, 1973) of the Denver Metropolitan Area	116
15	Improvement in the Training-Set Accuracy of Automated Interpretation of a Single-Date Landsat-1 Image of the Denver Metropolitan Area by Adding 30 Ancillary Variables	117
16	High Training-Set Accuracies and Economy Achieved by Automated Interpretation of a Selection of Three MSS Bands of a Single-Date Landsat Image (August 15, 1973) of the Denver Metropolitan Area with Overlays of Four Landscape Variables	119
17	Low Verified Land-Use Accuracy for Automated Interpretation Using Rectangular Training-Set Statistics (August 15, 1973, image)	123
18	Verification Accuracies of Automated Interpretation of the Four Original MSS Bands of an Image (August 15, 1973) of the Denver Metropolitan Area	125
19	Verification Accuracies of Automated Interpretation of the Four Original MSS Bands and Their Six Ratios for a Single-Date Landsat Image (August 15, 1973) of the Denver Metropolitan Area	127
20	Improvement in the Verification Accuracies of Automated Interpretation by the Addition of 31 Ancillary Variables for a Single-Date (August 15, 1973) Landsat Image of the Denver Metropolitan Area	130

TABLES (Continued)

Table		*Page*
21	Map-Verification Accuracy Achieved with the Grid-Sampling Approach to Assembling Training-Set Statistics (August 15, 1973, image)	132
22	Comparative Accuracies of Maximum-Likelihood and Linear-Discriminant Image-Classification Algorithms	143
23	Test of Final Ten-Channel Combination of Landsat and Landscape Variables for Full Image Classification (August 15, 1973, image)	145
24	Final Ten-Channel Linear-Discriminant Classification Results Using Four Landsat and Six Ancillary Landscape Variables	162
25	Cost/Time Tabulation for Single-Date 48-Variable Landsat/Ancillary Landscape Variable Classification of the Denver Metropolitan Area	165
26	Cost/Time Tabulation for Single-Date, Ten-Variable Landsat/Ancillary Landscape Variable Classification of the Denver Metropolitan Area	166
B-1	Cost Estimates for Operating the Landsat Mapping System	194
C-1	Machine Classification Decision Table	197
C-2	First-Order Land-Use Classification Matrix Using Final Ten-Variable Linear-Discriminant Analysis	198
C-3	Composite Second- and Third-Order Land-Use Classification Matrix Using Final Ten-Variable Linear-Discriminant Analysis	199
E-1	Within-Groups Correlation Matrix for the Four Original MSS Bands Derived from Rectangular Training Sets	207
E2	Within-Groups Correlation Matrix for the Four Original MSS Bands and Six Ratios Derived from Rectangular Training Sets	208
E-3	Within-Groups Correlation Matrix for the Optimal Set of Three MSS Spectral Bands and Four Landscape Variables Derived from Rectangular Training Sets	208
E-4	Within-Groups Correlation Matrix for the Four Original MSS Bands Derived from Grid-Sampled Training Points	209
E-5	Within-Groups Correlation Matrix for the Four Original MSS Bands and Six Ratios Derived from Grid-Sampled Training Points	209
E-6	Within-Groups Correlation Matrix for the Ten-Variable Combination of Landsat and Landscape Variables for Full-Image Classification Derived from Grid-Sampled Training Points	210

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

GENERAL

Planning is a process that systematically deals with the following (Reference 1):

- Formulation of objectives and standards with which to specify future conditions
- Collection and analysis of relevant data to accurately determine existing conditions and trends
- Development of alternative plans to achieve the desired conditions
- Selection, adoption, and implementation of an "optimal" plan from the available alternatives

Land-use planning, the name commonly applied to the tasks of designing a unified development plan, must consider the spatial arrangements of all land uses. Such land-use plans range from the location of transportation routes to the location of public facilities such as airports, dams, open space, parks, power plants, and schools.

The increasing size and complexity characterizing the modern, large-scale land-planning organization has made the historical managerial functions of planning, organizing, and controlling the future spatial distribution of land use much more difficult to achieve. At the same time, the successful application of these functions has become increasingly essential to the orderly development and stability of today's urban and rural communities.

Rapid growth pressures on the lands used for agriculture, mineral resources, open space, outdoor recreation, water supply and storage, transportation, and wildlands requires substantial advances in the collection, classification, and availability of appropriate land inventory and planning data in general.

The required inventory and planning data are not generally available to land planners at any level. When available, these data often lack completeness, quality, and timeliness. The availability of timely data in the form of a detailed, quantitative inventory and planning data bank provides land planners with a dynamic definition of the tradeoffs involved in evaluating the following:

- Alternative sites for the same type of development
- Proposed alternative sites to be earmarked for development
- Reciprocal effects between development/nondevelopment and activities in adjacent areas

2 PAGE INTENTIONALLY BLANK

Planning Process

Urban planning has traditionally emphasized efficient functioning of cities in a purely economic and engineering sense, with particular attention devoted to land-use activities, transportation, and zoning. This emphasis has been termed "physical" planning and seeks to achieve a viable and sound land-use pattern (Reference 2). A stress on "current" planning permits only a land regulation-type of urban planning in which subdivision codes, zoning, and parcel-oriented policy-making is paramount. "Advance" planning addresses the longer range, but equally urgent, problems such as blight, obsolescence, suburbanization, and sprawl and may be done both intermittently and incrementally. Physical planning works as an agent of private development, rather than as a normative instrument for the public welfare. It does not challenge the 19th century view of land as a speculative commodity (Reference 3), but merely imposes ground rules under which speculation continues.

Promoted through the Housing Act and other recent federal legislation, social planning adopts a socioeconomic, political, and physical approach to the development and functioning of the urban community in an effort to compensate for the past failures of physical planning and to address major issues and problems confronting the hitherto ignored human and social elements. Historically, reconciliation of the two philosophies has been unsuccessful.

Regional growth has been accomplished through near-universal phenomena of air, land, and water pollution, undesirable economic contingencies and land-use conversions, traffic congestion, and urban blight and sprawl. Although urban land uses command the greatest value in monetary terms, the "value" of open, nonurban areas remains uncontested and essential. This open space is being progressively preempted by other land uses because of a myriad of factors inherent in the employment of current planning practices. Therefore, it is both necessary and valuable to inventory large urban and urban fringe areas and to delimit land uses and potentials.

It is essential to first develop an understanding of the landscape in terms of its cultural and natural components. These are difficult to measure because of the complex interactions among air, water, land, biomass, and cultural resources. In addition, these landscape elements are always in a dynamic state of succession, however slow or fast that successional rate may be.

Needed Technology Transfer

Local and regional planning agencies represent a large number of potential users of the evolving remote sensing technology. Remote sensing may be defined as the noncontact collection, analysis, and interpretation of data from aerospace platforms. It should be realized that this remote sensing extends far beyond the realm of conventional aerial photography, which is only a small, representative facet of the increasingly expanding and sophisticated whole. Spatial remote sensing data are available over large areas and can already be cost-effectively obtained relative to traditional ground methods. Remote sensing data can be collected, interpreted for various purposes, stored for further use, and used as a historical record. It can also be employed more easily, consistently, and objectively, and at lower unit-area cost than most ground-based surveys.

Remote sensing offers considerable potential in advance planning procedures in which unique advantages are realized in dealing with environmental issues in which policy-making must be based on both quantitative and qualitative data. It permits for the first time the practical establishment of standardized observations of,

and criteria for, change. Observations at appropriate time intervals will reveal the precursors of change and trends sufficiently earlier so that land planners can probe for basic causes and assess latent probabilities. Continuous planning at regional scales becomes practical. Lastly, remote sensing presents a permanent record of urban phenomena unbiased by the planner's experience.

These advantages of remote sensing suggest practical applications in a different form than is currently practiced. The rising tide of environmental concern popularly manifested in the 1970's is exerting great pressure on "front-line" planning agencies to shed their historical burden of current administrative problems. They must now assume greater responsibilities for the quality of life, environmental quality, and other long-neglected urban problems. Because environmental deterioration and destruction can be traced to man's interference, there are increasing pressures to monitor, measure, and evaluate his activities. Surrogates for environmental indicators have been proposed, including the analysis of disparate and preemptive land uses (References 4 and 5), changes in land value (Reference 6), and the quality and care of residential lawns (Reference 7). While the greatest current demand on remote sensing is to generate land-use data, the technical capability simultaneously exists to extract quantitative and qualitative data of which land use is but a part.

Unfortunately, there has been relatively limited success as yet in incorporating remote sensing applications into the routine operations of urban and regional planners and decision-makers. The mechanics for accomplishing such a technology transfer are simply not known. The scarcity of such planning applications is not apparently related to either the quality or utility of remote sensing techniques, nor even simple ignorance. One major obstacle in realizing the full potential of remote sensing is that its form has been incompatible so far with the census tracts, blocks, and parcels and other land-use inventory schemes used by planners. While remote sensing data can be interpreted to yield both spatial and statistical parameters, these data must also be aggregated and interfaced, for the time being, to the areal units that are compatible with the planners' normal geographical definitions of the city or region.

A second aspect of the technology transfer problem is the necessity to create and maintain large geographic data bases over extensive areas. The various available data input sources are rectified to a common geographic base to meet social planning requirements. Basic social land-use planning concepts have not yet been widely researched or implemented. However, recent developments in electronic data processing have created the possibility that the voluminous quantity of available social planning data may be manipulated more easily and economically in the form of a geographic information system. Remote sensing can provide wide, regional coverage of inaccessible or sparsely monitored geographic data. Thus, the computer-based geographic information system can, in turn, effectively blend regional planning and remote sensing to simplify both physical and social planning data requirements and actions. Such data collection, retrieval, and analysis capabilities are essential for the monitoring, measurement, and evaluation of land-use changes before planners can undertake environmental planning and thereby assume large-scale environmental management responsibilities. Currently, remote sensing and geographic information systems remain as intermittent, piecemeal grafts on a persistently manual methodology of operational land-use analysis that is largely cosmetic in nature.

STUDY OBJECTIVES

The objectives of this specific research were to overlay ancillary map data onto a Landsat multispectral data base, create and test various models for the prediction of land-use change, and map land uses in the Denver

Metropolitan Area. Specific objectives included the following:

- Development of a landscape model with land-use, physiographic, socioeconomic, and transportation components
- Testing of a land-use trend model
- Development and testing of a spatial land-use projection model
- Overlaying of the Landsat image onto the landscape model
- Optimization of the Landsat image classification algorithm when used with landscape or ancillary variables
- Optimal land-use mapping of the study area using Landsat and landscape variables
- Display and verification of the machine-produced maps
- Tabulation of computer, labor, and material cost/time for the land-use maps

SCOPE OF THE RESEARCH

The first Landsat, formerly Earth Resources Technology Satellite (ERTS-1), has provided unexcelled opportunities to explore the utility of aerospace remote sensing data for large-scale analyses of land ecosystems. This research was specifically directed toward the extraction, use, and assessment of land-use inventory and planning data derived from such remote sensing imagery, ancillary map data, and geographic information systems. This endeavor sought to illustrate which features and characteristics of these data inputs and systems are useful in planning processes.

The research was also structured to provide an analytical framework for modeling land-use changes and for identifying the factors that were influential in controlling such changes. The results indicated the levels of accuracy to be expected under operational conditions and also provided the recommended procedures, data sources, and computational techniques.

GENERAL APPROACH

The digital landscape model organized and overlayed data from existing maps, census tables, and remote sensing imagery into a computer framework (figure 1). This assemblage provided a multivariate, multitemporal mathematical model that represented the landscape much as a three-dimensional model of the physical terrain is represented by a topographic map (Reference 8). Coupled with this composite of data overlays was a collection of computer techniques that permitted meaningful simulations of the spatial or map-like behavior of this landscape to either natural or man-induced alteration and control (Reference 9). The thrust of land-use modeling was the prediction and display in map form of the future landscape that would result from the continuation of current land-management practices or the lack

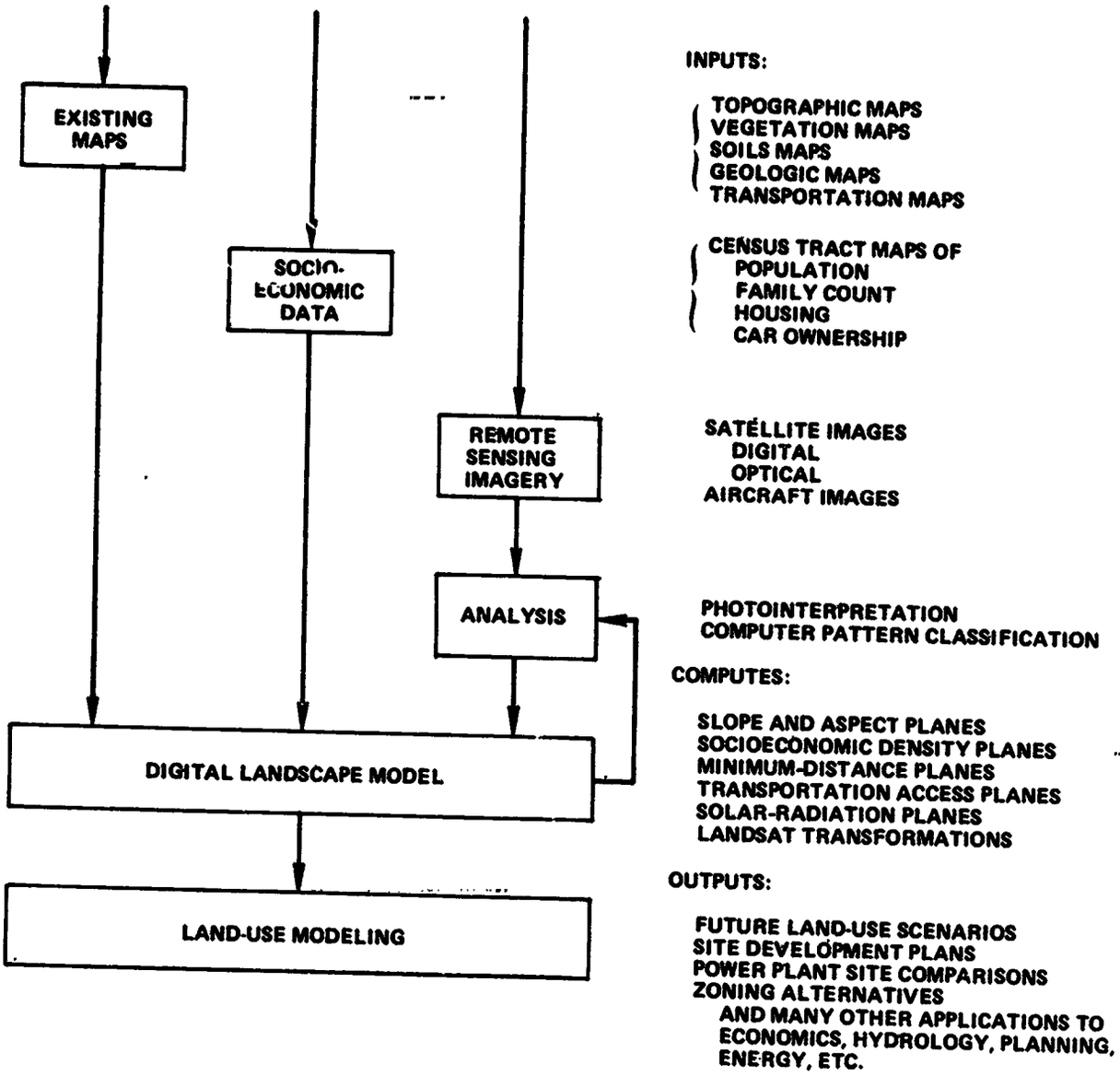


FIGURE 1. SIMPLE SCHEMATIC REPRESENTATION OF THE LANDSCAPE MODELING CONCEPT. Spatially referenced data from a variety of sources is overlaid in the landscape model. A symbiotic relationship exists between landscape modeling and remote sensing image analysis. Current and projected landscape scenarios provide new inputs to the land-use modeling and decision-making processes.

thereof. Success in this objective provided a basis for the proposal for the improvement of these techniques to predict how the landscape will evolve in a spatial sense to various scenarios of anticipated alterations.

Computer analysis of remote sensing imagery provided the important current and past land-use inputs to the land-use modeling and was totally compatible with the modeling process. Symbiotically, the accuracy of the computer interpretation of the remote sensing imagery was substantially improved by including landscape variables such as topographic elevation. Thus, combining the available remote sensing imagery with map data in the digital landscape model provided the basis for substantial improvements in both activities.

The following basic selections were made to provide benchmark resources for testing the basic hypotheses and for satisfying the stated objectives. Detailed interpretation of the most recent sets of low-altitude black-and-white airphotos provided the maps of the land use on various dates, which were used as quantitative measures of change. Landsat four-band imagery has been continuously obtained since July 1972, and the excellent August 15, 1973, image was selected for analysis.* The U.S. Geological Survey (USGS) Circular 671 (Reference 10) hierarchical land-use classification scheme was adopted for uniform interpretations in both manual and automated image analysis activities (table 1). Twenty-one second-order USGS land uses were originally photointerpreted for landscape model construction and spatial land-use projection. However, 13 second-order and 11 third-order categories were used in the Landsat land-use classification to conform to the classes used in a 1972-1973 USGS photointerpretation study of the Denver Metropolitan Area (Reference 11). Although USGS Circular 671 was revised by Professional Paper 964 (Reference 12), revision was not made here because a large amount of intercorrelated data (e.g., photointerpretations, etc.) had already been completed using the Circular 671 approach.

The Circular 671 system has required modification and redefinition because of more recent changes in remote sensing technology. However, the original classification system was sufficiently inclusive to have been proposed in 1972 as a standardized framework for land-use surveys by remote sensing. The original system was tested by the USGS Geographic Applications Program (GEOGAP) in several research programs that monitored land-use changes using satellite and high-altitude aerial photographic data. Three regional projects that used the scheme studied the Central Atlantic Regional Ecological Test Site (CARETS) (Reference 13), the Ozarks region, and southern Arizona (Reference 14). The Anderson land-use classification system was tested in each of these regions and was found to work satisfactorily, even when used with satellite imagery.

ANALYSIS PROCEDURES

This study made use of numerous sources of cultural and natural landscape data. These diverse sources and quantities of input data were used in this study, but would have overburdened manual data collection and analysis methodologies. It should be emphasized that large quantities of map and remote sensing data can serve no practical purpose unless both a rationale and a capability exists for their rapid, objective processing. Therefore, the data analysis plan was prepared early during the research to outline the orderly progression of steps needed for accomplishing the study objectives (figure 2). The following subsections describe in more detail the data resources, analysis procedures, and computer programming called for by this analysis plan. Specific inputs, operations, and outputs are cross-referenced to the data analysis plan by the bracketed numbers (i.e., [1] in the text indicates the position on the diagram indicated by [1] (figure 2)).

Data Resources

The data used in the conceptual design of the land-use modeling effort was acquired from the literature, ground surveys and maps, aircraft photography, and satellite line-scanner data (Reference 1). The initial tasks in this area included the collection of pertinent graphic- and numeric-format land-use, cultural, and physiographic data. Historical black-and-white, 1:20,000-scale panchromatic aerial photos of 1963 and 1:24,000-scale 1970 orthophotomaps [1] provided the multidate land-use "ground-truth" data base. High-altitude U-2 aircraft photos dated 1972-1973 were used by USGS to compile their published map [1]. Other

* The image processing research efforts were initiated in 1972, and the selected 1973 image was the first available high-quality summer image (Landsat scene 1388-17131) for the Denver Metropolitan Area.

TABLE 1

HIERARCHICAL LAND-USE/LAND-COVER CLASSIFICATION SCHEME USED FOR THE DENVER METROPOLITAN AREA. The various levels of the USGS Circular 671 system are shown (Reference 10). This standardized classification was used for manual airphoto interpretation and was modified by adding eleven third-order classes for automated Landsat image analysis. The system was revised and reissued with minor changes as Professional Paper 964 (Reference 12).

Digital Codes	First-Order Land Use/Land Cover Second-Order Third-Order
1	Urban and built-up land
11	Residential
12	Commercial and services
121	Recreational
13	Industrial
14	Extractive
15	Transportation, communications, and utilities
151	Utilities
16	Institutional
17	Strip and clustered development
18	Mixed urban
19	Open and other urban
191	Solid-waste dump
192	Cemetery
2	Agricultural land
21	Cropland and pasture
211	Nonirrigated cropland
212	Irrigated cropland
213	Pasture
22	Orchards, groves, and other horticultural areas
23	Feeding operations
24	Other agricultural land
3	Rangeland
31	Grass
32*	Savannas
33	Chaparral (taken as brushland)
34*	Desert shrub
4	Forest land
41	Deciduous
411	Deciduous/intermittent crown
42	Evergreen (coniferous and other)
421	Coniferous/solid crown
422	Coniferous/intermittent crown
43*	Mixed forest land
5	Water
51	Streams and waterways
52	Lakes
53	Reservoirs
54*	Bays and estuaries
55	Other water
6	Nonforested wetland
61	Vegetated
62*	Bare
7	Barren land
71*	Salt flats
72*	Beaches
73	Sand other than beaches
74	Bare exposed rock
741	Hillslopes
75	Other barren land
8*	Tundra
81*	Tundra
9*	Permanent snow and icefields
91*	Permanent snow and icefields

*Land-use/land-cover type not found in the Denver Metropolitan Area.

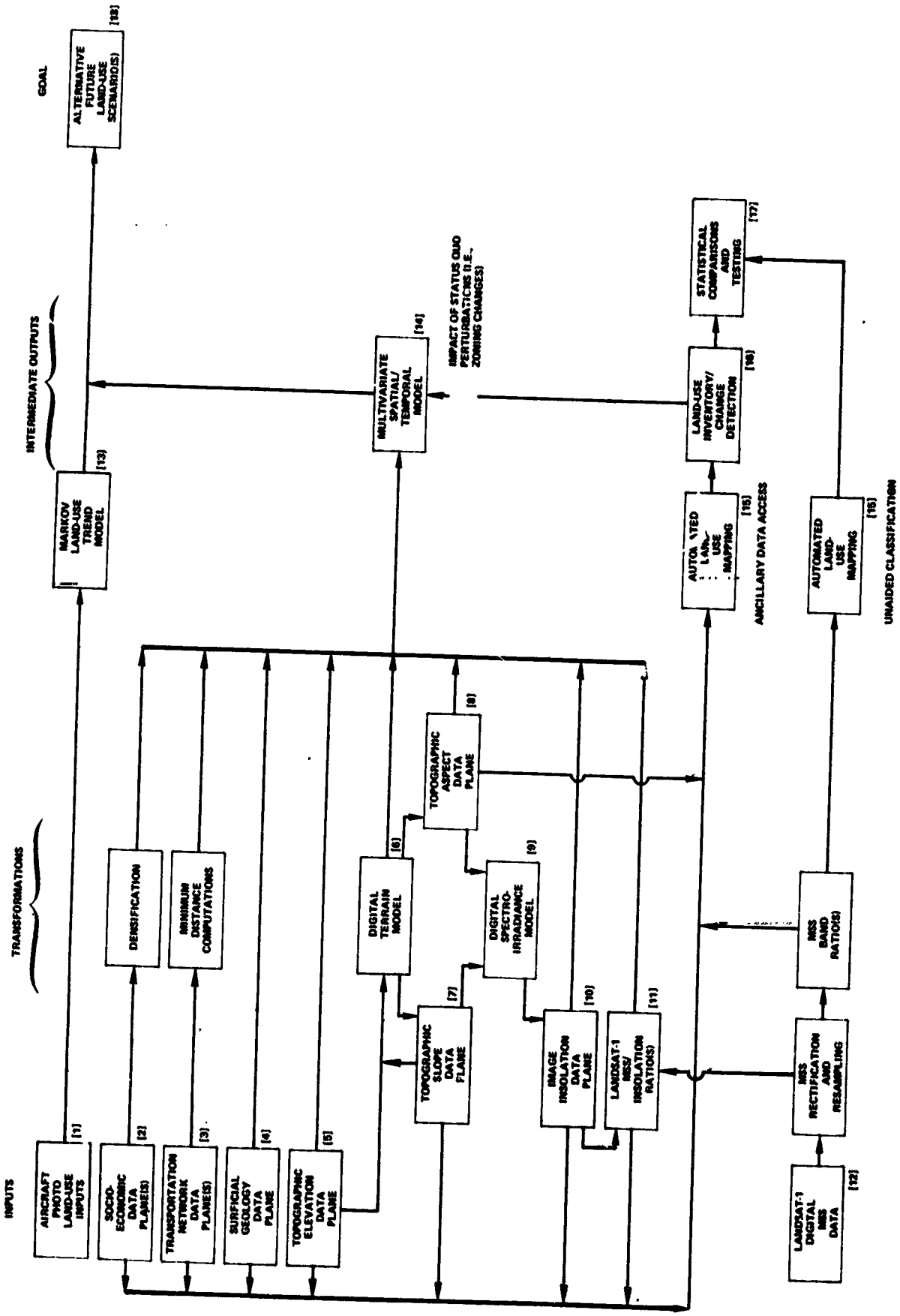


FIGURE 2. LAND-USE CHANGE ANALYSIS, MAPPING, MODELING, AND CHANGE-PREDICTION FLOWCHART. Specific inputs, operations, and outputs from the data analysis flowchart are cross-referenced to the text by bracketed numbers (*i.e.*, [1]).

cultural landscape data included census tract-aggregated socioeconomic data for 1970 [2] and a 1971 State Highway Department street/highway classification map with eight road categories [3].

Physiographic landscape data consisted of a USGS 1:62,500-scale surficial geology map [4], sixteen 1:24,000-scale topographic maps [5], and computer-generated slope [7], aspect [8], and insolation [10] data planes.

Landsat imagery resources [12] included the four basic multispectral scanner (MSS) images of a single-date 1973 scene, the six MSS image-to-image ratios, and four MSS/insolation-normalized ratios [11].

Analysis Procedures

The modeling activity was structured to analyze the interactions between land-use changes and the regulating physiographic, socioeconomic, and transportation components of the landscape. This involved the compilation, rectification, and generation of spatially registered data planes for the various landscape elements.

Each landscape data plane represented a spatially distributed, single-variable map for the 39- by 39-km* (24- by 24-statute mile) or 1491-km^2 (576-mi^2) study area centered on the Denver Metropolitan Area. The majority of these data planes were obtained by thematically sampling and computationally transforming the available maps using a fixed 192-row by 192-column dot grid yielding 36,864 cells of 4 hectares (10 acres) per plane. These planes, using the popular grid-cell technique, were accessed in overlaying operations that combined single-variable data planes with conformable spatial dimensions.

Ian McHarg's regional landscape analysis method is a well-known manual method of such overlaying processes (Reference 15). This approach physically overlays colored transparent acetate-plotted variables to view spatial correlations and juxtapositions as composite colors; however, as more variables are considered, the display becomes confusing, and the approach becomes less useful. The overlay process used in this research was carefully designed after a thorough review of the available systems (Reference 9). Implemented as a software package for use on a digital computer, it permits the easy retrieval, manipulation, modeling, and display of a large number of landscape phenomena.

Linear discriminant analysis was selected as the computational vehicle for implementing the dual land-use predictive modeling [14] and inventory mapping [15] capabilities. This dual use represented a new, multivariate statistical approach to land-use mapping. Discriminant analysis permitted each landscape cell to be represented as a point in a multidimensional, statistical framework in which each observation was a measured variable. The original observations of land use or change in land use were used to determine orthogonal axes (discriminant functions) to a possible total of one less than the number of land-use variables represented. Each new observation can be plotted and identified by means of a single discriminant score.

The mapping with Landsat MSS data or modeling future land-use patterns with ancillary map-format data required the computation of representative statistical signatures for each land use or land-use change of interest. Subsequently, each unknown individual land use, land-use change element, or landscape cell was classified according to the closest multivariate match with the nearest numeric signature subset or class.

*Throughout the text, English measurements given in parenthesis are precise, whereas their equivalents in the metric system are reasonable approximations.

Similarly, a landscape cell within this inventory/modeling system consisted of a series of spectral responses from MSS data and/or a statistical measure of several types of ancillary data from a given area within the Denver study area. This response was a vector of size equal to the total number of spectral bands and transformations and/or ancillary map variables. Both forms of data were interleaved to represent this vector on digital tapes in proper geometric relationship. Thus, the mode of analysis could be specified by an appropriate selection of image, land-use, socioeconomic, physiographic, or transportation variables.

Discriminant analysis was performed in a stepwise fashion (i.e., by entering one additional variable or attribute into the set of discriminating variables for each successive iteration). The variable to be added, but not already included in a previous iteration, was selected for inclusion on the basis of greatest F-value. Thus, the analysis operated in a "supervised" fashion on user-specified representative training sets of known land-use composition to: (1) evaluate the statistical utility of each variable as hierarchically determined by discriminant analysis, and (2) quantify the overall and incremental land-use mapping/modeling acccuracies for each variable entered. These analytical capabilities were perhaps the greatest advantages of a stepwise linear discriminant analysis approach.

Utilization of Computer Program Modules

When possible, available operational computer software was used to bypass potential "programming bottlenecks" in handling the large quantities of imagery and map data associated with the study.

The TOPOMAP program (References 16 and 17) created a digital terrain model [6] with the elevation data to compute topographic slope [7] and aspect [8] data planes. These topographic slope and aspect data planes were input to a digital spectroirradiance model (program INSOL2) [9] to compute a near-instantaneous incoming solar radiation or insolation data plane [10].

Program TRANSF2 from the Colorado State University (CSU) Landsat mapping system (LMS) (Reference 18) was used to create the six MSS ratios and four MSS/insolation ratios [11]. Other Landsat-1 preprocessing operations performed by the LMS package included data reformatting and tape merging [12].

Program PLANMAP (Reference 1), a geographic information system, compared the 1963 and 1970 land-use data planes and produced urbanization rates for the various land-use classes to drive a Markov chain successional trend model [13].

Automated image classification with and without ancillary data access [15] was tested with the channel selection feature of the EXTRACT and CLASSIFY linear discriminant analysis programs in LMS. CLASSIFY represented an adaptation of the BMD07M stepwise multidiscriminant analysis routine in the biomedical design (BMD) program series (Reference 19). This stepwise approach also permitted the selection of an optimal subset of image and map variables for a full-image classification [16].

The simple tabulation programs, CHECKER1 and CHECKER2, were written to facilitate a point-to-point comparison of the classification map file to the 1972-1973 USGS land-use reference plane for the multilevel tabulation of land-use mapping accuracy [17].

Finally, extended development and interrelated use of these land-use inventory and modeling programs indicated that it was possible to develop even more sophisticated models to assess the limitations and impacts from, or response to, various land-use scenarios [18].

DETAILED REPORT ORGANIZATION

Chapter 1, "Introduction," describes the study and provides general comments on land-use planning, background, and rationale, objectives, scope, approach, and the analysis plan.

Chapter 2, "Landscape Model Construction," presents the map sampling and compilation methodology involved in constructing the land-use, physiographic, transportation, and socioeconomic map data planes.

Chapter 3, "Spatial Land-Use Projection," describes the spatial land-use projection models, testing, and results, and outlines a revised spatial-change modeling strategy that combines both the Markov trend and discriminant analysis models.

Chapter 4, "Landsat Land-Use Classification," explains the Landsat land-use classification effort. It deals with image rectification and resampling, ancillary map data overlaying, feature extraction sampling methods and verification, and classification algorithm selection and testing. It also gives the costs of developing the data and optimal approach. Finally, this approach is applied to the total study area in a cost-effective format, and displays and verifies the results achieved.

Chapter 5, "Conclusions," summarizes the basic endeavors, activities, and results of the study. Applications areas are identified, and various recommendations are made. The recommendations generally concern ancillary data inputs and machine interpretation and processing.

Five appendixes support the main text:

- Appendix A, "Multiple Discriminant Analysis," presents a brief overview of the multivariate statistical technique of linear-discriminant analysis, used in this study for both spatial land-use projection and Landsat land-use classification.

- Appendix B, "Landsat Mapping System" (LMS), describes the set of programs making up the Landsat Mapping System that has been implemented at Colorado State University.

- Appendix C, "Machine Classification Error-Rate Estimation," tabulates omission/commission rates for the first- and composite second- and third-order USGS land-use classes.

- Appendix D, "Maximum-Likelihood Ratio," describes the widely used multivariate classification algorithm tested against linear-discriminant analysis in this study.

- Appendix E, "Correlation Matrices for Rectangular/Point-Sampled Training Sets," presents these statistics.

This Page Intentionally Left Blank

CHAPTER 2

LANDSCAPE MODEL CONSTRUCTION

PRECEDING PAGE BLANK-NOT FILMED

CHAPTER 2

LANDSCAPE MODEL CONSTRUCTION

INTRODUCTION

The evolution of land use in the United States is one of the most perplexing problems confronting our society. It involves a tangled web of biological, economic, environmental, institutional, physical, and political interactions. The land conversions that result can be studied in terms of aesthetics, aging, birth rates, climate, crowding, economics, industry, legislation, mineral extraction, mortality, public services, social amenities, technology, topography, transportation, vegetation, water, and zoning.

It was concluded that the most fruitful approach to understanding the evolution of urban land use was by the construction and manipulation of a landscape model of the Greater Denver Metropolitan Area. This landscape model provided a better understanding of Denver's physical environment from a joint analysis of the various remote sensing images, collateral landscape maps and data (e.g., land-use, physiographic, socio-economic, and transportation maps), and a literature survey for the historical period under study. These collateral data stored in the landscape model provided the data base for the analysis necessary for understanding the present and recent past patterns of land-use evolution in the Denver Metropolitan Area. The analysis also yielded insights into the biological, economic, environmental, and social driving forces that have induced the rural-to-urban land-use conversions observed. Most important, these collateral data were indispensable as both spatial and statistical geographic inputs to the evaluation, analysis, and presentation of a quantitative landscape model for predicting future spatial changes in Denver land use.

The cultural and natural landscape components of land use are not only undergoing a dynamic state of succession, but are also subject to quite varied and complex driving forces. Accordingly, the landscape modeling effort focused its time and resources on selecting ancillary data factors that were believed to be of broad applicability, usefulness, and general availability for the intensively urbanized areas of the United States. Thus, in this undertaking, special emphasis was placed on validating and utilizing meaningful and widely available data input sources to the landscape modeling to allow replication by interested researchers or planners for other metropolitan areas. Although some data sources will not be universally available throughout the country, it was anticipated that very close, if not identical, surrogates would exist in the majority of cases. Preference was always given to universally available data such as Department of Commerce census data, USGS topographic and geologic maps, National Aeronautics and Space Administration (NASA) satellite imagery, and historical sequences of airphotos in contrast to equally usable but more localized and volatile data sources. Even these more universally available data have been collected with diverse techniques and map scales by a wide assortment of public and private organizations. The landscape model provided a depository for interrelating such diverse information in a common format. A model variable or data plane is one overlay of spatially registered data in a common cellular network upon all other data planes or variables in the landscape model (figure 3).

ORIGINAL PAGE IS
OF POOR QUALITY

7 LAND-USE SUBMODEL VARIABLES:
- 1963 Photo Land-Use
- 1970 Photo Land-Use
- 1972 to 1973 USGS Photo Land-Use
- Two 1963 to 1970 Land-Use Changes
- Two 1963 to 1970 Alphanumeric Coded Land-Use Changes

9 PHYSIOGRAPHIC SUBMODEL VARIABLES:
- Topographic Elevation
- Topographic Slope
- Topographic Aspect
- USGS Surficial Geology
- Solar Insolation for Landsat Imagery
- Four Landsat MSS ÷ Insolation Ratios

5 TRANSPORTATION SUBMODEL VARIABLES:
- Composite Minor-Road MD
- Composite Major-Road MD
- Freeway MD
- Freeway Interchange MD
- Built-Up Urban-Area MD

17 SOCIOECONOMIC SUBMODEL VARIABLES:
- Four Population/Housing Densities Per Acre
- Five Population/Family/Housing-Unit Totals
- 1969 Mean Family Income
- Median Housing-Unit Rent/Value
- One-/Two-/Three-Car Family Totals
- Total Census Tract Acreage
- Average Number of Cars Per Family

10 LANDSAT IMAGE SUBMODEL VARIABLES:
- MSS-4 (Visible Green)
- MSS-5 (Visible Red)
- MSS-6 (Solar Infrared)
- MSS-7 (Solar Infrared)

FIGURE 3. CONCEPTUAL DIAGRAM OF THE LANDSCAPE MODEL OF THE DENVER STUDY AREA. Available maps, spatially referenced tabular data, and remote sensing imagery were assembled, interpreted, registered, and overlaid into a cellular landscape model with 48 image/map variables. This landscape model provided the data base for understanding the present and recent past patterns of land-use evolution and their causal factors (MD = minimum distance)

CONSTRUCTION OF A SPATIAL DATA PLANE

At the outset of the design of the landscape model, a complete inventory was prepared of all available maps, spatially referenced tabular data, and remote sensing imagery, ranging from the early low-altitude black-and-white airphotos of the mid-1930's to current Landsat imagery (Reference 1). The data planes to be input to the landscape model were selected from this thorough inventory. New data were continually sought and evaluated as available data were found to be incomplete, incompatible, or inadequate for the landscape model. Caution was taken to ensure that not too much money and energy were expended on collecting and organizing modeling data so that adequate resources remained for constructive analysis and synthesis of what it all meant. A marked tendency exists among those now using some form of landscape modeling, especially in smaller planning agencies, to think that assembling data planes in composite mapping systems constitutes planning and analysis, whereas it is in actuality only a necessary prerequisite. These planners and experimenters are currently enamored with sophisticated processes for inputting huge amounts of data into some cellular overlay framework. Some 95 percent of the thought goes into the input of the data, whereas a mere 5-percent, last-minute attempt is made to determine how to analyze it. This effort has selected the simplest means possible for overlaying each data plane onto the landscape model and has concentrated the bulk of its attention on analyzing the data planes.

Map Sampling Procedure

Potential landscape modeling variables were evaluated for general availability and significance. Thirty-four map variables were selected for inclusion in the landscape model and were divided into land-use, physiographic, socioeconomic, and transportation submodels (table 2).

Some explanation of the two general types of variables that were overlaid on the model is in order here. A categorical variable denotes the presence or absence of a category or class. An observation of this type is mutually exclusive (that is, it can fall into only one distinct preselected class or category). Thus, each land-use data plane has 24 specific classes, whereas the surficial geology variable in the physiographic submodel has 13 specific classes. A numerical variable refers to a purely digital sequence or range of continuously varying numbers. Observation values can be directly compared with one another in a mathematical sense. Topographic elevation and the minimum distance to minor roads are examples of numerical variables.

The landscape model covered an area of 39 by 39 km (24 by 24 mi) or 1491 km^2 (576 mi^2) centered on the city of Denver (figure 4). Most of this study site has relatively low relief except for about 5 percent of the area along the southwestern edge, which includes the eastern foothills of the Rocky Mountains. The Greater Denver Metropolitan Area is entirely contained within this area and includes Denver proper, which is a rapidly expanding population center of approximately 1,500,000 people with future expectations for continued growth.

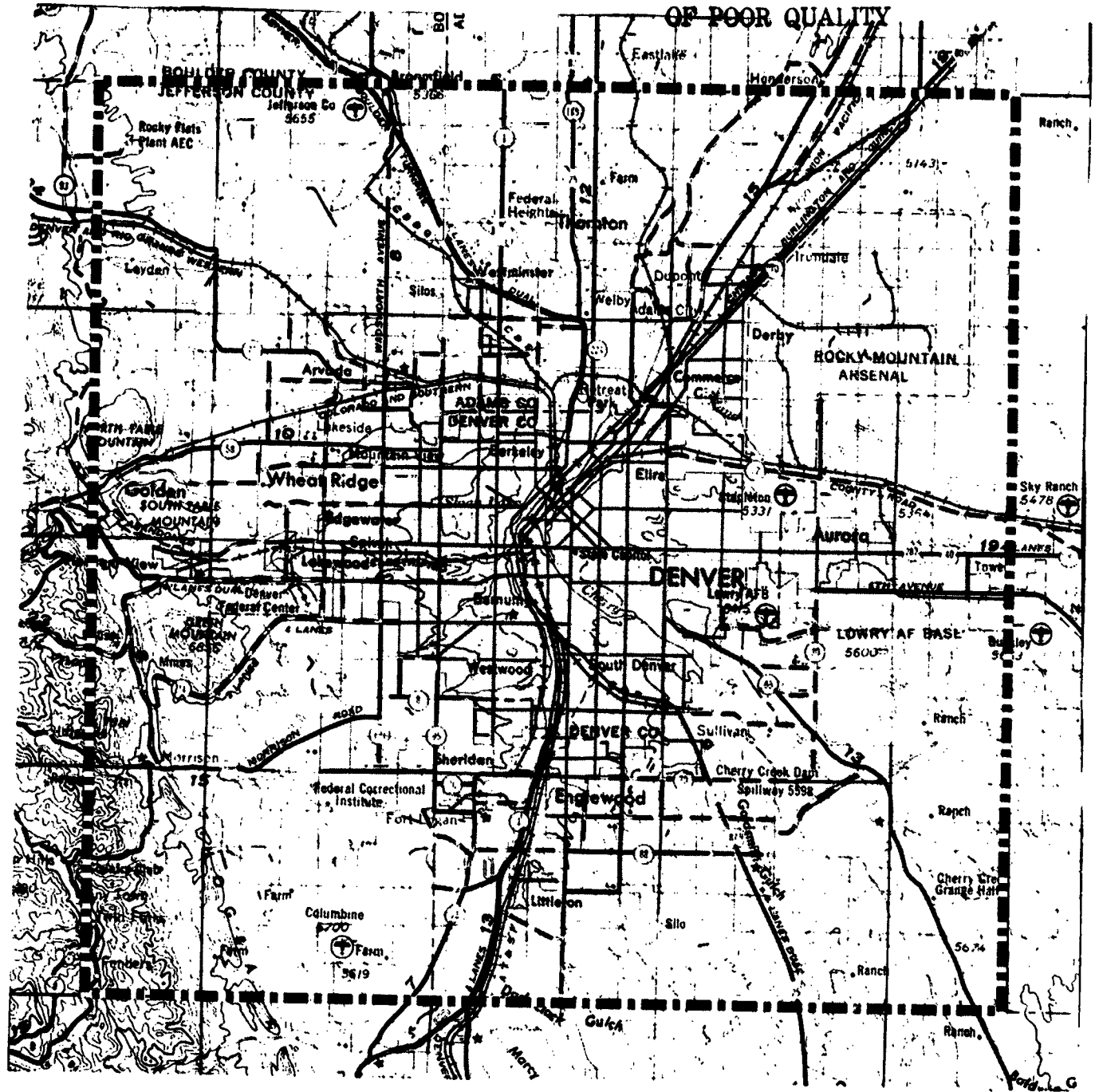
The 1491-km^2 (576-mi^2) study area was divided into 192 north-south rows and 192 east-west columns for a total of 36,864 4-$hectare^2$ (10-$acre^2$) mapping cells. This cellularization permitted any spatially distributed two- or three-dimensional landscape variable to be rectified, sampled, encoded, stored, retrieved, and displayed by its implicit position in this predetermined computer-compatible grid network. Adopting this landscape modeling approach made it practical and feasible to construct and overlay geographical data planes of multiple landscape attributes for subsequent computer analysis, modeling, and land-use change predictions.

TABLE 2

LIST OF SPATIAL LANDSCAPE MODELING VARIABLES. The 24- by 24-mile Denver Metropolitan Area was modeled by 38 collateral nonimage data planes with a 10-acre square-cell element. These lists of ancillary variables are arranged into four functional submodels. A "categorical variable" is a data plane that represents map classes or categories with no implicit numerical relationship within the numbers or characters selected to represent them. A "numerical variable" represents a data plane that varies continuously in a mathematical sense (i.e., an elevation of 4000 feet is two times an elevation of 2000 feet).

Landscape Submodel	Landscape Variable	Source of Data	Variable Type
Land use	1963 photo land use	1:20,000-scale B/W photos	Categorical
	1970 photo land use	1:24,000-scale orthophotos	
	1972-1973 USGS photo land-use	1:100,000-scale USGS map	
	1963 to 1970 land-use changes (from)	1963 to 1970 land-use data planes	
	1963 to 1970 land-use changes (to)		
	1963 to 1970 alphanumeric land-use changes (from)		
	1963 to 1970 alphanumeric land-use changes (to)		
Physiographic	Topographic elevation	1:24,000-scale USGS topographic maps	Numerical
	Topographic slope	Computed from elevations	Numerical
	Topographic aspect	Computed from elevations	Numerical
	Surficial geology	1:62,500-scale USGS map	Categorical
	Landsat image insolation	Computed from slope/aspect	Numerical
	Landsat MSS/insolation ratios	Computed as MSS÷insolation	Numerical
Transportation	Composite minor road minimum distance (MD)	Computed from 1:45,000-scale state highway map	Numerical
	Composite major road MD		Numerical
	Freeway MD		Numerical
	Freeway interchange MD		Numerical
	Built-up urban area MD		Numerical
Socioeconomic	Total population	1970 census reports and 1:84,500-scale census tract map	Numerical
	Total families		
	Total year-round housing units		
	Total vacant housing units		
	Total occupied housing units		
	1969 mean family income		
	Median housing-unit value		
	Median housing-unit rent		
	Total one-car families		
	Total two-car families		
	Total three-/three-plus car families		
	Total census tract acreage	Computed	
	Population density per acre	Computed	
	Average number of cars per family	Computed	
	Average number of families per acre	Computed	
	Average number of year-round housing units per acre	Computed	
	Average number of vacant housing units per acre	Computed	

ORIGINAL PAGE IS
OF POOR QUALITY



CONTOUR INTERVAL 200 FEET
WITH SUPPLEMENTARY CONTOURS AT 100 FOOT INTERVALS
TRANSVERSE MERCATOR PROJECTION

FIGURE 4. TOPOGRAPHIC MAP OF THE DENVER STUDY AREA. The outer boundary of the Denver study area is a square of 24 by 24 miles enclosing the Denver Metropolitan Area.

As noted earlier, the simplest way of inputting each data plane into the landscape model was adopted. Generally, this required overlaying a dot pattern representing a selected cell size on the map (Reference 19). Topographic elevation data, for example, is an essential part of any landscape model (figure 5). The input value of each elevation cell was estimated from 1:24,000-scale topographic maps at the position of the sample dot for each of the 36,864 cells that constitute the elevation data plane. Very careful procedures were used to ensure that the common cell pattern overlaid by hand in this fashion on each new map or set of maps provided a data plane that registered exactly, cell-by-cell, on all existing data planes (figure 6). This hand tabulation of all the map-format data for analysis was very laborious but was more accurate for these initial research efforts in landscape modeling. When the basic principles of the approach are better understood, new landscape modeling efforts can input the data with a variety of more sophisticated map digitizing procedures. Many earlier studies of this nature dealt with more complex machine entry of map data into the computer and, as a result, expended less effort on its constructive analysis.

A semitransparent multipurpose overlay data form (figure 7) was devised to combine both the coding and the keypunch operations. This 1:24,000-scale data form overlaid a quadrant of a township consisting of nine sections of land of 1.6- by 1.6-km (1- by 1-mi) squares, with 64 rectilinearly distributed sampling dots per section representing the 4-ha (10-acre) square resolution sampling unit. A 4-ha (10-acre) square cell was selected because it represented the smallest individual area that could be adequately sampled from the majority of the available 1:24,000-scale source maps. Data contained on each 1:24,000-scale map were cellularized by overlaying the data form on the map to match the boundaries of the appropriate township quadrant. The desired map information was next traced onto the overlay, and the category identification was recorded in each area (figure 8). This overlay form was the direct input to the keypunch operation, and no further transcription of the data was required.

Compilation of the Data Plane

Each horizontal row of dots on the semitransparent overlay form represented one punch card, and the keypuncher generated a deck of 24 cards directly from each 23-km^2 (9-mi^2) area of 24 by 24 cell elements (figure 9). The effort of keypunching the data from these sheets was minimized by punching only the leftmost identity of a land-use change at the proper position on the card representing its column. The cell code of identical row elements was identified with successive cells left blank until the left-hand boundary of a different land-use category was reached. This process facilitated keypunching and subsequent verification steps because fewer keystrokes were made, and the suppression of identical elements aided visual proofreading by comparing a line-printer display of the data deck with the original code form (figure 10a). Auxillary computer programs were available to fill in the blank cells in a data deck (figure 10b), to assemble and display each total data plane from a group of data decks (figure 11), and to display a selected group of characteristics in a data plane (figure 12).

When each source map has been sampled and assembled, it constitutes a data plane of the landscape model. Each of these data planes can be selectively displayed as a computer microfilm graphic (graymap) to illustrate the spatial distribution of one or more categories or numeric ranges. As an example of these high-resolution computer graphics, a 1:250,000-scale graymap can illustrate all first-order urban and built-up lands contained in the 1970 land-use data plane (figure 13). A second microfilm of the same data plane emphasizes all first-order agricultural land (figure 14).

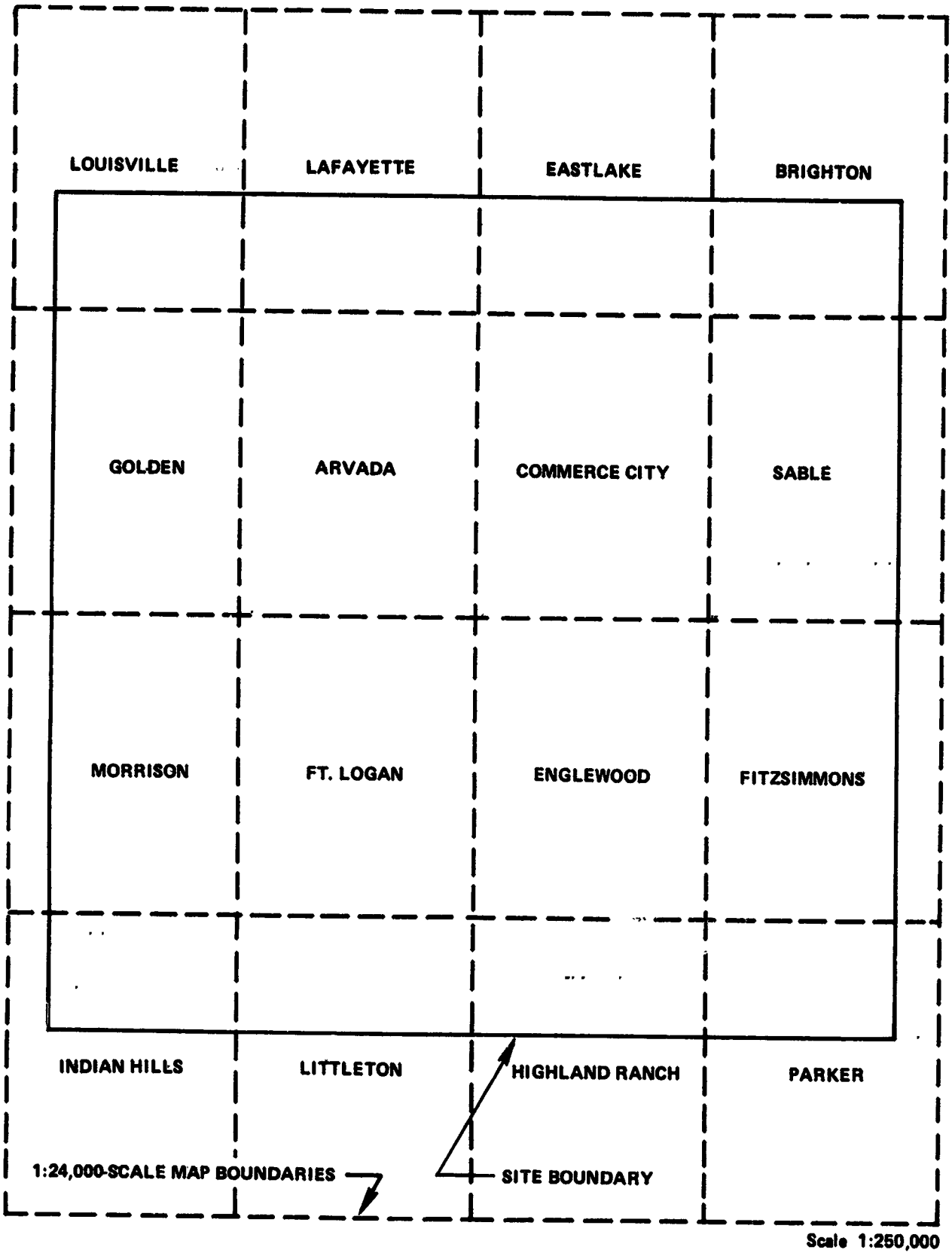


FIGURE 5. TOPOGRAPHIC MAP INDEX OF THE DENVER STUDY AREA. Sixteen 1:24,000-scale, 7.5-minute USGS topographic maps covered the Denver study area and were input into the elevation data plane.

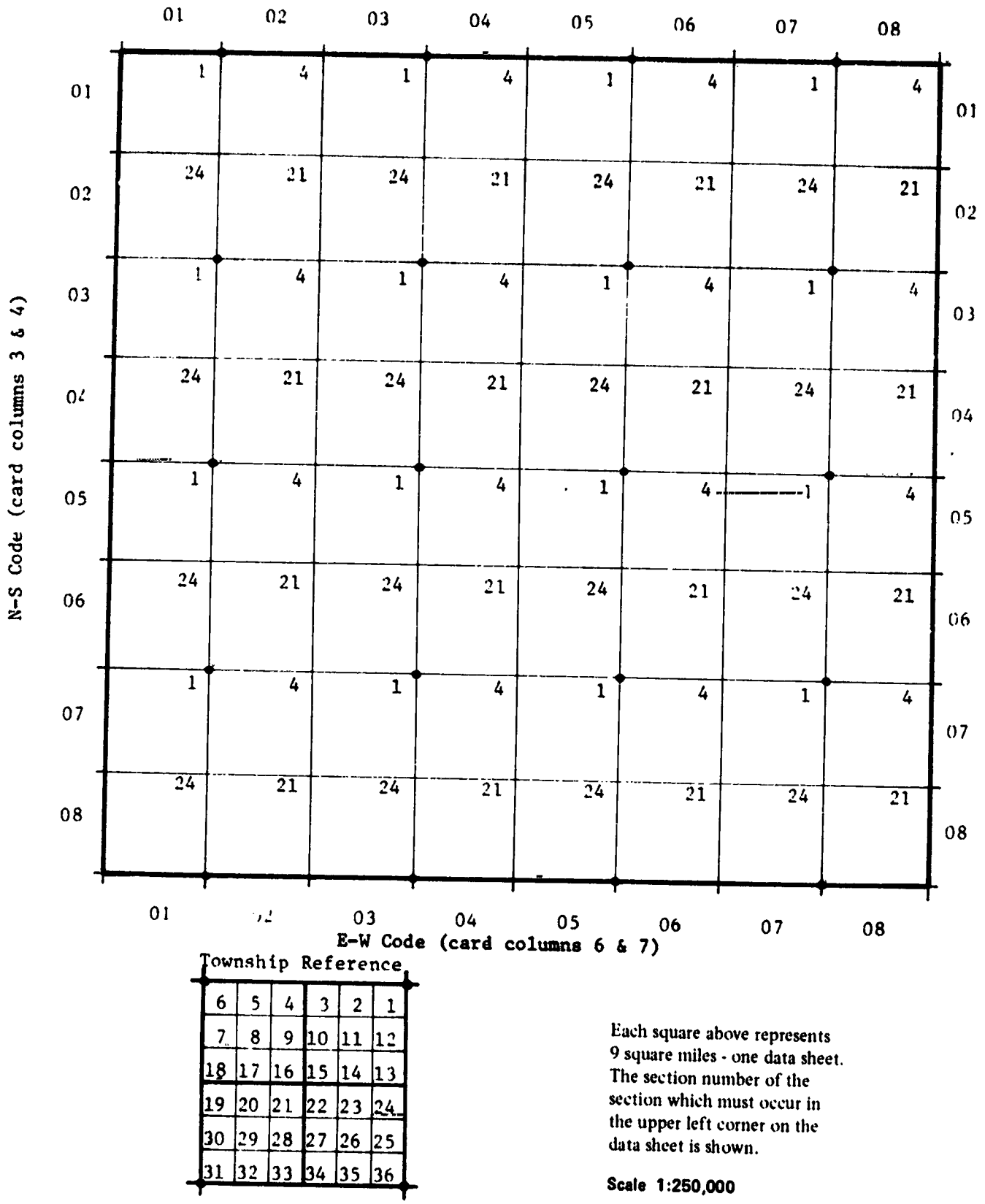


FIGURE 6. GEOGRAPHIC REFERENCE OVERLAY FOR DATA FORMS. The exact location and internal grid row and column identifiers for the data coding forms (figures 7 and 8) are shown relative to the USGS topographic map boundaries (figure 5). Four data coding forms are used for the four quadrants of each township.

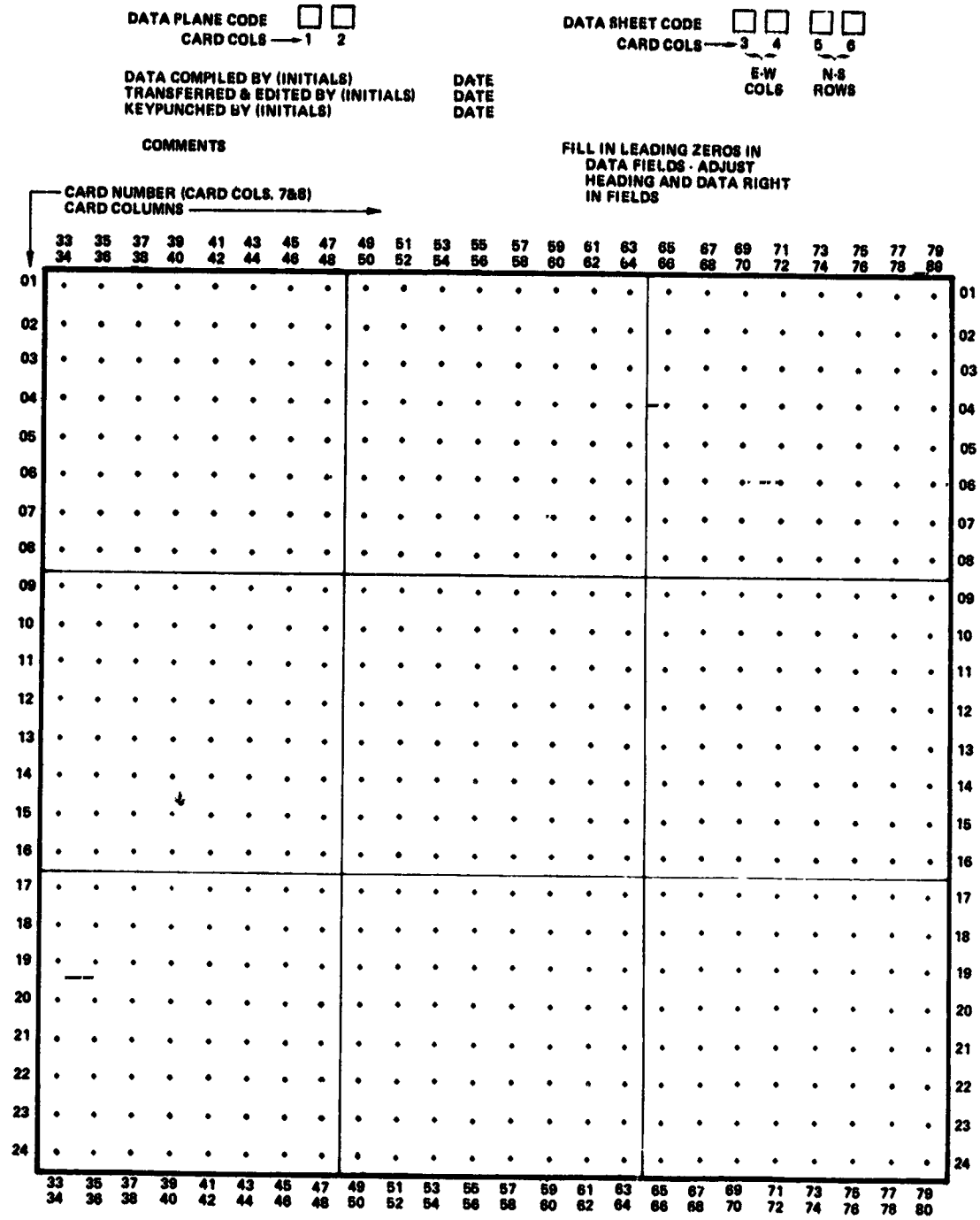


FIGURE 7. OVERLAY DATA FORM USED FOR SAMPLING A 1:24,000-SCALE MAP. The nine large squares in a three-by-three array overlay individual sections or square miles of land on the appropriate 1:24,000-scale map. The eight-by-eight array of dots in each of these nine sections represents a sampling of each 10 acres, or 1/64 square mile per cell. The original form is printed on transparent paper and can be drawn on while overlaying a particular source map. The form shown here has been reduced to a scale of about two-thirds of its original 1:24,000 scale.

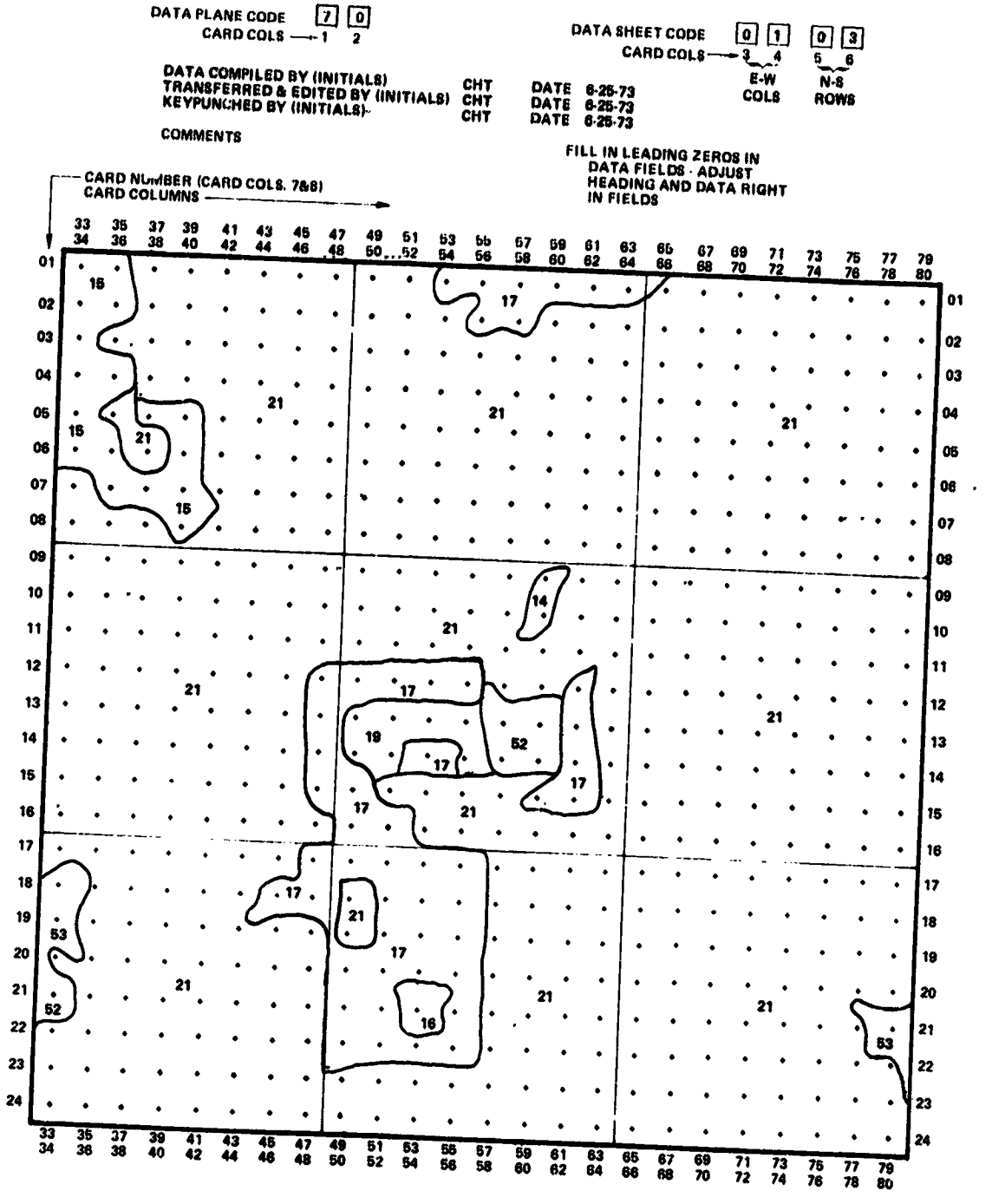


FIGURE 8. COMPLETE OVERLAY DATA FORM READY FOR KEYPUNCHING. Keypunching is easily completed directly from the form as shown. The "DATA PLANE CODE = 70" (upper left corner) indicates that the data results from the 1970 land-use maps. The "DATA SHEET CODES = 01 and 03" indicates the unique position of this block of 9 square miles of data, a township quadrant, with respect to the balance of the 576 square-mile Denver study area, as shown by this row and column counter obtained from the reference map of figure 6. Each of the 24 lines of dots represents two columns on a single computer card. To conserve keypunching effort, only the card columns to the right of any boundary indicating a change in land use are punched with the two-digit land-use code of the area to the right (figure 9). The form shown here has been reduced to a scale of about two-thirds of its original 1:24,000 scale.

ORIGINAL PAGE IS
OF POOR QUALITY

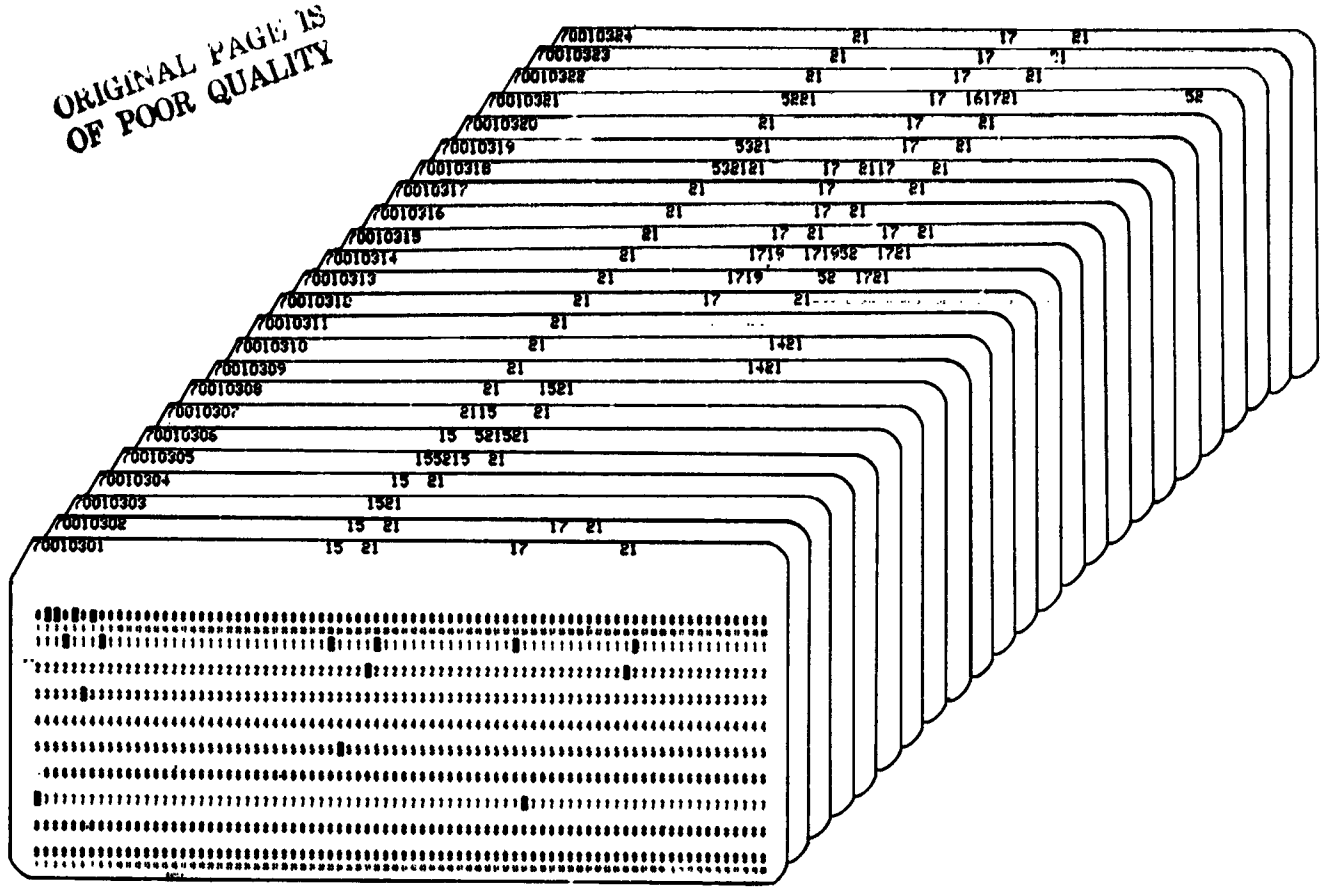
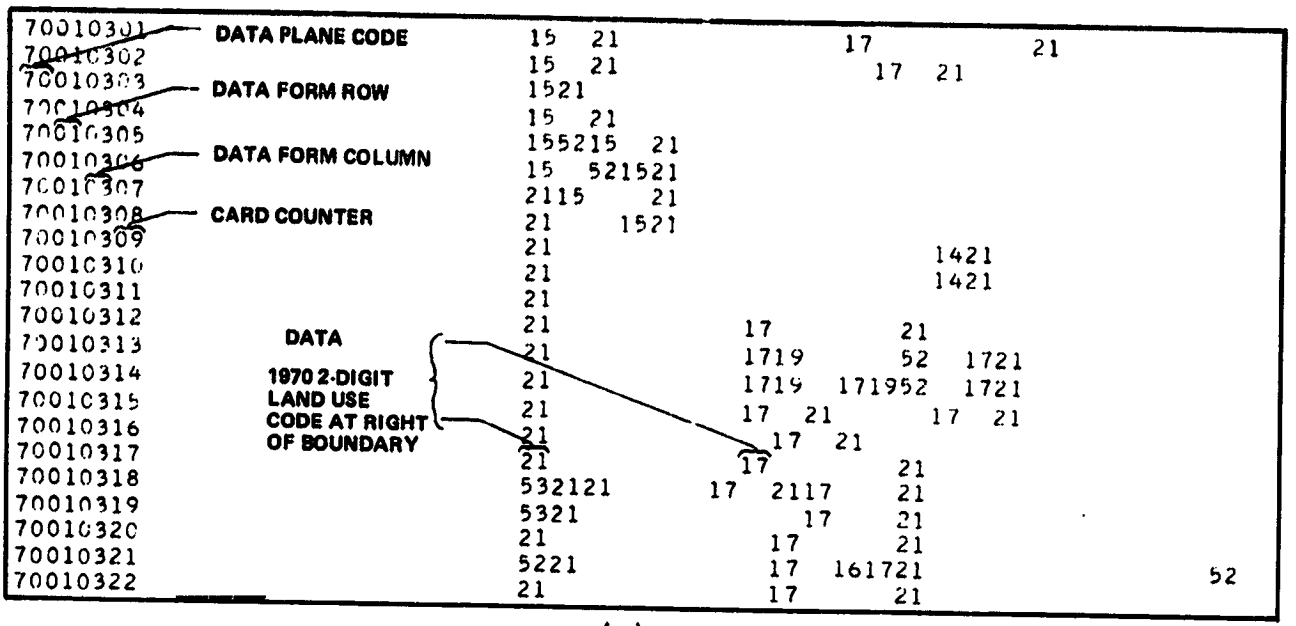


FIGURE 9. A BLOCK OF PUNCHED CARDS REPRESENTING ONE OVERLAY DATA FORM. Reading from front to back, these 24 cards represent exactly what is keypunched from the completed data form shown in figure 8. Note that only a minimal amount of keypunching and visual verification is needed to represent and check the map boundaries. Simple computer routines expand these data to fill every landscape cell.

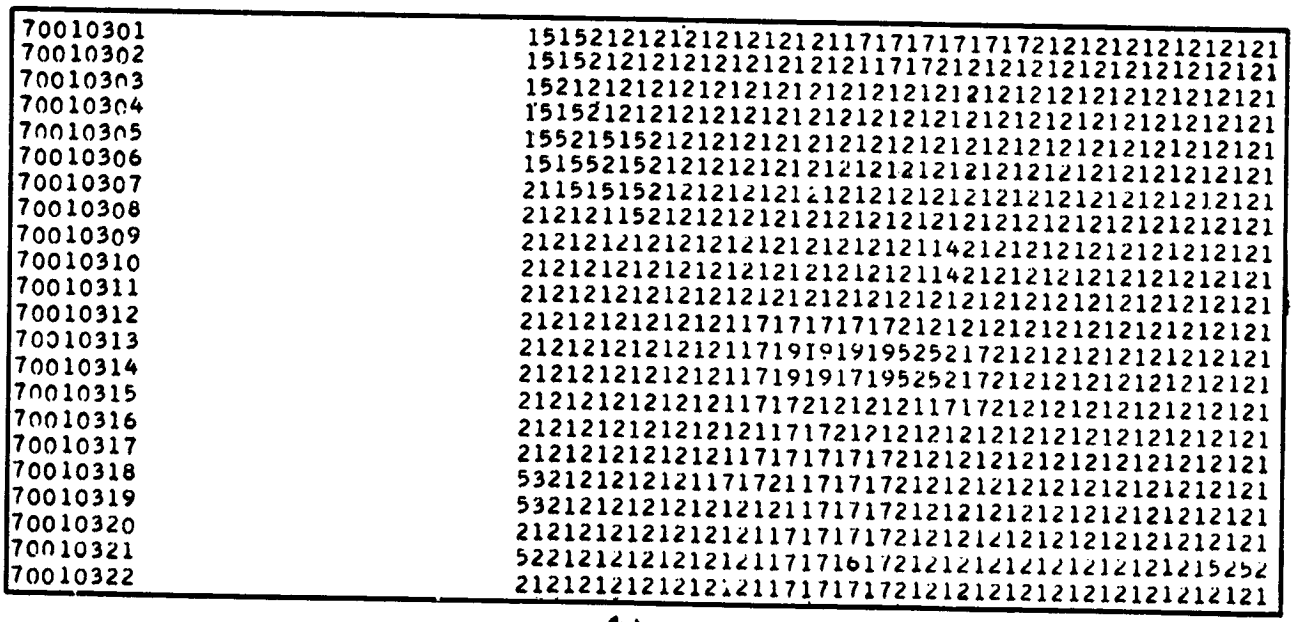
CONSTRUCTION OF LAND-USE SUBMODEL

The 1970 land-use data already illustrated was obtained from 1:24,000-scale orthophotomaps photointerpreted for thematic land use with the remote sensing-based classification scheme of Anderson, Hardy, and Roach (Reference 10), transferred to the semitransparent overlay data forms, and dot-sampled and assembled as described. When an unrepresentative classification was made on the basis of the dot designation alone in areas of finely mixed land uses, the dominant land use was assigned with priority to the lower-valued categorical urban codes (versus the higher-valued categorical rural codes).

Selective display graymaps of this data plane show that recent developments in land use, such as clustered development, occurred on the periphery of existing urban lands and placed new pressures on adjoining open-space lands by providing nodes or springboards for additional urbanization (figures 15 and 16).



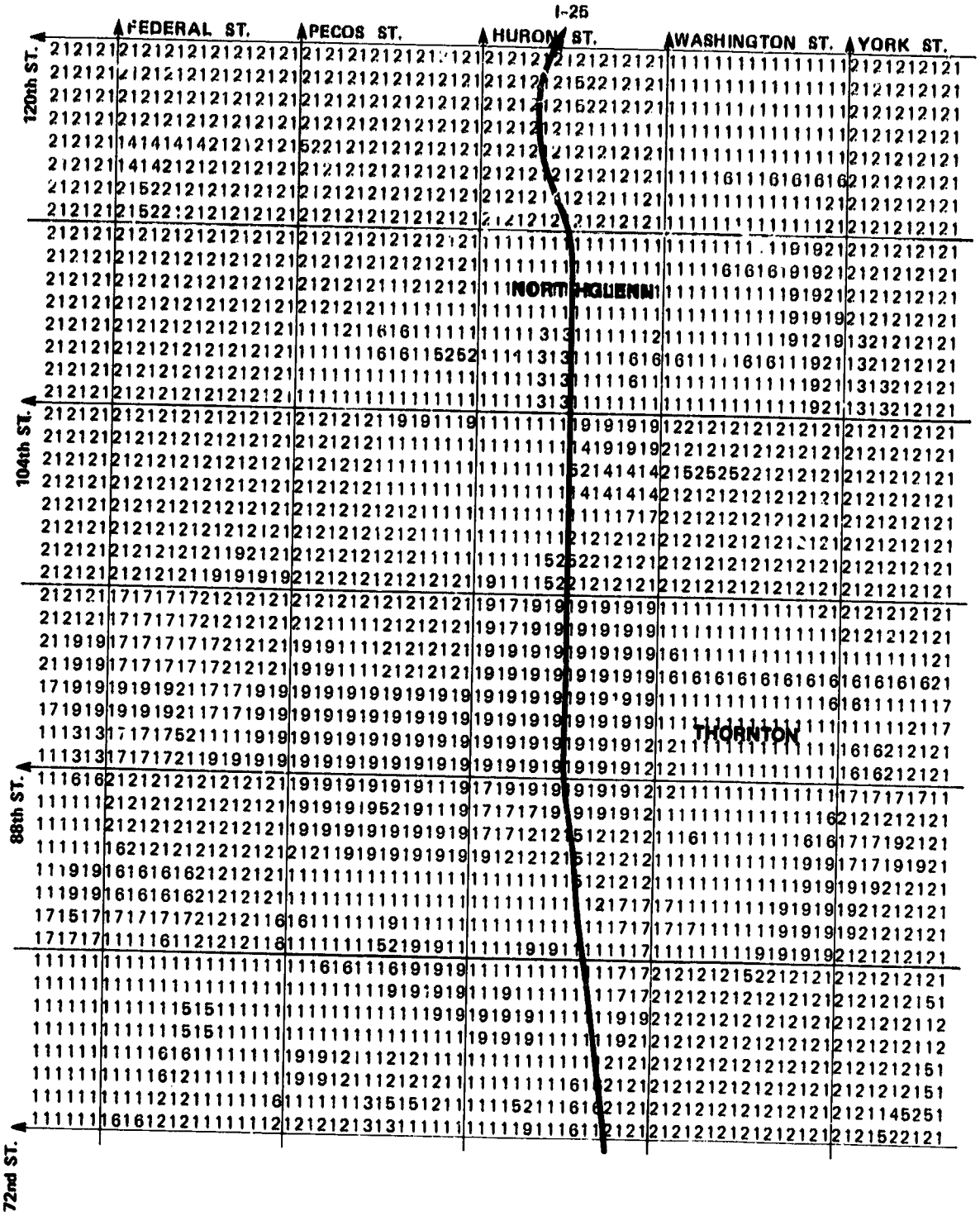
(a)



(b)

FIGURE 10. DISPLAY OF THE DATA ON ONE KEYPUNCHED OVERLAY DATA FORM. (a) A direct, offline listing of the card deck shown in figure 9. (b) A computer listing showing these data expanded to fill in the intermediate 10-acre cells with the appropriate two-digit land-use code.

ORIGINAL
OF POOR



Display scale ~ 1:33,000

FIGURE 11. A PORTION OF THE 1970 DENVER LAND-USE DATA PLANE. Each two-digit number pair represents a second-order land-use code (table 1) for a 10-acre cell of approximately 660 by 660 feet. The computer has aggregated or assembled together the eight by eight set of 64 data forms keypunched for each 9-square-mile area of the 576-square-mile site. Each of these 64 data forms yields a data deck of 24 computer punch cards, and the 1536 punch cards therefore represent one data plane such as the portion of the 1970 land-use map represented here.

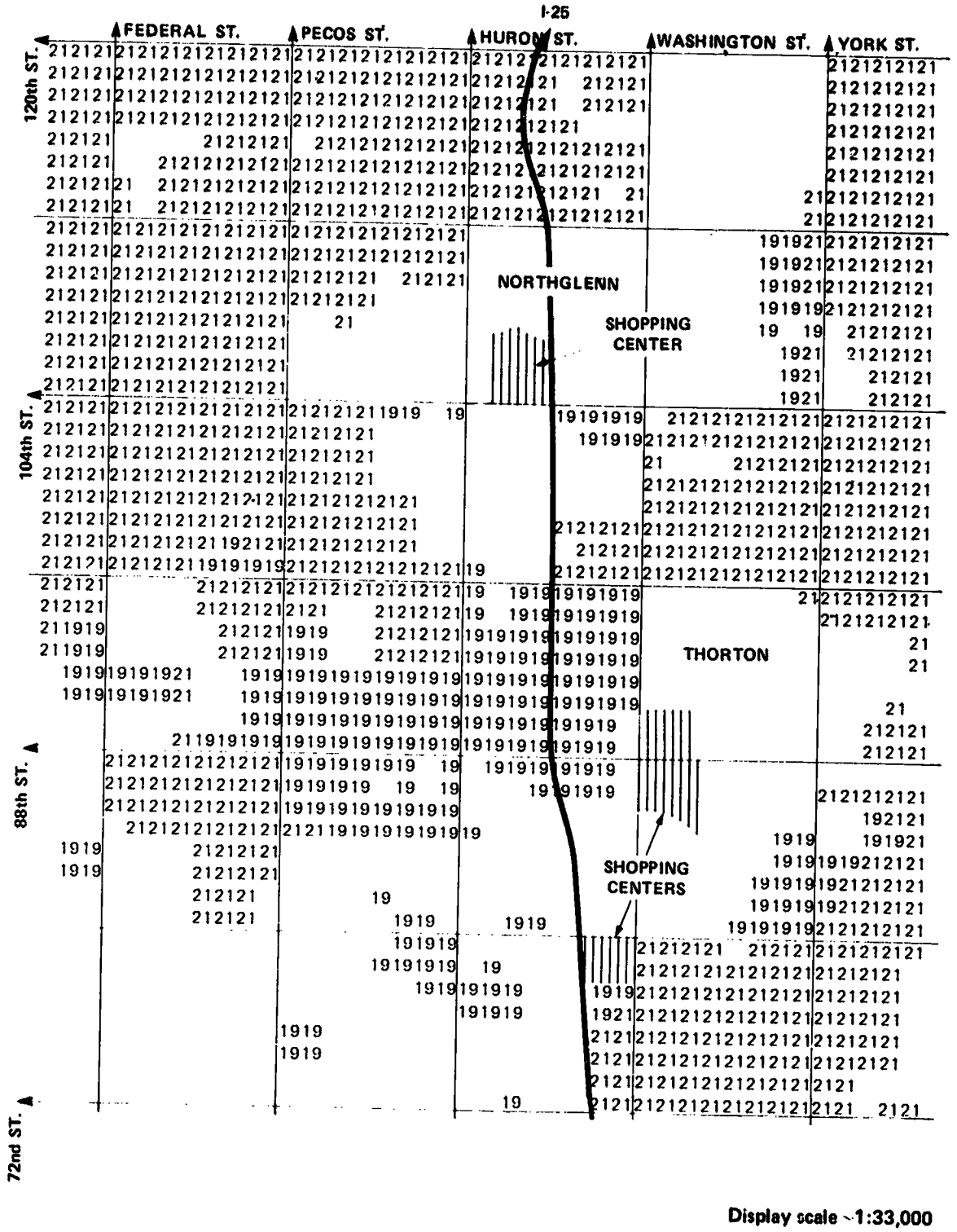
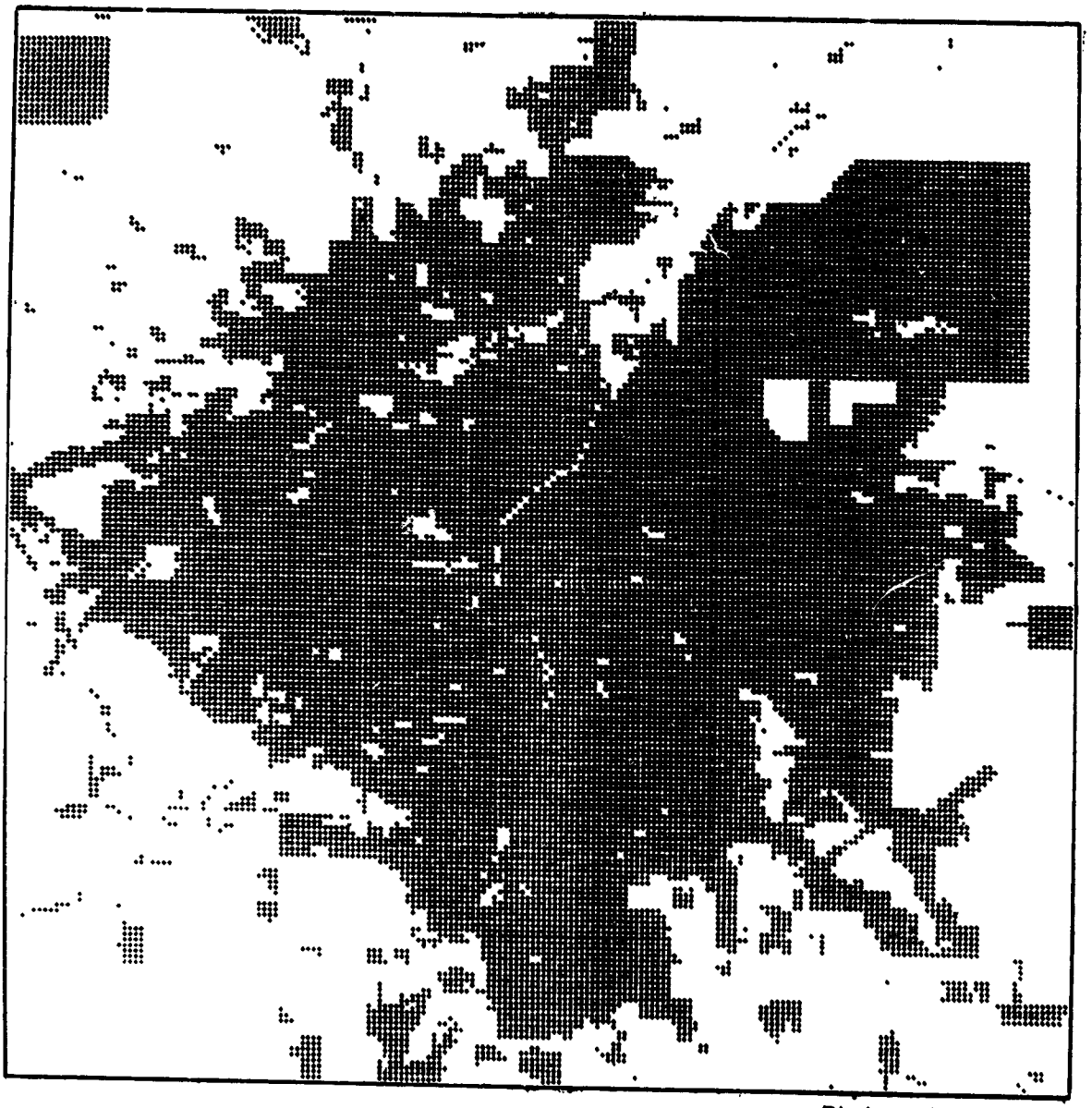


FIGURE 12. A SELECTIVE "OPEN-SPACE" DISPLAY OF A PORTION OF THE 1970 DENVER LAND-USE DATA PLANE. Two of the second-order land-use classifications, specifically 19 = Open and Other Urban, a subclass of Urban and Built-Up Land, and 21 = Cropland and Pasture, a subclass of Agricultural Land, have been selectively displayed to represent a rough categorization of open space. These two codes have been printed at the geometric position where they occur. The other 19 of the 21 existing second-order 1970 land-use codes have been suppressed and are not printed.

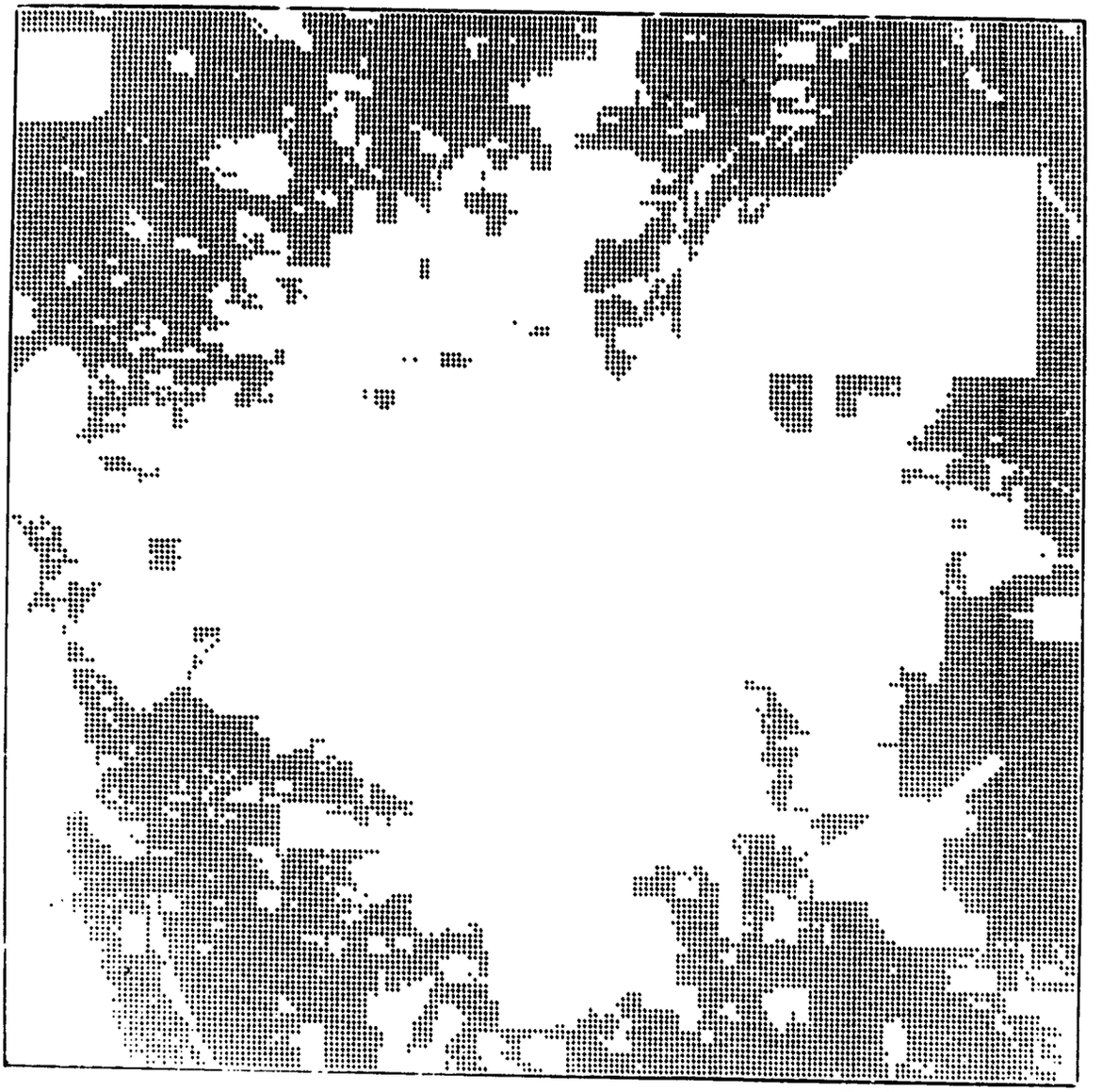
ORIGINAL
OF POOR QUALITY



Display scale 1 : 250,000

BLACK = All urbanized land-use classes (codes 11 through 19 inclusive)
WHITE = All other land uses

FIGURE 13. DISPLAY OF THE 1970 LAND-USE DATA PLANE EMPHASIZING ALL OF THE URBAN AND BUILT-UP AREAS (in black). The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

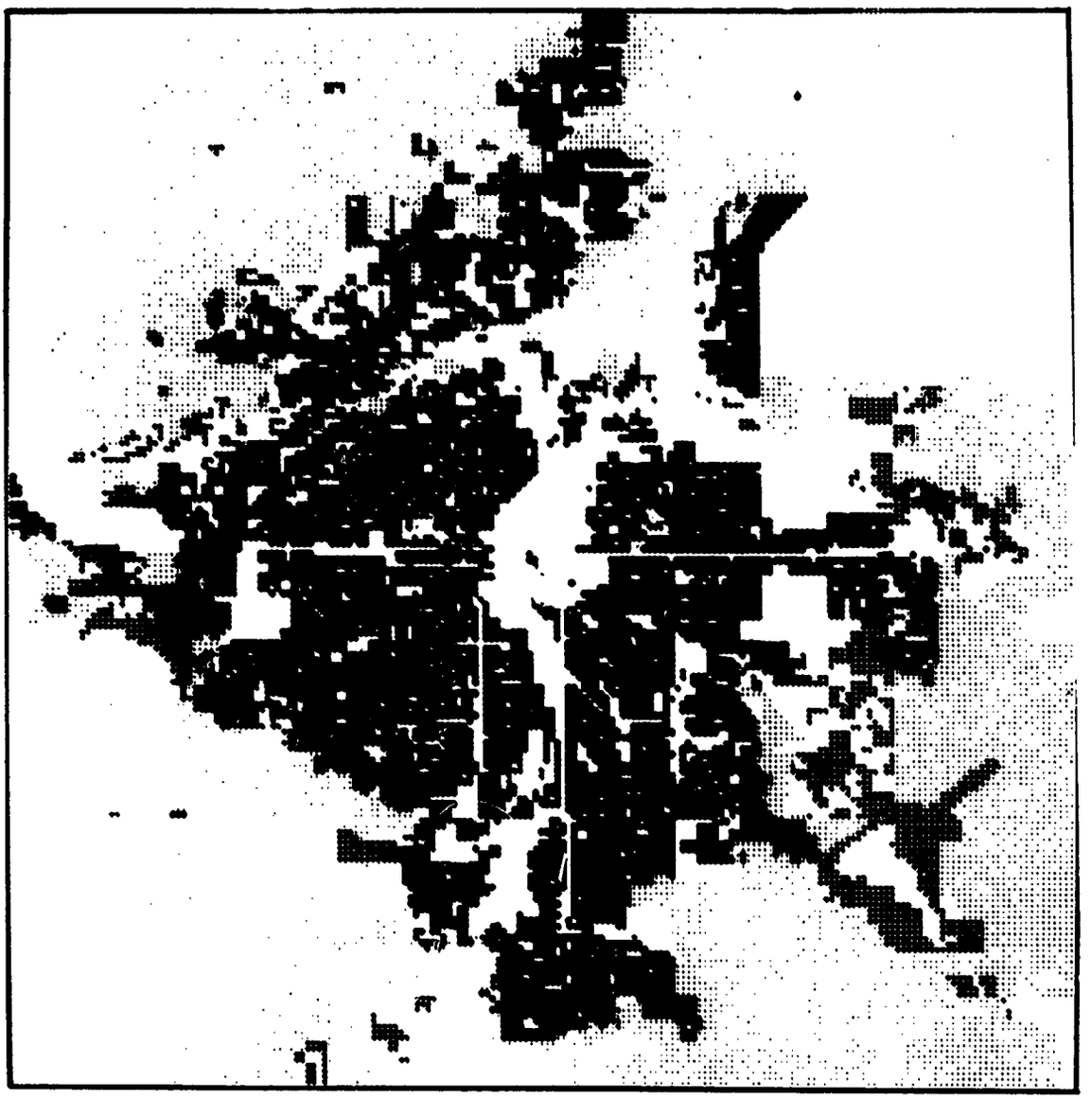


Display scale 1:250,000

BLACK = All agricultural land-use classes (codes 21 through 24 inclusive)
WHITE = All other land uses

FIGURE 14. DISPLAY OF THE 1970 LAND-USE DATA PLANE EMPHASIZING ALL REMAINING AGRICULTURAL LANDS (in black). These lands are prime candidates for preservation as greenbelts, open space, and parks. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

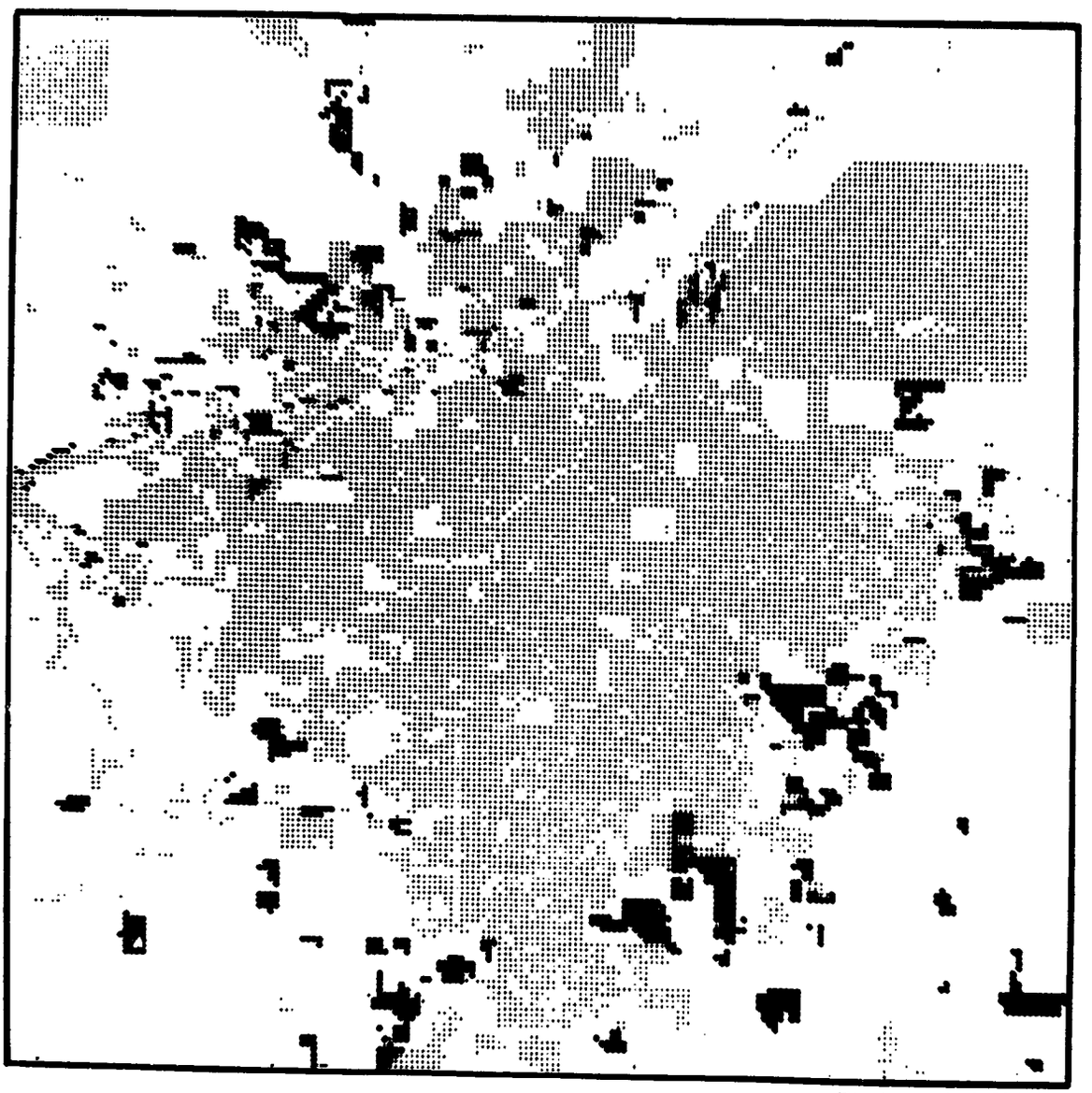
ORIGINAL PAGE IS
OF POOR QUALITY



Display scale 1:250,000

BLACK = Single- and multiple-unit residential dwellings (code 11)
GRAY = Open and other urban land uses (code 19) and cropland and pasture agricultural land (code 21)
WHITE = All other land uses

FIGURE 15. DISPLAY OF THE 1970 LAND-USE DATA PLANE EMPHASIZING SINGLE- AND MULTIPLE-UNIT RESIDENTIAL AREAS (in black). The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.



Display scale 1:250,000

BLACK = Strip and clustered development (code 17)
GRAY = Urban and built-up areas (codes 11, 12, 13, 14, 15, and 16)
WHITE = All other land uses

FIGURE 16. DISPLAY OF THE 1970 LAND-USE DATA PLANE EMPHASIZING THE LOCATION OF STRIP AND CLUSTERED DEVELOPMENTS (in black) RELATIVE TO OTHER URBAN AND BUILT-UP AREAS (in gray). The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

The 1963 land-use data reduction efforts were structured to minimize coding and keypunching duplication of the completed 1970 land-use data plane. This was accomplished by comparatively interpreting individual, nonstereo 1963 aerial photos to the 1970 orthophotomaps for land-use changes. Only the areas of change were coded and keypunched into a 1963 land-use change plane, on which the areas of change were punched as solid geometric blocks and unchanged areas were left blank. The resultant change data plane was logically merged with the 1970 land-use data plane to generate a complete 1' 3 land-use data plane (figures 17 and 18).

The USGS land-use map of Denver (figure 19) was also manually sampled from a 1:100,000-scale land-use classification map (figure 20) encoded in the USGS Circular 671 system (Reference 11). This USGS map was compiled from 1:121,000-scale, high-altitude NASA U-2 color infrared aerial photos of 1972-1973. A 90-percent minimum accuracy value for this photointerpreted land use was claimed for this product although no verification procedures were described. A 1:100,000-scale mylar copy of this map was obtained from the USGS Public Inquiries Office in Denver. It was run through an ozalid blueprint machine in contact with a matching dot sampling grid on mylar to produce a dot-overprinted paper map for the keypunching operation. Both map and reduced transparent grid were carefully registered before printing. Six major first-order categories were defined, with 21 second- and third-order land uses defi. in the Denver study area.

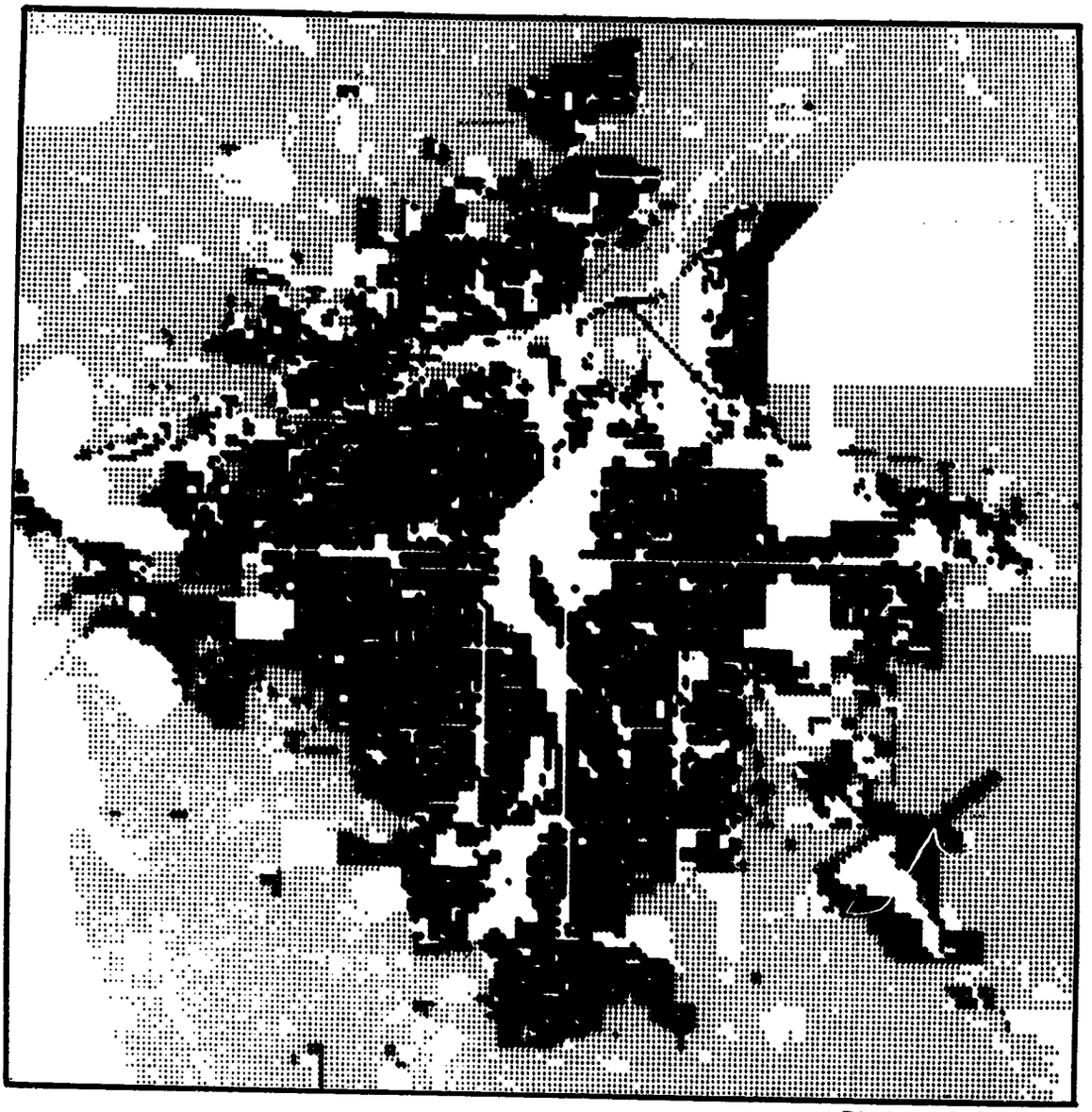
CONSTRUCTION OF PHYSIOGRAPHIC SUBMODEL

Topographic elevation data was manually sampled from sixteen 1:24,000-scale USGS 7.5-minute quadrangle sheets (figure 5) with the data overlay form (figure 7). The elevation of each cell was estimated to the nearest 3 meters (10 feet) and noted next to the appropriate sample dot. The 64 sheets of elevation values were punched and assembled to provide a completely filled topographic elevation data plane (figure 21).

This basic elevation data plane was input to separate computer programs to generate additional "derived" topographic slope, topographic aspect, and image insolation (solar radiation) data planes. The TOPOMAP program (References 16 and 17) computed both slope and aspect from the uniform grid of elevation points. This was accomplished by fitting a regression surface to three-by-three arrays of cells of the elevation data plane and assigning the slope and aspect to the center cell of each three-by-three array (Reference 20). The repetition of this process cell by cell over the entire elevation data plane yielded the slope (figures 22 and 23) and aspect (figure 24) data planes.

Three-dimensional perspective graphics provide familiar oblique viewing angles and convey the spatial concepts and relationships represented by the topographic data planes (figure 25). Additional computer graphics permit the preparation of vertical contouring (figure 26) and perspective contouring (figure 27) displays of the topographic elevation data. Physiographic relationships of landscape components to topography may be visualized by perspective graymapping of a second data plane; for example, all 1970 urban and built-up land (figure 13) on the oblique representation of the topographic data plane (figure 28).

Potential incoming solar radiation (insolation) was computed by program INSOL2 (References 21 and 22) for the Landsat-1 image to be subsequently introduced. This cell-by-cell computation was completed as a function of time and date (solar altitude and azimuth), location (latitude and longitude), and topographic elevation for ancillary data cells with a computed slope and aspect. This computation yielded a measure of both direct and indirect potential insolation for each landscape cell. The calculated flat-surface insolation was ratioed by an actual flat-surface measurement from the Pawnee National Grasslands just northeast of

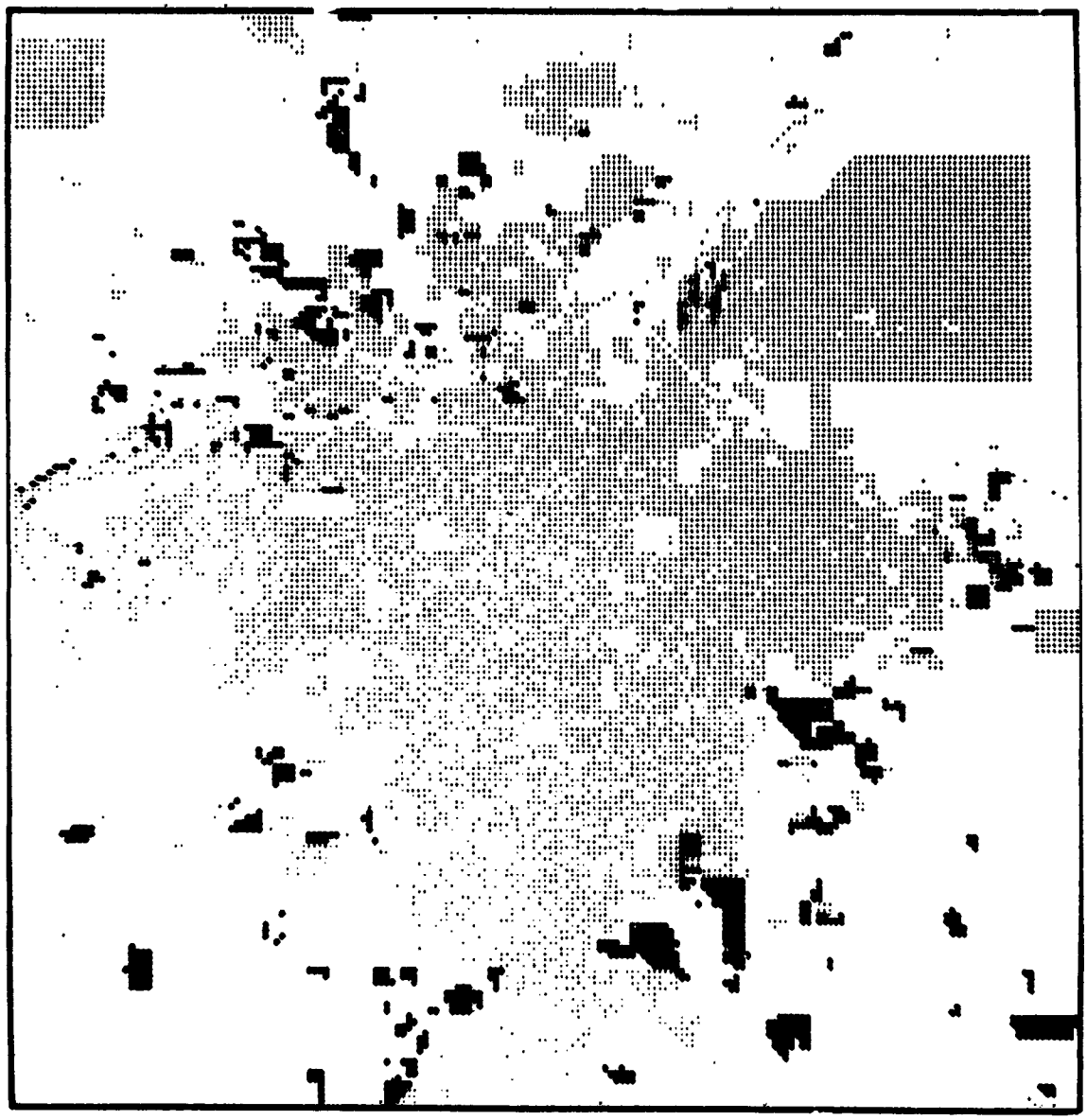


Display scale 1:250,000

BLACK = Single- and multiple-unit residential dwellings (code 11)
GRAY = Open and other urban land (code 19) and cropland and pasture agricultural land (code 21)
WHITE = All other land uses

FIGURE 17. DISPLAY OF THE 1963 LAND-USE DATA PLANE EMPHASIZING SINGLE- AND MULTIPLE-UNIT RESIDENTIAL AREAS (in black). The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

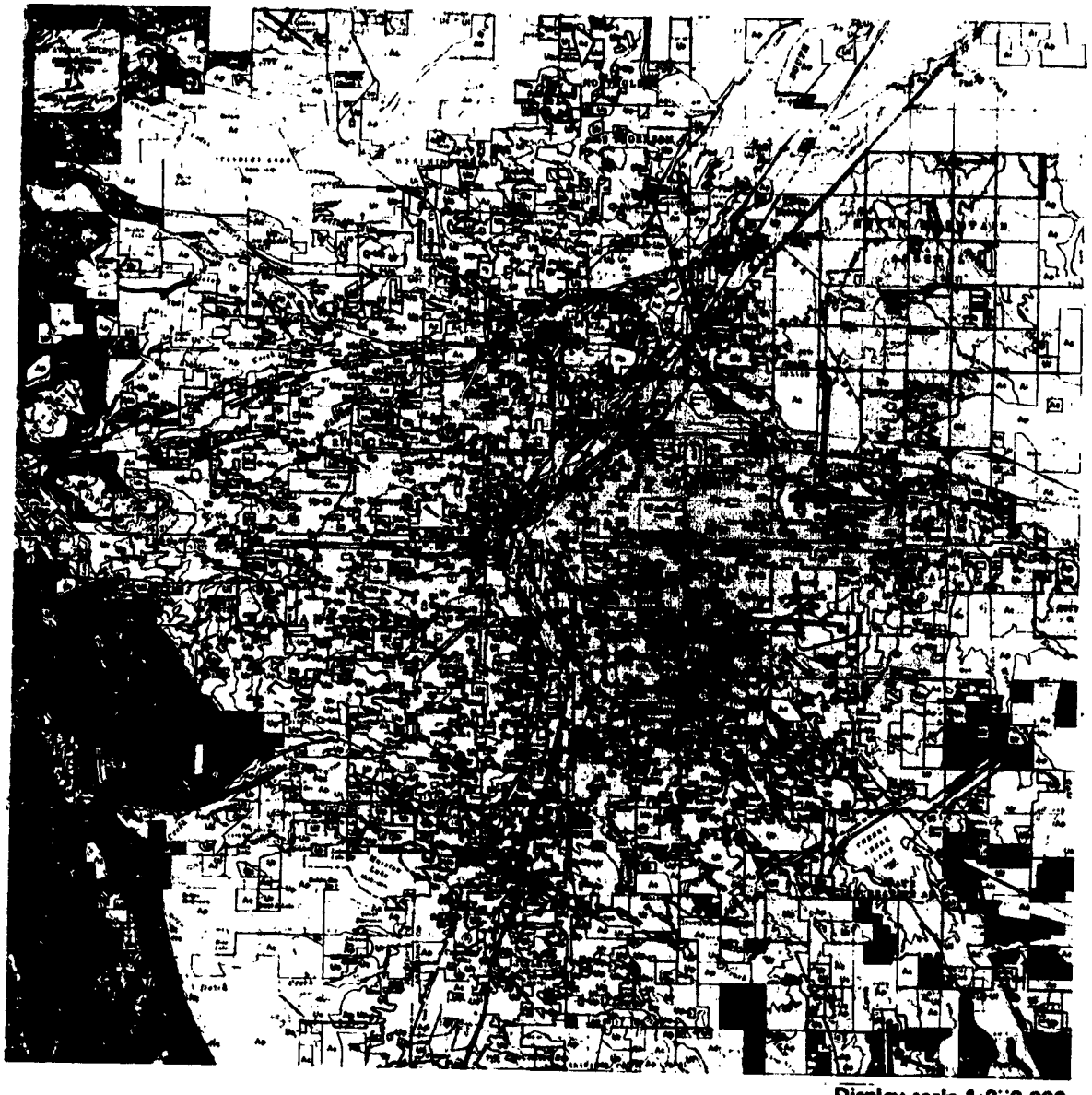
ORIGINAL PAGE IS
OF POOR QUALITY



Display scale 1:250,000

BLACK = Strip and clustered development (code 17)
GRAY = Urban and built-up areas (codes 11, 12, 13, 14, 15, and 16)
WHITE = All other land uses

FIGURE 18. DISPLAY OF THE 1963 LAND-USE DATA PLANE EMPHASIZING THE LOCATION OF STRIP AND CLUSTERED DEVELOPMENTS (in black) RELATIVE TO OTHER URBAN AND BUILT-UP AREAS (in gray). The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

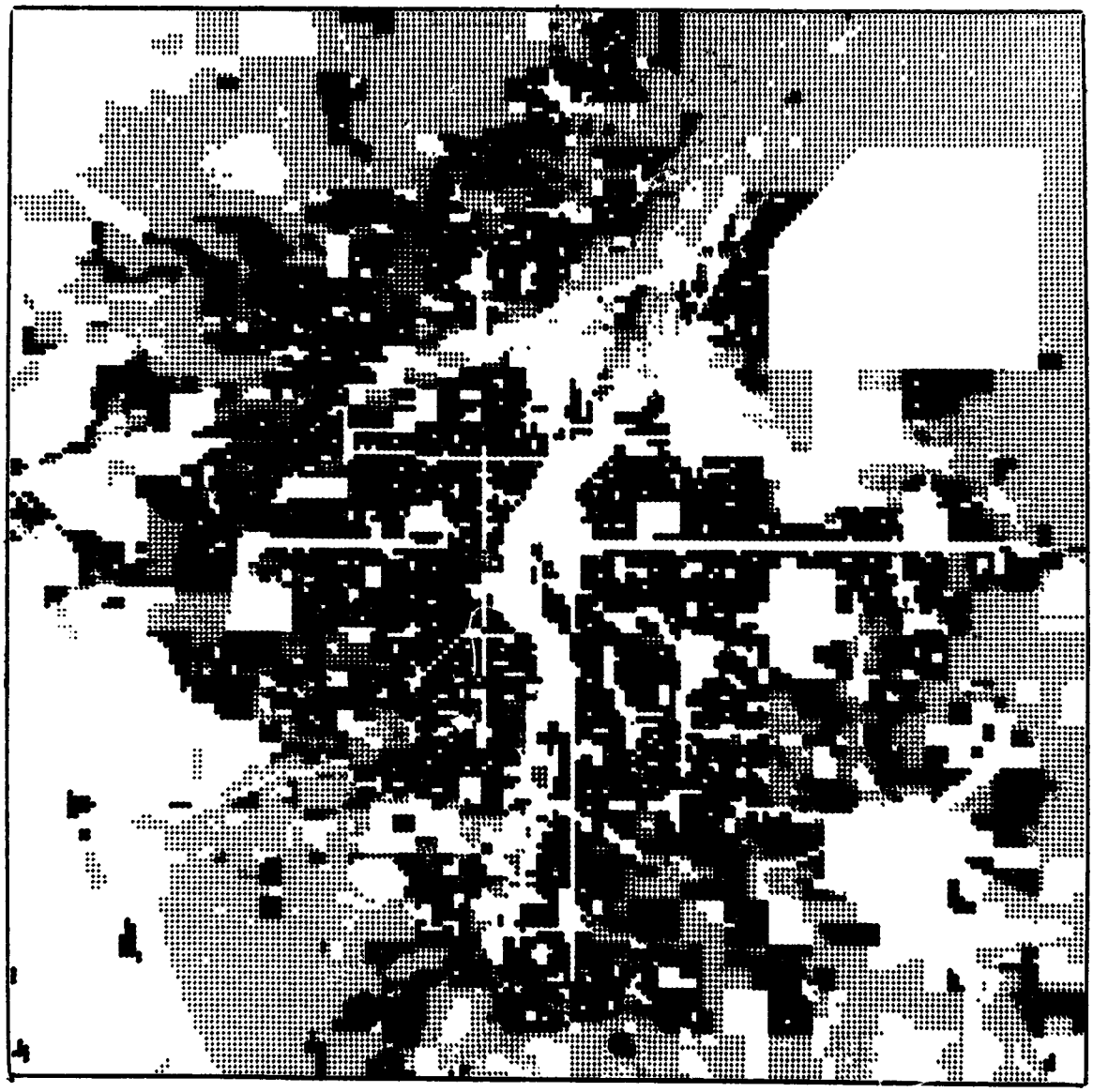


Display scale 1:?50,000

FIGURE 19. DISPLAY OF THE ORIGINAL 1972-1973 USGS 1:100,000-SCALE LAND-USE SOURCE MAP. This map provides a direct comparison with the cellular representation of this source map as a data plane in figure 20. This map was compiled from high-altitude, NASA U-2 color infrared aerial photos taken at a scale of 1:121,000. The USGS Circular 671 classification scheme was employed. A 90-percent minimum accuracy for manual photointerpretation was claimed (Reference 11).

ORIGINAL PAGE IS
OF POOR QUALITY

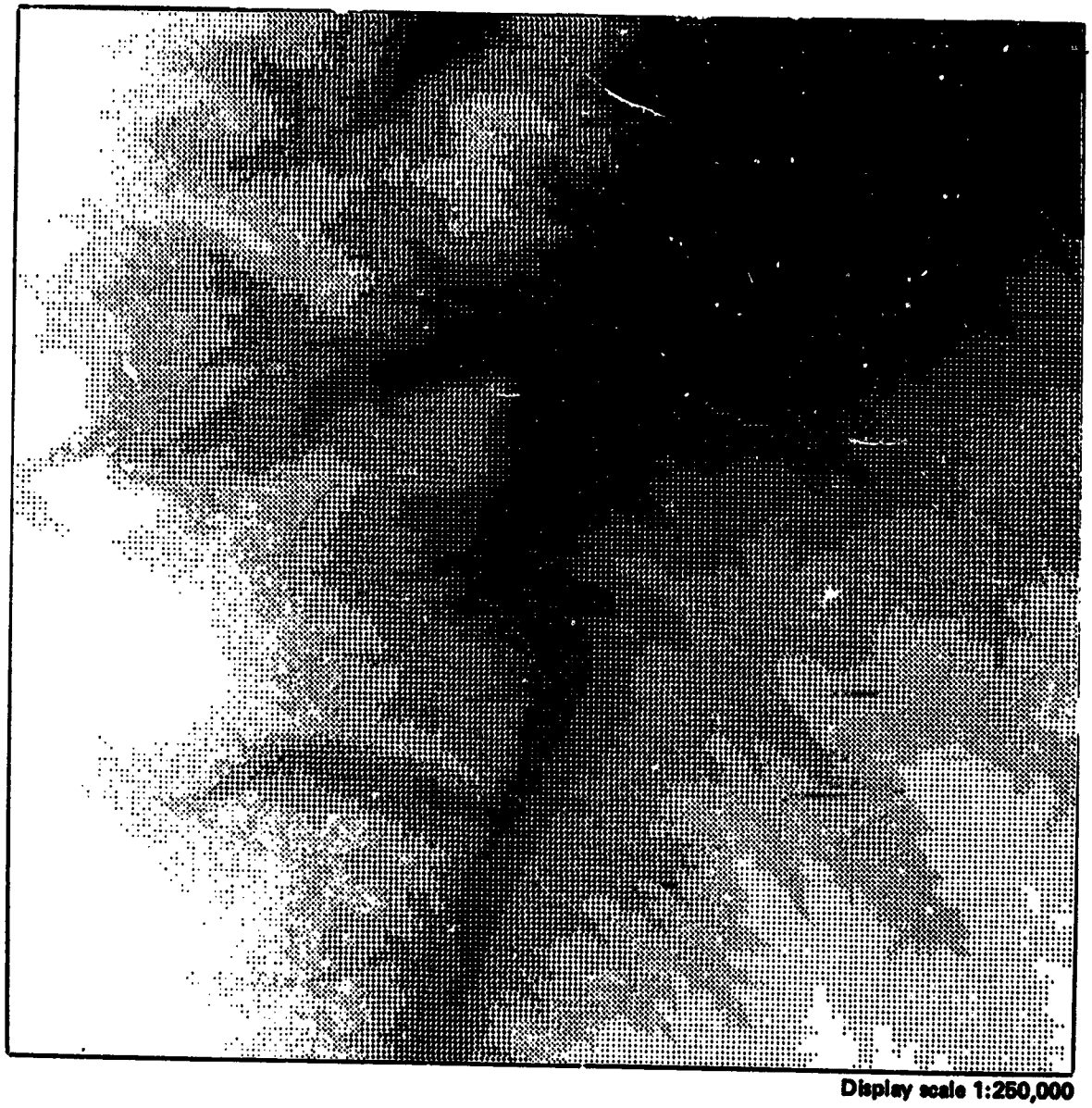
ORIGINAL PAGE IS
OF POOR QUALITY



Display scale 1:250,000

BLACK = Single- and multiple-unit residential dwellings (code 11)
GRAY = Open and other urban land uses (code 19), nonirrigated cropland (code 211), irrigated cropland (code 212), pasture (code 213), grass (code 31), and chaparral (code 33)
WHITE = All other land uses

FIGURE 20. DISPLAY OF THE 1972-1973 USGS LAND-USE DATA PLANE EMPHASIZING SINGLE- AND MULTIPLE-UNIT RESIDENTIAL AREAS (in black). The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area shown in the original map in figure 19.



Display scale 1:250,000

BLACK = Lowest elevations (5,110 feet and below)
GRAY = Intermediate elevations (5,110 to 5,780 feet)
WHITE = Highest elevations (5,780 feet and above)

FIGURE 21. DISPLAY OF THE TOPOGRAPHIC ELEVATION DATA PLANE EMPHASIZING THE LOWEST AREAS (in black). Elevation data was manually sampled from sixteen 1:24,000-scale USGS quadrangle sheets for 36,864 individual 10-acre cells with a 10-foot vertical resolution. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

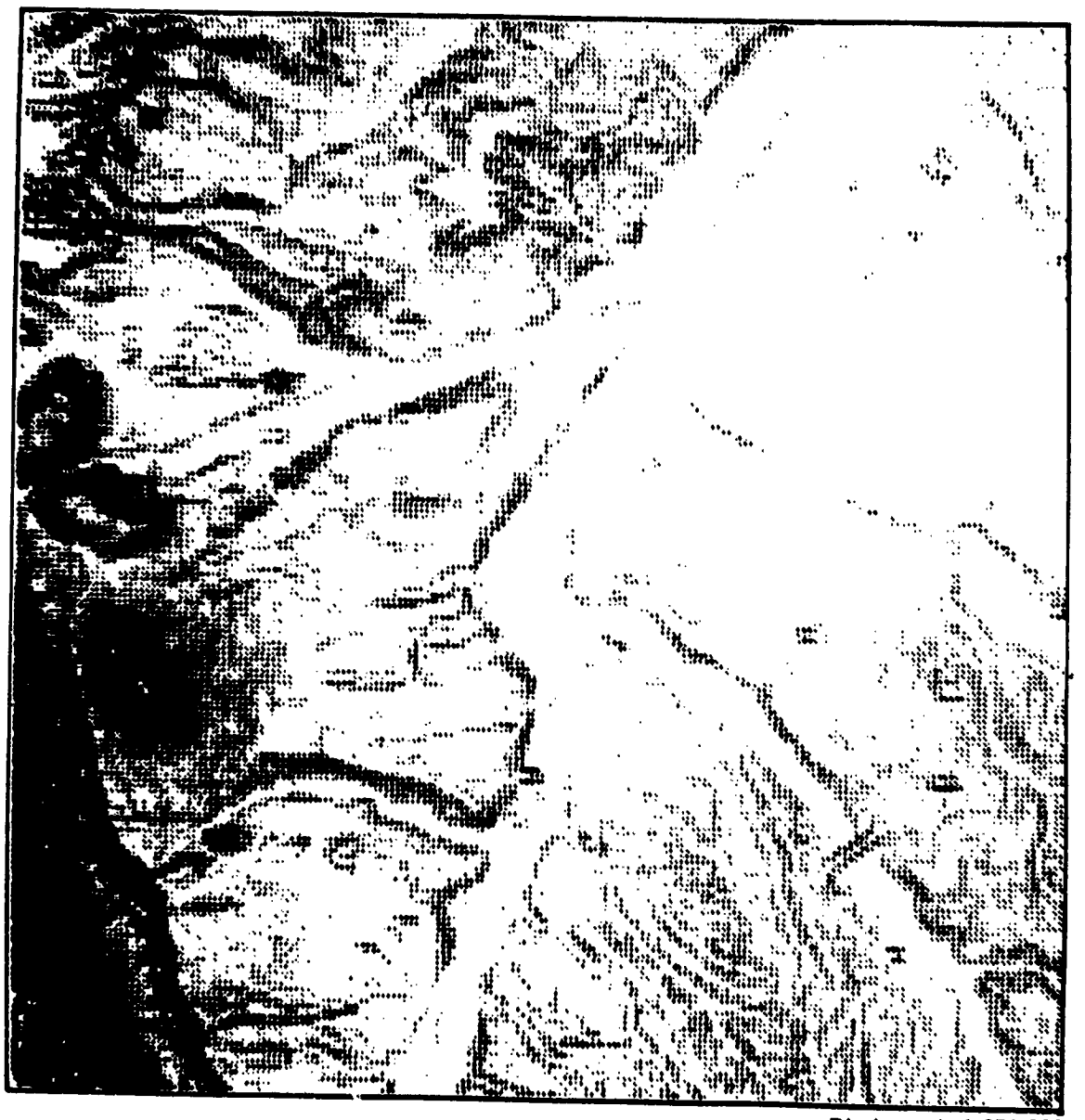
ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

Display scale 1:250,000

BLACK = Flat areas (0.4 percent and less)
GRAY = Intermediate slope areas (0.4 to 5.8 percent)
WHITE = Higher slope areas (5.8 percent and greater)

FIGURE 22. DISPLAY OF THE TOPOGRAPHIC SLOPE DATA PLANE EMPHASIZING SHALLOW SLOPES (in black). Hand-coded elevations displayed in figure 21 were used to compute topographic slope for each 10-acre cell. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.



Display scale 1:250,000

BLACK = Higher slope areas (5.8 percent and greater)
GRAY = Intermediate slope areas (0.4 to 5.8 percent)
WHITE = Flat areas (0.4 percent and less)

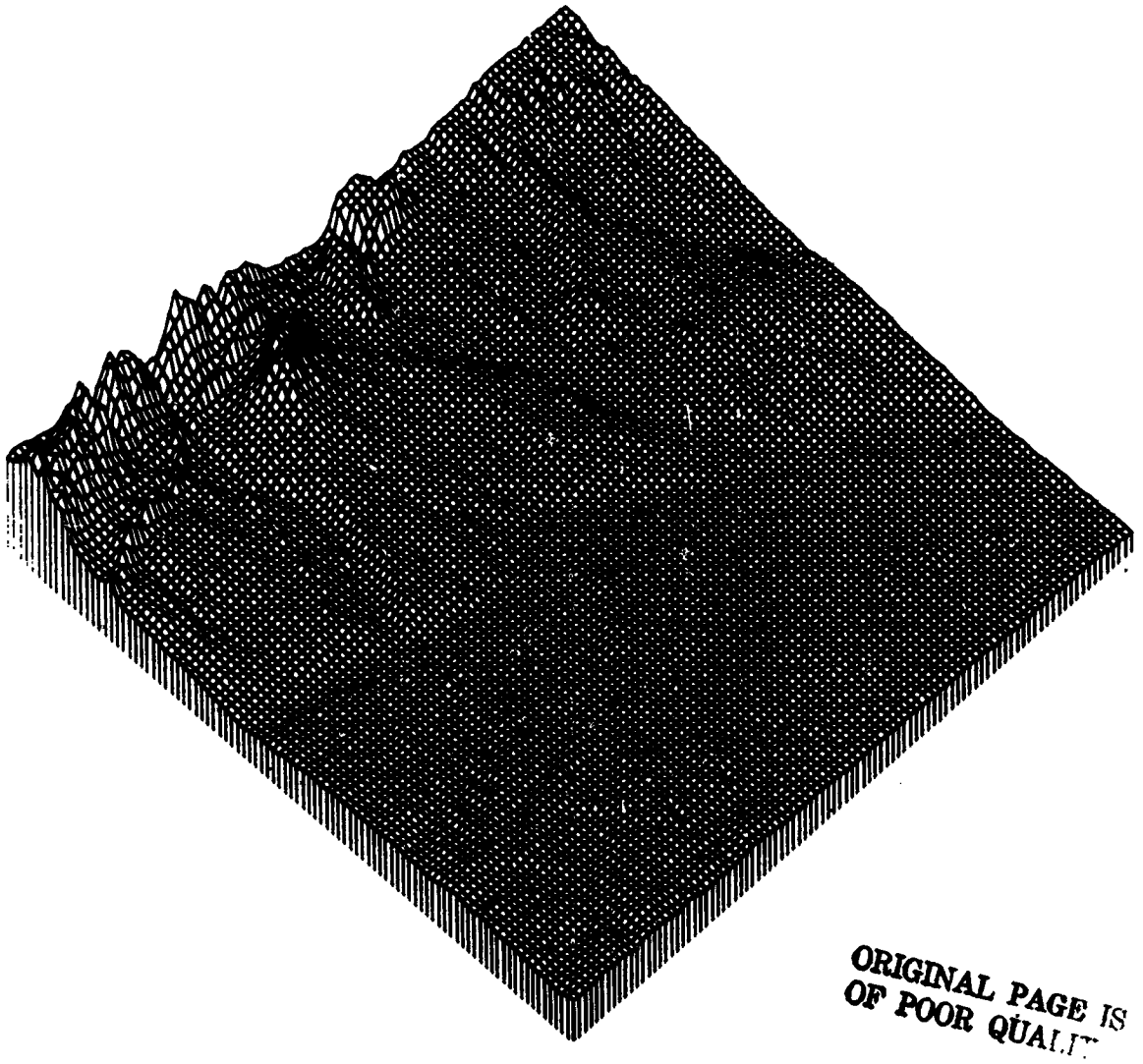
FIGURE 23. DISPLAY OF THE TOPOGRAPHIC SLOPE DATA PLANE EMPHASIZING STEEP SLOPES (in black). The graphic display values of figure 22 were simply reversed to emphasize different slope components. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

ORIGINAL PAGE IS
OF POOR QUALITY

Display scale 1:250,000

BLACK = Northwest-facing areas
GRAY = Intermediate aspects
WHITE = Southeast-facing areas

FIGURE 24. DISPLAY OF THE TOPOGRAPHIC ASPECT DATA PLANE EMPHASIZING NORTHWEST-FACING AREAS (in black). Hand-coded elevations displayed in figure 21 were used to compute topographic aspect for each 10-acre cell. The reader should interpret the display as illuminated from the lower right-hand corner (i.e., from the southeast) so that the terrain shadows fall away from the observer. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.



ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 25. THREE-DIMENSIONAL PERSPECTIVE DISPLAY OF THE TOPOGRAPHIC ELEVATION DATA PLANE LOOKING FROM THE SOUTHEAST. The data input consists of the elevation data plane. The user defines the line of sight by specifying the viewing point and the point looked at. In the resulting perspective plot, the hidden lines are removed. The 24- by 24-mile Denver Metropolitan Area is mapped by alternate rows and columns of 10-acre square cells.

Fort Collins, Colorado, at the time of the August 15, 1973, Landsat-1 overflight. This introduced an estimate of the atmospheric attenuation in the area at the time of the Landsat overflight. This attenuation factor was applied to derive final ground insolation estimates for each 4-ha (10-acre) ground data cell (figure 29). These collateral insolation data will be evaluated to determine if they can be used during machine land-use classification efforts in topographically shaded areas where the lack of uniform illumination altered the spectroirradiance measured from otherwise identical land-cover types and/or surface materials.

Generalized surficial geologic data was manually sampled from a USGS map prepared in cooperation with the Denver Board of Water Commissioners and the Colorado Water Conservation Board (Reference 23). This 1:62,500-scale map was sampled with a reduced version of the original dot grid. Twelve categorical geologic classes were originally defined on the map. Subsequently, a thirteenth class was added, consisting of the prominent water bodies shown on the map (figure 30).

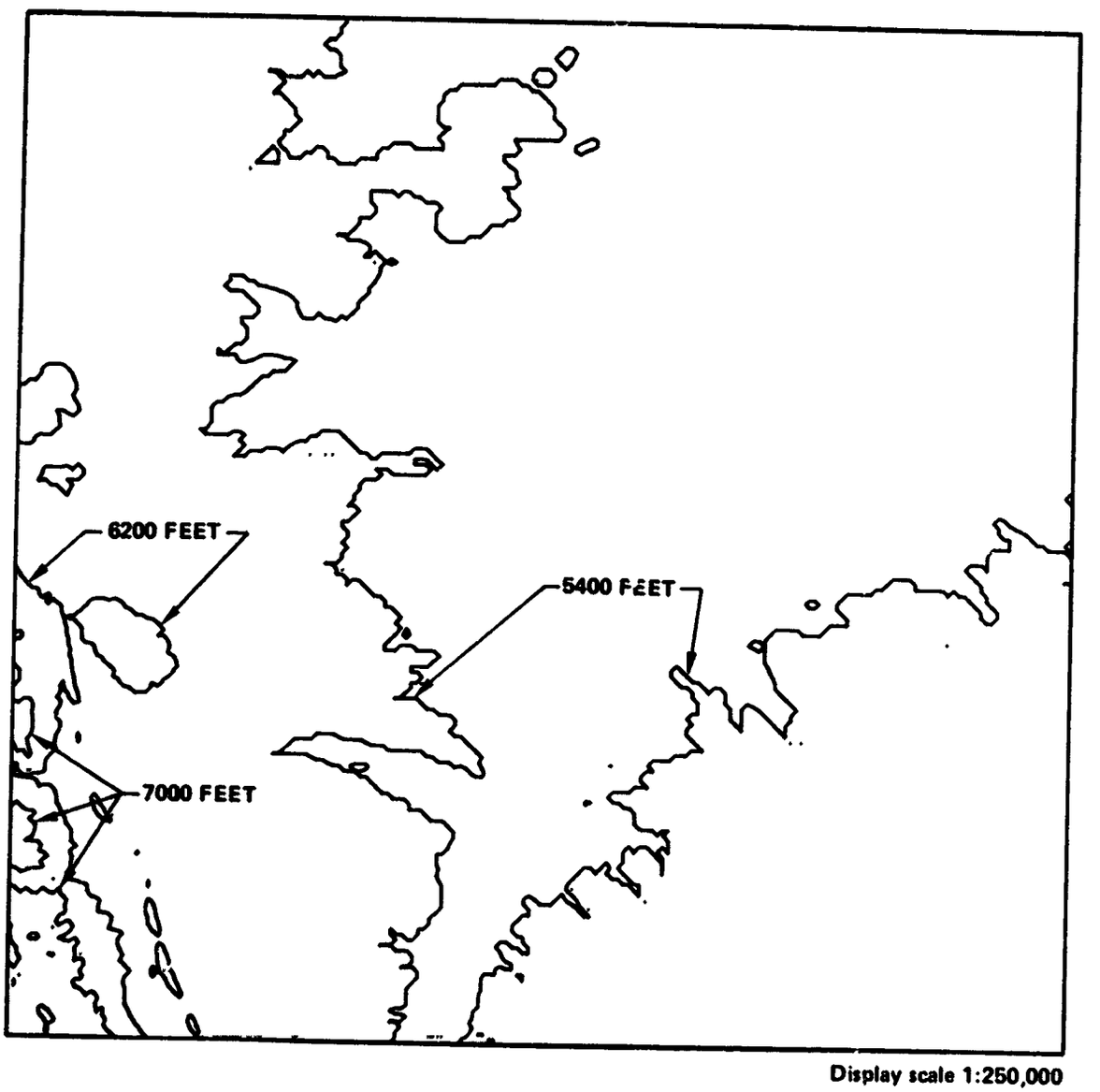
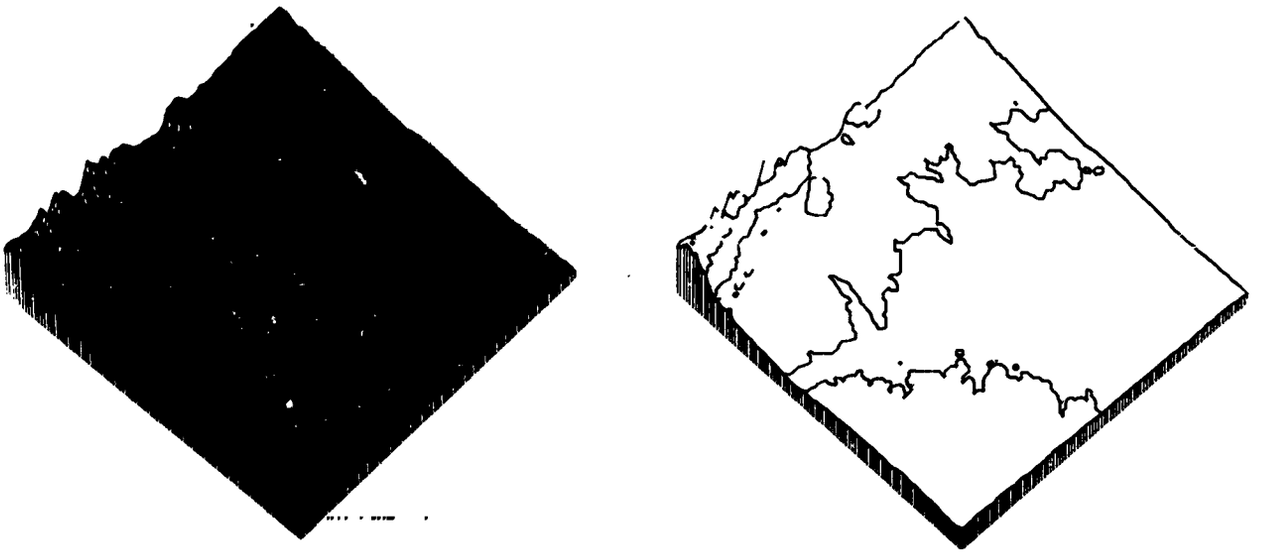


FIGURE 26. CONTOUR MAP OF THE TOPOGRAPHIC ELEVATION DATA PLANE. The 5400-, 6200-, and 7000-foot elevation contour lines are shown. These 800-foot elevation contours of figure 21 are viewed in the conventional vertical map format. The 24- by 24-mile Denver Metropolitan Area is mapped by alternate rows and columns of 10-acre square cells.



(a) Terrain contour display with surface grid

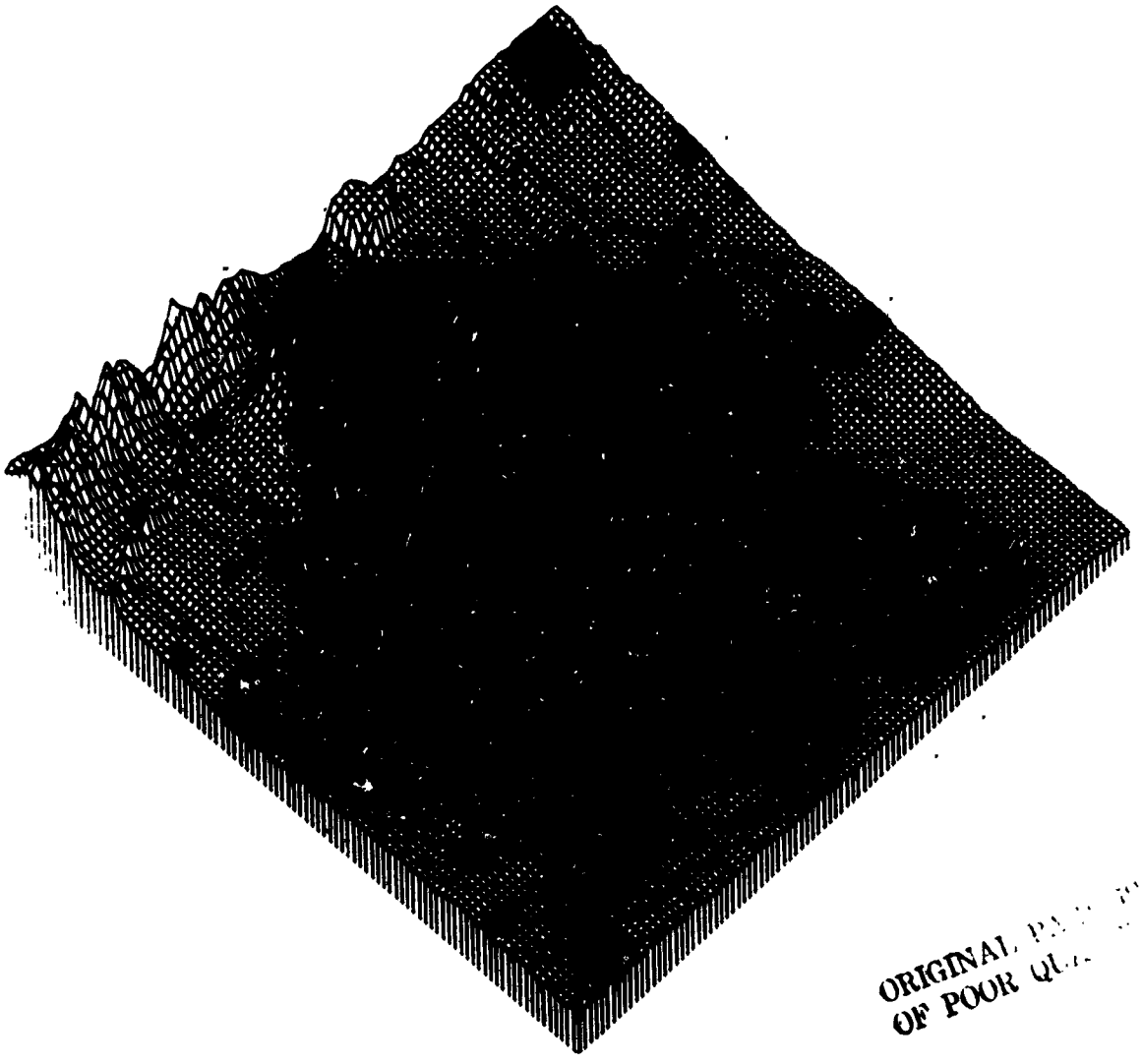
(b) Terrain contour display without surface grid

FIGURE 27. THREE-DIMENSIONAL PERSPECTIVE DISPLAY OF THE TOPOGRAPHIC ELEVATION DATA PLANE. The 5400-, 6200-, and 7000-foot elevation contour lines are shown. (a) Topographic display with surface grid lines and contours. (b) Topographic display with contours but without surface grid lines. Incomplete contour lines are due to hidden line removal. The 24- by 24-mile Denver Metropolitan Area is mapped by alternate rows and columns of 10-acre square cells.

CONSTRUCTION OF TRANSPORTATION SUBMODEL

A 1971 road-classification map for the Denver Metropolitan Area was obtained from the State Highway Department that differentiated freeways, expressways, principals, majors, minors, collectors, and isolated rural roads. This 1:45,000-scale map provided the point data plane for the computation of minimum distance planes and was sampled with a reduced scale, point-sampling grid. The linear road classes were assigned to the nearest-neighbor grid point to provide a continuous representation in cellular form. When multiple road classes intersected or existed in the nearest-neighbor vicinity of a dot point, the decision was made to choose the highest capacity road class.

These point and linear transportation features were transformed into area planes and overlaid on the landscape model. The locations of freeway interchanges are an example of a typical point feature that has an important impact on specific changes in the land use surrounding them. The impact of the interchange is a function of the distance away from the interchange and was best handled in the landscape model in terms of a data plane representing minimum distances. The freeway interchanges were initially tabulated into a point-type data plane that recorded their location in the nearest 4-ha (10-acre) cell (figure 31). This initial



ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 28. THREE-DIMENSIONAL PERSPECTIVE DISPLAY OF THE TOPOGRAPHIC ELEVATION DATA PLANE EMPHASIZING ALL URBAN BUILT-UP LANDS (in black). The display system generated this plot by first defining the terrain as a statistical surface on which the urban land uses (code 11 through 19, inclusively) were then overlaid as graymapped cells. Like 3-D contouring (figure 27), perspective graymapping is a useful graphic tool for visualizing the numeric and/or spatial relationships of landscape variables. The 24- by 24-mile Denver Metropolitan Area is mapped by alternate rows and columns of 10-acre square cells.

data plane was subjected to numeric computation so that the minimum distance in an east-west and/or north-south sense was computed for each cell in the data plane to the nearest cell occupied by a freeway interchange (Reference 24).

The minimum distance computed in this fashion was recorded at the position of the selected cell, and the computation was completed for each cell in the data plane. This transformed the original point plane into a useful minimum-distance area plane, which was overlaid onto the landscape model (figure 32). Similarly, the initial data planes representing linear road features (figures 33a, 33c, and 34a) were computationally converted to area planes for overlaying on the landscape model (figures 33b, 33d, and 34b). An initial urban built-up area classification was established in recognition of the area nature of the fully developed street network and the established land-use patterns found in extensively developed urban areas (figure 34c). Again, an urban built-up area minimum-distance plane was numerically derived from this source data plane (figure 34d).

Display scale 1:250,000-

BLACK = Shadowed northwest-facing areas (113.2 centilangleys and less)
GRAY = Intermediate sunlit areas (113.2 to 121.4 centilangleys)
WHITE = Fully sunlit southeast-facing areas (121.4 centilangleys and greater)

FIGURE 29. DISPLAY OF THE INSOLATION DATA PLANE EMPHASIZING AREAS OF LOWEST INCOMING SOLAR ENERGY (in black). Near-instantaneous solar radiation for the overflight time of the August 15, 1973, Landsat-1 image was generated from the computed slope and aspect for each 10-acre landscape cell. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

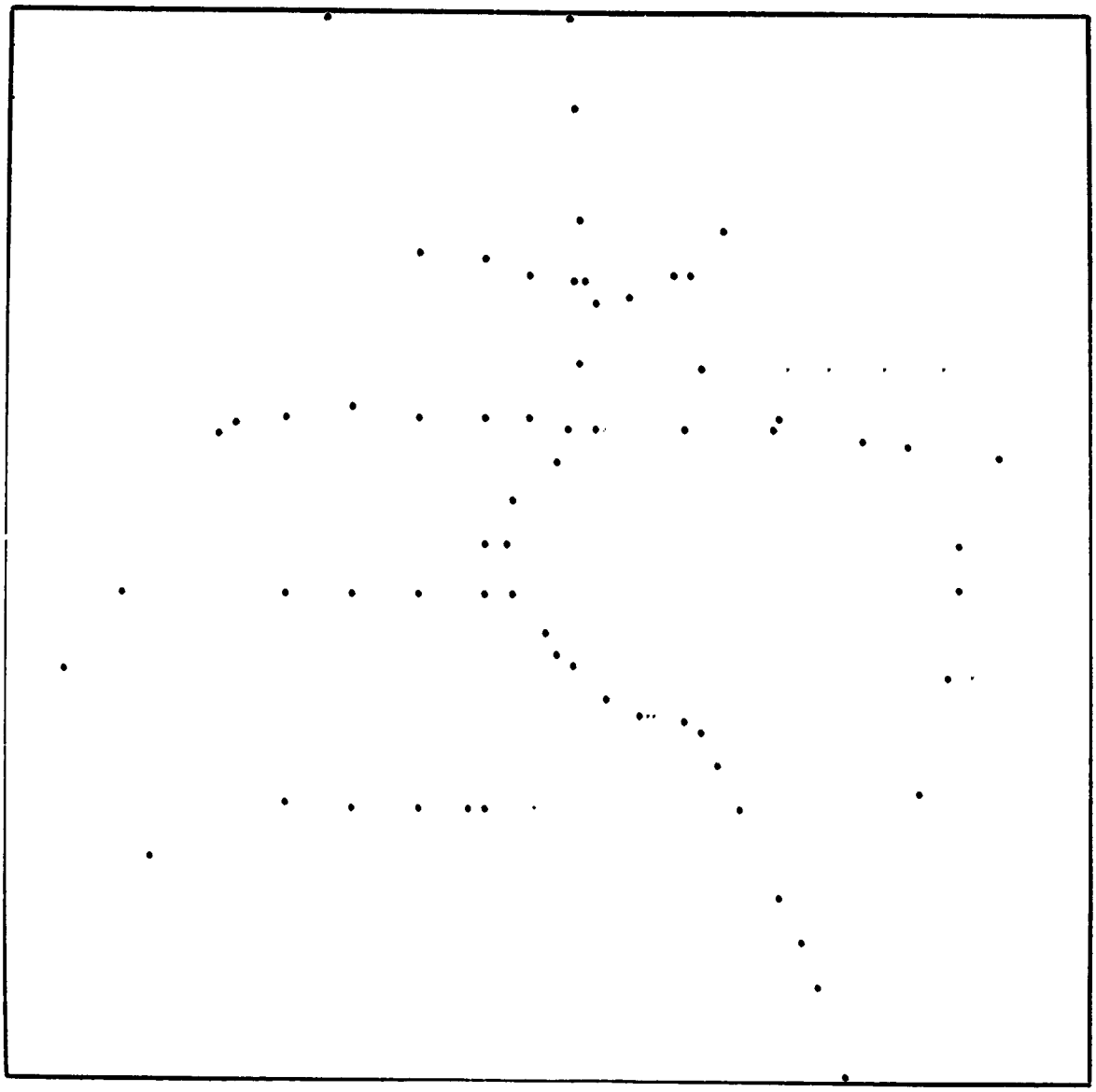
ORIGINAL PAGE IS
OF POOR QUALITY

Display scale 1:250,000

BLACK = Wind-deposited silt, sand, and cobbles on foothill slopes (Holocene period)

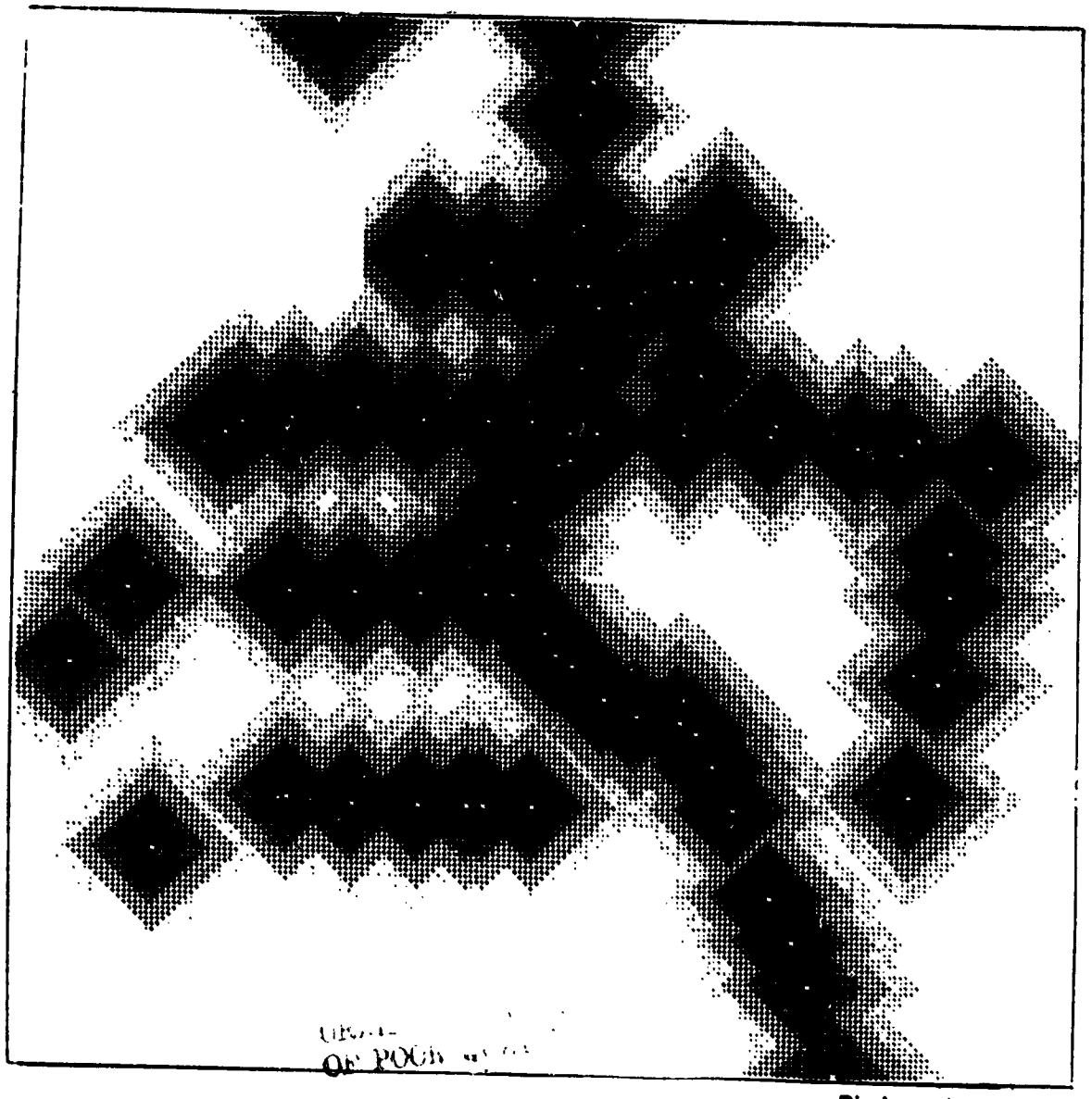
DARKEST GRAY = Sand, gravel, silt, and clay (Pleistocene and Holocene period Post-Piney Creek, Piney Creek, pre-Piney Creek, Broadway, and Louviers Alluviums)

FIGURE 30. DISPLAY OF THE SURFICIAL GEOLOGIC DATA PLANE EMPHASIZING AEOLIAN DEPOSITS (in black). Thirteen generalized surficial geologic units were manually sampled from a 1:62,500-scale USGS map. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.



Display scale 1:250,000

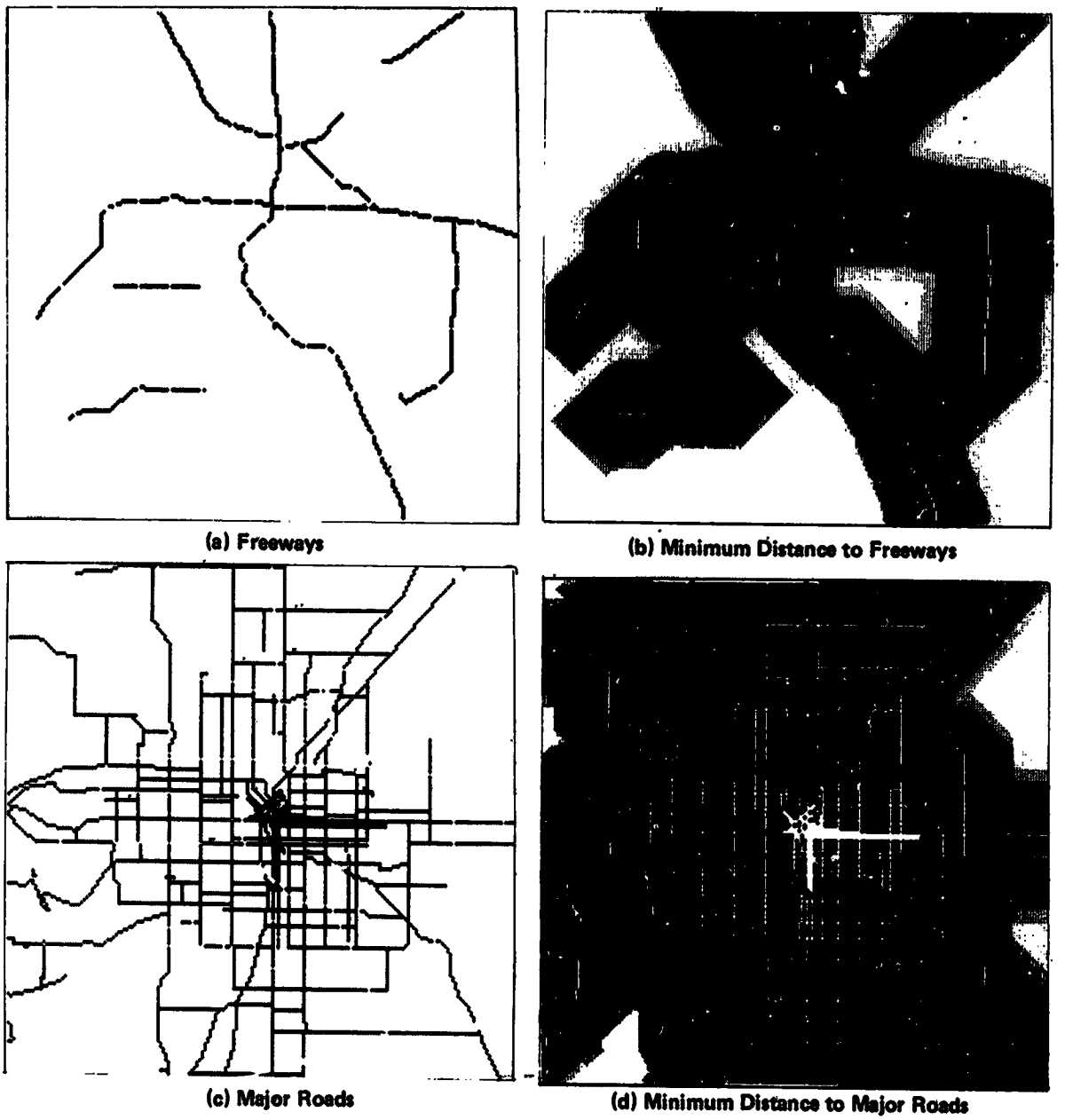
FIGURE 31. DISPLAY OF THE 1971 ROAD TRANSPORTATION DATA PLANE SHOWING FREEWAY INTERCHANGES. Eight other road access categories were also manually sampled from a 1:45,000-scale state metropolitan road classification map. Some planners have speculated that freeway development is a powerful inducement to land-use conversions. This hypothesis was statistically tested through the spatial landscape modeling process. The microfilm display shows each 10-acre cell containing a freeway interchange in the 24-by 24-mile Denver Metropolitan Area.



Display scale 1:250,000

BLACK = Minimum distances to freeway interchanges (3,300 feet or less)
GRAYS = Intermediate to furthest distances to freeway interchanges (five 3,300-foot intervals from 3,300 to 16,500 feet or more)
WHITE = Freeway interchanges

FIGURE 32. DISPLAY OF THE MINIMUM DISTANCE TO FREEWAY INTERCHANGE DATA PLANE EMPHASIZING MINIMUM DISTANCE (in black). Freeway interchange access data (figure 31) was used to compute minimum distance in a north-south and/or east-west traverse from each cell to the nearest freeway interchange. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.



(a) Freeways

(b) Minimum Distance to Freeways

(c) Major Roads

(d) Minimum Distance to Major Roads

Display scale 1:500,000

BLACK = Minimum distance to road type of interest (3,300 feet or less)
GRAYS = Intermediate to furthest distances (3,300 feet or greater)
WHITE = Road type of interest

FIGURE 33. DISPLAY OF THE 1971 FREEWAY AND MAJOR-ROAD TRANSPORTATION DATA PLANES AND THEIR ASSOCIATED TRANSFORMED PLANES THAT EMPHASIZE MINIMUM DISTANCES (in black). Freeways (a) and major roads (c) were manually sampled as linear features from a 1:45,000-scale road-classification map. These data were used to compute minimum distances [(b) and (d)] from each cell in the plane in a north-south and east-west traverse to the nearest cell containing the given road type. The microfilm display of each 10-acre cell maps the 24- by 24-mile Denver Metropolitan Area.

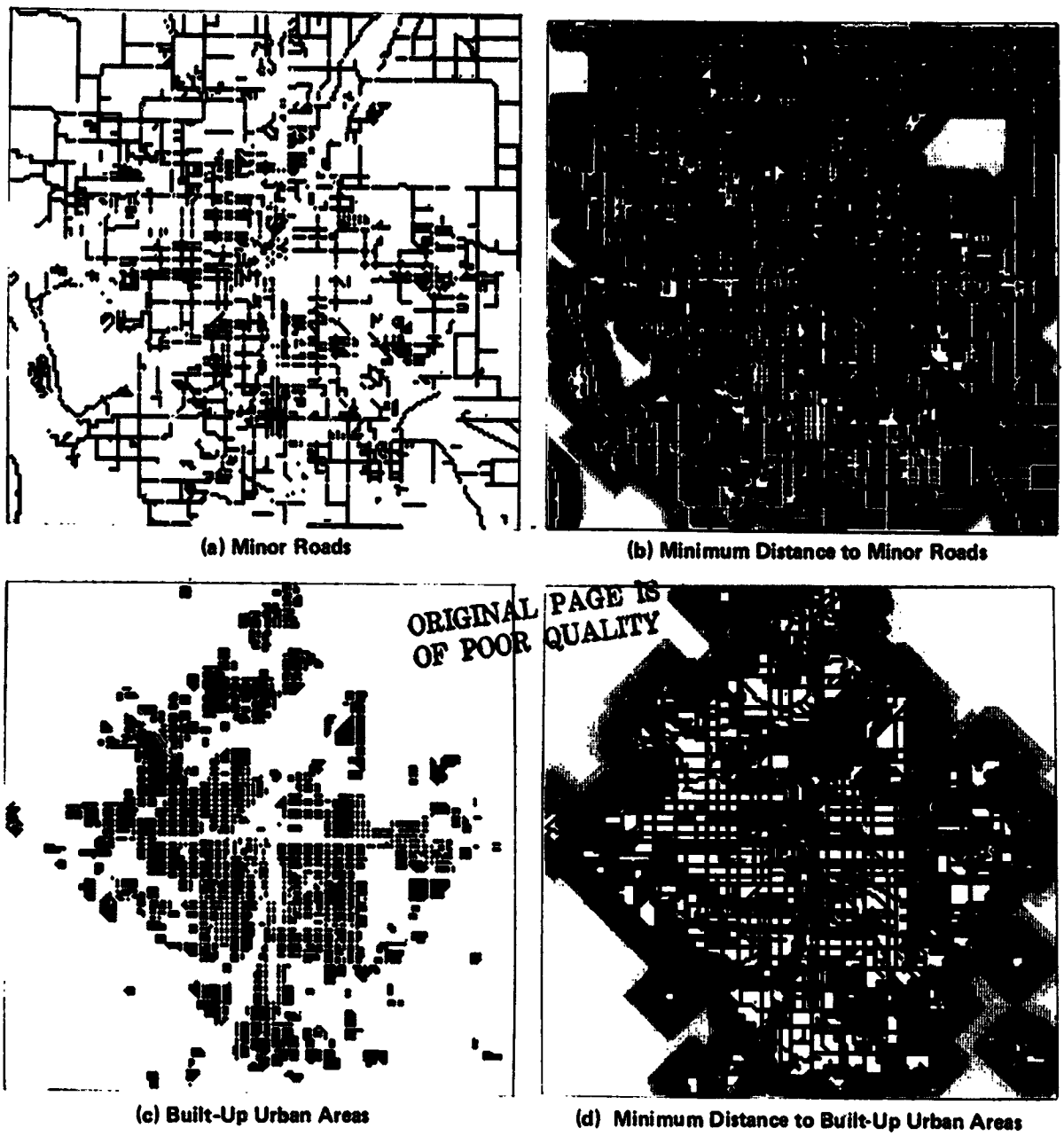


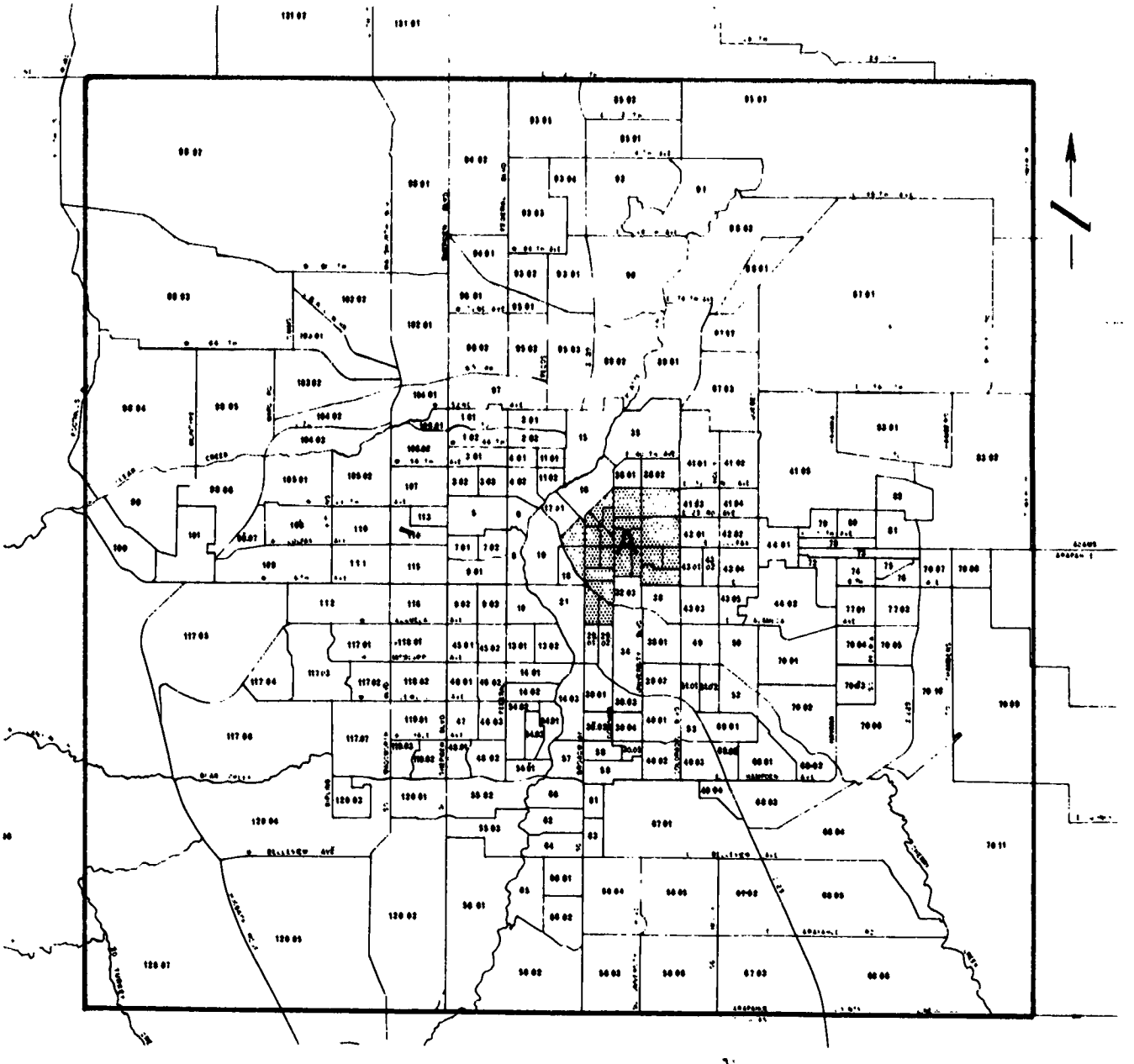
FIGURE 34. DISPLAY OF THE 1971 MINOR-ROAD AND BUILT-UP URBAN AREA TRANSPORTATION DATA PLANES AND THEIR ASSOCIATED TRANSFORMED PLANES THAT EMPHASIZE MINIMUM DISTANCE (in black). These road-access data [(a) and (c)] were used to compute minimum distance [(b) and (d)] from each cell in the plane in a north-south and east-west traverse to the nearest cell containing the given road class. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

CONSTRUCTION OF SOCIOECONOMIC SUBMODEL

Socioeconomic data for the United States is tabulated by census at 10-year intervals. These data are reported in a tabular form referenced to maps of the census tracts or smaller tabulation units called enumeration districts. These census-tract reference maps (figure 35) were sampled by a dot pattern so that each 4-ha (10-acre) cell was assigned to a specific census tract on the resulting data plane. This procedure permitted the tabular statistics to be input in a list format and projected into area-type, socioeconomic data planes of population, housing, income, car ownership, and census-tract acreage as follows:

Socioeconomic Variable	Intermediate Processing
Total population	None
Total families	None
Total year-round housing units	None
Total vacant housing units	None
Total occupied housing units	None
1969 mean family income	None
Median housing-unit value	None
Median housing-unit rent	None
Total one-car families	None
Total two-car families	None
Total three-/three-plus car families	None
Total census tract acreage	None
Population density per acre	Normalization by census tract acreage
Average number of families per acre	Normalization by census tract acreage
Average number of year-round housing units per acre	Normalization by census tract acreage
Average number of vacant housing units per acre	Normalization by census tract acreage
Average number of cars per family	Average value computed

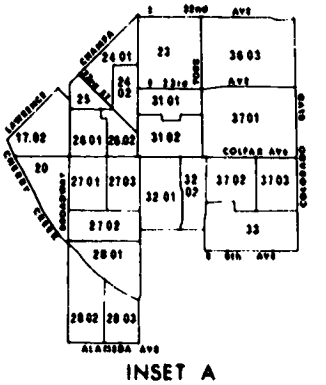
Unfortunately, the size of the census tracts is far coarser than the 4-ha (10-acre) resolution of the landscape model. Consequently, these socioeconomic data planes are not as highly resolved as desired in a spatial sense, but they provide a reasonable approximation of the spatial variation of these census variables.



SCALE IN MILES

4 2 0 2 4

ORIGINAL PAGE IS
OF POOR QUALITY



INSET A

FIGURE 35. CENSUS TRACT OVERLAY OF THE DENVER METROPOLITAN AREA. The boundaries of each of these census tracts as defined for the Denver Metropolitan statistical area for 1970 were overlaid onto the 10-acre cellular network. Socioeconomic census data were introduced into the landscape model in this fashion.

Seventeen socioeconomic data planes were generated for the 241 census tracts in the Denver study area. Total number of families (figure 36), total families, total year-round housing units, total vacant housing units, total occupied housing units, 1969 mean family income (figure 37a), median housing unit value (figure 37b), and median housing unit rent (figure 37c) were directly projected into the landscape model. Census-tract acreage was computed from manual dot counts and used to normalize the population, family, and housing totals. Thus, a preliminary data plane of total population (figure 38a) for each census tract was divided by that tract's acreage, and the resulting population density per acre (figure 38b) was projected into all of that tract's cells, yielding a meaningful population density data plane. Similarly, the initial data planes representing the average number of families, year-round housing units, and vacant housing units per acre were computed and projected into density-type, socioeconomic census data planes and were overlaid onto the landscape model.

One- (figure 39a), two- (figure 39b), three-, and three-plus-car family totals (figure 39c) per census tract were also transcribed into initial data planes, summed, and divided by total number of families data plane on a cell-by-cell basis to derive a data plane representing the average number of cars per family (figure 39d).

SUMMARY

Construction of a landscape model for the Denver Metropolitan Area involved 4-ha (10-acre) dot-grid sampling of selected land-use, physiographic, socioeconomic, and transportation map variables to generate a total of 34 meaningful data planes. These map data planes were spatially registered to each other so that they could be "stacked" or overlaid for subsequent computer analysis, modeling, and spatial/temporal prediction (figure 3).

The selection of collateral map data for inclusion in the landscape model was fundamental to the success of this effort. The data employed were extensive, dispersed, and complex, and were obtained from many widely varying sources. The variables ultimately chosen were selected for their suitability and availability.

These map variables were substantial in number and exhibited considerable diversity in the broad realms of cultural and physical landscape components. However, they were simply only a sample of all possible variables pertinent to landscape modeling. Such aspects as utilities service, domestic water supply, land ownership and parcel size, land values, soils, and zoning regulations were among the many variables considered, but not included in this study (Reference 1). If, by some happenstance, one or more of these data were available for inclusion, other data equally important to the study of land-use practices in the Denver Metropolitan Area would still be unavailable.

Because there were no immediate limitations to enlarging the set of variables, the underlying understanding of land-use practices constituted a limitation in the determination of which mapping variables should be included in the data set to fully inform the landscape analysis effort. Even if this considerable problem were resolved, the most significant remaining limitation is the simple lack of data or "lack" of data in any usable form. This is particularly true in developing countries (Reference 25), but was no less the case in regional studies such as this, in which extensive spatial data are required and generally believed to be readily available.

The simplest way of inputting each data plane into the landscape model was adopted. This ensured that the majority of the resources and effort would be channeled into analytical endeavors. The thrust of earlier studies involved more complex map data entry and computer data structures, resulting in curtailed analysis efforts.

Display scale 1:250,000

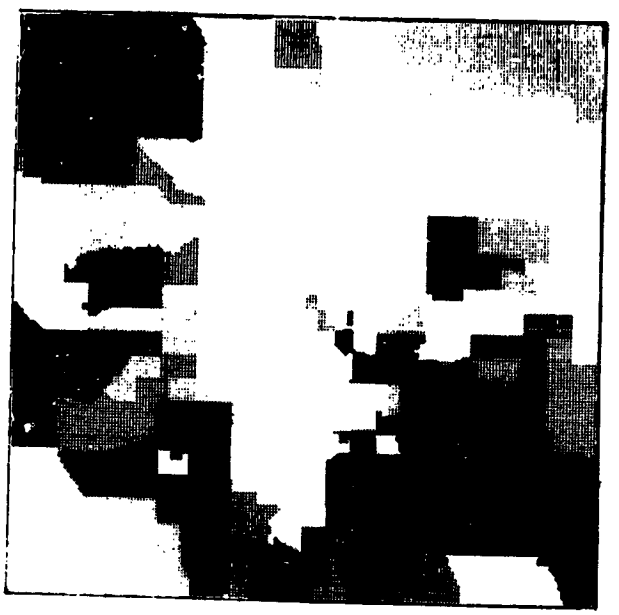
BLACK = Maximum total families per census tract (2772 to 3080 families)
GRAYS = Minimum to intermediate total families per census tract (7 to 2772 families)

FIGURE 36. DISPLAY OF THE 1970 SOCIOECONOMIC DATA EMPHASIZING THE HIGHEST TOTAL NUMBER OF FAMILIES PER CENSUS TRACT (in black). Eleven other socioeconomic parameters were also manually sampled from 1:84,500-scale census maps. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

This Page Intentionally Left Blank

ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME



Display scale 1:500,000

(a) 1969 Mean Family Income Per Census Tract

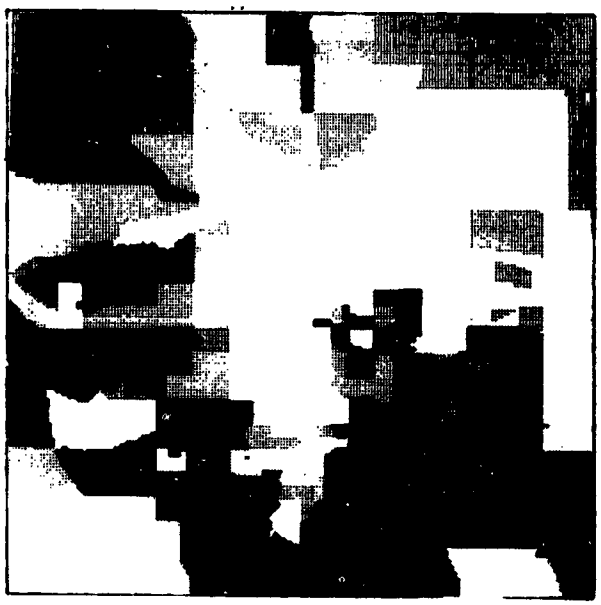
BLACK = Maximum mean family income per census tract ($31,890 to $35,433 per family)

GRAYS = Minimum to intermediate mean family income per census tract ($4,399 to $31,890 per family)

These three socioeconomic summary statistics were manuall
play of each 10-acre square cell maps the 24- by 24-mile Den

2 FOLDOUT FRAME

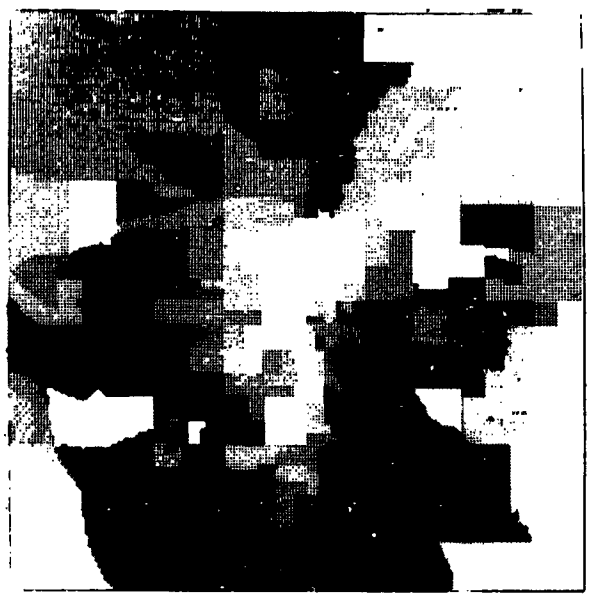
ORIGINAL PAGE IS
OF POOR QUALITY



Display scale 1:500,000

(b) Median Housing-Unit Value Per Census Tract

BLACK = Maximum median housing-unit value per census tract ($45,000 to $50,000 per housing unit)
GRAYS = Minimum to intermediate median housing-unit value ($7,700 to $45,000 per housing unit)



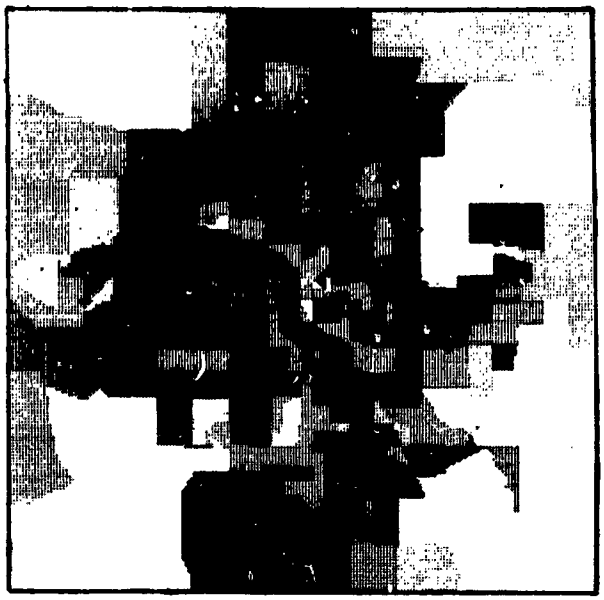
Display scale 1:500,000

(c) Median Housing-Unit Rent Per Census Tract

BLACK = Maximum median housing-unit rent per census tract ($270 to $300 per month)
GRAYS = Minimum to intermediate median housing-unit rent per census tract ($48 to $270 per month)

/ sampled from 1:84,500-scale census maps. The microfilm dis-
- er Metropolitan Area.

FIGURE 37. DISPLAY OF THE 1970 SOCIOECONOMIC DATA EMPHASIZING HIGHEST 1969 MEAN FAMILY INCOME, MEDIAN HOUSING-UNIT VALUE, AND MEDIAN HOUSING-UNIT RENT (in black).



(a) Total Population Per Census-Tract

BLACK = Maximum total population per census tract (11,240 to 12,484 persons)
GRAYS = Minimum to intermediate total populations per census tract (39 to 11,240 persons)

Display scale 1:500,000

(b) Population Density Per Census-Tract Acre

BLACK = Maximum population density per census-tract acre (52.9 to 58.8 persons per acre)
GRAYS = Minimum to intermediate population densities per census-tract acre (0.004 to 52.9 persons per acre)

Display scale 1:500,000

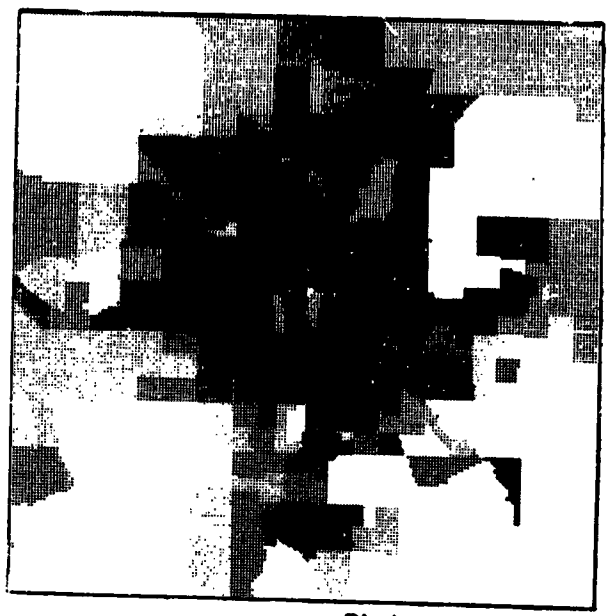
FIGURE 36. DISPLAY OF THE 1970 SOCIOECONOMIC DATA EMPHASIZING HIGHEST POPULATION TOTALS AND AVERAGES (in black). Total population (a) was manually sampled from 1:84,500-scale census maps. These totals were divided by census-tract acreage to compute population density (b). The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

PRECEDING PAGE BLANK-NOT FILMED

Tliis Page Intentionally Left Blank

ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME

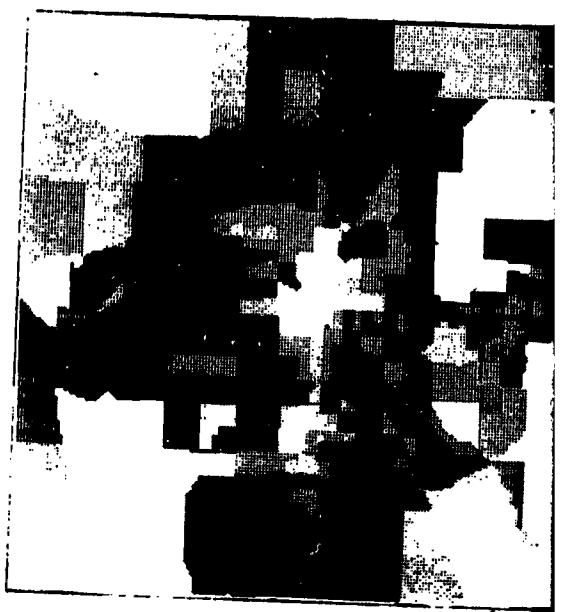


Display scale 1:500,000

(a) Total One-Car Families Per Census Tract

BLACK = Maximum total one-car families per census tract (1735 to 1928 families)

GRAYS = Minimum to intermediate total one-car families per census tract (6 to 1735 families)



Display scale 1:500

(b) Total Two-Car Families Per Census Tract

BLACK = Maximum total two-car families per census tract (1584 to 1760 families)

GRAYS = Minimum to intermediate total two-car families per census tract (11 to 1584 families)

The one-, two-, and three-car ownership totals were manually sampled from 1:84,50(of cars per family was computed from the one-, two-, three- and three-plus-car data 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME

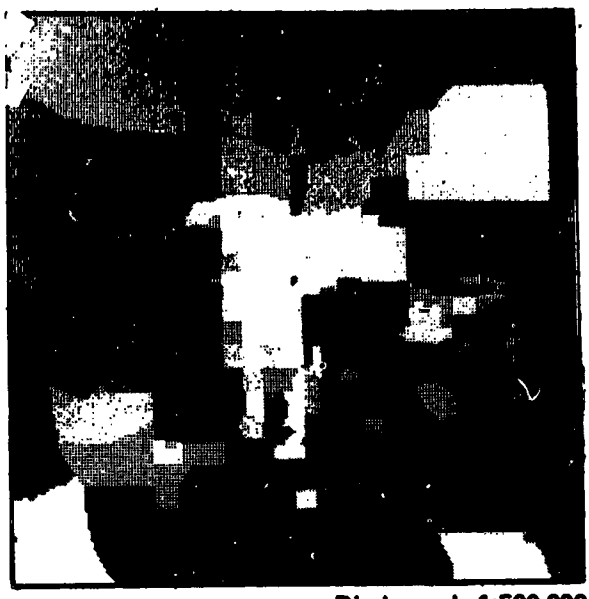
000

Display scale 1:500,00

(c) Total Three and Three-Plus-Car Families Per Census Tract

BLACK = Maximum total three-, and three-plus-car families per census tract (425 to 472 families)

GRAYS = Minimum to intermediate total three-,and three-plus-car families per census tract (5 to 425 families)



Display scale 1:500,000

(d) Average Number of Cars Per Family

BLACK = Maximum average number of cars per family (2.9 to 3.2 cars per family)

GRAYS = Minimum to intermediate average number of cars per family (0.7 to 2.9 cars per family)

l-scale census maps. The average number
planes. The microfilm display of eeeh

FIGURE 39. DISPLAY OF THE 1970 SOCIOECONOMIC DATA EMPHASIZING HIGHEST ONE-, TWO-, THREE-, AND THREE-PLUS-CAR FAMILY OWNERSHIP AND AVERAGE NUMBER OF CARS PER FAMILY (in black).

The collateral data for the landscape model consisted of measurements on 34 map variables for 36,864 observations. These observations were point locations selected by a systematic 4-ha (10-acre) dot-grid sample. An overlay data form was created to combine and simplify the map coding and keypunching functions for 1:24,000-scale maps. Simple computer routines also simplified the editing and creation of completed map overlays or data planes.

The 34 map variables were grouped into four general land-use, physiographic, transportation, and socio-economic submodels. The USGS Circular 671 classification system was used for manual photointerpretation of 1963 and 1970 photographic imagery. A 1972-1973 USGS land-use map was also dot-sampled and overlaid onto the land-use submodel.

Topographic elevation—the basic data input to the topographic submodel—was encoded from 16 USGS quadrangles. Slope, aspect, and insolation data planes were computed with auxillary computer programs. Surficial geologic data were also added to the landscape model from a USGS map.

A 1971 road-classification map with eight defined road classes was the basic input to the transportation submodel. The five original point data planes representing composite minor roads, composite major roads, freeways, freeway interchanges, and urban built-up areas were transformed into minimum-distance area planes by numeric computation to be overlaid in the landscape model.

Finally, the tabular socioeconomic data and maps of the 1970 Census were used to create twelve area-type data planes of population, housing, income, car ownership, and census-tract acreage. Population density per acre, average number of families per acre, average number of year-round/vacant housing units per acre, and average number of cars per family were computed and projected into five density-type data planes.

PRECEDING PAGE BLANK-NOT FILMED

This Page Intentionally Left Blank

CHAPTER 3

SPATIAL LAND-USE PROJECTION

CHAPTER 3

SPATIAL LAND-USE PROJECTION

INTRODUCTION

PRECEDING PAGE

This chapter deals with the structure, testing, and verification of the land-use projection models that were developed and tested. The development of the full spatial projection model proceeded in three major steps. First, an initial multivariate model was developed. The model was then refined by improving its logical structure. Finally, the model was further refined on the basis of the results of spatial-change replications.

The development of the landscape modeling data set formed the nucleus of the initial spatial projection model. Many potential variables were eliminated from consideration for reasons of suitability and availability. The initial projection model therefore consisted of all land-use change classes and landscape variables. This model was used as a benchmark for further refinements and analyses.

Detailed examination of the initial model run showed flaws in its logical structure, especially the projection of changes into infeasible combinations that ignored *a priori* knowledge of prior land use. Therefore, the second phase of spatial-model development focused on improving the logical structure of the model to restrict its projection to those changes that could actually occur, based on their current land use.

Modeling refinements were based on the testing of alternative formulations on the basic data set. The chief criterion used for refining the spatial model was the ability to adequately replicate its own training or calibration data. This was accomplished by using the model to project 1970 land-use changes on the basis of statistics of the known 1963 to 1970 changes. The various spatial projections tested were evaluated by examining the total number of changes that were correctly classified over the total number tested. Each projection was made using the training set sample of known changes on a purely statistical basis, and only afterward was the actual land-use change status of each individual cell revealed in a classification accuracy table.

DETERMINATION OF RECENT LAND-USE DYNAMICS

Projections for future land use in the Denver Metropolitan Area were based on observations of the changes that occurred in the area in the recent past. This required the overlay onto the landscape model of accurate, detailed current and near-past land-use patterns as described in Chapter 2. Remote sensing imagery with accurate interpretation provided the data for the land-use modeling process. It will be subsequently shown how these inputs can be obtained and overlaid in a timely and accurate fashion through computer analysis of Landsat multispectral digital imagery. However, the development and initial testing of the land-use modeling process used carefully prepared, accurate land-use maps interpreted from low-altitude black-and-white airphotos for both a current (1970)* and past (1963) dates. A uniform land-use classification scheme covering 24 land uses was first adopted (table 1). As described earlier, a single photointerpreter analyzed

*When this effort was first undertaken in 1972, 1970 was the most current airphoto data available.

the large collection of airphotos for each date, annotating the location of each of the 24 land uses to a 4-ha (10-acre) resolution. The two interpretations on the airphotos were then transferred to the sixteen 1:24,000-scale map overlays covering the site. The 4-ha (10-acre) dot-grid patterns were imposed on these map overlays, and the 24-class land-use maps were tabulated and overlaid onto the digital landscape model (Reference 1).

Recent changes in land use were computed and displayed from the 1963 and 1970 land-use data planes overlaid in the model. This employed the cell-by-cell comparison of the 24 land uses interpreted for the two different dates, which yielded a third data plane that recorded the changes in land use between the two dates. A direct visual display of this new plane can illustrate its contents by highlighting the areas of gain or loss of each land use (figures 40 and 41). Summation of the cells that changed between the two dates for each of the 24 land uses provided a quick insight into those categories that were rapidly changing during the 7-year interval (table 3).

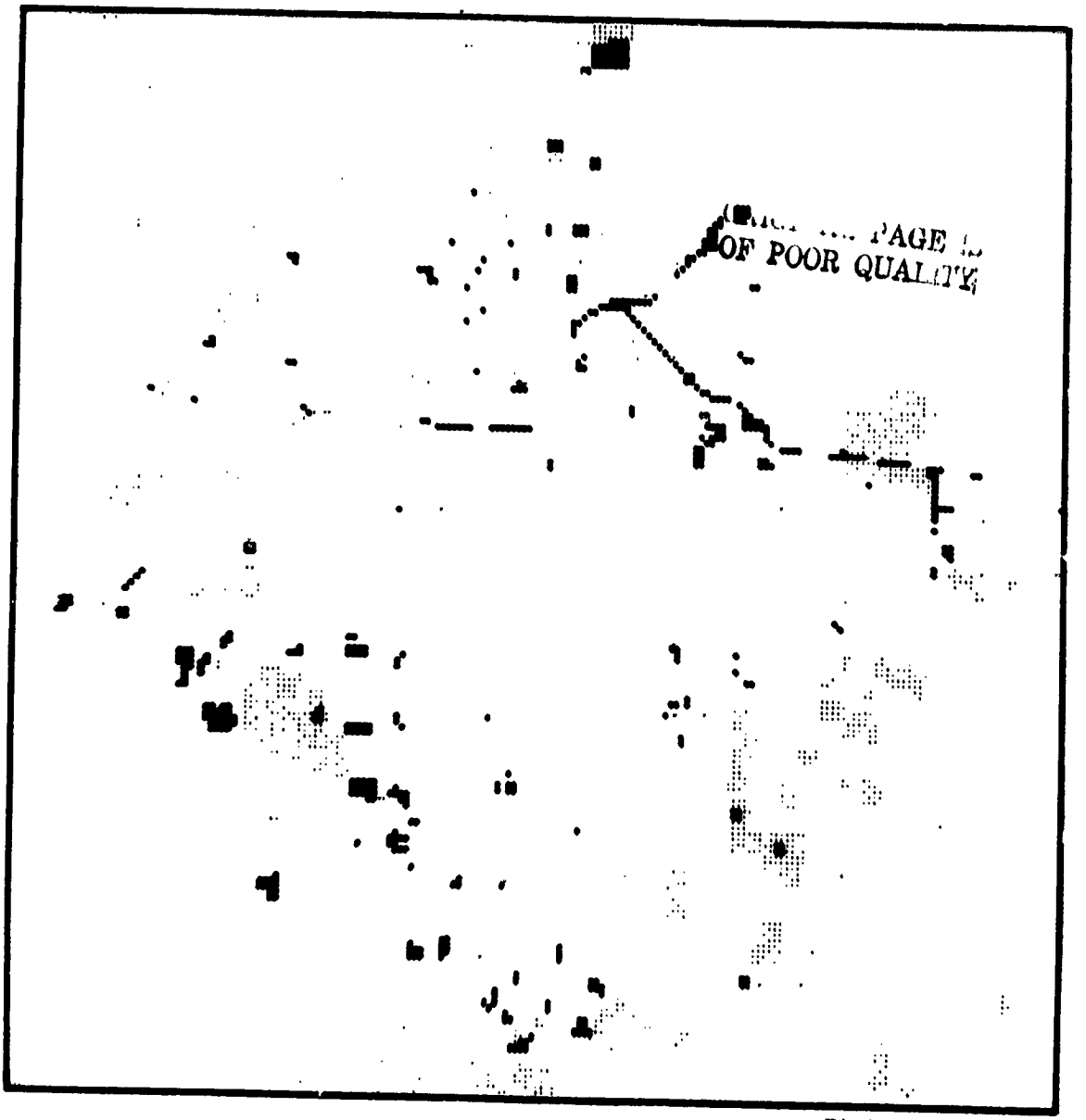
The detailed cell-by-cell comparisons of the land-use type for each of the 36,864 cells on each date also provided a simple matrix that contained the number of 4-ha (10-acre) cells of each of the 24 land uses of the earlier date (1963), which converted to another land use by the second date (1970) (table 4). This matrix tabulation of recent changes in land use provided the basis for computing the tendency for additional change in the near future. These measured tendencies were the essence of the initial land-use trend model when converted into a probability transition matrix or a square two-dimensional array of probabilities of change arranged in rows and columns.

The assumption that future changes in land use can be measured in terms of those that occurred in the recent past permitted a simple projection of the future trends in land use (References 8 and 26). This assumption does not truly represent the evolution of real-world land use, which is constantly subjected to new and often unanticipated stimuli. However, assuming no change in practices from the past, the techniques for projecting future land use must be perfected before the impact of new, unmeasured, and unobserved trends can be incorporated into the process.

MARKOV LAND-USE TREND MODEL

The Denver Metropolitan Area was photointerpreted into a finite number of observable states or land classes that could be numbered 1, 2, 3, . . ., m. A given mapping cell originally classified as *other barren land*, for example, might be considered as state i; state j might denote *open and other urban land*, and so forth, with state m (perhaps *urban residential land use*) representing the last state in the land-use model. In this study, the total number of states (m) was equal to 34--the number of second-order land classification types in the USGS Circular 671 system. Although only 21 classes were found in the Denver area during photointerpretation and 13 land-use classes were absent in a geographical sense (e.g., *tundra*), the full 34 classes were retained for a full conceptual application elsewhere.

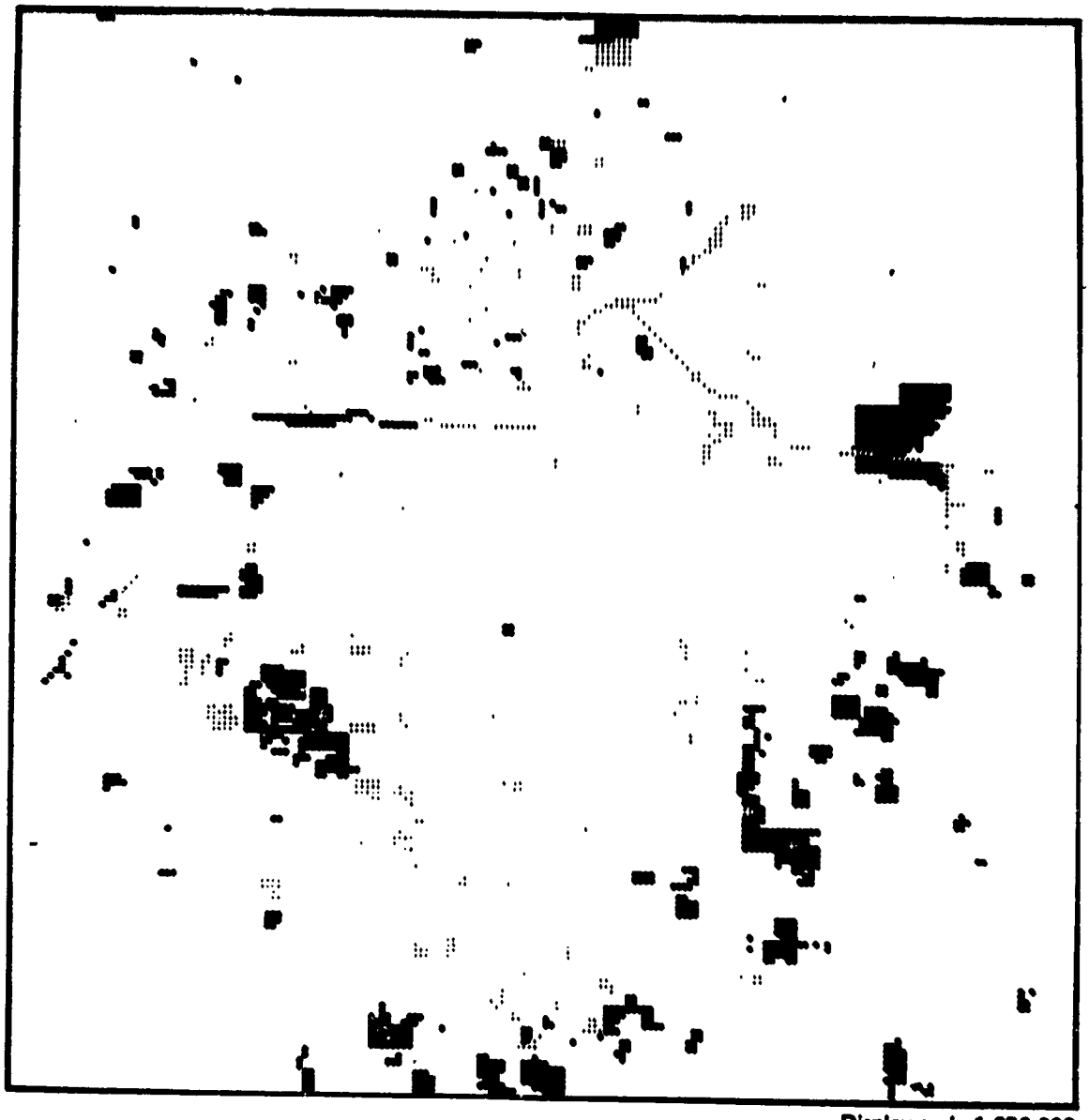
The probability transition matrix contains the average probability that each of the 34 land uses will remain the same or change to some other land use over the time represented by the dates of the two input land-use



Display scale 1:250,000

BLACK = 1963 urban open space (code 19) that was converted to other land uses by 1970.
GRAY = 1963 cropland and pasture agricultural land (code 21) that was converted to other land uses by 1970.
WHITE = All other land uses.

FIGURE 40. DISPLAY OF THE LOSS IN OPEN SPACE BETWEEN 1963 AND 1970 EMPHASIZING THE OPEN SPACE EMBEDDED IN THE URBAN AREA THAT WAS CONVERTED INTO OTHER LAND USES (in black). The amount and characteristics of the acreages converted can be found in tables 3 and 4. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.



Display scale 1:250,000

BLACK = 1963 cropland and pasture agricultural land (code 21) that was converted to other land uses by 1970.
GRAY = 1963 urban open space (code 19) that was converted to other land uses by 1970.
WHITE = All other land uses.

FIGURE 41. DISPLAY OF THE LOSS IN EMBEDDED OPEN SPACE BETWEEN 1963 AND 1970 EMPHASIZING THE AGRICULTURAL LAND THAT WAS CONVERTED INTO OTHER LAND USES (in black). The amount and characteristics of the acreages converted can be found in tables 3 and 4. The microfilm display of each 10-acre square cell maps the 24- by 24-mile Denver Metropolitan Area.

TABLE 3

NET CHANGES IN 1963 DENVER LAND USE RELATIVE TO 1970. Specific net transitions, such as those illustrated in figures 40 and 41, that occurred between various specific land-use classes are given here. Computer comparison of the land-use data planes illustrated in figures 13 through 16 provided these simple comparisons.

Land-Use Type	1963 Acreage	1970 Acreage	1970 Acreage (net gain (+) or loss (-))
Residential	64,210	70,500	+ 6,290
Commercial and services	11,020	11,820	+ 800
Industrial	8,870	10,610	+ 1,740
Extractive	4,630	6,360	+ 1,730
Transportation, communications, and utilities	7,290	8,650	+ 1,360
Institutional	31,250	31,590	+ 340
Strip and clustered development	13,500	16,550	+ 3,050
Mixed urban	40	0	-40
Open and other urban	37,410	35,620	- 1,790
Urban subtotal	178,220	191,700	+13,480
Cropland and pasture	160,090	146,240	-13,850
Orchards, groves, and other horticultural areas	60	60	0
Livestock feeding operations	20	20	0
Other agricultural land	330	70	- 260
Agricultural subtotal	160,500	146,390	-14,110
Deciduous forest land	180	180	0
Streams and waterways	960	1,010	+ 50
Lakes	5,930	6,410	+ 480
Reservoirs	1,580	1,750	+ 170
Other water	50	50	0
Water subtotal	8,520	9,220	+ 700
Vegetated nonforested wetland	1,710	1,710	0
Sand other than beaches	640	520	- 120
Other barren land	18,870	18,920	+ 50
Barren subtotal	19,510	19,440	- 70
Grand Total	368,640	368,640	±16,060

TABLE 4

CHANGES IN DENVER LAND USE, 1963 TO 1970. Entries show the number of 10 acre cells of each 1963 land use that converted to another land use by 1970. A blank denotes that no conversion occurred between the respective land uses during the period covered. Entries along the principal diagonal represent the number of cells of the land use that did not change to another land use between 1963 and 1970.

1963 Land-Use Type	Code	Converting to 1970 Land Use																					1963 Land-Use Totals
		11	12	13	14	15	16	17	18	19	21	22	23	24	41	51	52	53	55	61	73	75	
Residential	11	6393	11	2			2			12						1							[illegible]
Commercial and service	12	2	1089							6												[illegible]	[illegible]
Industrial	13	3	1	882			1																[illegible]
Extractive	14	3		[illegible]	444						[illegible]						[illegible]						[illegible]
Transportation, communication, and utilities	15					729																	[illegible]
Institutional	16	14	2		1		[illegible]			14							15						[illegible]
Strip and clustered development	17							[illegible]															[illegible]
Mixed urban	18							4															4
Open and other urban land	19	239	52	46	[illegible]	[illegible]	42	[illegible]		[illegible]	46						[illegible]					[illegible]	[illegible]
Cropland and pasture	21	394	27	125	[illegible]	2	[illegible]	[illegible]		[illegible]	[illegible]					4	22	[illegible]				[illegible]	[illegible]
Orchards and other horticultural areas	22											[illegible]											[illegible]
Feeding operations	23																						[illegible]
Other agricultural land	24	[illegible]			[illegible]	[illegible]				[illegible]													[illegible]
Deciduous forest land	41																						[illegible]
Streams and waterways	51	1																					[illegible]
Lakes	52																[illegible]						[illegible]
Reservoirs	53																	[illegible]					[illegible]
Other water	55																		[illegible]				[illegible]
Vegetated non-forested wetland	61																			[illegible]			[illegible]
Sand other than beaches	73			1	[illegible]					2													4
Other barren land	75						4				[illegible]												[illegible]
1970 Land-Use totals		7050	1182	1061	[illegible]	865	3159	[illegible]	[illegible]	[illegible]	14624	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	

data planes. These stochastic elements are summarized in a matrix of the form:

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,34} \\ p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,34} \\ \cdot & \cdot & p_{i,j} & \cdots & \cdot \\ p_{34,1} & p_{34,2} & p_{34,3} & \cdots & p_{34,34} \end{bmatrix} ,$$

where $p_{i,j}$ is the average probability that a given 4-ha (10-acre) cell in state i as of 1963 would evolve into state j in 1970. Alternately, a briefer form is

$$P = \{p_{i,j}\} \text{ for } i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, m$$

where the braces indicate that $p_{i,j}$ is a typical element of the stochastic matrix, P, the limits of i and j being m. The element, $p_{i,j}$, is called the i, j^{th} element, and the first subscript refers to the element's row and the second subscript, to the element's column.

The probability transition matrix, $P = \{p_{i,j}\}$, between 1963 and 1970 for the Denver Metropolitan Area, is reproduced with nonoccurring, or zero land-use change, probabilities deleted for more convenient reference (table 5). The elements on the principal diagonal of the probability transition matrix, $p_{i,i}$, are significant in that they represent the proportion of the earlier (1963) land use that remained in the same land use in the later (1970) date. Elements not on the principal diagonal are transition probabilities (or proportions) for a given land use to change in the given time interval. All rows in the matrix are stochastic vectors, that is, the entries sum to one across any row, or in dot notation,

$$p_{i} = \sum_{j=1}^{m} p_{i,j} = 1$$

The probability transition matrix was input to a Markov projection process that operated on the relative amounts of each of the 21 land-uses present in the two most current land-use data planes (1963 and 1970) to project future trends in land use (figure 42). Fifteen 7-year simulation periods were run from 1963 to 2068 and yielded (table 6) rapid decreases in two functional open-space categories–open and other urban land (code 19) and cropland and pasture agricultural land (code 21). These two land uses in the Markov model declined from 15,145 and 64,813 ha (37,410 and 160,090 acres), respectively, in 1963 to 1,280 and 7,753 ha (15,340 and 43,180 acres), respectively, in 2068. Urban residential land use (code 11) nearly doubled, increasing from 26,000 ha (64,210 acres) in 1963 to 50,700 ha (121,600 acres) in 2068.

The Markov trend model was intended to augment the land manager's knowledge and experience of the future implications of current actions and to permit more thorough searches for "best" ecomanagement decisions. Although these current trends can be mathematically projected as far into the future as desired, only the first few time periods are reliable. A major limitation of all such models is the assumption of homeostasis in that the probability transition matrix of conversions, $P = \{p_{i,j}\}$, of land from type i to type j is invariant and constant during all future simulated time periods. A more serious shortcoming is the fact that it was simply a descriptive trend model, and, although it indicated general trends over time,

TABLE 5

DENVER LAND-USE PROBABILITY TRANSITION MATRIX, 1963 TO 1970. Entries in the matrix denote the fraction of each 1963 land use that converted to another land use in 1970 as computed from table 4. A blank in the matrix denotes that no conversion occurred between the respective land uses during the period covered. A 1.0 in the matrix denotes that the area of that land use either did not change or increased by conversion to it from another land use between 1963 and 1970. Entries less than 0.001 are indicated by a dash.

1963 Land-Use Type	Code	Converting to 1970 Land Use																				
		11	12	13	14	15	16	17	18	19	21	22	23	24	41	51	52	53	55	61	73	75
Residential	11	0.996	0.002	-						0.002												
Commercial and services	12	0.002	0.988																			
Industrial	13	0.003	0.001	0.994																		
Extractive	14	0.006		0.011	0.959																	
Transportation, communications and utilities	15																					
Institutional	16																					
Strip and clustered development	17																					
Mixed urban	18																					
Open and other urban land	19																					
Cropland and pasture	21																					
Orchard and other horticultural areas	22																					
Feeding operations	23																					
Other agricultural land	24																					
Deciduous forest land	41																					
Streams and waterways	51																					
Lakes	52																					
Reservoirs	53																					
Other water	55																					
Vegetated nonforest wetland	61																					
Sand other than beaches	73																					
Other barren land	75																					

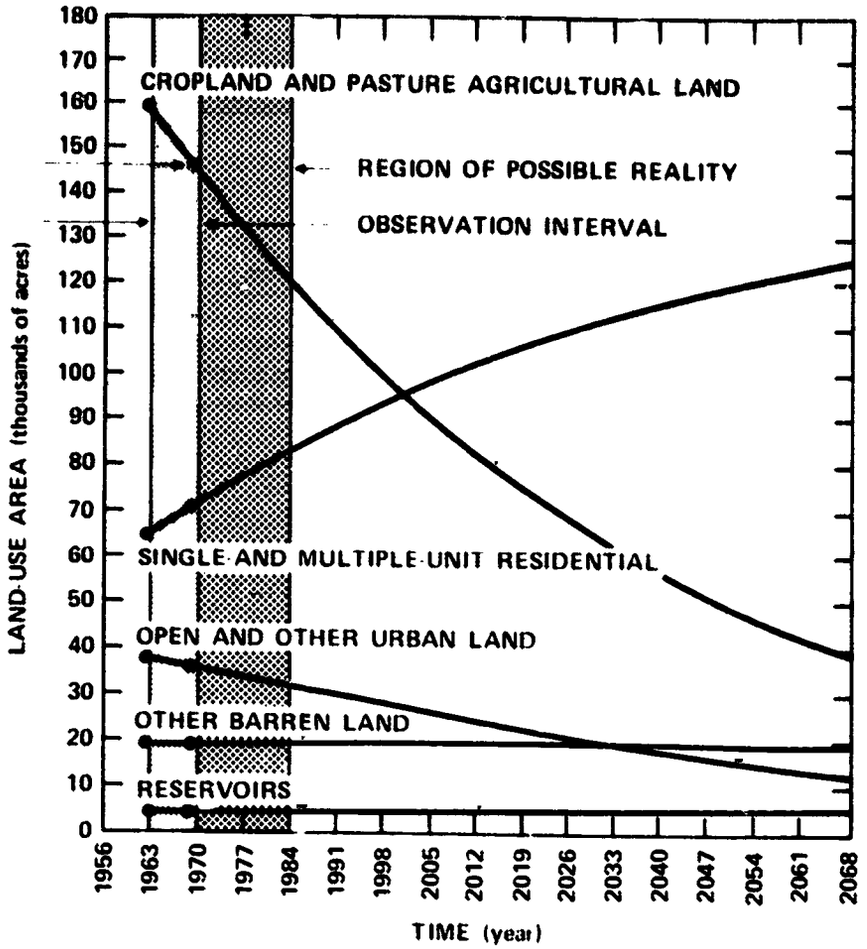


FIGURE 42. PREDICTION OF FUTURE TRENDS IN THE AMOUNT OF OPEN SPACE AND COMPETING LAND USE IN THE DENVER METROPOLITAN AREA. A constant matrix of transfers, P $\{p_{i,j}\}$, from land use, i, to land use, j, was assumed. Acreages displayed relate to the original area of 24- by 24-statute miles 576 square statute miles.

it did not provide spatial land-change information on a point-by-point basis. Finally, only gross interpretations can be made of the long-term evolution of land use from this model. New and unanticipated controls on the use of the land can quickly occur and impact future urban growth and attendant land-use patterns.

MULTIVARIATE LAND-USE SPATIAL PREDICTION MODEL

The land-use modeling effort was next addressed to the design and structure of a spatial temporal land-use change model that could be used for map predictions. The Markov chain process, in which the probability of transition from one state to another is the conditional probability, p (j|i), was a useful point of departure. However, its simplistic matrix multiplication of land-cover types was completely inappropriate for spatial analysis and prediction. Like most regionalization problems, this next level of complexity must characteristically be addressed to more numerous geographic observations where each consists of measurements on a large number of variables. It was therefore apparent that any spatial land-use modeling would be impossible without commitment to a full multivariate statistical approach. This approach was largely a classification exercise and employed the techniques of modern numerical taxonomy (Reference 27).

TABLE 6

DENVER AGGREGATE LAND-USE PROJECTIONS BY A MARKOV TREND MODEL, 1963 TO 2068. Specific land-use total areas, such as those illustrated in figure 42, are detailed here. Photointerpreted second-order land-use changes between 1963 and 1970 were used to drive this generalized trend model for the 24- by 24-mile Denver Metropolitan Area. The basic assumption of the Markov model is that the rates of changes are constant over time for all land-use classes. This model can project only aggregate land-use areas and cannot spatially predict the actual sites of conversion.

Land-Use Type	Measured Acreages		Projected Acreages													
	1963	1970	1977	1984	1991	1998	2005	2012	2019	2026	2033	2040	2047	2054	2061	2068
Residential	64,710	70,500	76,320	81,280	86,680	91,280	95,830	99,450	103,080	106,400	109,470	112,310	114,920	117,330	119,550	121,600
Commercial and services	11,420	12,520	12,980	13,300	13,960	14,590	15,180	15,730	16,260	16,750	17,210	17,640	18,040	18,430	18,760	19,120
Industrial	9,810	10,610	12,200	13,300	15,170	16,470	17,690	18,810	19,860	20,830	21,730	22,570	23,350	24,070	24,740	25,360
Extractive	4,060	5,660	7,800	9,000	10,090	10,980	11,730	12,350	12,860	13,260	13,580	13,810	13,980	14,080	14,130	14,120
Transportation, communication, and utilities	7,240	8,650	9,920	10,600	11,890	12,810	13,660	14,450	15,190	15,870	16,500	17,090	17,630	18,140	18,610	19,040
Institutional	31,750	31,590	31,880	32,130	32,330	32,500	32,620	32,710	32,770	32,810	32,840	32,790	32,740	32,680	32,600	32,500
Strip and isolated development	13,500	16,550	19,310	21,890	24,190	26,330	28,300	30,110	31,770	33,300	34,710	36,010	37,200	38,300	39,320	40,260
Mixed urban	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Open and other urban	37,430	35,620	33,280	31,900	30,200	28,470	26,810	25,220	23,700	22,270	20,920	19,650	18,460	17,340	16,310	15,340
Cropland and pasture	160,080	146,240	133,600	122,120	111,670	102,150	93,500	85,610	78,430	71,900	65,940	60,510	55,560	51,050	46,940	43,180
Orchards and other horticultural areas	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
Feedlot operations	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Other agricultural land	330	70	20	20	20	20	20	10	10	10	10	10	10	10	10	10
Deciduous forest land	180	180	180	180	180	180	180	180	180	180	180	180	180	180	180	180
Streams and waterways	980	1,010	1,060	1,100	1,140	1,180	1,210	1,250	1,280	1,300	1,330	1,360	1,380	1,400	1,420	1,440
Lakes	5,930	6,410	6,800	7,340	7,800	8,260	8,700	9,140	9,540	10,000	10,420	10,830	11,240	11,630	12,020	12,400
Reservoir	1,580	1,750	1,910	2,000	2,180	2,300	2,400	2,500	2,590	2,680	2,750	2,820	2,890	2,950	3,000	3,050
Other water	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
Vegetated nonforested wetland	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710	1,710
Salt flats that produce	640	520	420	340	280	230	180	150	120	100	80	70	50	40	30	30
Other nonurban land	18,870	18,920	18,960	19,080	19,040	19,070	19,090	19,110	19,130	19,140	19,150	19,160	19,170	19,170	19,170	19,160
Grand Total	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640	368,640

Observation interval Region of possible reality

Relationship of Land-Use Change to Landscape Variables

A particular land use can be considered as a class in a classification system, and, further, each class is defined by its similarity to other class members and some level of differentiation from nonclass members. Likewise, each type of land-use change that was observed between 1963 and 1970 was quantitatively defined by its associated landscape parameters for the Denver Metropolitan Area. It was assumed that they exhibited some similarity to other cells in the change class and differentiation from nonchanged cells.

The important assumption that cells of land use that have undergone similar changes in a particular period have some common landscape features was intuitive. Intuition was borne out by the subsequent analysis and was increased by visual display. Three histograms of the cells of agricultural land use (code 21) were generated as a function of topographic elevation from the 1963 and 1970 land-use data planes (figures 17 and 15) and for the cells converted out of this class between 1963 and 1970 from the change plane (figure 43). These graphics were not scaled identically in the vertical plotting dimension and cannot be directly visually compared. Statistical values computed with these figures indicated that the average elevation of agricultural land use was 1,692 and 1,693.4 meters (5,550 and 5,556 feet) for the 1963 and 1970 land use, respectively. The average elevation of the agricultural land loss during this period was 1,670 m (5,480 feet). A skewed frequency distribution in the histogram of the loss versus elevation (figure 43c) signified a higher likelihood or probability of conversion for agricultural lands at lower elevations. The majority of the agricultural land transition between 1963 and 1970 was into urban land uses (table 4); 5,414 of 5,607 total hectares (13,373 of 13,850 total acres) were lost, predominantly at the lower, more suitable elevations. It was clear that the lower agricultural lands adjacent to existing urban lands were more prone to suburbanization processes. Thus, a multivariate rationale for quantifying the mathematical properties of land-use change types for all of their physiographic, socioeconomic, and transportation variables provided a basis for a spatial/temporal land-use projection model.

Spatial Modeling Concepts

One of the most widely used multivariate procedures is discriminant analysis (Appendix A). *A priori* data in the form of selected samples of known identity are extracted from the entire sample space and are used as training data to structure the discriminant function. This discriminant function is then used to classify the remaining balance or unknown portion of the pattern space. The discriminant function is a powerful statistical tool that can be applied completely free of statistical knowledge or assumptions. This overall approach is commonly referred to as distribution-free or nonparametric classification (Reference 28). All classification algorithms in discriminant-function analysis can be reduced to either fixed or varying hyperplanes of pattern or feature spaces. These hyperplane boundaries may or may not have been defined in the context of known statistical distributions (Reference 29).

A linear discriminant function transforms an original set of sample measurements into a single discriminant score. This score, or transformed variable, represents the sample's position on the line defined by the linear discriminant function. Consequently, the discriminant function can be visualized as a method of telescoping a multivariate problem into a univariate, linearly ordered situation.

Discriminant-function analysis seeks to find a transform that gives the minimum ratio of the difference between a pair of class univariate means to the multivariate variance within the two classes in the simple linear case. These two classes may be visualized as consisting of two swarms of data points in multivariate space; the one optimum orientation is sought along which the two clusters have the greatest separation while

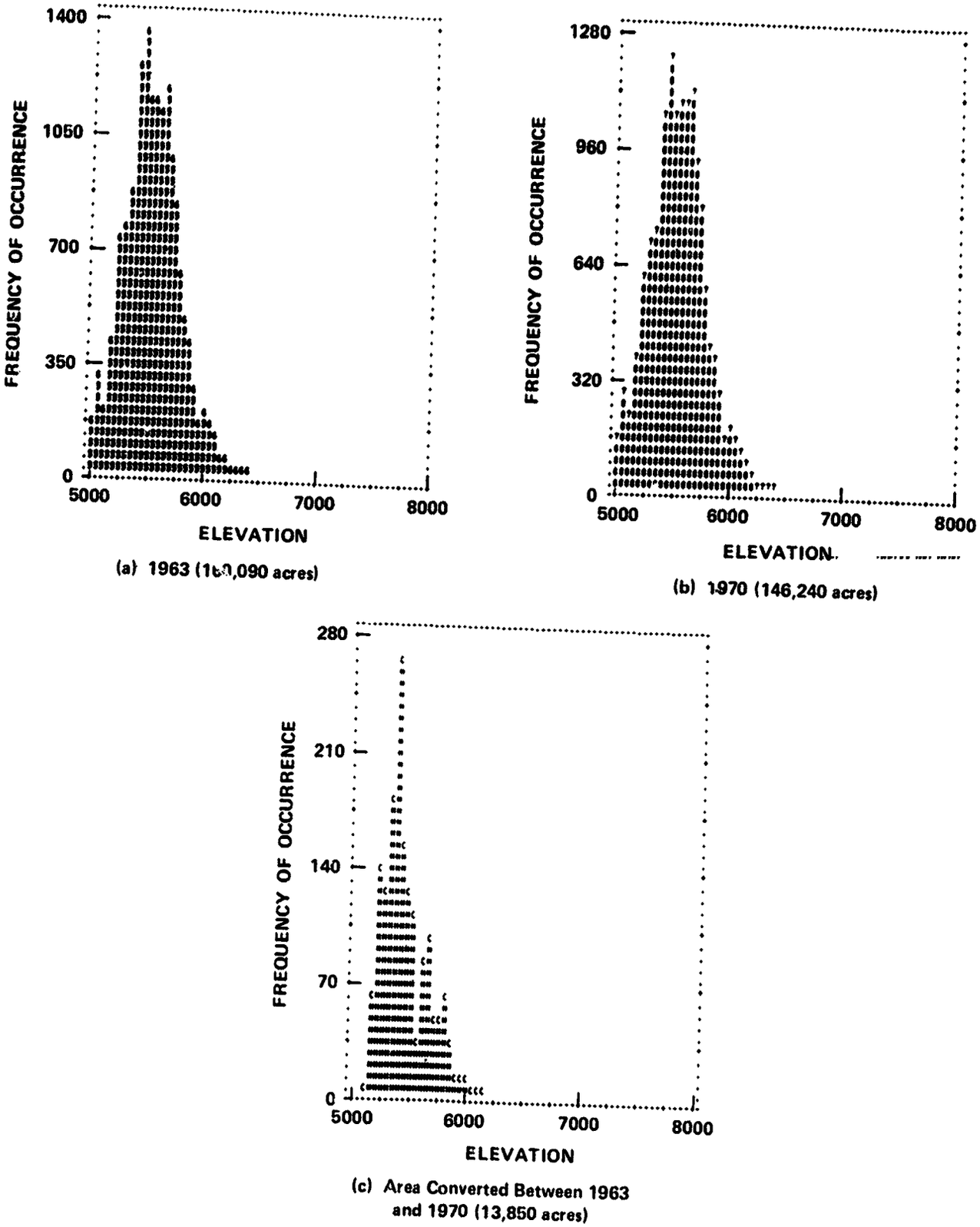


FIGURE 43. CORRELATION OF THE CHANGE IN AGRICULTURAL LAND USE WITH ELEVATION. (a) Histogram of the agricultural land use in 1963 versus elevation. (b) Histogram of the agricultural land use in 1970 versus elevation. (c) Histogram of that portion of the agricultural land use that converted to some other land use between 1963 and 1970 as a function of elevation.

simultaneously minimizing the spread or inflation of the distribution of each cluster. An adequate separation between groups A and B for the two-dimensional case cannot be made with either variable X_1 or X_2 (figure 44). However, it is feasible to find an orientation along which the two clusters are separated the most and inflated the least, with the coordinates of this axis of orientation being the linear discriminant function.

CLASSIFY was the modified BMD07M (Reference 18) stepwise linear discriminant-analysis program used in this final phase of spatial land-use projection. Linear discriminant functions were computed by entering ancillary landscape variables with a largest F-value-to-enter criterion. The *a posteriori* probabilities of each land-use case belonging to each of the possible change combinations in land use between 1963 and 1970 were calculated using these functions and either *a priori* or equal (1/n) group probabilities. Additional programming changes to provide tape input and output features, a more useful classification matrix, and a timing function to evaluate the cost-effectiveness of each additional stepwise variable resulted in the present CLASSIFY program.

Initial Spatial Modeling Test

The data planes of the earlier (1963) and current (1970) land use and the derived 1963 to 1970 land-use changes were overlaid in the landscape model with the 34 ancillary variables. The 2,039 of the 36,864 cells in the model that were observed to have changed between the two land-use dates provided the basis for determining how the landscape variables correlated with changes in land use in a multivariate sense. The 2,039 cells that made a transition from one land use to another over the 7-year test period provided a group of observations for each type of change that occurred. These observations were used as a statistical sample to model how the changes in land use will proceed in the future based on physiographic, socioeconomic, and transportation landscape variables.

The Denver Metropolitan Area was mapped with 21 land-use classes on both the initial (1963) and subsequent (1970) dates. Thus, 21^2-21 or n(n-1) change combinations were possible. However, only 38 of these possible 420 changes actually occurred as the earlier probability transition matrix shows (table 5). Many kinds of possible land-use changes either seldom or never occur (e.g., the backward conversion from some higher state of land use, such as industrialized land, back to a lower state, such as agricultural land). All test cells that had undergone a specific transition were assembled so that the 2,039 changes observed between the two dates were grouped into the 38 observed combinations. The statistical technique of discriminant analysis was applied to these observations to form a simulation model by training it to recognize the range in a multivariate sense of each landscape variable associated with each of the 38 change combinations.

A single, simplified test model will serve to clarify this approach. The first spatial land-use projection model used all 38 change combinations, together with the landscape variables and equal (1/n) class probabilities. The seven land-use data variables of the 34 variables overlaid in the landscape model were excluded from the independent variables in this and all subsequent analyses so that they represented the results being modeled. Linear discriminant analysis combined the remaining 27 landscape variables for each change point by constructing a linear discriminant function of the form

$$Y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \quad (n=27)$$

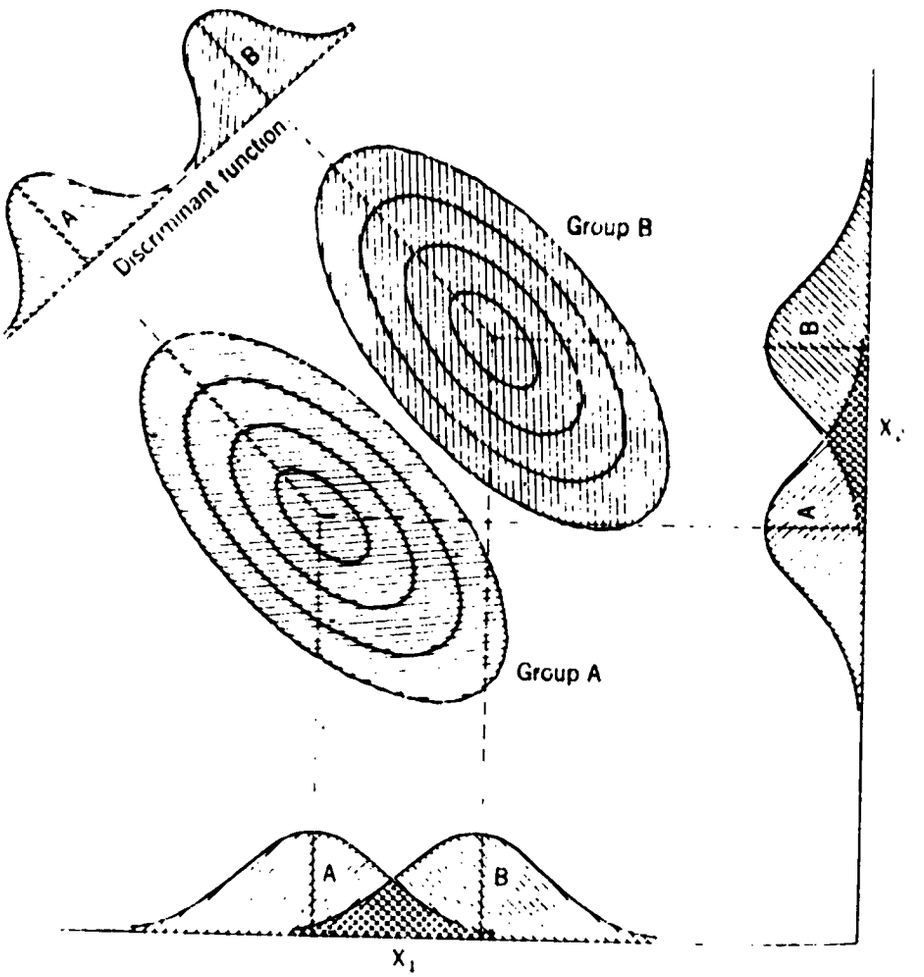


FIGURE 44. SIMPLE LINEAR DISCRIMINANT-FUNCTION DIAGRAM. This plot of two bivariate distributions shows an overlap between groups A and B along both variables, X_1 and X_2. Discriminant-function analysis determines a transform that gives the minimum ratio of the difference between a pair of group multivariate means to the multivariate variance within the two groups. The orientation is computed along which the two clusters are separated the most and inflated the least. The coordinates of this orientation are the linear discriminant function, and the clusters become distinguishable by projecting members of the two populations onto the discriminant function line (Reference 30).

where $x_1, x_2, \ldots, x_n$ were the landscape variable values, and $a_1, a_2, \ldots, a_n$ were coefficients computed to determine a value for Y, the linear compound, that minimized misclassification of the 2,039 changed cells into the 38 change combinations. The function was checked by seeing how well it classified its calibration data set of 1963 to 1970 changed cells. The final discriminant analysis step printed a classification matrix showing how the 2,039 changed cells were classified into their respective 38 change combinations. The evaluation of this and each subsequent model's value was based on a figure of merit (FOM), calculated as the sum of the correct change classification along the principal diagonal of the classification matrix divided by the total number of changed cells and expressed as a percent.

The FOM for the first model was 41.9 percent (table 7). Overall, it correctly identified the expected change in land use of 854 of the 2,039 observed cells that had changed based upon the 27 landscape variables (table 8). Because a choice was made from 38 different land-use change combinations, an accuracy of 1/n = 38 groups times 100 percent = 2.6 percent would be expected from random assignment of the 2,039 changed cells to 38 combinations. Although, by comparison, the 41.9-percent figure-of-merit was far from perfect, it clearly indicated the feasibility of this spatial land-use projection approach.

The entry order of the change variables may be of some interest to planners. Examination of the order of the "F-value to enter" also provides insight into the relative value of the landscape variables in predicting land-use change (table 9). The distance to freeways was found to be an influentual factor and was selected early by the classifier. However, the proximity to the periphery of the urban built-up area was even more important. Average number of families per acre, population density per acre, and topographic elevation were also important descriptors of change for the Denver Metropolitan Area.

Tests of Other Generic Spatial Models

An examination of the initial test classification matrix (table 7) showed that it failed to exploit the known *a priori* 1963 land use of all change cells. For example, the land-use change code 21/19, representing 1963 cropland and pasture agricultural land (code 21) changing to open and other urban land (code 19) in 1970, could be classified in any of 37 other remaining land-change combinations. This specific change, in fact, was predicted to occur in 27 other land-use change combinations, 15 of which were inconsistent with the fact that the 1963 initial land use was cropland and pasture agriculture (code 21). Because the initial or starting land use would always be known, the choice of a type of change must be made to be consistent with the known, initial land use.

Six additional spatial land-use projection models incorporating the logic of a known, beginning land use were tested with varying degrees of improvement over the initial model (table 10). The second and third models were derivatives of the initial full 38-class model restricted to the 1963 cropland and pasture agricultural change cases, 1,437 cells, and represented 13 of the 38 change combinations under alternative *a priori* and equal class occurrence probabilities.

The second model correctly reclassified 822 of the 1,437 agricultural change cases for a 57.2-percent FOM (figure 45) with *a priori* class probabilities (tables 11 and 12), a significant increase over the comparable 854 correct cases from among all 2,039 cells (41.9 percent) achieved in the initial model. The third model used the same 1,437 agricultural change cells but differed in that it used equal (1/n = 13) class probabilities and yielded 756 correct cases for a 52.6-percent FOM.

The only difference between the second and third models was the use of the *a priori* versus equal-class occurrence probabilities, respectively, which yielded a slight classificational advantage to the second model. Equal-class likelihood was used where the actual proportion of the modeled land-use (1970) changes was unknown. The *a priori* probabilities of the second model could be either estimated by remote sensing or extrapolated from mathematical simulation models, such as the Markov trend model. Here, the *a priori* probabilities were simply derived as the ratio of the total cases of each change class to the grand total of change cases.

This Page Intentionally Left Blank

INITIAL 38-CLASS 1963 TO 1970 LAND-USE CHANGE MODEL-CLASSIFICATION MATR
bined physiographic, socioeconomic, and transportation characteristics of the 2,039 changed ce
to linear discriminant analysis, and 864 cells were correctly classified for an average 42-percent

Actual 1963 to 1970 Change Combinations	Number of
Change Code	(14/11)	(14/13)	(14/21)	(14/52)	(16/52)	(18/17)	(19/11)	(19/12)	(19/13)	(19/14)	(19/15)	(19/16)	(19/17)	(19/21)
(14/11)	2	1												
(14/13)		5												
(14/21)			3											
(14/52)	1			4										
(16/52)					15									
(18/17)						4								
(19/11)	2	5	5	1			40	20		1		39	4	8
(19/12)		1	1				2	16	1	4		2	1	3
(19/13)				1				2	21	2	2	1		
(19/14)	1	2								18				3
(19/15)								1	2	12	19		1	1
(19/16)		1					2					14		
(19/17)			1	1			1			1			23	2
(19/21)			3				3	2		3	3	2	3	8
(19/51)														
(19/52)										1				
(21/11)		7		3			7	2	6	7		20	10	
(21/12)		1					1					1		
(21/13)		1									14			2
(21/14)				2					1	4	26		3	
(21/15)			2						3	6			3	
(21/16)		1							2					
(21/17)	2	4				5	4		4			6	8	4
(21/19)	2	3		3			15	3	15	5	10	2	3	5
(21/24)														
(21/51)										1				
(21/52)			2							1				
(21/53)														
(21/75)														
(24/11)														
(24/14)														
(24/15)														
(24/19)														
(24/21)														
(24/52)														
(73/13)														
(73/14)										1				
(73/19)														

ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME

TABLE 7

IX. The 27 com-
lls were subjected
accuracy.

Land-use data were not used, except for identifying the cells that had changed. On the axes, the numerator is the 1963 land use, and the denominator is the 1970 land use. Numeric land-use codes are identified in table 1.

10-Acre Cells Classified Into 38 Second-Order Land-Use Combinations

(19 51)	(19 /52)	(21 /11)	(21 /12)	(21 /13)	(21 /14)	(21 /15)	(21 /16)	(21 /17)	(21 /19)	(21 /24)	(21 /51)	(21 /52)	(21 /53)	(21 /75)	(24 /11)	(24 /14)	(24 /15)	(24 /19)	(24 /21)	(24 /52)	(73 /13)	(73 /14)	(73 /19)
		27	5		7		30	4	1		4	5		1					3	3	5	1	2
3	5		2					4	1			3					2		2				
	1			2								1									3		9
																							9
5	1			6							2											1	
		2			3		3	1															
3											2	1				1		2	1				
5		1		4			1	2			2							1					
1																							
	4																						
		218	7	6	1		29	17	19			13				9	3		7		1	1	4
1	4	5	7				3		1			2					1						
		1		57	10		2	2	23		5	3				1			1	1			
				13	52			1	1		6	2	11	2		4	5	1		2		3	
				1		35	2	2		5			1	2		1		1	8				
		3	1				11	3			1										1		
		10	4	20	3		14	103	53	6		3				6	3						
		30	20	10	2		11	41	115	1	2	17		8		4	7	1	5			1	3
										2													
										3													
		1			1	1			2	1	3	9							1				
												17											
									1	1													
												3											
												1											
												5					1			1			
												1											
												2				3	7			1			
																		2					
																		3					
																		1					
																		1					
																						3	5
																							2

ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME

This Page Intentionally Left Blank

TABLE 8

ACCURACY OF PREDICTION BY THE INITIAL MODEL OF FUTURE CHANGES IN LAND USE ON A CELL-TO-CELL OR SPATIAL BASIS FOR THE DENVER METROPOLITAN AREA. This table is based on predicting the change in land use from 1963 to 1970 for the 2,039 cells in the initial landscape model that underwent a change during this period. The change was predicted by discriminant analysis from the landscape variables. The accuracy was based on the number of cells whose future or 1970 land use was correctly predicted (table 7).

1963 Land Use Predicted to Change to 1970 Land Use		Total Sample Cells	Total Cells Correct	Prediction Accuracy (percent)
Extractive				
Δto	Residential	3	2	66.7
	Industrial	5	5	100.0
	Cropland and pasture	3	3	100.0
	Lakes	5	4	80.0
Institutional				
Δ to	Lakes	15	15	100.0
Mixed urban				
Δto	Strip and clustered development	4	4	100.0
Open and other urban				
Δto	Residential	231	40	17.3
	Commercial and services	53	16	30.2
	Industrial	45	21	46.7
	Extractive	33	18	54.5
	Transportation, communications, and utilities	51	19	37.3
	Institutional	26	14	53.8
	Strip and clustered development	39	23	59.0
	Cropland and pasture	43	8	18.6
	Streams and waterways	1	1	100.0
	Lakes	5	4	80.0
Cropland and pasture				
Δto	Residential	397	218	54.9
	Commercial and services	27	7	25.9
	Industrial	123	57	46.3
	Extractive	139	53	37.4
	Transportation, communications, and utilities	72	35	48.6
	Institutional	23	11	47.8
	Strip and clustered development	262	103	39.3
	Open and other urban	344	115	33.4
	Other agricultural land	2	2	100.0
	Streams and waterways	4	3	75.0
	Lakes	22	9	40.9
	Reservoirs	17	17	100.0
	Other barren land	5	3	60.0
Other agricultural land				
Δto	Residential	1	1	100.0
	Extractive	8	5	62.5
	Transportation, communications, and utilities	13	7	53.8
	Open and other urban	3	3	100.0
	Cropland and pasture	3	3	100.0
	Lakes	1	1	100.0
Sand other than beaches				
Δto	Industrial	1	1	100.0
	Extractive	9	3	33.3
	Open and other urban	3	3	100.0
Total		2039	851	41.9

TABLE 9

LANDSCAPE VARIABLE ENTRY ORDER OF THE INITIAL SPATIAL LAND-USE PROJECTION MODEL. The numerical utility of the landscape variables as determined in the initial land-use change model (table 10) is given in descending order. The F-values reported were determined during the stepwise portion of the discriminant analysis.

Step Number	Ancillary Landscape Variable Entered	F-Value to Enter
1	Average number of families per acre	48.3188
2	Built-up urban-area minimum distances	31.1547
3	Population density per acre	26.2660
4	Topographic elevation	19.6477
5	Freeway minimum distances	17.8388
6	Median housing-unit rent	11.5608
7	Composite minor-road minimum distances	11.3720
8	Average number of year-round housing units per acre	10.1748
9	Median housing-unit value	8.9827
10	Total occupied housing units	8.9332
11	Average number of housing units per acre	6.5551
12	Total one-car families	7.0151
13	Topographic slope	6.3456
14	Freeway interchange minimum distances	5.8250
15	Landsat-1 image insolation	5.2111
16	Total three-car families	3.9773
17	Total year-round housing units	9.3293
18	Total census-tract acreage	9.8015
19	Total two-car families	6.4901
20	Total vacant housing units	6.2008
21	Total families	5.0273
22	Average number of cars per family	4.4601
23	Composite major-road minimum distances	3.7221
24	Topographic aspect	3.5609
25	Surficial geology	3.3950
26	Total population	1.5531
27	1969 Mean family income	0.8823

TABLE 10

COMPARISON OF SEVEN TEST LAND-USE MODELS OF SPATIAL CHANGE IN THE DENVER METROPOLITAN AREA, 1963 TO 1970. This table is based on predicting the change in land use for the cells that had undergone a change during the test period. The change was predicted by discriminant analysis from the landscape variables. The accuracy was based on the number of cells whose future or 1970 land use was correctly predicted.

Model Number	Model Description	Total Sample Cells	Total Cells Correct	Average Accuracy (percent)
1	Full 38-change class model with 1970 equal-change probabilities	2,039	854	41.9
	(Subportion of thirteen cropland and pasture agricultural changes)	(1,437)	(633)	(44.1)
2	Second composite 38-change class model with 1970 *a priori* change probabilities	2,039	1,208	52.9
	(Subportion of thirteen cropland and pasture agricultural changes)	(1,437)	(822)	(57.2)
3	Thirteen cropland and pasture agricultural changes of 38 total change classes with 1970 equal-change probabilities	1,437	756	52.6
4	Pure 13 cropland and pasture agricultural change classes with 1970 *a priori*-change probabilities	1,437	799	55.6
5	Pure 13 cropland and pasture agricultural change classes with 1970 equal-change probabilities	1,437	747	52.0
6	Fourteen cropland and pasture agricultural change nonchange classes with 1970 equal-change probabilities	2,347	1,241	52.9
7	Fourteen cropland and pasture agricultural change nonchange classes with 1970 *a priori*-change probabilities	2,347	1,128	48.1

The fourth and fifth models were similar to the second and third models, respectively, in that they used the same 1,437 agricultural change cases representing thirteen 1970 change combinations with *a priori* and equal-class occurrence probabilities, respectively. Unlike the second and third models, however, the sample case files contained only the 13 agricultural change cases, whereas the three previous models derived their discriminant functions from the full 38-class change combinations.

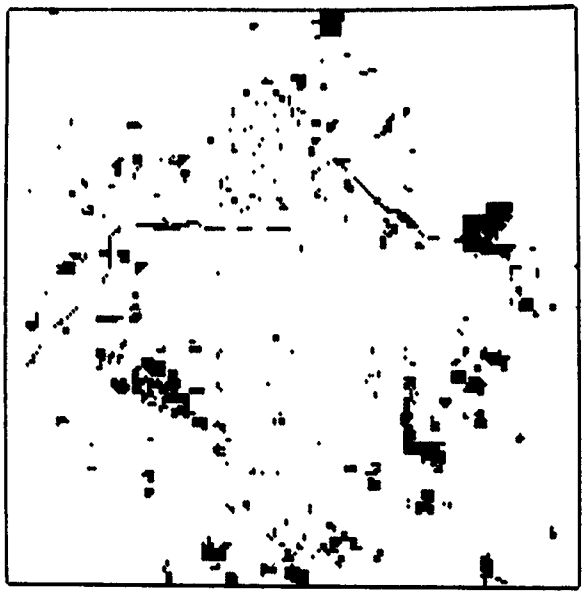
The fourth model correctly reclassified 799 agricultural change cases for a 55.6-percent FOM (figure 45), which was somewhat less than the 57.2 percent FOM (822 correct cases) achieved in the comparable second model.

This Page Intentionally Left Blank

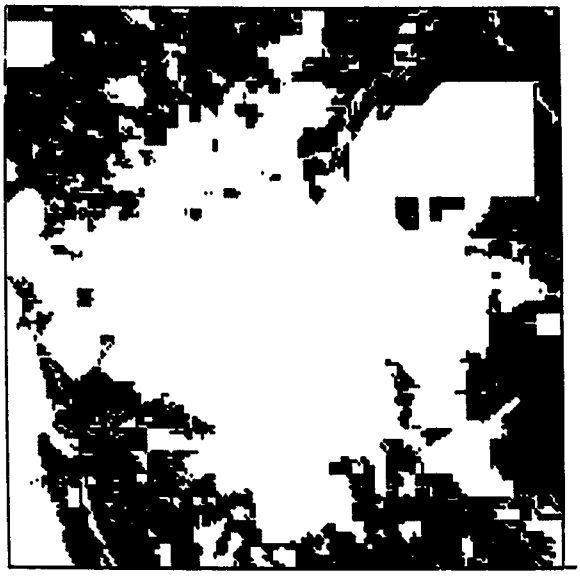
(c) Thirteen agricultural change classes predicted to have changed from agricultural land as of 1970 (second model)

FOLDOUT FRAME

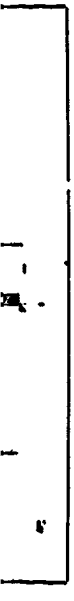
ORIGINAL PAGE IS
OF POOR QUALITY

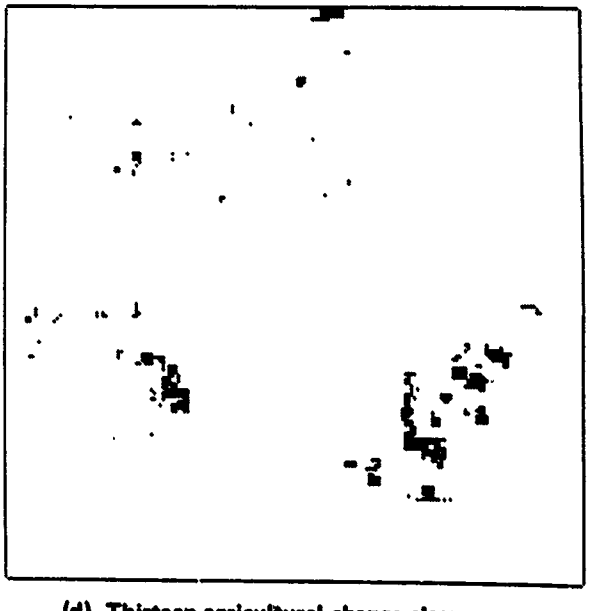


(b) All observed changes from agricultural land from 1963 to 1970

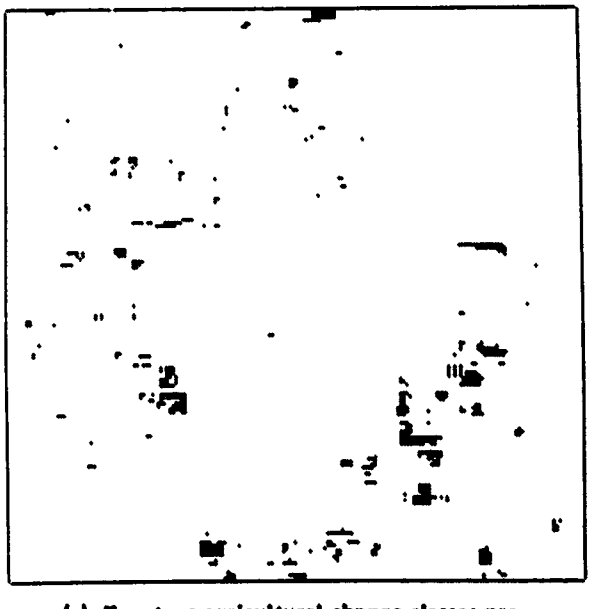


(a) All 1970 Agricultural Land-Use





(d) Thirteen agricultural change classes predicted to have changed from agricultural land as of 1970 (fourth model)



(e) Fourteen agricultural change classes predicted to have changed from agricultural land as of 1970 (sixth model)

(a) Ground-truth map derived from the 1970 aerial photointerpretation showing only the first-order agricultural land use. (b) Change detection map derived from a point-to-point comparison of the 1963 and 1970 land-use data planes. (c, d and e) Landscape-modeled maps showing only these cells predicted to evolve from agricultural land use as of 1970.

ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 45. COMPARATIVE DISPLAYS OF THE SPATIAL PREDICTIONS OF THREE LAND-USE MODELS EMPHASIZING ALL CHANGES FROM AGRICULTURAL LANDS (in black).

This Page Intentionally Left Blank

The initial model was partitioned into seven runs, corresponding to each of the basic 1: classes. Blank entries indicate unfeasible change combinations; dashes indicate that no cell the feasible combinations. Land-use data were not used except to select the change cells. identified in table 1.

Change Code	Number													
	(14 /11)	(14 /13)	(14 /21)	(14 /52)	(16 /52)	(18 /17)	(19 /11)	(19 /12)	(19 /13)	(19 /14)	(19 /15)	(19 /16)	(19 /17)	(19 /21)
(14/11)	2	1	–	–										
(14/13)	–	5	–	–										
(14/21)	–	–	3	–										
(14/52)	1	–	–	4										
(16/52)					15									
(18/17)						4								
(19/11)							184	6	13	8	–	2	2	8
(19/12)							16	20	2	5	–	2	1	5
(19/13)							2	3	24	11	2	1	–	1
(19/14)							2	–	1	27	–	–	–	3
(19/15)							2	1	4	12	20	–	1	7
(19/16)							18	–	–	–	–	7	–	–
(19/17)							3	–	3	2	–	–	24	6
(19/21)							10	3	–	6	3	–	5	13
(19/51)							–	–	–	–	–	–	–	–
(19/52)							–	–	–	1	–	–	–	–
(21/11)														
(21/12)														
(21/13)														
(21/14)														
(21/15)														
(21/16)														
(21/17)														
(21/19)														
(21/24)														
(21/51)														
(21/52)														
(21/53)														
(21/75)														
(24/11)														
(24/14)														
(24/15)														
(24/19)														
(24/21)														
(24/52)														
(73/13)														
(73/14)														
(73/19)														

Δ's from 1963 extractive land use (code 14)

Δ's from 1963 open and other urban land use (code 19)

Δ's from 1963 cropland and pastu (code 21)

ORIGINAL PAGE IS OF POOR QUALITY

FOLDOUT FRAME

TABLE 11

3 to 1970 change
were classified into
and-use codes are

SECOND COMPOSITE 38-CLASS 1963 TO 1970 LAND-USE CHANGE MODEL CLASSIFICATION MATRIX. The restriction of change predictions to feasible 1963 classes increased the classification accuracy from 41.9 (tables 8 and 10) to 52.9 percent, using the same 27 combined physiographic, socioeconomic, and transportation variables of the initial model.

10-Acre Cells Classified Into 38 Second-Order Land-Use Combinations																								
(19 51)	(19 /52)	(21 /11)	(21 /12)	(21 /13)	(21 /14)	(21 /15)	(21 /16)	(21 /17)	(21 /19)	(21 /24)	(21 /51)	(21 /52)	(21 /53)	(24 /75)	(24 /11)	(24 /14)	(24 /15)	(24 /19)	(24 /21)	(24 /52)	(73 /13)	(73 /14)	(73 /19)	
–	184																							
–	20																							
–	24																							
–	27																							
3	20																							
–	7																							
1	24																							
3	13																							
1	1																							
–	4																							
		314	2	3	2	–	7	22	43	–	–	–	–	–										
		9	6	–	–	–	3	–	8	–	–	1	–	–										
		6	1	42	22	–	1	1	46	–	3	1	–	–										
		5	–	19	86	1	–	3	7	–	6	4	7	1										
		12	–	2	5	36	–	2	3	5	–	5	1	1										
		8	3	–	4	–	2	2	3	–	1	–	–	–										
		28	–	8	3	–	3	108	101	11	–	–	–	–										
		72	–	3	12	–	5	42	199	1	–	3	–	7										
		–	–	–	–	–	–	–	–	2	–	–	–	–										
		–	–	–	–	–	–	–	–	–	3	1	–	–										
		2	–	–	1	2	–	–	10	–	3	4	–	–										
		–	–	–	–	–	–	–	–	–	–	–	17	–										
		–	–	–	–	–	–	–	1	1	–	–	–	3										
															1	–	–	–	–	–				
															–	5	1	1	1	–				
															–	3	9	–	–	1				
															–	–	–	2	–	–				
															–	–	–	–	3	–				
															–	–	–	–	1	–				
																					1	–	–	
																					–	6	3	
																					–	–	2	

and use

Δ's from 1963 other agricultural land use (code 24)

Δ's from 1963 and other than beaches land use (code 73)

FOLDOUT FRAME

ORIGINAL PAGE IS OF POOR QUALITY

This Page Intentionally Left Blank

TABLE 12

ACCURACY OF PREDICTION OF FUTURE CHANGES IN LAND USE ON A CELL-TO-CELL OR SPATIAL BASIS FOR THE DENVER METROPOLITAN AREA. This table is based on predicting the change in land use from 1963 to 1970 for the 2,039 cells in the second landscape model that underwent a change during the test period. The change was predicted by discriminant analysis from the landscape variables with reference to the initial (1963) land use. The accuracy was based on the number of cells whose future (1970) land use was correctly predicted (table 11).

1963 Land-Use Predicted to Change to 1970 Land-Use		Total Sample Cells	Total Cells Correct	Prediction Accuracy (percent)
Extractive				
Δto	Residential	3	2	66.7
	Industrial	5	5	100.0
	Cropland and pasture	3	3	100.0
	Lakes	5	4	80.0
Institutional				
Δto	Lakes	15	15	100.0
Mixed Urban				
Δto	Strip and clustered development	4	4	100.0
Open and other urban				
Δto	Residential	231	184	79.7
	Commercial and services	53	20	37.7
	Industrial	45	24	53.3
	Extractive	33	27	81.8
	Transportation, communications, and utilities	51	20	39.2
	Institutional	26	7	26.9
	Strip and clustered development	39	24	61.5
	Cropland and pasture	43	13	30.2
	Streams and waterways	1	1	100.0
	Lakes	5	4	80.0
Cropland and pasture				
Δto	Residential	397	314	79.1
	Commercial and services	27	6	22.2
	Industrial	123	42	34.2
	Extractive	139	86	61.8
	Transportation, communications, and utilities	72	36	50.0
	Institutional	23	2	8.7
	Strip and clustered development	262	108	41.8
	Open and other urban	344	199	57.9
	Other agricultural land	2	2	100.0
	Streams and waterways	4	3	75.0
	Lakes	22	4	18.2
	Reservoirs	17	17	100.0
	Other barren land	5	3	60.0
Other agricultural land				
Δto	Residential	1	1	100.0
	Extractive	8	5	62.5
	Transportation, communications, and utilities	13	9	69.2
	Open and other land	2	2	100.0
	Cropland and pasture	3	3	100.0
	Streams and waterways	1	0	0.0
Sand other than beaches				
Δto	Industrial	1	1	100.0
	Extractive	9	6	66.7
	Open and other urban	2	2	100.0
Total		2,039	1,208	59.9

The fifth model correctly reclassified 747 cases for a 52.0-percent FOM, which was also somewhat less than the 52.6-percent FOM (756 correct cases) achieved in the comparable third model.

Two additional types of models were tested to determine if it were possible to include an additional class representing those cells that did not change. The 1963 to 1970 land-use change plane contained 14,624 cells of cropland and pasture that remained unchanged, but 1,437 cells changed to 13 other land uses by 1970. A 910-cell sample of the 14,624 unchanged cells was taken for economy and consisted of every 1[illegible] row and column in the 1963 to 1970 land-use change plane that represented the agricultural class. These 910 cells, representing a nonchange class, together with the 1,437 cells in 13 change classes were combined into a composite test file of 14 classes and 2,347 cells. The sixth and seventh spatial land-use projection models used this composite file and 14 agricultural land-use change/nonchange classes with *a priori* and equal-class occurrence probabilities, respectively.

The sixth spatial model correctly reclassified 1,241 cases for a 52.9-percent FOM (figure 45). Only 437 of the known agricultural change cases were correctly reclassified for a 30.4-percent accuracy, whereas 804 of the sampled nonchange agricultural cases were correctly classified as unchanged for a substantial 98.4-percent accuracy using *a priori* class probabilities. Unfortunately, this error of 11.6 percent appeared [illegible]olerable [illegible] first glance, but, when extended to the original population of 14,624 unchanged agricultural cells, it represented 1,638 unchanged cells erroneously predicted to undergo some change. This number was [illegible] than the 1,437 known changes observed for cropland and pasture agricultural land.

Using equal ($1/n = 1/14$) class probabilities, the last model correctly [illegible] 1,198 cases for a 48.1-percent FOM. Only 663 change cases were correctly classified for a 46.1-percent accuracy, and 465 nonchange cases were correctly modeled for a 51.1-percent FOM that must be interpreted as for the sixth model.

Proposed Advanced Spatial Model

A proposed spatial land-use projection strategy (figure 46) emerged as a synthesis of the Markov trend model and the discriminant-analysis model. The most logical and successful discriminant model (model 2 of table 10) is first applied to all 36,864 cells in the landscape model to predict the next most likely change in land use for all 36,864 4-ha (10-acre) cells. The earlier Markov trend model provides the number of these cells that will convert to a different land use in a given future time increment. After the discriminant model predicts the next change in land use and its posterior probability for each cell in the landscape, the actual changes in a future time period can be determined by assembling all changes of each given type from the 36,864 predictions constituting the entire landscape. The group of cells representing each type of change can be ordered by their posterior probability of occurrence. The correct number of transitions supplied by the Markov trend model can be selected on the basis of the highest posterior probability at the top of each of the ordered list of change types. Cells not selected can be assumed to be unchanged and can be noted as such. The exact spatial location of each cell is preserved by carrying along their respective rows and columns. The sorted cells can be reassembled by row and column, and a predicted map of the future distribution of each land use for a particular date can be displayed. The total modeling process can be iteratively performed to yield a time succession of spatial projections of future land-use maps.

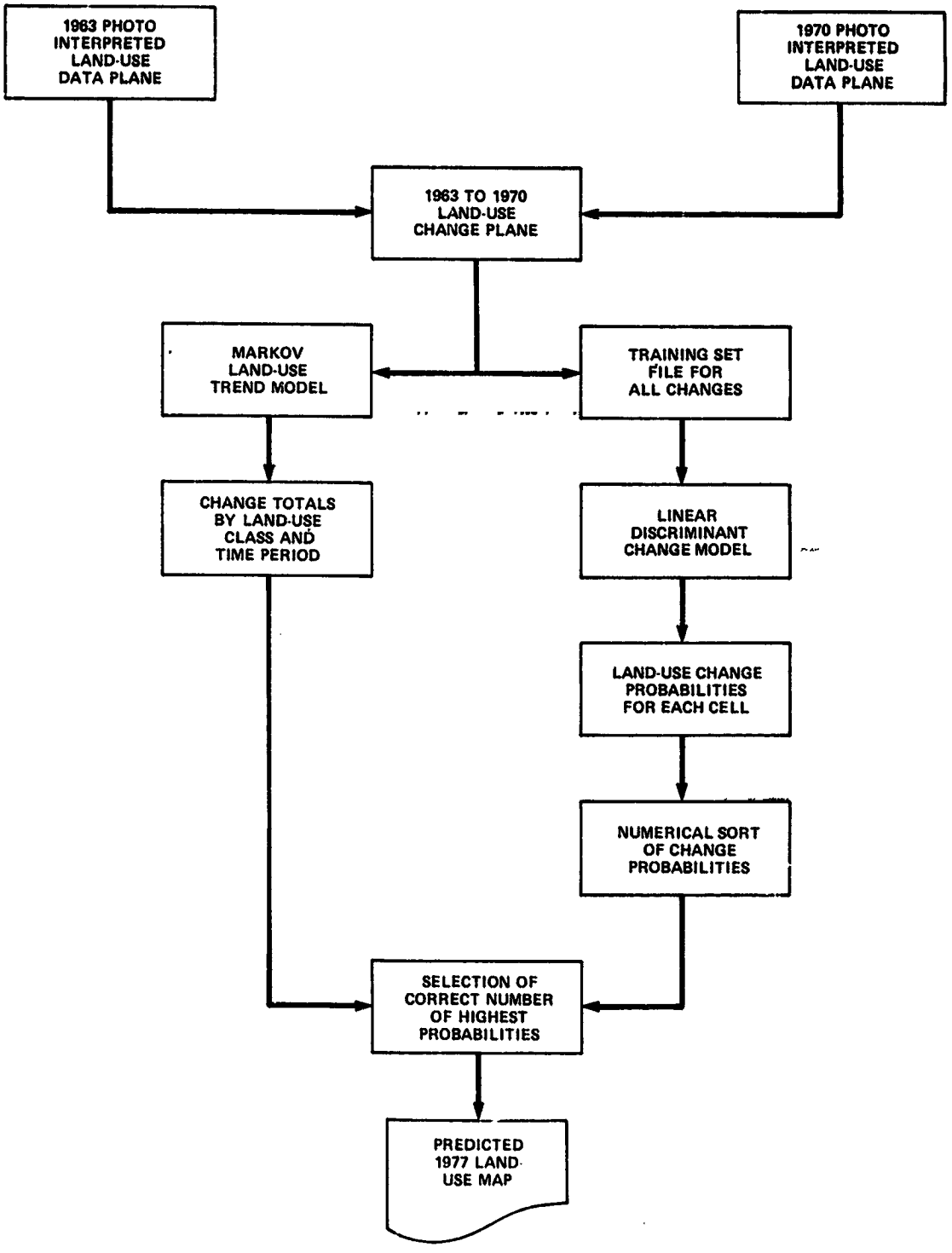


FIGURE 46. PROPOSED COMBINATION MARKOV AND LINEAR DISCRIMINANT MODELS FOR IMPROVED SPATIAL-CHANGE PREDICTION. The Markov trend model provides the correct number of change cells by type. These can be selected from a sorted list of discriminant-computed posterior probabilities of change. The selected cells can be assembled into a map of future land use. Spatially registered Landsat digital imagery can serve as future land-use inputs in lieu of the 1963 to 1970 aerial photography.

This Page Intentionally Left Blank

CHAPTER 4

LANDSAT LAND-USE CLASSIFICATION

CHAPTER 4

LANDSAT LAND-USE CLASSIFICATION

INTRODUCTION

PRECEDING PAGE BLANK NOT FILMED

The first pair of Earth Resources Technology Satellites, designated Landsat-1 and -2,* are providing unexcelled opportunities to explore the utility of spacecraft remote sensing data for regional land-use mapping and analysis. Remote sensing is the detection and evaluation of objects without any direct contact. A number of devices collectively known as remote sensors have been developed for collecting and recording the electromagnetic (EM) energy emitted, reflected, and scattered from terrestrial objects at a number of different angles, frequencies, and polarizations.

The level of energy emitted, reflected, and scattered from objects of terrain varies with wavelength throughout the EM spectrum. The spectral signature of an object by which it is both detected and recognized is governed by the amount of energy transmitted to the remote sensor in the EM region in which that sensor operates. Therefore, an identifying spectral signature can often be developed if the spectral energy observed is partitioned into carefully selected bands for analysis. Complementary electromagnetic sensors operating in adjacent spectral regions are commonly used for this purpose (Reference 1).

Whether by man, machine, or interactive man/machine symbiosis, remote sensing data analysis involves pattern recognition, which categorizes phenomena into classes of interest from a set of measurements. The recognition of patterns and the delineation of boundaries on an image are tasks that a photointerpreter does quite well on a single image. Different land-cover and land-use types and their patterns are recognized through the use of numerous contextual and inferential clues. These include the spatial relationships of tone, texture, and size, shape, and geometry of objects, as well as spectral information obtained by given aerial film/filter combinations.

Conversely, a collection of images of varying spectroradiance can be input to digital computers to adaptively improve their numerical pattern recognition/classification efforts using either supervised or unsupervised "learning" algorithms. Consequently, the output from either man or machine is a series of decisions yielding a map on the nature and probabilistic relation of unknown patterns to known or learned patterns.

Although an all-inclusive automated remote sensing data-analysis system will not be possible for some time, digital computers can perform a number of specialized tasks with a large degree of success. Although these processing functions are relatively simple in a machine sense, they are not trivial in a manual interpretation sense. When an interpretation problem demands the simultaneous comparison of multiple statistical patterns, as in multispectral remote sensing, the human analyst's capability is severely limited, whereas the computer does this without difficulty. Some specific justifications for developing automated image processing techniques in remote sensing are:

* Originally designated ERTS-1 and -2.

- The demand for remote sensing inputs to real-time resource management decision-making is rapidly growing.
- The rapid changes caused by population growth and technological development necessitates change detection at short intervals, especially in swiftly urbanizing areas.
- The derivation of usable information from remote sensing data is increasingly needed.
- Weather and resource satellite programs are generating an overwhelming volume of data.
- Short supplies of skilled image-analyst manpower obviate against conventional manual image processing.
- Simple human functions can be duplicated by digital computers to achieve higher input/throughput rates, higher degrees of objectivity, and higher accuracies.
- The liberation of human analysts from routine work through the use of computers makes them available for more challenging tasks.
- Computers can easily manipulate digital input data of high dimensionality, thereby extending the realm of image processing beyond limited human capabilities.

Pattern recognition is concerned with the classification of input data provided in the form of measurements. In a remote sensing context, pattern recognition deals with the classification of pictorial/numeric data describing terrestrial phenomena and consists of five principal steps:

1. Input of the measured patterns (e.g., multispectral imagery)
2. Pattern preprocessing (e.g., calibration and geometric restitution)
3. Feature extraction (e.g., training set selection)
4. Decision/classification (e.g., identification of land use)
5. Classification output (e.g., map display)

Only the first and fifth pattern recognition steps are reasonably discrete. Steps 2, 3, and 4 overlap to a greater or lesser extent, depending on the processing problem and the specific procedures used. However, the sequence of these steps is generally the proper order of the analytical phases.

A dramatic new dimension for terrestrial remote sensing and automatic image analysis was added with the launch of Landsat-1 on July 23, 1972, and was further enhanced with the addition of Landsat-2 on January 23, 1975. Both Landsats are in 920-km (572-mi) altitude Sun-synchronous polar orbits where they complete a full observational cycle of the Earth between 81 degrees north and 81 degrees south latitude every 9 days. The former 18-day repetitive Landsat-1 cycle was halved when Landsat-2 was launched into a 180-degree diametrically opposite orbit. Both satellites carry a similar payload of multispectral sensors

and return-beam vidicon (RBV) cameras for recording visible and solar infrared reflectance data to monitor and inventory agricultural, atmospheric, geologic, hydrologic, and oceanic resources.

Landsat multispectral scanner (MSS) data are simultaneously sensed and recorded in four spectral bands:

- 0.5 to 0.6 μm (micrometer bandwidth of MSS band 4 representing visible green)
- 0.6 to 0.7 μm (MSS-5 representing visible red)
- 0.7 to 0.8 μm (MSS-6 representing solar infrared)
- 0.8 to 1.1 μm (MSS-7 representing solar infrared)

RBV data are recorded in three spectral bands as follows:

- 0.475 to 0.575 μm (visible blue-green interval)
- 0.580 to 0.680 μm (visible red interval)
- 0.690 to 0.830 μm (solar infrared interval)

Unfortunately, Landsat-1's RBV sensor system malfunctioned after only 1 month in orbit, and this device on Landsat-2 has provided only limited coverage.

The Landsat MSS is a line-scanning device that uses an oscillating mirror to continuously scan an east-west track perpendicular to the spacecraft's north-to-south motion (Reference 31). Six lines are simultaneously scanned in each of the four MSS bands for each mirror sweep, with the vehicle's forward movement providing the along-track progression of the six scanning lines. Subsequently, during NASA's image data conversion, this continuous-strip imagery is transformed into framed images with a 10-percent overlap between consecutive frames. As framed, each Landsat image covers a 185- by 185-km (100- by 100-n. mi.) square and consists of 2340 east-west scan lines of data representing 3240 north-south columns of data. This provides 7,581,000 picture elements for the four-band image, which, with four spectral bands, yields more than 30 million data values per 34,225-km^2 (10,000-n. mi.^{2}) image. Landsat images are also available as photographic prints and transparencies.

Each digital Landsat image is disseminated as a set of four 730-m (2400-ft) computer-compatible tapes (CCT's) that segment the scene into four vertical sections of 810 columns of data 46 km (25 n. mi. wide) and 2340 scan lines of data representing 185 km (100 n. mi.) for all four image channels or spectral bands. The analysis of Landsat imagery in this study was extensively structured around machine classification of this digital MSS data. Little direct photointerpretation of the photographic form of Landsat imagery was performed, except in a support role to the computer analysis activities, such as supervised training-site selection and general geographic reference.

OVERLAYING LANDSAT IMAGES ON THE LANDSCAPE MODEL

Computer analysis of remote sensing imagery is symbiotic with the process of landscape modeling described previously. It provides the important current and past land-cover inputs to the landscape model. The accuracy of the computer interpretation of the remote sensing imagery can then be improved by including landscape variables, such as topographic elevation. Combining the available remote sensing imagery with map information in the landscape model provides a basis for improvements in both activities. Coincident spatially registered overlays of readily available map information on the Landsat multispectral imagery provides a basis for improving the accuracy of machine-interpreted land-use or land-cover maps. These improved automatically interpreted maps of land use or land cover also "feed back" directly into the landscape model to provide a timely measure of past and present dynamic tendencies for change in the land use or land cover.

Review of Image Processing

Image preprocessing may be defined as one or more transformations used to sample, calibrate, restructure, or reformat the basic analog/digital scene so that it acquires new properties that facilitate subsequent man or machine interpretation. These transformations can be either image-enhancement or image-restoration operations. The former emphasizes certain features of the image's radiometric content; the latter compensates for image-degradation factors. Ten general sources of image degradation and attendant factors have been identified as follows (Reference 32):

- Illumination (luminance, terrain features, and viewing geometry)
- Terrain (geometry, height profile, Earth curvature, and viewing perspective)
- Atmosphere (absorption, refraction, and background illuminance)
- Spacecraft motion (orbit rates, attitude, and attitude rates)
- Sensor optics (lens, aberrations, mean time before failure, and boresighting)
- Sensor electronics (nonlinear sweep, scene-induced distortions, and detector response)
- Spacecraft recording (signal-to-noise ratio (SNR), nonlinear tape dynamics, and timing signals)
- Data links (SNR, transmitter/receiver, modulation, and signal noise)
- Ground recording (SNR, nonlinear tape dynamics, and timing signals)
- Image processing (recording/playback, storage, and image transformation)

Image transformation operations may be either analog or digital. Although analog operations often achieve speed-increasing parallelism, they have a limited number of processing options and tend to sacrifice accuracy. Digital preprocessing, on the other hand, has advantages in near-total flexibility and higher subsequent accuracy but is usually more time-consuming as currently executed on conventional serial-processing computers.

Image smoothing and image sharpening are two local image-enhancing operations. The former reduces the high-frequency content of an image, thus eliminating noise. Conversely, the latter suppresses low frequencies for edge or boundary enhancement. Digital, electronic analog, electro-optical, noncoherent optical, and photographic media are all used to implement these operations in the spatial domain by convolving the image-enhancing function with the untransformed scene. The frequency domain can also be used for spatial filtering operations by using digital Fourier transforms to generate the image's frequency spectrum and then reconstructing an image after selective suppression of undesired frequency spectrum components. These concepts of image-enhancing operations can be generalized to any form of bandpass filtering to include intermediate frequencies as well.

Image-restoring point operations seek to remove radiometric errors. Analog or digital corrections can be applied if the errors are systematic and the causative factors, such as luminance, terrain features, and viewing geometry are known (Reference 32).

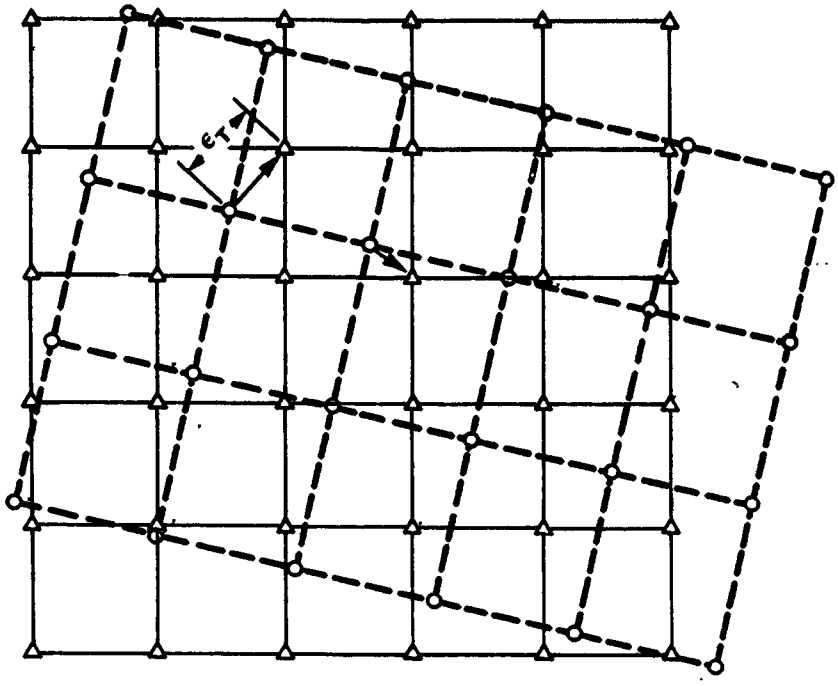
Although these preprocessing techniques are effective in reducing the effects of image radiometric and degradation factors, they were not used because the principal thrust of this study was to test a specific hypothesis. However, to the extent that they could potentially improve land-use classification results, these techniques should be kept in mind for future application.

Preprocessing for Geometric Rectification

The primary image preprocessing used in this study was geometric image transformation whereby a set of picture elements collected on a geometrically distorted grid were converted to samples on another precisely rectified grid (figure 47). This method was necessary because the digital CCT's of the Landsat images were not geometrically corrected, and their utility for mensuration, precise location, and spatial/temporal registration was therefore limited. The correction of geometric errors was imperative because this study required Landsat digital imagery that spatially matched ancillary landscape data planes derived from multivariate maps of the Denver Metropolitan Area.

The following major sources of Landsat geometric errors have been cited (Reference 33):

- Rectangular picture element
- Spacecraft altitude variations
- Spacecraft attitude variations
- Earth rotation skew
- Orbital velocity changes
- Line-scanner time skew



○ Original Landsat Data Input X-Grid
△ New Transformed Output Y-Grid

FIGURE 47. GEOMETRIC RELATIONSHIP OF ORIGINAL AND NEAREST-NEIGHBOR TRANSFORMED LANDSAT IMAGE CELLS. The new or output grid represents a counter clockwise rotation and rescaling of the original input grid. ϵ_T is the total Euclidean error distance introduced by the nearest-neighbor resampling technique (from Reference 34).

- Nonlinear line-scanner sweep
- Line-scanner viewing geometry
- Non-north-oriented image.

Resampling techniques for geometric image transformation include the bilinear and cubic convolution, nearest neighbor (NN), and various truncated versions of sine x/x. It has been shown (Reference 32) that the cubic convolution function gave substantially higher-quality images for typically encountered scenes. Nonetheless, for economy in the use of computer resources, it was decided to proceed with an NN resampling of the unrectified digital scene. Little geometric error is introduced by using this process when the scene content changes slowly or when the interpolated value lies close to an initial X-grid point (figure 47). Further, errors induced by NN resampling should be regarded as residual geometric errors rather than as intensity errors. Lastly, initial efforts at developing an NN resampling computer program had already produced a rudimentary but workable and economical piece of software.

The NN resampling program was improved to accommodate the full four-channel, 600- by 600-element image needed to cover the Greater Denver Metropolitan Area. These alterations also provided for square resampling units of any specified acreage or rectangular sampling units representing the correct geometry for both six

or eight vertical lines per inch displays at any given scale (figure 48). The improved software (ROTATE2) operated in an "open-loop" geometric correction mode without feedback from any ground control points. However, this initial rectification (figure 49) of three corrections (scale adjustment, Earth rotation skew, and image rotation) was further improved with the addition of the nonlinear line-scanner sweep correction.

Landsat image height and width dimensions we adjusted to match pictorial detail on 1:24,000-scale rectified displays to the corresponding control mosaic of four by four 7.5-minute USGS topographic map sheets (figure 5). The rectilinear road network appeared dark on MSS-7 solar infrared graymaps, and was the chief geometric registration factor. Major roads were located over the entire base-map mosaic to appraise the image height and width parameters. Spot checks of the geometric fit on the extensive network of lakes, reservoirs, and parks were also made on these distinctive features. A spatial registration between the rotated MSS-7 graymaps and the base-map mosaic was finally achieved, matching the desired number of picture elements to within one cell over the 39-km (24-mi) image height and width.

Each of the four basis MSS bands was rectified for square 0.4-ha (1.111-acre) picture elements yielding nine Landsat-1 cells in a three-by-three array exactly equal to the 4-ha (10-acre) landscape data cell. This was done one band at a time, because a somewhat larger scene was transformed to ensure that the desired study area was enclosed, and the entire core capacity of the computer was occupied in the 600- by 600-element rectification. Hence, four MSS single-channel files, 615 lines by 654 columns, were generated, subsequently manipulated to produce a composite overlay of the four-band file, and then trimmed to yield the exact study area of 576 lines by 576 columns (figures 50 and 51).

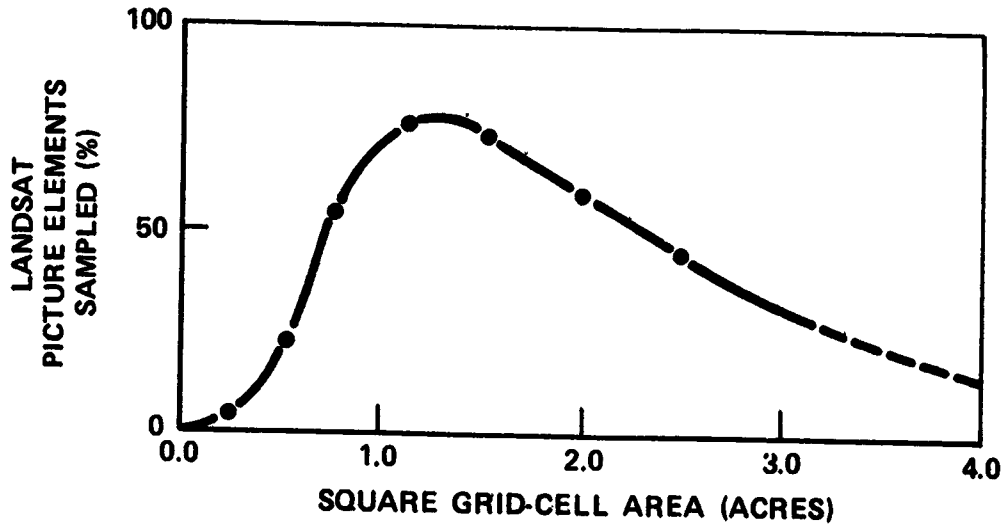
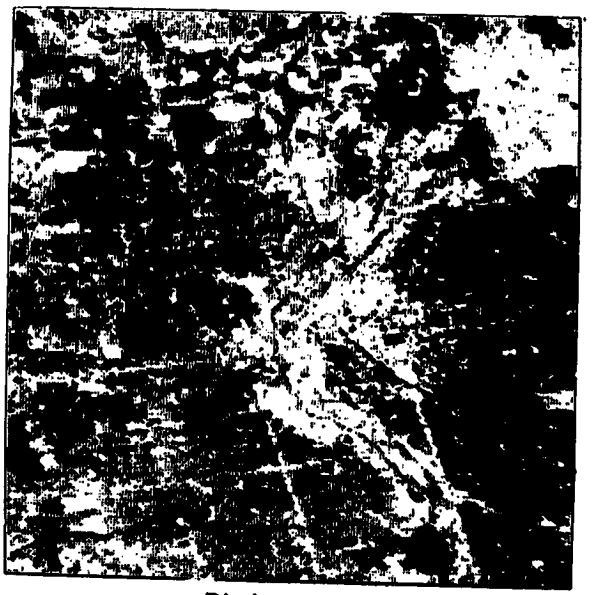


FIGURE 48. RESAMPLING EFFICIENCIES OF THE IMAGE GEOMETRIC TRANSFORMATION. The application of the nearest-neighbor approach in resampling to a square grid transfers percentages of the samples shown from the input grid (X-grid) to the output grid (Y-grid) as shown in figure 47.

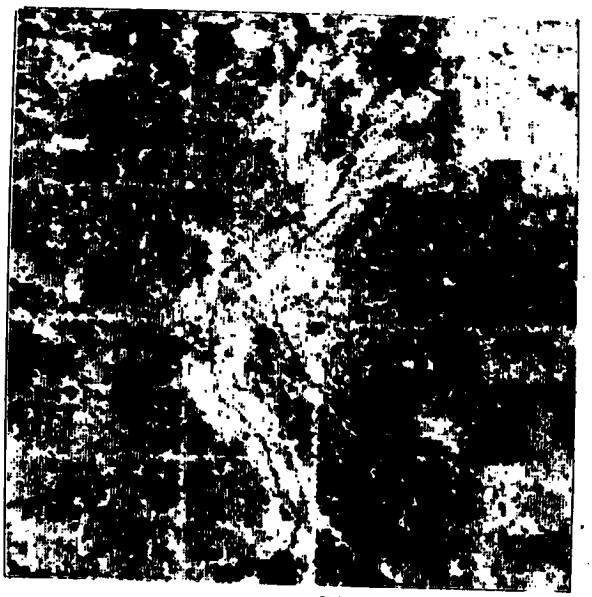


Display scale nominally 1:173,600

(a) Unrectified red image
(MSS band 5 = 0.6 to 0.7 μm).

Display scale nominally 1:173,600

(b) Unrectified solar IR image
(MSS band 7 = 0.8 to 1.1 μm).



Display scale 1:173,600

(c) Rectified red image
(MSS band 5 = 0.6 to 0.7 μm).

Display scale 1:173,000

(d) Rectified solar IR image
(MSS band 7 = 0.8 to 1.1 μm).

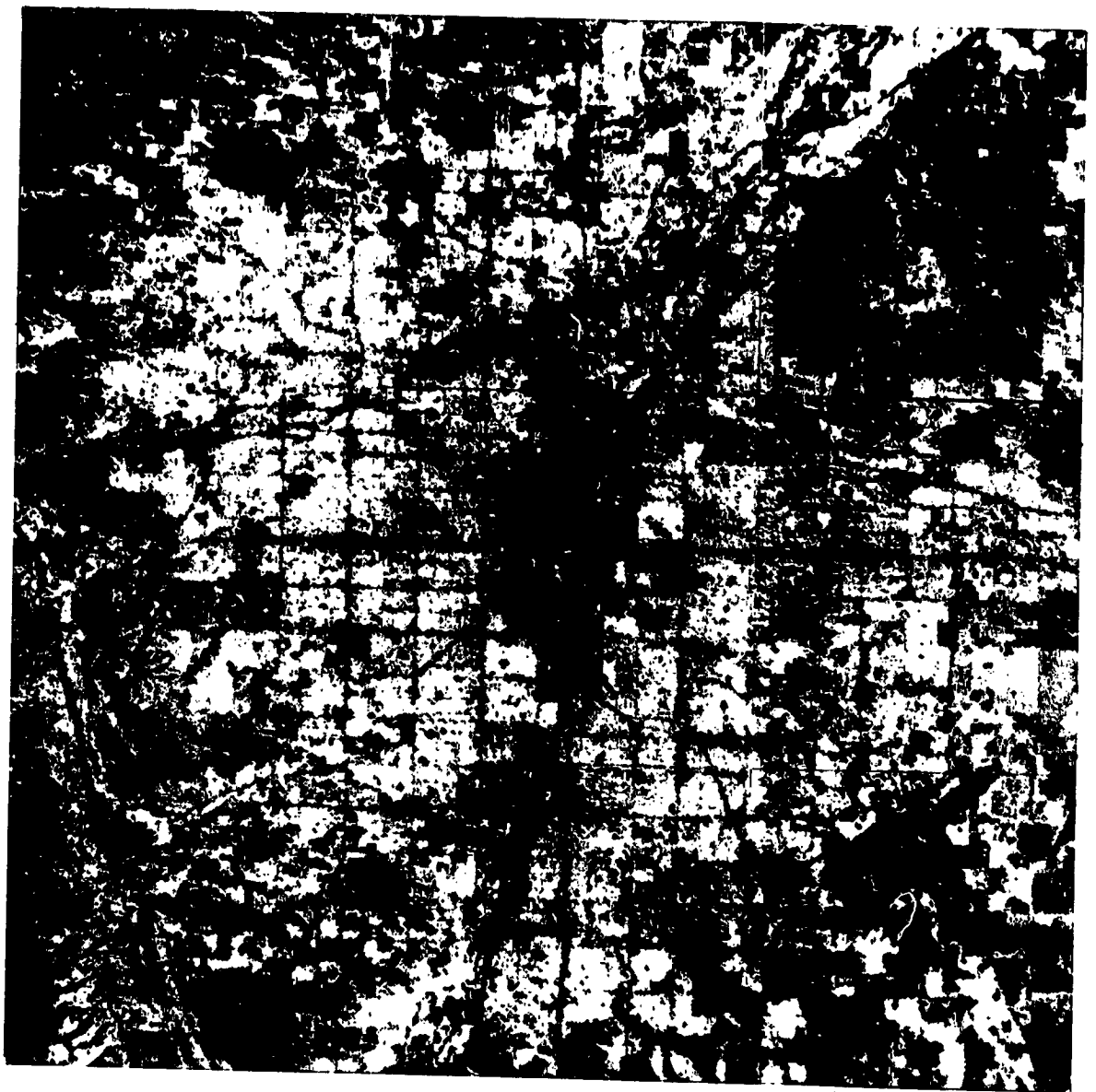
FIGURE 49. COMPARATIVE DISPLAY OF THE ORIGINAL UNRECTIFIED AND THE TRANSFORMED (RECTIFIED/RESAMPLED) MULTISPECTRAL LANDSAT IMAGERY OF THE DENVER METROPOLITAN AREA. The raw or unrectified Landsat picture elements, (a) and (b), are 192- by 259-foot inclined rectangles. The resampled picture elements are rectified 210-foot north-south squares, (c) and (d), that are displayed from computer-compatible tapes as microfilm graymaps with a cellular resolution of exactly nine picture elements per 10 acres, or approximately 1.1 acres per picture element. The display scale of (c) and (d) is nominally 1:173,600.

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

Display scale 1:250,000

FIGURE 50. LANDSAT-1 MSS BAND 5 VISIBLE RED IMAGE OF THE DENVER METROPOLITAN AREA. Graymap of the August 15, 1973 image showing 576 by 576 image cells exactly as they were overlaid from the Landsat computer-compatible tapes to yield 1.111 acre per square cell. A square three-by-three array of these image cells of 10 acres exactly overlaid the square 10-acre cells of the landscape data planes.



Display scale 1:250,000

FIGURE 51. LANDSAT-1 MSS BAND 7 SOLAR INFRARED IMAGE OF THE DENVER METROPOLITAN AREA. Graymap of the August 15, 1973 image showing 576 by 576 image cells exactly as they were overlaid from the Landsat computer-compatible tapes to yield 1.111 acre per square cell. A square three-by-three array of these image cells of 10 acres exactly overlaid the square 10-acre cells of the landscape data planes.

ORIGINAL PAGE IS
OF POOR QUALITY

Preprocessing to Form MSS-Band Ratios

Ratioing has been advocated as a means of effectively reducing random fluctuations of MSS reflectance values caused by source variations and changing atmospheric conditions (References 34 and 35). Ratioing is simply the division of the digital radiance value of one MSS band by that of another on a cell-to-cell basis. A ratio of the near-infrared and chlorophyll absorption bands proved to be well correlated with the functioning green biomass in grasslands (Reference 36). The ratio of MSS band 7 to band 5 enhanced the effect of biomass changes (Reference 37). The main advantage of band ratios in vegetation classification is an improved signal-to-noise ratio (Reference 37).

Twelve possible ratios can be computed for the four primary MSS bands taken two at a time; however, six of these are merely the inverse of the other six. Because the spatial variation in the ratio of any two spectral bands is the same as the ratio of the inverse of the two bands, except in an inverse sense, the inverse ratios provided no unique differences and were therefore omitted. Six ratios between the four original MSS bands were thus computed and interspliced back into the multichannel image/landscape variable file using the LMS programs. Each picture element in this file was represented by ten image values, one each for MSS bands 4, 5, 6, and 7 and for ratios 5/4, 6/4, 7/4, 5/6, 7/5, and 7/6.

Merging of Landsat and Landscape Data Planes

Digital Landsat imagery was geometrically corrected and resampled to overlay the landscape model with a square cell resolution of 0.4-ha (1.111 acre) that nested a three-by-three array of Landsat cells exactly into 4-ha (10 acres). It was next necessary to transform each of the 36,864 data cells in each of the 34 landscape planes, the assembly of which was described in Chapter 2, into a three-by-three array of 0.4-ha (1.111-acre) cells with the original 4-ha (10-acre) value duplicated in all nine cells. This zoom-like transformation preserved the spatial registration because the Landsat image data planes were three times larger in height/width elements than the landscape data planes—576 picture elements to 192 landscape cells. This expansion function was performed on the 34 landscape variables with an auxiliary program (ANCILLARY). The resultant expanded or zoomed landscape planes appeared to be somewhat blocky in comparison to their original form (figure 52), but they were now in perfect one-to-one registration with the Landsat imagery at the 0.4-ha (1.111-acre) resolution representing 331,776 cells. Subsequently, the 10-variable Landsat imagery and the zoomed or expanded 34-variable landscape data were merged into a composite 44-variable data file (figure 53). These 44 variables were embedded in a standardized packed binary format of the landscape-modeling/image-processing system for analysis.

Preprocessing to Remove Terrain Effects

Normalizing is proposed as a preprocessing technique for specifically removing terrain effects. Normalizing is simply the division of the digital radiance of an MSS band by the computed Landsat overflight insolation on a cell-to-cell basis. Although normalizing is mathematically similar to ratioing, the denominator is an ancillary variable, and the resultant quotient is also an ancillary landscape variable.

Identical surface-cover materials may have been imaged at different radiance values in a given MSS band because they occurred on varying terrain (i.e., differing slope, aspect, or topographically shaded areas).

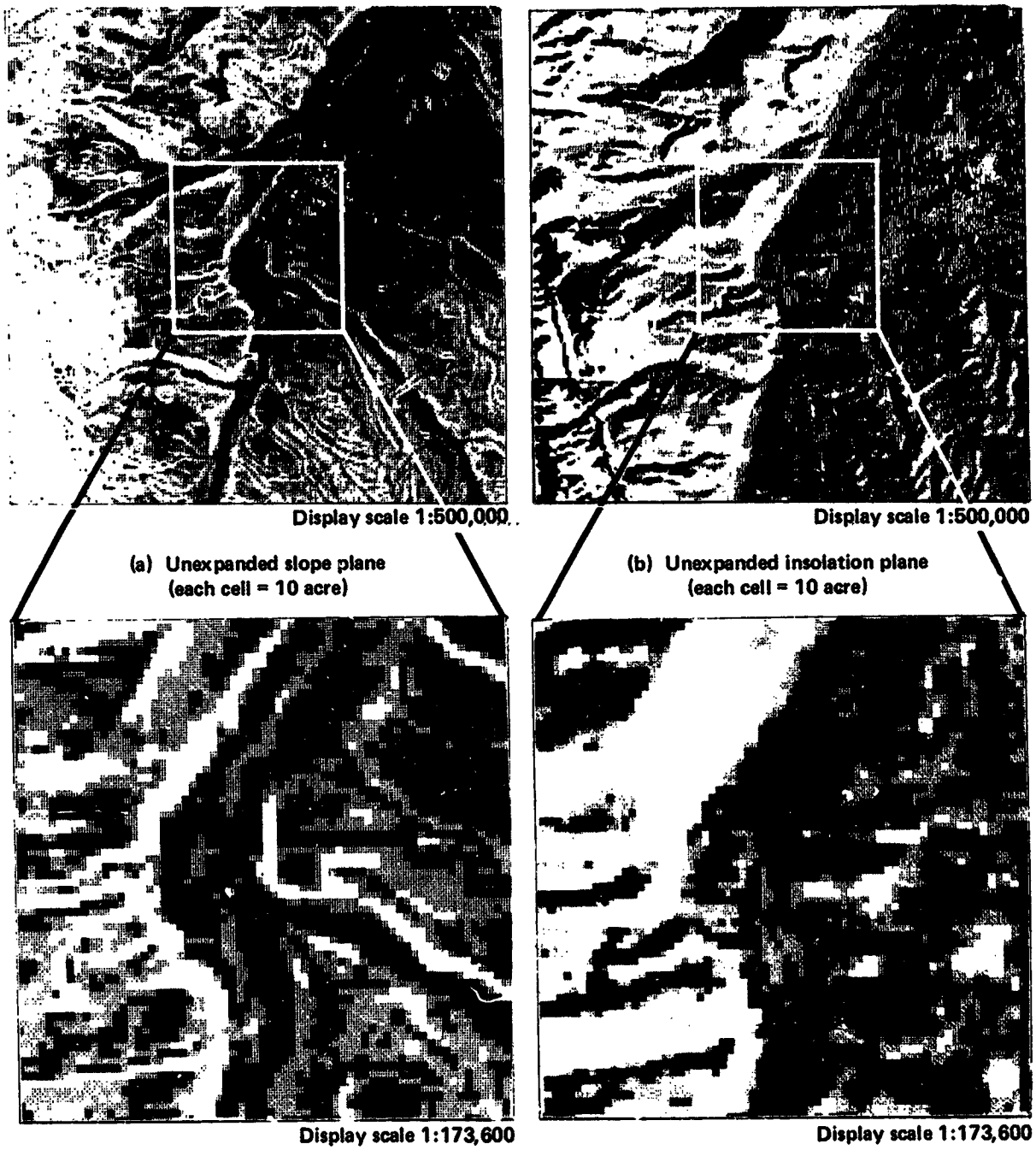
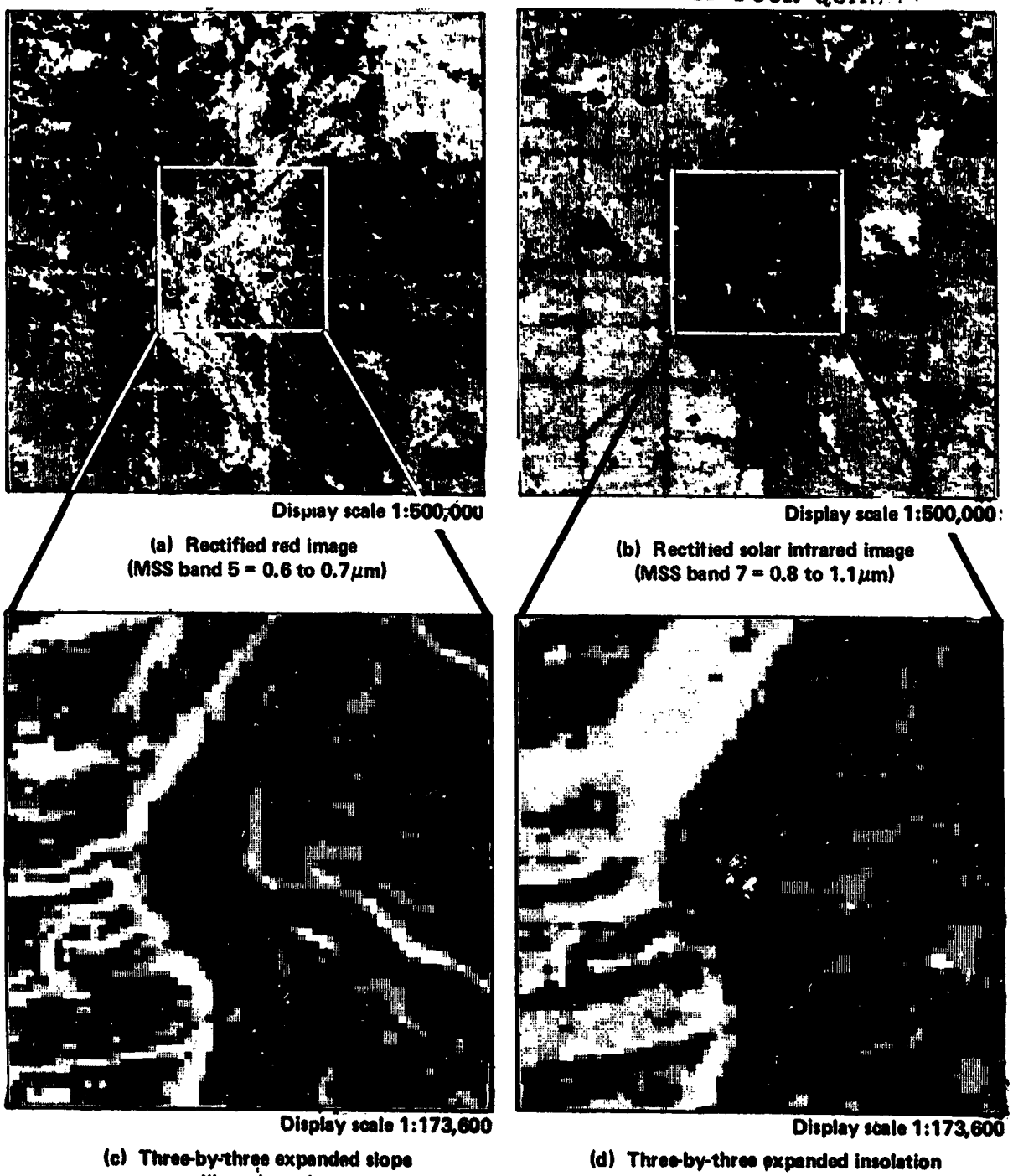


FIGURE 52. DISPLAY OF THE EFFECT OF THE EXPANSION OF THE ANCILLARY MAP DATA PLANES IN THE LANDSCAPE MODEL FOR OVERLAY ON LANDSAT IMAGE DATA PLANES. Nonimage map data cells representing 10 acres each were duplicated in a three-by-three picture-element array to exactly spatially overlay the higher-resolution (1.111-acre) rectified/resampled Landsat picture elements (figure 50). Unexpanded ancillary data displays, (a) and (b), scale 1:500,000. Expanded ancillary data displays, (c) and (d), show the expanded center one-ninth of (a) and (b) at a nominal scale of 1:173,600.

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY



Display scale 1:500,000

(a) Rectified red image
(MSS band 5 = 0.6 to 0.7 μm)

Display scale 1:500,000

(b) Rectified solar infrared image
(MSS band 7 = 0.8 to 1.1 μm)

Display scale 1:173,600

(c) Three-by-three expanded slope ancillary data plane

Display scale 1:173,600

(d) Three-by-three expanded insolation ancillary data plane

FIGURE 53. DISPLAY ILLUSTRATING THE OVERLAY OF SELECTED PHYSIOGRAPHIC MAP DATA ON THE LANDSAT MULTISPECTRAL IMAGERY OF THE DENVER METROPOLITAN AREA. Thirty-eight ancillary land-use, physiographic, socioeconomic, transportation, and four of MSS/insolation ratio data planes were thus combined with the basic four Landsat multispectral channels and six image transformations (i.e., channel ratios) into a composite 48-channel Landsat image/ancillary data file for automated image processing.

If these illumination effects are proportionally constant for both the MSS band and insolation value, the ratio of the two parameters will negate the terrain effect because it appears in both the numerator and denominator.

The LMS programs (Appendix B) were used to create (TRANSF2) four MSS/insolation normalized bands and to intersplice (COMBINE) them back into the composite image/landscape data file as the last four ancillary landscape variables. Therefore, the complete 48-variable data file was composed of the four original MSS bands, six MSS band ratios, and 38 land-use, physiographic, socioeconomic, transportation access, and MSS/insolation normalized ancillary landscape variables (figure 3).

MULTIVARIATE CLASSIFICATION WITH RECTANGULAR TRAINING SETS

Feature selection with multispectral scanner data commonly uses training sets. A training set is a subsample whose identification is known to an acceptable level of accuracy. The training set is used to generate subpopulation statistics for efficiently implementing decision rules in the following classification phase. Overall utility of the feature extraction process is contingent on the quality of samples selected to serve as a training set (i.e., sample size, spatial distribution, and efficiency). The basic problem is the identification and selection of samples with the correct probability distributions in n-dimensional space, where n is the number of variables of the multispectral imagery. These probability distributions are then used to construct decision rules that can be used to represent the prediction sample set.

A number of data displays are used with MSS data to facilitate feature selection, such as histograms for each class and data plane or spectral overlay, as well as correlation and covariance matrices between classes. Parametric discriminants, such as Bayesian probabilities or maximum likelihood functions, are usually used with training-set data to develop the classification algorithms. Here, linear-discriminant analysis was used as in Chapter 3 to reduce a multivariate problem to a univariate situation.

Others have tested many alternative procedures because of the great amount of computational time required for the maximum-likelihood ratio. For example, it is possible to reduce the number of data through sampling and to save time with the decision rule itself (Reference 38). The decision rule commonly used is founded on two assumptions: (1) that the data are Gaussian or normally distributed; and (2) that the training data for each class adequately represents the entire class. Under these assumptions, the maximum-likelihood decision rule becomes a quadratic rule. This rule requires many multiplications for each decision, especially when many channels of data and many desired identification classes are used. The linear-decision rule also compared favorably in accuracy and was 50 times faster than the maximum-likelihood function (Reference 38).

Very fast nonparametric procedures have been developed (Reference 39). These linear-discriminant functions achieved satisfactory results with multiclass/multispectral data as long as the order of class separability was observed during the classification process. Their algorithm took 5 minutes for one area, as compared to 70 minutes for six channels and 5 hours for 12 channels with the maximum-likelihood algorithm. Economies like these will be important in making the computer more available (and acceptable) for future machine processing of multispectral imagery.

Other nonparametric classification studies that either cited increased speed or satisfactory predictions used a composite sequential clustering algorithm (Reference 40), an elliptical boundary condition model (Reference 41), and a lookup table procedure that was 30 times faster than the maximum-likelihood ratio and generally as accurate (Reference 42).

Initial feature-extraction activities in this study included careful selection of 72 rectangular training sets totaling 2,413 picture elements, or three rectangles each to represent the 24 second- and third-order USGS land-use classes to be mapped. Geometrically rectified and resampled MSS-5 and -7 graymaps overlaid on the 1:24,000-scale topographic/land-use maps facilitated the entire process of selection and identification. Stepwise linear-discriminant analysis was again used to test machine recognition and mapping of the 24 land-use categories.

The initial or baseline rectangular training-set classification used only the original four MSS channels and no other information to correctly reclassify 65.2 percent of the training sample back into their correct land use (table 13). This percentage (FOM) will be referred to hereafter as training-set accuracy. An overall FOM or training-set accuracy is defined as the total number of cells correctly classified, divided by the total number of cells, times 100 to obtain a percentage (Reference 34). A second baseline classification used the original four MSS bands plus the six transformed image channels consisting of the ratios of the basic four bands to yield an improvement of 2.1 percentage points to an overall average training-set accuracy of 67.3 percent. This indicated that the six ratios of the basic four spectral bands contributed little to this particular land-use classification (table 14).

Next, the 31 landscape data planes, exclusive of any categorical-type land-use data planes, were included with the ten image bands, raising the accuracy of the correct training-set identification to 99.7 percent for the 24 classes (table 15). Therefore, 99.7 percent of the 2,413 test image cells could be correctly assigned to their known land-use category. The random assignment of these cells to the 24 land-use categories would have yielded 1/n = 24 classes times 100 percent = 4.1-percent accuracy. Thus, this represented a significant increase in the accuracy of automatically interpreted Landsat imagery with the symbiotic inclusion of spatially overlaid ancillary map information. Only 40 variables were included in the final model because total year-round housing units failed the stepwise tolerance test for inclusion in the function at step 23.

TABLE 13

TRAINING-SET ACCURACY OF AUTOMATED INTERPRETATION OF THE FOUR ORIGINAL MSS BANDS OF A LANDSAT IMAGE (August 15, 1973) OF THE DENVER METROPOLITAN AREA. This table is based on identifying 24 land uses (table 1) with three rectangular training sets per land use that also provided the test sample of 2,413 picture elements. The Landsat-1 image variables were added in a free stepwise fashion and were classified using linear-discriminant analysis.

Step Number	Landsat Variable Entered (free)	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
1	MSS-7 (solar infrared)	1,208	50.06	8.74	138.22	138.22	1,640.13
2	MSS-5 (visible red)	1,518	62.91	9.68	156.82	147.99	562.87
3	MSS-4 (visible green)	1,585	65.69	10.48	151.24	149.17	115.96
4	MSS-6 (solar IR)	1,573	65.19	24.78	63.48	109.61	10.67

*C-P = central processor.

TABLE 14

TRAINING-SET ACCURACY OF AUTOMATED INTERPRETATION OF THE FOUR ORIGINAL LANDSAT MSS BANDS AND SIX RATIOS OF A SINGLE-DATE LANDSAT IMAGE (August 15, 1973) OF THE DENVER METROPOLITAN AREA. This table is based on identifying 24 land uses (table 1) with three rectangular training sets per land use that also provided the test sample of 2,413 picture elements. The Landsat-1 image variables were added in a free stepwise fashion and were classified using linear-discriminant analysis.

Step Number	Landsat Variable Entered (free)	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
1	MSS-7 (solar infrared)	1,208	50.06	11.49	105.13	105.13	1,640.13
2	MSS-5/MSS-4 ratio	1,479	61.29	12.12	121.93	113.76	573.23
3	MSS-7/MSS-6 ratio	1,522	63.08	12.83	118.63	115.47	227.78
4	MSS-5 (visible red)	1,552	64.32	13.63	113.87	115.04	93.89
5	MSS-4 (visible green)	1,607	66.60	14.42	111.44	114.23	192.63
6	MSS-6/MSS-4 ratio	1,621	67.18	15.30	105.95	112.64	108.88
7	MSS-7/MSS-4 ratio	1,623	67.26	16.22	100.06	110.19	54.06
8	MSS-5/MSS-6 ratio	1,615	66.93	16.93	95.39	108.25	28.90
9	MSS-6 (solar infrared)	1,624	67.30	17.45	93.07	106.22	10.15
10	MSS-7/MSS-5 ratio	1,624	67.30	34.30	47.35	93.96	6.05

*C-P = central processor.

The stepwise linear-discriminant analysis algorithm automatically added each landscape variable in the order in which it added the most to the land-use classification accuracy achieved at that step. Clearly, the addition of many of the less sensitive landscape variables did not measurably increase the final accuracy achieved, but significantly increased the total cost and decreased the cost-efficiency of the test classification (figure 54). These accuracy and cost-performance values were used as the basis for selecting three optimal MSS spectral bands and four landscape variables from the 41 initial variables for a final test of a more efficient and economic procedure. These seven optimal variables served to correctly assign all 24 land uses with an average training-set accuracy of 96.6 percent (table 16), as compared with the 99.7-percent accuracy achieved with 40 variables at almost two times the cost in central-processor time.

The utility of the landscape data planes overlaid on the Landsat spectral data base was evident in all classification tests. However, a potential time bias problem was inherent in most of the ancillary data planes. Specifically, it was felt that the time-related cultural data (i.e., 1970 Census data) might unduly influence the discriminant functions toward the historical date represented by these landscape variables, particularly in view of their high F-values. Therefore, it was considered appropriate to statistically force in the current (1973) Landsat imagery variables in their free order before the totally free stepwise entry of ancillary variables to reduce the ancillary data time bias and to generate a more representative land-use map.

TABLE 15

IMPROVEMENT IN THE TRAINING-SET ACCURACY OF AUTOMATED INTERPRETATION OF A SINGLE-DATE LANDSAT-1 IMAGE OF THE DENVER METROPOLITAN AREA BY ADDING 30 ANCILLARY VARIABLES. This table is based on identifying 24 land uses (table 1) with three rectangular training sets per land use that also provided the test sample of 2,413 picture elements. The Landsat-1 image variables were forced in the predetermined order (table 14) and the landscape variables were added in a free stepwise fashion and were classified using linear-discriminant analysis.

Step Number	Variable Entered	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
	Landsat (forced)						
1	MSS-7 (solar infrared)	1,208	50.06	24.22	49.88	49.88	1,640.13
2	MSS-7/MSS-4 ratio	1,479	61.29	24.98	59.21	54.61	573.23
3	MSS-7/MSS-6 ratio	1,522	63.08	25.76	59.08	56.15	227.78
4	MSS-5 (visible red)	1,552	64.32	26.51	58.54	56.78	93.89
5	MSS-4 (visible green)	1,607	66.60	27.27	58.93	57.23	192.63
6	MSS-6/MSS-4 ratio	1,621	67.18	28.01	57.87	57.35	108.88
7	MSS-7/MSS-4 ratio	1,623	67.26	29.00	55.97	57.13	54.06
8	MSS-5/MSS-6 ratio	1,615	66.93	29.82	54.16	56.72	28.90
9	MSS-6 (solar infrared)	1,624	67.30	30.36	53.49	56.32	10.15
10	MSS-7/MSS-5 ratio	1,624	67.30	31.14	52.15	55.85	6.05
	Landscape (free)						
11	Total census-tract acreage	2,036	84.38	31.72	64.19	56.71	133,372.71
12	Built-up urban area MD†	2,162	89.60	32.49	66.54	57.64	1,294.12
13	Topographic elevation	2,278	94.41	33.17	68.68	58.62	1,060.76
14	Freeway interchange MD	2,343	97.10	34.00	68.91	59.48	349.27
15	Freeway MD	2,353	97.51	34.73	67.75	60.13	626.06
16	Median housing-unit rent	2,349	97.35	35.51	66.15	60.57	349.55
17	Composite major-road MD	2,353	97.51	36.27	64.87	60.88	221.08
18	Median housing-unit value	2,371	98.26	36.98	64.12	61.09	224.90
19	Total vacant housing units	2,380	98.63	37.75	63.05	61.22	182.63
20	Average number of cars per family	2,376	98.47	38.56	61.62	61.24	119.81
21	Average number of vacant housing units per acre	2,387	98.92	39.33	60.69	61.21	110.39
22	Population density per acre	2,382	98.72	39.86	59.76	61.13	97.27
23	Total population	2,383	98.76	40.61	58.68	61.00	96.18
24	1969 mean family income	2,379	98.59	41.34	57.55	60.82	163.80
25	Total two-car families	2,384	98.80	42.08	56.65	60.60	135.20
26	Total three-car families	2,390	99.05	42.77	55.88	60.37	102.20
27	Total year-round housing units	2,396	99.30	45.73	52.39	59.98	105.41
28	Average number of year-round housing units per acre	2,403	99.59	45.17	53.20	59.66	97.37
29	Total families	2,398	99.38	45.10	53.17	59.37	102.18
30	Total one-car families	2,395	99.25	45.77	52.33	59.06	142.24
31	Average number of families per acre	2,395	99.25	46.84	51.13	58.73	91.83
32	Landsat-1 image insolation	2,399	99.42	47.55	50.45	58.39	93.71
33	Topographic slope	2,405	99.67	48.10	50.00	58.05	100.84
34	Surficial geology	2,405	99.67	48.69	49.39	57.71	38.35
35	Composite minor-road MD	2,405	99.67	49.60	48.49	57.36	35.28
36	Landsat-1 MSS-5/insolation ratio	2,405	99.67	50.31	47.80	57.00	9.45
37	Landsat-1 MSS-4 insolation ratio	2,405	99.67	51.01	47.15	56.64	24.72
38	Landsat-1 MSS-7 insolation ratio	2,405	99.67	51.77	46.46	56.28	8.36
39	Topographic aspect	2,405	99.67	52.54	45.77	55.91	7.43
40	Landsat-1 MSS-6 insolation ratio	2,405	99.67	53.35	45.08	55.54	3.96

*C-P = central processor.

†MD = minimum distance.

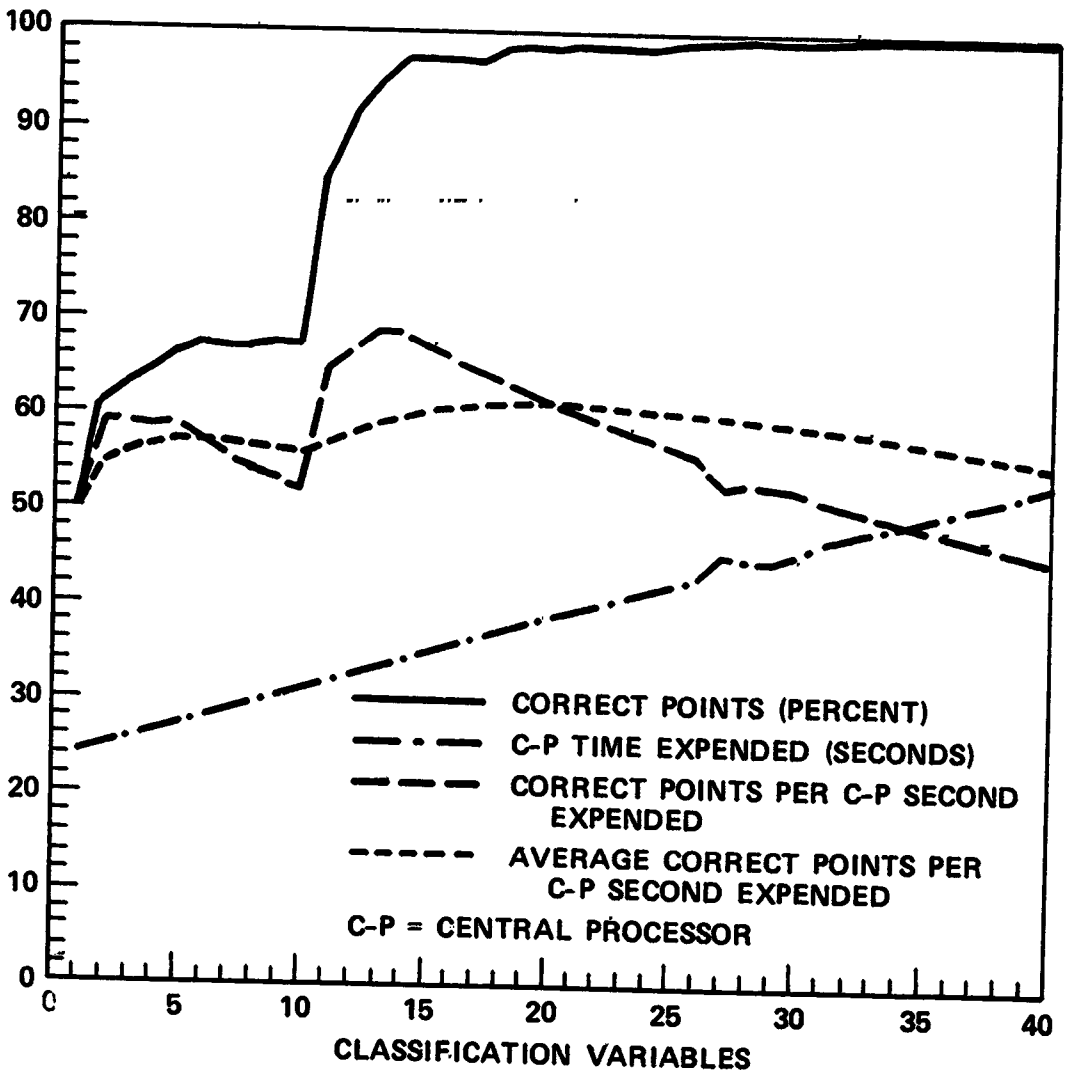


FIGURE 54. TRAINING-SET ACCURACIES AND COSTS OF PROCESSING SINGLE-DATE LANDSAT IMAGERY OF THE DENVER METROPOLITAN AREA WITH 30 ANCILLARY LANDSCAPE VARIABLES. This figure is based on identifying 24 land uses (table 1) with three rectangular training sets per land use that also provided the test sample of 2,413 picture elements. The Landsat-1 image variables representing the first ten variables were forced in the predetermined order (table 14), and the ancillary variables were next added in a free stepwise fashion and were classified using linear-discriminant analysis. The variable numbers shown coincide with step numbers and data-plane identities in table 15.

TABLE 16

HIGH TRAINING-SET ACCURACIES AND ECONOMY ACHIEVED BY AUTOMATED INTERPRETATION OF A SELECTION OF THREE MSS BANDS OF A SINGLE-DATE LANDSAT IMAGE (August 16, 1973) OF THE DENVER METROPOLITAN AREA WITH OVERLAYS OF FOUR LANDSCAPE VARIABLES. This table is based on identifying 24 land uses (table 1) with three rectangular training sets per land use that also provided the test sample of 2,413 picture elements. The Landsat variables were forced in the predetermined order (table 13), and the ancillary landscape variables were added in a free stepwise fashion and were classified using linear-discriminant analysis.

Step Number	Variable Entered	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
	Landsat (forced)						
1	MSS-7 (solar infrared)	1,208	50.06	10.17	118.78	118.78	1,640.13
2	MSS-5 (visible red)	1,518	62.91	11.04	137.50	128.52	562.87
3	MSS-4 (visible green)	1,585	65.69	11.70	135.47	130.99	115.96
	Landscape (free)						
4	Total census-tract acreage	1,992	82.55	12.32	161.69	132.35	113,029.23
5	Built-up urban area MD†	2,143	88.81	13.19	162.47	146.96	1,332.26
6	Topographic elevation	2,242	92.91	13.77	162.82	149.98	2,007.72
7	Freeway interchange MD	2,332	96.64	27.57	84.58	131.91	356.92

*C-P = central processor.

†MD = minimum distance.

It has been carefully observed that, by the use of the term "training-set accuracy," these classification tests indicated only the rectangular training-set accuracy for the use of Landsat and landscape or ancillary data variables for mapping land use. Further testing was done to determine how accurately these feature extraction procedures extended to the mapping of the entire Denver study site—the "verification" accuracy. The verification test was performed by classifying the 36,864 cells of a larger one-ninth sample image with the discriminant function developed from the 72 rectangular training sets for each variable added in the stepwise fashion previously determined (table 15). This one-ninth training set was generated by rectilinearly sampling every third row and third column for a total sample of 192 rows and 192 columns from the full 576- by 576-picture element composite image. The accuracy of each successive map produced by this function in the step-by-step fashion was then verified on a cell-by-cell basis by comparison with the known 1972-1973 USGS land-use map stored in the landscape model (figure 55). The cost of computing and comparing a total classification map at each step prevented the continuation of this process for all 40 steps. The average image-verification accuracy was a surprisingly low 24.1 percent at the addition of the 28th step (figure 56 and table 17), as compared to the predicted training-set accuracy of 99.6 percent at the same step (table 15). The rectangular training sets, carefully selected to statistically represent the 24 land-use classes by 2,413 points, were clearly inadequate samples to represent the total sampled image of 36,864 points. Apparently, a fortuitous choice of training-set samples gave extremely high training-set accuracies in this restricted pattern space. However, the training sets were statistically insufficient descriptors for classifying the larger image sample and, for that matter, the full image because the one-ninth sample image statistically represented the full image.

This exercise clearly indicated that extreme care must be used in interpreting the many examples of training-set accuracy in the technical literature. Similarly, this caution extends to test-set accuracy when the test fields are selected in much the same way as the training sets and do not represent a truly random sample of the image that will be subsequently classified. Judicious selection of training sets can be used to manipulate the final training-set accuracy to be anywhere from very poor to very good, depending on the desired results. Better methods must be developed for selecting training sets if the most accurate end product is desired, and a true verification test on randomly sampled points must be included with the map produced.

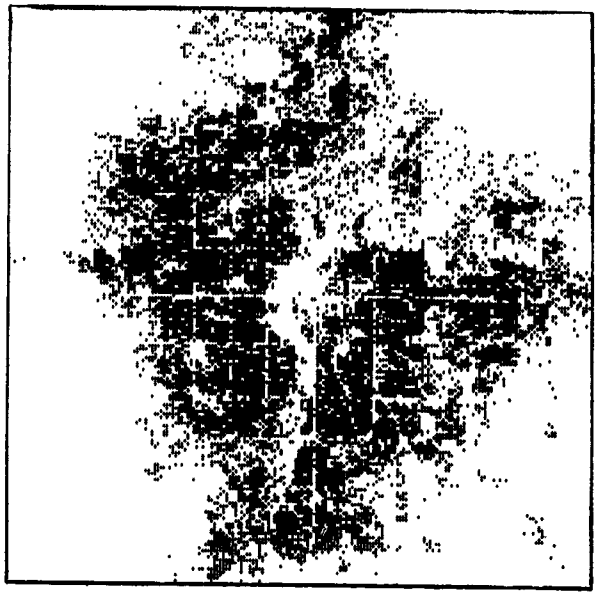
MULTIVARIATE CLASSIFICATION WITH GRID-SAMPLED TRAINING SETS

Demonstration of the potential bias in conventional rectangular training-set selection (i.e., the wide disparity between the feature extraction and the one-ninth sampled image-classification accuracies) led to the development of a self-verifying grid-sampled training-point approach. The search for the essence of the features of the multivariate data set centered on the 1972-1973 USGS land-use data plane that identified the same 24 land-use classes already tested. Proper use of this data plane provided control of training-set selection and gave a representative sampling of all multivariate variables from the entire image for each of the 24 land uses.

At first glance, it may appear that using a map that represented a sample of the final desired output of the classification effort was self-defeating in that the "answer" was necessary for solving the problem. The collection of training-set data for Landsat image processing has invariably been the one step that the ultimate user of the classification maps undertakes; it is not the skilled image interpreter who completes this task. It is just as easy, if not easier, to instruct the user to prepare a sample map of one or more subportions of the total areas to be classified. He can prepare this map with conventional ground and airphoto methods with which he is usually familiar. He can prepare it to represent the type of final product desired, conditioned by what is reasonable to expect from Landsat image classification. These sample maps might be two or three USGS 7.5-minute quadrangles that are representative of the materials that occur in 2600- to 5200-km^2 (1000- to 2000-mi^2) area to be mapped. It is claimed that this procedure is more clearly understood and completed than instructing the user to collect illustrative training sets for each surface-cover type sought. This is especially true if the image analysts are not certain of the exact nature of the training sets desired or their subsequent impact on the classification accuracy of the final map products sought.

The test of the hypothesis employed a grid-sampled point training set that was created by resampling every third row and third column of the one-ninth sample image. This yielded a one-ninth times one-ninth (1/81-sampled) image of 4,100 training points of known land use by reference to the overlay of 1972-1973 USGS land use. This systematic point-sampling process represented an efficient statistical distillation of the data set while providing uniform coverage of the study area. Further, the systematic sample chose land-use sample points in proportion to their representation in the entire image. Hence, the probability of the selection of points was proportional to land-use type frequency. This assumed, of course, that the user was proportionally interested in all final classes roughly in proportion to the amount of their occurrence in the sample map. Specific interest in locating a given surface cover or covers in the context of all other materials constituting a background would require the reexamination of this sampling approach.

Machine recognition and mapping of the 24 land-use classes was again tested by stepwise linear-discriminant analysis. The 4,100 grid-sampled points were grouped into a training set to represent the 24 of 1972-1973 USGS mapped land uses. The first classification with this training data employed the four basic MSS bands with no other information. The 4,100 samples grouped in the training set were used to compute the discriminant function in a stepwise fashion. The samples were then test-classified using the same function and



ORIGINAL PAGE IS OF POOR QUALITY

FIGURE 55. VERIFICATION OF THE MAP ACCURACY FOR THE DENVER METROPOLITAN AREA FOR ALL 1972-1973 USGS LAND-USE CLASSES (August 15, 1973 Landsat-1 image). A point was plotted in black if the individual cell was correctly classified when checked against the corresponding cell on the digital 1972-1973 USGS land-use reference map. The linear-discriminant algorithm developed from the rectangular training sets was used to classify this one-ninth rectilinearly sampled image. It used the first 20 variables and correctly classified 9,740 points into 24 second- and third-order classes for a 26.4-percent accuracy.

step order. A total of 1,554 points were correctly classified after the four steps, yielding an FOM,* or overall accuracy of 37.9 percent (figure 57 and table 18). Since the one-ninth by one-ninth rectilinear-point training-set sample of 4,100 points constituted a statistical sample of the points in the final map, this accuracy closely represented the final map accuracy that might be expected. It therefore directly represented map or verification accuracy in lieu of the training-set accuracy usually achieved by this type of activity. The representative nature of the verification accuracy achieved here with the 1/81-sample relative to the final, total accuracy achieved by using the same discriminant function will subsequently be clearly established.

The second test made use of the ten-band combination of four original MSS bands and six transformation ratios and achieved a slightly improved 38.4-percent verification accuracy. Again, the six MSS ratios contributed little to the improvement of the classification (figure 58 and table 19).

The full array of landscape variables, less the land-use data planes, were next included with the ten image bands. Verification accuracy was raised 16 percentage points to 53.9 percent by using these 31 nonland-use

*The FOM–the average classification accuracy–is defined as the total number of correctly classified cells, divided by the total number of cells times 100 to obtain a percentage value.

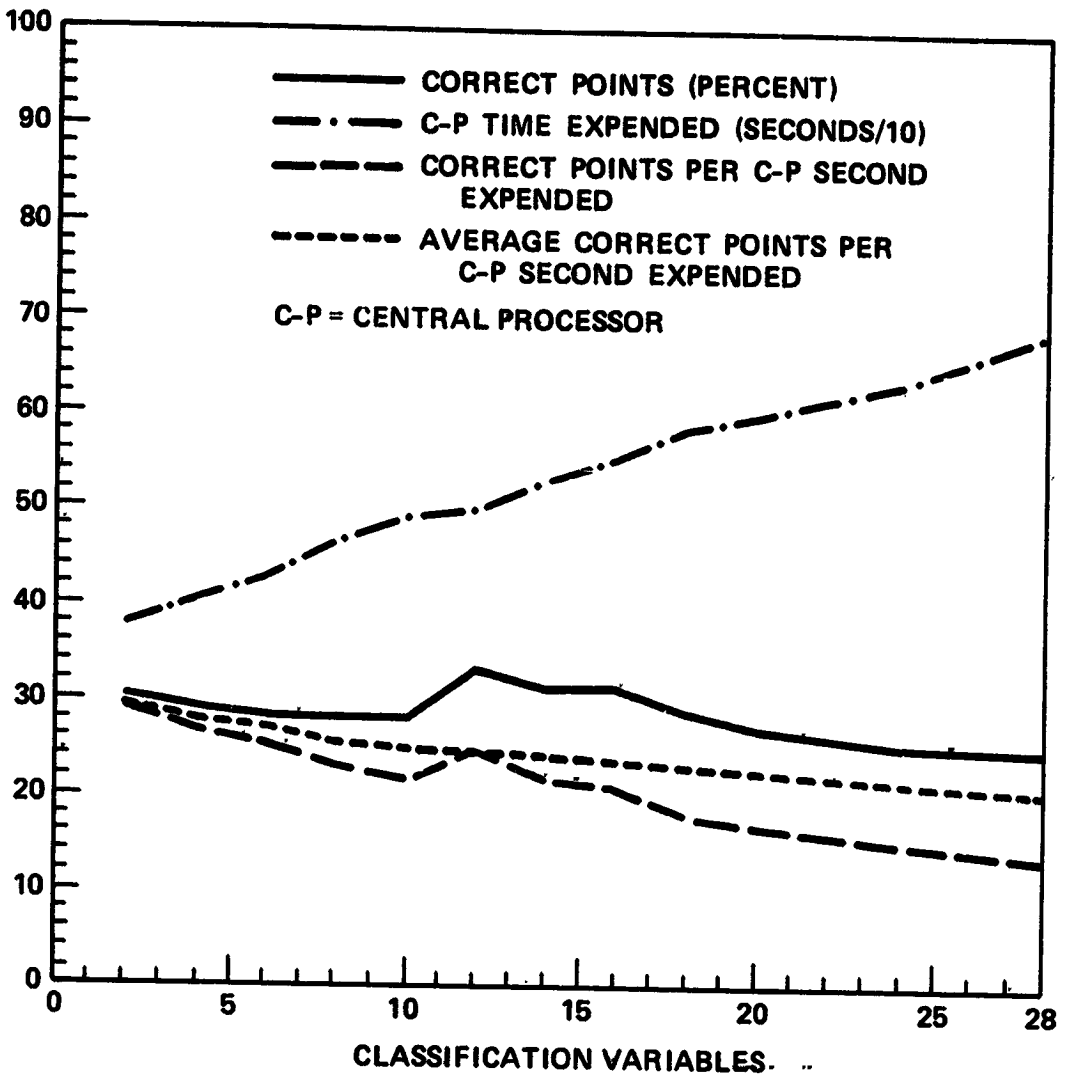


FIGURE 56. LOW VERIFIED LAND-USE ACCURACY FOR AUTOMATED INTERPRETATION USING RECTANGULAR TRAINING-SET STATISTICS. The discriminant function was derived from the three rectangular training sets of each land use representing a 2,413-point sample. It was applied in the same stepwise fashion as previously determined (table 15) to classify a rectilinearly sampled one-ninth image of 36,864 points. This sample image-classification accuracy was only 24.1-percent correct, based on a point-to-point comparison to the 1972-1973 USGS land-use reference map, whereas 99.6-percent accuracy was achieved on the rectangular training-set sample at step number 28. The potential bias of manual training-set selection was clearly evident in this case. The variable numbers coincide with step numbers in table 15. Intermediate variables were included in the classification, but their accuracies were not checked.

TABLE 17

LOW VERIFIED LAND-USE ACCURACY FOR AUTOMATED INTERPRETATION USING RECTANGULAR TRAINING-SET STATISTICS (August 15, 1973, image). The fortuitous selection of rectangular training areas yielded unrepresentative statistics that could sharply separate the sample blocks (table 15), but that were inapplicable to the rectilinearly sampled image (figures 55 and 56). This table is based on identifying 24 land uses (table 1) of 36,864 picture elements on a point-to-point comparison with the digital 1972-1973 USGS land-use reference map. The variables were added in the predetermined order of table 15, using the discriminant function developed from the 2,413-point sample rectangular training set. The image accuracy was a disappointing 24.1 percent after 28 variables added versus a predicted 99.6 percent. Intermediate variables were included, but their accuracies were not checked in the classification.

Step Number	Variable Entered	Training-Set Classification		C-P* Time Expended (Seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
	Landsat (forced)						
2	MSS-5/MSS-4 ratio	11,185	30.34	378.81	29.53	29.53	573.23
4	MSS-5 (visible red)	10,708	29.05	403.74	26.52	27.98	93.89
6	MSS-6/MSS-4 ratio	10,598	28.75	426.92	24.82	26.86	108.88
8	MSS-5/MSS-6 ratio	10,365	28.12	461.85	22.44	25.64	28.90
10	MSS-7/MSS-5 ratio	10,369	28.13	478.69	21.66	24.76	6.05
	Landscape (free)						
12	Built-up urban area MD†	12,075	32.76	499.91	24.15	24.64	1,294.12
14	Freeway interchange MD	11,425	30.99	527.16	21.67	24.15	349.27
16	Median housing-unit rent	11,445	31.05	554.67	20.63	23.63	349.55
18	Median housing-unit value	10,417	28.26	580.60	17.94	22.86	224.90
20	Average number of cars per family	9,740	26.42	599.09	16.26	22.06	119.81
24	1969 mean family income	9,162	24.85	638.01	14.36	21.17	163.80
28	Average number of year-round housing units per acre	8,870	24.06	682.75	12.99	20.28	97.37

*C-P = central processor.
†MD = minimum distance.

landscape variables to classify the 24 land uses sought. Although not as spectacular as the increase in training-set accuracy achieved when using the rectangular training sets, the 53.9-percent verification accuracy was realistically representative of larger-scale image-classification results (figure 59 and table 20).

Verification

The discriminant function for each step was saved from the 41-variable discriminant analysis of the 4,100 grid-sampled point-training set for verification testing on the one-ninth sampled image. The classification results and resultant step-by-step display maps confirmed the utility of the 1/81-grid-sampled training set as a valid and representative sampling of the larger image and illustrated a successful display product of the new systematic, point-feature extraction process. The 52.9-percent verification accuracy for this one-ninth

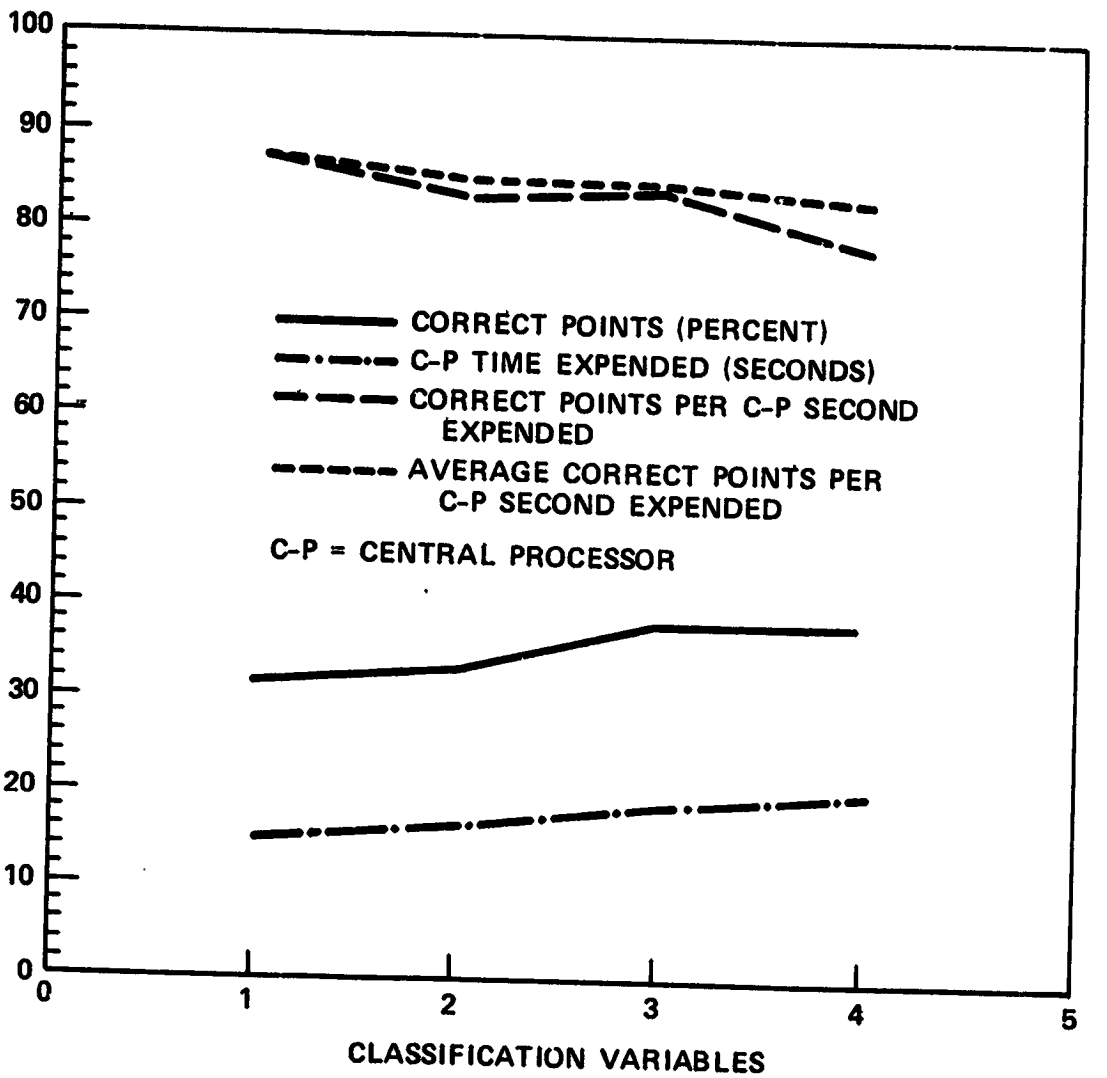


FIGURE 57. VERIFICATION ACCURACIES AND COSTS OF PROCESSING SINGLE-DATE (August 15, 1973) LANDSAT IMAGERY OF THE DENVER METROPOLITAN AREA WITH FOUR MSS BANDS. This figure is based on identifying the 24 land uses (table 1) with 4,100 grid-sampled picture elements and the 1972-1973 USGS land use data plane. The Landsat channels were added in a free stepwise fashion. The variable numbers coincide with the step numbers in table 18.

TABLE 18

VERIFICATION ACCURACIES OF AUTOMATED INTERPRETATION OF THE FOUR ORIGINAL MSS BANDS OF AN IMAGE (August 15, 1973) OF THE DENVER METROPOLITAN AREA. This table is based on identifying 24 land uses (table 1) with a test sample of 4,100 picture elements. These elements were systematically sampled from the center of each nine-by-nine array of picture elements and identified and grouped by the land-use codes specified by the 1972-1973 USGS land-use data plane. The Landsat-1 image variables were added in a free stepwise fashion and classified using linear-discriminant analysis.

Step Number	Landsat Variable Entered (free)	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
1	MSS-7 (solar infrared)	1,290	31.46	14.79	87.22	87.22	68.56
2	MSS-4 (visible green)	1,356	33.07	16.32	83.09	85.05	33.21
3	MSS-5 (visible red)	1,554	37.90	18.54	83.82	84.59	35.41
4	MSS-6 (solar infrared)	1,554	37.90	19.95	77.89	82.67	1.23

*C-P = central processor.

sampled image after the 41st step was substantially the same as the 53.9-percent verification accuracy achieved with the 1/81 point-sampled training set (figure 60 and table 21).

Classification maps of the one-ninth sampled image were preserved for each of the variables tested as they were entered in the predetermined stepwise order (table 19) into the discriminant function. The accuracy of each successive map produced in this stepwise fashion was verified on a point-to-point basis (36,864 points) by comparison with the known six first-order USGS classes* (figure 61) and 24 second- and third-order USGS classes (figure 62). These verified classification maps displayed urban, agricultural, and combined agricultural/urban land-use themes for every fourth variable added. The incremental verified accuracy contributed by the stepwise-added variables (table 21), together with its spatial distribution, can be visually assessed in this fashion relative to the 1972-1973 USGS reference theme maps (figures 61 and 62).

Finally, two additional verified classification maps were displayed for the last or 41st step results for the one-ninth sample image. The average composite second- and third-order verified accuracy was 53.9 percent for the 19,482 correctly classified points (figure 63a). The average first order* verified accuracy was 78.3 percent for the 28,871 correctly classified points (figure 63b) out of a population of 36,864 picture elements.

MAXIMUM-LIKELIHOOD CLASSIFICATION WITH GRID-SAMPLED TRAINING SETS

The solution of the feature-extraction problem provided an opportunity to compare parametric versus non-parametric classification algorithms for both accuracy and time using the maximum-likelihood ratio and the linear-discriminant function respectively.

*Classified at the second- or third-order USGS categories, but checked or verified against the correct first order of six-gross land uses only.

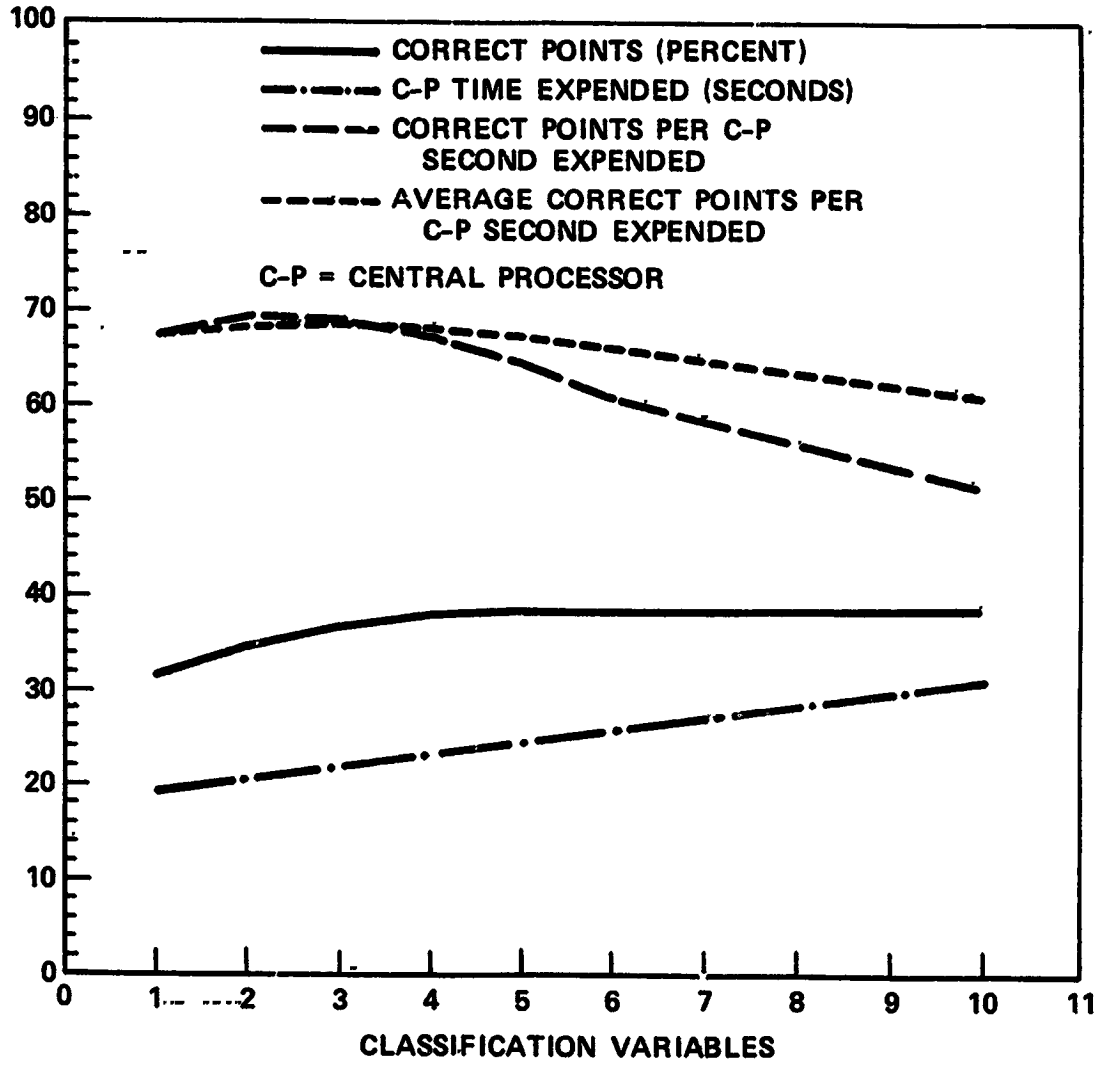


FIGURE 58. VERIFICATION OF ACCURACIES AND COSTS OF PROCESSING SINGLE-DATE (August 15, 1973) LANDSAT IMAGERY OF THE DENVER METROPOLITAN AREA WITH THE FOUR PRIMARY AND SIX MSS-BAND RATIOS. This figure is based on identifying the 24 land uses (table 1) with 4,100 grid-sampled picture elements and the 1972-1973 USGS land-use data plane. The Landsat variables were added in a free stepwise fashion. The variable numbers coincide with the step numbers and MSS-band/ratio identities in table 19.

TABLE 19

VERIFICATION ACCURACIES OF AUTOMATED INTERPRETATION OF THE FOUR ORIGINAL MSS BANDS AND THEIR SIX RATIOS FOR A SINGLE-DATE LANDSAT IMAGE (August 15, 1973) OF THE DENVER METROPOLITAN AREA. This table is based on identifying the 24 land uses (table 1) with a test sample 4,100 picture elements. These elements were systematically sampled from the center of each nine-by-nine array of picture elements and identified and grouped by the land-use codes specified by the 1972-1973 USGS land-use data plane. The Landsat-1 image variables were added in a free stepwise fashion and were classified using linear-discriminant analysis.

Step Number	Landsat Variable Entered (free)	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
1	MSS-7 (solar infrared)	1,290	31.46	19.14	67.40	67.40	68.56
2	MSS-5/MSS-4 ratio	1,417	34.56	20.50	69.12	68.29	37.02
3	MSS-7/MSS-5 ratio	1,495	36.46	21.76	68.70	68.44	38.61
4	MSS-4 (visible green)	1,551	37.83	23.15	67.00	68.04	25.46
5	MSS-5 (visible red)	1,570	38.29	24.41	64.32	67.21	52.56
6	MSS-7/MSS-4 ratio	1,562	38.10	25.75	60.66	65.96	10.54
7	MSS-5/MSS-6 ratio	1,562	38.10	26.86	58.15	64.66	9.10
8	MSS-7/MSS-6 ratio	1,570	38.29	28.09	55.89	63.36	3.34
9	MSS-6 (solar infrared)	1,575	38.41	29.39	53.59	62.05	1.08
10	MSS-6/MSS-4 ratio	1,575	38.41	30.77	51.19	60.71	2.28

*C-P = central processor.

The same 4,100 entities of the 1/81 grid-sampled point-training set were used to generate the mean and covariance matrices for classifying the one-ninth sampled image by maximum-liklihood ratioing with four, six, and 22 variables. The four-variable combination was the basic four MSS bands. The six-variable set included MSS-4, MSS-7, MSS-7/MSS-5 ratio, MSS-5/MSS-4 ratio, topographic elevation, and urban built-up minimum distances. The 22-variable run contained the maximum number of the 41 variables that could be used (that is, those variables with nonzero variance in all classes and therefore capable of matrix inversion). These were the four basic MSS bands, six MSS ratios, four MSS/insolation ratios, topographic elevation and aspect, insolation, and five transportation minimum-distance variables.

These three combinations were applied to the one-ninth sample image using maximum-likelihood ratioing and were checked for first- and composite second-/third-order accuracy on a point-to-point basis for the 36,864 points against the 1972-1973 USGS land-use data plane. These map-verification accuracies checked and displayed to the first order were 53.6, 68.1, and 61.5 percent, respectively, for four, six, and 22 variables (figure 64), whereas the composite second- and third-order accuracies were 4.3, 15.4, and 2.4 percent, respectively (figure 65).

The 1/81 grid-sampled-point data set was classified by stepwise discriminant analysis for the same combination of variables to provide comparative results at both the first- and composite second- and third-order levels for the nonparametric comparison (table 22). Although these linear-discriminant analyses were run on a smaller data set than that of the maximum-likelihood ratio tests, the close correspondence of the 1/81 data set to the one-ninth sample-image results has already been documented.

The poor performance of the maximum-likelihood ratio, especially for the second- and third-order classes, was surprising. A detailed explanation for this occurrence is not currently possible. However, the fact that the first-order accuracies were comparatively high relative to the composite second- and third-order accuracies may provide a clue. It is believed that the capability for specifying *a priori* class probabilities for discriminant analysis materially improved all its classifications. No such capability existed in the maximum-likelihood algorithm used. Hence, it could achieve comparable results to discriminant analysis for only six first-order classes; however, the statistical similarity of 24 second- and third-order USGS land-use classes proved too much for it to handle.

A prerequisite for the use of the maximum-likelihood ratio algorithm is that the data are multivariate normally distributed (Appendix D). The non-Gaussian distributions of the 4,100 grid-sampled training-set points could have conceivably accounted for the poor performance of the maximum-likelihood ratio in classifying the 24 land-use classes of the one-ninth image sample. However, the central limit theorum essentially states that the distribution of either the sums of averages of n measurements drawn from any population tend to possess, approximately, a normal distribution in repeated sampling when n is large (Referenec 43). The exact form of the distribution of the grid-sampled points is not now known, but the normal approximation was justifiable in the four- and six-variable combinations in which the Landsat image bands were numerically dominant.

This presented the perennial problem of equal versus weighted probabilities for predicting from a training set to an unknown image scene. The use of equal *a priori* class probabilities presumes equal likelihood of each land use in the larger scene, whereas weighted probabilities presume that the image analysis has "*a priori*" knowledge of the approximate amount of each land use within the classification set. Some knowledge of the amount of each land use in the scene is invariably available or could be achieved by a cursory inspection of Landsat imagery in the photographic form. The 1/81 point-sampled training set represented a probability sampling proportional to total land-use area in the larger sampled image. These *a priori* class probabilities were used for both training-set and image-classification efforts. The statistical benefits of inputing this *a priori* knowledge to the linear-discriminant analysis, versus the assumption of equal probabilities, was also tested and verified here for the different combinations and at both the first- and composite second- and third-order levels (table 22).

Finally a clear economic advantage was also illustrated irrespective of classification accuracies as the maximum-likelihood ratio algorithm took 321.2, 583.7, and 1,574.6 seconds, respectively, for the four-, six-, and 22-variable combinations. A computational time advantage emerged for the linear-discriminant analysis function, which took 417.6, 441.4, and 624.0 seconds, respectively, for five-, seven-, and 22-variable combinations (table 21). Clearly, the linear-discriminant function was much less sensitive to the increasing number of channels. A further cost advantage emerges if cost is weighted by classificational accuracies for a comparative analysis.

PRODUCTION OF LAND-USE MAPS

After the feature-extraction step is completed, the fourth step in the pattern-recognition process is that of decision and classification. All cells of interest are tested against all features selected for representation in the feature-extraction step. A probabilistic decision is then made to determine into which class each cell is placed.

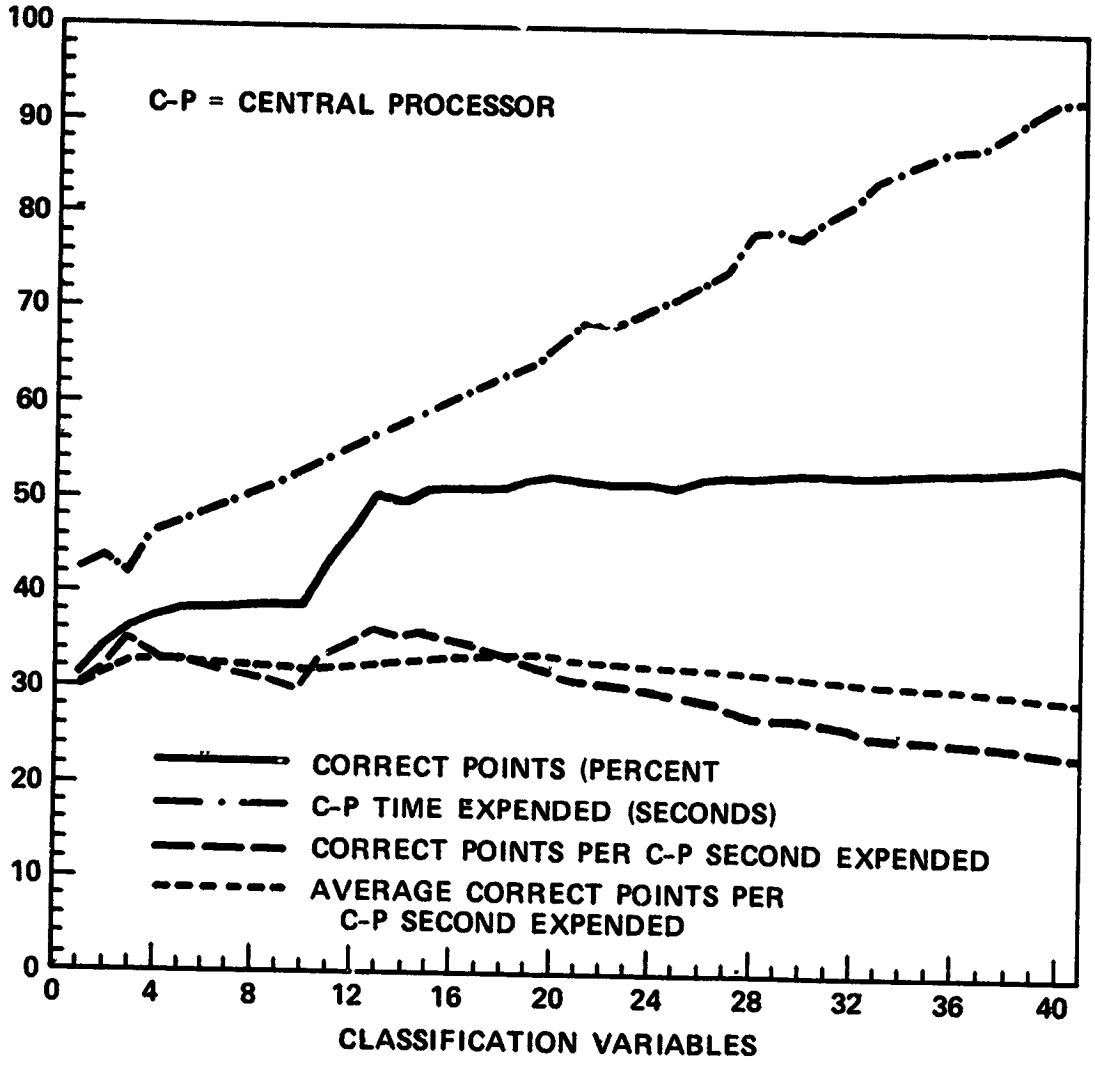


FIGURE 59. VERIFICATION ACCURACIES AND COSTS OF PROCESSING SINGLE-DATE (August 15, 1973) LANDSAT IMAGERY OF THE DENVER METROPOLITAN AREA WITH TEN MSS-BAND/RATIO VARIABLES AND 31 ANCILLARY LANDSCAPE VARIABLES. This figure is based on identifying the 24 land uses (table 1) with 4,100 grid-sampled picture elements and the 1972-1973 USGS land-use data plane. The Landsat image variables were forced in a predetermined order (table 19), and the landscape variables were added in a free stepwise fashion. The variable numbers coincide with the step numbers and MSS-band and data-plane identities in table 20.

TABLE 20

IMPROVEMENT* IN THE VERIFICATION ACCURACIES OF AUTOMATED INTERPRETATION BY THE ADDITION OF 31 ANCILLARY VARIABLES FOR A SINGLE-DATE (August 15, 1973) LANDSAT IMAGE OF THE DENVER METROPOLITAN AREA. This table is based on identifying 24 land uses (table 1) with a test sample of 4,100 picture elements. These elements were systematically sampled from the center of each nine-by-nine array of picture elements and identified and grouped by the land-use codes specified by the 1972-1973 USGS land-use data plane. The Landsat-1 variables were forced in a predetermined order (table 19), and the landscape variables were added in a free-stepwise fashion and were classified using linear-discriminant analysis.

Step Number	Variable Entered	Training-Set Classification		C-P† Time Expended (seconds)	Correct Points/C-P Second Expended (%)		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
	Landsat (forced)						
1	MSS-7 (solar infrared)	1,290	31.46	42.45	30.39	30.39	68.56
2	MSS-5/MSS-4 ratio	1,417	34.56	43.84	32.32	31.37	37.02
3	MSS-7/MSS-5 ratio	1,495	36.46	42.25	35.38	32.69	38.61
4	MSS-4 (visible green)	1,551	37.83	46.52	33.34	32.86	25.46
5	MSS-5 (visible red)	1,570	38.29	47.48	33.07	32.91	52.56
6	MSS-7/MSS-4 ratio	1,562	38.10	48.48	32.22	32.78	10.54
7	MSS-5/MSS-6 ratio	1,562	38.10	49.37	31.64	32.61	9.10
8	MSS-7/MSS-6 ratio	1,570	38.29	50.52	31.08	32.40	3.44
9	MSS-6 (solar infrared)	1,575	38.41	51.25	30.73	32.40	1.08
10	MSS-6/MSS-4 ratio	1,575	38.41	52.60	29.94	31.95	2.28
	Landscape (free)						
11	Topographic elevation	1,792	43.71	53.88	33.26	32.08	259.46
12	Average number of cars per family	1,917	46.76	55.24	34.70	32.33	74.34
13	Built-up urban area MD†	2,056	50.15	56.52	36.38	32.69	65.53
14	Topographic slope	2,040	49.76	57.67	35.37	32.91	56.67
15	Average number of families per acre	2,120	51.71	58.95	35.96	33.15	32.39
16	1969 mean family income	2,118	51.66	60.17	35.20	33.30	25.11
17	Median housing-unit value	2,127	51.88	61.44	34.62	33.39	30.21
18	Surficial geology	2,115	51.59	62.72	33.72	33.41	16.66
19	Composite minor-road MD	2,128	51.90	63.96	33.27	33.40	16.50
20	Landsat-1 image insolation	2,136	52.10	66.23	32.25	33.33	14.39
21	Freeway interchange MD	2,131	51.98	68.51	31.10	33.20	11.26
22	Freeway MD	2,106	51.37	68.00	30.97	33.07	13.99
23	Median housing-unit rent	2,117	51.63	69.21	30.59	32.94	10.49
24	Landsat-1 MSS-5 insolation ratio	2,122	51.76	70.49	30.10	32.79	8.62
25	Total two-car families	2,125	51.83	71.28	29.81	32.64	8.14
26	Landsat-1 MSS-4 insolation ratio	2,121	51.73	73.10	29.02	32.46	6.45
27	Total one-car families	2,140	52.20	74.50	28.72	32.28	6.45
28	Total population	2,154	52.54	78.33	27.50	32.06	6.60
29	Total families	2,152	52.49	78.70	27.34	31.84	7.93
30	Total three-car families	2,162	52.73	78.30	27.61	31.66	8.06
31	Total census-tract acreage	2,169	52.90	80.04	27.10	31.46	5.52
32	Composite major-road MD	2,164	52.78	81.35	26.60	31.26	5.38
33	Total occupied housing units	2,166	52.83	84.03	25.78	31.04	4.59
34	Average number of vacant housing units per acre	2,173	53.00	85.12	25.58	30.82	3.38
35	Total vacant housing units	2,185	53.29	86.42	25.28	30.60	3.42
36	Topographic aspect	2,194	53.51	87.53	25.15	30.40	3.02
37	Total year-round housing units	2,201	53.68	87.60	25.13	30.20	3.02
38	Landsat-1 MSS-6 insolation ratio	2,206	53.80	88.93	24.81	30.01	2.16
39	Average number of year-round housing units per acre	2,211	53.93	90.07	24.39	29.81	1.85
40	Population density per acre	2,216	54.05	91	24.04	29.61	1.30
41	Landsat-1 MSS-7 insolation ratio	2,208	53.85	91.78	23.80	29.41	0.96

*C-P = central processor

†MD = minimum distance

ORIGINAL PAGE IS
OF POOR QUALITY

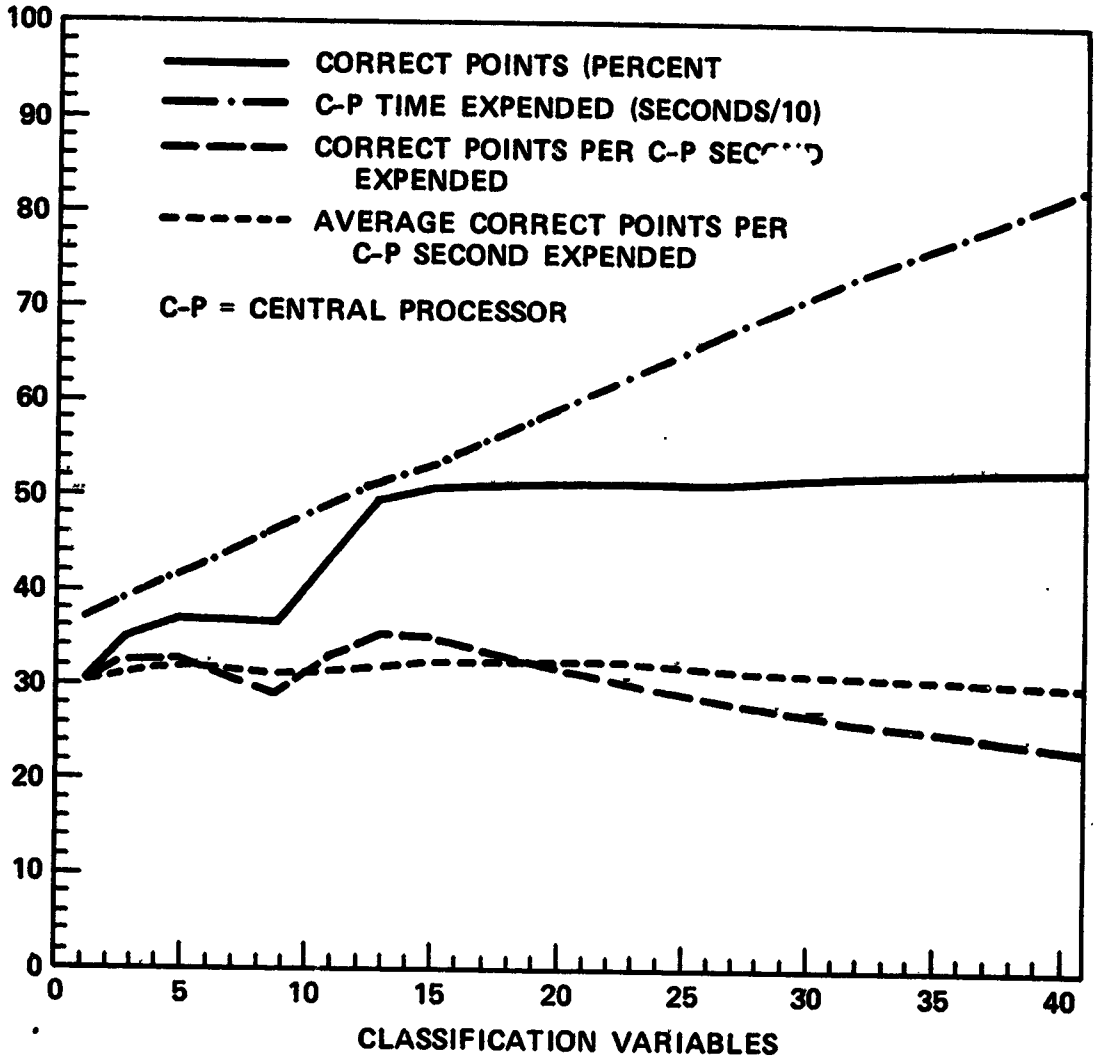


FIGURE 60. MAP-VERIFICATION ACCURACIES AND COSTS OF PROCESSING SINGLE-DATE LANDSAT IMAGERY OF THE DENVER METROPOLITAN AREA WITH TEN MSS-BAND/RATIO VARIABLES AND 31 ANCILLARY LANDSCAPE VARIABLES. This figure is based on identifying 24 land uses (table 1) for 36,864 grid-sampled picture elements representing the 1972-1973 land-use data plane. The Landsat image variables were forced in a predetermined order (table 19) and the landscape variables were added in a free stepwise fashion and were classified using linear-discriminant analysis. The variable numbers coincide with the step numbers and MSS-band and data-plane identities in table 21.

TABLE 21

MAP-VERIFICATION ACCURACY ACHIEVED WITH THE GRID-SAMPLING APPROACH TO ASSEMBLING TRAINING-SET STATISTICS (August 15, 1973, image). The one-ninth map accuracy of the sample image classification detailed below used the 4,100-point sample statistics and was in close agreement with the 41-variable training-set results (table 20). This table is based on identifying the 24 land uses (table 1) with a sample of 36,864 picture elements systematically selected as the center of each three-by-three array of picture elements and checked by direct cell-by-cell comparison with the 1972-1973 USGS land-use data plane. The 41 variables were entered in exactly the same stepwise order as determined by the grid-sampled training set (table 20).

Step Number	Variable Entered	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
	Landsat (forced)						
1	MSS-7 (solar infrared)	11,246	30.31	368.55	30.51	30.51	68.56
3	MSS-7/MSS-5 ratio	12,939	35.10	395.10	32.75	31.67	38.61
5	MSS-5 (visible red)	13,687	37.13	417.55	32.78	32.06	52.56
7	MSS-5/MSS-6 ratio	13,652	37.03	441.35	30.93	31.75	9.10
9	MSS-6 (solar infrared)	13,645	37.01	467.69	29.18	31.18	1.08
	Landscape (free)						
11	Topographic elevation	16,027	43.48	499.18	32.76	31.48	259.46
13	Built-up urban area MD†	18,161	49.26	514.49	35.30	32.11	65.53
15	Average number of families per acre	18,664	50.63	530.46	35.18	32.56	32.39
17	Median housing-unit value	18,786	50.96	552.37	34.01	32.75	30.21
19	Composite minor-road MD.	18,920	51.32	576.98	32.79	32.79	16.50
23	Median housing-unit rent	18,849	51.13	623.99	30.21	32.46	10.49
27	Total one-car families	18,986	51.50	672.05	28.25	31.99	6.45
31	Total census-tract acreage	19,210	52.11	721.17	26.64	31.42	5.52
36	Topographic aspect	19,405	52.64	772.18	25.13	30.78	3.02
41	Landsat-1 MSS-7/insolation ratio	19,482	52.85	825.90	23.59	30.07	0.96

*C-P = central processor.

†MD = minimum distance.

Selection of Optimal Mapping Variables

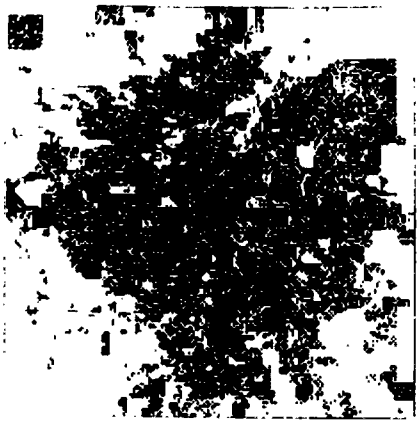
Before the full 576- by 576-cell image/landscape scene was actually classified, a search was made for an efficient subset of the 41 possible classification variables. This optimization was made necessary by the combination of the large number of variables and the size of the image. It was predetermined that a number of image variables would be included in this subset so that the resultant classification map would not be unduly influenced by the ancillary landscape variables mapped 2 to 3 years earlier.

Examination of results achieved with the 1/81 grid-sampled point-training set were also useful in this capacity (figures 57, 58, and 59). Only four bands (MSS-7, MSS-5/MSS-4 ratio, MSS-7/MSS-5 ratio, and MSS-4) accounted for practically all of the accuracy in the ten-image channel test (figure 58 and table 19). Likewise, the first six to ten ancillary variables contributed the greatest accuracy gains to the 41-channel

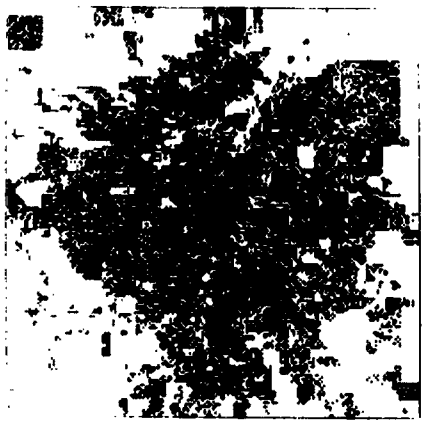
OUT FRAME

ORIGINAL PAGE IS
OF POOR QUALITY

Variable 1 Added



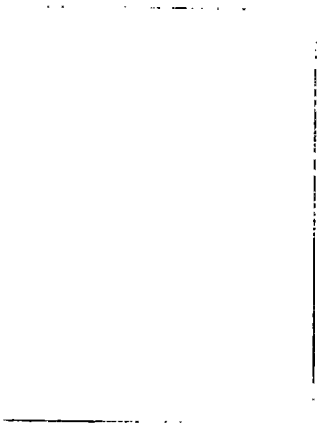
Variable 5 Added



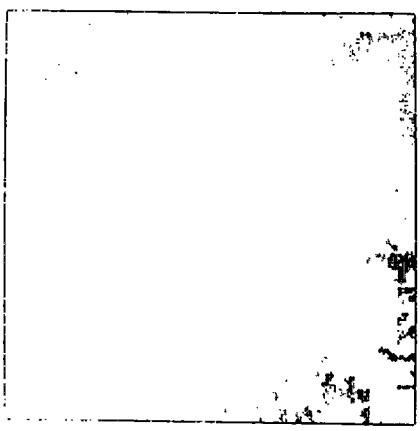
Variable 9 Added

Variable 13 Added

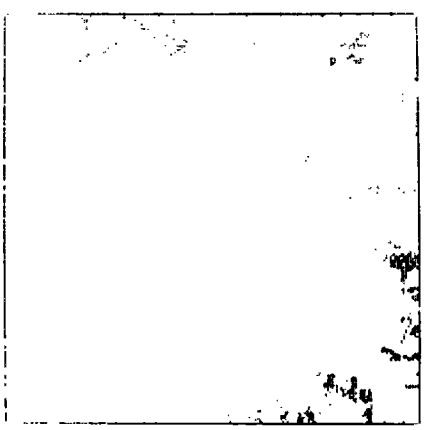
land use verified at the first order



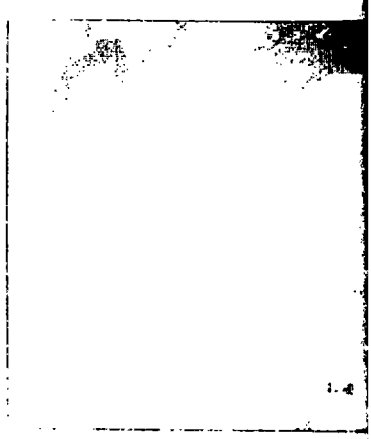
Variable 1 Added



Variable 5 Added



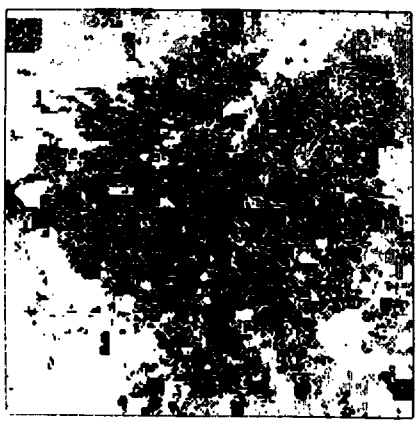
Variable 9 Added



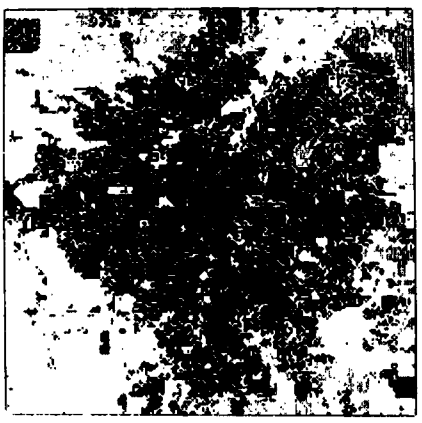
Variable 13 Added

ltural land use verified at the first order

Variable 1 Added



Variable 5 Added



Variable 9 Added

Variable 13 Added

tural (gray) and urban (black) land use verified at the first order

gray point was plotted on the display if the individual cell was correctly classified for the variable added. The correctness was made against the corresponding cell on the 1972-1973 USGS land-use data plane. The classification for the 24 second- and third-order land-use classes and was then checked at the six first-order categories only. which were used in the classification are identified in table 21. Display scale is 1:650,000.

ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME

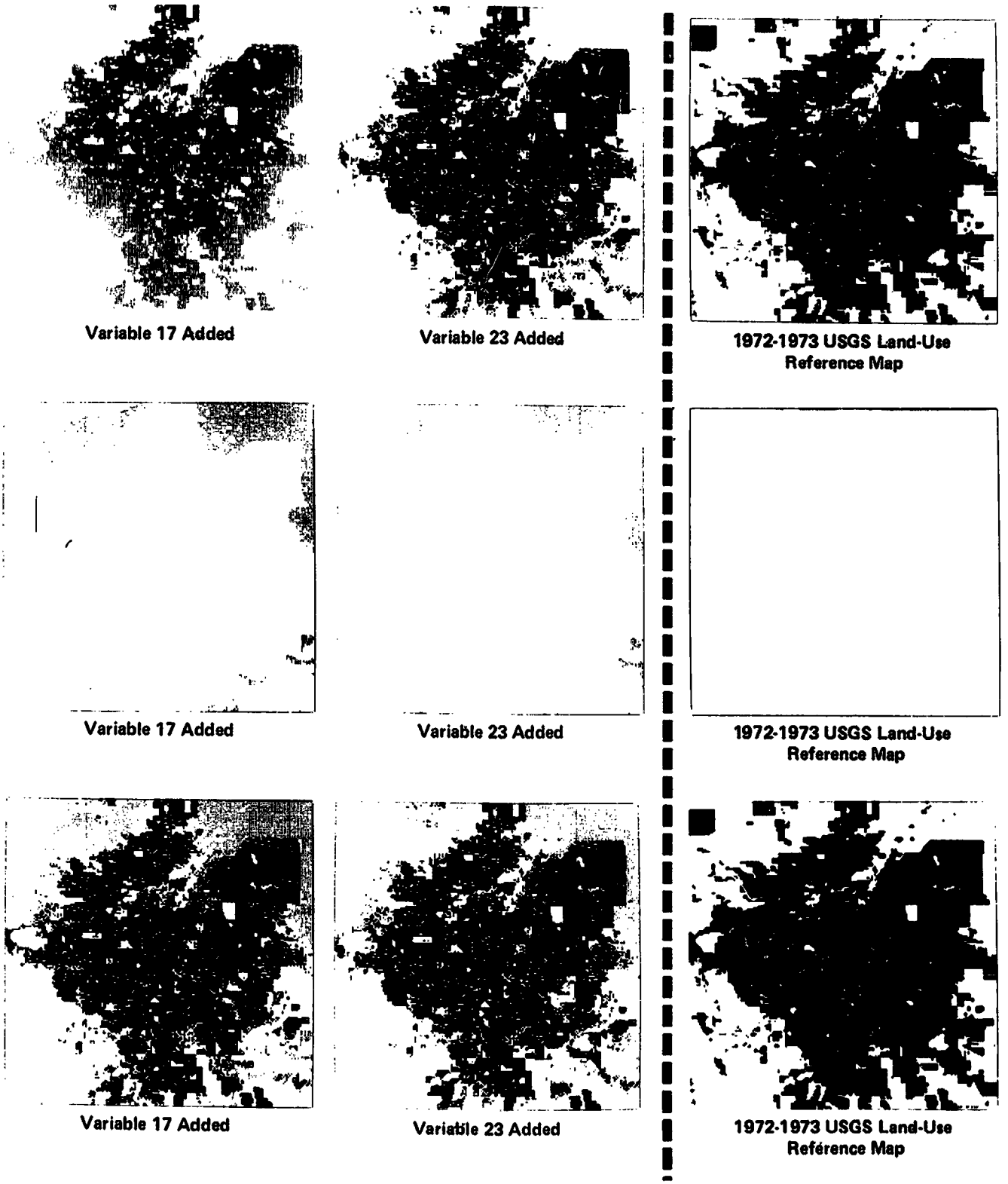


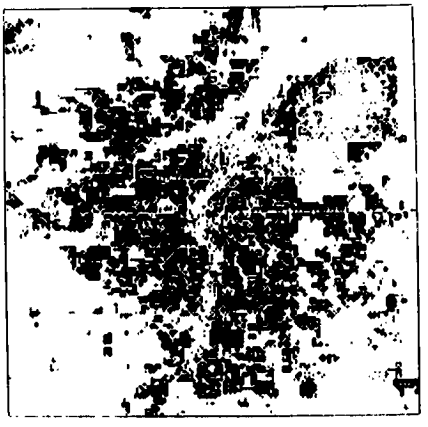
FIGURE 61. STEP-BY-STEP CLASSIFICATION OF THE LAND USE OF THE DENVER METROPOLITAN AREA VERIFIED AT THE FIRST ORDER (August 15, 1973, Landsat-1 image).

This Page Intentionally Left Blank

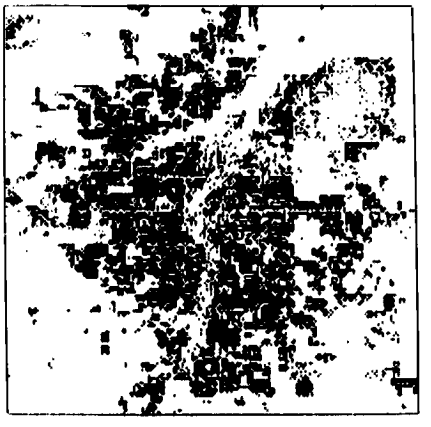
FOLDOUT FRAME

ORIGINAL PAGE IS
OF POOR QUALITY

Variable 1 Added



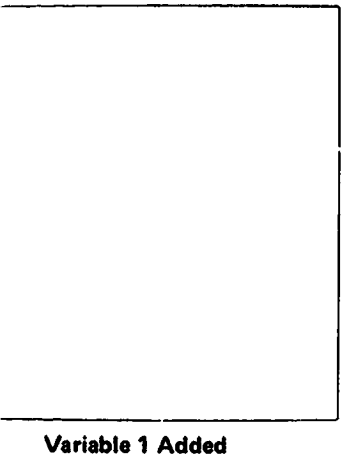
Variable 5 Added



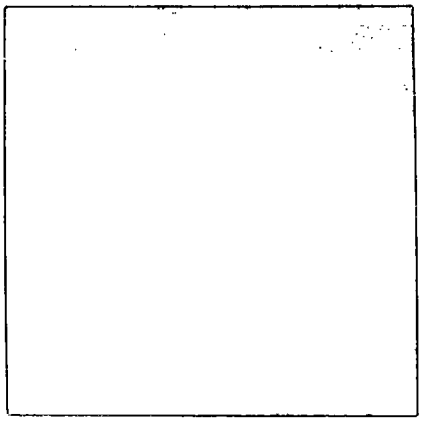
Variable 9 Added

Variable 13 Added

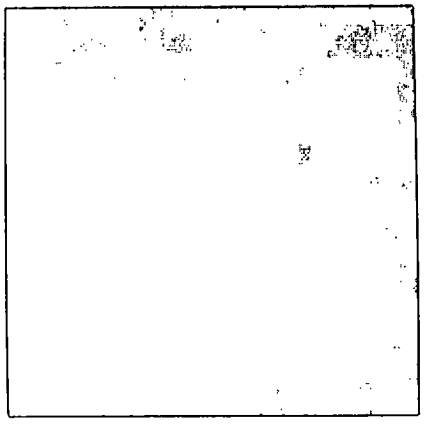
n land use verified at the second and third orders



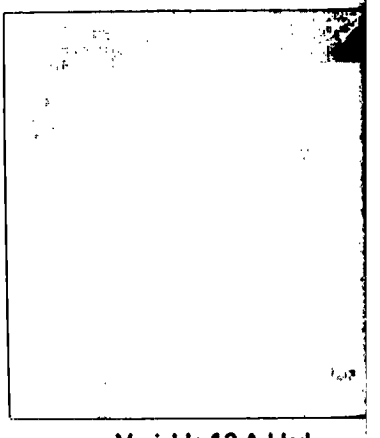
Variable 1 Added



Variable 5 Added



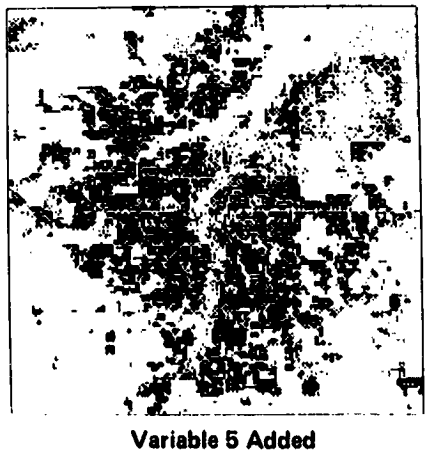
Variable 9 Added



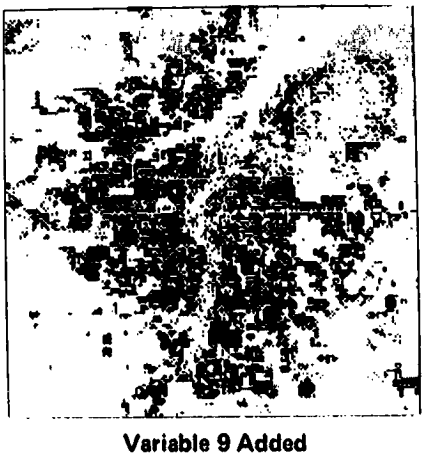
Variable 13 Added

cultural land use verified at the second and third orders

Variable 1 Added



Variable 5 Added



Variable 9 Added

Variable 13 Added

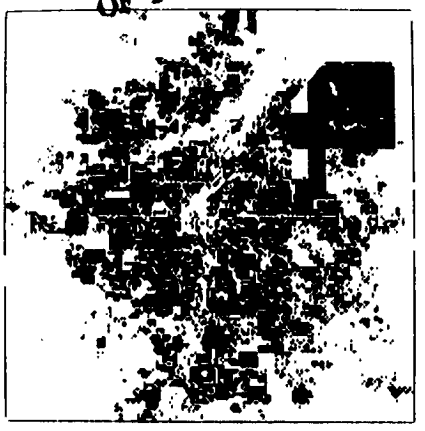
cultural (gray) and urban (black) land use verified at the second and third orders

or gray point was plotted on the display if the individual cell was correctly classified for the variable added. The
r correctness was made against the corresponding cell on the 1972-1973 USGS land-use data plane. Variables which
d in the classification are identified in table 21. Display scale is ~ 1:650,000.

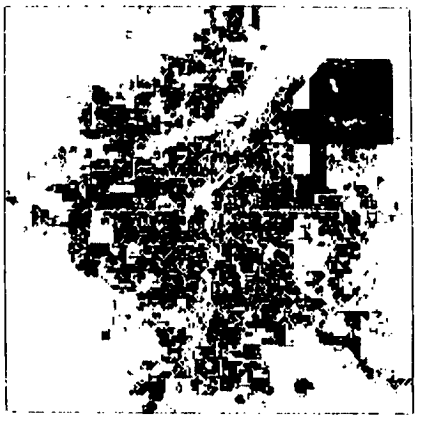
ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT

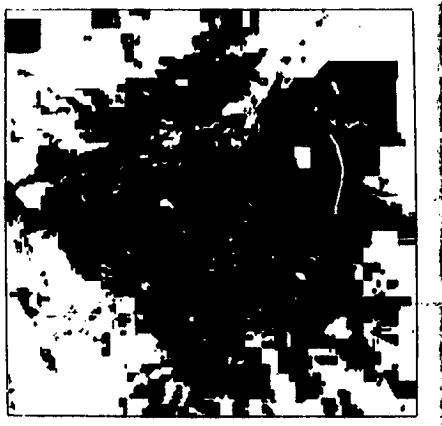
3 Added



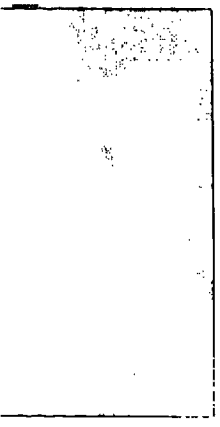
Variable 17 Added



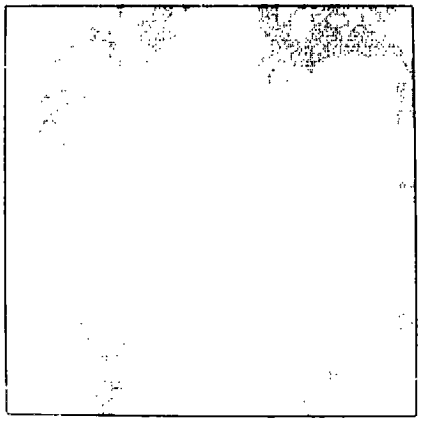
Variable 23 Added



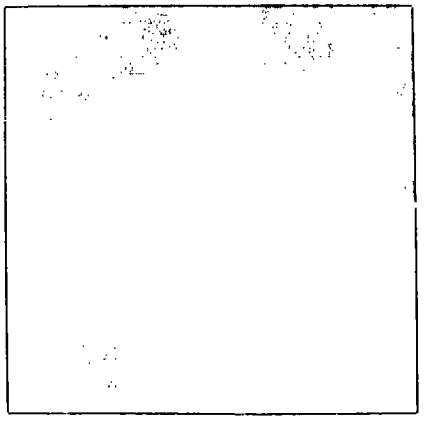
Display scale 1:650,000
1972-1973 USGS Land-Use
Reference Map



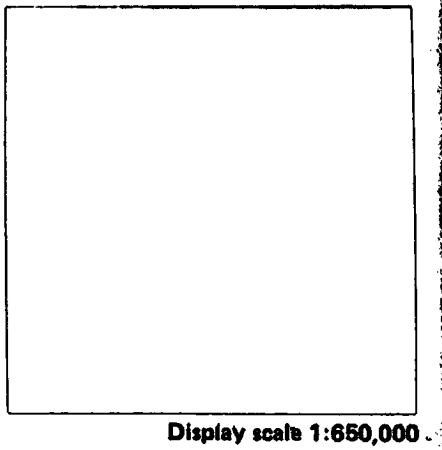
3 Added



Variable 17 Added

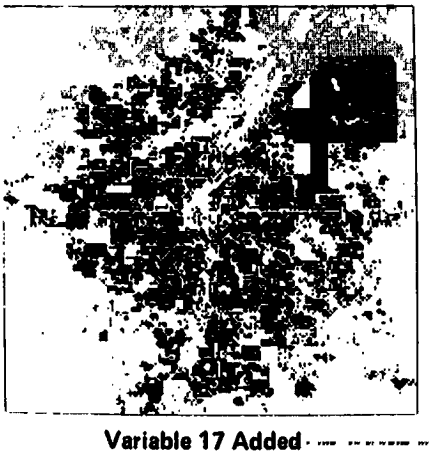


Variable 23 Added

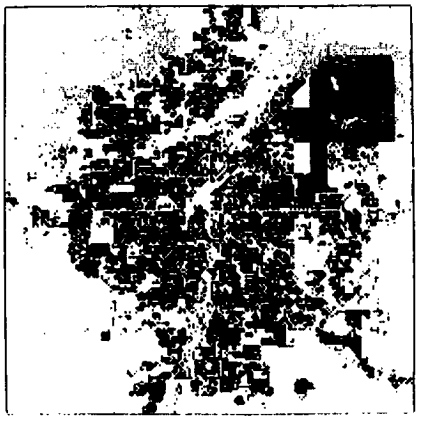


Display scale 1:650,000
1972-1973 USGS Land-Use
Reference Map

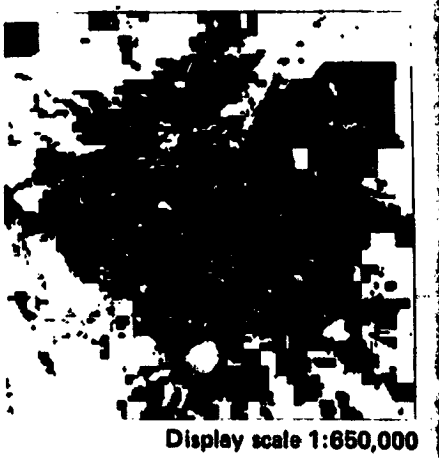
3 Added



Variable 17 Added



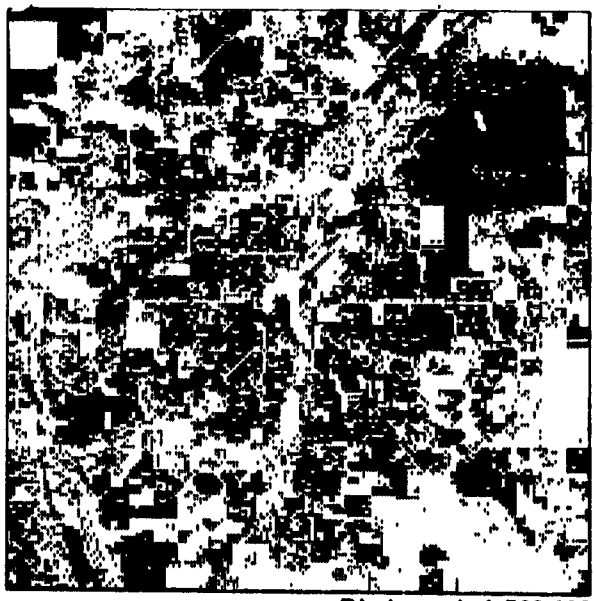
Variable 23 Added



Display scale 1:650,000
1972-1973 USGS Land-Use
Reference Map

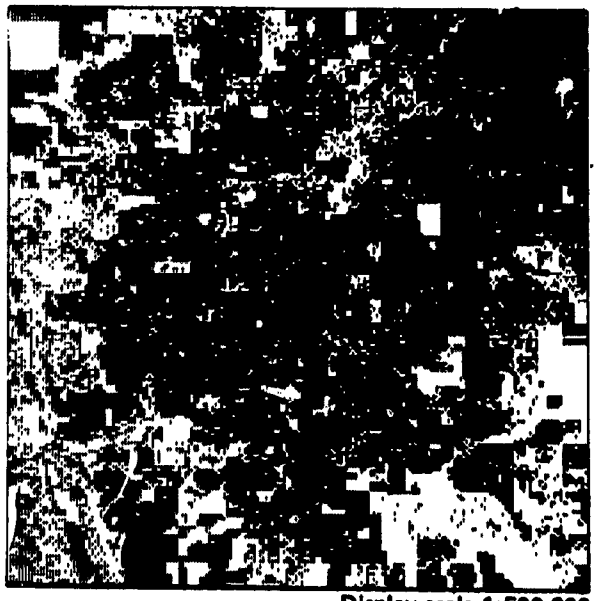
FIGURE 62. STEP-BY-STEP CLASSIFICATION OF THE LAND USE OF THE DENVER METROPOLITAN AREA VERIFIED AT THE SECOND AND THIRD ORDERS (August 15, 1973, Landsat-1 image).

This Page Intentionally Left Blank



Display scale 1:500,000

(a) Original 24 Second- and Third-Order USGS Classes



Display scale 1:500,000

(b) Composite Six First-Order USGS Classes

FIGURE 63. VERIFIED MAP ACCURACY FOR THE DENVER METROPOLITAN AREA FOR ALL USGS LAND-USE CLASSES AT THE 41st STEP (August 15, 1973, Landsat-1 image). The linear-discriminant classification tested here employed the 41-image/map variables exclusive of land use. A point was plotted in black if the individual cell was correctly classified when checked against the corresponding cell on the 1972-1973 USGS land-use data plane. This image (b) represents the same 24 classes but was checked for correctness at the six first-order classes only.

training-set test (figure 59 and table 20). Five test runs with no forced variables were made on the 1/81 training set with the four image bands and the six, seven, eight, nine, and ten landscape variables, progressively adding another landscape variable for each successive test. The ten-channel combination using six landscape variables proved to be the most accurate, correctly classifying 2,091 points for a 51-percent verified accuracy (figure 66 and table 23).

The use of the ten channels achieved 94 percent of the maximum accuracy that could have been obtained with all 41 variables. However, it represented an expenditure of only 34 percent of the processing time that would have been needed.

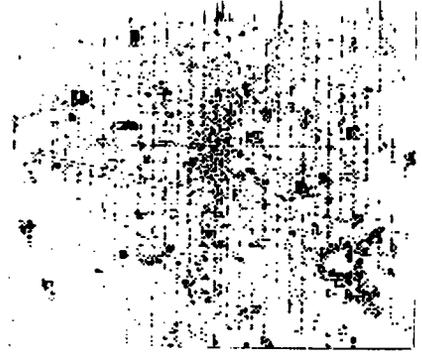
First-Order Theme Map

ORIGINAL PAGE IS
OF POOR QUALITY

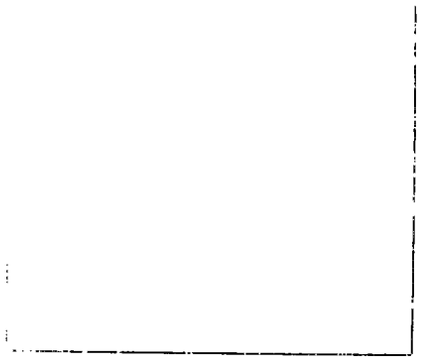
Each of the following first-order classification displays that resulted is composed of three thematic maps: (1) the 1972-1973 USGS land-use data plane for reference; (2) the verified discriminant-classified map showing only correctly mapped 0.4-ha (1.111-acre) picture elements; and (3) the actual discriminant-classified map. Unfortunately, the lack of an appropriate color-display device necessitated the cumbersome display of each separate first-order theme map in black-and-white. Using this general approach, the six aggregate first-order categories were displayed for the zoomed USGS reference data plane, the verified classification map, and the machine-processed land-use map (figure 67).

This Page Intentionally Left Blank

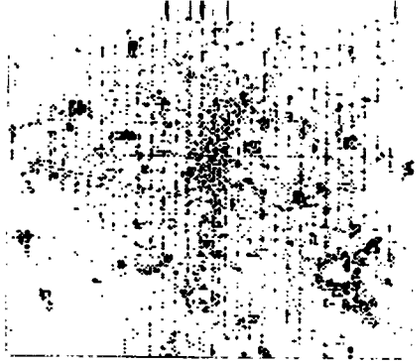
ORIGINAL PAGE IS
POOR QUALITY



All Correct First-Order
Urban Land Use

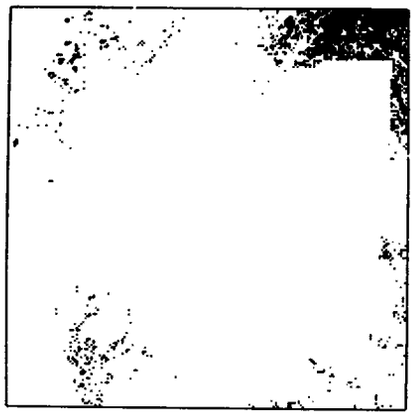


All Correct First-Order
Agricultural Land Use

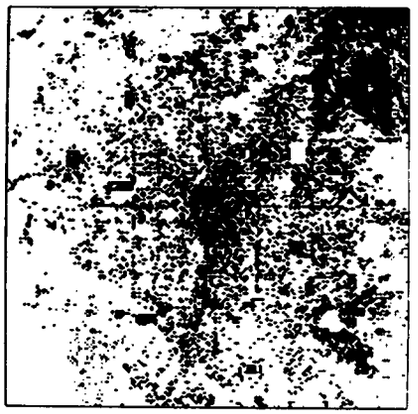


(a) Verified four original Landsat-1 MSS band classification

All Correct First-Order
Urban Land Use

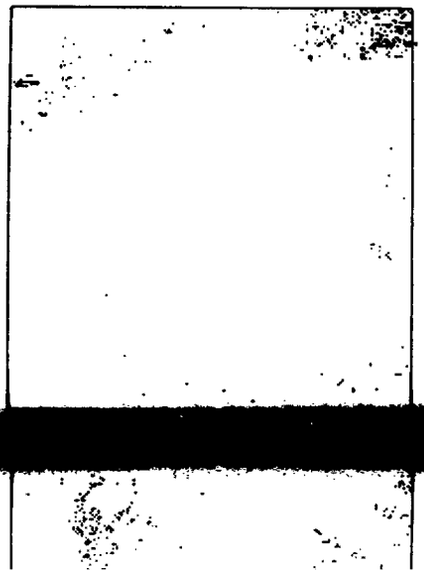


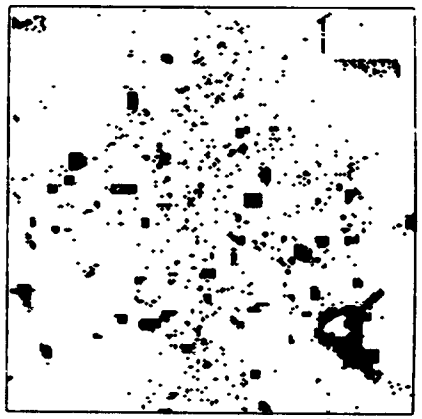
All Correct First-Order
Agricultural Land Use



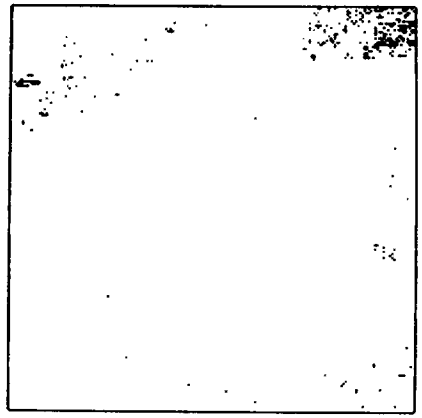
All Correct First-Order
Urban/Agricultural Land Use

(b) Verified six-channel Landsat-1/ancillary variable classification

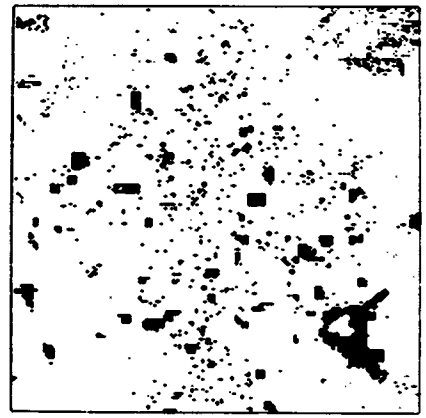




All Correct First-Order
Urban Land Use

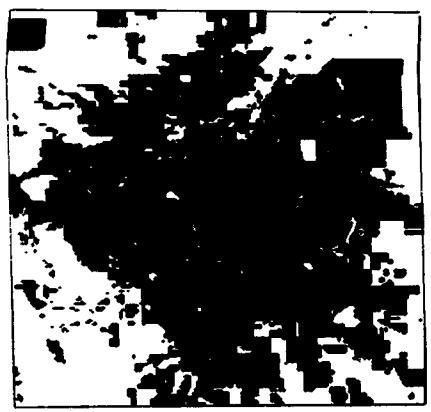


All Correct First-Order
Agricultural Land Use

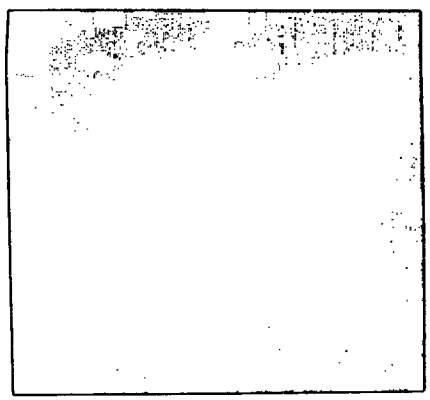


All Correct First-Order
Urban/Agricultural Land Use

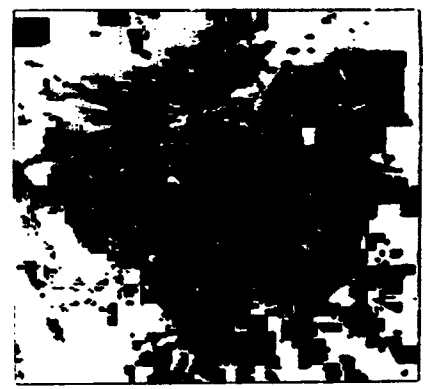
(c) Verified 22-channel Landsat-1/ancillary variable classification



All Correct First-Order
Urban Land Use



All Correct First-Order
Agricultural Land Use



All Correct First-Order
Urban/Agricultural Land Use

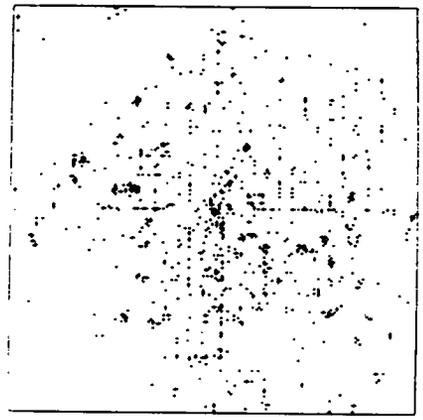
(d) 1972/73 USGS land-use reference Data Plane

FIGURE 64. MAXIMUM-LIKELIHOOD RATIO (OR GAUSSIAN) IMAGE CLASSIFICATION OF THE DENVER METROPOLITAN AREA VERIFIED AT THE FIRST ORDER (August 15, 1973, Landsat-1 image). A black (urban) or gray (agricultural) point was printed on the display if the individual cell was correctly classified. The check for correctness was made against the corresponding cell on the 1972-1973 USGS land-use data plane. The classification was made for the 24 second- and third-order land-use classes, and was then checked at the first-order categories only. Variables which were used in the classification are identified in table 22. Display scale is ~ 1:650,000.

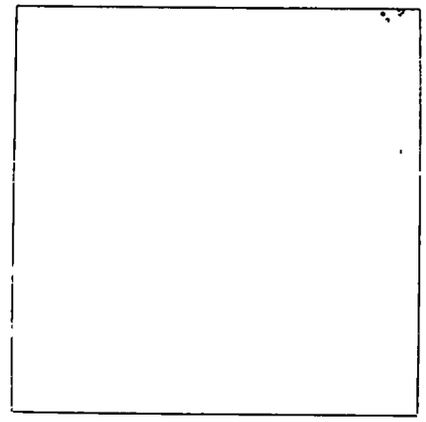
ORIGINAL PAGE IS
OF POOR QUALITY

This Page Intentionally Left Blank

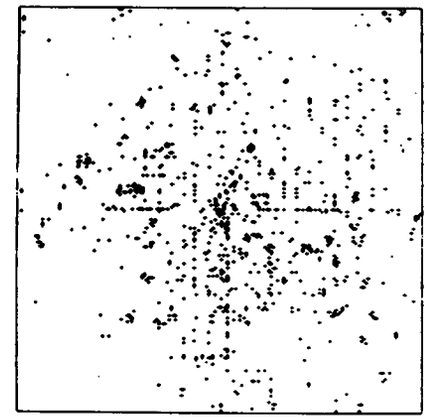
ORIGINAL PAGE IS
OF POOR QUALITY



All Correct Second-Order
Urban Land Use



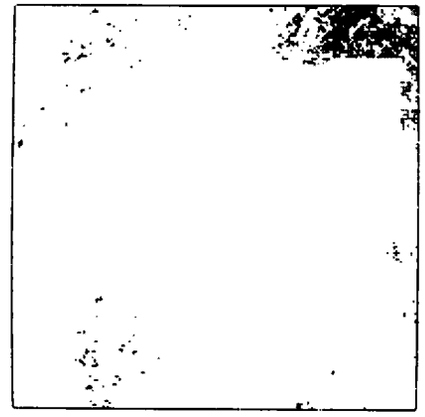
All Correct Second-Order
Agricultural Land Use



All Correct Second-Order
Urban/Agricultural Land Use

(a) Verified four original Landsat-1 MSS-band classification

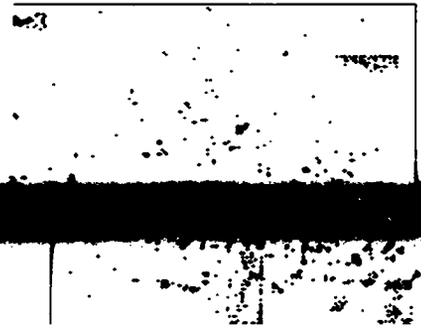
All Correct Second-Order
Urban Land Use

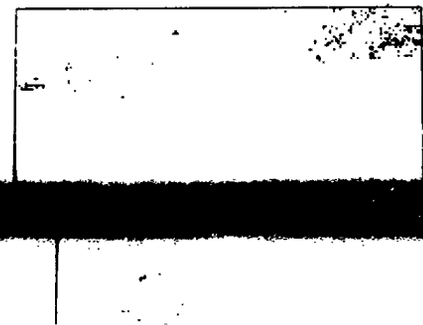


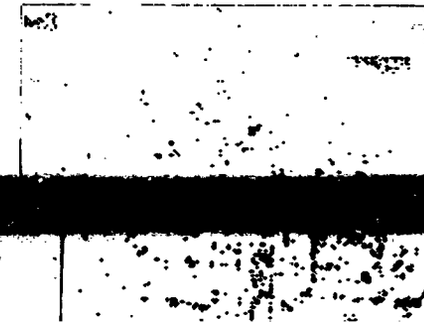
All Correct Second-Order
Agricultural Land Use

All Correct Second-Order
Urban/Agricultural Land Use

(b) Verified six-channel Landsat-1/ancillary variable classification





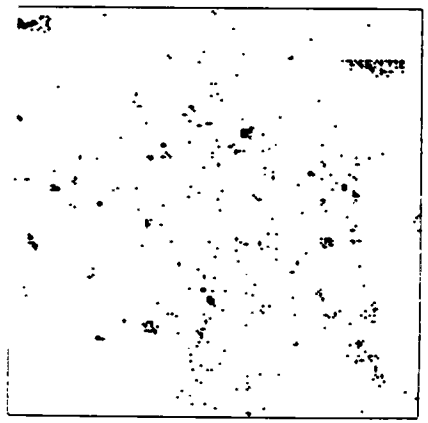


Urban Land Use

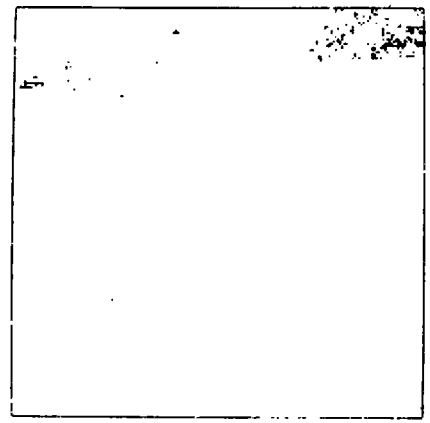
Agricultural Land Use

All Correct Second-Order
Urban/Agricultural Land Use

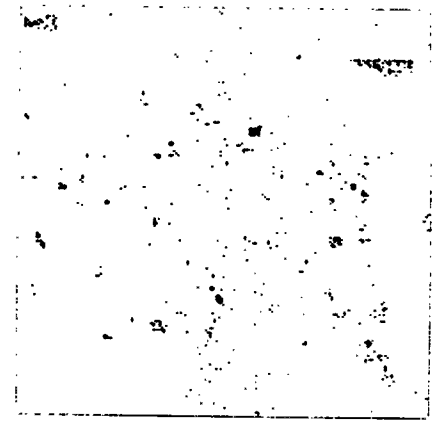
(b) Verified six-channel Landsat-1/ancillary variable classification



All Correct Second-Order
Urban Land Use

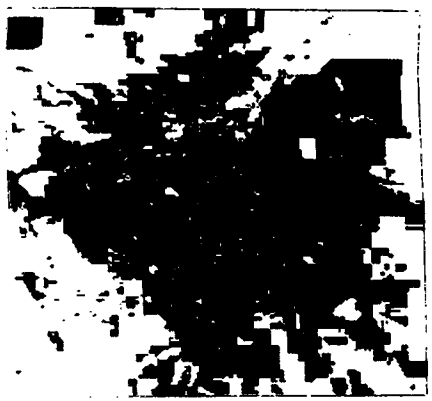


All Correct Second-Order
Agricultural Land Use

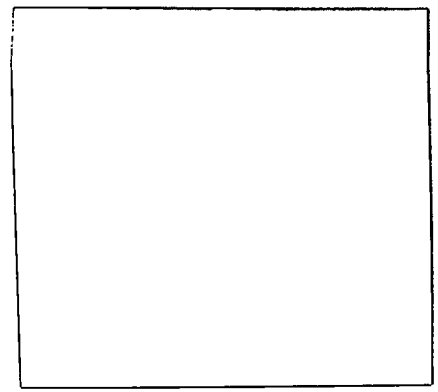


All Correct Second-Order
Urban/Agricultural Land Use

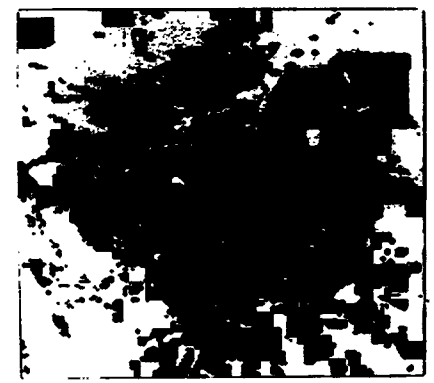
(c) Verified 22-channel Landsat-1/ancillary variable classification



All Correct Second-Order
Urban Land Use



All Correct Second-Order
Agricultural Land Use



All Correct Second-Order
Urban/Agricultural Land Use

(d) 1072-1973 USGS land-use reference data plane

FIGURE 65. MAXIMUM-LIKELIHOOD RATIO (OR GAUSSIAN) IMAGE CLASSIFICATION OF THE DENVER METROPOLITAN AREA VERIFIED AT THE SECOND AND THIRD ORDERS (August 15, 1973, Landsat-1 image). A black (urban) or gray (agricultural) point was printed on the display if the individual cell was correctly classified. The check for correctness was made against the corresponding cell on the 1972-1973 USGS land-use data plane. The classification was made for the 24 second- and third-order land-use classes, and was then checked at the same level. Variables which were used in the classification are identified in table 22. Display scale is ~1:650,000.

ORIGINAL PAGE IS
OF POOR QUALITY

This Page Intentionally Left Blank

TABLE 22

COMPARATIVE ACCURACIES OF MAXIMUM-LIKELIHOOD AND LINEAR-DISCRIMINANT IMAGE-CLASSIFICATION ALGORITHMS. The accuracy of each of these two automated image classifications was checked point-by-point at a first and composite second and third order against the digital 1972-1973 USGS land-use data plane for three different combinations of Landsat-1 image and/or ancillary map variables. The maximum-likelihood classification was performed on the sampled image of 36,864 points, and for economy, the linear-discriminant tests were run on the 4,100 systematically sampled points only.

Number of Variables	First-Order Verification (percent)			Second-Order Verification (percent)		
	Maximum-Likelihood Ratio	Linear-Discriminant Analysis		Maximum-Likelihood Ratio	Linear-Discriminant Analysis	
		Equal Probabilities	*A Priori* Probabilities		Equal Probabilities	*A Priori* Probabilities
Four*	53.6	50.9	65.2	4.3	19.1	37.9
Six†	68.1	70.6	73.1	15.4	32.3	46.3
Twenty-Two‡	61.5	73.1	75.6	2.4	37.2	48.3

*Four original Landsat-1 MSS bands.

†Six channels: MSS-4, MSS-7, MSS-7/MSS-5 ratio, MSS-5/MSS-4 ratio, topographic elevation, and built-up urban-area minimum distance.

‡Twenty-two channels: four original Landsat-1 MSS bands, six Landsat-1 band ratios, four Landsat-1 MSS/insolation ratios, Landsat-1 image insolation, topographic aspect, topographic elevation, composite minor-road MD, composite major-road MD, freeway MD, freeway interchange MD, and built-up urban-area MD (MD = minimum distance).

Second- and Third-Order Theme Maps

The eleven second- and third-order *urban* land-use classes were displayed for the USGS land-use reference data plane, the point-verified discriminant-classified land-use map, and the actual discriminant-classification map (figure 68).

The three third-order *agricultural* land-use classes were displayed for the USGS land-use reference data plane, the point-verified discriminant-classified land-use map, and the actual discriminant-classification map (figure 69).

The three second-order *water* land-use classes were displayed for the USGS land-use reference data plane, the point-verified discriminant-classified land-use map, and the actual discriminant-classification land-use map (figure 70).

The two second-order *range* land-use classes were displayed for the USGS land-use reference data plane, the point-verified discriminant-classified land-use map, and the actual discriminant-classification land-use map (figure 71).

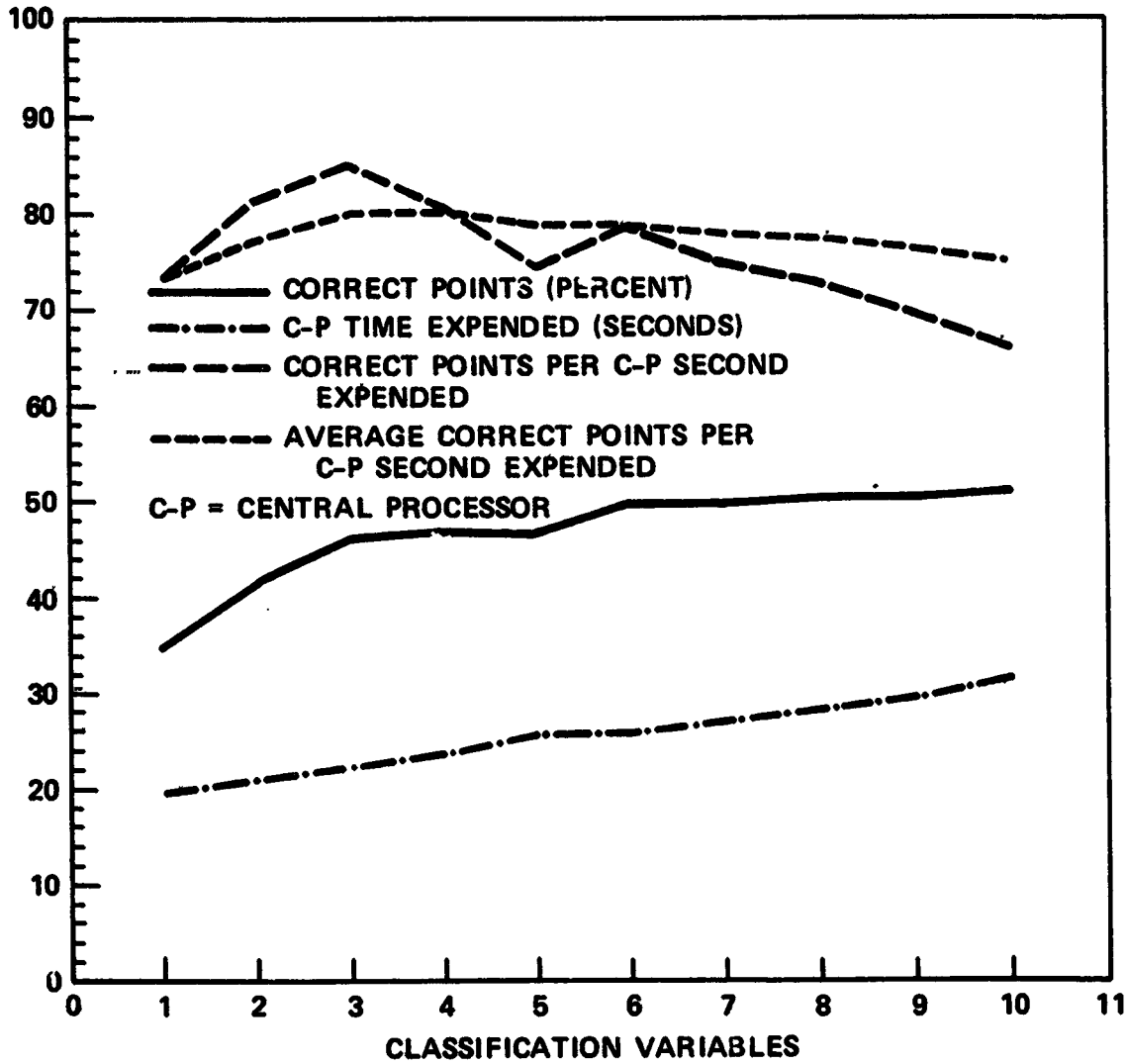


FIGURE 66. TEST OF FINAL TEN-CHANNEL COMBINATION OF LANDSAT AND LANDSCAPE VARIABLES FOR THE FULL IMAGE CLASSIFICATION. This figure is based on identifying 24 land uses (table 1) with 4,100 grid-sampled picture elements. This ten-variable combination was selected as an optimal subset of the 41 available image/ancillary variables. It realized over 94 percent of the maximum accuracy that could have been obtained with all 41 variables, but represented an expenditure of only 34 percent of the processing time. All variables were added in a free stepwise fashion and were classified using linear-discriminant analysis.

TABLE 23

TEST OF FINAL TEN-CHANNEL COMBINATION OF LANDSAT AND LANDSCAPE VARIABLES FOR FULL IMAGE CLASSIFICATION (August 15, 1973 image). This table is based on identifying 24 land uses (table 1) with 4,100 grid-sampled picture elements. This ten-variable subset retained over 94 percent of the full 41-variable training-set accuracy while expending only 34 percent of the computer time. All variables were added in a free stepwise fashion and were classified using linear-discriminant analysis.

Step Number	Landsat and Landscape Variables Entered	Training-Set Classification		C-P* Time Expended (seconds)	Correct Points/C-P Second Expended (%)		F-Value to Enter
		Total Points Correct	Correct Points (%)		Step	Average	
1	Topographic elevation	1,428	34.83	19.51	73.19	73.19	315.73
2	Built-up urban-area MD†	1,701	41.49	20.92	81.31	77.39	88.79
3	Average number of families per acre	1,892	46.15	22.26	85.00	80.09	75.66
4	Landsat-1 MSS-7 (solar infrared)	1,921	46.85	23.76	80.85	80.30	61.53
5	Topographic slope	1,911	46.61	25.66	74.47	78.97	57.42
6	Average number of year-round housing units per acre	2,038	49.71	25.95	78.54	78.89	40.50
7	Landsat-1 MSS-5/MSS-4 ratio	2,035	49.63	27.16	74.93	78.24	29.75
8	Median housing-unit value	2,059	50.22	28.26	72.86	77.45	26.33
9	Landsat-1 MSS-7/MSS-5 ratio	2,063	50.32	29.55	69.81	76.44	20.44
10	Landsat-1 MSS-4 (visible green)	2,091	51.00	31.66	66.05	75.15	21.77

*C-P = central processor.

†MD = minimum distance.

The second- and third-order *barren* land-use classes were displayed for the USGS land-use reference data plane, the point-verified discriminant-classified land-use map, and the actual discriminant-classfiication land-use map (figure 72).

The three third-order *forest* land-use classes were displayed for the USGS land-use reference data plane, the point-verified discriminant-classified land-use map, and the actual discriminant-classification land-use map (figure 73).

Verification

The verification of classification results in this study was a highly structured procedure. An advantage of the quantitative orientation taken throughout this endeavor was the explicit quantitative verification of results. The great utility of the computer lay not only in the capability of massive data manipulation, but also in the face-to-face confrontation of the uncertainties in the results and in a conscientious examination of the various sources of possible errors.

This Page Intentionally Left Blank

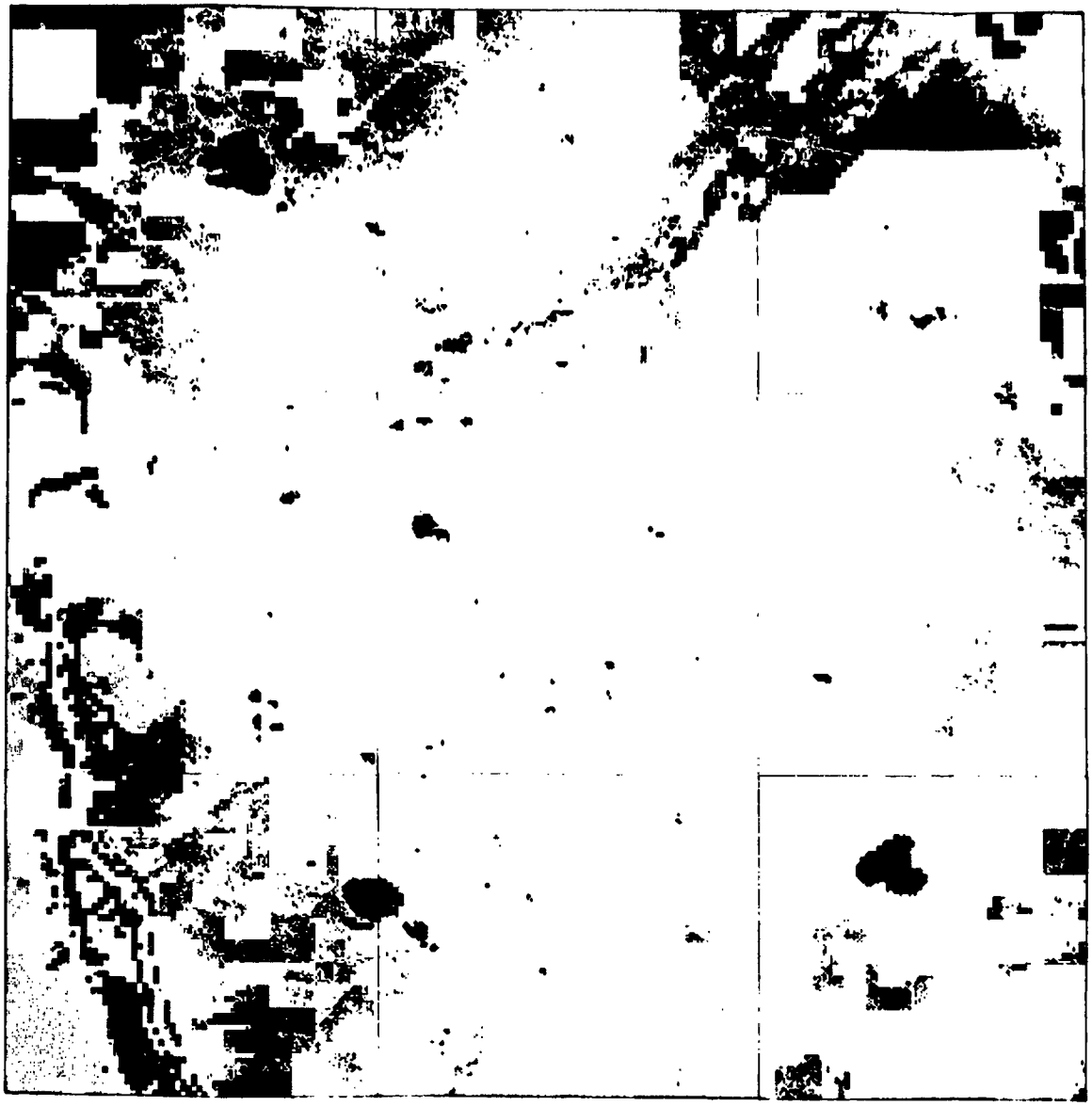
Display scale 1:250,000

(a) Six first-order land-use classes identified by USGS airphoto interpretation.

BLACK = All water areas
DARKEST GRAY = All range areas
DARK GRAY = All agricultural areas
GRAY = All forest areas
LIGHT GRAY = All barren areas
LIGHTEST GRAY = All urban areas

ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME



Display scale 1:250,000

(b) Verification map of the six first-order classifications (c) where points are plotted only if they check with USGS map (a).

BLACK = All water areas (62.0% correct)
DARKEST GRAY = All range areas (56.3% correct)
DARK GRAY = All agricultural areas (56.7% correct)
GRAY = All forest areas (66.3% correct)
LIGHT GRAY = All barren areas (35.0% correct)
LIGHTEST GRAY = All urban areas (89.3% correct)

FOLDOUT FRAME

2

ORIGINAL PAGE IS
OF POOR QUALITY

Display scale 1:250,000

(c) Six first-order land-use classes identified using discriminant analysis with four image and six ancillary variables.

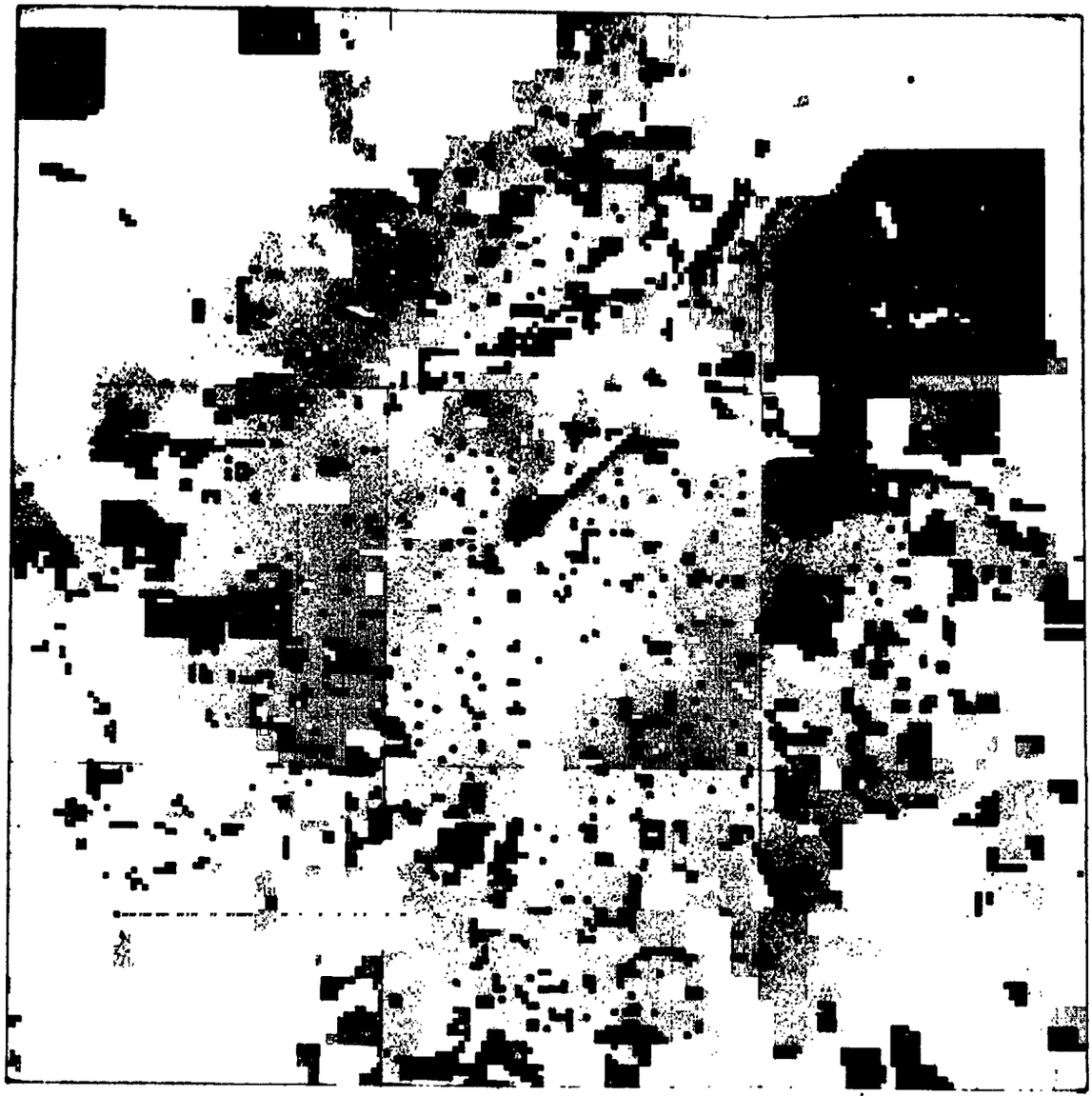
BLACK = All water areas
DARKEST GRAY = All range areas
DARK GRAY = All agricultural areas
GRAY = All forest areas
LIGHT GRAY = All barren areas
LIGHTEST GRAY = All urban areas

FIGURE 67. COMPARATIVE DISPLAYS OF THE DISCRIMINANT CLASSIFICATION, A VERIFICATION OF THAT CLASSIFICATION, AND THE USGS MAP OF THE SIX *FIRST-ORDER LAND-USES* OF THE DENVER AREA. Scale 1:250,000.

ORIGINAL PAGE IS OF POOR QUALITY

FOLDOUT FRAME 3

This Page Intentionally Left Blank.



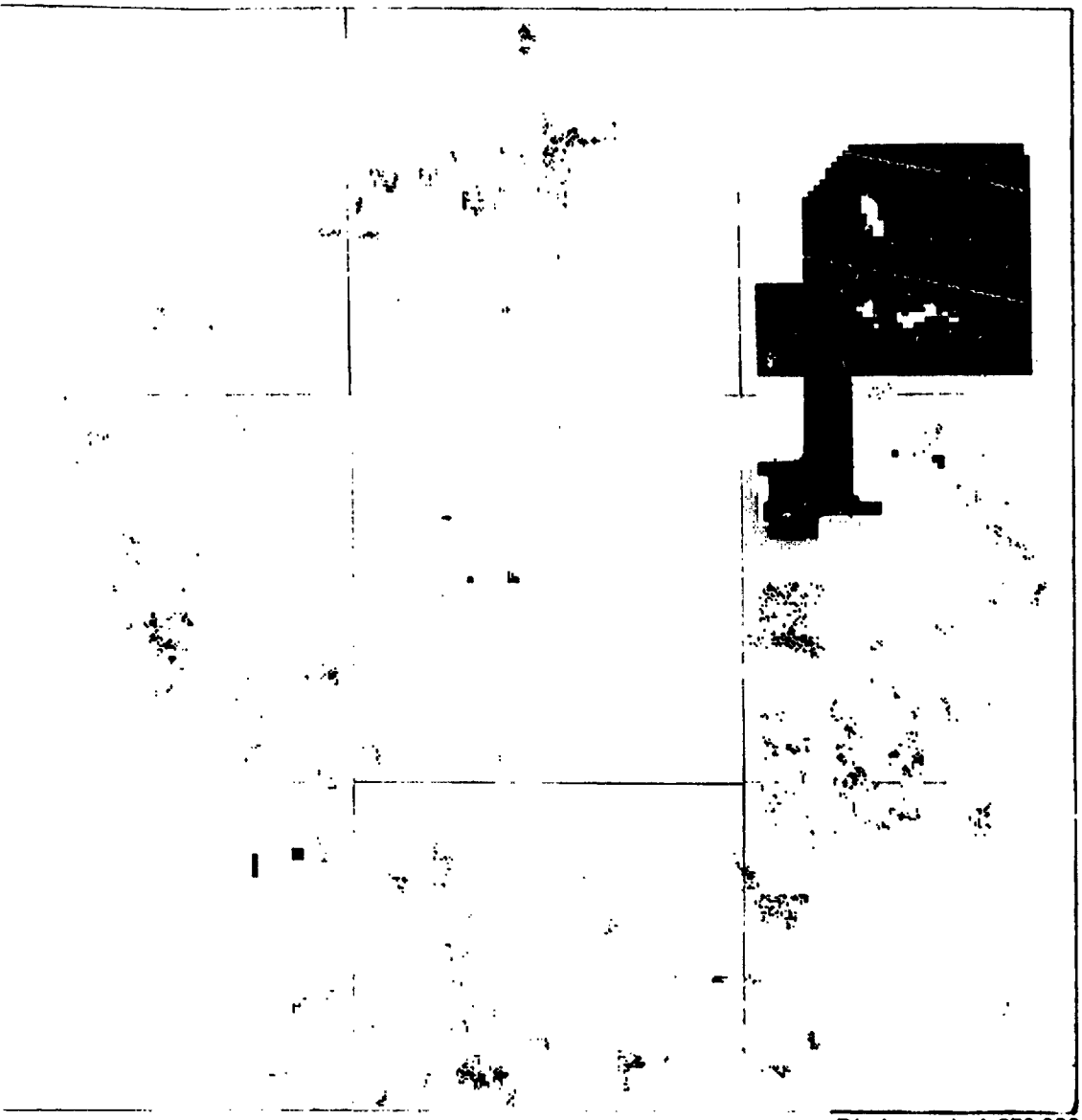
Display scale 1:250,000

(a) Eleven second- and third-order urban land-use classes identified by USGS airphoto interpretation.

ORIGINAL PAGE IS
OF POOR QUALITY

BLACK = Industrial/transportation
DARK GRAY = Cemetery/recreational/open land
MEDIUM GRAY = Utility/public and institutional
GRAY = Solid-waste dump/extraction
LIGHT GRAY = Residential/commercial and services

FOLDOUT FRAME



Display scale 1:250,000

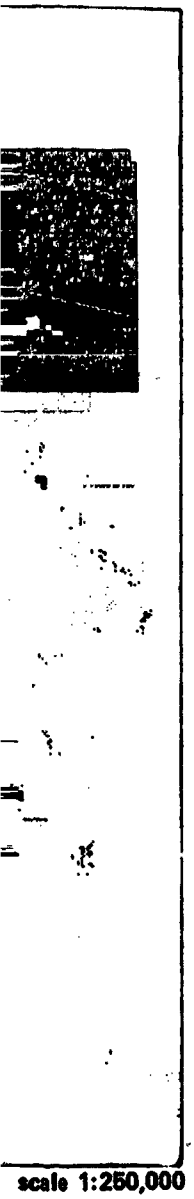
(b) Verification map of the eleven second- and third-order urban land-use classifications (c) where points are plotted only if they check with USGS map (a).

(c) Eleven se

BLACK = Industrial/transportation (42.1% correct)
DARK GRAY = Cemetery/recreational/open land (0.1% correct)
MEDIUM GRAY = Utility/public and institutional (49.2% correct)
GRAY = Solid-waste dump/extraction (0.0% correct)
LIGHT GRAY = Residential/commercial and services (78.0% correct)

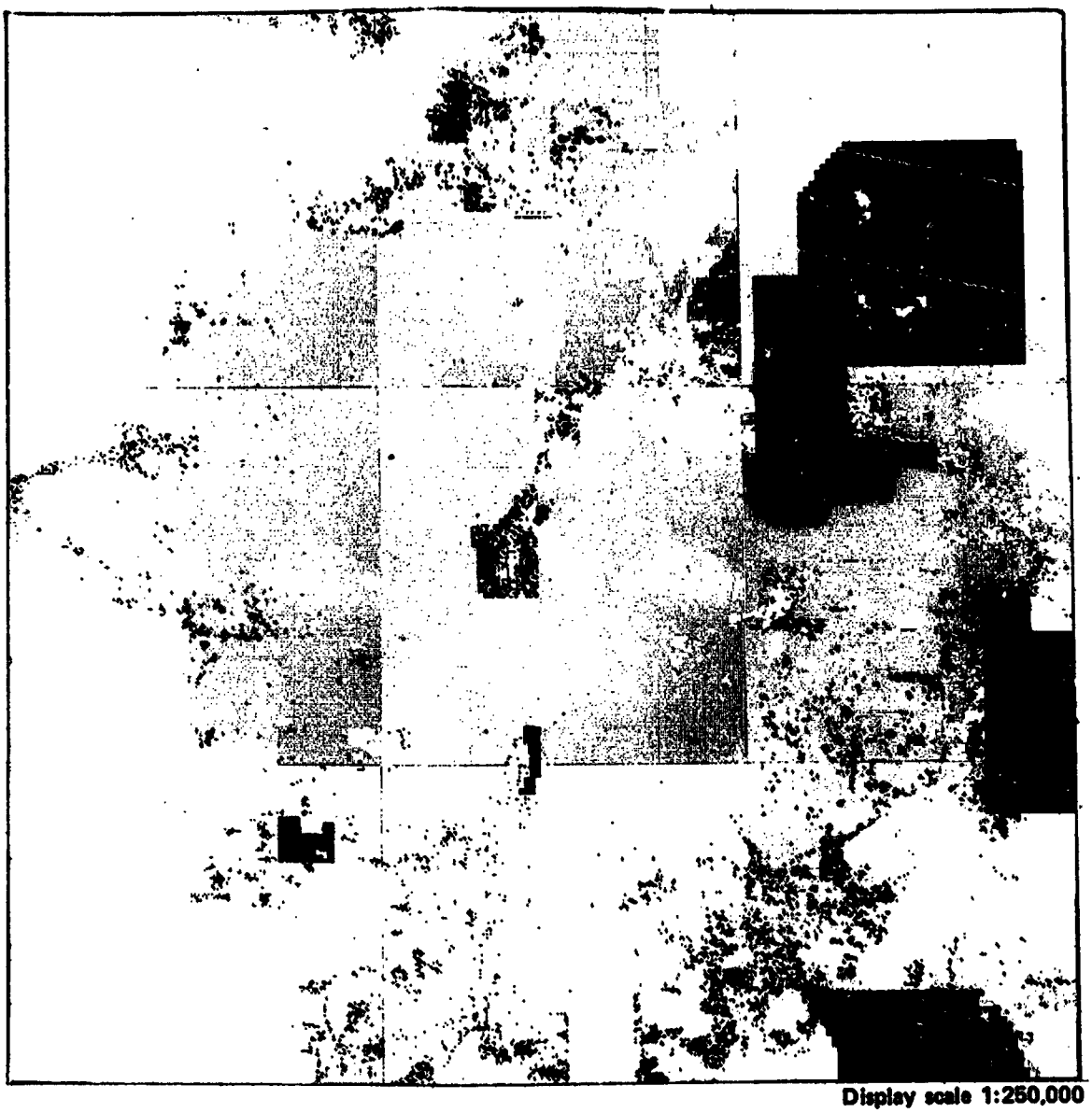
ORIGINAL PAGE IS
OF POOR QUALITY

FOLDOUT FRAME



scale 1:250,000

ons (c)



Display scale 1:250,000

(c) Eleven second- and third-order urban land-use classes identified using discriminant analysis with four image and six ancillary variables.

BLACK = Industrial/transportation
DARK GRAY = Cemetery/recreational/open land
MEDIUM GRAY = Utility/public and institutional
GRAY = Solid-waste dump/extraction
LIGHT GRAY = Residential/commercial and services

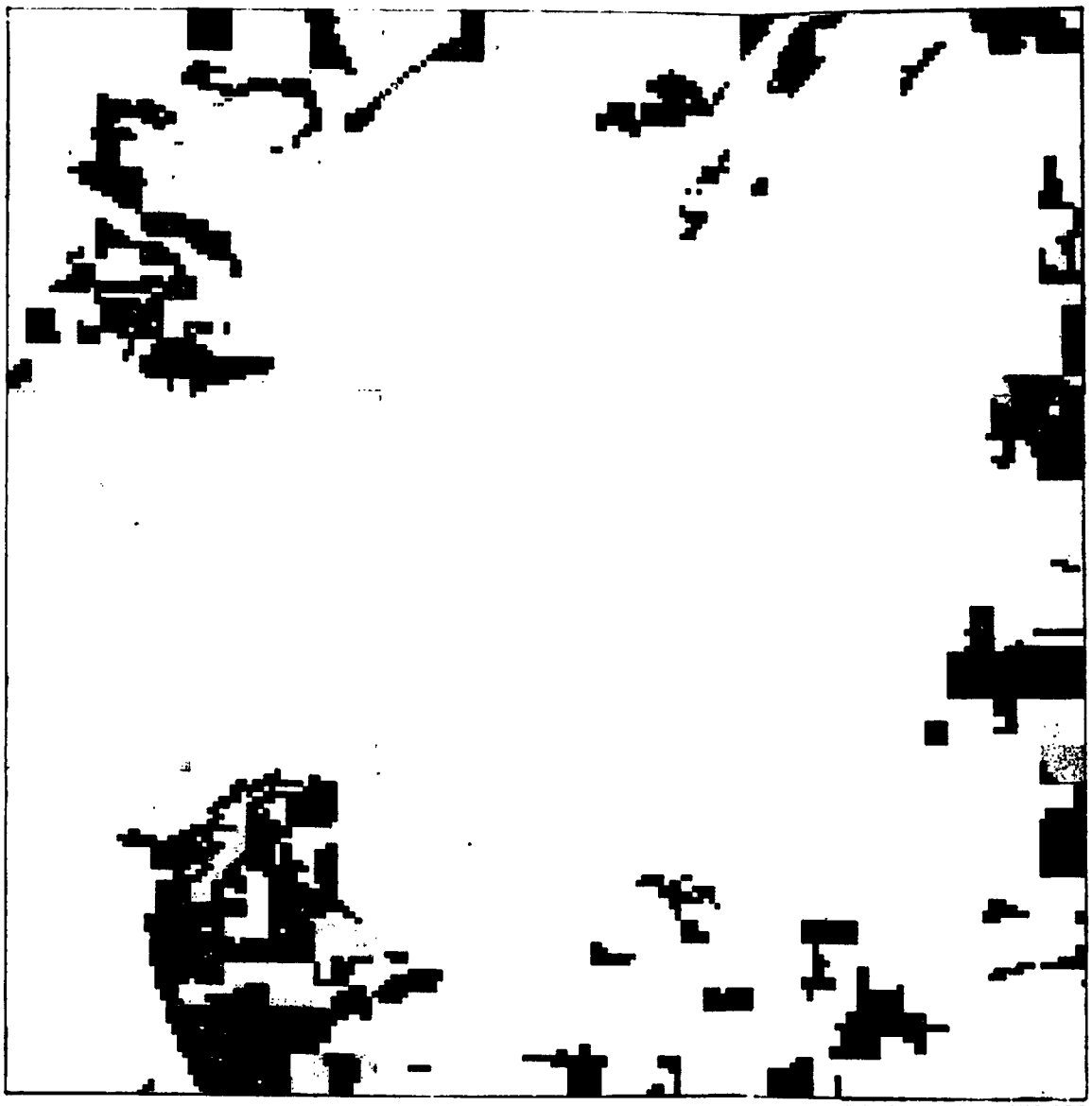
ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 68. COMPARATIVE DISPLAYS OF THE DISCRIMINANT CLASSIFICATION, A VERIFICATION OF THAT CLASSIFICATION, AND THE USGS MAP OF THE *URBAN LAND USES* OF THE DENVER AREA. Scale 1:250,000.

3

FOLDOUT FRAME

This Page Intentionally Left Blank



Display scale 1:250,000

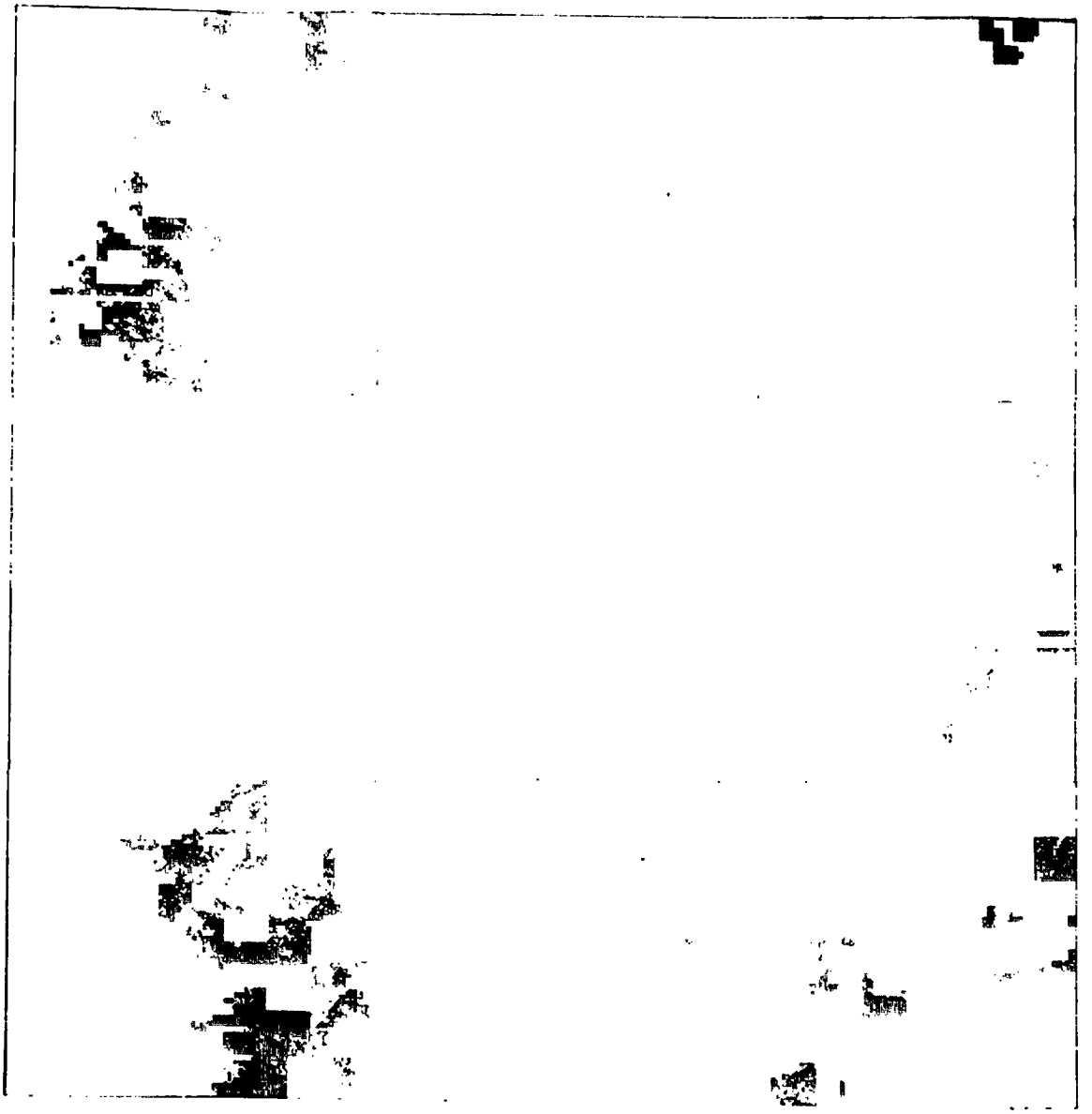
(a) Three second-order agricultural land-use classes identified by USGS airphoto interpretation.

(b) Ver

BLACK = Irrigated cropland
DARK GRAY = Pasture
LIGHT-GRAY = Nonirrigated cropland

FOLDOUT FRAME

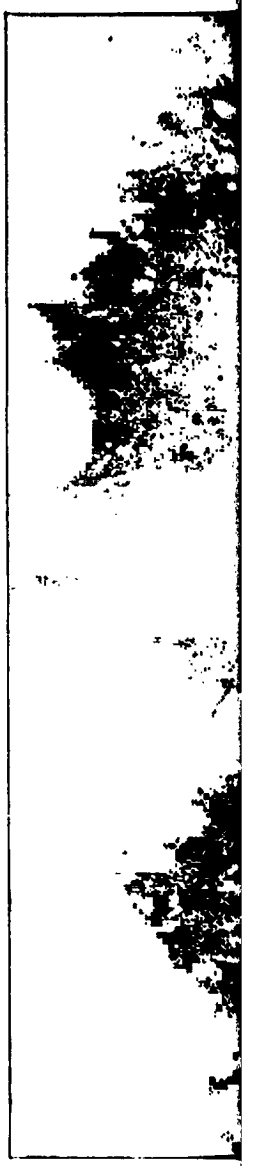
ORIGINAL PAGE IS
POOR QUALITY



Display scale 1:250,000

(b) Verification map of the three second-order agricultural land-use classifications (c) where points are plotted only if they check with USGS map (a).

BLACK = Irrigated cropland (65.4% correct)
DARK GRAY = Pasture (28.5% correct)
LIGHT GRAY = Nonirrigated cropland (38.8% correct)

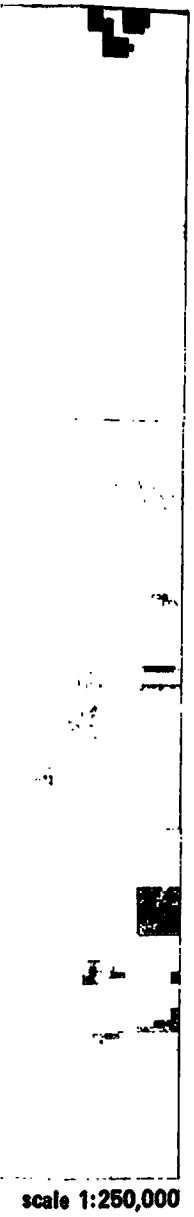


(c) Th

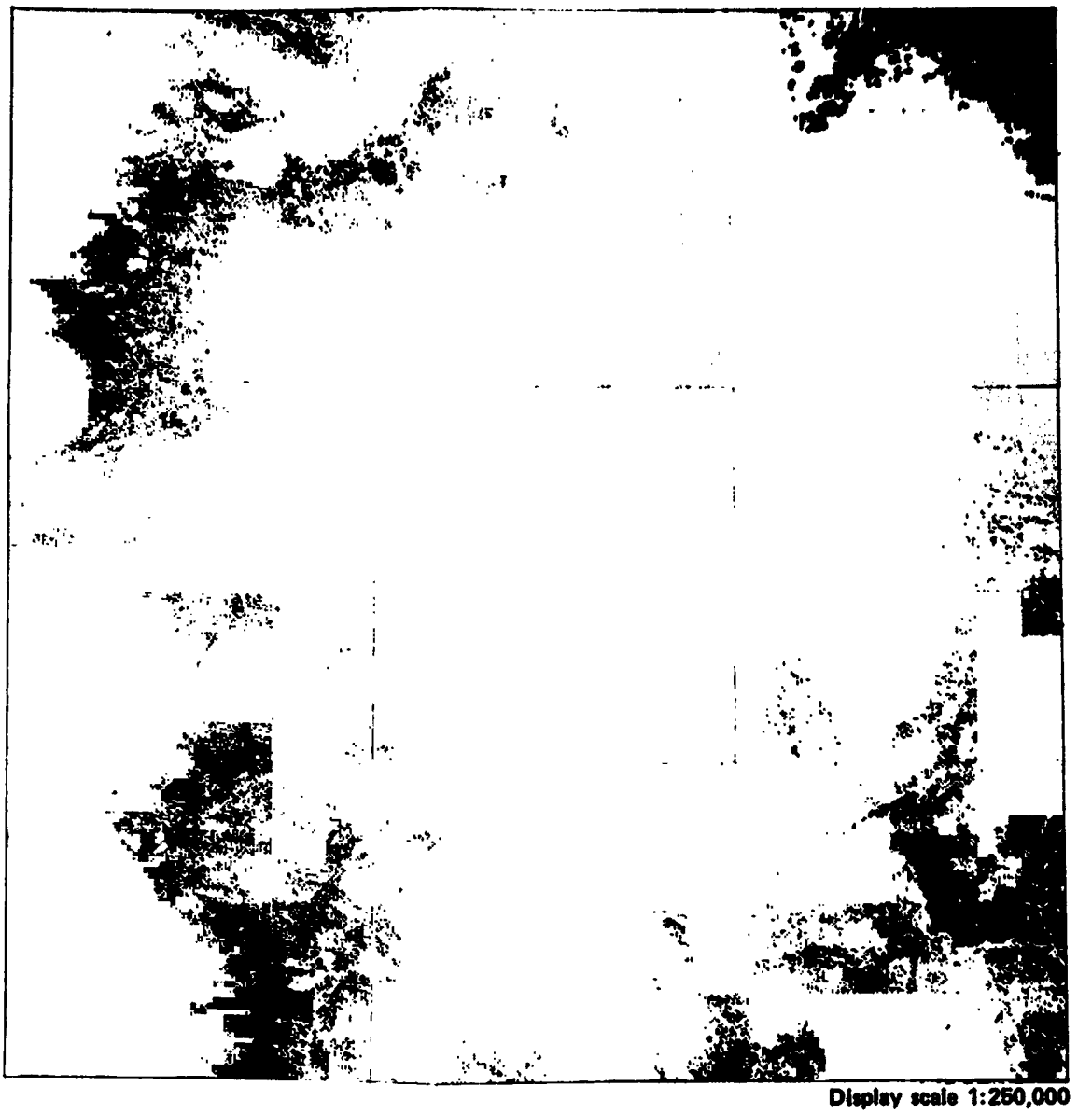
2

BOLDOUT FRAME

ORIGINAL PAGE IS
OF POOR QUALITY



scale 1:250,000



Display scale 1:250,000

-ns (c)

(c) Three second-order agricultural land-use classes identified using discriminant analysis with four image and six ancillary variables.

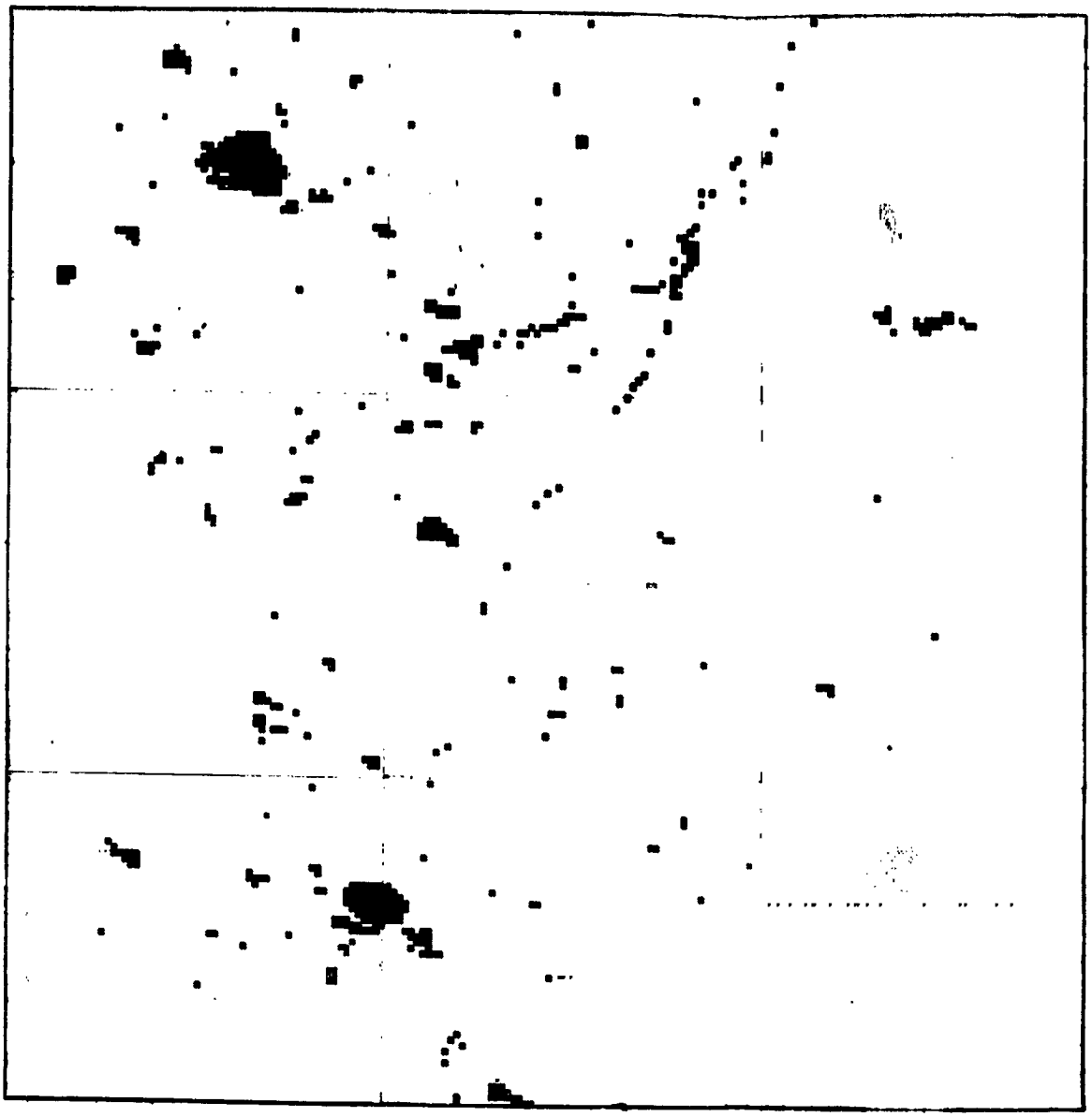
BLACK = Irrigated cropland
DARK GRAY = Pasture
LIGHT GRAY = Nonirrigated cropland

FIGURE 69. COMPARATIVE DISPLAYS OF THE DISCRIMINANT CLASSIFICATION, A VERIFICATION OF THAT CLASSIFICATION, AND THE USGS MAP OF THE *AGRICULTURAL LAND-USES* OF THE DENVER AREA. Scale 1:250,000.

3 [redacted] 1

ORIGINAL PAGE
OF POOR QUALITY

This Page Intentionally Left Blank



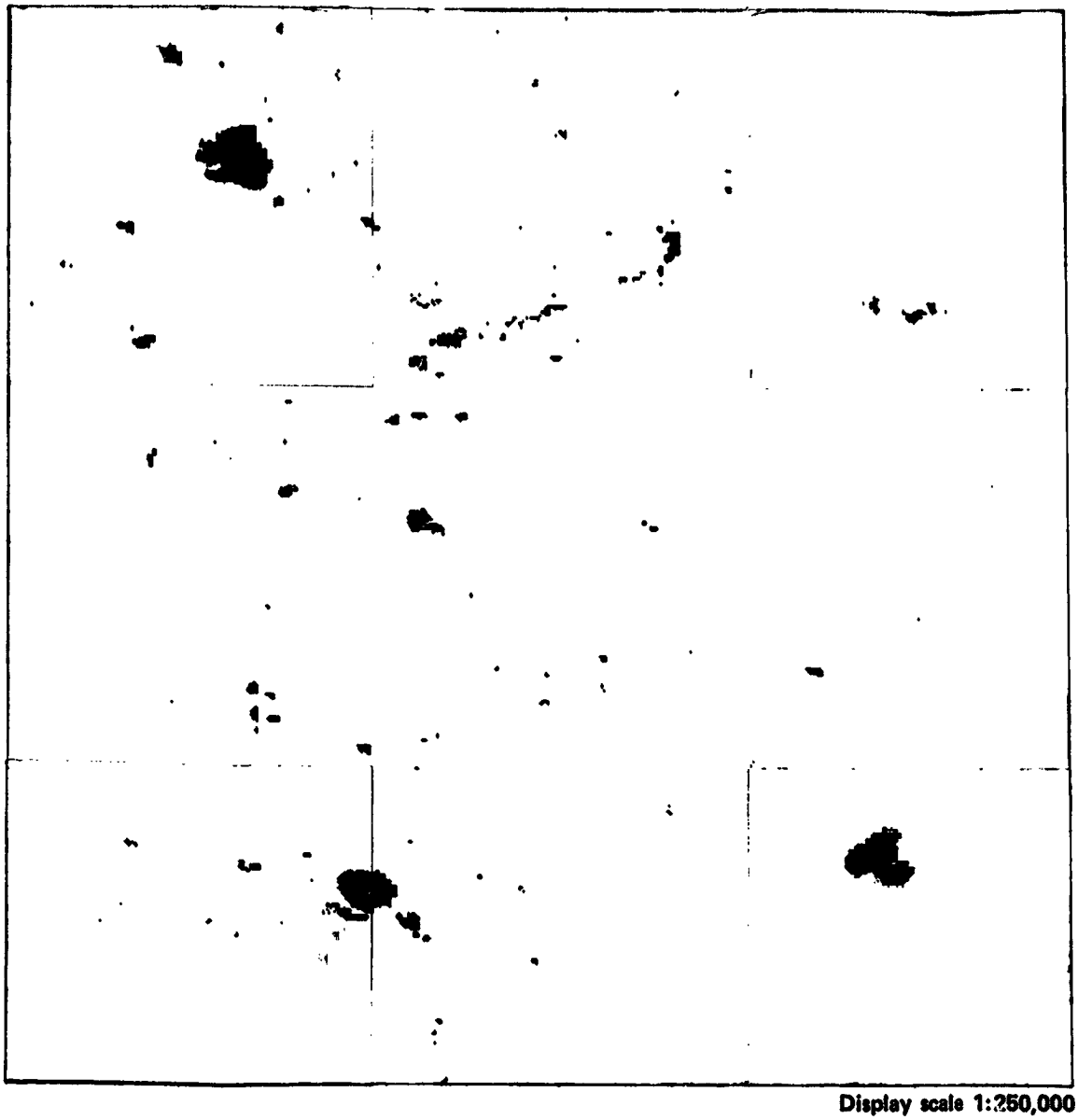
Display scale 1:250,000

(a) Three second-order water-type classes identified by USGS airphoto interpretation.

ORIGINAL PAGE IS
OF POOR QUALITY

BLACK = Streams and waterways
DARK GRAY = Lakes
LIGHT GRAY = Reservoirs

FOLDOUT FRAME



Display scale 1:250,000

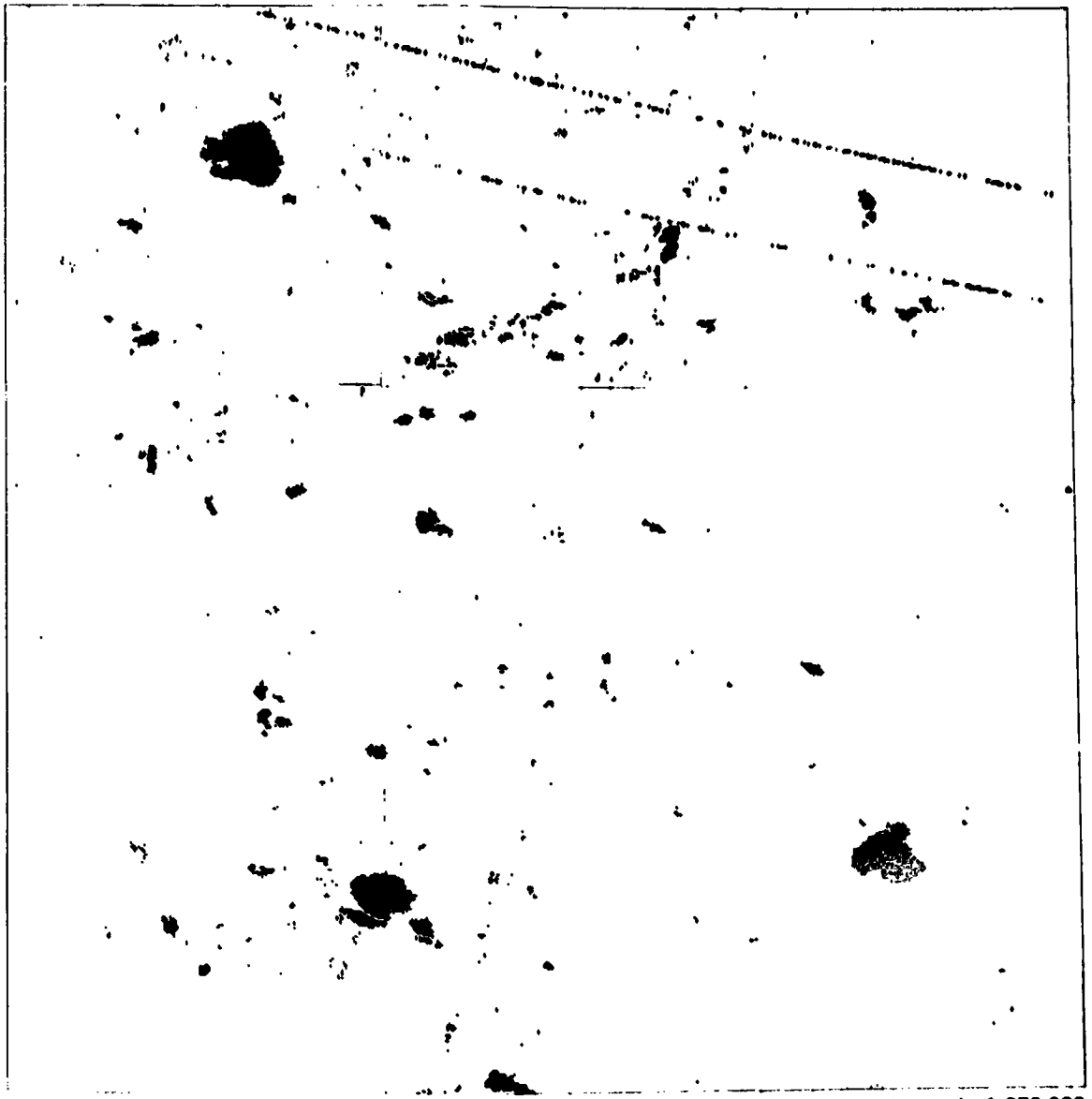
(b) Verification map of the three second-order water-type classifications (c) where points are plotted only if they check with USGS map (a).

BLACK = Streams and waterways (0.35% correct)
DARK GRAY = Lakes (56.0% correct)
LIGHT GRAY = Reservoirs (58.1% correct)

ORIGINAL PAGE IS
OF POOR QUALITY

2

FOLDOUT FRAME



Display scale 1:250,000

(c) Three second-order water-type classes identified using discriminant analysis with four image and six ancillary variables.

BLACK = Streams and waterways
DARK GRAY = Lakes
LIGHT GRAY = Reservoirs

ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 70. COMPARATIVE DISPLAYS OF THE DISCRIMINANT CLASSIFICATION, A VERIFICATION OF THAT CLASSIFICATION, AND THE USGS MAP OF THE *WATER-TYPE CLASSES* OF THE DENVER AREA. Scale 1:250,000.

FOLDOUT FRAME

This Page Intentionally Left Blank

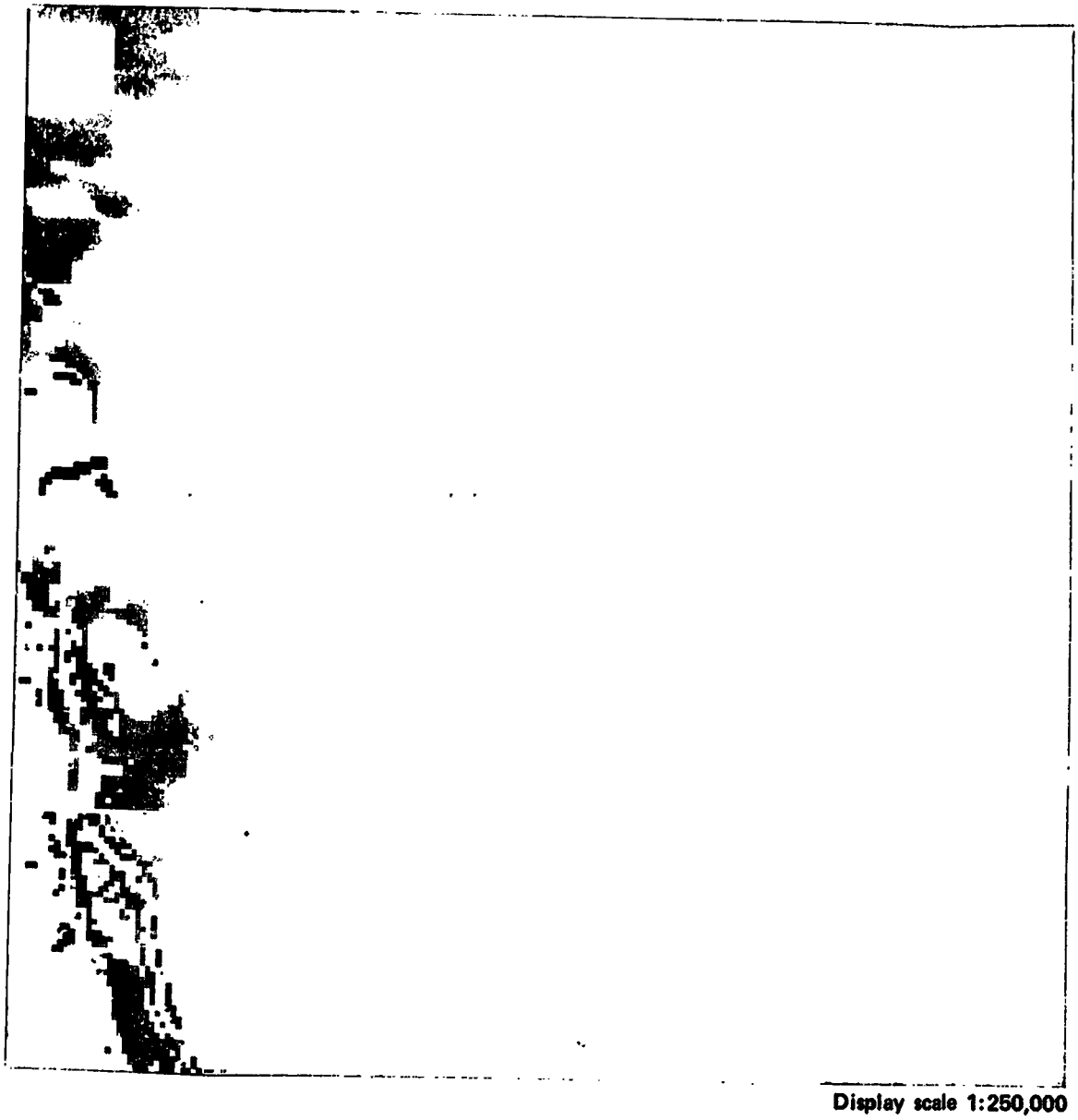
Display scale 1:250,000

(a) Two second-order rangeland classes identified by USGS airphoto interpretation.

BLACK = Chapparal (taken as brushland)
DARK GRAY = Grassland

FOLDOUT FRAME

ORIGINAL PAGE IS
OF POOR QUALITY



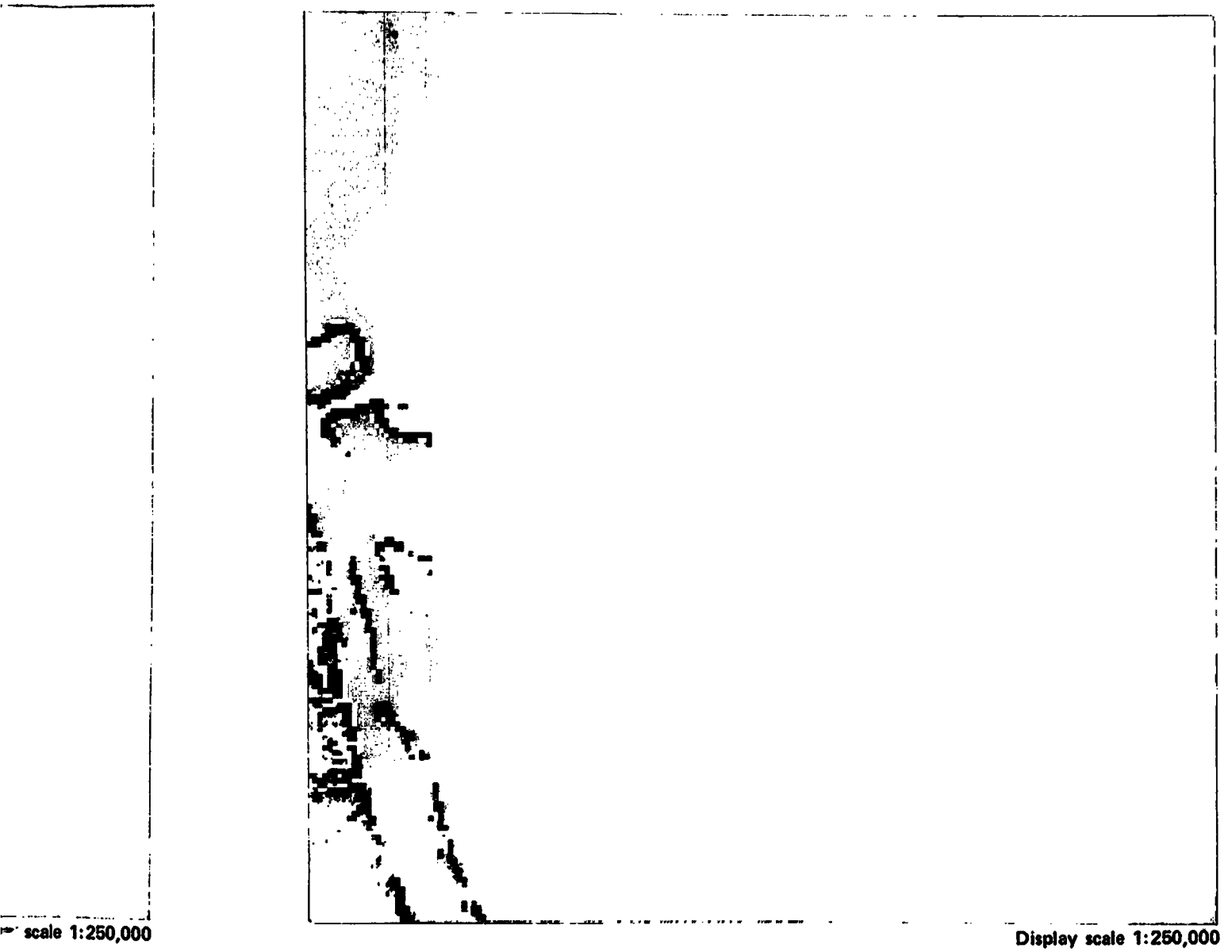
Display scale 1:250,000

(b) Verification map of the two second-order rangeland classifications (c) where points are plotted only if they check with USGS map (a).

BLACK = Chapparal (taken as brushland) (40.9% correct)
DARK GRAY = Grassland (54.9% correct)

FOLDOUT FRAME

ORIGINAL PAGE IS
OF POOR QUALITY



scale 1:250,000

Display scale 1:250,000

(c) Two second-order rangeland classes identified using discriminant analysis with four image and six ancillary variables.

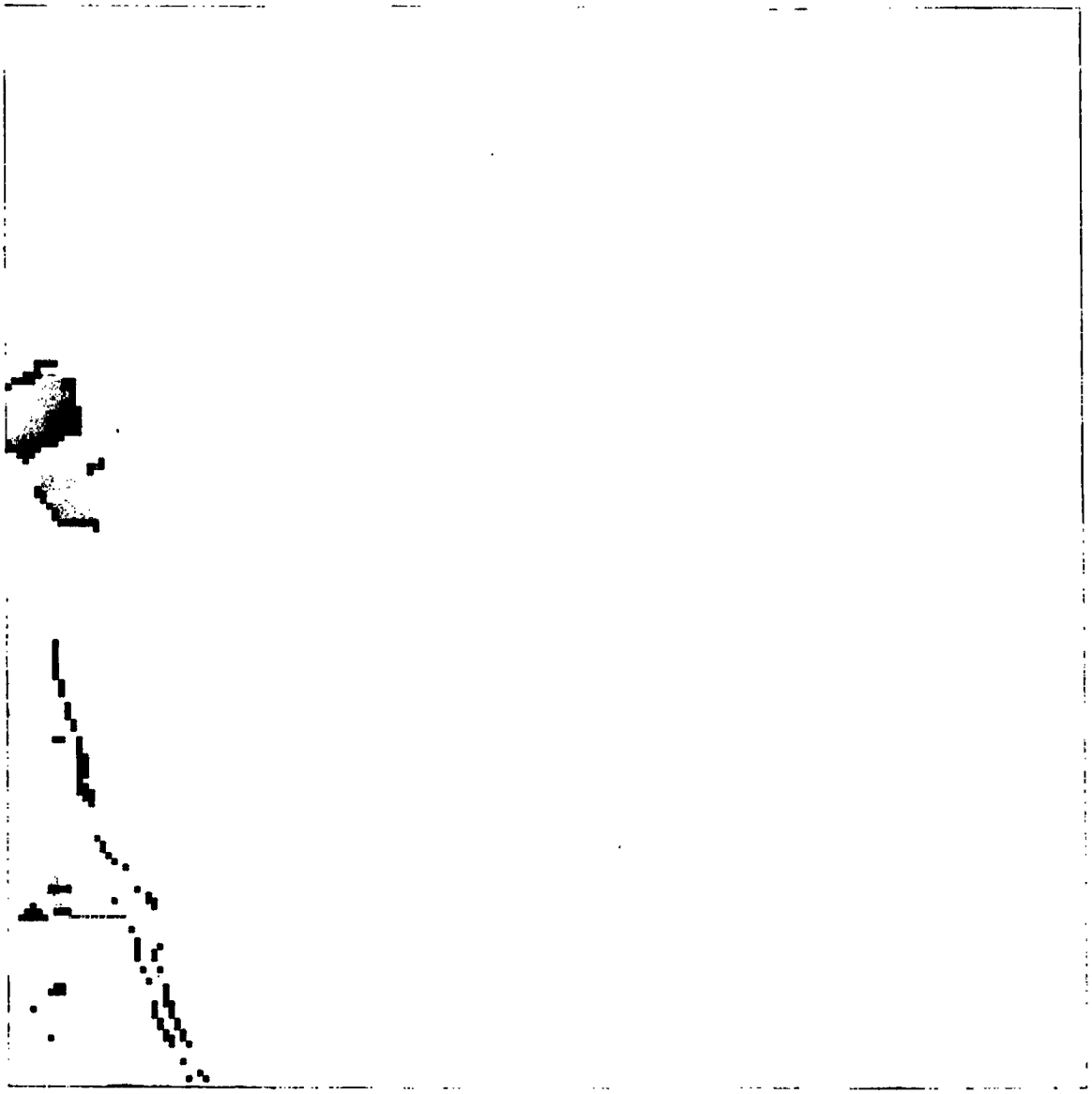
BLACK = Chapparal (taken as brushland)
DARK GRAY = Grassland

ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 71. COMPARATIVE DISPLAYS OF THE DISCRIMINANT CLASSIFICATION, A VERIFICATION OF THAT CLASSIFICATION, AND THE USGS MAP OF THE *RANGELANDS* OF THE DENVER AREA. Scale 1:250,000.

3 FOLDOUT
FOLDOUT FRAME

This Page Intentionally Left Blank



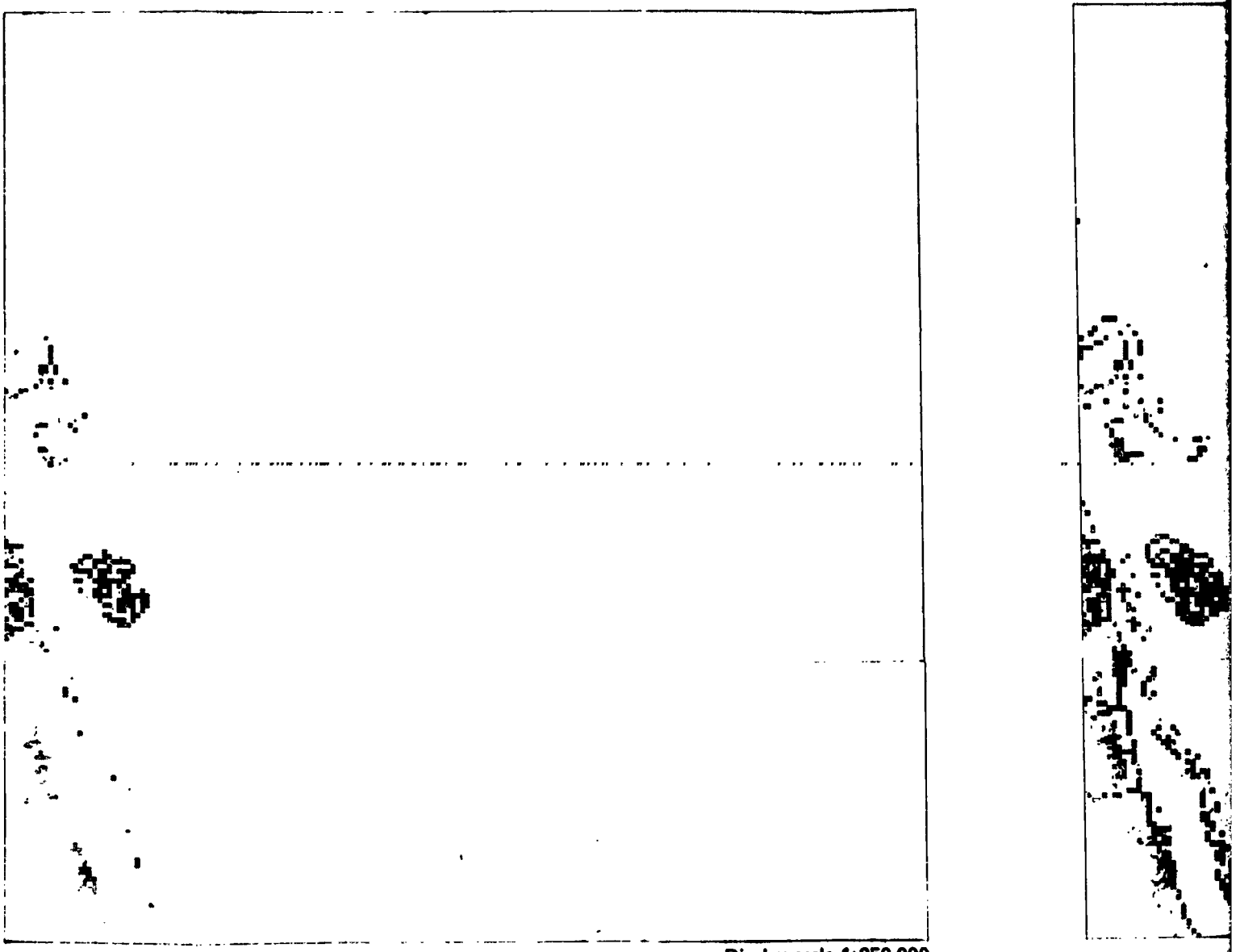
Display scale 1:250,000

(a) Two second-order barrenland classes identified by USGS airphoto interpretation.

BLACK = Exposed rock
DARK GRAY = Hillslopes

FOLDOUT FRAME

ORIGINAL PAGE
OF POOR QUALITY



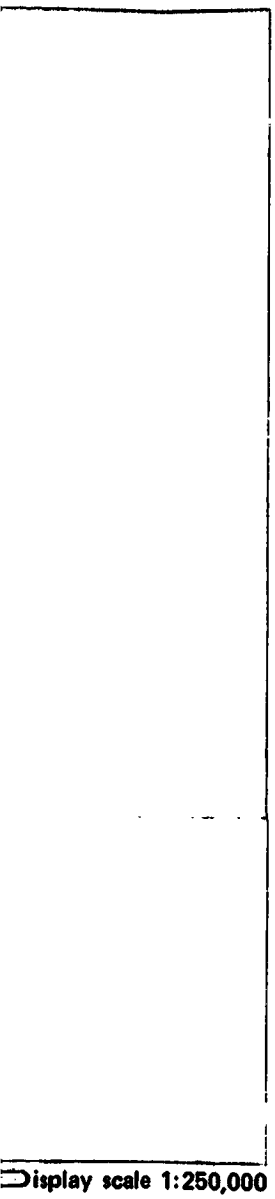
Display scale 1:250,000

(b) Verification map of the two second-order barrenland classifications (c) where points are plotted only if they check with USGS map (a).

BLACK = Exposed rock (8.7% correct)
DARK GRAY = Hillslopes (38.5% correct)

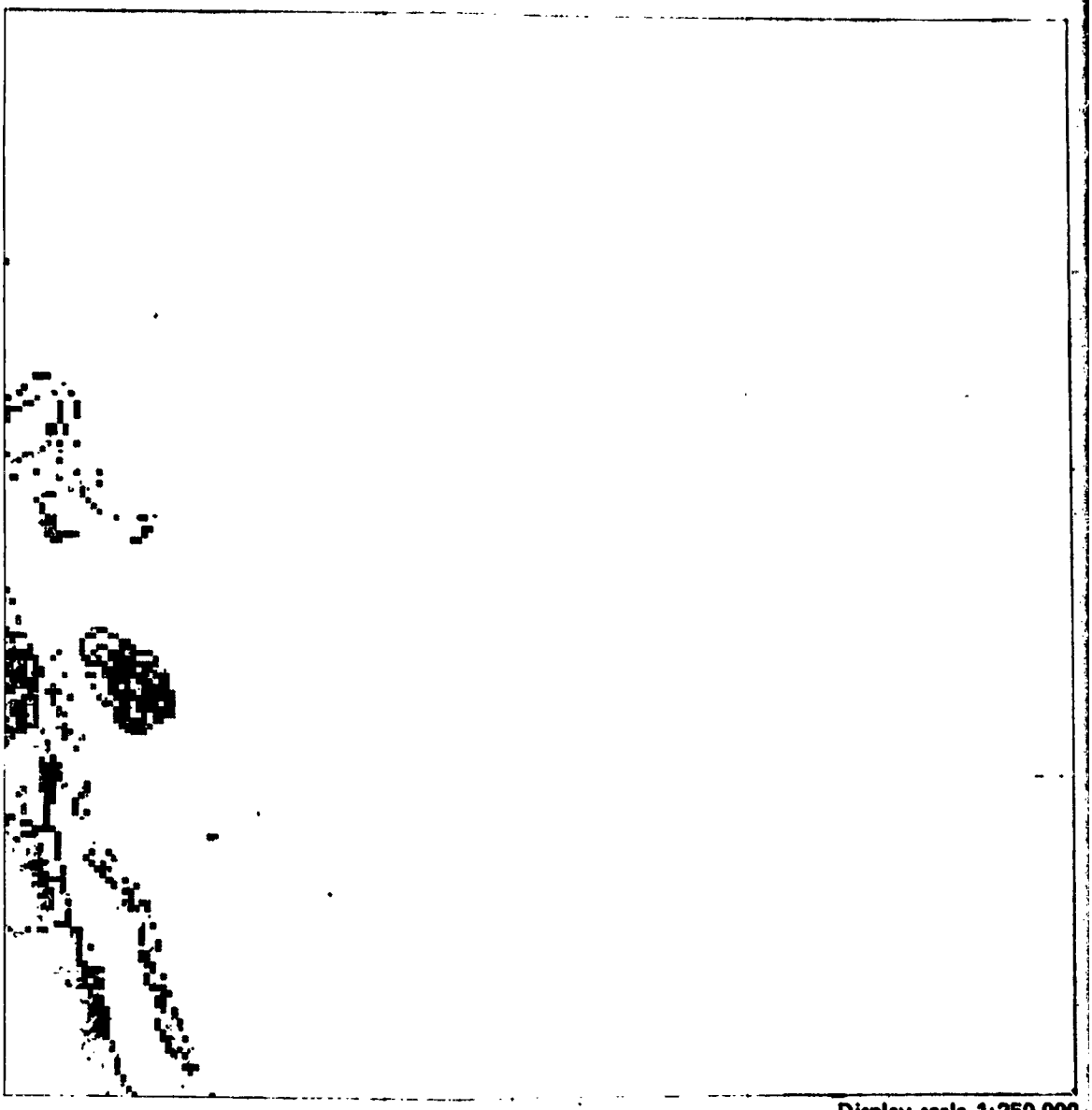
FOLDOUT FRAME

ORIGINAL PAGE IS OF POOR QUALITY



Display scale 1:250,000

ons (c)
).



Display scale 1:250,000

(c) Two second-order barrenland classes identified using discriminant analysis with four image and six ancillary variables.

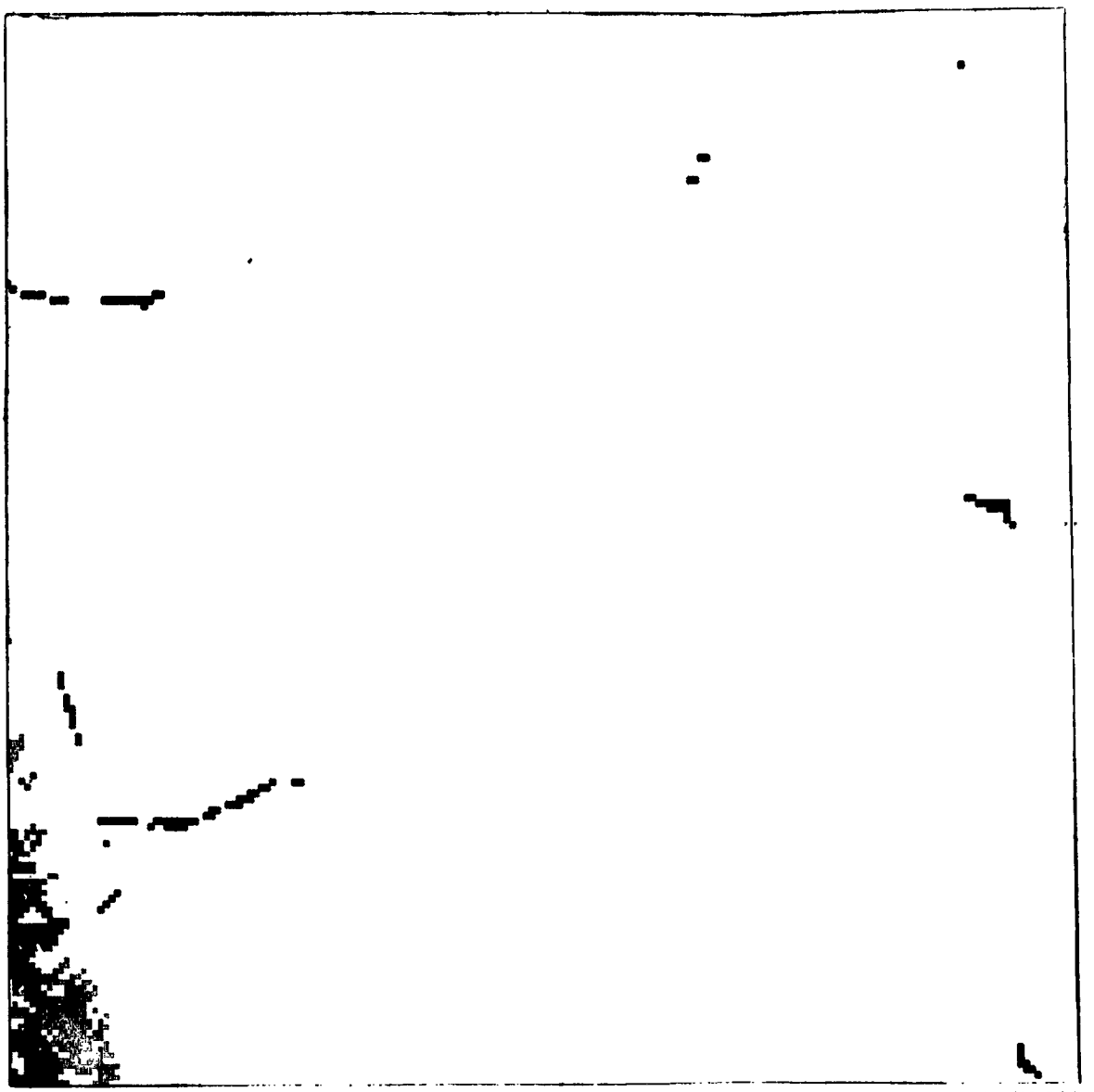
BLACK = Exposed rock
DARK GRAY = Hillslopes

FIGURE 72. COMPARATIVE DISPLAYS OF THE DISCRIMINANT CLASSIFICATION, A VERIFICATION OF THAT CLASSIFICATION, AND THE USGS MAP OF THE *BARRENLANDS* OF THE DENVER AREA. Scale 1:250,000.

3 FOLDOUT FRAME

ORIGINAL PAGE IS OF POOR QUALITY

This Page Intentionally Left Blank

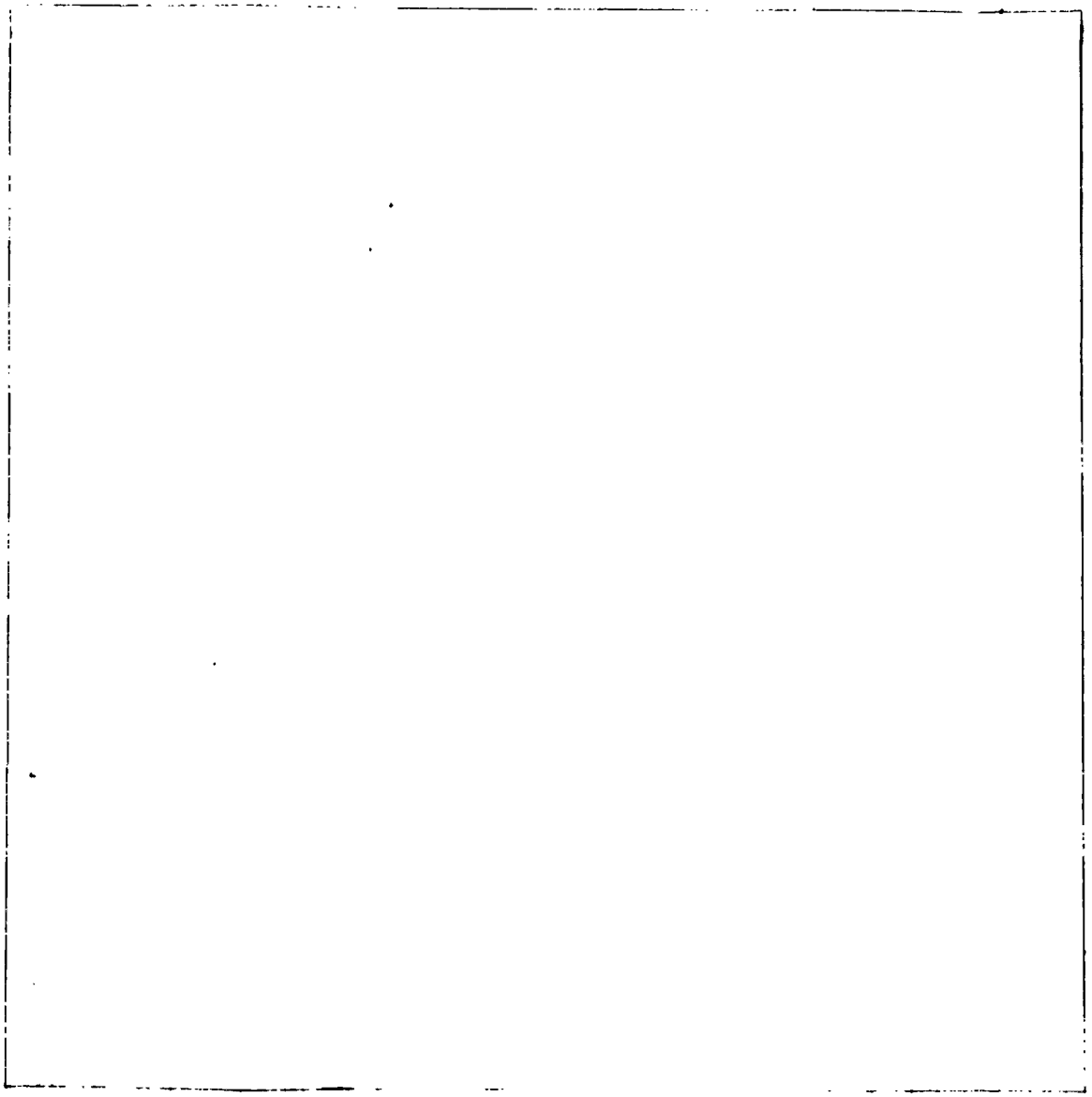


Display scale 1:250,000

(a) Three second-order forestland classes identified by USGS airphoto interpretation.

BLACK = Deciduous/intermediate crown
DARK GRAY = Coniferous/intermediate crown
LIGHT GRAY = Coniferous/solid crown

FOLDOUT FRAME
ORIGINAL PAGE
OF POOR QUALITY

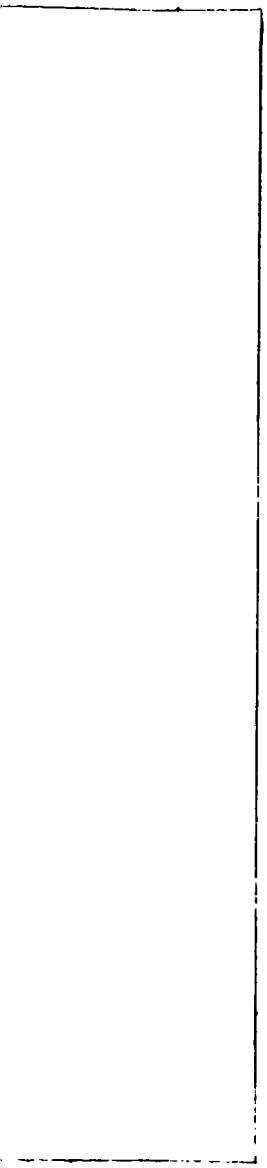


Display scale 1:250,000

(b) Verification map of the three second-order forestland classifications (c) where points are plotted only if they check with USGS map (a).

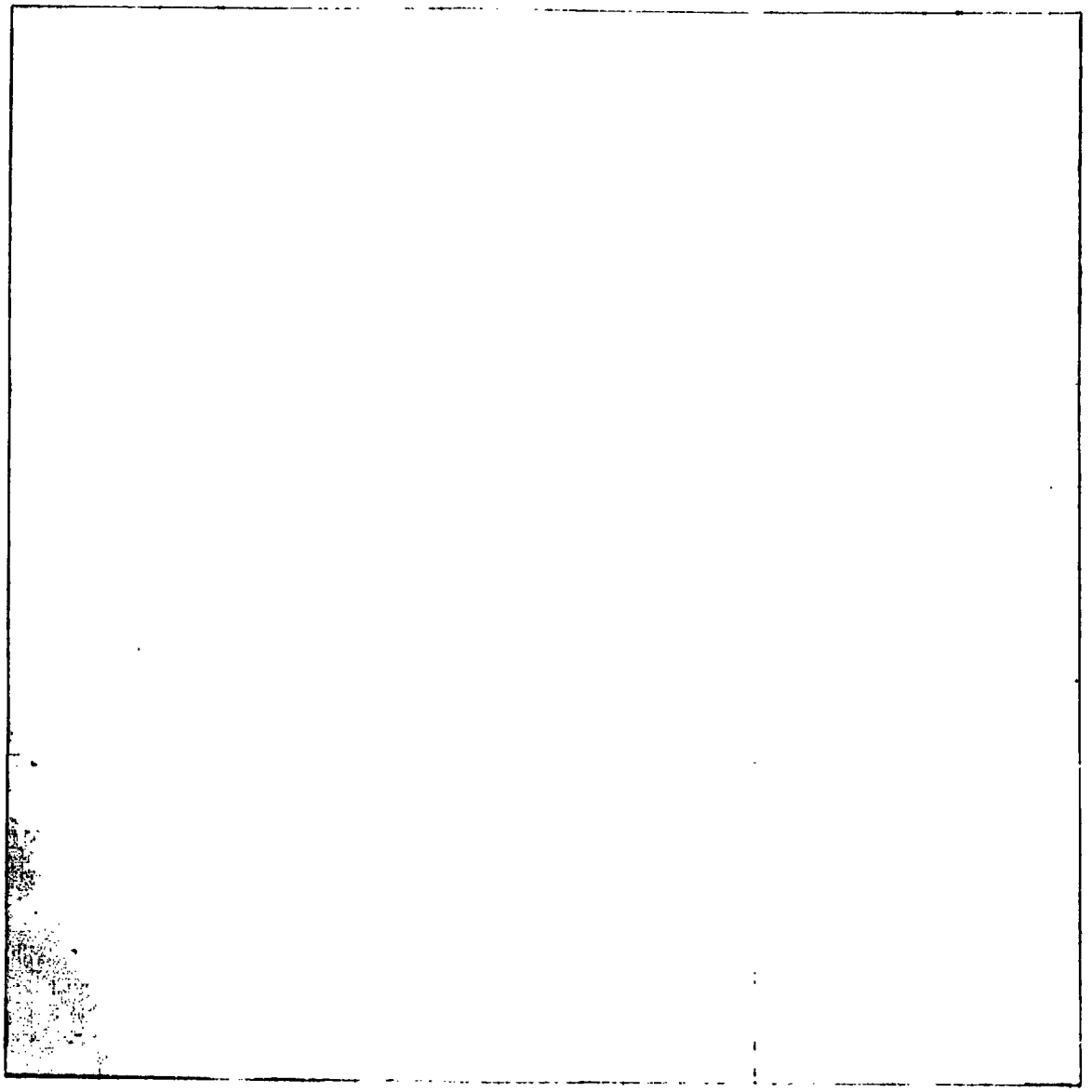
BLACK = Deciduous/intermediate crown (0.0% correct)
DARK GRAY = Coniferous/intermediate crown (2.8% correct)
LIGHT GRAY = Coniferous/solid crown (81.3% correct)

2 FOLDOUT FRAME ORIGINAL PAGE OF POOR QUALITY



isplay scale 1:250,000

ns (c)



Display scale 1:250,000

(c) Two second-order forestland classes identified using discriminant analysis with four image and six ancillary variables.

BLACK = Deciduous/intermediate crown
DARK GRAY = Coniferous/intermediate crown
LIGHT GRAY = Coniferous/solid crown

ORIGINAL PAGE
OF POOR QUALITY

GINAL PAGE
POOR QUALITY

FIGURE 73. COMPARATIVE DISPLAYS OF THE DISCRIMINANT CLASSIFICATION, A VERIFICATION OF THAT CLASSIFICATION, AND THE USGS MAP OF THE *FOREST-LANDS* OF THE DENVER AREA. Scale 1:250,000.

3

BOLDOUT FRAME

This Page Intentionally Left Blank

The final ten-channel discriminant function-classified land-use map was checked cell-by-cell against the expanded 1972-1973 USGS reference map (Appendix C). The average verification accuracy for all second- and third-order land uses was 50.1 percent, whereas the average verification accuracy for all first-order land uses was 76.3 percent (table 24). The 50.1-percent final verification accuracy was in close agreement with the 51.0-percent verification accuracy achieved with the 1/81-grid sample of 4,100 points during the development of the discriminant function (table 23).

The first-order classification results were, in order of decreasing accuracy:

- Urban land use (89.3 percent)
- Forest land use (66.3 percent)
- Water land use (62.0 percent)
- Agriculture land use (56.7 percent)
- Rangeland use (56.3 percent)
- Barren land use (35.0 percent)

The second-order classification results were, in order of decreasing accuracy:

- Residential (84.9 percent)
- Reservoirs (58.1 percent)
- Lakes (56.0 percent)
- Grassland (54.9 percent)
- Institutional (51.1 percent)
- Transportation (45.3 percent)
- Chapparal (taken as brushland) (40.9 percent)
- Industrial (40.5 percent)
- Commercial and services (25.3 percent)
- Bare exposed rock (8.7 percent)
- Open and other urban (7.4 percent)
- Streams and waterways (0.4 percent)
- Extractive (0.1 percent)

TABLE 24

FINAL TEN-CHANNEL LINEAR-DISCRIMINANT CLASSIFICATION RESULTS USING FOUR LANDSAT AND SIX ANCILLARY LANDSCAPE VARIABLES. The verification accuracy of the machine-processed single-date Landsat-1 image (figures 67 through 73) are given. The tabulations were based on checking 24 land uses on a point-to-point comparison with the 1972-1973 USGS land-use data plane. The composite second- and third-order average classification accuracy of 50.1 percent substantially agreed with the 51.0-percent value predicted by the 4,100-point grid-sampling training-set sample (table 23).

First-Order Land Use/Land Cover Second-Order Land Use/Land Cover Third-Order Land Use/Land Cover	Total USGS Cells	Total Cells Correct	Classification Accuracy (percent)
Urban and Built-Up Land			
Residential	98,676	83,761	84.88
Commercial and services	12,825	3,238	25.25
Recreational	13,113	2,073	15.81
Industrial	12,330	4,997	40.53
Extractive	4,824	6	0.12
Transportation	5,922	2,685	45.34
Utilities	1,116	0	0.00
Institutional	29,313	14,968	51.06
Open and other urban	22,239	1,642	7.38
Solid-waste dump	234	0	0.00
Cemetery	1,971	0	0.00
Subtotal/Average	202,563	180,781	89.25
Subtotal/Average	186,129	111,297	59.80
Subtotal/Average	16,434	2,073	12.61
Agricultural Land			
Nonirrigated cropland	46,926	18,228	38.84
Irrigated cropland	729	477	65.43
Pasture	34,542	9,850	28.52
Subtotal/Average	82,197	46,619	56.72
Subtotal/Average	82,197	28,555	34.74
Rangeland			
Grassland	25,281	13,876	54.89
Chapparal	2,763	1,129	40.86
Subtotal/Average	28,044	15,788	56.30
Subtotal, Average	28,044	15,005	53.50
Forest Land			
Deciduous/intermittent crown	810	0	0.00
Coniferous/solid crown	4,068	3,309	81.34
Coniferous/intermittent crown	144	4	2.78
Subtotal/Average	5,022	3,328	66.27
Subtotal/Average	5,022	3,313	65.97
Water			
Streams and waterways	288	1	0.35
Lakes	4,950	2,773	56.02
Reservoirs	1,521	883	58.05
Subtotal/Average	6,759	4,190	61.99
Subtotal/Average	6,759	3,657	54.11
Barren Land			
Bare exposed rock	1,710	149	8.71
Hillslopes	5,481	2,112	38.53
Subtotal/Average	7,191	2,513	34.95
Subtotal/Average	1,710	149	8.71
Subtotal/Average	5,481	2,112	38.53
Aggregate Grand Total/Average	331,776	253,219	76.32
Aggregate Grand Total/Average	222,642	130,108	58.44
Aggregate Grand Total/Average	109,134	36,053	33.04

The third-order classification results were, in order of decreasing accuracy:

- Coniferous/solid crown (81.3 percent)
- Irrigated cropland (65.4 percent)
- Nonirrigated cropland (38.8 percent)
- Hillslopes (38.5 percent)
- Pasture (28.5 percent)
- Recreational (15.8 percent)
- Coniferous/intermittent crown (2.8 percent)
- Cemetery (0.0 percent)
- Deciduous/intermittent crown (0.0 percent)
- Solid-waste dump (0.0 percent)
- Utilities (0.0 percent)

Several problem areas influenced the potential classification accuracy. One substantial difficulty was finding an image or ancillary landscape variable to adequately represent some of the USGS land uses as land-cover types.

Another major problem involved the dual use of the 1972-1973 USGS land-use reference data plane for both training-set feature selection and classification verification. It should be recognized that this map was not 100-percent correct. The stated 90-percent minimum accuracy of the map may be true for the first-order classes, but it was not verified for any level and therefore is doubtful. A 90-percent accurate USGS reference map checked point-by-point against an equally accurate classification map would yield an 81-percent accurate output map; this hypothetical 81-percent verification value is close to the actual first-order verified classification accuracy of 76.3 percent. A random-point field check of the final classification maps is the only sure verification technique; in using the USGS map solely as a training data source, a 10- or 20-percent error may be statistically forgiven by the statistical nature of the classification procedure, but it cannot be overcome in a final point-to-point comparison.

However, the most serious question about the 1972-1973 USGS map accuracy arose from a companion USGS study in which a 1:100,000-scale land-use map was similarly prepared from 1:120,000-scale high-altitude color-infrared imagery. Land use was subsequently verified by observation from a low-flying aircraft, and the average photoclassification accuracy was found to be 77.4 percent for 18 second-order classes of the USGS Circular 671 system (Reference 44).

Total classification time for the ten-variable discriminant function was 2,517.97 central-processor (C-P) seconds to map the 331,776 picture elements into 24 USGS land-use classes, which averaged ~ 132 points

processed per C-P second. An accuracy-weighted classification rate would be ~ 52 correct second-order points per C-P second for the 130,108 second-order points correctly classified. Similarly, a rate of ~ 101 correct first-order points per C-P second was realized for the 253,219 points correctly classified at the first order.

Direct Cost/Time Analysis

A detailed analysis was performed to assess the various direct computer, labor, and material costs and times (table 25). These parameters were broken down by submodel and task area as accurately as possible. Direct computing costs represented $5,509.05 of the total $6,981.55, or almost 79 percent of the entire direct-analysis costs. Only $1,472.50 (21 percent) was expended for direct labor and materials. More important, the equivalent average cost per unit area for 48-variable multivariate image classification, verification, and display was calculated as either 1.9 cents per acre, 4.7 cents per hectare, $12.12 per square statute mile, or $704.19 per 1:24,000-scale USGS quadrangle.

It was not considered appropriate to allocate direct or indirect labor costs other than those reflected because of the extensive iterative nature of any research effort. The time consumed in searching for, examining, and discarding potential imagery or maps, for example, was considerable, as was the computer program coding, debugging, and development. A production land-use mapping mode using these estimates would assume that all input data were on hand and that the computer software was fully implemented.

These costs also potentially represent the development, assembly, and testing of a very large data base on one or more limited map areas in order to optimize the selection of specific Landsat/landscape data planes. The extension of this classification to a larger area involves the smaller costs represented by analysis of the ten variables finally classified here. Total costs for the classification used can be assembled for the same area as a subset of development costs (table 25). These parameters are also broken down by submodel and task area and represent direct computer, labor, and material costs and times (table 26). The equivalent average cost per unit area for ten-variable multivariate image classification, verification, and display was calculated as either 1.3 cents per acre, 3.2 cents per hectare, $8.21 per square statute mile, or $477.17 per 1:24,000-scale USGS quadrangle.

SUMMARY

The focus of this phase of the research investigation was the automated identification and mapping of the various land-use/land-cover types that make up the Denver, Colorado landscape. This endeavor used digital Landsat imagery and ancillary landscape variables to test different feature-extraction procedures and machine-classification algorithms on a 48-variable data base. The grid-sampled-point feature-selection process made use of an existing ground-truth reference map and proved to be superior to the conventional rectangular supervised field selection procedure.

A nonparametric classification technique, linear-discriminant analysis with *a priori* class probabilities, clearly proved to be superior to the maximum-likelihood ratio method in common use today. These two algorithms were tested for identical combinations of four, six, and 22 variables. The 4,100-point 1/81 grid-sampling image was used as the statistical training set for both approaches.

The incremental contribution of ancillary spatial landscape variables to the classification of Landsat spectral data was quantified with a stepwise linear-discriminant function. A ten-channel combination of Landsat

C-3

TABLE 25

COST/TIME TABULATION FOR SINGLE DATE 48-VARIABLE LANDSAT ANCILLARY LANDSCAPE VARIABLE CLASSIFICATION OF THE DENVER METROPOLITAN AREA. These figures are based on the Control Data Corporation 6400 computer and system used at Colorado State University at the basic campus research rate of $290 per machine hour and an hourly work ratio of $5 per man hour. Quoted figures represent only direct computer, labor, and material costs/times. They account for one-time costs only and do not represent redoing anything. Many of the computer program names refer to programs or modification of programs in the LMS package; others refer to programs discussed elsewhere in this report.

Task Description	Time: Man Hours	Time: C.P.* Seconds	Cost
Land use submodel (seven variables)			
Aerial photoacquisition (procurement of existing prints)			$ [illegible]
1970 land use photointerpretation	[illegible]		[illegible]
1963 land use photointerpretation	[illegible]		[illegible]
1972-1973 USGS land use map sampling	[illegible]		[illegible]
PLANMAP land use data plane filling manipulation		[illegible]	[illegible]
Total			[illegible]
Physiographic submodel (nine variables)			
Topographic elevation coding	[illegible]		[illegible]
TOPOMAP slope aspect computations		[illegible]	[illegible]
INSOL[2] solar radiation computations		[illegible]	[illegible]
Surficial geology coding	[illegible]		[illegible]
Total			[illegible]
Transportation submodel (five variables)			
Transportation route coding	[illegible]		[illegible]
Binary road data plane filling		[illegible]	[illegible]
PAN[2] minimum distance computations		[illegible]	[illegible]
Total			[illegible]
Socioeconomic submodel ([illegible] variables)			
Census tract boundary coding	[illegible]		[illegible]
Census tract acreage dot sampling	[illegible]		[illegible]
Census data plane filling ratioing		[illegible]	[illegible]
Total			[illegible]
Landsat image submodel (ten variables)			
Digital Landsat data acquisition (one frame and date)			[illegible]
Image preprocessing			
EROS USGS data reformatting		[illegible]	[illegible]
TWOTAPE image file merging		[illegible]	[illegible]
ROTATE[2] image rectification rotation		[illegible]	[illegible]
COMBINE image ancillary data overlaying		[illegible]	[illegible]
TRANSE[2] image channel ratioing		[illegible]	[illegible]
ANCILLARY data reformatting		[illegible]	[illegible]
Subtotal			[illegible]
Feature Extraction			
SAMPLER[2] LST image grid sampled point file		[illegible]	[illegible]
EXTRACT[2] LST image point file creation		[illegible]	[illegible]
CLASSIFY LST image point file classification		[illegible]	[illegible]
EXTRACT[2] LO image file creation		[illegible]	[illegible]
CLASSIFY LO image file classification		[illegible]	[illegible]
DECODER[2] coded map file creation		[illegible]	[illegible]
CHECKER[2] LO image first order verification		[illegible]	[illegible]
PHASE[2] LO image microfilm graymaps		[illegible]	[illegible]
Subtotal			[illegible]
Image classification			
EXTRACT[2] image classification file creation		[illegible]	[illegible]
CLASSIFY full image classification		[illegible]	[illegible]
DECODER[2] coded map file creation		[illegible]	[illegible]
Subtotal			[illegible]
Classification output			
ANCILLARY map file data reformatting		[illegible]	[illegible]
PHASE[2] land use theme microfilm graymaps		[illegible]	[illegible]
CHECKER[2] CHECKER[2] full image verification		[illegible]	[illegible]
Subtotal			[illegible]
Total (Landsat image submodel)			[illegible]
Total Cost			[illegible]

*C.P. = central processor.

TABLE 26

COST/TIME TABULATION FOR SINGLE-DATE, TEN-VARIABLE LANDSAT/ANCILLARY LANDSCAPE VARIABLE CLASSIFICATION OF THE DENVER METROPOLITAN AREA. These figures are based on the Control Data Corporation 6400 computer and system used at Colorado State University at the basic campus research rate of $290 per machine hour and an hourly work rate of $5 per man-hour. Quoted figures represent only direct computer, labor, and material costs/times. They account for one-time costs only and do not represent redoing anything. Only the steps that would actually be employed in extending the concepts learned with the 48-variable tests were costed. For ease of comparison, the area analyzed here was assumed to be the same size.

Task Description	Time		Cost
	Man-Hours	CP* Seconds	
Land-use submodel			80.00
Physiographic submodel			
Topographic elevation coding	74.8		$ 374.00
TOPOMAP slope aspect computations		750	61.14
Total			435.14
Transportation submodel			
Transportation route coding	11.9		59.60
Binary road data-plane filling		16	1.31
PAX2 minimum-distance computation		215	17.29
Total			78.20
Socioeconomic submodel			
Census-tract boundary coding	26.2		131.50
Census-tract acreage dot sampling	14.1		70.50
Census data-plane filling rationg		56	4.54
Total			206.54
Landsat image submodel			
Digital Landsat data acquisition (1 frame)			200.00
Image preprocessing			
FROS USGS data reformatting		732	58.92
TWOTAPE image-file merging		622	50.11
ROTATE2 image rectification rotation		19,282	1,553.27
COMBINE image ancillary data overlaying		61	4.93
TRANSF2 image-channel rationg		404	32.52
ANCILLARY data reformatting		281	29.24
Subtotal			1,929.04
Feature Extraction			
Image Classification			
EXTRACT2 image-classification file creation		13,556	1,092.01
CLASSIFY full-image classification		7,656	616.73
Subtotal			1,708.74
Classification Output			
ANCILLARY map-file data reformatting		411	33.11
PHASE1 land-use theme microfilm gray maps		4,218	339.78
Subtotal			372.89
Total (Landsat image submodel)			4,010.67
Total (five submodels)			4,730.55

*CP = central processor.

basic bands and ratios produced an average verification accuracy of 38.4 percent. The addition of three ancillary variables (topographic elevation, average number of cars per family, and built-up urban-area minimum distance) increased thi. verification accuracy to over 50 percent, but the successive 28 nonland-use spatial variables raised the final average accuracy to only 53.9 percent.

An optimal ten-channel combination of four Landsat image bands and six ancillary landscape variables was used to classify the full 576- by 576-element image into 24 land-use categories. Point-by-point comparison with the 1972-1973 USGS land-use reference data plane showed an overall accuracy of 76.3 percent for the six aggregate first-order classes. Average accuracy for the 24 second- and third-order classes was 58.4 and 33.0 percent, respectively, but three urban and one forest third-order categories received no correct classifications.

A detailed cost/time tabulation for the multivariate data compilation, registration, classification, verification, and display showed the average cost of developing the model to be 1.9 cents per acre, 4.7 cents per hectare, $12.12 per square mile, or $704.19 per 7.5-minute USGS quadrangle. Actual production application of the selected optimal channels and signatures yielded costs of 1.3 cents per acre, 3.2 cents per hectare, $8.21 per square statute mile, or $477.17 per 7.5-minute USGS quadrangle.

This Page Intentionally Left Blank

CHAPTER 5

CONCLUSIONS

PRECEDING PAGE BLANK NOT FILMED

CHAPTER 5

CONCLUSIONS

SUMMARY

Significant findings were made in the related study of spatial land-use projection and Landsat image classification. The first line of investigation was an attempt to estimate land-use changes using multivariate map variables and known points of change from aerial photointerpretation. Statistical relationships were developed for both change and nonchange land-use types with 27 physiographic, socioeconomic, and transportation variables.

The general land-use change model was formulated as a linear-discriminant function using 4-ha (10-acre) square cells as prediction units. Map-derived independent variables were numerically related in the multivariate structure of this model to known imagery-derived cells of change.

Seven models were tested for focusing upon cropland and pasture agricultural land—the primary 1963 land-use change class. Change prediction accuracies ranged from 42 to 57 percent. Although these results were encouraging, they were difficult to assess because of the scarcity of similar studies. Although the current analysis was tentative and exploratory, nonetheless it indicated the feasibility of developing spatial land-use projection models. Clearly, their application could be of considerable utility in future improved land-use planning activities.

The parallel application of linear-discriminant analysis for Landsat multivariate image classification has confirmed the value of this flexible statistical technique. Various feature-extraction methods and supervised machine-classification algorithms were tested on a single-date Landsat image augmented by 38 additional channels of ancillary map data spatially registered to the spectral-image data base. The overall objective was to systematically test these factors in producing a land-classification map, using a current USGS land-use map as a reference.

The comparison of two sampling methods for statistical feature extraction indicated a wide variation in trial classification test results and applicability to image processing (Appendix E). The conventional rectangular training field selection process resulted in a very high test-classification accuracy, but was invalidated after poor classification results on a larger one-ninth image sample. However, the systematic point-grid sampling gave highly repeatable results. This 1/81-image sample provided not only sample points distributed over the entire image, but also useful *a priori* land-use probabilities.

These *a priori* class probabilities enabled the linear-discriminant function to far outclassify the widely used maximum-likelihood ratio algorithm. Both procedures were compared on three test sets of image and ancillary landscape variables using the same point-training set developed earlier.

Verification of the 576- by 576-point Landsat image classification with the USGS reference map showed an average accuracy of over 76 percent at the generalized first-order categories. The ten-variable discriminant function used also classified the 24 second- and third-order classes at accuracies of 58.4 and 33.0 percent respectively.

A significant achievement of this study was the creation of a composite Landsat image file consisting of ten spectral bands and ratios overlaid by 38 map-derived ancillary land-use, physiographic, socioeconomic, transportation, and MSS/insolation ratio variables. Statistical testing showed these collateral variables (excluding land-use data variables) increased classification accuracy by over 40 percent; however, just three selected additional ancillary variables added over 30 percent of the increased accuracy.

The USGS Circular 671 scheme, developed for high-altitude aircraft and satellite data inputs, worked reasonably well as a hierarchical classification system at both the first- and second-order levels in this supervised machine-interpretation effort. However, it appeared that much more detailed ancillary data will be needed for accurately interpreting the third-order urban classes. The inherent remaining problem in urban settings is the repetitive use of manmade building materials in a diversity of urban land uses. Fortunately, the USGS land-use classification system is predominately land-cover oriented and is therefore useful for remote sensing discrimination. The ancillary data compensated, in large part, for the land-use versus land-cover confusion factor in the seven second-order urban classes but did not do well for the four third-order urban classes. Average classification accuracy was 59.8 percent for the second-order urban classes but only 1?.6 percent for the third-order urban classes. Lastly, the fixed *a priori* classes of the USGS system t! enable the results of its application to be intercompared from area to area pose a potential problem for the unsupervised clustering algorithms coming into general use (Reference 45).

The average direct cost for the optimized model for 24-class land-use mapping of the Landsat scanner imagery and associated ancillary map overlays was $0.0468 per hectare ($0.0189 per acre). Based on these costs, the mapping of a 7.5-minute USGS quadrangle would be $704.19. These unit-area costs included extensive geometric correction preprocessing, ancillary data registration, feature extraction, classification, verification, and display operations. The production costs of employing this approach were $0.0128 per acre, $0.0317 per hectare, $8.21 per square statute mile, or $477.17 per 7.5-minute USGS quadrangle. Computer time was billed at $290 per hour, and personnel wages were set at $5 per man-hour.

RELATED APPLICATIONS AREAS

A considerable quantity of landscape data was assembled during this study. An increasingly important facet of land-use planning is facilities siting, particularly fossil fuel, hydroelectric, geothermal, and nuclear-power plants. Thus, a new, comprehensive, land-use planning data base must encompass even more ecological, environmental, hydrological, land-use, and natural resource data.

These long-term construction projects are enormously costly, heavily regulated by governmental agencies, scrutinized by citizen watchdog groups, and vulnerable to location errors. Inadequate data may result in structures of the wrong size or scope with consequent losses from destruction or damage, failure to take advantage of economic potential, or excess capacity (Reference 25). Consequently, these engineering projects are planned with consummate knowledge and care, and they provide an even more logical place to apply the landscape assessment and modeling methods developed here.

More appropriate planning data and modeling of this type could be used to avoid agricultural investment where the climate is inappropriate or in designing dams, bridges, and irrigation works more in accord with the true characteristics of the site (Reference 25).

Most important, however, these comprehensive data bases will permit full development of the computer modeling techniques necessary for predicting how the landscape will spatially evolve in response to various scenarios of anticipated alternatives (References 8 and 46). The scenarios simulated and evaluated could include diverse objectives, such as the control of shifting cultivation in underdeveloped countries, new zoning boundaries for urban land planning, alternate sites for a new power plant, or the environmental impact of siting a new dam. Land management in general, and land-use planning in particular, would be substantially improved if the future spatial impacts of a contemplated action could be accurately modeled before any commitment to a long-term, rigid course of action.

RECOMMENDATIONS

When this 5-year study was completed, six recommendations appeared to be appropriate. The first three deal with ancillary data inputs, and the remainder deal with machine interpretation and processing. All six recommendations impact on the landscape modeling approach to land-use planning outlined in this study.

Socioeconomic Census Data Densification

The use of census tracts in preparing socioeconomic data planes produced reasonable results in both the spatial land-use projection modeling and the Landsat land-use/land-cover classification efforts. Unfortunately, although a 4-ha (10-acre) ancillary data mapping cell was superimposed, these data planes were not as highly resolved in a spatial sense as desired for urban planning applications. The use of the smaller enumeration districts for enumerating densely populated urban areas would improve the data resolution three- to five-fold.

Topographic Elevation Data Digitization

Topographic elevation has been shown to be a primary constituent of the landscape model. Derivative data planes included topographic slope, topographic aspect, and solar radiation or insolation. Experimentation with Defense Mapping Agency (DMA) digital terrain tapes showed glaring deficiencies in the interpolated 61-m (200-ft) contours (Reference 16). Thus, the only alternative was to tediously hand-cellularize the topographic maps for the precise elevation data needed for this study. Considerable savings in both time and money can be realized if the digital terrain models created by the USGS in constructing the 1:24,000-scale, 7.5-minute topographic maps were saved on magnetic tape.

Collateral Soils Data Inputs

A soil survey is a detailed physical inventory of the soil, showing its depth, texture, structure, drainage, stoniness, slope, erosion, and other land features that are helpful in determining its use and capability. These highly useful data were not available for the Denver study area. Soil type is certainly one of the principal determinants of land use (Reference 25). Its inclusion in the data base would be a useful addition to the ancillary physiographic submodel for predicting spatial land-use changes under given circumstances and for improving the Landsat image-classification results.

Multidate Land-Use Change Detection

Land-use change detection involves the comparison of remote sensing imagery or other inventory data for two different points in time. Change detection on a spatial basis provides important planning data in managing the various landscape components. The use of high-altitude or orbital remote sensing data provides a near-orthographic view with attendant synoptic large-area overview. These images must be in point geometric congruence if multiple images are to be processed in a digital computer and point classifications are to be made using statistical pattern recognition techniques. Thus, a principal advantage of Landsat digital imagery is that time-differing scenes can be transformed to usable congruence much easier and at less cost per unit area than any type of aircraft imagery. It is therefore recommended that continued effort be expended to supply geometrically corrected data at various selected cell sizes. Because this need is common to all users, it should be centrally addressed for all data distributed.

Symbiotic Use of Ancillary Map Data

The use of remote sensing images in generating change-detection data can provide metropolitan areas with the necessary information for structuring viable planes for the guidance and development of the evolution of these areas. However, for maximizing potential utility, remote sensing should be used with ancillary data sources. Although spectral Landsat data alone has worked well on natural cover types, spatial map data has proved its value in urban settings in which a single land use encompassed a plethora of artificial cover materials. It appears, therefore, that any reliable machine interpretation system will have to embody spectral, spatial, and temporal discriminants to function well for all anticipated urban land uses and conditions. The basic input to these automatic techniques is, and will continue to be, spectral data. However, both spatial and multitemporal map data must be incorporated into the spectral data base on an increasingly larger basis for further improvement in classification accuracy.

Future Information Systems Development

Proper use of available data, as well as new methods of data collection and analysis, are required for defining and quantifying the capabilities and limitations of the landscape components for local or regional level control and management. Electronic data processing (EDP) provides extensive computational power for storing, manipulating, retrieving, displaying, and updating resource data (Reference 9). Automated resource information systems, which spatially identify resource data, increasingly appear to be the most feasible means of providing complete, objective, and consistent information and analyses (Reference 47). These systems will provide the basis for enlightened land-management decisions relative to the following questions:

- What is the current land use?
- How is the land use defined and evaluated?
- What are its characteristics, dynamics, and limitations?
- How much single- versus multiple-purpose land use is there?
- What are the evolving socioeconomic spatial patterns?

- How much open space exists and how is it preserved?
- How can land-use data be efficiently collected and used?

A common natural-resource inventory data system is needed, as well as an automated resource-information system for keeping the inventory current. Inventory and benchmark data components for such a system for improving land planning include map and other spatial data to specify the following (Reference 1):

Climate
Demography
Economics
Geology
Land Use
Minerals
Population
Recreation
Soils
Topography
Vegetation
Water
Wildlife

The need for more timely and higher quality land-use data for evaluating, managing, and planning shrinking natural resources is clear. The perfection and application of automated resource-information systems is needed for interfacing the multiple sources of physiographic, remote sensing, socioeconomic, and transportation data into a coherent analytical machine structure. The computational power thus achieved can be used to store, manipulate, display, and update the resulting resource information and to enable the effective transfer of spatial landscape models and related operations research methods into improved land resource-management decisions.

This Page Intentionally Left Blank

REFERENCES

1. Tom, C.H., L.D. Miller, J.S. Krebs, and R. Aukerman, "The Design of a Model to Project Land Uses and Predict Open Space Encroachment Patterns/Denver Metropolitan Area," Contract 3-14-07-03 for Bureau of Outdoor Recreation, Department of Interior; Department of Civil Engineering, Colorado State University, Fort Collins, 1974, 195 pp.

2. Perloff, Harvey S., *Education for Planning: City, State, and Regional*, Johns Hopkins Press, Baltimore, Maryland, 1957, 189 pp.

3. Bosselman, F., and D. Callies, *Summary Report — The Quiet Revolution in Land Use Contr* Council on Environmental Quality, U.S. Government Printing Office, Washington, D.C., 1971 34 pp

4. Moore, E.G., and B.S.-Wellar, "Remote Sensor Imagery in Urban Research: Some Po[illegible]ialit[illegible] Problems," Technical Letter NASA-118, Lyndon B. Johnson Space Center, H[illegible]xas, 1968, 32 pp.

5. Mullins, R.H., "Analysis of Urban Residential Environments Using Color Infrared Aerial Photography: An Examination of Socioeconomic Variables and Physical Characteristics of Selected Areas in the Los Angeles Basin," Department of Geography, University of California, Los Angeles, 1969, 185 pp.

6. Anderson, R.J., Jr., and T.D. Crocker, "Air Polution and Residential Property Values," *Urban Studies*, 8(3), 1971, pp. 171-180.

7. Goehring, Darryl R., "Monitoring the Evolving Land Use Patterns on the Los Angeles Metropolitan Fringe Using Remote Sensing," Technical Report T-71-5, Department of Geography, University of California, Riverside, 1971, 107 pp.

8. Miller, L.D., C.H. Tom, and K. Nualchawee, "Remote Sensing Inputs to Landscape Models Which Predict Future Spatial Land Use Patterns for Hydrologic Models," NASA TM X-71330, Goddard Space Flight Center, Greenbelt, Maryland, 1977, 41 pp.

9. Tom, C.H., and L.D. Miller, "A Review of Computer-Based Resource Information Systems," Land Use Planning Information Report 2, Colorado State University, Fort Collins, 1974, 50 pp.

10. Anderson, J.R., E.E. Hardy, and J.T. Roach, *A Land-Use Classification System for Use with Remote Sensor Data*, Geological Survey Circular 671, U.S. Government Printing Office, Washington, D.C., 1972, 16 pp.

11. Driscoll, Linda B., *Land-Use Classification Map of the Greater Denver Area, Front Range Urban Corridor*, U.S. Geological Survey Miscellaneous Inventory Map I-856-I, 1975.

12. Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, Geological Survey Professional Paper 964, U.S. Government Printing Office, Washington, D.C., 1976, 28 pp.

13. Alexander, Robert H., "Central Atlantic Regional Ecological Test Site," In: *Fourth Annual Earth Resources Program Review, Vol. III*, Lyndon B. Johnson Space Center, Houston, Texas, 1972, pp. 72-1 to 72-9.

14. Place, John L., "An Automated Map and Model of Land Use in the Phoenix Quadrangle," In: *Fourth Annual Earth Resources Program Review, Vol. III*, Lyndon B. Johnson Space Center, Houston, Texas, 1972, pp. 71-1 to 71-19.

15. McHarg, Ian, *Design with Nature*, Natural History Press, Garden City, New York, 1969, 179 pp.

16. Tom, C.H., and J.R. Getter, "Computer Mapping of Wildfire Hazard Areas: A User-Oriented Case Study," Land Use Planning Information Report 4A, Colorado State Forest Service, Resources Division, Fort Collins, 1975, 44 pp.

17. Tom, C.H., "Technical Documentation for the Colorado State Forest Service "TOPOMAP" Slope/Aspect Mapping System," Land Use Planning Information Report 4B, Colorado State Forest Service, Resources Division, Fort Collins, 1975, 23 pp.

18. Dixon, W.J., "BMD Biomedical Programs," *Publications in Automatic Computation 2*, University of California Press, Berkeley, 1967, pp. 214a-214s.

19. Miller, L.D., "Satellite Monitoring of Regional Open Space Encroachment (Denver, Colorado)," Progress Report 1, Department of Civil Engineering, Colorado State University, Fort Collins, 1973, 20 pp.

20. Sharpnack, D.A., and G. Akin, "An Algorithm for Computing Slope and Aspect from Elevations," *Photogram. Eng.*, 35(3), 1969, pp. 247-248.

21. Buffo, J., L.J. Fritschen, and J.L. Murphy, "Direct Solar Radiation on Various Slopes from 0 to 60 Degrees North Latitude," Research Paper PNW-142. Pacific Northwest Forest and Range Experiment Station, Portland, Oregon, 1972, 74 pp.

22. Holben, Brent N., *The Development and Sensitivity Analysis of a Model for Estimating Insolation Climate in Mountainous Topography*, thesis, Department of Earth Resources, Colorado State University, Fort Collins, 1975, 159 pp.

23. Chase, G.H., and J.A. McConaghy, *Generalized Surficial Geologic Map of the Denver Area, Colorado*, U.S. Geological Survey Miscellaneous Inventory Map I-731, 1972.

24. Johnston, Emily G., "The PAX User's Manual," Computer Note CN-10, Computer Science Center, University of Maryland, College Park, 1970, 304 pp.

25. Herfindahl, Orris C., *Natural Resource Information for Economic Development*, Johns Hopkins Press, Baltimore, Maryland, 1969, 212 pp.

26. Miller, L.D., "Land Use and Open Space Encroachment Patterns," *Slide Rule*, 19(2), College of Engineering, Colorado State University, Fort Collins, 1976, pp. 1-5.

27. Getter, J.R., and C.H. Tom, *Forest Site Index Mapping and Yield Model Inputs to Determine Potential Site Productivity*, Colorado State Forest Service, Resources Division, Fort Collins, 1977, 20 pp.*

28. Duda, Richard O., and Peter E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973, 482 pp.

29. Andrews, Harry C., *Introduction of Mathematical Techniques in Pattern Recognition*, John Wiley and Sons, New York, 1972, 242 pp.

30. Davis, John C., *Statistics and Data Analysis in Geology*, John Wiley and Sons, New York, 1973, 550 pp.

31. Williams, D.L., and G.F. Haver, "Forest Land Management by Satellite: LANDSAT-Derived Information as Input to a Forest Inventory System," NASA Intralab Project 75-1, Goddard Space Flight Center, Greenbelt, Maryland, 1976, 35 pp.

32. Taber, John E., "Evaluation of Digitally Corrected ERTS Images," *Third Earth Resources Technology Satellite-1 Symposium*, Vol. I, Technical Presentations, Section B, Goddard Space Flight Center, Greenbelt, Maryland, 1973, pp. 1837-1841.

33. Anuta, Paul E., "Geometric Correction of ERTS-1 Digital Multispectral Data," LARS Information Note 103073, Purdue University, West Lafayette, Indiana, 1973, 23 pp.

34. Sung, Quo-Cheng, and L.D. Miller, "Land Use/Land Cover Mapping (1:25,000) of Taiwan, Republic of China by Automated Multispectral Interpretations of LANDSAT Imagery," NASA TM X-71382, Goddard Space Flight Center, Greenbelt, Maryland, 1977, 168 pp.

35. Maxwell, Eugene L., "A Remote Rangeland Analysis System," Contract 14-08-0001-13561 for U.S. Geological Survey, Report 1885-F, Department of Earth Resources, Colorado State University, Fort Collins, 1974, 110 pp.

36. Pearson, R.L., C.J. Tucker, and L.D. Miller, "Spectral Mapping of Short Grass Prairie Biomass," *Photogram. Eng. and Remote Sensing*, **42**(4), 1976, pp. 317-323.

37. Maxwell, Eugene L., "Multivariate Systems Analysis of Multispectral Imagery," *Photogram. Eng. and Remote Sensing*, **42**(9), 1976, pp. 1173-1186.

38. Crane, R.N., and W. Richardson, "Rapid Processing of Multispectral Data Using Linear Techniques," In: *Remote Sensing of Earth Resources, Vol. I*, F. Shahrokhi, ed., University of Tennessee, Tullahoma, 1972, pp. 581-585.

39. Bond, A.D., and R.J. Atkinson, "An Integrated Feature Selection and Supervised Learning Scheme for Fast Computer Classification of Multispectral Data " In: *Remote Sensing of Earth Resources, Vol. I*, F. Shahrokhi, ed., University of Tennessee, Tullahoma, 1972, pp. 645-672.

*Reprinted in *Resource Inventory Notes*, Bureau of Land Management, U.S. Department of Interior, Denver, Colorado, September 1977, pp. 1-10.

40. Su, M.Y., R.R. Jayroe, and R.E. Cummings, "Unsupervised Classification of Earth Resources Data," In: *Remote Sensing of Earth Resources, Vol. 1*, F. Shahrokhi, ed., University of Tennessee, Tullahoma, 1972, pp. 673-694.

41. Richardson, A.J., R.J. Torline, and W.A. Allen, "Computer Identification of Ground Pattern from Aerial Photographs," *Proceedings of the Seventh International Symposium on Remote Sensing of Environment*, University of Michigan, Ann Arbor, 1971, pp. 1357-1376.

42. Eppler, W.G., C.A. Helmke, and R.H. Evans, "Table Look-Up Approach to Pattern Recognition," *Proceedings of the Seventh International Symposium on Remote Sensing of Environment*, University of Michigan, Ann Arbor, 1971, pp. 1415-1425.

43. Mendenhall, William, *Introduction to Linear Models and the Design and Analysis of Experiments*, Wadsworth Publishing Company, Belmont, California, 1968, 465 pp.

44. Fitzpatrick, Katherine A., *CARETS: A Prototype Regional Environmental Information System, Vol. 6: Cost, Accuracy and Consistency Comparisons of Land Use Maps Made from High-Altitude Aircraft Photography and ERTS Imagery*, U.S. Geological Survey, Reston, Virginia, 1975, 62 pp.

45. Krebs, P.V., and R.M. Hoffer, Multiple Resource Evaluation of Region 2 U.S. Forest Service Lands Utilizing LANDSAT MSS Data, NASA CR-149595, Institute of Arctic and Alpine Research, University of Colorado, Boulder, 1976, 298 pp.

46. Wacharakitti, S., L.D. Miller, and C.H. Tom, "Tropical Forest Land Use Evolution/Twenty-Year Landscape Model with Inputs from Existing Maps, Historical Air Photos, and ERTS Satellite Imagery." Land Use Planning Monograph 1, Colorado State University, Fort Collins, 1975, 217 pp.

47. Tom, C.H., *Technical Documentation for Colorado State Forest Service "COMBINE2" Cellular Data Plane Overlaying/Plotting Program*, Colorado State Forest Service, Resources Division, Fort Collins, 1977, 28 pp.

APPENDIX A

MULTIPLE DISCRIMINANT ANALYSIS

PRECEDING PAGE BLANK NOT FILMED

APPENDIX A

MULTIPLE DISCRIMINANT ANALYSIS

Multiple discriminant analysis finds the transform that yields the minimum ratio of the difference between group multivariate means to the multivariate variance within the groups. The statistical algorithm embodied in program CLASSIFY* (Appendix B) computes a discriminant function for each of the land-use classes by selecting the independent variables—the 48 Landsat image and ancillary map variables—in a stepwise fashion. The new variable entered at each step is selected on the basis of largest F-value to enter. An original set of observations on a picture element or cell, for example, is transformed into a single discriminant score by the discriminant function. The score represents the position of the picture element along the line defined by the linear-discriminant function. It is seen, then, that the discriminant function reduces a multivariate problem down into a univariate situation.

The discriminant function is found by solving an equation of the form

$$\left[s_p^2\right] \cdot [\lambda] = [D]$$

where $\left[s_p^2\right]$ is an m-by-m matrix of pooled variances and covariances of the m variables. The coefficients of the discriminant function are represented by a column vector of the unknown lambdas. Lowercase Greek lambdas (λ's) are used by convention to represent the coefficients of the discriminant function. These are exactly the same as the betas (β's) also used by convention in regression equations. These should not be confused with the lambdas used to represent eigenvalues in principal components or factor analyses, nor the lambdas used to represent wavelength in a remote sensing sense.

The right-hand side of the equation consists of the column vector of m differences between the means of the two groups in the simple discriminant-analysis case. The equation can be solved by inversion and multiplication, such as

$$[\lambda] = \left[s_p^2\right]^{-1} \cdot [D]$$

or by use of a simultaneous equation solution.

The various entries in the matrix equation must be determined in order to compute the discriminant function. The mean differences are simply found as

$$D_j = \bar{A}_j - \bar{B}_j = \frac{\sum_{i=1}^{n_a} A_{ij}}{n_a} - \frac{\sum_{i=1}^{n_b} B_{ij}}{n_b}$$

*CLASSIFY occurs in the LMS package and is the modified version of BMD07M, which is part of the UCLA biomedical statistical package available on most major computers (Reference 1). The stepwise linear-discriminant analysis remains unmodified; however, the input/output and internal control code was modified to handle the much larger data bases of remote sensing imagery.

In this notation, A_{ij} is the ith observation of variable j in group A; A_j is the mean of variable j in group A, or the average of n_a observations. Similarly, these conventions apply to group B. The multivariate means of groups A and B can be regarded as forming two vectors. The difference between these multivariate means, therefore, also forms the vector

$$|D_j| = |\bar{A}_j| - |\bar{B}_j|$$

A matrix of sums of squares and cross-products of all variables in group A, as well as a similar matrix for group B, is computed in order to construct the matrix of pooled variances and covariances. This is done by conventional means for group A as

$$SPA_{jk} = \sum_{i=1}^{n_a} (A_{ij}A_{ik}) - \frac{\sum_{i=1}^{n_a} A_{ij} \sum_{i=1}^{n_a} A_{ik}}{n_a}$$

Here, A_{ij} denotes the ith observation of variable j in group A as before, and A_{ik} denotes the ith observation of variable k in the same group. Whenever j = k, this quantity becomes the sum of squares of variable k. Similarly, a matrix of sums of squares and cross-products is found for group B as

$$SPB_{jk} = \sum_{i=1}^{n_a} (B_{ij}B_{ik}) - \frac{\sum_{i=1}^{n_b} B_{ij} \sum_{i=1}^{n_b} B_{ik}}{n_b}$$

The sums of products matrix from group A is denoted as [SPA], and that from group B as [SPB]. The matrix of pooled variance is now found as

$$\left[s_p^2\right] = \frac{[SPA] + [SPB]}{n_a + n_b^{-2}}$$

This equation for pooled variance is exactly the same as that used in T-tests for the equality of multivariate means. Although the number of calculations involved in deriving the coefficients of a discriminant function appear to be large, they are less formidable than at first glance. The set of λ coefficients are entries in the discriminant-function equation of the form

$$R = \lambda_1 \psi_1 + \lambda_2 \psi_2 + \cdots + \lambda_m \psi_m$$

This is a linear function whose terms are summed to yield a single number, the discriminant score. In this two-dimensional example, the discriminant function can be plotted as a line on the scatter diagram of the two original variables (figure 44 of the main text). It is a line through the plot whose slope, α, is

$$\alpha = \lambda_2/\lambda_1$$

Substitution of the midpoint between the two group means in this equation yields the discriminant index, R_0. Specifically, for each value of ψ_j, the following terms are inserted

$$\psi_j = \bar{A}_j \bar{B}_j / 2$$

The discriminant index, R_0, is the point along the discriminant-function line that is exactly halfway between the center of group A and the center of group B. Next, the multivariate mean of group A can be substituted into the equation to obtain R_A (that is, $\psi_j = \bar{A}_j$), and the mean of group B to obtain R_B ($\psi_j = \bar{B}_j$). These define the centers of the two original groups along the discriminant function for group A,

$$R_A = \lambda_1 \bar{A}_1 + \lambda_2 \bar{A}_2$$

and for group B,

$$R_B = \lambda_1 \bar{B}_1 + \lambda_2 \bar{B}_2$$

The members of group A that are located on the group B side of R_0 and the members of group B that are located on the group A side of R_0 are misclassified by the discriminant function.

The significance of the separation between the two groups can be tested if certain assumptions are made regarding the nature of the data used in the discriminant function. These five basic test assumptions are:

(1) The observations in each group are randomly chosen.

(2) The probability of an unknown observation belonging to either group is equal.

(3) The variables are normally distributed within each group.

(4) The variance-covariance matrices of the groups are equal in size.

(5) None of the observations used to calculate the function were misclassified.

The most difficult assumptions to justify are (2), (3), and (4). However, the function is not seriously affected by limited departures from normality or by limited inequality of variances. The justification of (2) depends on an *a priori* assessment of the relative abundance of the groups under examination (Reference 2).

A test for the significance of the discriminant function is developed from the T-statistic mentioned earlier. A "distance" measure between the two multivariate means can be calculated by simply subtracting R_A from R_B. This is equivalent to substituting the vector of differences between the two group means into the discriminant equation, or setting the individual values of ψ_j equal to D_j. This distance measure is called "Mahalanobis' distance," or generalized distance, D^2. It is a measure of the separation between the two multivariate means expressed in units of the pooled variance. The T-test of this distance has the form

$$T^2 = \frac{n_a n_b}{n_a + n_b} D^2$$

The T-test can be transformed into an F-test, becoming

$$F = \frac{(n_a + n_b - m - 1)}{(n_a + n_b - 2)\, m} \frac{n_a n_b}{n_a + n_b} D^2$$

with m and ($n_a + n_b - m - 1$) degrees of freedom. The null hypothesis tested by this statistic is that the two multivariate means are equal or that the distance between them is zero; that is,

$$H_0: [D_j] = 0$$

against

$$H_1: [D_j] > 0$$

The utility of this as a test of a discriminant function should be clear. If the means of the two groups are very close together, it will be difficult to separate them, especially if both groups have large variances. On the other hand, if the two means are well-separated and scatter around the means is small, discrimination is relatively easy.

Not all of the variables included in the discriminant function are equally useful in distinguishing one group from another. Those variables that are not particularly useful can be isolated and eliminated from future analyses. Because discriminant analysis is so closely related to multiple regression, most of the procedures for selecting the most effective set of predictors can also be used to find the most effective set of discriminators. For example, the relative contribution of variable j to the distance between the two group means may be measured by a quantity, E_j,

$$E_j = \lambda_j D_j / D^2$$

where D_j is the difference between the jth means of the two groups. This is only one measure of the direct contribution of the variable j, and it does not consider interactions between variables. If two or more of the variables in the discriminant function are not independent, their interactions may contribute to D^2 to a greater extent than the value of E_j suggests. This measure serves roughly the same purpose as standardized partial-regression coefficients in multiple regression. Values of E_j may be simply converted to percentages by multiplying by 100.

REFERENCES

1. Dixon, W. J., *BMD Biomedical Programs*, Publications in Automatic Computation 2, University of California Press, Berkeley, 1967, pp. 214a-214s.

2. Davis, John C., *Statistics and Data Analysis in Geology*, John Wiley and Sons, New York, 1973, 550 pp.

APPENDIX B.

LANDSAT MAPPING SYSTEM

PRECEDING PAGE BLANK NOT FILMED

PRECEDING PAGE BLANK NOT FILMED

APPENDIX B

LANDSAT MAPPING SYSTEM

The Landsat Mapping System (LMS) represents a total rewrite of the RECOG or RECOGnition Mapping System (References 1 through 3).* RECOG was intended principally for instruction at Colorado State University, and the new LMS software is compatible with it. However, the new system is specifically structured for both Landsat imagery inputs and composite mapping. Advantages cited for LMS include economy of operation, flexibility, exportability to other CDC computers, and high-volume production (Reference 4).

The LMS software is subdivided into four major image-processing steps. The first step prepares geometrically rectified digital scenes in a given scale or picture-element ground area with Landsat computer-compatible tape inputs (figure B-1). The second step interleaves multitemporal Landsat images, with the added option of overlaying multiple ancillary data planes derived from maps onto the multidate/multitemporal spectral data base (figure B-2). The third phase performs feature extraction computation and optimization of the mapping materials' statistical signatures (figure B-3). The fourth and final step maps the spatial distribution of each desired material as rectified and scaled displays (figure B-4).

The computer cost of LMS for preparing a single classification map of a 1:24,000-scale quadrangle was estimated at $200. (See table B-1.) The LMS cost example involved three multidate images (twelve bands of multispectral data), but no ancillary data was involved. Significant additional economies of scale were realized in the application of many of these LMS modules to the larger-scale Denver land-use/land-cover inventory effort reported here. Forty-eight Landsat image and ancillary map variables were overlayed, classified, and displayed for an average cost of $704.19 per 7.5-minute USGS quadrangle.

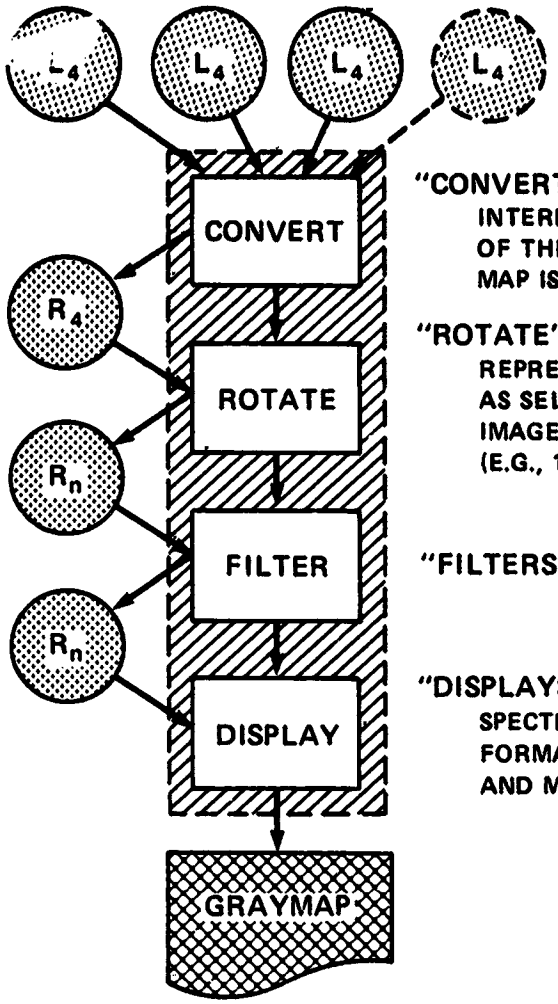
REFERENCES

1. Smith, J. A., L. D. Miller, and T. D. Ells, *Pattern Recognition Routines for Graduate Training in the Automatic Analysis of Remote Sensing Imagery--RECOG*, Science Series 3A, Department of Watershed Sciences, Colorado State University, Fort Collins, 1972, 86 pp.

2. Ells, T., L. D. Miller, and J. A. Smith, *User's Manual for RECOG (Pattern RECOGnition Programs)*. Science Series 3B, Department of Watershed Sciences, Colorado State University, Fort Collins, 1972a, 85 pp.

3. Ells, T., L. D. Miller, and J. A. Smith, *Programmer's Manual for RECOG (Pattern RECOGnition Programs)*. Science Series 3C, Department of Watershed Sciences, Colorado State University, Fort Collins, 1972b, 216 pp.

*The LMS software package was a joint effort of L. D. Miller, E. L. Maxwell, and R. L. Riggs at Colorado State University.



UP TO 4 TAPES, REPRESENTING ~25 MILE E-W SEGMENTS OF A GIVEN LANDSAT IMAGE, MAY BE INPUT SIMULTANEOUSLY.

"CONVERTS" THE LANDSAT FORMAT TAPE(S) INTO THE INTERNAL, SINGLE RECOG TAPE. ONLY THE PORTION OF THE IMAGE NEEDED TO OVERLAY THE SELECTED MAP IS CONVERTED AND POOLED TOGETHER.

"ROTATE" RESAMPLES THE ORIGINAL IMAGE CELLS TO REPRESENT ANY SIZE RECTANGULAR OR SQUARE CELL AS SELECTED BY THE USER. ADJUSTS FOR ORIGINAL IMAGE DISTORTIONS. SCALES IMAGE TO MAP SCALE (E.G., 1:24,000).

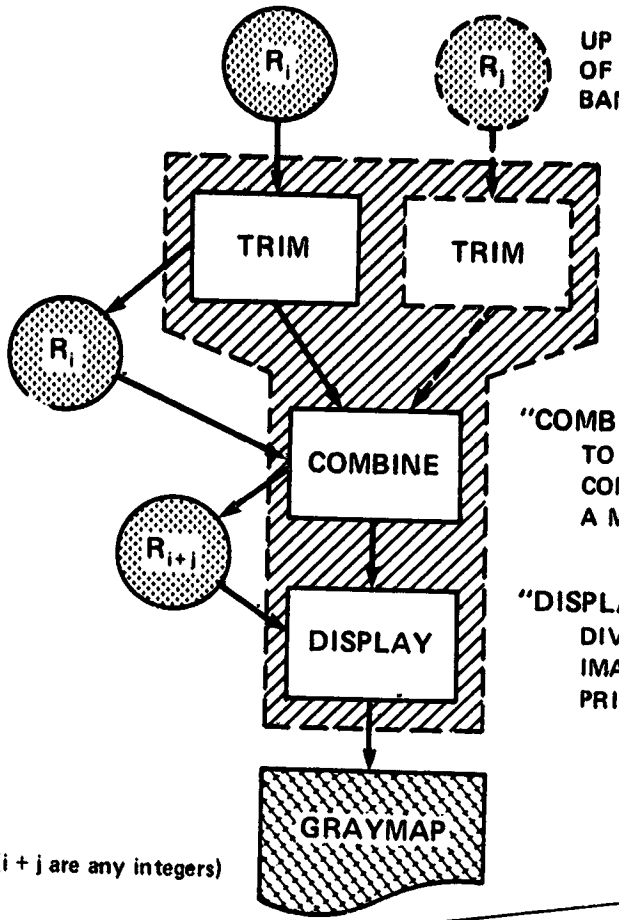
"FILTERS" THE IMAGE.

"DISPLAYS" 1, 2, 3 . . . OR ALL OF THE INDIVIDUAL SPECTRAL BANDS IN THE ORIGINAL OR MAP OVERLAY FORMAT. DISPLAY OPTIONS INCLUDE LINEPRINTER AND MICROFILM GRAYMAPS.

L_4 "LANDSAT" COMPUTER COMPATIBLE TAPE (CCT) AS SUPPLIED BY EROS DATA CENTER.

R_4 "RECOG" FORMATTED TAPE (OR DISK) FILE - AS STANDARD FORMAT TAPE USED THROUGHOUT THE IMAGE PROCESSING ACTIVITY. (n = 1 to 4)

FIGURE B-1. STEP 1: IMAGE PREPARATION/MAP OVERLAY.



UP TO 10 RECOG FORMATTED TAPES OF A VARYING NUMBER OF SPECTRAL BANDS ARE INPUT.

"TRIMS" EACH RECOG FORMATTED TAPE (OR FILE) TO A SELECTED NUMBER OF LINES AND COLUMNS DESIGNATED BY THE USER, USUALLY THOSE NEEDED TO COVER MAP SELECTED. LINES AND COLUMNS ARE RENUMBERED, BEGINNING AT 1, 1.

"COMBINES" RECOG FORMATTED DATA FROM THE 1 TO 10 SEPARATE INPUT TAPES (FILES) INTO 1 COMPOSITE RECOG TAPE (FILE) REPRESENTING A MULTIDATE, MULTISPECTRAL IMAGE.

"DISPLAYS" 1, 2, 3 . . . OR ALL OF THE INDIVIDUAL SPECTRAL BANDS IN COMBINED IMAGE. DISPLAY OPTIONS INCLUDE LINE-PRINTER AND MICROFILM GRAYMAPS.

STEP 2. AUXILIARY PROGRAMS.

"ANCILLARY" CREATES RECOG FORMATTED DATA FROM CELLULARIZED MAP DATA PLANES INPUT IN CARD OR MAGNETIC TAPE FORMAT. MAP CELLS MUST BE THE SAME SIZE OR SOME INTEGER MULTIPLE OF THE CELLS ON THE RECOG FORMATTED DATA WITH WHICH THE ANCILLARY DATA WILL BE COMBINED.

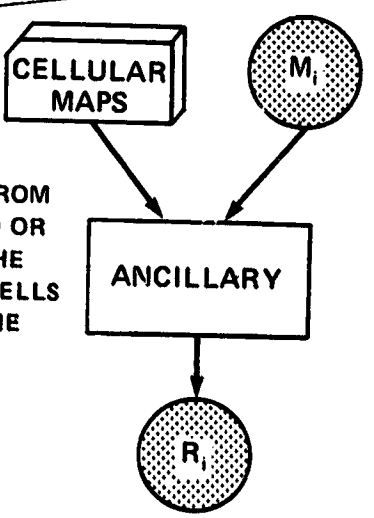
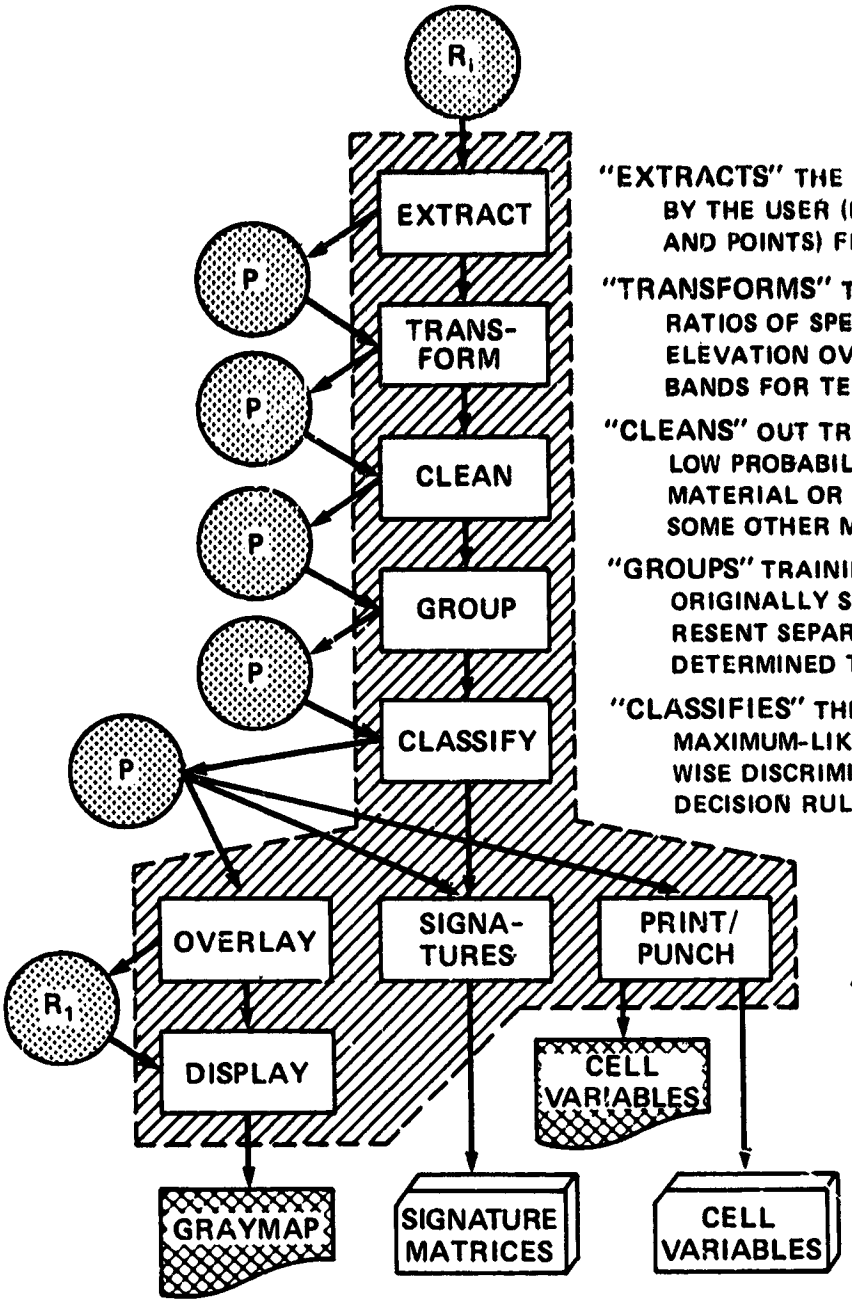


FIGURE B-2. STEP 2: INTERLEAVES IMAGES FROM VARIOUS DATES.



(i is any integer)

"EXTRACTS" THE TRAINING FIELD DATA IDENTIFIED BY THE USER (RECTANGLES, IRREGULAR AREAS, AND POINTS) FROM THE RECOG IMAGE FORMAT.

"TRANSFORMS" THE TRAINING FIELD DATA. FORMS RATIOS OF SPECIFIED SPECTRAL BANDS, USES ELEVATION OVERLAYS TO ADJUST SPECTRAL BANDS FOR TERRAIN SHADOWING, ETC.

"CLEANS" OUT TRAINING FIELD DATA POINTS WITH LOW PROBABILITY OF BEING THE SELECTED MATERIAL OR HIGH PROBABILITY OF BEING SOME OTHER MATERIAL, ETC.

"GROUPS" TRAINING SETS TOGETHER WHICH WERE ORIGINALLY SELECTED IN EXTRACT TO REPRESENT SEPARATE MATERIALS BUT ARE NOW DETERMINED TO BE STATISTICALLY SIMILAR.

"CLASSIFIES" THE TRAINING FIELDS USING MAXIMUM-LIKELIHOOD APPROACH (STEPWISE DISCRIMINANT ANALYSIS). OTHER DECISION RULES CAN BE SUBSTITUTED HERE.

"OVERLAYS" ANY VARIABLE OR RESULT IN POINT FILE INTO A RECOG FORMAT FOR DISPLAY AND MAP OVERLAY.

"SIGNATURES" COMPUTES STATISTICAL REPRESENTATION OF EACH MATERIAL SPECIFIED BY THE USER FOR USE IN MAPPING THESE MATERIALS ON ANY DATA TAKEN FROM THE SAME ORIGINAL IMAGE.

"PRINTS" OR "PUNCHES" OUT ANY VARIABLE(S) IN THE POINT FILE FOR FURTHER ANALYSIS IN ADDITIONAL PROGRAMS WRITTEN BY THE USER.

"POINT" BY POINT TAPE (OR DISK) FILE. - AN INTERNAL TAPE, DISK, AND/OR CARD FILE FORMAT WHICH CONTAINS ONLY THE EXTRACTED TRAINING FIELD DATA AND DOES NOT MAINTAIN ITS CORRECT MAP OVERLAY POSITION.

FIGURE B-3. STEP 3: COMPUTES STATISTICAL "SIGNATURES" OF MATERIALS TO BE MAPPED.

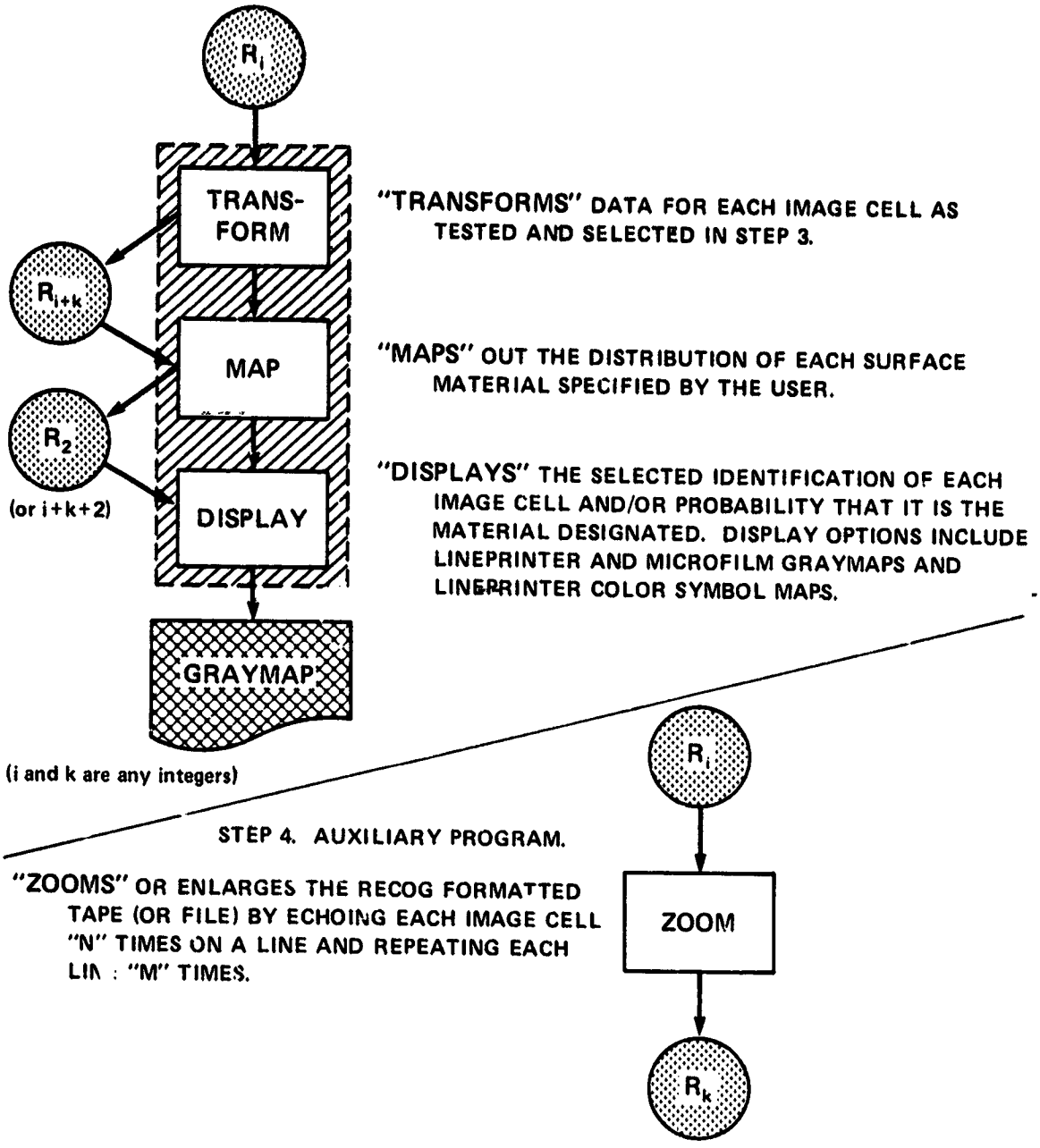


FIGURE B-4. STEP 4: MAPS DISTRIBUTION OF EACH MATERIAL.

TABLE B-1
COST ESTIMATES FOR OPERATING THE LANDSAT MAPPING SYSTEM

Item		Cost Estimate*
CONVERT		$5/date
ROTATE		$7/date
FILTER		$6/date
DISPLAY	$1/band/date X 2 bands =	$2/date
Step 1		$20/date
Assuming three dates involved gives $20/date X 3 =		$60
TRIM		$3/date
COMBINE	3 dates combined =	$1
DISPLAY	$1/band/date X 1 band =	$1
ANCILLARY		optional
Step 2		
Assuming three dates gives $3/date X 3 dates + $1 + $1 =		$11
EXTRACT		$10 (approx.)
TRANSFORM		$5 (approx.)
CLEAN	$2/iteration X 3 iterations =	$6 (approx.)
CLASSIFY	$8/iteration X 3 iterations =	$24 (approx.)
SIGNATURES		$2 (approx.)
OVERLAY		optional
GROUP		optional
PRINT/PUNCH		optional
Step 3		
Based on 2,000 points =		$50 (approx.)
TRANSFORM		$5 (approx.)
MAP	Based on mapping 30 material types	$73 (approx.)
DISPLAY	Black-and-white lineprinter symbol map	$1 (approx.)
ZOOM		optional
Step 4		
Based on 30 classes mapped =		$79 (approx.)
STEP TOTAL*		$200 (approx.)

*Estimated computer costs for 1 of 1:24,000 quad map with ~1-acre cells, three dates (12 spectral bands), 2,000 cells defining training fields, 30 material types, and black-and-white lineprinter display.

PRECEDING PAGE BLANK-NOT FILMED

APPENDIX C

MACHINE CLASSIFICATION ERROR-RATE ESTIMATION

APPENDIX C

MACHINE CLASSIFICATION ERROR-RATE ESTIMATION

The knowledge of the error rate of a machine classifier is desirable for two principal reasons. One is to evaluate its performance relative to other classifiers; the other is to verify the utility of the classifier. Both rationals have been explored in this endeavor.

One approach to error-rate estimation is to compute it from an assumed parametric model. However, this approach presents three problems (Reference 1). First, error-rate estimates of this nature are almost always overoptimistic because characteristics that make the design samples peculiar or unrepresentative will not be revealed. Second, the validity of an assumed parametric model is always suspect. Lastly, it is difficult to compute the error rate exactly, even if the probabilistic structure is completely known.

An empirical approach that obviates these questions is the experimental testing of the classifier. In practice, this is generally done by evaluating the classifier on a test data set and using the proportion of the samples that are misclassified as an estimate of the error rate. The use of the 1972-1973 USGS land-use data plane as a ground-truth surrogate permitted error-rate estimation to be performed on the full Landsat-1 image itself in this endeavor.

The linear-discriminant approach used throughout this study converts an n-dimensional problem to a more manageable one-dimensional problem. This many-to-one mapping, at least in theory, cannot reduce the minimum achievable error rate. However, in general, some of the theoretically attainable accuracy can be sacrificed for the advantages of working solely in one dimension.

The classification algorithm just described is subject to two types of errors prevalent in any multiclass yes-no decision problem. The correct class may be rejected when it is actually correct or when a picture element is designated a particular class when it is not and some other category is correct. These statistical errors are called type I and type II errors, respectively. More commonly, they are referred to as omission and commission errors, respectively, in pattern-recognition literature. The two states for the category (that is, true or false) in a multiclass situation, along with the two decisions that the classifier can make, are indicated in a two-way decision table (table C-1).

TABLE C-1

MACHINE CLASSIFICATION DECISION TABLE. The outcomes for given machine actions in a two-choice decision problem are outlined. (After Reference 2).

Decision	Test Hypothesis	
	True	False
Reject	Omission (type I) error	Correct Decision
Accept	Correct decision	Commission (type II) error

TABLE C-2

FIRST-ORDER LAND-USE CLASSIFICATION MATRIX USING FINAL TEN-VARIABLE LINEAR-DISCRIMINANT ANALYSIS. The verified omission/commission errors of the machine-processed single-date Landsat-1 image are given. The tabulations are based on identifying 24 second- and third-order land uses, and checking only at the first order on a point-to-point comparison with the 1972-1973 USGS land-use data plane. Land-use codes are identified in table 1 of the main text. Correctly classified class totals are italicized. Verification accuracies are summarized in table 24 of the main text.

USGS Land-Use Classes	Number of Picture Elements							Omission Errors (type I)		Commission Errors (type II)	
	Urban	Agriculture	Range	Forest	Water	Barren	Total	Total	Percent	Total	Percent
Urban	*180,781*	13,563	6,287	175	1,292	465	202,563	21,782	10.75	35,383	16.37
Agriculture	29,299	*46,619*	5,296	0	826	157	82,197	35,578	43.28	20,663	30.71
Range	4,005	5,968	*15,788*	307	53	1,923	28,044	12,156	43.70	16,683	51.38
Forest	246	462	780	*3,328*	2	204	5,022	1,694	33.73	906	21.40
Water	1,830	667	72	0	*4,190*	0	6,759	2,569	38.01	2,173	34.15
Barren	3	3	4,248	424	0	*2,513*	7,191	4,678	65.05	2,749	52.24
Total Elements	216,164	67,282	32,471	4,234	6,363	5,262	331,776	78,557	23.68	78,557	23.68

Classification matrices for the ten-variable image analysis of the 576- by 576-element August 15, 1973, Landsat scene of the Denver Metropolitan Area were generated for the six first-order classes (table C-2) and for the 24 second- and third-order categories (table C-3).

REFERENCES

1. Duda, Richard O., and Peter E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973, 482 pp.

2. Mendenhall, William, *Introduction to Linear Models and the Design and Analysis of Experiments*, Wadsworth Publishing Company, Belmont California, 1968, 465 pp.

COMPOSITE SECOND- AND THIRD-
LINEAR-DISCRIMINANT ANALYSIS
Landsat-1 image are given. The tabulat
comparison with the 1972-1973 USGS
rectly classified class totals are italiciz

USGS Land-Use Classes	Number of											
	11	12	121	13	14	15	151	16	19	191	192	21
11	*83,761*	2,140	1,428	1,670	25	783	0	378	2,102	3	0	1,
12	6,891	*3,238*	50	1,544	7	132	0	299	496	1	0	
121	6,129	181	*2,073*	156	24	36	0	185	586	0	0	
13	3,797	949	37	*4,997*	12	606	0	610	350	6	0	
14	1,529	119	34	1,119	*6*	9	0	13	269	2	0	
15	442	199	7	405	30	*2,685*	0	950	4	10	0	
151	486	35	3	177	1	133	*0*	87	35	1	0	
16	8,316	498	196	646	12	36	0	*14,968*	279	20	0	
19	15,027	266	327	971	9	792	0	585	*1,642*	0	0	
191	61	6	0	37	1	0	0	0	0	*0*	0	
192	1,192	3	357	16	0	0	0	0	46	0	*0*	
211	8,637	45	1,665	657	14	2,151	0	1,160	1,890	2	0	*18,*
212	20	0	25	0	0	0	0	0	0	0	0	
213	7,841	17	910	148	26	1,854	0	1,278	959	0	0	6,
31	1,630	21	70	7	10	864	0	1,107	276	0	0	
33	15	0	2	1	2	0	0	0	0	0	0	
411	141	0	19	21	0	0	0	54	11	0	0	
421	0	0	0	0	0	0	0	0	0	0	0	
422	0	0	0	0	0	0	0	0	0	0	0	
51	78	30	1	83	0	0	0	11	0	0	0	
52	780	66	64	239	10	9	0	151	29	0	0	
53	153	2	1	0	16	0	0	84	23	0	0	
74	3	0	0	0	0	0	0	0	0	0	0	
741	0	0	0	0	0	0	0	0	0	0	0	
Total Elements	146,929	7,815	7,269	12,894	205	10,090	0	21,920	8,997	45	0	30,

FOLDOUT FRAME

TABLE C-3

ORDER LAND-USE CLASSIFICATION MATRIX USING FINAL TEN-VARIABLE

. The verified omission/commission errors of the machine-processed, single-date ons are based on identifying 24 second- and third-order land uses on a point-to-point land-use data plane. Land-use codes are identified in table 1 of the main text. Cor- ed. Verification accuracies are summarized in table 24 of the main text.

Picture Elements													Total Elements	Omission Errors (Type I)		Commission Errors (Type II)	
1	212	213	31	33	411	421	422	51	52	53	74	741		Total	Percent	Total	Percent
401	0	3,337	1,100	117	0	87	0	26	235	21	51	11	98,676	14,915	15.12	63,168	42.99
11	0	109	8	0	0	0	0	8	24	7	0	0	12,825	9,587	74.75	4,577	58.57
356	8	2,348	364	154	0	46	0	2	176	58	119	112	13,113	11,040	84.19	5,196	71.48
511	0	155	98	18	0	0	0	1	65	1	8	9	12,330	7,333	59.47	7,897	61.25
500	3	372	480	45	0	42	0	8	177	51	9	37	4,824	4,818	99.88	199	97.07
282	0	702	140	17	0	0	0	11	10	9	19	0	5,922	3,237	54.66	7,405	73.39
51	0	30	36	0	0	0	0	0	41	0	0	0	1,116	1,116	100.00	0	0.00
155	0	788	3,058	108	0	0	0	102	74	3	28	26	21,920	14,345	48.94	6,952	31.72
941	0	1,145	343	9	0	0	0	10	125	11	35	1	22,239	20,597	92.62	7,355	81.75
5	5	1	117	0	0	0	0	0	1	0	0	0	234	234	100.00	45	100.00
89	0	158	75	0	0	0	0	0	34	1	0	0	1,971	1,971	100.00	0	0.00
228	4,327	6,436	1,266	18	0	0	0	88	294	48	0	0	46,926	28,698	61.16	11,796	39.29
207	477	0	0	0	0	0	0	0	0	0	0	0	729	252	34.57	4,984	91.27
459	635	9,850	3,862	150	0	0	0	28	305	63	157	0	34,542	24,692	71.48	21,947	69.02
321	0	5,646	13,876	307	0	0	0	0	27	26	562	531	25,281	11,405	45.11	13,618	49.53
0	0	1	476	1,129	0	307	0	0	0	0	204	626	2,763	1,634	59.14	3,848	77.32
121	6	335	100	0	0	0	0	0	2	0	0	0	810	810	100.00	0	0.00
0	0	0	22	572	0	3,309	0	0	0	0	12	153	4,068	759	18.66	917	21.70
0	0	0	0	86	0	15	4	0	0	0	0	0	144	140	97.22	4	50.00
69	0	0	0	0	0	0	0	1	15	0	0	0	288	287	99.65	286	99.65
162	0	337	66	0	0	0	0	2	2,773	262	0	0	4,950	2,177	43.98	1,859	40.13
55	0	44	6	0	0	0	0	0	254	883	0	0	1,521	638	41.94	561	38.85
0	0	3	633	729	0	64	1	0	0	0	149	128	1,710	1,561	91.29	1,328	89.91
0	0	0	1,368	1,518	0	356	3	0	0	0	124	2,112	5,481	3,369	61.47	1,673	44.20
024	5,461	31,797	27,494	4,977	0	4,226	8	287	4,632	1,444	1,477	3,785	331,776	165,615	49.92	165,615	49.92

2

FOLDOUT

This Page Intentionally Left Blank

PRECEDING PAGE BLANK-NOT FILMED

APPENDIX D

MAXIMUM-LIKELIHOOD RATIO

APPENDIX D

MAXIMUM-LIKELIHOOD RATIO

The maximum-likelihood ratio classifier, such as the classification subroutine GLIKE in CSU's RECOG system (Reference 1), assumes that the data are multivariate normally distributed. It utilizes a Bayesian decision rule of the form

$$p(X|i) = \frac{1}{(2\pi^{N/2} |\Sigma_i|^{1/2}} e^{(1/2)(X-A_i)^T \Sigma_i^{-1} (X-A_i)}$$

where

$p(X|i)$ = the likelihood of occurrence of the feature vector, X, if it belongs to the ith class

Σ_i = the covariance matrix for the ith class

$|\Sigma_i|$ = the determinant of Σ_i

Σ^{-1} = the inverse of Σ_i

A_i = the mean feature vector for the ith class

These conditional probabilities for classes are taken two at a time, ratioed, and evaluated to assign each picture-element vector to the class for which the likelihood, $p(X|i)$, of the unknown vector is the highest, or

$$g_i(X) = p(X|i)$$

The class probabilities, $p(i)$, are assumed to be equal in the maximum-likelihood ratio (Reference 2). However, the modified maximum-likelihood ratio can exploit *a priori* class probabilities by taking

$$g_i(X) = p(i)\, p(X|i)$$

The benefits of having *a priori* class probabilities versus equal class probabilities were significant, as pointed out by comparative tests using the 4,100-point training set (table 22) of the main text.

PRECEDING PAGE BLANK-NOT FILMED

REFERENCES

1. Ells, T., L. D. Miller, and J. A. Smith, *Programmer's Manual for RECOG (Pattern RECOGnition Programs)*, Science Series 3C, Department of Watershed Sciences, Colorado State University, Fort Collins, 1972, 216 pp.

2. Maxwell, Eugene L., "Multivariate Systems Analysis of Multispectral Imagery," *Photogram. Eng. and Remote Sensing*, **42** (9), 1976, pp. 1173-1186.

PRECEDING PAGE BLANK-NOT FILMED

APPENDIX E

CORRELATION MATRICES FOR RECTANGULAR/POINT-SAMPLED TRAINING SETS

APPENDIX E

CORRELATION MATRICES FOR RECTANGULAR/POINT-SAMPLED TRAINING SETS

The CLASSIFY linear-discriminant analysis program operates on a single, within-groups, variance-covariance matrix to compute discriminant functions and for optional canonical analyses. Another useful table generated and printed by this Landsat Mapping System (LMS) program is the d-by-d correlation matrix, $R = [\rho_{ij}]$, where the correlation coefficients, ρ_{ij}, are related to the covariances (or sample covariances) by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\,\sigma_{jj}}}$$

Because $-1 \leq \rho_{ij}^2 \leq 1$, with $\rho_{ij}^2 = 0$ for uncorrelated features and $\rho_{ij}^2 = \pm 1$ for completely correlated features, the correlation matrix can be useful, particularly in a nonstepwise analysis in which many variables are considered. The presence of two features, for which ρ_{ij}^2 is relatively large, provides redundant information and creates possibilities for dimensionality reduction.

Even with the stepwise variable-entry feature of CLASSIFY, the correlation matrix is useful for perceiving numerical relationships between variables. Therefore, correlation matrices were prepared for the four original MSS bands (table E-1), the ten MSS bands/ratios (table E-2), and the seven-variable optimal classifier (table E-3), using the rectangular training-set concept tested and discarded as being unrepresentative. It was also instructive for comparison to include corresponding correlation matrices for the four original MSS bands (table E-4) and the ten MSS bands/ratios (table E-5) derived from the 1/81 grid-sampled point training set, as well as for the final ten-variable correlation matrix (table E-6) used to classify the 576- by 576-element Landsat/ancillary variable scene. The size of the 48-variable correlation matrices precluded their inclusion here.

TABLE E-1

WITHIN-GROUPS CORRELATION MATRIX FOR THE FOUR ORIGINAL MSS BANDS DERIVED FROM RECTANGULAR TRAINING SETS. This set of variables correctly classified 65.19 percent of the 2,413 total samples, but the rectangular training sets later proved to be unrepresentative of the larger scene. Only the lower diagonal of the matrix is shown because the upper and lower diagonals of the matrix are symmetrical. The Landsat-1 image variables were added in a free stepwise fashion. The entry order is indicated by superscript numbers on the variable names.

Bands	Bands			
	4^3	5^2	6^4	7^1
4	1.000			
5	0.891	1.000		
6	0.457	0.423	1.000	
7	0.255	0.201	0.635	1.000

TABLE E-2

WITHIN-GROUPS CORRELATION MATRIX FOR THE FOUR ORIGINAL MSS BANDS AND SIX RATIOS DERIVED FROM RECTANGULAR TRAINING SETS. This set of variables correctly classified 67.30 percent of the 2,413 total samples, but the rectangular training sets later proved to be non-representative of the larger scene. Only the lower diagonal of the matrix is shown because the upper and lower diagonals are symmetrical. The Landsat-1 image variables were added in a free stepwise fashion. The entry order is indicated by superscript numbers on the variable names.

Bands/ Ratios	Bands/Ratios									
	4^5	5^4	6^9	7^1	$7/6^3$	$7/5^{10}$	$7/4^7$	$5/4^2$	$5/6^8$	$6/4^6$
4	1.000									
5	0.891	1.000								
6	0.456	0.423	1.000							
7	0.255	0.201	0.635	1.000						
7/6	-0.195	-0.233	-0.179	0.452	1.000					
7/5	-0.289	-0.418	0.179	0.507	0.394	1.000				
7/4	-0.377	-0.382	0.301	0.731	0.517	0.732	1.000			
5/4	0.368	0.719	0.137	-0.040	-0.193	-0.514	-0.302	1.000		
5/6	0.249	0.358	-0.365	-0.238	-0.082	-0.336	-0.374	0.391	1.000	
6/4	-0.320	-0.297	0.642	0.496	0.027	0.537	0.755	-0.197	-0.578	1.000

TABLE E-3

WITHIN-GROUPS CORRELATION MATRIX FOR THE OPTIMAL SET OF THREE MSS SPECTRAL BANDS AND FOUR LANDSCAPE VARIABLES DERIVED FROM RECTANGULAR TRAINING SETS. This set of variables correctly classified 96.64 percent of the 2,413 total samples, but the rectangular training sets later proved to be unrepresentative of the larger scene. Only the lower diagonal of the matrix is shown because the upper and lower diagonals are symmetrical. The Landsat-1 image variables were forced in the predetermined order (table 13 of the main text), and the ancillary landscape variables were added in a free stepwise fashion. The entry order is indicated by superscript numbers on the variable names. ACRES = total census tract acreage. ELEVS = topographic elevation, INTER = freeway interchange minimum distances; and URBAN = built-up urban-area minimum distances.

Variables	Variables						
	MSS-4^3	MSS-5^2	MSS-7^1	ACRES^4	ELEVS^6	INTER^7	URBAN^5
MSS-4	1.000						
MSS-5	0.891	1.000					
MSS-7	0.255	0.201	1.000				
ACRES	-0.035	-0.034	-0.011	1.000			
ELEVS	-0.032	0.015	-0.001	0.057	1.000		
INTER	0.100	0.111	0.042	0.396	-0.089	1.000	
URBAN	0.026	0.049	0.017	0.508	0.014	0.713	1.000

TABLE E-4

WITHIN-GROUPS CORRELATION MATRIX FOR THE FOUR ORIGINAL MSS BANDS DERIVED FROM GRID-SAMPLED TRAINING POINTS. This set of variables correctly classified 37.90 percent of the 4,100 total samples. Only the lower diagonal of the matrix is shown because the upper and lower diagonals are symmetrical. The Landsat-1 image variables were added in a free stepwise fashion. The entry order is indicated by superscript numbers on the variable names.

Bands	Bands			
	4^2	5^3	6^4	7^1
4	1.000			
5	0.914	1.000		
6	0.492	0.439	1.000	
7	0.169	0.101	0.770	1.000

TABLE E-5

WITHIN-GROUPS CORRELATION MATRIX FOR THE FOUR ORIGINAL MSS BANDS AND SIX RATIOS DERIVED FROM GRID-SAMPLED TRAINING POINTS. This set of variables correctly classified 38.41 percent of the 4,100 total points. Only the lower diagonal of the matrix is shown because the upper and lower diagonals are symmetrical. The Landsat-1 image variables were added in a free stepwise fashion. The entry order is indicated by superscript numbers on the variable names.

Bands/ Ratios	Bands/Ratios									
	4^4	5^5	6^9	7^1	$7/6^8$	$7/5^3$	$7/4^6$	$5/4^2$	$5/6^7$	$6/4^{10}$
4	1.000									
5	0.914	1.000								
6	0.492	0.439	1.000							
7	0.169	0.101	0.770	1.000						
7/6	-0.172	-0.180	0.105	0.256	1.000					
7/5	-0.530	-0.626	0.233	0.626	0.305	1.000				
7/4	-0.482	-0.496	0.359	0.758	0.334	0.934	1.000			
5/4	0.557	0.834	0.226	-0.047	-0.139	-0.639	-0.416	1.000		
5/6	0.350	0.449	-0.151	-0.425	-0.103	-0.608	-0.601	0.460	1.000	
6/4	-0.382	-0.375	0.595	0.683	0.282	0.794	0.871	-0.294	-0.534	1.000

TABLE E-6

WITHIN-GROUPS CORRELATION MATRIX FOR THE TEN-VARIABLE COMBINATION OF LANDSAT AND LANDSCAPE VARIABLES FOR FULL-IMAGE CLASSIFICATION DERIVED FROM GRID-SAMPLED TRAINING POINTS. This set of variables correctly classified 51.00 percent of the 4,100 total points. Only the lower diagonal of the matrix is shown because the upper and lower diagonals are symmetrical. The Landsat-1 image bands and the ancillary landscape variables were added in a free stepwise fashion. The entry order is indicated by superscript numbers on the variable names. CDP10 = median housing-unit value; CDP18 = average number of families per acre; CDP19 = average number of year-round housing units per acre; SLOPE = topographic slope; ELEVS = topographic elevation; and HI9 = built-up urban-area minimum distances.

Variables	Variables									
	MSS-4[10]	MSS-7[4]	CDP10[8]	CDP18[3]	CDP19[6]	SLOPE[5]	ELEVS[1]	HI9[2]	7/5[9]	5/4[7]
MSS-4	1.000									
MSS-7	0.169	1.000								
CDP10	-0.128	-0.015	1.000							
CDP18	0.069	0.059	0.447	1.000						
CDP19	-0.057	-0.084	-0.120	0.016	1.000					
SLOPE	-0.035	-0.026	0.100	0.072	-0.064	1.000				
ELEVS	-0.042	-0.057	0.259	0.166	-0.218	0.319	1.000			
HI9	-0.022	-0.041	-0.030	-0.076	-0.196	0.103	0.099	1.000		
7/5	-0.527	0.626	0.055	-0.011	0.038	-0.014	-0.055	-0.023	1.000	
5/4	0.557	-0.047	-0.059	0.014	-0.140	0.006	0.064	0.060	-0.639	1.000