PATTERN RECOGNITION APPLIED TO URANIUM PROSPECTING

by

PETER LAURENCE BRIGGS

S.B. Physics, M.I.T. (1971)

S.B. Earth and Planetary Sciences, M.I.T. (1971)

S.M. Earth and Planetary Sciences, M.I.T. (1976)

SUBMITTED IN PARTIAL

FULFILLMENT OF THE REQUIREMENTS

FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February, 1978

Signature of Author .
Department of Earth and Planetary Sciences, February, 1978

Certified by ....

and ....
Thesis Supervisors

Accepted by ....
Chairman, Departmental Committee on Graduate Students

PATTERN RECOGNITION APPLIED TO URANIUM PROSPECTING

by

Peter Laurence Briggs

Submitted to the Department of Earth and Planetary Sciences on February 7, 1978, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

In exploration for minerals and hydrocarbons, prospecting targets are commonly selected on the basis of a geologist's subjective interpretation of a combination of diverse geological data. This thesis investigates the possibility of automating and standardizing this interpretation task. Four pattern recognition algorithms that provide quantitative, reproducible means of coding, organizing, and interpreting geological data are used here to make exploration decisions. Although data collection is still a fundamental, somewhat subjective input, the remainder of a combined interpretation problem is handled in an automated, algorithmic fashion.

Reconnaissance level data are used to estimate favorability for sandstone-type uranium deposits on the Colorado Plateau and in the Casper Quadrangle of central Wyoming. Pattern recognition procedures are used to identify geological features that mark areas favorable for ore. Pattern recognition algorithms provide a logical framework for organizing these features for the recognition of areas favorable for uranium ore occurrence. Automated data evaluations have produced geologically reasonable predictions of new exploration targets.

Variations in the performance of these four algorithms suggest guides to the use of these and other pattern recognition procedures in geological problems. Control experiments test the predictive potential of these techniques and verify the stability of pattern recognition analyses.

Pattern recognition techniques may be useful in a variety of exploration problems where large amounts of diverse data must be winnowed, integrated, and interpreted for decision making.

Thesis Supervisors:
B.C. Burchfiel, Professor of Geology
I.T. Young, Associate Professor of Electrical Engineering

ACKNOWLEDGEMENTS

If the customary litany of indebtedness is foregone, my gratitude to Roxanne Regan and Steven Hellinger may seem more in proportion to the kindness and encouragement they have shown me. I would like to thank also my thesis supervisors, B.C. Burchfiel and I.T. Young, who oversaw this project and tolerated my occasionally reclusive work habits (particularly Ian, whose enthusiastic teaching stimulated my interest in pattern recognition problems). David Emilia supervised data collection for the Casper Quadrangle at E.R.D.A., Grand Junction. This research was sponsored by the United States Energy Research and Development Administration.

Hopefully my parents will take some pleasure in this document.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

CHAPTER 1

Computer programs that simulate modes of human learning can provide a quantitative, logical framework for the combined interpretation of diverse geological data. This work considers uranium deposits of the Casper, Wyoming Quadrangle and of the Colorado Plateau for testing the ability of four pattern classification procedures to provide computerized combined interpretations of geological data for mineral exploration. Each of the four algorithms offers a different model of the way an exploration geologist might think. So that their performance characteristics and individual merits may be compared, each algorithm is applied to the same tasks.

Reconnaissance level geological data covering the entire Colorado Plateau and the entire Casper Quadrangle are searched by machine in much the same way that a geologist might scan individual surveys to find geologic signatures that mark uranium-producing areas. The work of multiple survey interpretation to designate unexplored ground as favorable or unfavorable for exploration is taken over by pattern recognition algorithms that use these geologic signatures in quantified decision processes that replace conventional subjective interpretation.

Mineral deposits are commonly discovered in three ways - by accident, by exhaustive search, or by inference from geological data. Ancient ore finds were made without input from a reliable body of geological knowledge, and must have resulted from accidental discovery. Today, readily located ore bodies with surface shows of ore are decreasing in number. As shallow ore bodies are exhausted, new discoveries must be made at depth, and surface shows of a mineral may no longer be expected to mark ore occurrences. Accidental discovery cannot be relied upon to secure the deeper ore deposits because work at depth is not within the means of amateur prospectors.

In historical times, prospectors have relied on exhaustive search to find ores. The ephemeral boom towns of the American frontier offer many examples of small land areas saturated with prospectors searching the ground for gold, silver, or other metals. Because exhaustive search guarantees results, some propose, even today, that mineral or hydrocarbon exploration be based on extensive programs of closely-spaced drilling. Saturation ground coverage by exhaustive drilling programs may be logistically and economically the least feasible method for locating ore.

Because reserves of many minerals remain finite while consumption is increasing, a variety of exploration survey techniques, some newly borrowed from petroleum exploration, are now being used to assess the favorability-for-ore of land parcels. Despite the numerous exploration tools available to the mining industry, and the work of many expert geologists, accidental discovery still accounts for a substantial fraction (perhaps a majority*) of new discoveries of uranium and other minerals. Accidental discovery is neither an aggressive exploration policy nor one that can be relied upon to provide a continuous supply of mineral resources. To make exploration more economical and efficient, standard predictive geologic capabilities should be developed and augmented by new analytical techniques whenever possible.

The intercomparison and interpretation of exploration surveys by geologists is now of increasing importance in the search for many types of ore deposits. In modern mineral exploration work, a large amount of data must be organized, winnowed, and interpreted before promising targets for exploration can be selected. No techniques

---

* This suggestion is difficult to prove, but reflects the opinion of several geologists in the mineral industry interviewed by the author.

other than drilling are available for the direct detection of ore; geological and geophysical surveys observe the objects of interest only indirectly. The diversity and ambiguity of data available from geological and geophysical surveys often make a well-integrated and orderly interpretation difficult to obtain. The emergence of new survey techniques, such as LANDSAT imagery, water and soil gas analyses, and new logging tools may further complicate exploration work, as these survey tools may provide data of unfamiliar or uncertain interpretation and of unknown relation to more conventional surveys. The quantity and diversity of data now available in exploration may conspire to make conventional, subjective attempts at combined interpretations suboptimal.

Pattern recognition techniques can unite very diverse data types (numeric, quantized, and descriptive data) into a single logical framework for learning and interpretation. The pattern recognition algorithms used here offer simple but useful simulations of human learning habits. Computers may "learn" about ore deposits and predict their occurrence in somewhat the same way that a geologist does. Clearly defined learning algorithms may be used to enhance the interpretive power of the geologist by using computers to produce quantitative, reproducible evaluations of prospects.

To select targets for prospecting, information from geological, geochemical, and geophysical surveys must all be integrated into a decision process. Conventionally, combined interpretation is performed by a geologist who uses rules learned consciously or unconsciously from his previous experience with ore occurrences to organize information and arrive at a qualitative evaluation of a prospect's potential.

Subjective interpretation of data has some undesirable features that may be obviated by use of computerized combined interpretation. For example, a given geologist's interpretation of data may be non-unique, and possibly non-reproducible. Different experts commonly reach varying conclusions from the same body of data. Different geologists may also use somewhat different sets of learned rules to guide their interpretations; they may weight the significance of particular pieces of information differently, each according to his personal learning history and experience. Also, even a skilled interpretor may not be able to state explicitly, or even be fully aware of, all the rules for combined interpretation that have contributed to his evaluation. Nor is he likely to know just how his mind has combined these rules to work together; hunches and "geologic intuition" may significantly color exploration recommendations and decisions. Finally, and perhaps most importantly, there is a combinatorial problem involved in most modern exploration work. There is rather little difficulty in picking the outstanding features of a single survey. When several types of surveys are available over an area, however, the opportunity to consider each data item in the context of all other data arises. Exploring all the possible interrelations among surveys may be beyond the computing and memory capabilities of even skilled geologists.

Given the complexity of ore-forming systems, quantified, computerized methods for combined interpretation of exploration surveys offer a reasonable way of synthesizing diverse geological information toward exploration goals. Two complimentary strategies are now emerging to model geologic interpretation. These two strategies may be classed broadly as artificial intelligence techniques and pattern recognition techniques. The artificial intelligence approach uses rules for interpretation taken by interview from geologists with expertise in a particular geological problem. These expert interpretors or prospectors also provide estimates of the relative importance and the interrelationships of their rules for interpretation. To the extent that it is known, the structure of the logical or probabilistic interrelationships among the rules is programmed into computer software as a network or net of evidences and hypotheses (observations and plausible conclusions) about geological objects (Duda et al., 1976, 1977). The information net is complete before any geological objects are analyzed. Data describing an area of interest with unknown resource potential are given to the net which then generates an estimate of the probability of finding an ore deposit in the unexplored area. In the artificial intelligence approach, experts specify the interpretation rules, their weights, and their interrelationships. In contrast, the pattern recognition approaches to interpretation presented here start with data describing known resource areas and attempt to recognize in these data regularities that can form a basis for interpretation rules. The specific rules for interpretation generated by generalized learning algorithms can then be used to recognize the resource potential of unfamiliar geological objects. If the data base permits, the computer-generated rules for combined interpretation may lend new insight into geological processes. In a sense, the artificial intelligence approach tries to apply the extant, conventional wisdom to whatever data are at hand, while pattern recognition tries to draw forward the most effective set of rules for interpretation of the available data.

The strategy used here for recognition of resource potential proceeds as follows. Geological data describing areas of known mineral production are contrasted with data describing surrounding areas that are initially presumed to be barren. The computer collects salient features of the available data that contribute to the distinction of barren from producing areas. Some of these features within a given data base might well be overlooked by human interpretors. Each piece of geologic evidence relating to the favorability-for-ore is then weighed in a quantitative, reproducible way, according to quantitative decision criteria. After data assembly, learning algorithms that consider individual pieces of evidence or synergistic combinations of evidence can be used to find signatures of uranium-bearing areas within data. Contrasting characteristic signatures of barren areas are also sought. If a signature exists, and the computer has shown an ability to correctly classify a training set of locations as uranium-producing or barren, a predictive phase may be entered wherein new uranium prospects can be selected by applying the learned interpretation rules to geologic data describing new locations with unknown uranium reserves.

In Chapter 2 of this thesis, current theories of uranium ore genesis are reviewed for the peneconcordant Colorado Plateau-type ores and for the roll-type deposits of Wyoming. This review provides a background for understanding and evaluating the results of pattern recognition surveys of the Wyoming and Colorado Plateau areas. Chapter 3 describes the selection of features from a large data base, and the feature coding procedures used here. The significance of features and their individual interpretations are also discussed. Chapter 4 presents the individual characteristics of the four pattern classification algorithms applied to the feature data base. Chapter 5 presents the results of computerized combined interpretation for the entire Colorado Plateau and Wyoming study areas. The geological "reasonableness" of the four algorithms' performance are discussed. Chapter 6 investigates the stability of recognition and the performance of classification algorithms with various combinations of features. Chapter 7 offers control experiments that test for self-deception in automated combined interpretations and simulate predictive use of these classifiers. Chapter 8 presents conclusions from this work and suggests other geological problems that might profitably be approached by pattern recognition treatment.

CHAPTER 2

The two areas surveyed here for uranium favorability are the Colorado Plateau and the Casper Quadrangle of Wyoming. Together, the Colorado Plateau and Wyoming basins have accounted for about 90% of uranium production in the United States to date. In addition to considerable past production, these two areas are of current interest to prospectors because they are credited with large possible and speculative uranium reserves.

The Colorado Plateau and Wyoming study areas also provide particularly interesting venues for pattern recognition prospecting because the results of three decades' active field exploration can serve as a control or standard of comparison for these pattern recognition surveys. Although the uranium deposits of these two areas have been objects of geologic interest for some time, their origins are still incompletely understood. The basis for computerized recognition of uranium-favorable areas should show a reasonable similarity to existing theories of ore genesis, and might also provide new insights into ore genesis.

This chapter presents a brief geologic history of both study areas, describes the origin of the penecon-cordant and roll-front types of uranium deposits characteristic of the Colorado Plateau and Wyoming basins, respectively. In the context of modern theories of ore genesis, the features for recognition appear to have reasonable geologic interpretations. One of the learning algorithms used here synthesizes these individual features into compound features that are easily interpreted. These compound features are the computer's way of generating a model of ore deposition, and can be understood as parallels to parts of the theories of ore genesis developed by geologists. The data available here (see Chapter 3) are not fully adequate to locate ore-bearing areas. The incomplete ability of the present data to resolve uranium-favorable and -unfavorable areas can also be understood in light of accepted models of ore deposition.

2.1 The Colorado Plateau

The Colorado Plateau structural province covers 140,000 square miles of Colorado, Utah, Arizona, and New Mexico (Figure 2-1). The plateau contains a number of well-developed mining areas and has produced more uranium than any other province in the United States (197,800 tons U_3O_8 to 1976). Despite previous production, even pessimistic analyses suggest that considerable ore may remain undiscovered (Lieberman, 1976), and current

FIGURE 2-1: Major tectonic blocks of the Colorado Plateau study area.

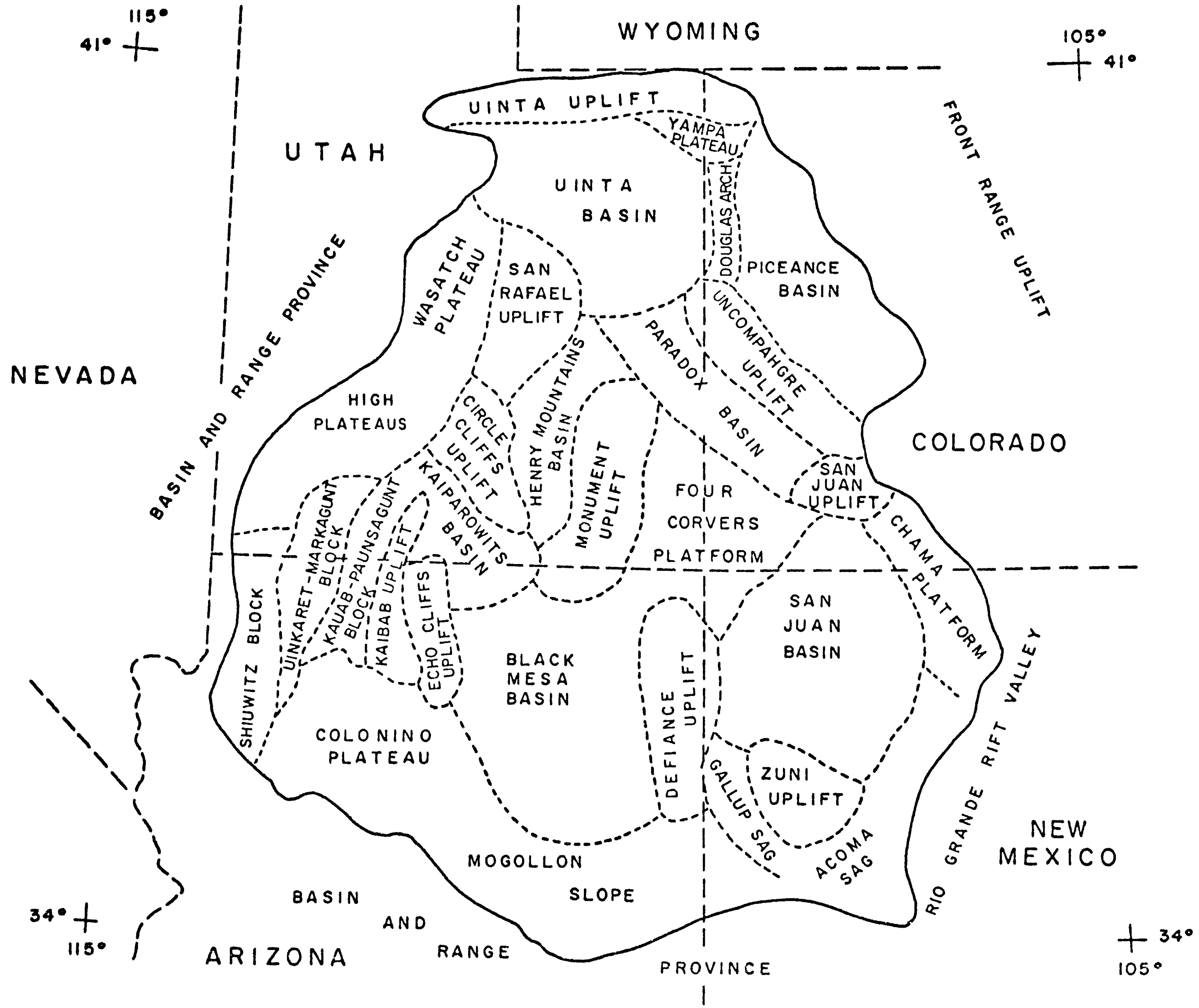

115°
41°
WYOMING
105°
41°
UINTA UPLIFT
YAMPA PLATEAU
UTAH
UINTA BASIN
DOUGLAS ARCH
FRONT RANGE UPLIFT
BASIN AND RANGE PROVINCE
WASATCH PLATEAU
SAN RAFAEL UPLIFT
PICEANCE BASIN
UNCOMPAHGRE UPLIFT
PARADOX BASIN
NEVADA
HIGH PLATEAUS
CIRCLE CLIFFS UPLIFT
HENRY MOUNTAINS BASIN
MONUMENT UPLIFT
COLORADO
SAN JUAN UPLIFT
FOUR CORVERS PLATFORM
KAIPAROWITS BASIN
UINKARET-MARKAGUNT BLOCK
KAUAB-PAUNSAGUNT BLOCK
KAIBAB UPLIFT
ECHO CLIFFS UPLIFT
CHAMA PLATFORM
SAN JUAN BASIN
SHIUWITZ BLOCK
BLACK MESA BASIN
DEFIANCE UPLIFT
RIO GRANDE RIFT VALLEY
COLONINO PLATEAU
ZUNI UPLIFT
GALLUP SAG
ACOMA SAG
NEW MEXICO
MOGOLLON SLOPE
BASIN AND RANGE PROVINCE
34°
115°
ARIZONA
34°
105°

FIGURE 2-1

reserve estimates offer hope of both continued production and new discoveries on the Plateau in the future. Potential resources of $30/lb. U_3O_8 on the Plateau break down as follows: probable, 433,000 tons; possible, 632,000 tons; speculative, 103,000 tons (U.S.E.R.D.A., 1976).

The Colorado Plateau has existed as a structural unit since the Cambrian. Today, high plateaus form a roughly circular province about 500 miles in diameter. Seven major sedimentary basins cover 1/3 of the area and nine major uplifts cover another 1/5 of the area. A sequence of sedimentary rocks of Paleozoic age and younger are widespread; metamorphic and granitic rocks of Precambrian age crop out in the Uncompahgre and Zuni uplifts and in the Black and Grand Canyons. Younger volcanic and intrusive rocks, mostly of Upper Cretaceous age, punctuate the sedimentary cover of the Plateau.

Since Cambrian time, the Colorado Plateau has undergone two major episodes of deformation. The Laramide orogeny of Late Cretaceous and Early Tertiary time was contemporaneous with uranium deposition on the Plateau, and had a direct influence on uranium deposition. An earlier episode of deformation during Pennsylvanian and Permian time also influenced ore deposition. Pre-ore tectonic events influenced not only the paleohydrology and erosional history of the Plateau, but also the environment in which younger ore-bearing sediments were deposited. Older structural features also guided later Laramide igneous intrusions and the trends and forms of Laramide uplifts and basins. Pre-existing structures have influenced later geologic evolution of the Plateau for the last 300 million years (Kelly, 1955).

The Colorado Plateau was stable during most of the Paleozoic. In the Pennsylvanian, the first episode of deformation began, elevating the Zuni and Uncompahgre uplifts. Through the Triassic, sedimentation proceeded over much of the Plateau as major basins and uplifts began to appear. Principal uranium-bearing formations, the Chinle and Morrison, were deposited in the Late Triassic and Late Jurassic, respectively. In the Late Cretaceous, the Laramide orogeny rejuvenated major features such as the San Juan, Black Mesa, Uinta and Piceance Basins, the Monument, Kaibab, Defiance, San Rafael, Circle Cliffs, Zuni, and Uncompahgre uplifts (Figure 2-1). Numerous laccoliths and smaller igneous bodies intruded Plateau sediments in the Late Cretaceous; volcanic fields were also active during this time. Many smaller structures including folds, fractures, anticlines, and synclines that may have locally influenced ore deposition also formed in Laramide time. Uranium-lead dates indicate a Late Cretaceous, Early Tertiary episode of uranium mineralization throughout the Plateau. (Shoemaker, 1955). A period of crustal quiescence persisted from Larimide time to the Miocene. During the Miocene, the entire Colorado Plateau was uplifted. This uplift rejuvenated streams and groundwater circulation etching the paleodrainage pattern into the Plateau. Some minor structural features formed at this time, guided by pre-existing structural trends. Minor structural adjustments continued into the Quaternary.

For much of the last half aeon, ancient structural features have persistently influenced structural change on the Colorado Plateau. Paleostructural and hydrologic factors critically affected ore emplacement, because uranium was transported by and precipitated from circulating groundwaters. The persistence of geological structures on the Plateau suggests that there should be some indication in pattern recognition features and results that the structures now visible on the Colorado Plateau are related in a geologically reasonable way to paleostructural features and to the locations of uranium ores. Structural features, though not the sole controls of ore deposition, should combine with lithologic information to play an important role in qualifying areas as favorable for uranium ore deposition.

Uranium deposits are widespread in the continental sedimentary rocks that cover the Colorado Plateau. Although sedimentary formations ranging in age from Pennsylvanian to Tertiary are known to contain ore-grade uranium, most production has been from the Morrison Formation (in the Grants, New Mexico, and Uravan, Colorado/Utah mineral belts) and from the Triassic Chinle Formation (particularly the Shinarump Member, Lisbon Valley and White Canyon, Utah and Monument Valley, Arizona). Significant production has also come from the Permian Cutler Formation (Paradox Basin, Colorado) and the Jurassic Todilto Limestone (Grants, New Mexico). Ore-bearing rocks of the Morrison, Chinle, and Cutler Formations are quartzose or arkosic, lenticular cross-bedded fluvial sandstones with interbedded clay and mudstone lenses; the Shinarump Member of the Chinle is a sandstone and conglomerate. Local variations in the sedimentary environment within these formations appear to have had a primary influence on the flow of uranium-bearing groundwaters and the deposition of uranium.

By far, the most numerous and productive ore bodies on the Colorado Plateau are the peneconcordant type that occur in gently dipping sandstone beds. These ore bodies are typically tabular - a few feet in thickness, but extending laterally for hundreds of feet, often in ancient stream channels, between enclosing sedimentary members. Within these tabular volumes, uranium minerals coat and, in some instances, replace grains in the host rocks. Most deposits occur in fluvial sediments that are micaceous or arkosic. Fossil organic matter is usually associated with the ore zones. The most favored host rocks are locally thickened sedimentary members containing mudstone, shale, or clay lenses interbedded with the sandstone. Large volumes of lithologically favorable rock contain no known ore.

Four possibilities (summarized by Kerr, 1957) have been advanced to explain the origin of Colorado Plateau uranium ores. Each has its strengths and shortcomings, so that the question of ore genesis is neither fully understood nor finally settled.

In one scenario, igneous activity associated with the Laramide orogeny is of genetic importance. Heated, mineralized solutions derived from magmas carried, among others, compounds of uranium. These solutions mixed with groundwaters, enriching them in uranium. As groundwaters circulated laterally away from uranium sources, they encountered fractures that provided conduits for vertical movement between beds. Thus uranium-rich groundwaters could flow through a great volume of the more-or-less permeable sedimentary beds of the Plateau. Occasionally, these fluids passed through volumes of rock enclosing chemical constituents (e.g. fossil humic material) that caused the reduction of uranium compounds. These reduced compounds of uranium are relatively insoluble in water and precipitated, forming the interstitial ore minerals found in sediments.

A second theory contends that normal groundwater concentrations of uranium derived from leaching of basement rock were adequate to form ore deposits as they circulated through the sediments. These waters encountered zones containing carbonaceous reductants that precipitated uranium compounds. In some deposits, however, solutions clearly heated above normal groundwater temperatures were involved in ore formation. Also, the volumes of basement rock postulated to have been leached and the volume of groundwater needed to form deposits may be prohibitively large (Fisher, 1974).

A third theory notes that deposits of volcanic ash are widespread in both Upper and Lower Mesozoic formations, including the Chinle and Morrison. Uranium, assumed to be present in the ash, could have been leached by meteoric waters and precipitated in deeper strata that were locally favorable. Some workers claim that this theory is stratigraphically and/or geometrically inadequate. The widespread mineralization of collapse features, breccia pipes and fractures, the influence of structural controls on deposition, and the mineralization accompanying uranium in some deposits all cast doubt on this theory.

A syngenetic origin of ores is also possible. The large areas of Precambrian basement that were exposed during the evolution of the Plateau may have been eroded and leached repeatedly, providing surface waters with abnormally high concentrations of uranium. As these solutions circulated through the near-surface, uranium deposits may have formed contemporaneously with the enclosing sediments. Decaying plant material in stream channels might have served to precipitate uranium, as mineralized logs and plants are not uncommon in Plateau ore bodies. This theory has difficulty, however, explaining the many similar uranium deposits of nearly coincident U/Pb ages that occur through a large part of the stratigraphic column.

Though each of these theories differ substantially in detail, they have essential elements in common. Uranium in an oxidized state is transported down-dip and laterally, possibly for great distances, in aqueous solution to the site of deposition. Ore grade deposits developed within formations when uranium-bearing solutions invaded locally favorable environments. Precipitation of uranium minerals occurred where oxidation/reduction reactions involving uranium compounds could proceed - in geochemically reducing environments and/or perhaps in areas of groundwater flow stagnation. Structural and lithologic controls, an adequate source of uranium, and a reducing environment are critical elements in all these theories.

For prospecting, one would like to predict the occurrence of all these features. Subtle lithologic variations at depth and the accidental occurrence of locally reducing zones may never be predictable. Structural influences on groundwater circulation and some possible sources of uranium, however, should be recognizable and useful for prediction. These elements do appear in the features for recognition presented in Chapter 3.

2.2 The Casper, Wyoming Quadrangle

The Casper Quadrangle, within the Wyoming Basins structural province, covers the area in central Wyoming from 42° and 43°N and from 106° to 108°W (Figure 2-2). There are four well-developed mining districts within the Quadrangle - Gas Hills and Crooks Gap in the west, and Shirley Basin and Poison Spider in the east. These, together with other mining areas within the Wyoming Basins, trail only the Colorado Plateau province in uranium production. Fifty percent (165,000 tons U_3O_8) of probable reserves for the Wyoming Basins are within the Quadrangle; most possible and speculative reserves of the Wyoming Basins are believed to lie outside the Quadrangle (U.S.E.R.D.A., 1976).

The uranium deposits of the Wyoming Basins are reviewed by Harshman (1972), Rackley (1972), Melin (1964), and Sharp and Gibbons (1964). During the Precambrian, rocks in the Casper Quadrangle were folded, metamorphosed, and intruded by granitic batholiths and mafic dikes. A long period of erosion reduced the area to a nearly flat surface that was repeatedly transgressed by epicontinental seas during the Paleozoic. During these transgressions, a thick series of marine, littoral, and continental sediments accumulated in the area, culminating in the Jurassic with the Morrison Formation. At the end of the Jurassic, the seas migrated eastward for a final time, and clastic material eroded from highlands to the west

FIGURE 2-2: Outline map of Wyoming showing location of Casper Quadrangle. The major tectonic units in and near the Casper Quadrangle are: 1) Powder River Basin; 2) Bighorn Uplift; 3) Bighorn Basin; 4) Owl Creek Uplift; 5) Wind River Basin; 6) Wind River Uplift; 7) Green River Basin; 8) Rock Springs Uplift; 9) Red Desert Basin; 10) Rawlins Uplift; 11) Hanna Basin; 12) Laramie Basin; 13) Laramie Uplift; 14) Hartville Uplift; 15) Casper Arch; 16) Sweetwater Arch; 17) Shirley Basin; 18) Shirley Mountains.
Uranium Mining Areas: A) Gas Hills; B) Crooks Gap; C) Shirley Basin; D) Poison Spider.

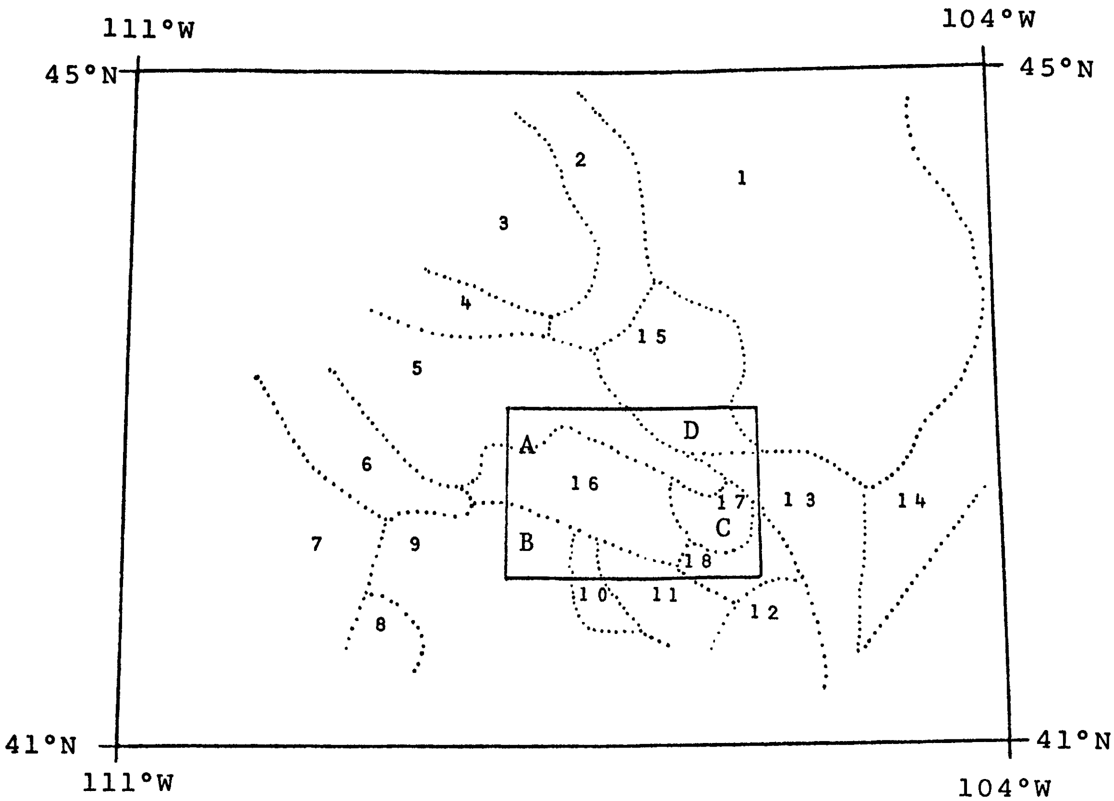

111°W
104°W
45°N
45°N
1
2
3
4
15
5
A
D
6
16
17
13
14
C
7
9
B
18
10
11
12
8
41°N
41°N
111°W
104°W

FIGURE 2-2

was deposited near the shores of the Cretaceous seas. The Larimide orogeny deformed the area in the Late Cretaceous, forming basins and uplifts. After the seas withdrew for a final time, erosion of the newly formed mountains began in Paleocene time. Detritus eroded from the mountains partly filled basins during the Paleocene and early Eocene. The resulting sediments are predominantly sands and silts with occasional arkosic sands and gravels. The Wind River Formation, which is the most important uranium host unit in the Quadrangle, was deposited in the early Eocene. Over the Wind River were deposited arkose and clay followed by tuffaceous sediments that filled the basins. During the Pliocene, broad regional uplift occurred, Mid-Tertiary sediments were removed, and mountain ranges were uncovered exposing Lower Tertiary strata. Quaternary stream gravels and alluvium now cover some areas in the Quadrangle.

Uranium ore in the Casper Quadrangle is found in roll front deposits in Tertiary fluvial, arkosic sandstones and conglomerates containing interbedded clay, silt, and occasional limestone. Deposits are characteristically tongue-shaped accumulations of uranium minerals within a single host unit. The uranium deposits mark the interfaces between altered host rock updip of the deposit and unaltered rock downdip. The roll fronts are generally only a few tens of feet thick, but may extend laterally for up to thousands of feet.

Uranium leached from Precambrian granites or from young, tuffaceous sediments appears to have been taken up by slightly acidic groundwaters during the Miocene. The weathering of pyrite or bacterial decomposition of hydrocarbons may have produced the acidic groundwater conditions. These uranium-bearing waters moved downdip within permeable strata, altering feldspars, removing calcite and numerous trace elements from the host beds. Eventually these waters entered higher Ph environments, and uranium was rapidly precipitated. In contrast to the Colorado Plateau case, immobile carbonaceous reducing material seems to have enhanced only slightly the chemical conditions favoring deposition. Downdip from zones of uranium deposition, there is little alteration of the host beds, so that ores marked the Ph boundary of a geochemical cell within the host formations. As new acidic groundwaters continued to enter the zones of deposition, uranium would reenter solution and again be deposited, a short distance downdip, on the alkaline side of slowly migrating roll fronts. Today, groundwater flow is reduced, and the deposits of uranium have taken up stable positions in the host beds.

As on the Colorado Plateau, structural features that influenced the flow of groundwater and the sedimentary environment should play an important role in recognition of zones favorable for uranium deposition. Subtle chemical and lithologic variations at depth that controlled ore emplacement perhaps cannot be predicted or related to structural features.

CHAPTER 3

This chapter defines the objects for recognition and presents the features used in recognition. Features are extracted from geologic data in a partly subjective, yet algorithmic way. In addition to statistical estimates of the significance of features, geologic interpretations are offered here to explain the features selected for recognition. If features can be reasonably understood in the context of current theories of ore genesis, this will lend confidence both in the present features and in the feature selection procedure when it is used in recognition problems with less familiar characteristics (e.g. less well understood mineral deposits).

Without assuming a model of ore deposition, the feature selection procedure has picked many features for recognition that are recognizable as elements of ore deposition models. Many unfamiliar features are also found, however, and tentative geologic interpretations of these may suggest less familiar influences upon ore deposition.

3.1 Objects for Recognition

The area of the Colorado Plateau is divided into 7-1/2 minute cells for recognition. The resulting grid of 508 cells is shown in Figure 3-1. This division of the plateau yields cells of a size appropriate to the resolution available in the data, and provides a manageable, yet statistically adequate, number of objects for computerized classification.

There are two logical ways to divide these 508 cells into uranium producing, U, and non-producing, U*, classes a priori. U and U* classes may be assigned according to the amount of ore-grade uranium in each cell - total production plus probable reserves. Alternatively, since uranium deposits tend to be found in either Triassic strata or in Jurassic/Cretaceous rocks within a given cell, this division may be used. Feature selection and several pattern recognition experiments were run to determine the relative merits of these two divisions. Although many features are common to the class of all producing cells, cells with Jurassic/Cretaceous production, and cells with Triassic production, the division according to age of host rock produces more interesting and stronger sets of features, and more area-specific recognition results than a division according to produc-

FIGURE 3-1: Grid of 508 15-minute cells for evaluation of uranium favorability on the Colorado Plateau. Each cell is coded according to its uranium production as follows:

1 - uranium production less than 1000 tons (ore $\geq$0.1% U_3O_8);
2 - uranium production from 1000 to 1,000,000 tons (ore $\geq$0.1% U_3O_8);
3 - uranium production greater than 1,000,000 tons (ore $\geq$0.1% U_3O_8);
4 - no uranium production to date;
C - uranium produced from Jurassic and/or Cretaceous host beds;
R - uranium produced from Triassic or older host beds;
T - uranium produced from Tertiary host beds;
V - uranium produced from vein or breccia pipe deposits.

COLORADO PLATEAU - URANIUM DEPOSITS

111.0°W 40.5°N — 109.0°W 40.5°N

								4	4	4	4	4	1C	4	1C	1C	4							
								4	4	4	1T	1T	4	4	4	1C	4							
						4	4	4	4	4	4	4	1T	4	4	4	4	4	4					
						4	4	4	4	4	4	4	4	4	4	4	4	4	4					
				4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4			
				4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4			
				4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4	
				4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4	
	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4			
	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4			
	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4			
	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V			
R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4	
	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	
R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4
V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4
	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4
	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4
	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C	
	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4	
		4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4	
		4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4	
						4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4			
						4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4			

111.5°W 34.5°N — 107.5°W 34.5°N

FIGURE 3-1

tion alone. The Colorado Plateau is studied here in two ways. One set of recognition experiments used features obtained from contrasting 45 cells with production from Triassic strata against barren cells. A second set of recognition experiments uses features generated from contrasting cells with production from Jurassic/ Cretaceous host rocks against barren cells.

For the Casper Quadrangle of Wyoming, recognition is based on data of somewhat finer resolution than for the Colorado Plateau. Accordingly, the Quadrangle is divided into smaller objects for recognition. An 18 x 26 grid of points on a 4 x 4 mile spacing covers the Quadrangle. Some data are incomplete in the southern three rows of points and in 10 points in the northeast corner of the Quadrangle; these areas were not considered in this study. Neighborhoods around the 380 remaining points are the objects for recognition. Uranium producing areas are shown on the grid of objects for recognition in Figure 3-2. Of 380 points, 21 are associated with known uranium production, 30 are uranium prospects, 26 are reported uranium occurrences, and 303 are presumed barren. The designations of "prospect" and "occurrence" are very uncertain indicators of uranium deposits; prospects are generally so designated on the basis of

FIGURE 3-2: Grid of 390 points spaced at 4 x 4 miles for evaluation of uranium favorability in the Casper, Wyoming Quadrangle. Points are coded according to their uranium production as follows:

1 - uranium mined within 3 miles of the point;

2 - uranium prospect within 3 miles of the point;

3 - uranium occurrence within 3 miles of the point;

4 - presumed barren;

5 - not used in this study.

CASPER, WYOMING QUADRANGLE - URANIUM DEPOSITS

108°W 43°N

106°W 43°N

4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	2	1	1	2	4	4	4	4	4	3	2	4	4	4	4	4	4	4	4	5	5
4	4	2	4	1	1	1	1	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	5
2	4	4	2	1	1	2	4	4	4	4	4	4	4	3	4	4	2	3	4	4	4	4	4	4	5
4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4
4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	4	4	4	4	4	4	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	1	4	4	3	4	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	2	1	4	4	3	4	4	4	4	4	4	4	4	2	4	3	4	4	4	4	4	4	1	4	1
4	3	1	1	3	2	2	4	4	4	4	4	4	4	2	4	4	4	4	3	3	4	4	1	1	4
3	4	2	4	2	2	4	4	4	4	4	4	4	4	1	3	3	4	2	4	4	4	4	4	1	4
4	3	4	3	2	2	4	4	4	4	4	4	4	4	4	4	3	2	3	4	4	4	4	4	4	4
4	4	3	4	3	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	4	4	3	4	4	4

108°W 42°N

FIGURE 3-2

106°W 42°N

somewhat stronger field evidence of ore potential than are occurrences. In feature selection, features that would characterize not only mining areas but also prospects and occurrences were sought. Statistically, the prospect and occurrence classes closely resemble barren areas, and so features used for recognition are those that distinguish the 21 producing points from the non-producing points.

3.2 Feature Selection

A large body of data was processed for both study areas to find features relevant to discrimination of U from U* objects.

One-dimensional projections of the disposition of objects within this feature space were used to select features for use in recognition from the large bodies of raw data. Although some U* objects are undoubtedly yet unrecognized U objects, the samples of U and U* objects as they are now known were used to define U and U* classes. The range of values for each feature was divided into 10 equal regions, and a 10-cell histogram of the feature values for U and U* objects was formed for each feature. These histograms were used as estimates of feature probability density functions (PDF's) conditioned on the state of nature, U or U*. If the histogram estimates of PDF's for U and U* objects on a feature indicate, according to the tests described below, that the U and U* objects are two distinct populations, the feature was used in recognition. Features that showed essentially identical PDF's for U and U* objects, and features that could separate U and U* objects only if the range of feature values were partitioned into a number of disconnected subregions, were not used.

Figure 3-3 illustrates typical types of state-conditional PDF estimates. A feature such as F_a, Figure 3-3a, with essentially indistinguishable U and U* PDF's is not used in recognition. It is possible that such a feature might actually be useful in recognition, but it appears uninformative when the locations of objects in the multidimensional feature space are projected onto the F_a axis. The features F_b and F_c are typical of those used in recognition. For example, in Figure 3-4, features 3, 7, and 8 have one-dimensional distributions similar to F_b; features 4, 10, and 11 have distributions similar to F_c. A tendency toward separation of the U and U* classes is apparent along the F_b and F_c axes of the feature space. The F_b and F_c axes may be broken into two and three regions, respectively, that are populated mostly by U or mostly by U* objects. The

<u>FIGURE 3-3</u>: Typical types of state-conditional probability density functions for recognition features.

A.- Producing and barren objects indistinguishable on feature F_a;

B - U objects are often characterized by high values of F_b, U* objects by low values of F_b;

C - U objects are often characterized by the highest or lowest values of F_c, U* objects tend to have intermediate values of F_c;

D - Multimodal, interleaved density functions for U and U* objects on feature F_d.

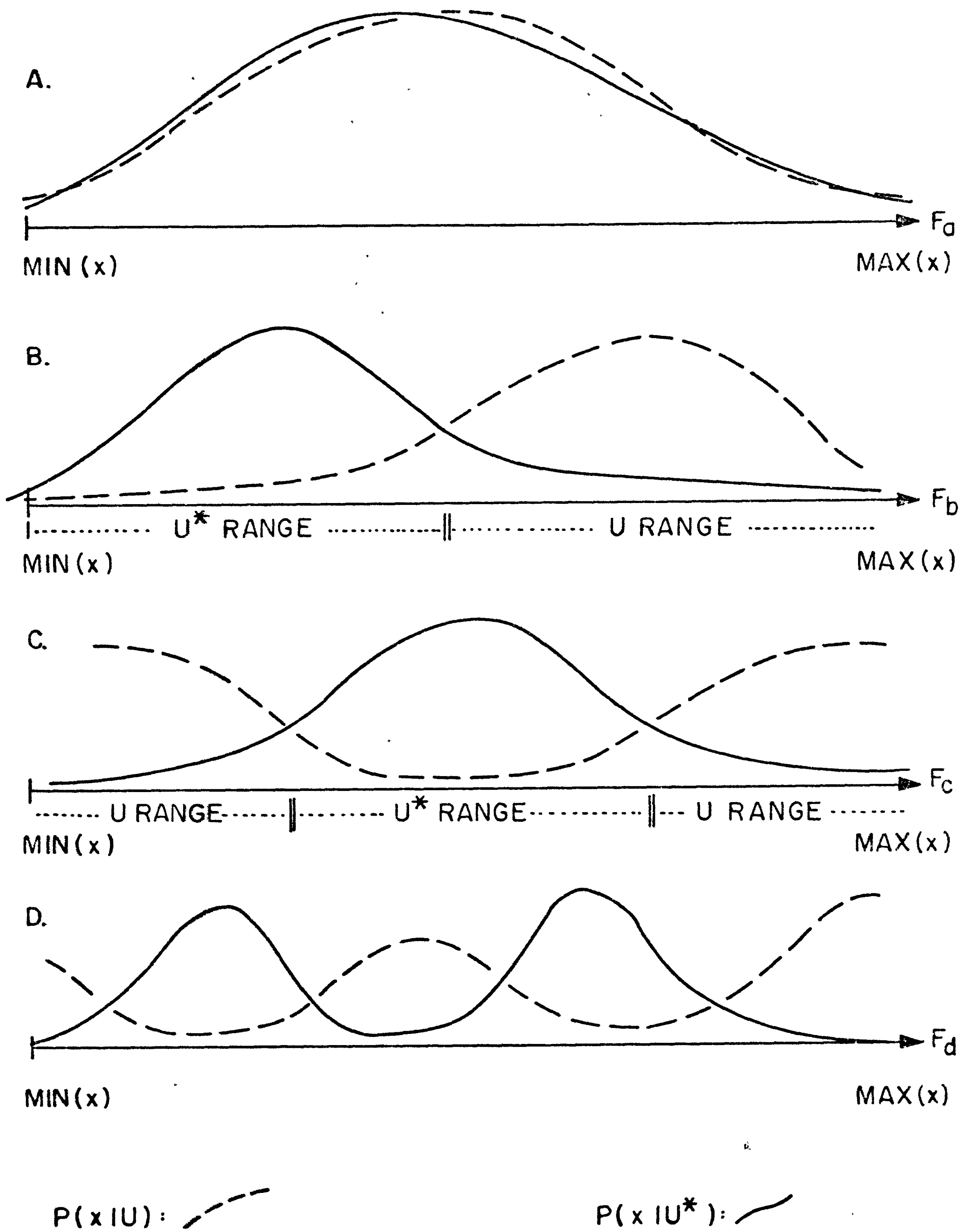

A.
Fa
MIN (x)
MAX(x)
B.
Fb
U* RANGE
U RANGE
MIN (x)
MAX(x)
C.
Fc
U RANGE
U* RANGE
U RANGE
MIN(x)
MAX(x)
D.
Fd
MIN(x)
MAX(x)
P(x|U):
P(x|U*):

FIGURE 3-3

boundaries of U and U* regions are estimated to be at feature values where the U and U* PDF's cross. If the histograms provide accurate estimates of feature PDF's, these boundaries are optimal in the sense that they minimize the probability of classification error with a feature, if objects with feature values in the U range are classified as U and objects with feature values in the U* range are classified as U*. In the features F_b and F_c, large and small, or intermediate and extreme, values of a parameter are contrasted. Such features are useful in distinguishing U and U* objects, and may often have a simple geologic interpretation. By contrast, feature F_d of Figure 3-3d exhibits more complex, multi-modal, and interleaved density functions. The F_d axis may be partitioned into a number of subregions, each populated almost exclusively by U or U* objects. Such a division of the F_d axis could serve to separate the U and U* objects, but no plausible geological interpretation of such a division is usually apparent, and such features are not used in this recognition procedure.

3.3 Feature Coding

The four algorithms used in this study make use of two extreme, and in a sense, opposite methods of feature coding. In one case, the exact values of data for each feature are used to give the most precise description possible of each object in the feature space. Objects are described by multidimensional data vectors with numerical components. Geological objects, such as ore bodies and areas favorable for ore deposition, while in many ways similar, will show some variability not clearly related to their ore potential. To emphasize the essential characteristics of geological objects, and to remove some differences in detail, the second data coding scheme records for an object not the exact values of its features, but only whether the values of the object's features fall in a characteristic U or U* range of that feature (where the U or U* PDF is greater than the other, respectively). Objects are thus described by multidimensional data vectors with binary components. Although this scheme produces the greatest possible quantization noise, it eliminates differences between objects within the U and U* ranges of a feature parameter. The major benefit of binary coding is that both qualitative and quantitative data that are continuous, quantized, coded, and intrinsically binary can be reduced to a uniform data format and united for use in recognition. Coded, quantized, and qualitative descriptive geological data are not readily integrated into the vector space classification framework typical of recognition algorithms that use exact values of features. It is an open question to what extent a geologist uses the exact values of all features when producing a subjective combined interpretation, and to what extent he notes the range of values into which an object's features fall. It is certain, however, that descriptive, semiquantitative, and non-numerical data may carry important informaation about geological objects, and are normally included in geological evaluations.

3.4 Feature Ranking

The feature spaces for the Colorado Plateau and Casper study areas are of high dimensionality, 17, 32, and 36, for Triassic, Jurassic/Cretaceous, and Wyoming deposits, respectively. This high dimensionality hampers analysis because a 30-dimensional space is much more difficult to explore than, say, a 10-dimensional space; one is reluctant, however, to discard information that can contribute to geological evaluations. This high dimensionality is also a problem when few samples are available. For example, with 36 features but only 21 samples of U objects for the Casper Quadrangle, it is difficult to obtain accurate estimates of PDF's or to determine the extent to which features are correlated or redundant (Chapter 6 investigates this problem). One would like a method of measuring the significance or usefulness of the features; such a ranking, in combination with geological criteria, could suggest subsets of the features that might be particularly useful in recognition.

Because the geological data used here are of various natures (numerical, coded, or intrinsically binary), three non-parametric techniques were used to measure the significance of features and their ability to separate the known populations of U and U* objects.

The first measure indicates the ability of binary-coded versions of features to separate U and U* objects. If $p_i = \text{Prob}(f_i = 1|U)$ and $q_i = \text{Prob}(f_i = 1|U^*)$, then a measure of a binary-valued feature's ability to separate U and U* objects along that feature axis is

$$w = \log \frac{p_i(1 - q_i)}{q_i(1 - p_i)}$$

(this factor is used to weight features in the simplest of the four classification algorithms used here; see Chapter 4, section 4.1). This ranking measures a feature's ability to separate U and U* objects along one feature axis. It does not, however, measure a feature's ability to separate objects when combined with other features in a multidimensional decision process.

A rank sum test was also applied to exact values of the features. U and U* objects are arranged and ranked in order of increasing magnitude of their values for a feature. The sum of these ranks is a normally distributed random variable with expected value

$$E_1 = n_1(n_1 + n_2 + 1)/2 \quad \text{or} \quad E_2 = n_2(n_1 + n_2 + 1)/2$$

and variance

$$\sigma^2 = n_1 n_2(n_1 + n_2 + 1)/12$$

where n_1 and n_2 are the numbers of U and U* objects. The output of this test is a confidence level at which one can presume that the U and U* objects do form different populations with respect to the feature. One difficulty with this test, however, is that it is not applicable to intrinsically binary or coded features. Also, a feature such as that of Figure 3-3c, which is clearly a good discriminant, may have a low confidence rank in this test. When using this confidence rank, therefore, the shape of the feature PDF must be considered.

Finally, each feature's information content, estimated from the 10-place one-dimensional histograms, was computed. The information content of a feature, F, is

$$I(F) = \sum_{d=1}^{10} P(d) \sum_{c=1}^{2} P(d|c)\log P(d|c) - \sum_{c=1}^{2} P(c)\log P(c)$$

where d represents one of ten deciles in the histograms, and c is one of two classes, U or U*. When natural logrithms are used, I(F) has a maximum possible value of 0.693. This test indicates a feature's ability to separate U from U* objects in classification outcome space rather than feature space (Boyle, 1976; Gallager, 1968). This rank is applicable to binary, coded, and continuous features, but a PDF, such as that of Figure 3-3d, may have a high information rank even though its estimated PDF may make no geologic sense. Though this measure of feature strength proved to be the most reliable of the three tests used here, the shape of estimated PDF's should be considered when using this rank.

The three ranks of features for the Colorado Plateau Jurassic/Cretaceous, Colorado Plateau Triassic, and Casper Quadrangle uranium deposits are given in Tables 3-1, 3-2, and 3-3. The rank ordering of the features often varies considerably between the three measures, though two of the three ranks are usually nearly the same for a feature. In geological problems such as this, there appears to be no single "best" way of measuring feature strength, though the information measure is the most versatile and seems the most useful for classification problems. In general, several complimentary tests, such as these three, should be used to insure the statistical significance of feature PDF's and to determine an ordering of feature strength, should one wish to discard some features from recognition.

3.5 Features for Recognition

The features used in recognition for the Casper Quadrangle and for the Colorado Plateau are listed below. A brief interpretation of each feature is also given to suggest a reasonable geologic connection between the features and uranium deposition.

The geological data used in classification are taken entirely from public sources. There are many features relevant to uranium favorability that are not used here. Data such as geochemical soil and water analyses, well log data, evidence of alteration, etc., are available only over small subregions of the study areas and/or are proprietary to private concerns. The features used here, then, do not form an ideal feature set, and do not contain all information relevant to uranium favorability. Several important influences on uranium deposition are contained within the features, however, and useful syntheses of this data and reasonable predictions have emerged from pattern recognition treatment.

Data for the Colorado Plateau were taken from the Geologic Atlas of the Rocky Mountain Region (R.M.A.G., 1972), from new U.S.G.S. isopach maps of Cretaceous sediments on the Colorado Plateau, and from standard U.S.G.S. topographic maps. Data for the Casper, Wyoming, Quadrangle were assembled for analysis by E.R.D.A., Grand Junction, from subcontracted work. A radiometric survey flown at 2-mile spacing, surface geological, structural, topographic, geothermal gradient, gravity, and magnetic maps, and lineament analysis of LANDSAT imagery were available to generate features for recognition.

The features used in recognition were selected from a larger set of candidate features derived from these data sources using a pre-determined algorithm independent of the problem at hand. It is noteworthy, therefore, that so many features are explicable in terms of modern theories of ore genesis, particularly in the case of the Colorado Plateau. Also note that although the present feature selection procedure has found features for the Colorado Plateau that refer to geological events from the Pennsylvanian to the present, an emphasis on Late

Cretaceous time, the time of ore deposition, has naturally emerged in the features.

Several prospecting guides to sandstone-type uranium deposits have been recognized by geologists (Fisher, 1974; McKay, 1955; Grutt, 1972). These include the age, thickness, and depositional environment of the host rock, interbedding of thin lenses of shale, mudstone, and conglomerate with the host sandstones, bleaching or other discoloration of host rocks, occurrence of unconformities and small scale faults, trace element anomalies in groundwaters or host rocks, occurrence of fossil carbonaceous material, and mineralogical characteristics of the host sandstones. Some of these diagnostics cannot be determined without extensive field work; others will be of little use for detecting deposits at depth without drilling. Wherever data permit, a reflection of these well-known guides occurs in the features for recognition, e.g. reference to bed dip angle, proximity to tuffaceous beds, proximity to small faults, LANDSAT lineaments, radiometric anomalies, occurrence of carbonates, depositional environment of the host sediments, and thickness of host beds.

Because some well-known guides for prospecting emerge naturally from pattern recognition techniques as features, it is reasonable to suppose that the unfamiliar features may also be of geological significance and may be useful as prospecting guides. Further, when deeper deposits are sought, and objective feature selection process may be an effective way of generating prospecting guides from geophysical and other survey data.

The features for recognition follow; histogram estimates of their state conditional PDF's are given in Figures 3-4, 3-5, and 3-6. Most features are continous-valued, many referring to the distance from a recognition object to the nearest geological entity of a given type (e.g. features #1 and #3 from Figure 3-4); other features are coded (e.g. feature #2 from Figure 3-4); other features are intrinsically binary (e.g. feature #6 from Figure 3-4). State-conditional PDF's for all features are given in 10-place histogram formats.

3.5.1 Features for Recognition of Uranium Deposits in Cretaceous Sediments on the Colorado Plateau

1. Proximity to Cretaceous shoreline of Eagle time. (Areas near the paleocoastline are favored for uranium. The series of Cretaceous marine transgressions were major geological events occurring nearly contemporaneously with ore deposition. Near the interface between marine and continental environments, sands, shales, silt, and limestone may be interbedded. Oxidation/reduction reactions might proceed quite differently from a clean sand environment in these interbedded sediments of varying porosities, permeabilities, surface areas per unit volume, and chemical characteristics. The presence of an ancient ocean may have influenced paleohydrology on both sides of a paleocoastine. Other features suggest that relative elevation was an influence on uranium deposition; a paleocoastline marks a reference elevation that may be reconstructed regardless of subsequent elevation changes within a province.)

2. Lithofacies of the Entrada and Carmel Formations. (Areas of marine, rather than continental, sandstone

are favored for uranium. These formations are not major uranium producers, but they may provide barriers to the vertical movement and escape of groundwaters circulating through adjacent beds.)

3. <u>Proximity to local maxima in thickness of the Brushy Basin Shale Member of the Morrison Formation</u>.
(Areas near thickenings of the Brushy Basin are favored for uranium. Locally thickened areas mark depositional foci; after deposition of the Brushy Basin, groundwaters may have continued to be focused toward these areas. As shale is relatively impermeable, flow through the enclosing beds may have been affected by variations in the thickness of the Brushy Basin.)

4. <u>Proximity to Cretaceous shoreline of Late Skull Creek time</u>.
(Proximity to the paleoshoreline is favored for uranium. See feature 1.)

5. <u>Thickness of the Pennsylvanian System</u>.
(Thicker areas are favored for uranium. Depositional foci of the Pennsylvanian may mark long-lived structural features that persisted throughout the evolution of the Plateau.)

6. <u>Presence of a Miocene/Eocene intrusive body within the cell</u>.
(Cells with intrusions are favored for uranium. Mineralizing solutions may have been derived from intrusive magmas.)

7. <u>Proximity to local maximum of thickness of the entire Morrison Formation</u>.
(Proximity to locally thickened areas is favorable for uranium; this is a familiar prospecting guide. Thickened areas are depositional foci, and perhaps groundwater flow foci. Thickened areas may have more numerous or thick interbedded shales, mudstones, and sandstones that are known to be favorable lithologic influences on ore deposition.)

8. <u>Proximity to Cenozoic volcanic and intrusive rocks</u>.
(Cenozoic igneous bodies post-date ore emplacement. As they occur in areas quite different from Cretaceous intrusives, it is possible that new tectonic influences activated zones of crustal weakness and a system of magma conduits distinct from those associated with ores. Possibly pre-existing uranium deposits may have been exposed to oxygenated waters where these volcanics disrupted Plateau sediments. These

volcanics ring the Plateau in the south and east. Proximity to them is unfavorable for uranium. It may be, therefore, that this feature merely reflects the fact that a majority of known ore bodies in Cretaceous sediments are clustered in the center of the Plateau.)

9. <u>Minimum thickness of the Morrison Formation in the cell</u>.
(Thinnest areas are unfavorable for ore. This feature reflects both the fact that the Morrison is the major host unit for uranium (zero thickness is unfavorable for ore), and that thicker areas within the Morrison are particularly favorable for ore.)

10. <u>Minimum thickness of the Brushy Basin Member of the Morrison in the cell</u>.
(Thicker areas are favorable for uranium. See features 3 and 9.)

11. <u>Proximity to a pinch-out of the Entrada/Carmel Formations</u>.
(Areas where these formations are present, at moderate distances from the pinch-outs, are favored for ore. Paleo-topographic highs are ringed by pinch-outs.)

12. Thickness of Swift Age sediments (Summerville and Todilto Formations).

    (Areas where these sediments are present, but relatively thin, are favored for ore.)

13. Proximity to Cretaceous shoreline of Late Moury time.

    (Proximity is favored for uranium. See feature 1.)

14. Proximity to a pinch-out of the Westwater Member of the Morrison Formation.

    (Proximity to a pinch-out is favored for ore.)

15. Maximum thickness of the Salt Wash Member of the Morrison Formation in the cell.

    (Thicker areas are favored for uranium. See feature 7.)

16. Proximity to a local maximum in thickness of the Salt Wash Member of the Morrison Formation.

    (Proximity to maxima of thickness are favored for uranium. See features 3 and 9.)

17. Proximity to Cretaceous shoreline of Judith River time.

    (Proximity to the paleoshore is favored for uranium. See feature 1.)

18. Maximum thickness of Brushy Basin Member of the Morrison Formation in the cell.

    (Areas of greatest thickness are favored for uranium. See features 3 and 9.)

19. Thickness of Triassic rocks in the cell.

    (Areas where these rocks are present, but thin, are favorable for uranium.)

20. Proximity to a pinch-out of lowest Cretaceous sediments.

    (Proximity to a pinch-out is favored for uranium.)

21. Proximity to the Cretaceous shoreline of Early Clagett time.

    (Proximity to the paleoshoreline is favored for uranium. See feature 1.)

22. Proximity to Upper Cretaceous intrusive rocks.

    (Proximity to intrusive bodies is favorable for uranium. Intrusives may have been a source of mineralizing solutions.)

23. Minimum thickness of the Salt Wash Member of the Morrison Formation in the cell.

    (Areas of greater thickness are favored for uranium. See feature 7.)

24. Proximity to a pinch-out of the Triassic Moenkopi Formation.

(Proximity to the pinch-outs are favored for uranium. This formation is older than either the Chinle or Morrison; enduring paleotopographic highs are marked by the pinch-outs.)

25. Proximity to a pinch-out of Swift Age rocks.

(Rather large distances from the pinch-outs are favored for uranium. These Jurassic sediments mark the paleotopographic highs present just before the Cretaceous.)

26. Proximity to a pinch-out of the Recapture Member of the Morrison Formation.

(Proximity to the pinch-outs is favored for uranium.)

27. Proximity to the Cretaceous shoreline of Early Belle Fourche time.

(Proximity to the shoreline is favored for uranium. See feature 1.)

28. Proximity to the Cretaceous shoreline of Middle Green Horn time.

(Proximity to this shoreline is unfavorable for uranium, in contrast to other shoreline features listed above. No interpretation of this feature is obvious.)

29. <u>Proximity to major anticlines</u>.

(Major anticlines and uplifts may have offered structural controls on circulation of groundwaters. Uplifts may have exposed Precambrian basement rocks, possibly providing uranium source material. Vertical movements of fluids may have been facilitated by faults near the axes of anticlines. Proximity to these structures is favored for uranium.)

30. <u>Maximum thickness of the Westwater Member of the Morrison Formation in the cell</u>.

(Areas of greater thickness are favored for uranium. See features 3 and 9.)

31. <u>Proximity to a local maximum of thickness of the Recapture Member of the Morrison Formation</u>.

(Proximity to thickened areas is favorable for uranium. See features 3 and 9.)

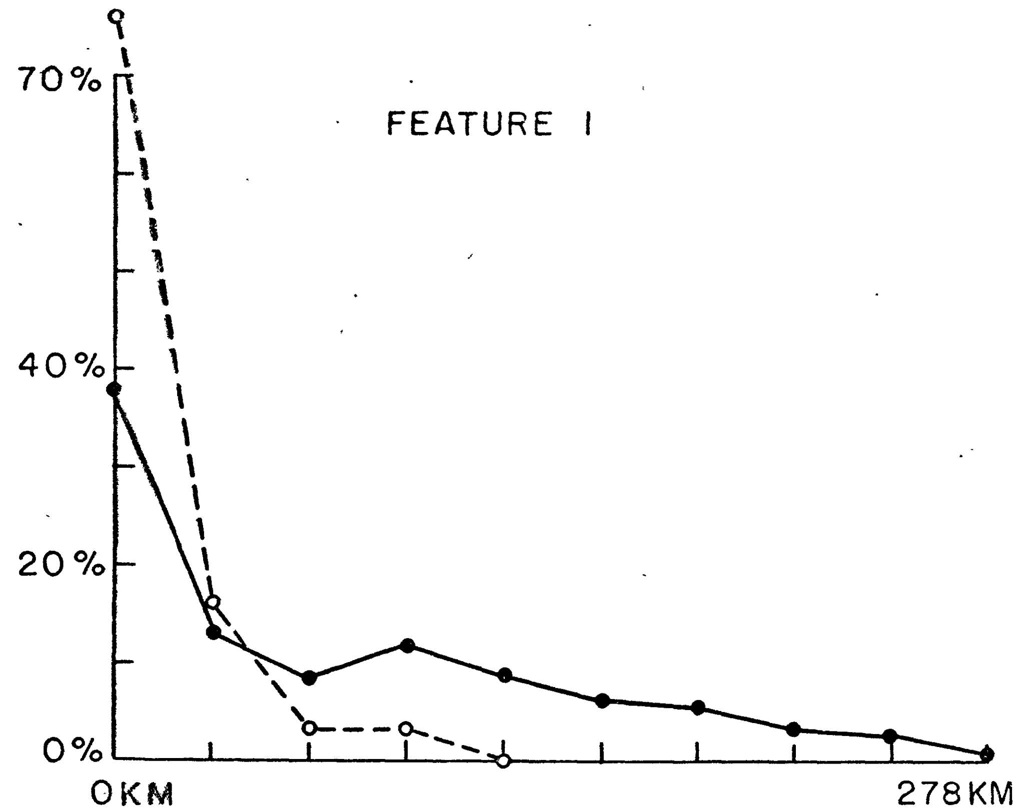
32. <u>Proximity to a local maximum of thickness of the Westwater Member of the Morrison Formation</u>.

(Proximity to thickened areas is favorable for uranium. See features 3 and 9.)

TABLE 3-1: Ranks of Features for Recognition of Uranium Deposits in Jurassic Sediments on the Colorado Plateau.

Feature	Bayes Weight	Rank Sum Confidence	Information	Information Rank
1	2.6	99%	.039	32
2	2.2	NA	.159	2
3	2.2	99%	.104	12
4	1.9	99%	.046	30
5	1.8	99%	.150	3
6	1.8	NA	.239	1
7	1.8	99%	.083	20
8	1.7	99%	.080	21
9	1.7	99%	.075	26
10	1.7	99%	.137	9
11	1.6	99%	.079	24
12	1.6	96%	.077	25
13	1.6	99%	.072	27
14	1.6	99%	.092	17
15	1.6	99%	.147	5
16	1.6	99%	.121	10
17	1.5	99%	.046	31
18	1.5	99%	.101	14
19	1.4	96%	.144	7
20	1.4	99%	.146	6
21	1.4	99%	.060	28
22	1.4	99%	.059	29
23	1.4	96%	.139	8
24	1.3	99%	.101	15
25	1.3	99%	.080	22
26	1.3	96%	.097	16
27	1.1	99%	.088	18
28	1.0	99%	.149	4
29	0.9	99%	.111	11
30	0.9	NA	.080	23
31	0.8	92%	.085	19
32	0.6	95%	.103	13

FIGURE 3-4: Histogram estimates of state conditional probability density functions for features of uranium deposits in Colorado Plateau Jurassic/Cretaceous sediments. These 10-place histogram estimates are based on 58 U objects, 450 U* objects. The vertical axis measures percentages of U and U* populations; U PDF's are indicated by dashed lines, U* PDF's are indicated by solid lines.


FEATURE 1
70%
40%
20%
0%
0KM
278KM

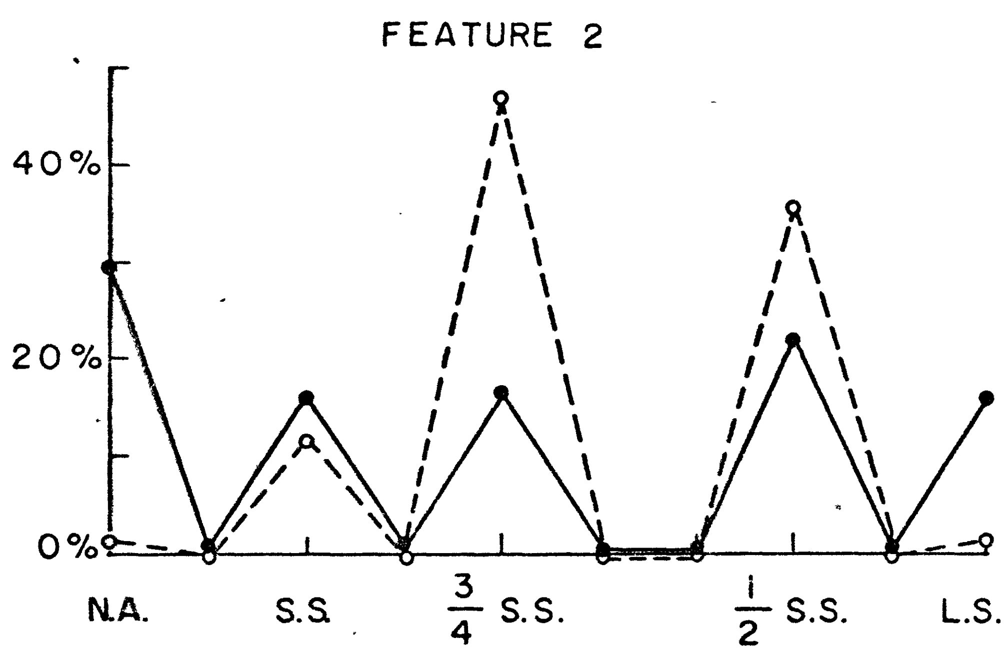
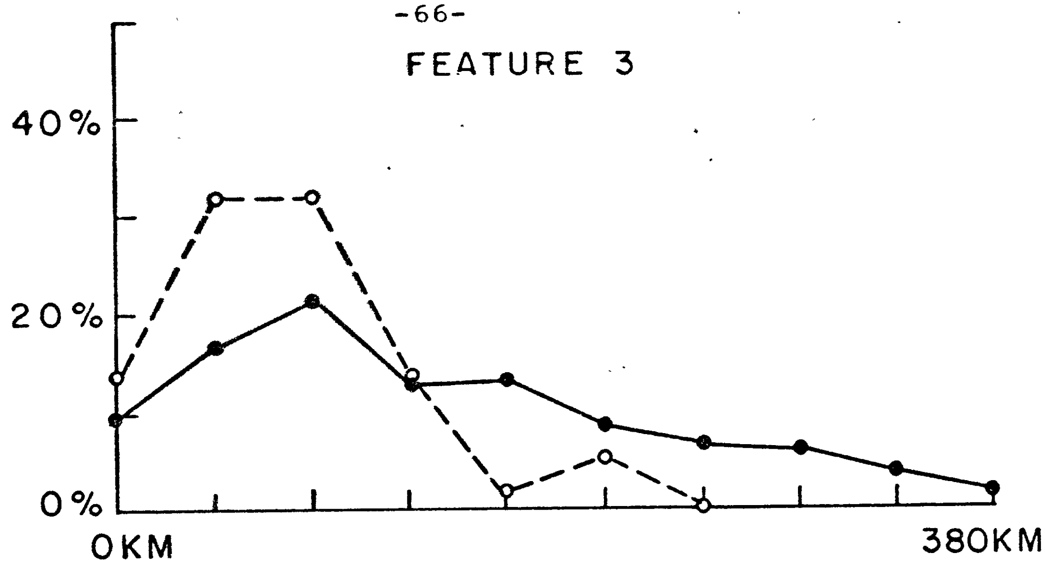
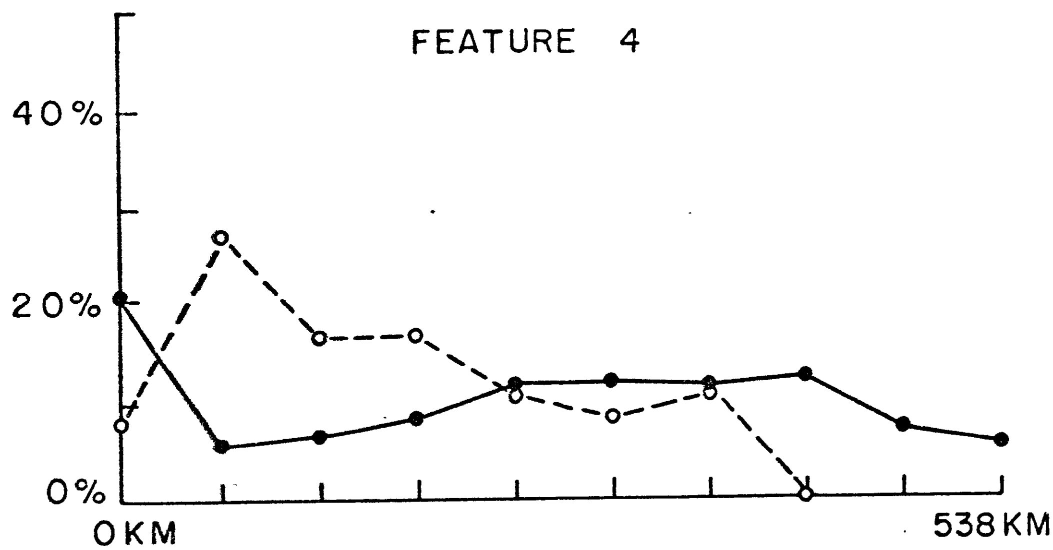
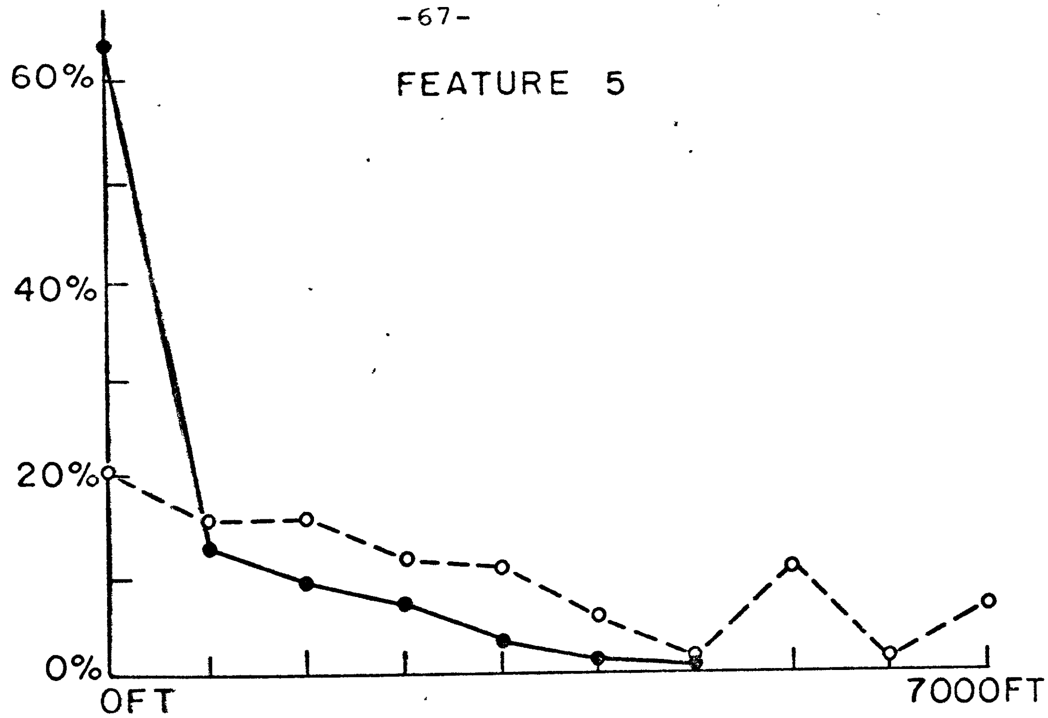
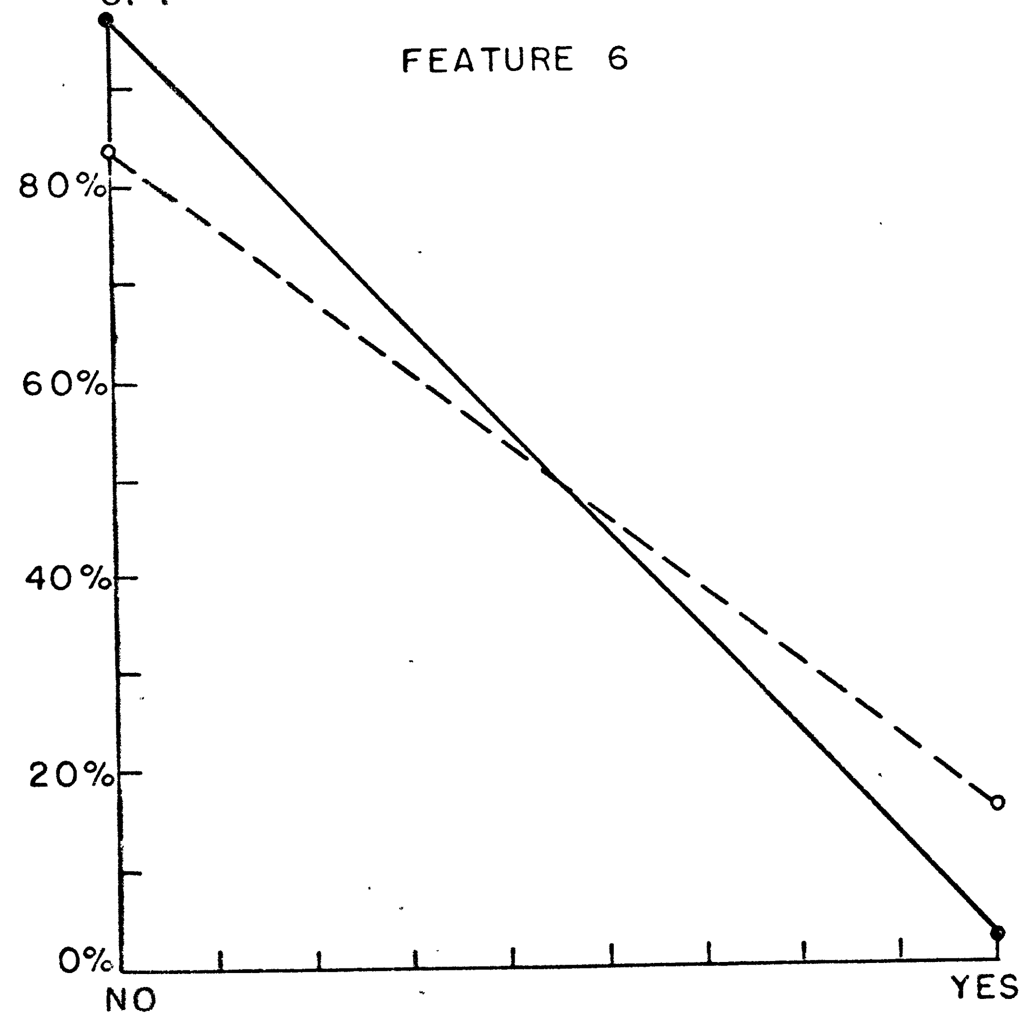

FEATURE 2
40%
20%
0%
N.A.
S.S.
3/4 S.S.
1/2 S.S.
L.S.

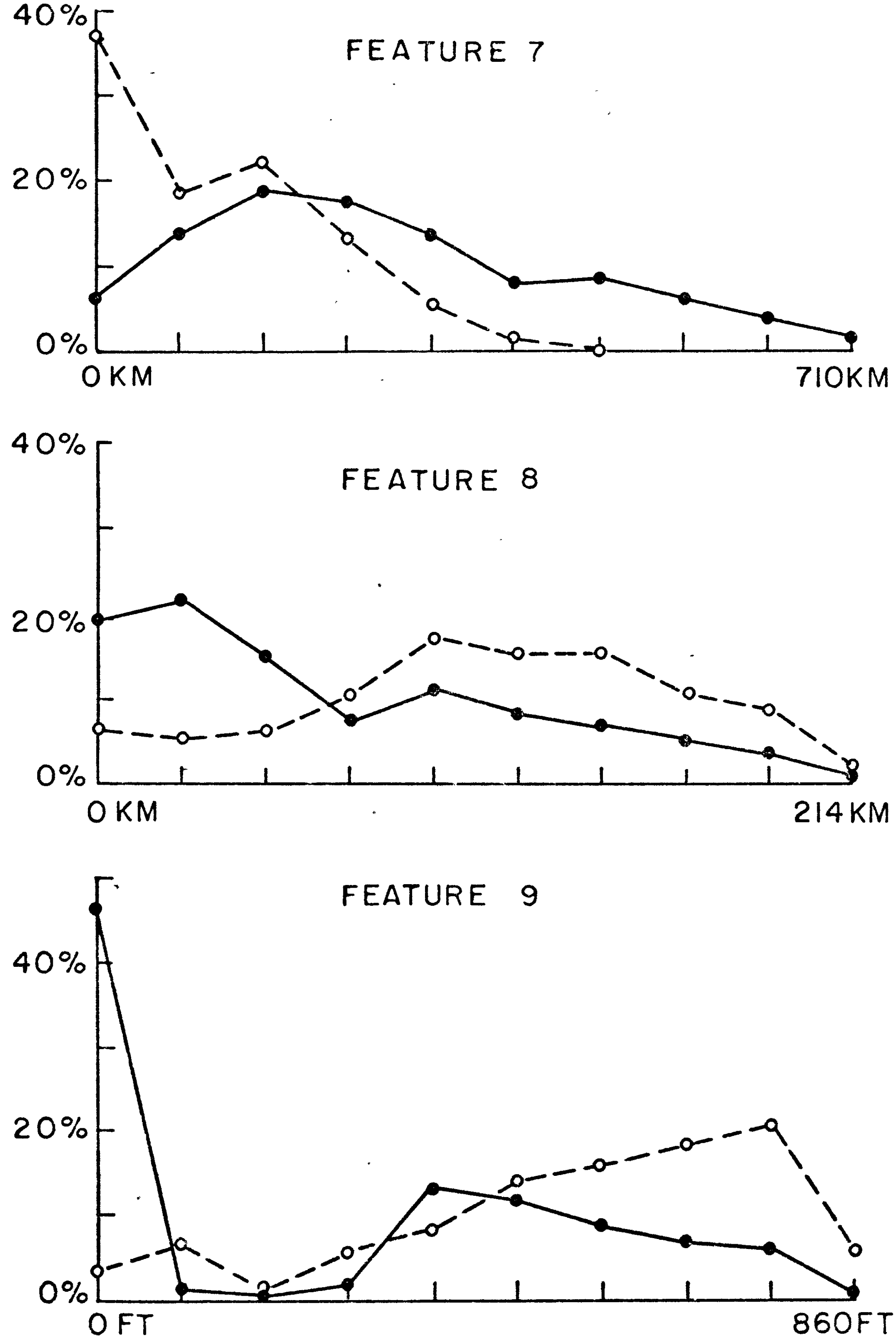
FIGURE 3-4

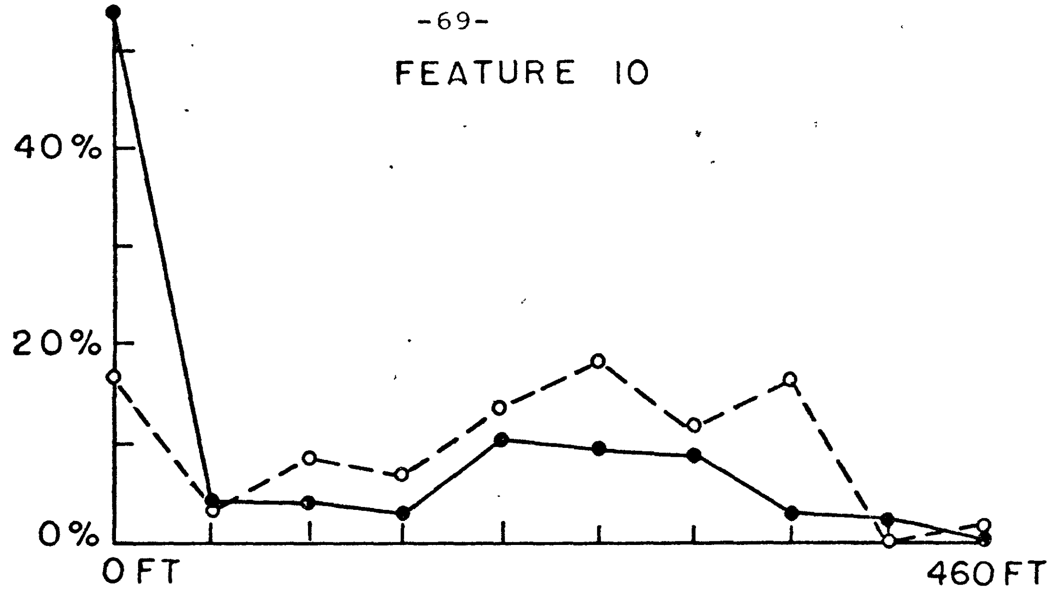

FEATURE 3
40%
20%
0%
0KM
380KM

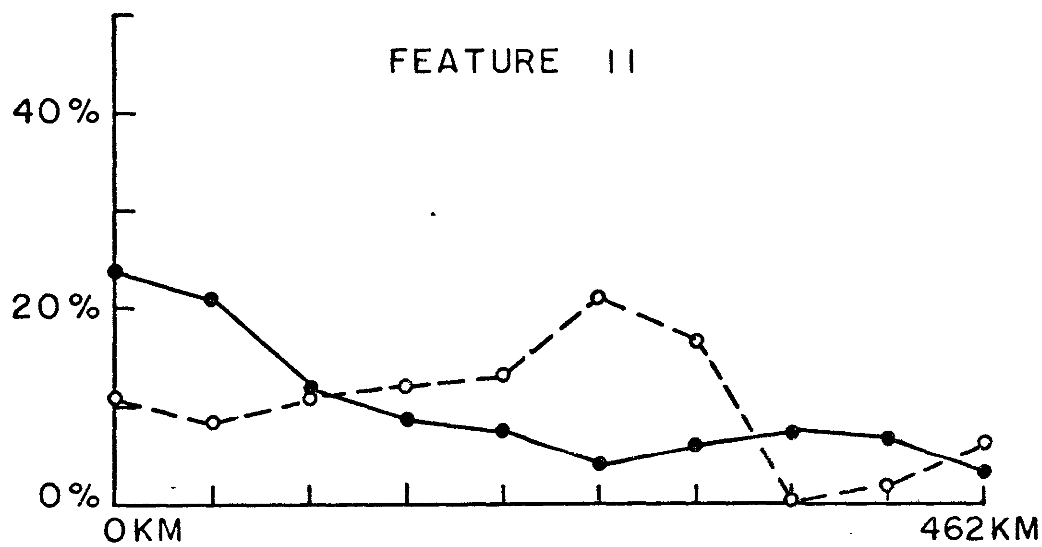

FEATURE 4
40%
20%
0%
0KM
538KM

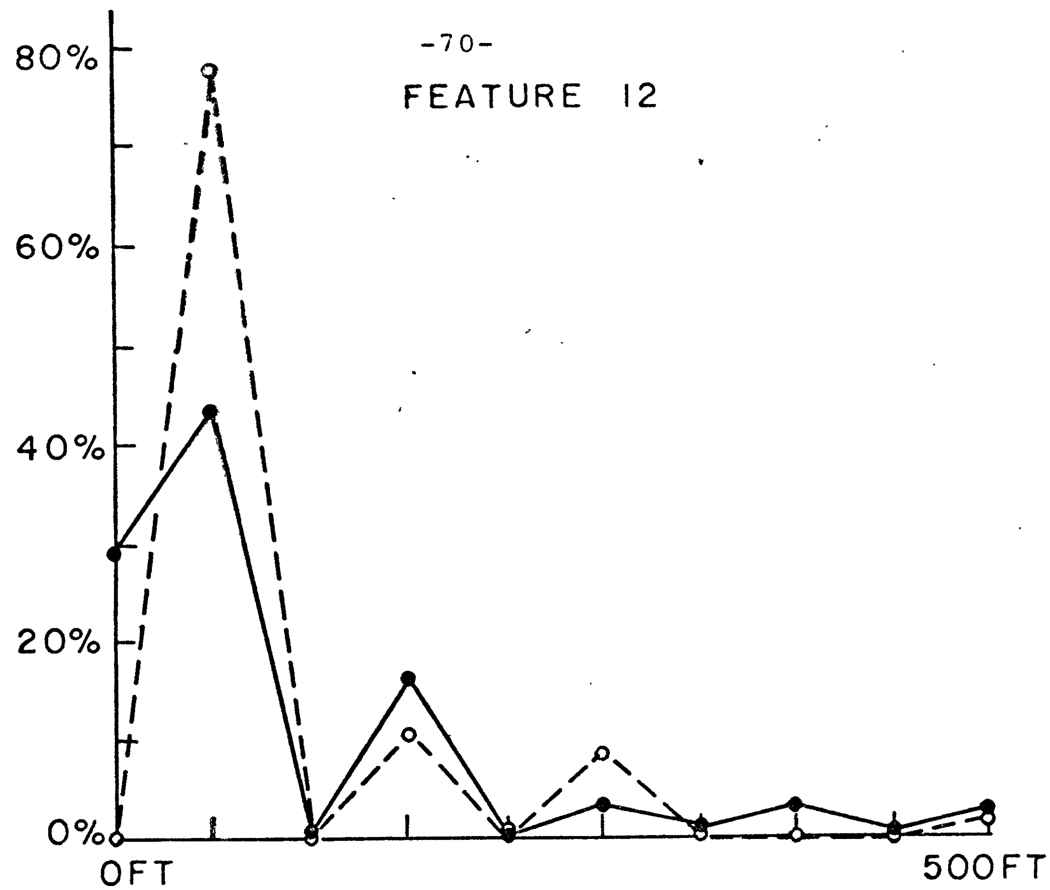

FEATURE 5
60%
40%
20%
0%
OFT
7000FT

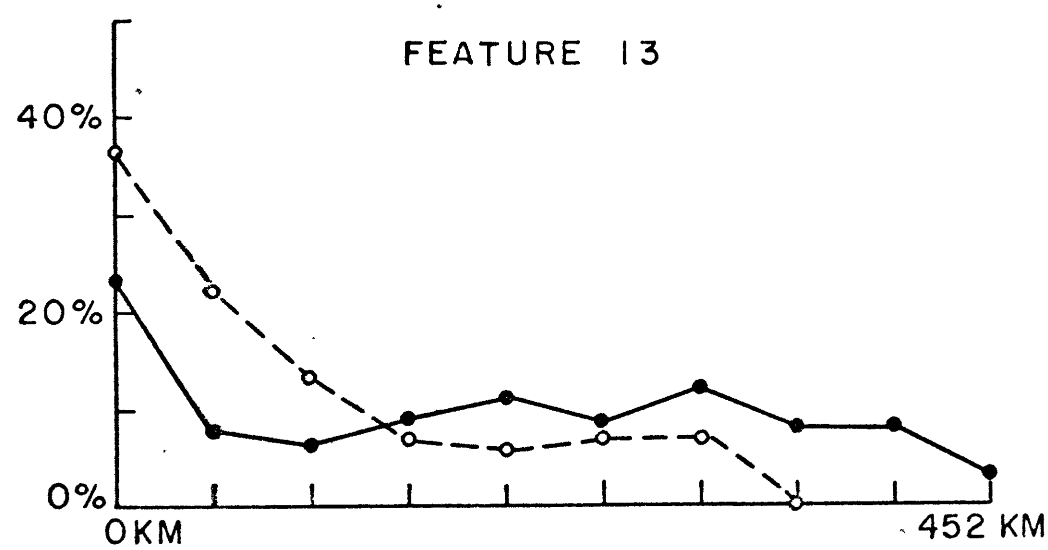

FEATURE 6
80%
60%
40%
20%
0%
NO
YES

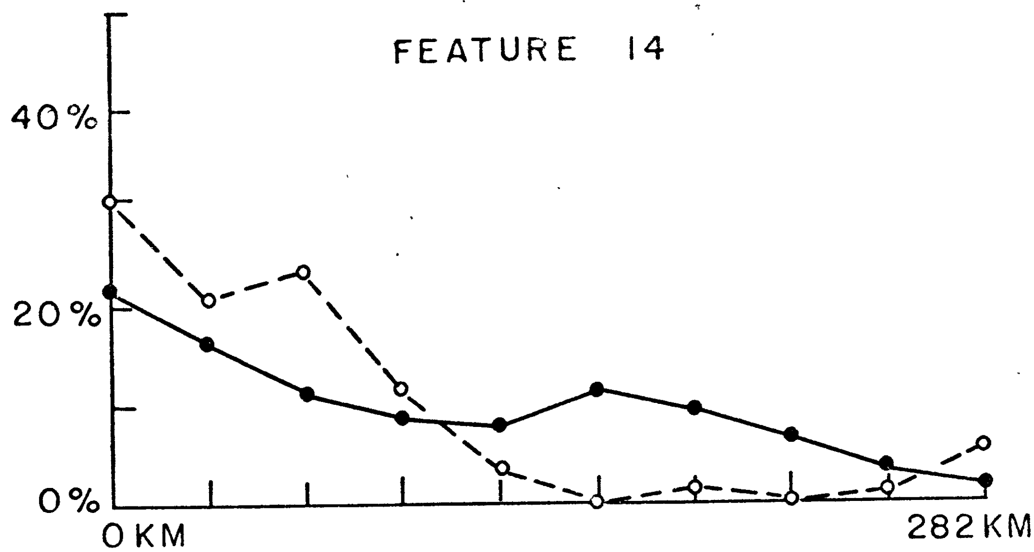

FEATURE 7
40%
20%
0%
0 KM
710KM
FEATURE 8
40%
20%
0%
0 KM
214 KM
FEATURE 9
40%
20%
0%
0 FT
860FT

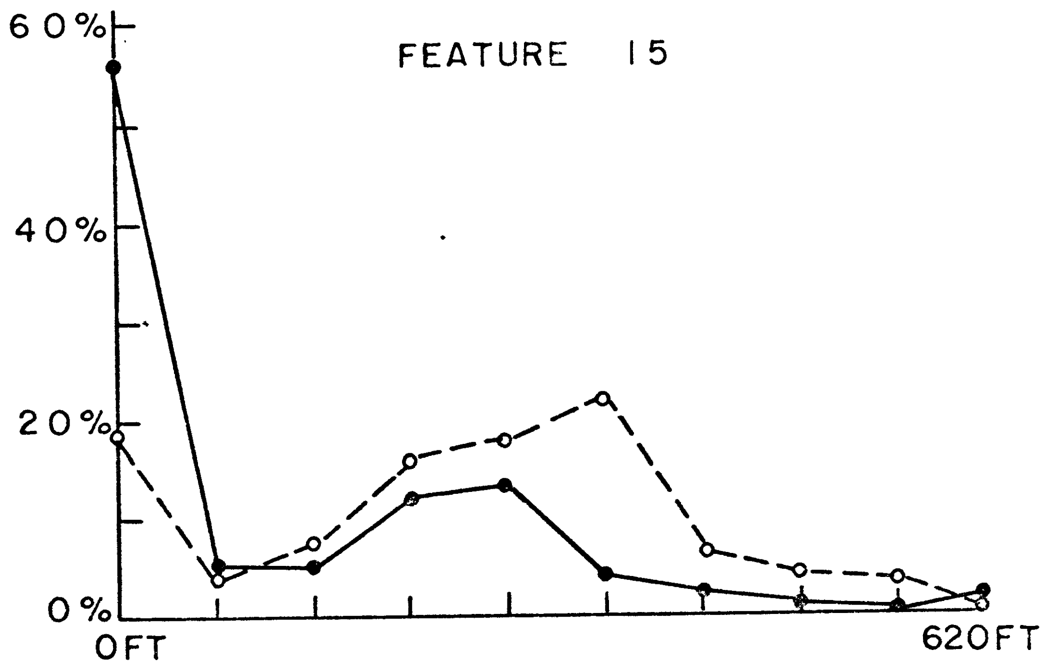

FEATURE 10
40%
20%
0%
0 FT
460 FT

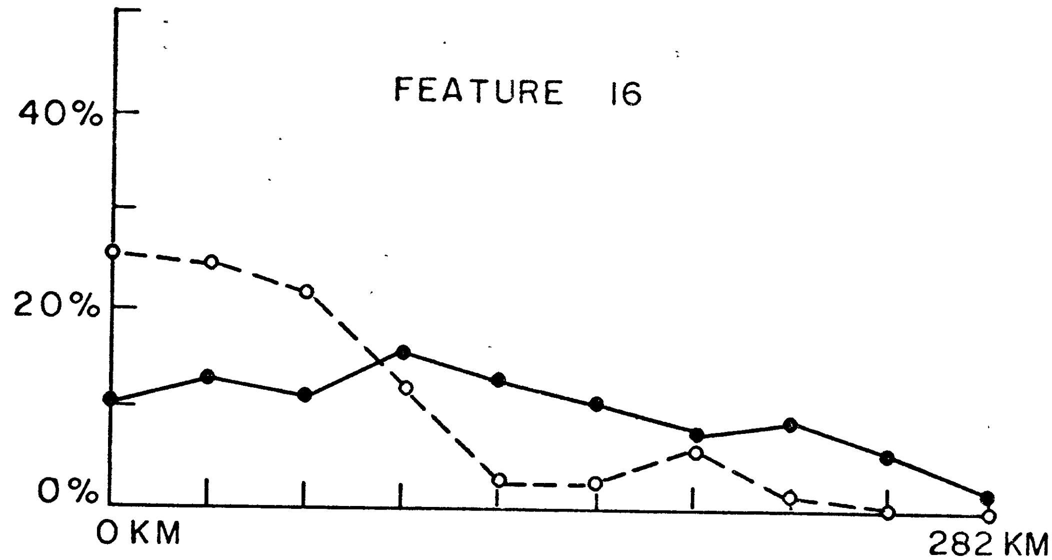

FEATURE 11
40%
20%
0%
0 KM
462 KM

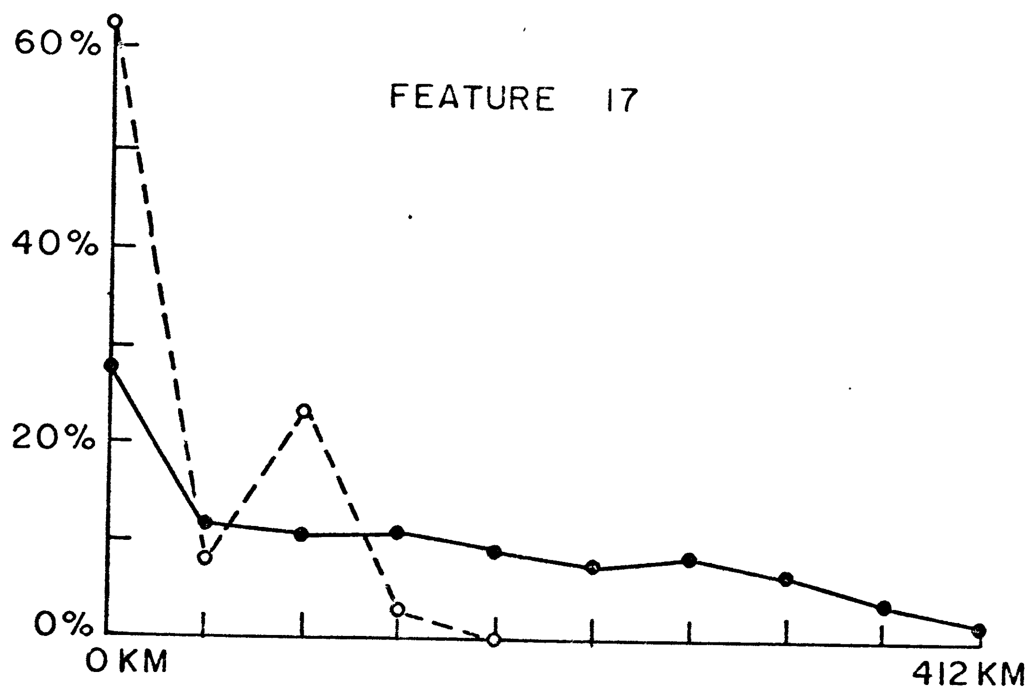

FEATURE 12
80%
60%
40%
20%
0%
0FT
500FT

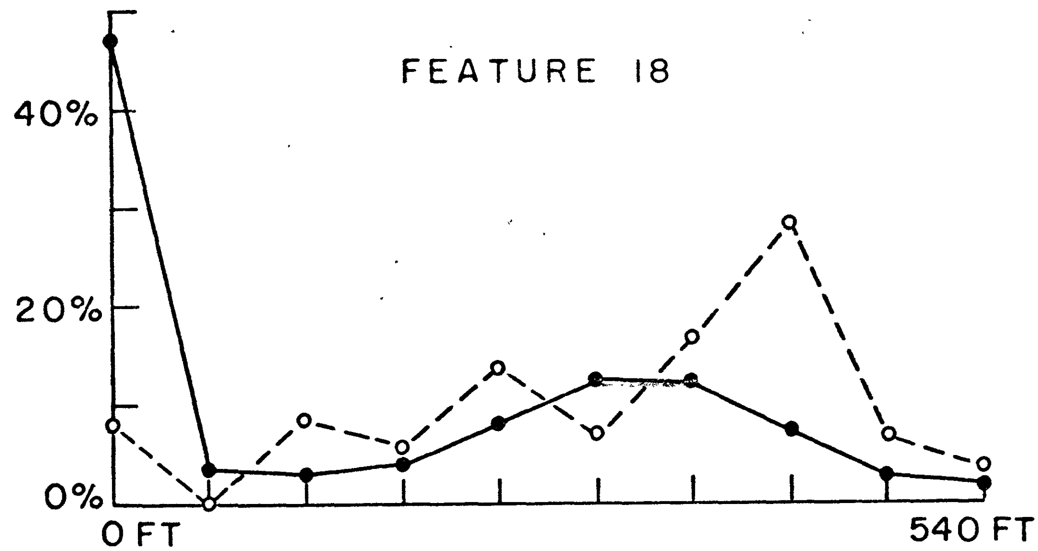

FEATURE 13
40%
20%
0%
0KM
452 KM

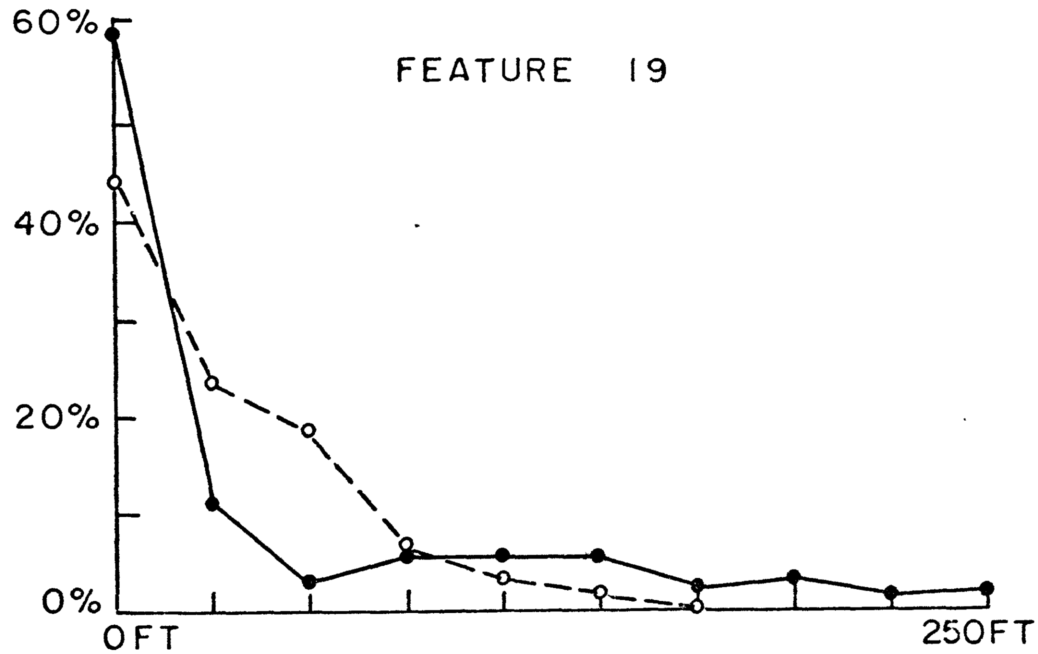

FEATURE 14
40%
20%
0%
0 KM
282 KM

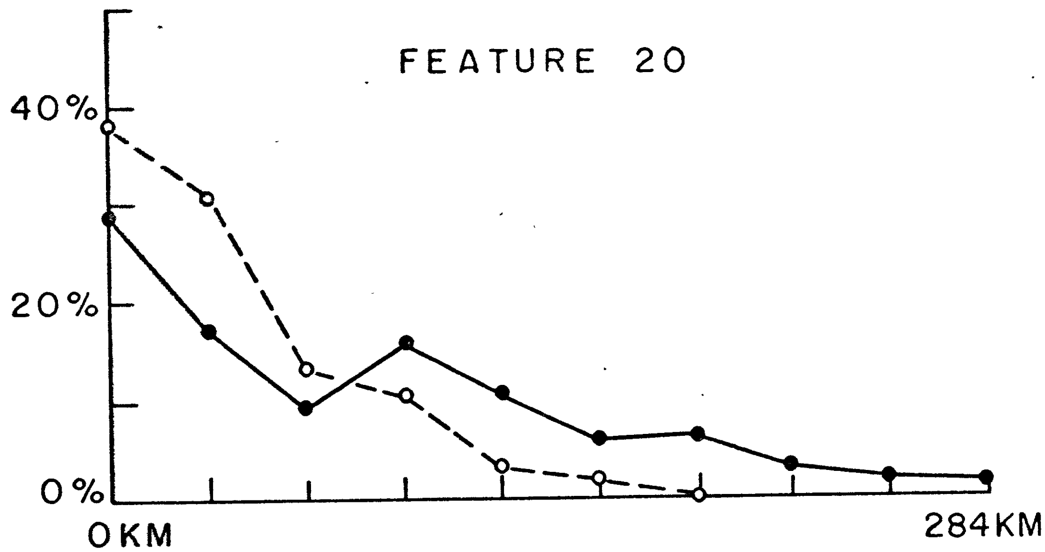

60%
FEATURE 15
40%
20%
0%
0 FT
620 FT

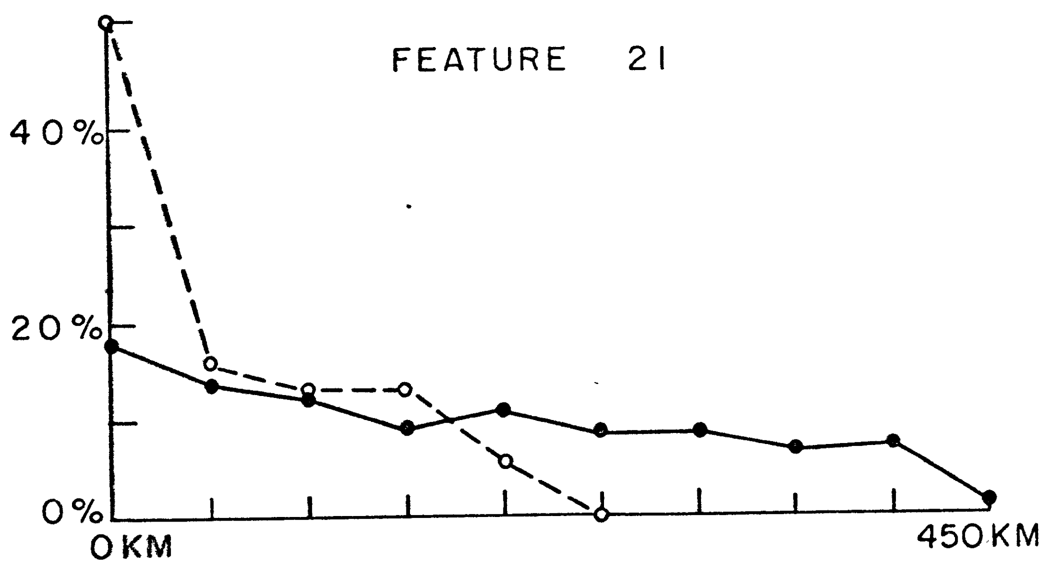

FEATURE 16
40%
20%
0%
0 KM
282 KM

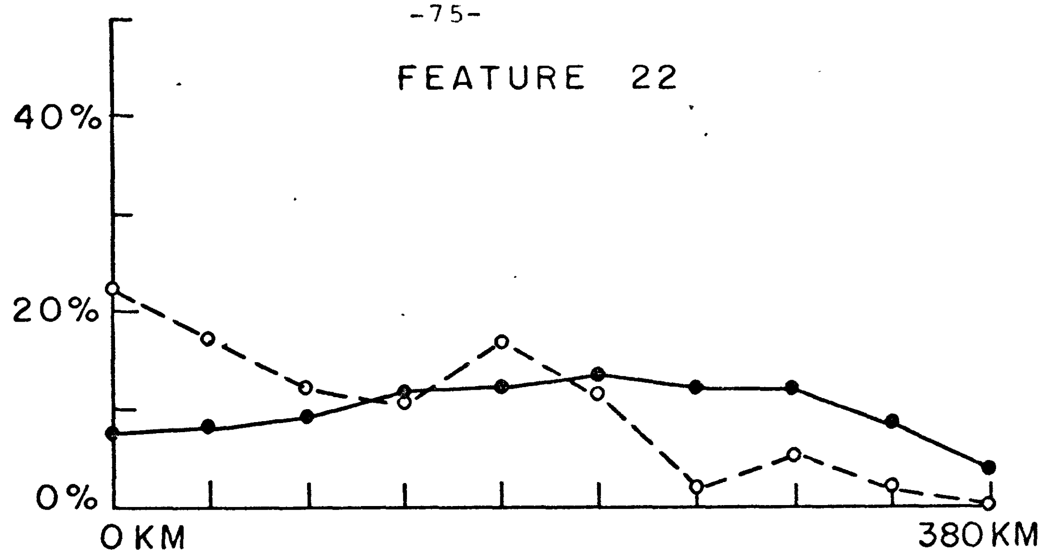

FEATURE 17
60%
40%
20%
0%
0 KM
412 KM

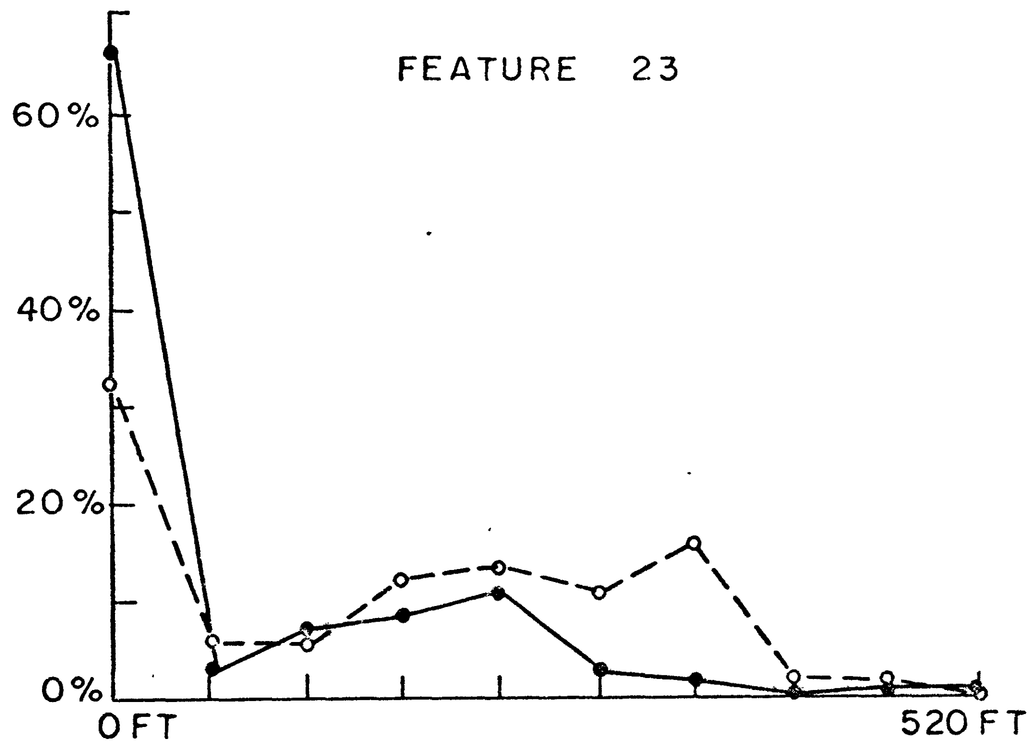

FEATURE 18
40%
20%
0%
0 FT
540 FT

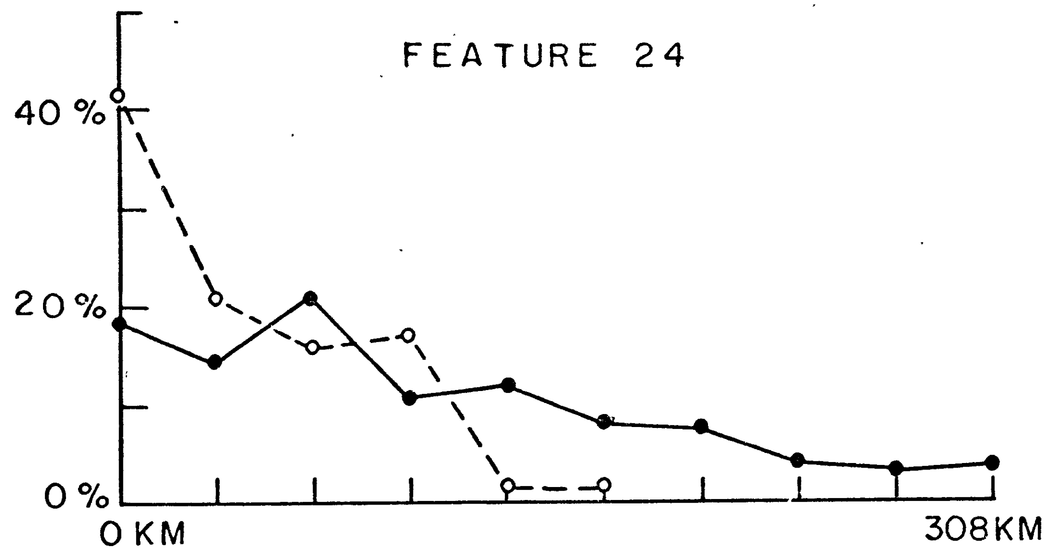

60%
FEATURE 19
40%
20%
0%
0 FT
250FT

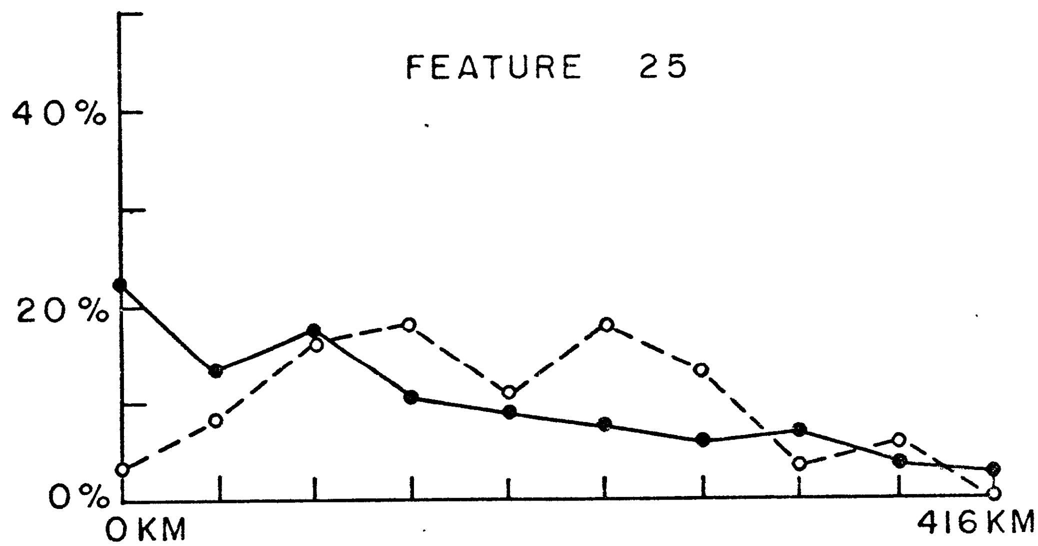

FEATURE 20
40%
20%
0%
0 KM
284 KM

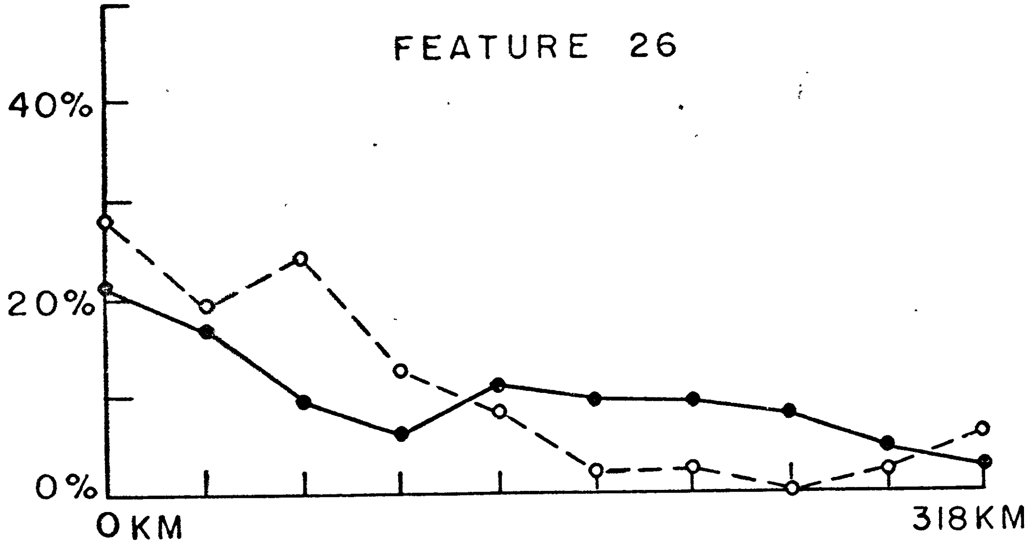

FEATURE 21
40%
20%
0%
0 KM
450 KM

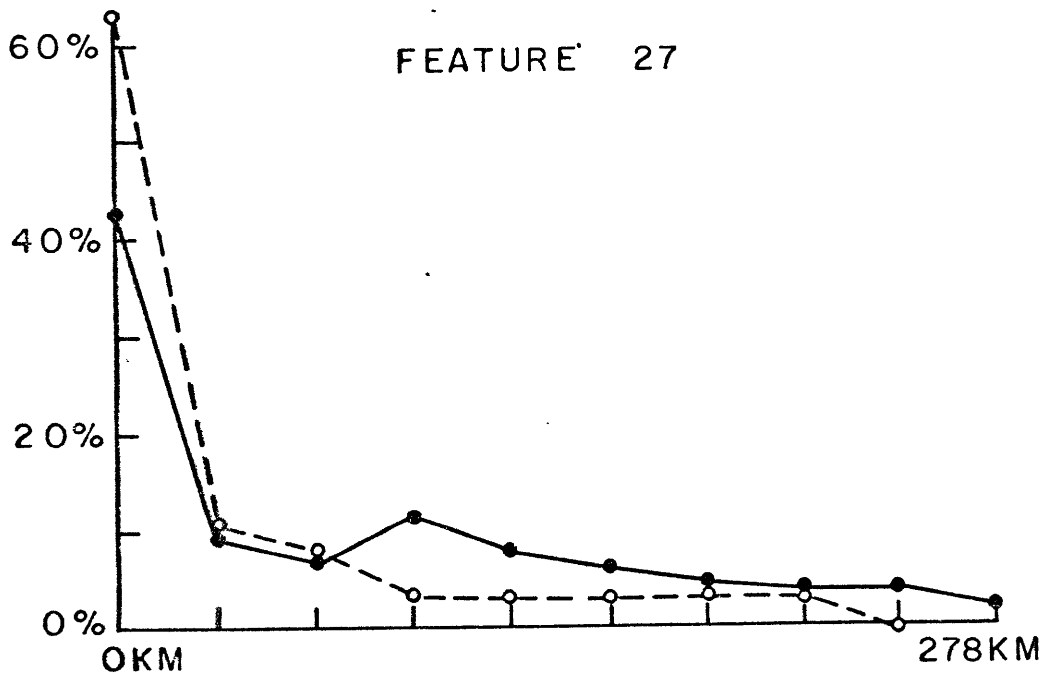

FEATURE 22
40%
20%
0%
0 KM
380 KM

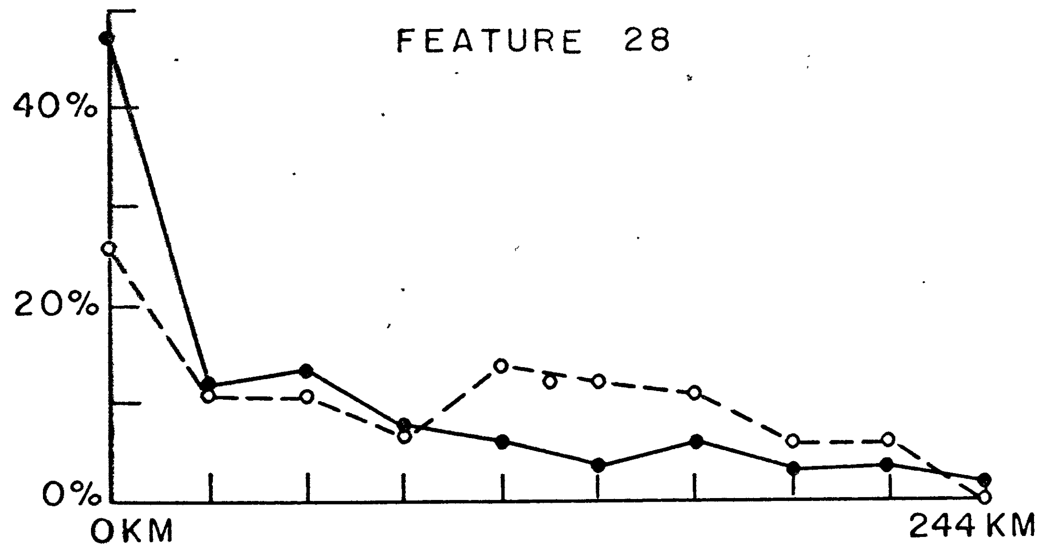

FEATURE 23
60%
40%
20%
0%
0 FT
520 FT

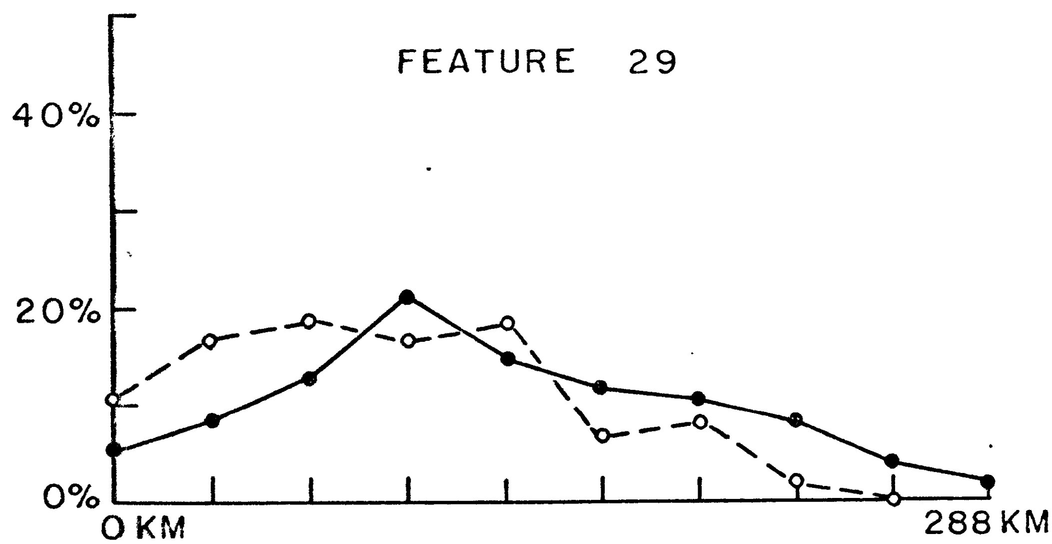

FEATURE 24
40 %
20 %
0 %
0 KM
308KM

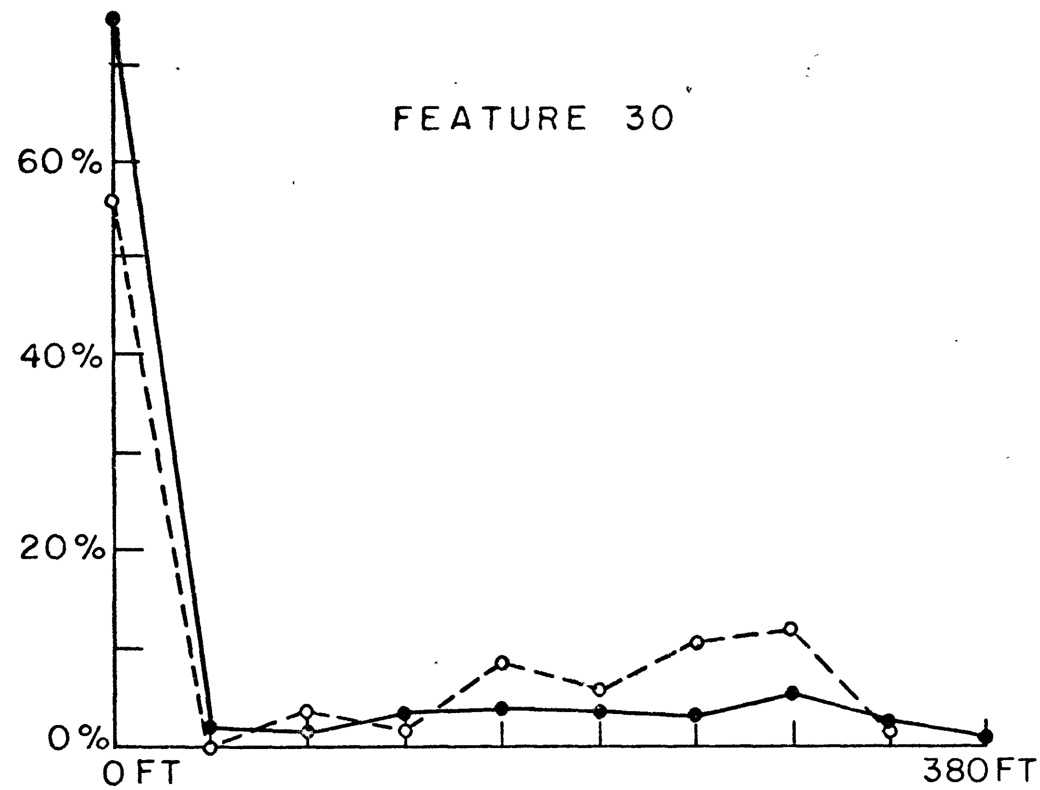

FEATURE 25
40 %
20 %
0 %
0 KM
416 KM

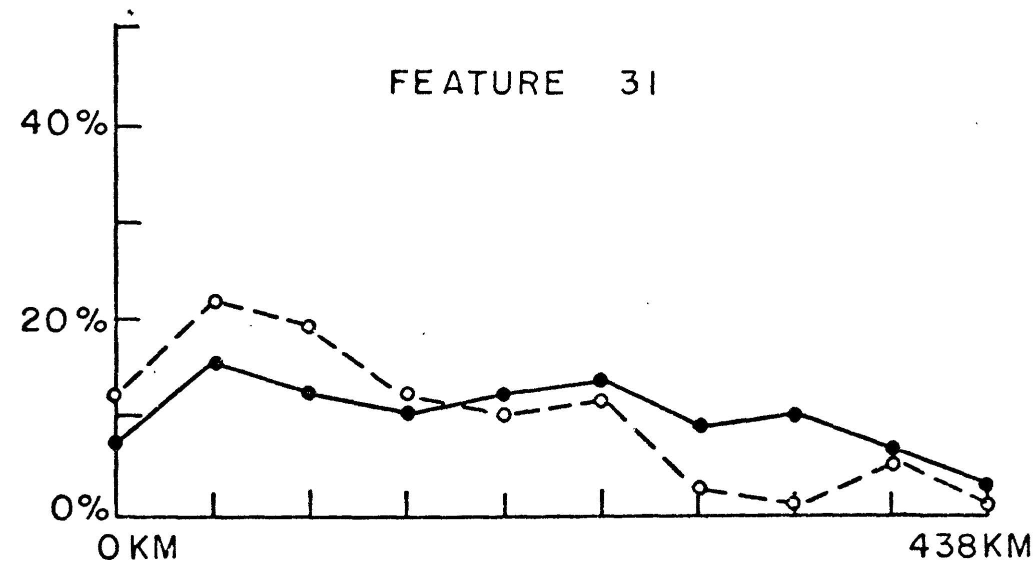

FEATURE 26
40%
20%
0%
O KM
318 KM

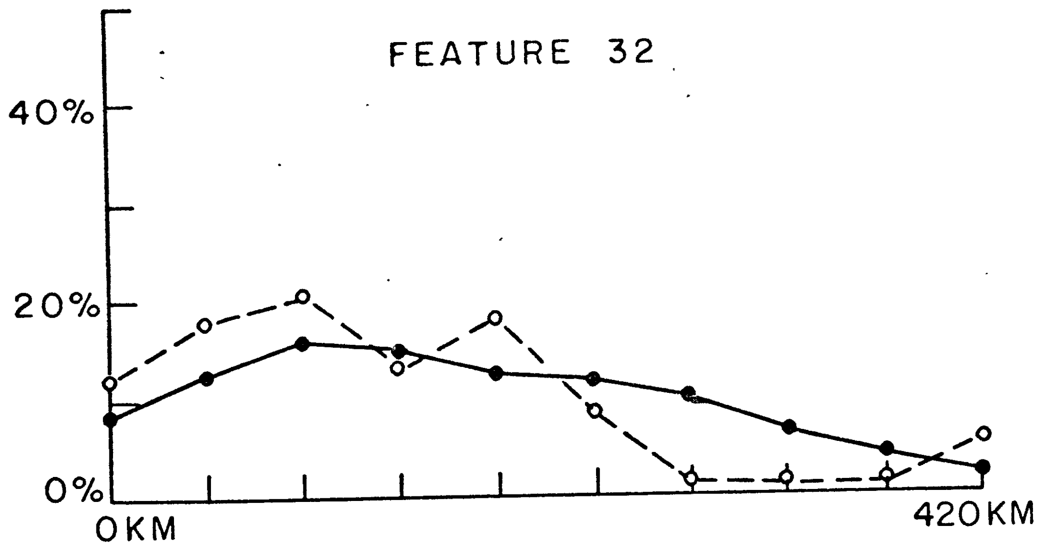

FEATURE 27
60%
40%
20%
0%
OKM
278 KM


FEATURE 28
40%
20%
0%
0 KM
244 KM


FEATURE 29
40%
20%
0%
0 KM
288 KM


FEATURE 30
60%
40%
20%
0%
0 FT
380 FT


FEATURE 31
40%
20%
0%
0 KM
438 KM


FEATURE 32
40%
20%
0%
0KM
420KM

3.5.2 Features for Recognition of Uranium Deposits in Triassic Sediments on the Colorado Plateau

1. Combined thickness of Navajo Sandstone and Kayenta Formation.

   (These are of Upper Triassic Age, and do not extend over the entire Plateau. Thickness of these formations carries information about ancient surface water flow. Groundwater flow may have been concentrated in areas of surface water flow concentration. Uranium is favored where these formations exceed 500 feet in thickness.)

2. Thickness of the Devonian System.

   (Areas of greater thickness are favored for uranium. Thickness of this system reflects, in part, distance from the Uncompahgre and Defiance and Central New Mexico uplifts, which influenced paleohydrology and sedimentation on the Plateau for considerable geologic time. Smaller scale irregularities in thickness may reflect pre-ore topographic or structural features.)

3. Proximity to a major anticline structure.

   (These large structural features, 100-200 km in length, such as the Zuni, Kaibab, and Uncompahgre uplifts, are pre-ore and exerted a considerable influence on paleohydrology. Proximity (< 100 km) to these uplifts is favored for ore. The flanks of these structures may have had a more stable hydrologic regime through time than other areas on the Plateau.)

4. <u>Proximity to the marine/non-marine sedimentary interface of Belle Fourche time</u>.

   (Areas nearer the paleocoastline are favored for uranium. The changing sedimentary environment near the coast may have allowed increased interbedding of sands, shales, and marine sediments. Interbedding of sediments of different porosities and permeabilities might increase the tortuosity of groundwater flow paths and slow groundwater flow so that oxidation/reduction reactions could proceed toward completion. Marine deposits may offer constituents, such as limestone, that influence the chemistry of uranium deposition.)

5. <u>Minimum elevation in the cell</u>.

   (Minimum, maximum, and average elevation in cells are strongly correlated; this feature offered the best separation of the classes. Lower modern elevations are favored for uranium. Modern elevations are not unrelated to paleoelevations on the Plateau, and erosion reflects the effects of post-ore surface water flow. Lower areas are expected to be areas of groundwater concentration. Possible sources of uranium include Precambrian rocks in uplifted areas that are now denuded of sediments. If groundwaters acquired uranium as they moved from highlands to lowlands, waters circulating in lowlands might, in general, have been exposed to more source material and might carry more uranium in solution.)

6. <u>Proximity to pinch-out of Carmel and Arapian Formations of Piper-Nesson Age</u>.

   (These Jurassic units resulted from a marine invasion of the Plateau from the north and west. Proximity to the boundary is favored for uranium. See feature 4.)

7. <u>Proximity to pinch-outs of the Entrada and Carmel Formations of Rierdon Age</u>.

   (These sediments are more widespread than those of Piper-Nesson Age. They result from a second Jurassic marine invasion. Pinch-outs indicate paleotopographic highs. Areas at moderate distance from pinch-outs are favored for uranium.)

8. Proximity to a pinch-out of Swift Age sediments, Summerville and Todilto Formations.

   (Another Jurassic marine invasion. Pinch-outs reflect paleo-topographic highs as well as areas of negligible sediment thickness. Areas at moderate distance from pinch-outs are favored for uranium.)

9. Thickness of upper Triassic sediments, Chinle and Dolores Formations.

   (Areas of moderate thickness of the principal Triassic uranium host formation are favored for uranium.)

10. Minimum thickness of the Salt Wash Member of the Morrison Formation in the cell.

    (Areas of thicker sediment cover are favored for uranium in underlying Triassic sediments. Thickness of the sediments overlying the host beds offer an indication of ground and surface water circulation after the deposition of the uranium host beds.)

11. Proximity to the Cretaceous shoreline of Late Skull Creek time.

    (Marine transgression may have influenced groundwater circulation in near-shore sediments. Proximity to this coast is favored for uranium. See feature 4.)

12. <u>Proximity to the Cretaceous shoreline of Late Maury time.</u>

    (Proximity to the shoreline is favored for uranium. See feature 4.)

13. <u>Maximum thickness of the Salt Wash Member of the Morrison Formation in the cell</u>.

    (Areas of greater thickness should represent foci for surface water flow, and are favored for uranium. See feature 10.)

14. <u>Maximum elevation in the cell</u>.

    (Areas at lower elevations are favored for the accumulation and deposition of uranium. See feature 5.)

15. <u>Proximity to a pinch-out of lowest Cretaceous sediments</u>.

    (Paleo-topographic highs are indicated.)

16. <u>Proximity to a pinch-out of the Recapture Member of the Morrison Formation</u>.

    (See feature 15.)

17. Proximity to the Cretaceous shoreline of Telegraph Creek time.

(See feature 4.)

TABLE 3-2: Ranks of Features for Recognition of Uranium Deposits in Triassic sediments on the Colorado Plateau.

Feature	Bayes Weight	Rank Sum Confidence	Information	Information Rank
1	2.2	99%	.155	7
2	2.1	99%	.202	2
3	1.7	99%	.128	11
4	1.6	99%	.075	16
5	1.6	99%	.210	1
6	1.5	99%	.133	10
7	1.4	<67%	.159	5
8	1.4	96%	.158	6
9	1.3	<67%	.016	17
10	1.3	95%	.134	9
11	1.2	<67%	.087	15
12	1.2	<67%	.122	13
13	1.2	99%	.142	8
14	1.1	99%	.199	3
15	0.8	97%	.181	4
16	0.8	92%	.125	12
17	0.7	99%	.105	14

FIGURE 3-5: Histogram estimates of state-conditional probability density functions for features of uranium deposits in Colorado Plateau Triassic sediments. These 10-place histogram estimates are based on 45 U objects and 463 U* objects. The vertical axis measures percentages of U and U* populations; U PDF's are indicated by dashed lines, U* PDF's are indicated by solid lines.

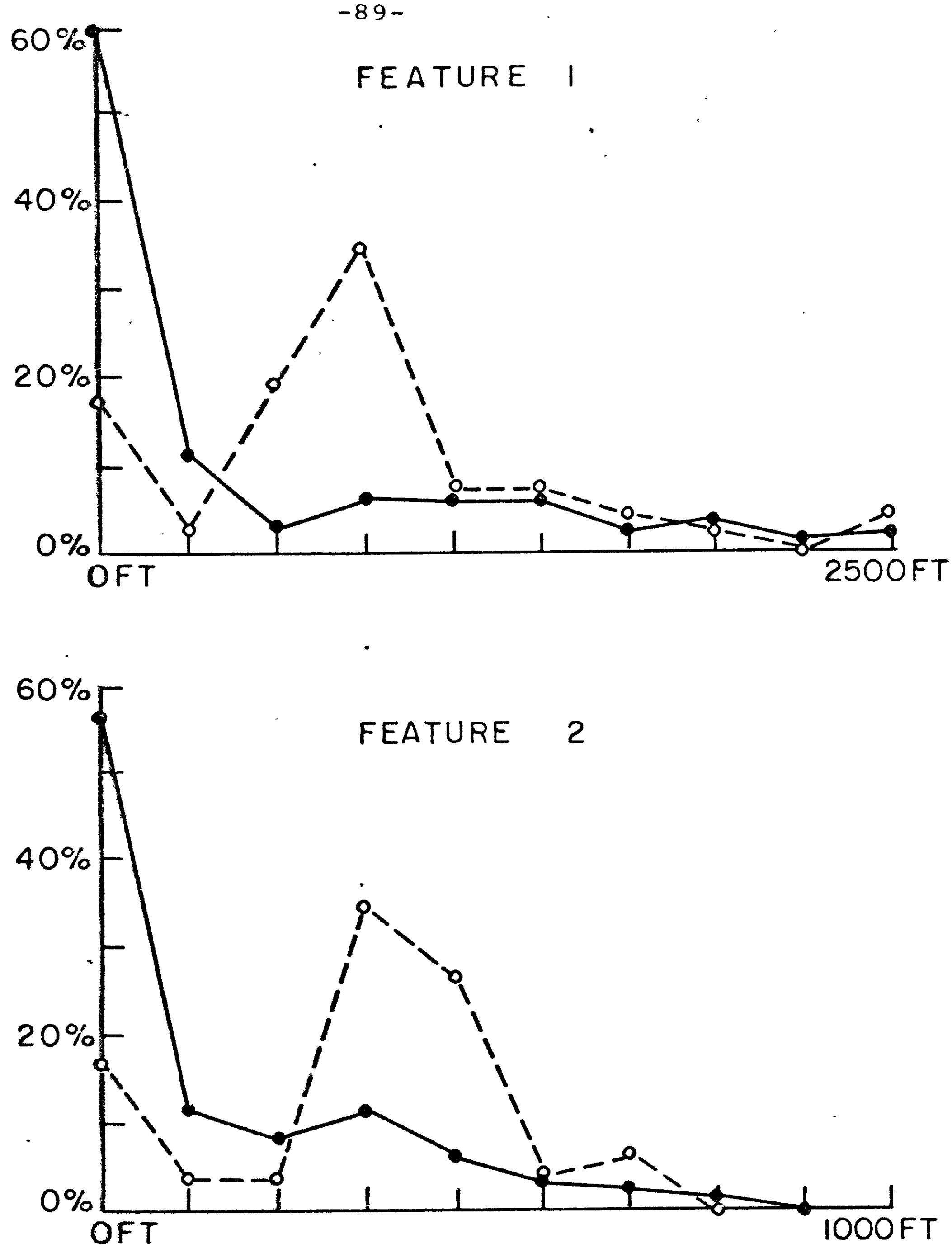
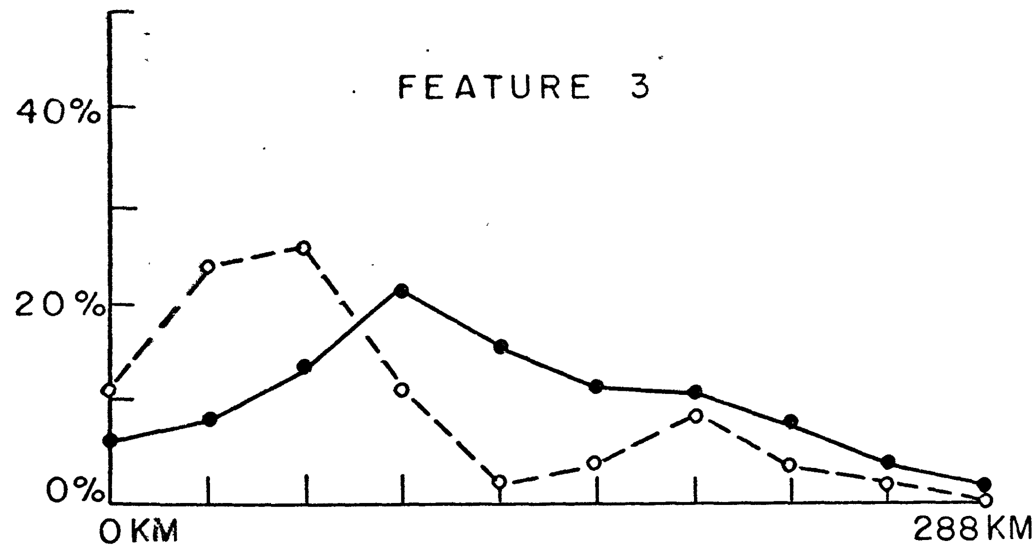
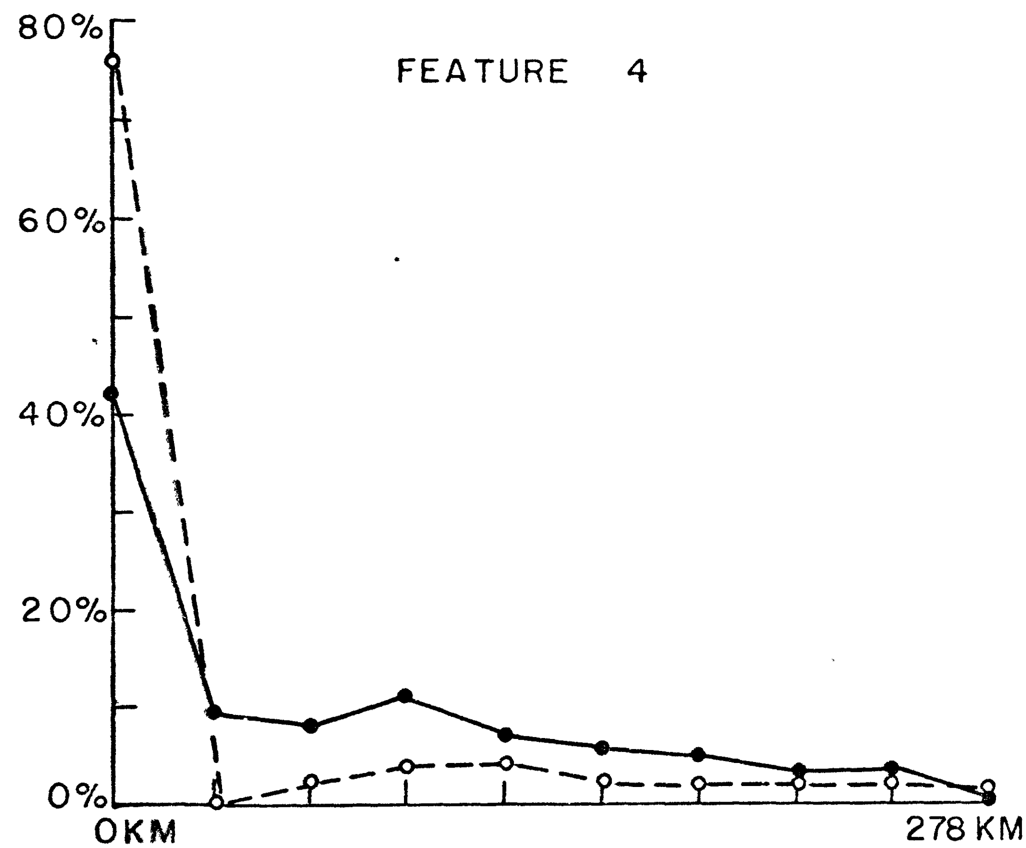
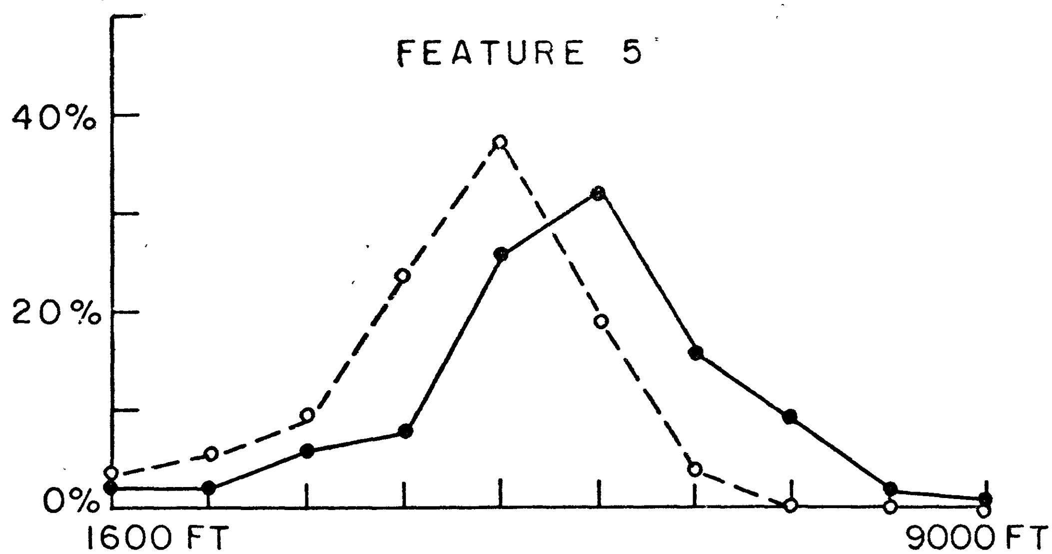
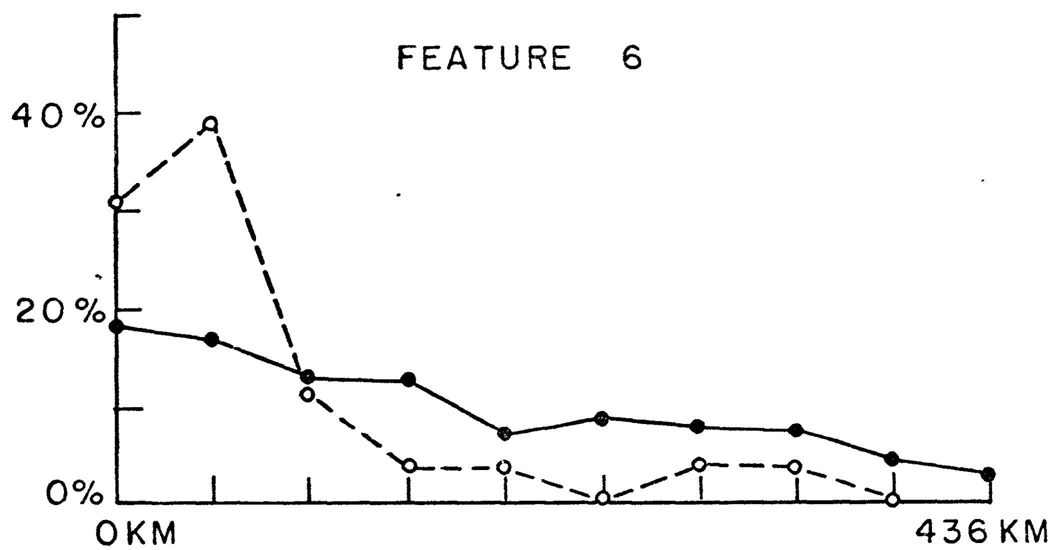


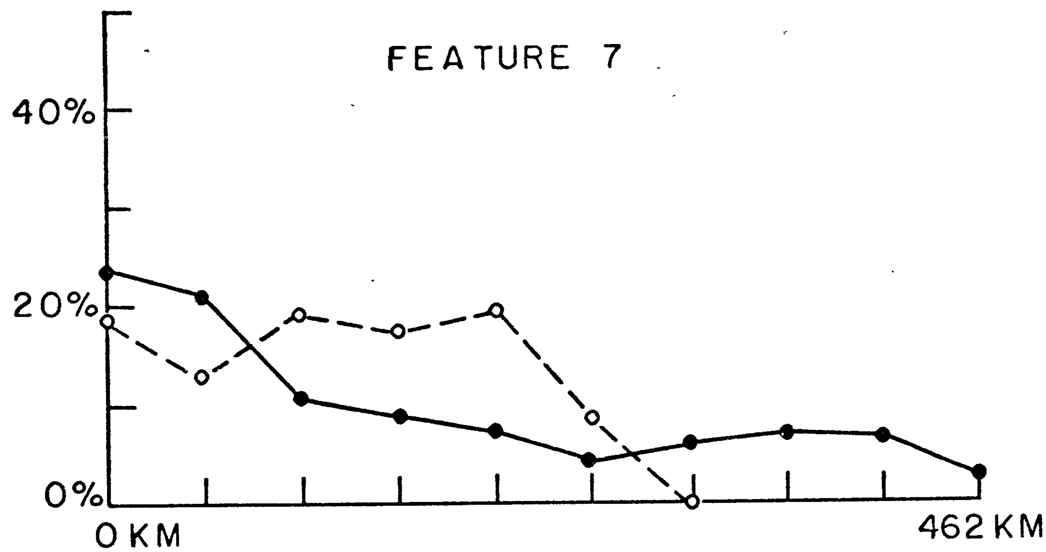
FIGURE 3-5

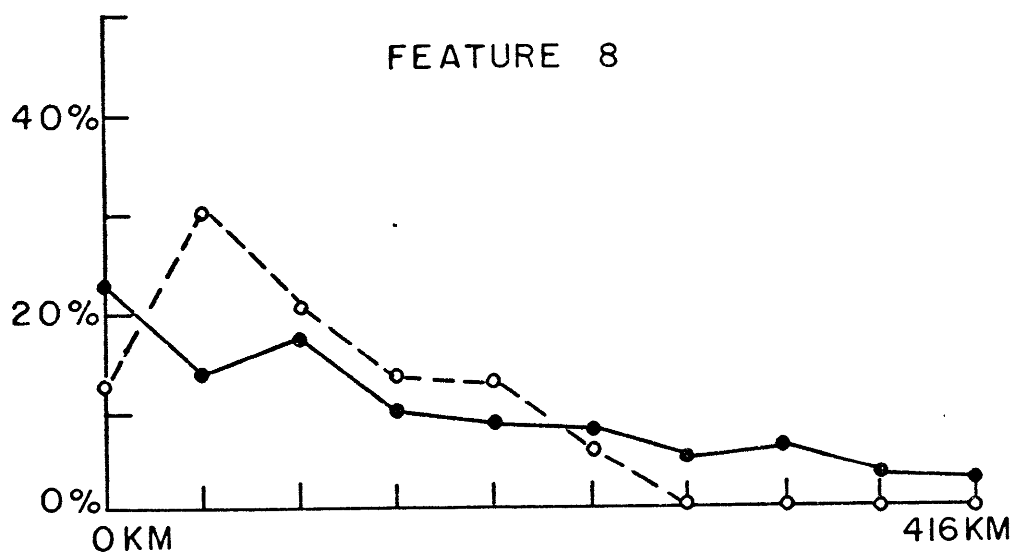

FEATURE 3
40%
20%
0%
0 KM
288 KM

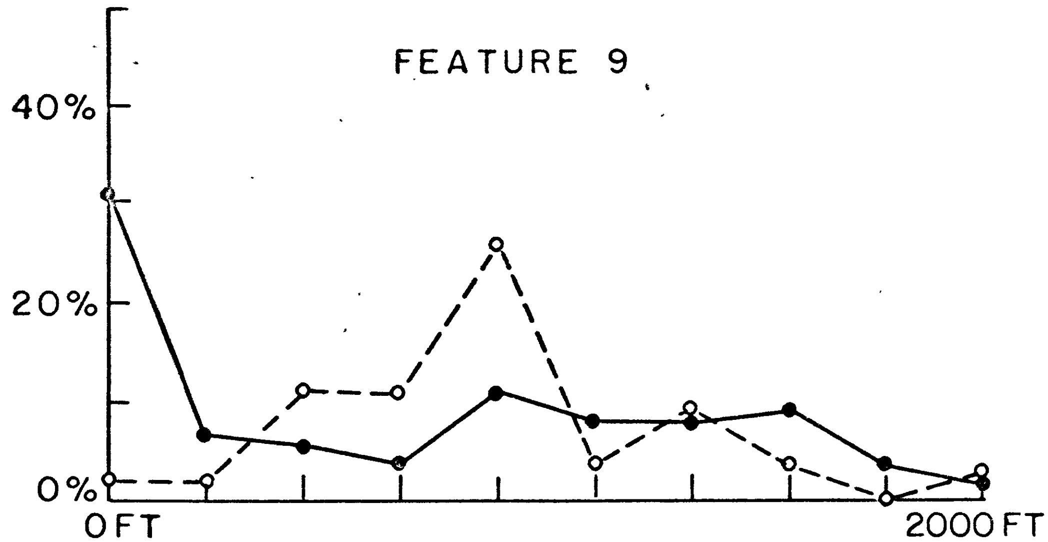

80%
FEATURE 4
60%
40%
20%
0%
0 KM
278 KM

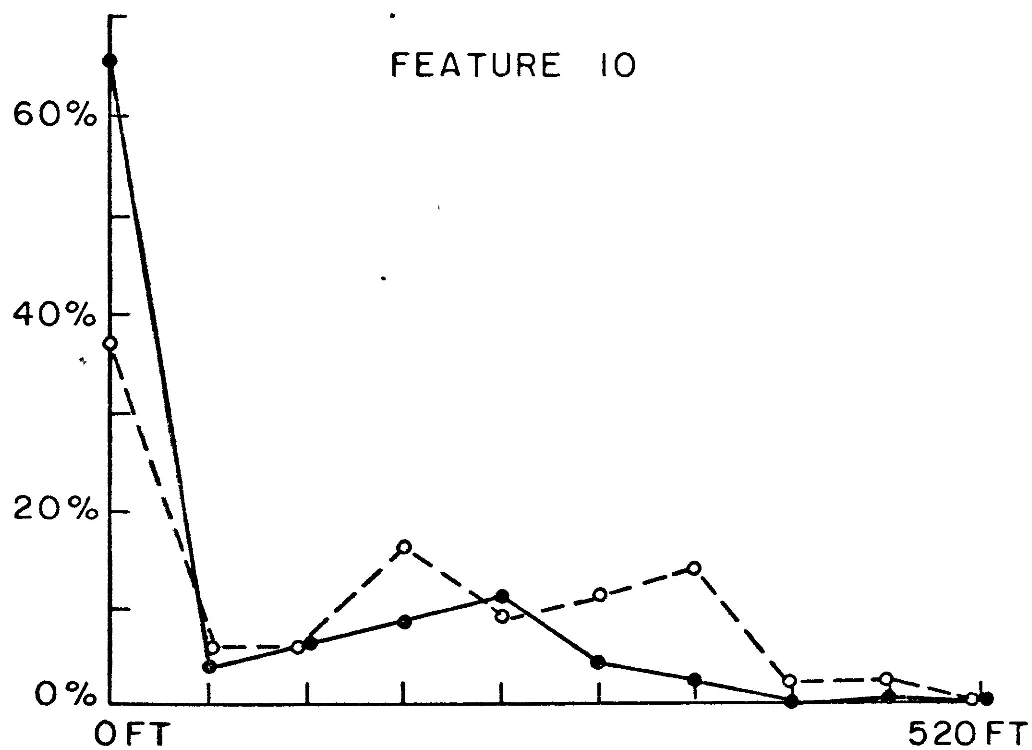

FEATURE 5
40%
20%
0%
1600 FT
9000 FT

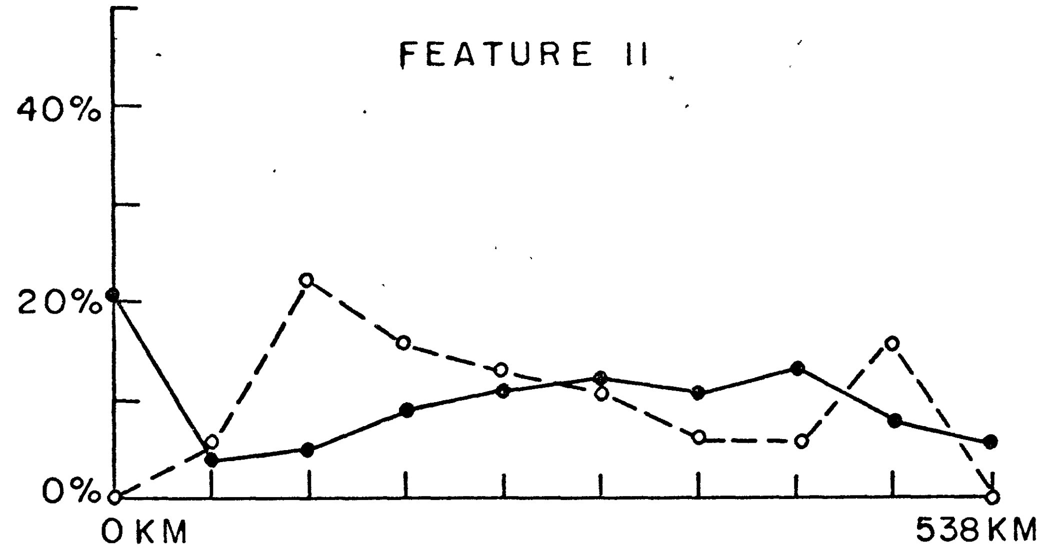

FEATURE 6
40%
20%
0%
0 KM
436 KM

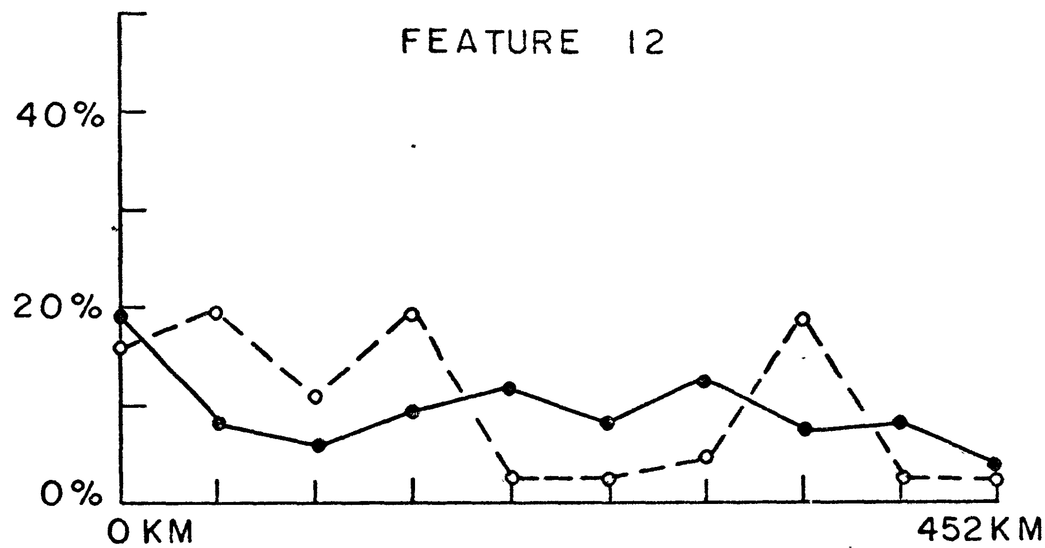

FEATURE 7
40%
20%
0%
0 KM
462 KM

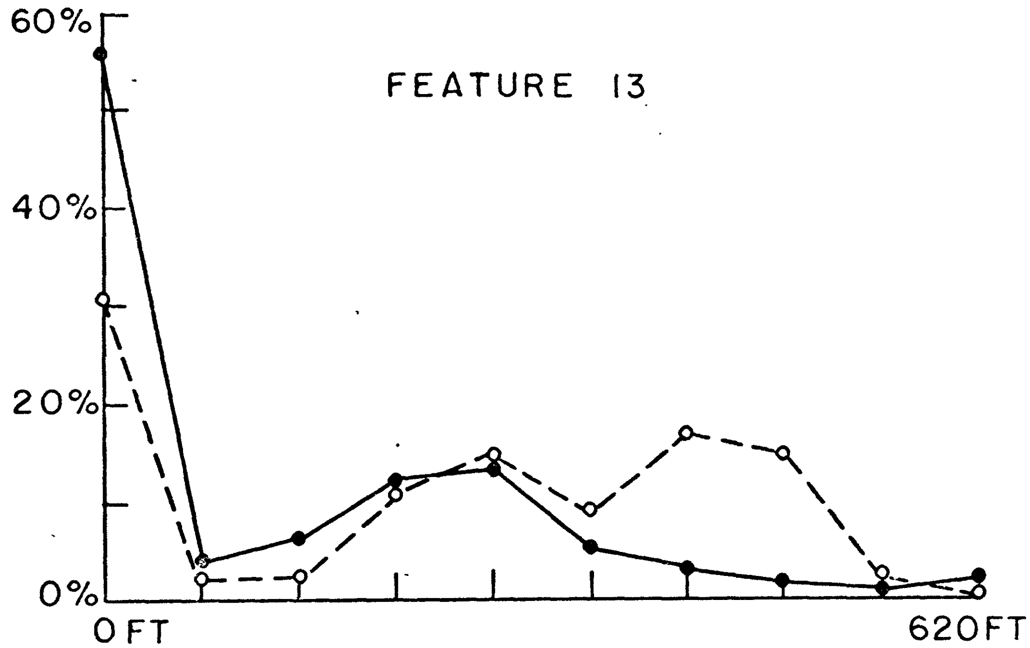

FEATURE 8
40%
20%
0%
0 KM
416 KM

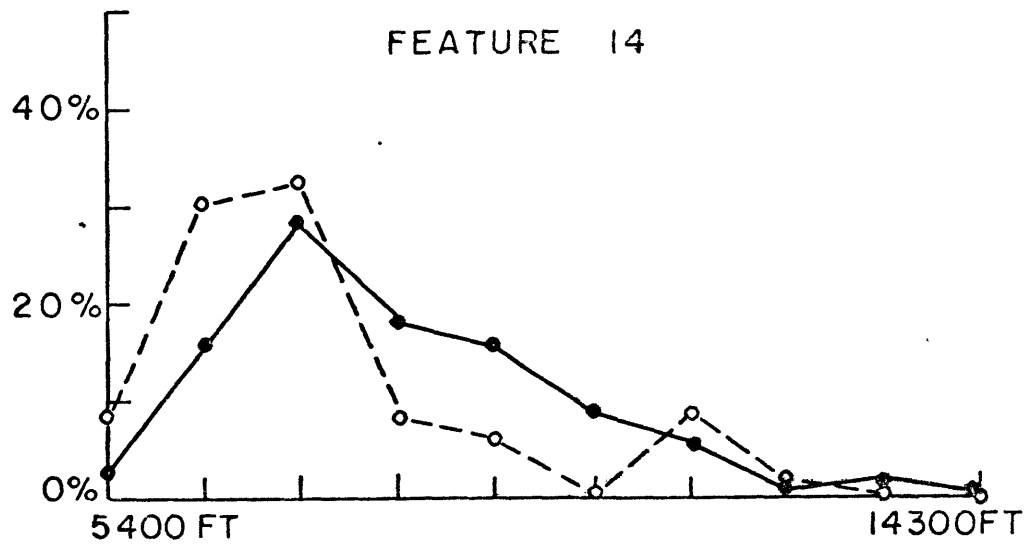

FEATURE 9
40%
20%
0%
0 FT
2000 FT

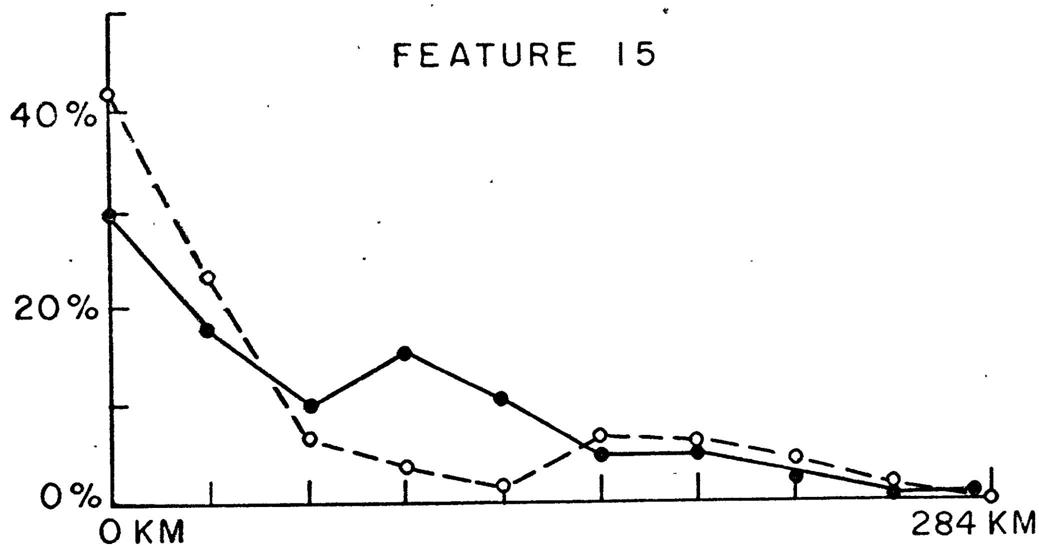

FEATURE 10
60%
40%
20%
0%
0 FT
520 FT

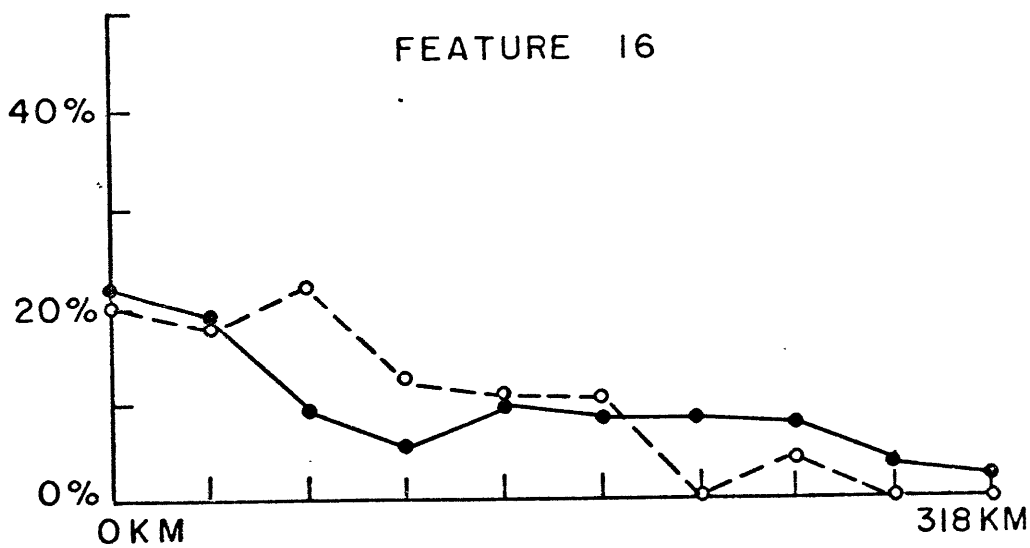

FEATURE 11
40%
20%
0%
0 KM
538 KM

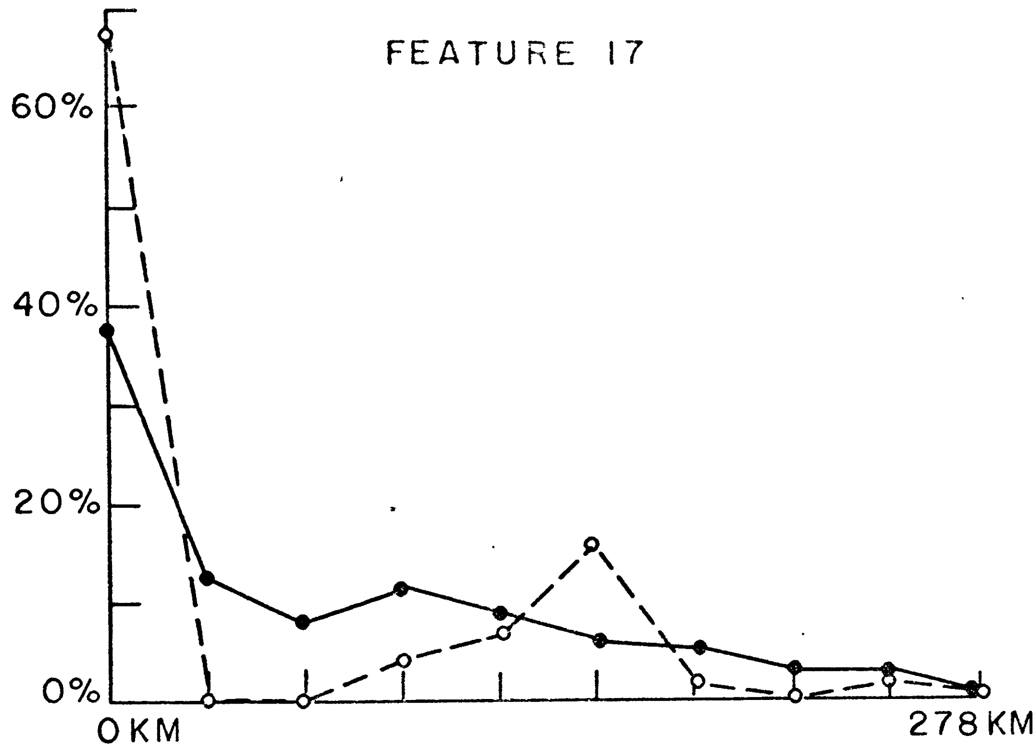

FEATURE 12
40%
20%
0%
0 KM
452 KM


60%
FEATURE 13
40%
20%
0%
0 FT
620FT


FEATURE 14
40%
20%
0%
5400 FT
14300FT


FEATURE 15
40%
20%
0%
0 KM
284 KM


FEATURE 16
40%
20%
0%
0 KM
318 KM


FEATURE 17
60%
40%
20%
0%
0 KM
278 KM

3.5.3 Features for Recognition of Uranium Deposits in the Casper, Wyoming Quadrangle

1. Proximity to major anomaly in airborne radiometric survey.

   (Will pick up surface shows of uranium, shallow deposits that have enriched overlaying soil with uranium; will not record deeper deposits; will pick up workings from mine operations.)

2. Proximity to an anomaly of any strength in airbone radiometric survey.

   (Includes all areas of feature 1, and smaller anomalies. There is no predictable relationship between the size of a deposit and uranium enrichment of overlaying soil; an aerorad anomaly of any size may be associated with a deposit of any size.)

3. Proximity to outcrop of mixed sandstone/limestone/ clay of Tertiary Age.

   (This may be a possible source of uranium; limestone and clay may have provided compounds favorable for uranium deposition. Despite Quaternary erosion, the location of modern surface exposures may help locate areas where meteoric waters could most easily have entered the formation.)

4. <u>Proximity to outcrop of Tertiary Wagonbed Formation</u>.

   (The Wagonbed is a possible uranium source bed; outcrops may mark areas where water could have entered the formation.)

5. <u>Proximity to a basin boundary</u>.

   (The interiors of basins mark concentrations of sediment and surface and groundwater flow.)

6. <u>Proximity to an extremum of terrain-corrected gravity anomaly</u>.

   (Highs mark mountain cores; lows mark basins; this feature may be of no other significance.)

7. <u>Proximity to a major LANDSAT lineament</u>.

   (These subcontinental scale linears may mark deep zones of basement weakness or mobility. Such zones might have influenced the depositional environment of uranium host beds, and could have provided conduits for the vertical migration of reducing gases derived from the decay of carbonaceous or petroliferous material.)

8. <u>Proximity to the axis of the drainage unit in which the point is located</u>.

   (Uranium source material collected from throughout a drainage unit will be concentrated by groundwater flow and sediment transport toward the axis of the drainage unit. Drainage patterns have not changed markedly here since the latest Tertiary.)

9. Proximity to the major axis of a sedimentary basin.
(This feature is similar to feature 8, but picks out with more certainty foci of paleodrainage in the Casper Quadrangle; drainage units are larger than the basins and include them.)

10. Proximity to nearest thrust fault.
(Faults with vertical displacement may provide important structural influences on groundwater flow. Even if faulting post-dates initial ore deposition, post-ore groundwater circulation may preserve, destroy, or relocate roll front deposits. Thrust faults are especially numerous in this area of largely vertical crustal movement.)

11. Proximity to outcrops of host formations, Fort Union and Wind River.
(Areas near outcrops appear favored for uranium because the most readily located ore bodies are those with surface shows. There appears to be no reason why these host beds should not hold reserves at depth

oxygenated water could assess buried strata around these linears, possibly removing uranium.)

16. <u>The total number of structural elements within 4 miles</u>.

    (The most complex regions are somewhat favored for uranium.)

17. <u>Proximity to second-closest extremum of terrain-corrected gravity anomaly</u>.

    (Areas close to two extrema are favored for uranium. This feature may only measure proximity to basins and ancient, granitic mountain cores.)

18. <u>Proximity to a minor LANDSAT linear of northwesterly trend</u>.

    (Areas closer to this group of sub-parallel linears are less favorable for uranium. This feature overlaps with feature 15.)

19. <u>Number of structural elements between point and nearest major anticline axis</u>.

    (Groundwaters will tend to move downdip away from anticlinal folds. Uranium deposits tend to occur closer to the anticlines, before numerous structural features intervene on water ciculation.)

in areas away from outcrops.)

12. <u>Proximity to Precambrian rock outcrop</u>.

    (Areas near uplifted Precambrian rock are favored for uranium. The ancient granites may have been a source of uranium.)

13. <u>Proximity to an aeromagnetic lineament</u>.

    (Areas near aeromagnetic lineaments are favored for uranium. These magnetic anomalies may mark basement faults, which, if continually remobilized, could have distributed accumulating sediments and may mark zones along which chemical species could migrate vertically.)

14. <u>Proximity to an uplift boundary</u>.

    (Areas near uplift boundaries are favored for uranium. Uranium-bearing waters may have entered host beds at these points.)

15. <u>Proximity to both northeast- and northwest-trending minor LANDSAT linears</u>.

    (This feature measures lineament density. Areas near linears from both of these suborthogonal sets of linears are not favored for uranium, although minor linears may enhance groundwater flow. If the smallest, least developed linears are post-ore, it may be that

20. <u>Average bed dip within 4 miles of a point</u>.

(Areas with gently dipping beds are favored for uranium over more steeply dipping or chaotic areas.)

21. <u>Number of dip-slip faults within 4 miles</u>.

(Areas of greater structural complexity are favored for uranium. This feature and feature 10 both suggest important structural controls on ore emplacement by faults with vertical offset.)

22. <u>Proximity to areas of Quaternary sedimentation</u>.

(Areas that have continued to receive sediment since the Tertiary are favored for uranium - another suggestion that the drainage characteristics of the area have not changed markedly since ore deposition.)

23. <u>Proximity to a minor LANDSAT linear of any azimuth</u>.

(Areas immediately surrounding minor lineaments are not favored for uranium, although minor linears may enhance groundwater flow. If the smallest, least developed linears are post-ore, it may be that oxygenated water could access buried strata around these linears, possibly removing uranium. This feature overlaps with feature 15.)

24. <u>Number of anticline axes within 4 miles</u>.

(Areas near anticlinal folds are favored for uranium. A structural control on groundwater is suggested.)

25. <u>Proximity to an intersection of major LANDSAT lineaments</u>.

(Areas near intersections of major lineaments that may reflect deep, pre-ore crustal structures are favored for ore.)

26. <u>Number of thrust faults between point and the nearest major anticline axis</u>.

(Areas that do not have thrust faults intervening between them and the nearest anticline are favored for uranium.

27. <u>Proximity to a minor syncline axis</u>.

(Areas closer to synclinal features are favored for uranium. A focusing or concentration of groundwater flow is suggested.)

28. <u>Gradient of terrain-corrected gravity anomaly</u>.

(Areas of high lateral gradient are favored for uranium. This feature may only pick out uplift/ basin interfaces.

29. Maximum terrain-corrected gravity anomaly within 4 miles.

    (Areas near positive anomalies are favored for uranium. This feature probably picks out areas near granitic source rocks.)

30. Maximum elevation within 3 miles.

    (Highlands within the study area are favored for uranium. Perhaps known deposits are those closer to possible uranium sources in ancient mountain areas.)

31. Proximity to outcrop of Tertiary White River Formation.

    (Proximity weakly favored for uranium. Features 31, 32, 33, and 34 relate known uranium deposits to outcrops of those strata that were deposited most nearly contemporaneously with the ores.)

32. Proximity to outcrop of Tertiary Moonstone Formation.

    (Proximity weakly favored for uranium.)

33. Proximity to outcrop of Tertiary Fort Union Formation.

    (Proximity weakly favored for uranium.)

34. Proximity to outcrop of Mesozoic Lance Formation.

    (Proximity weakly favored for uranium.)

35. Number of syncline axes within 4 miles.

(The axes of major synclines are not favored for uranium. Smaller synclinal folds and the flanks of larger synclines are favored; see feature 27).

36. Proximity to drainage divide.

(Proximity favored for uranium. This feature may recognize uplift boundaries; see feature 14.)

TABLE 3-3: Ranks of Features for Recognition of Uranium Deposits in the Casper, Wyoming Quadrangle.

Feature	Bayes Weight	Rank Sum Confidence	Information	Information Rank
1	2.3	99%	0.15	35
2	2.1	99%	0.17	31
3	2.0	< 68%	0.19	29
4	2.0	95%	0.23	27
5	1.9	90%	0.25	19
6	1.8	68%	0.19	30
7	1.7	99%	0.29	8
8	1.6	97%	0.26	17
9	1.5	94%	0.28	11
10	1.4	< 68%	0.27	13
11	1.4	95%	0.23	28
12	1.3	86%	0.32	4
13	1.2	96%	0.34	3
14	1.2	80%	0.25	20
15	1.2	95%	0.25	21
16	1.1	88%	0.36	2
17	1.1	< 68%	0.26	18
18	1.1	95%	0.27	14
19	1.1	85%	0.29	9
20	1.1	NA	0.32	5
21	1.0	94%	0.16	33
22	1.0	95%	0.16	34
23	1.0	86%	0.31	7
24	1.0	NA	0.38	1
25	1.0	96%	0.25	22
26	0.9	NA	0.28	12
27	0.9	68%	0.24	24
28	0.9	68%	0.29	10
29	0.9	84%	0.07	36
30	0.9	92%	0.25	23
31	0.9	73%	0.27	15
32	0.9	< 68%	0.17	32
33	0.8	68%	0.24	25
34	0.7	< 68%	0.27	16
35	0.7	NA	0.24	26
36	0.6	92%	0.32	6

FIGURE 3-6: Histogram estimates of state-conditional probability density functions for features of uranium deposits in the Casper Quadrangle of Wyoming. These 10-place histogram estimates are based on 21 U objects and 359 U* objects. The vertical axis measures percentage of U and U* populations; U PDF's are indicated by dashed lines, U* PDF's are indicated by solid lines.

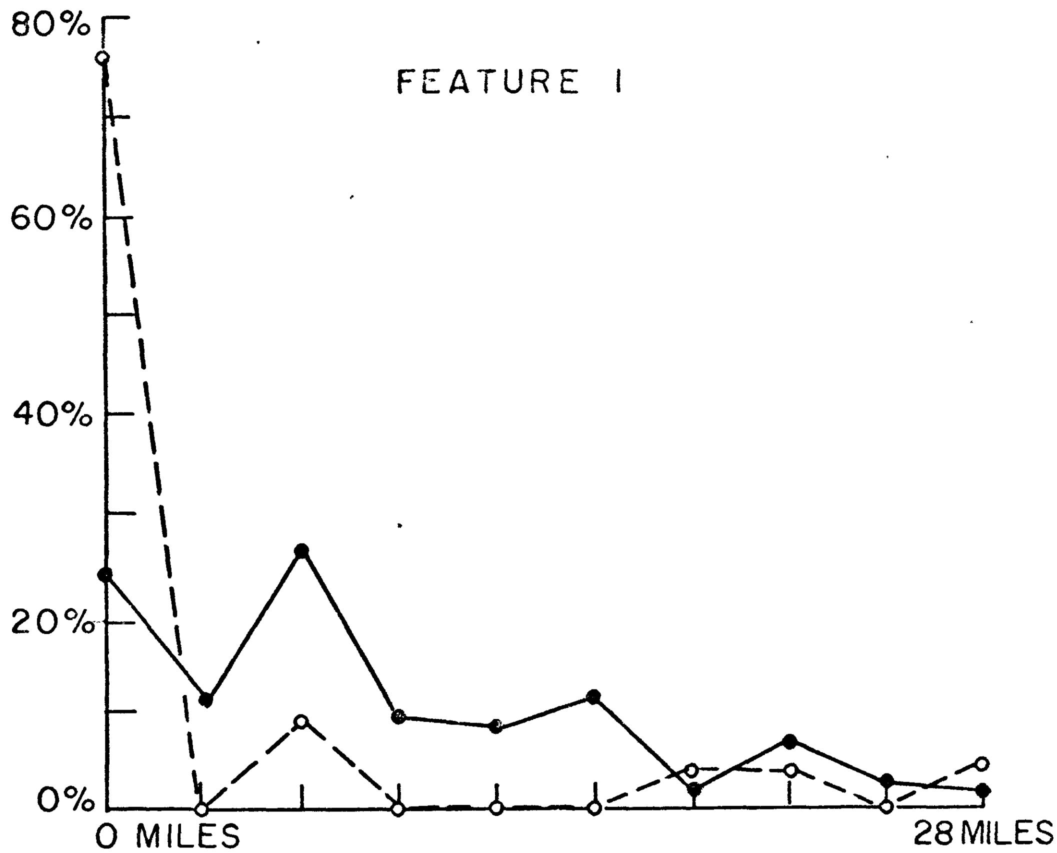
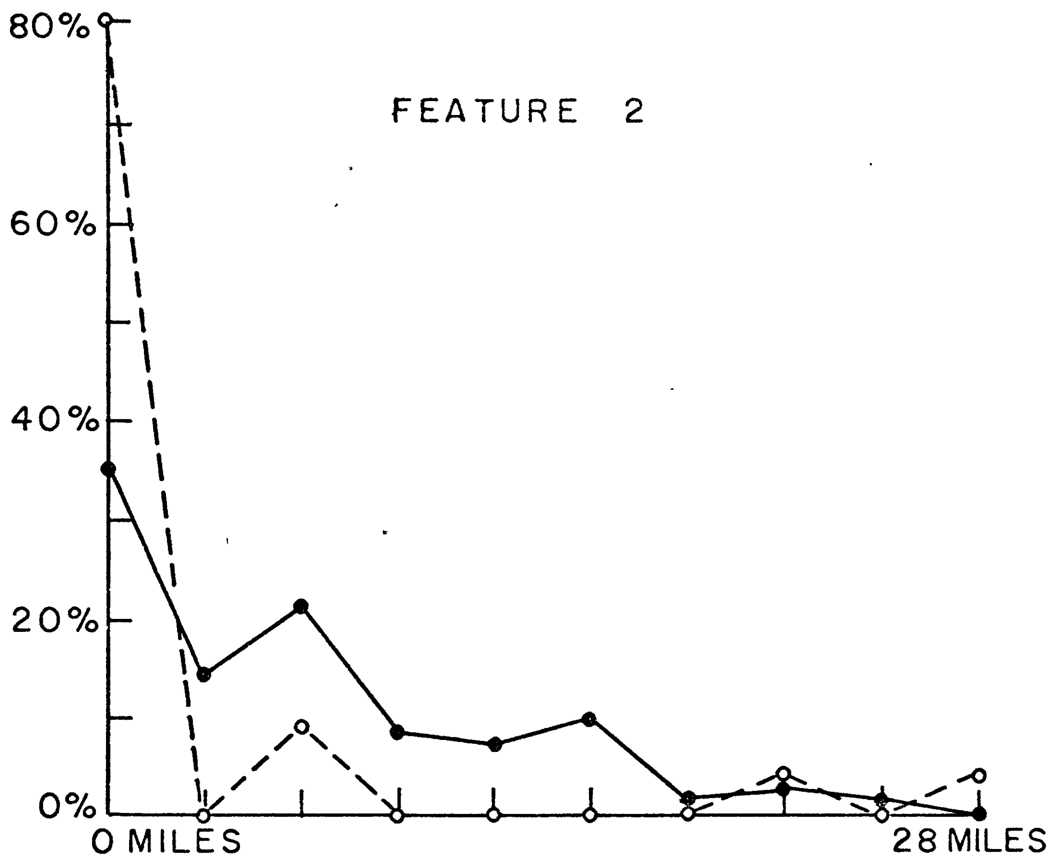
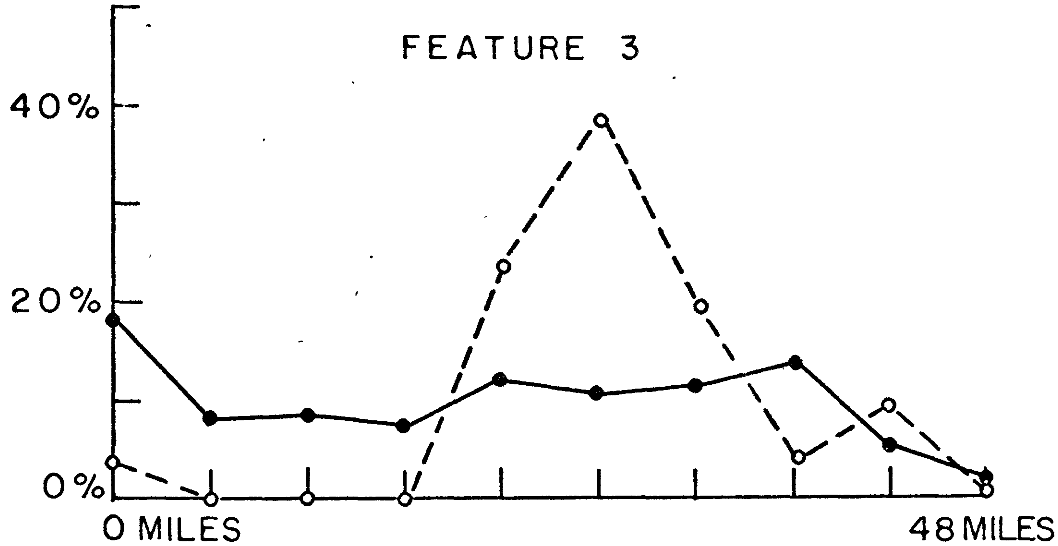
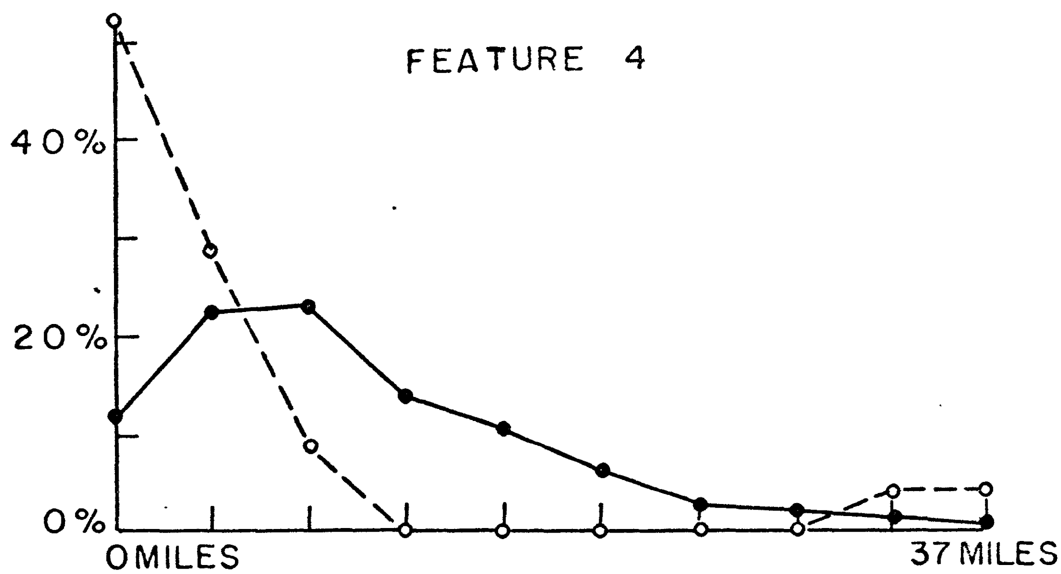
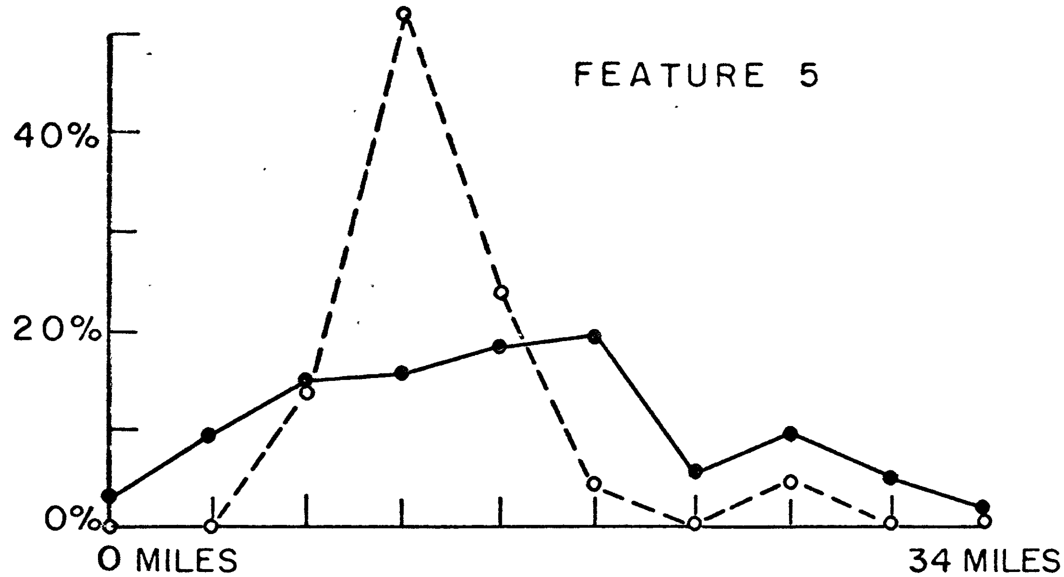


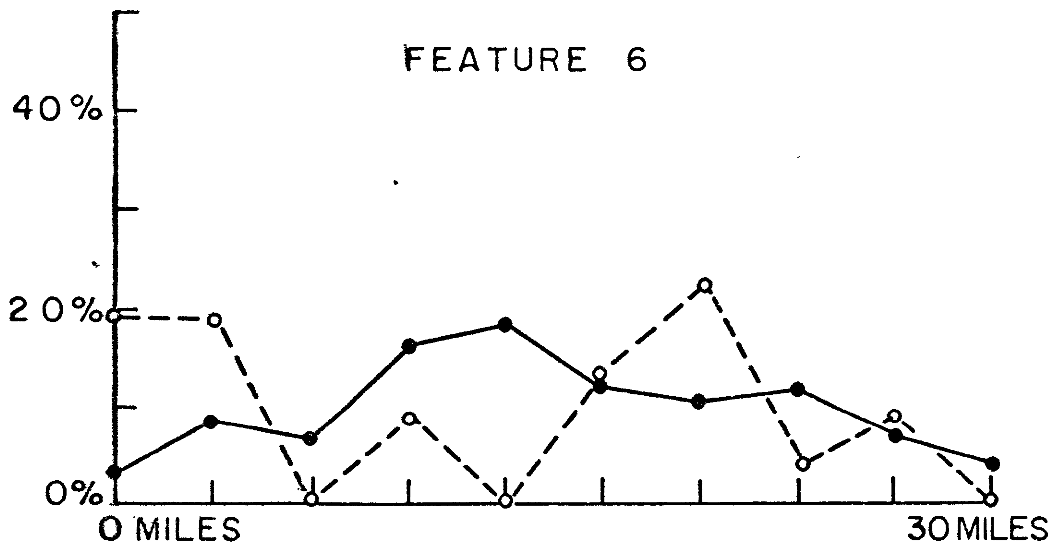
FIGURE 3-6

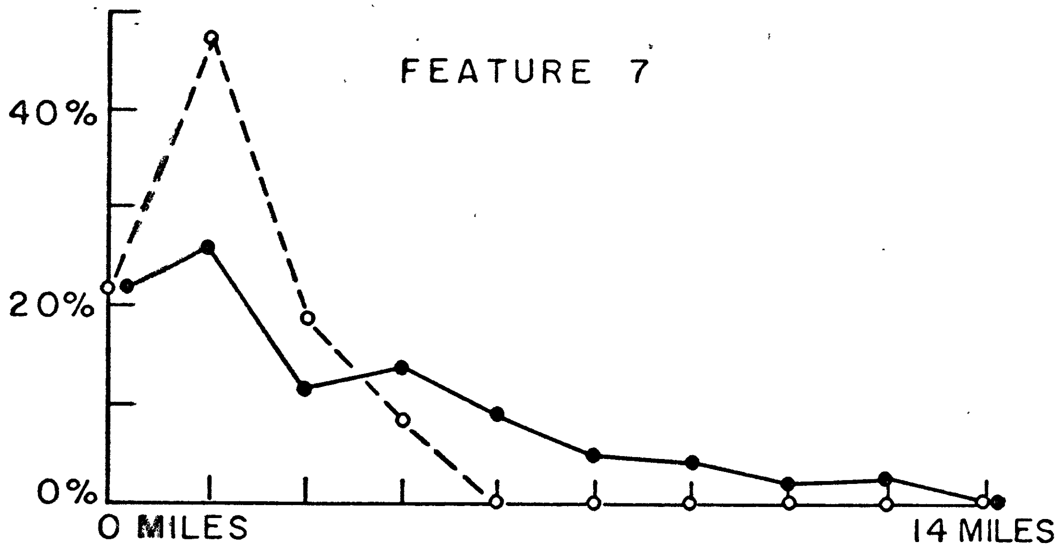

FEATURE 2
80%
60%
40%
20%
0%
0 MILES
28 MILES

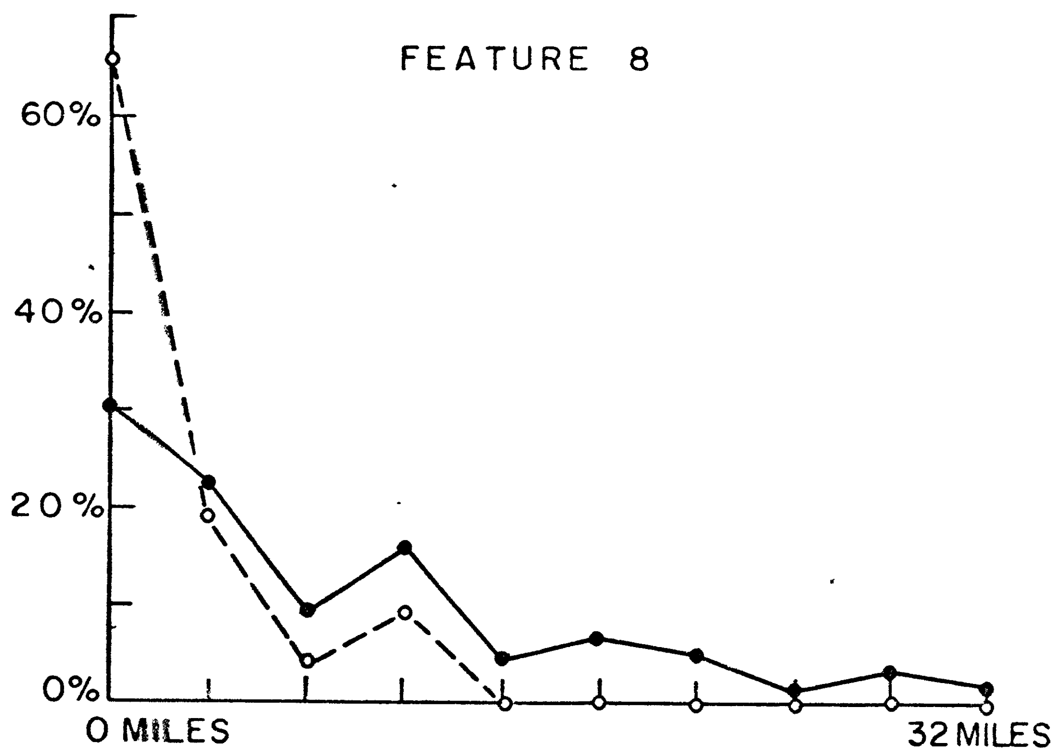

FEATURE 3
40%
20%
0%
0 MILES
48 MILES

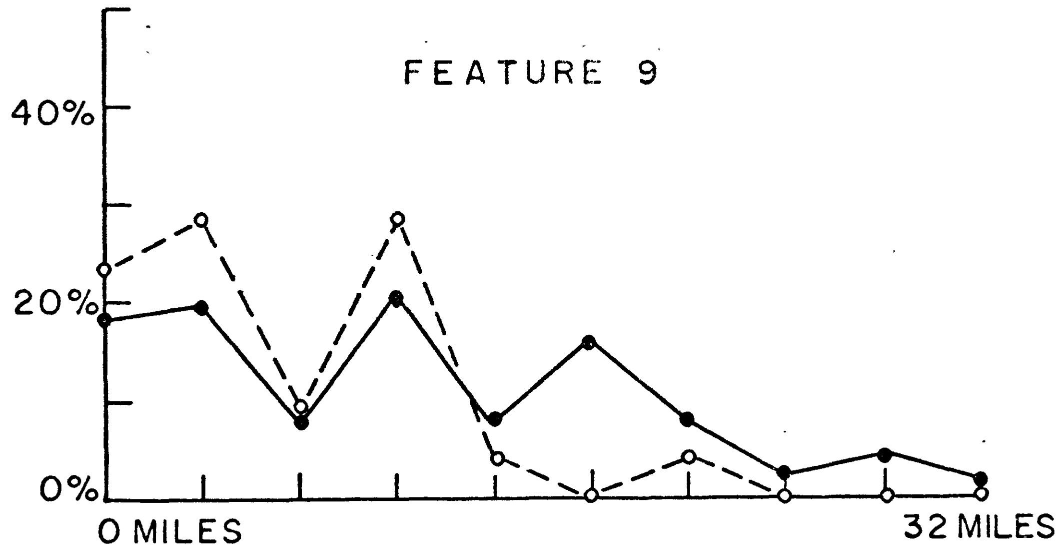

FEATURE 4
40%
20%
0%
0 MILES
37 MILES

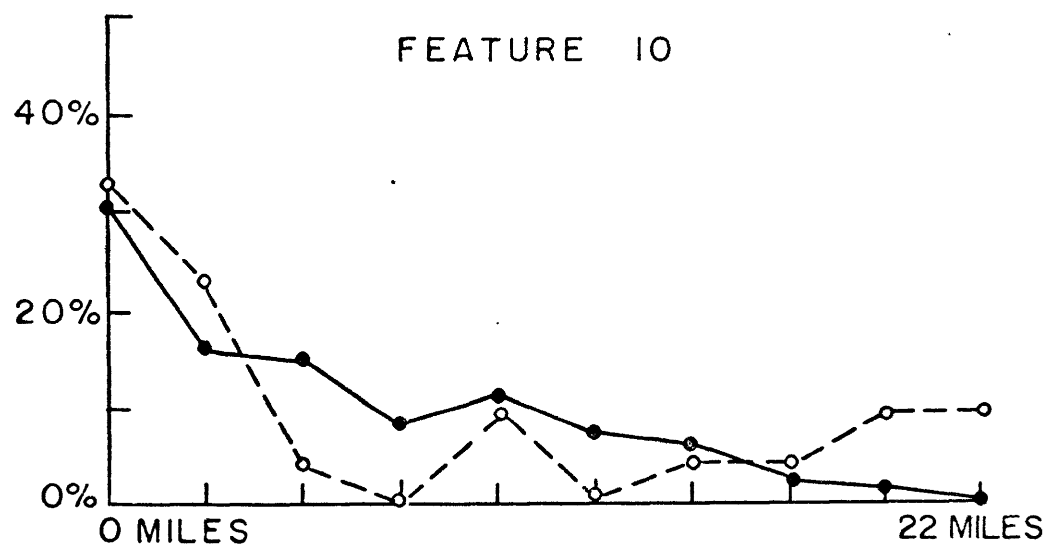

FEATURE 5
40%
20%
0%
0 MILES
34 MILES

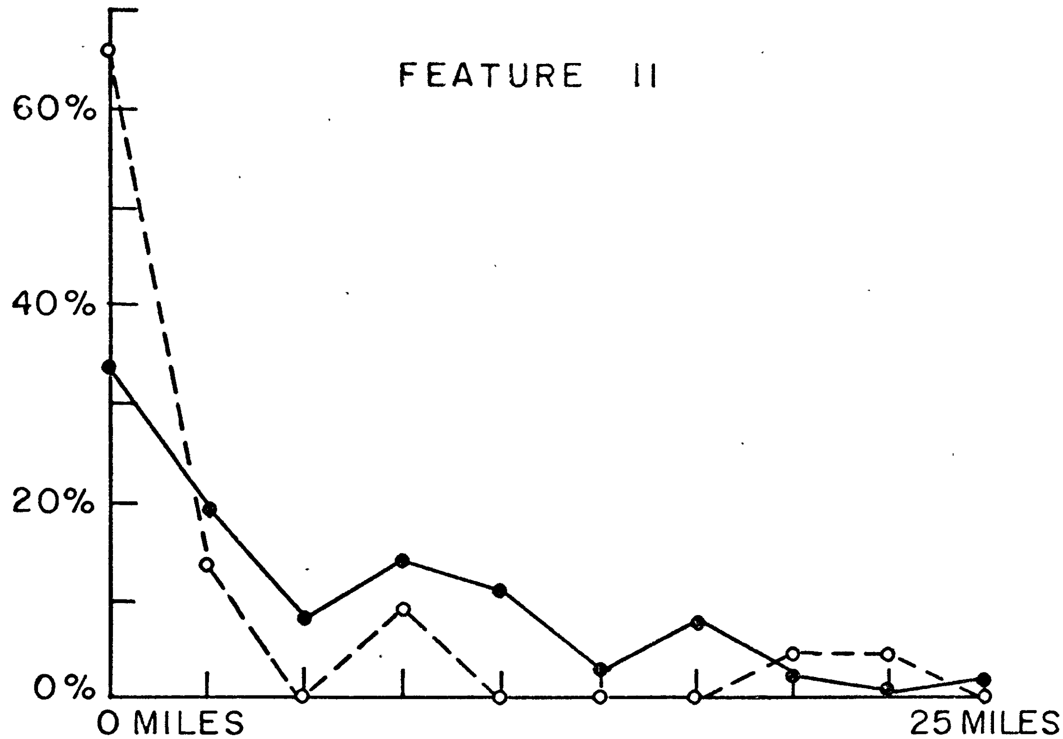

FEATURE 6
40%
20%
0%
0 MILES
30 MILES

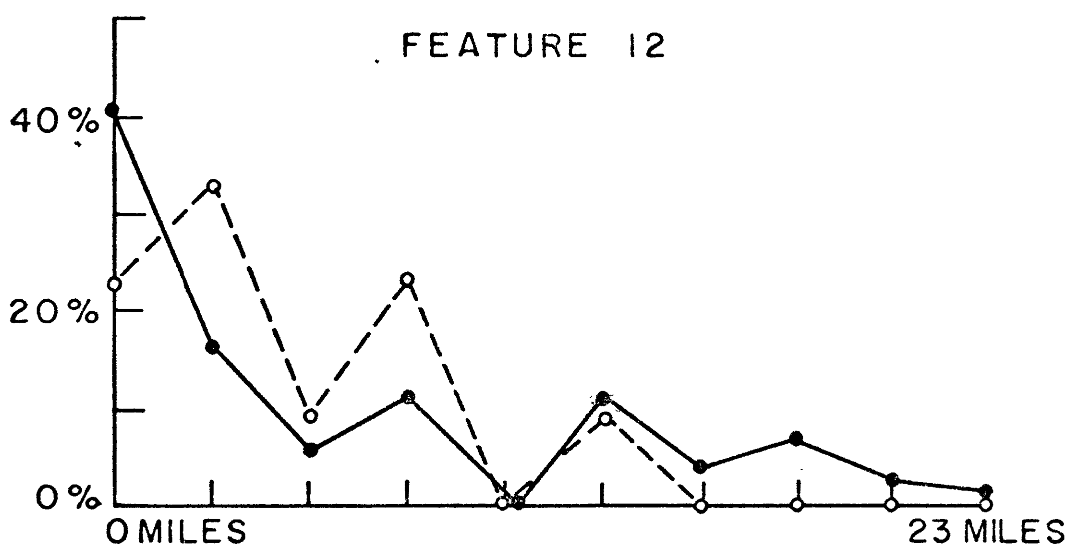

FEATURE 7
40%
20%
0%
0 MILES
14 MILES

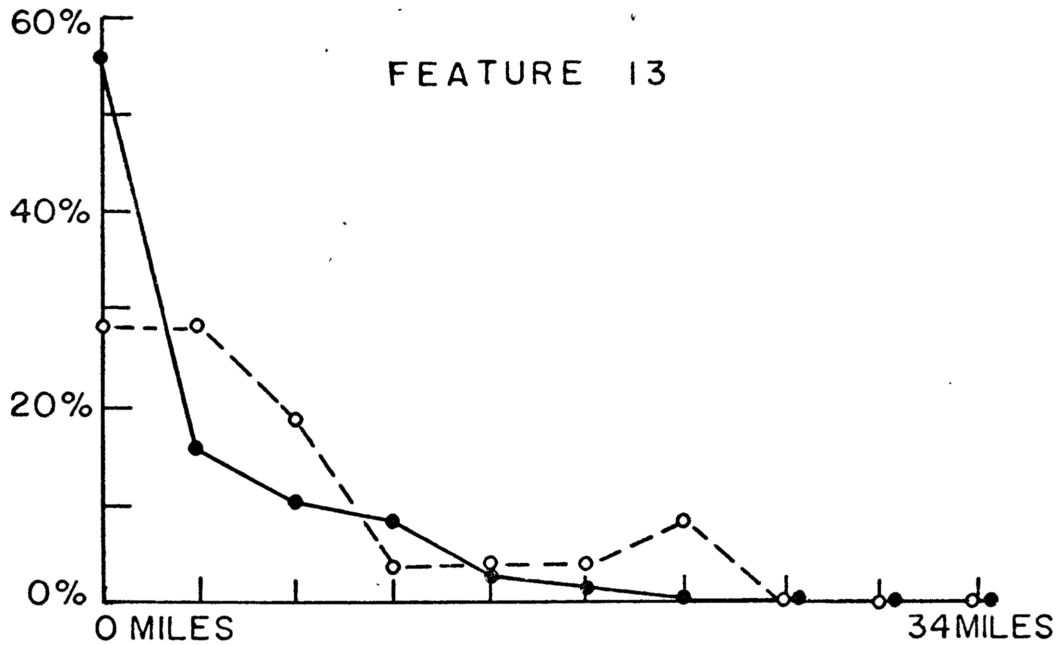

FEATURE 8
60%
40%
20%
0%
0 MILES
32 MILES

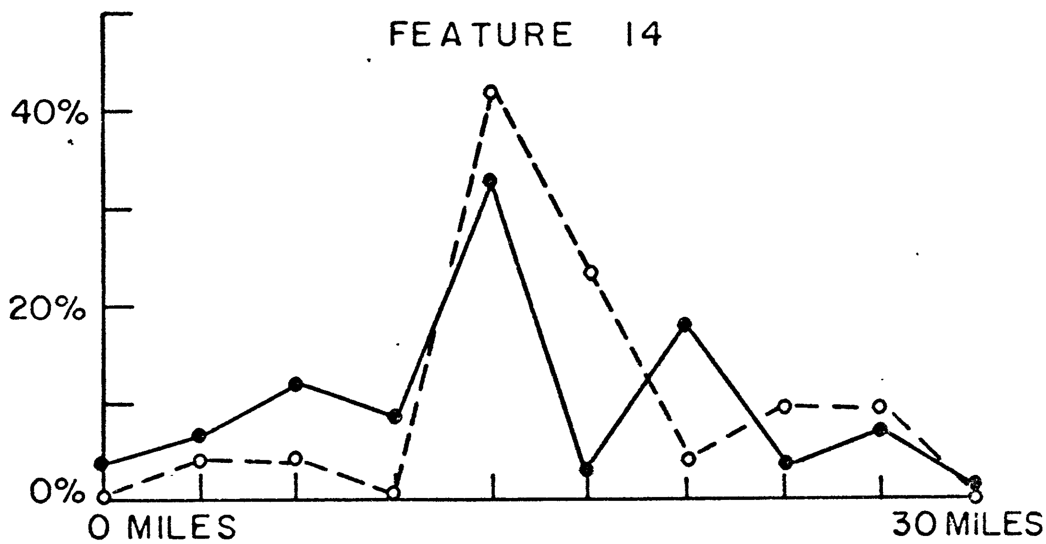

FEATURE 9
40%
20%
0%
0 MILES
32 MILES

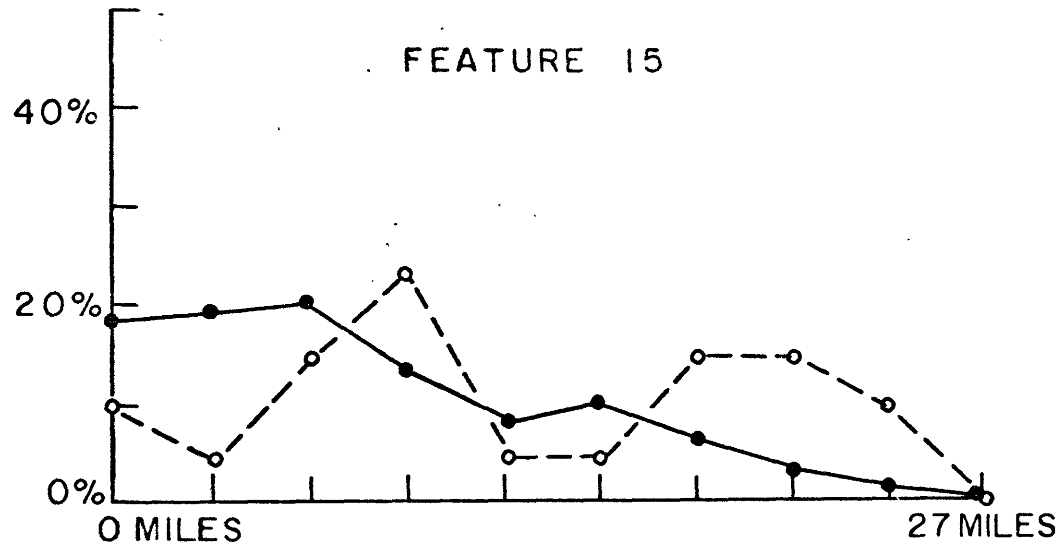

FEATURE 10
40%
20%
0%
0 MILES
22 MILES

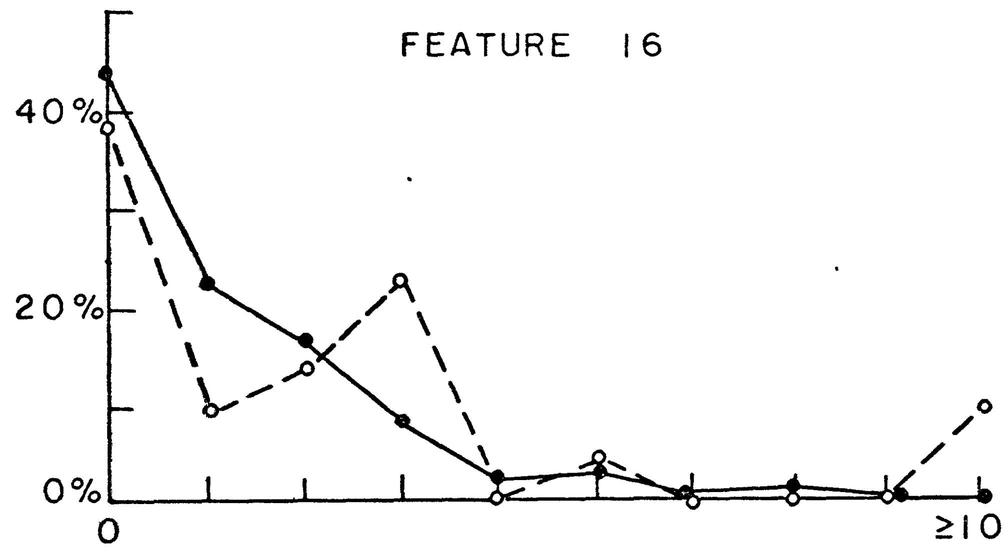

FEATURE 11
60%
40%
20%
0%
0 MILES
25 MILES

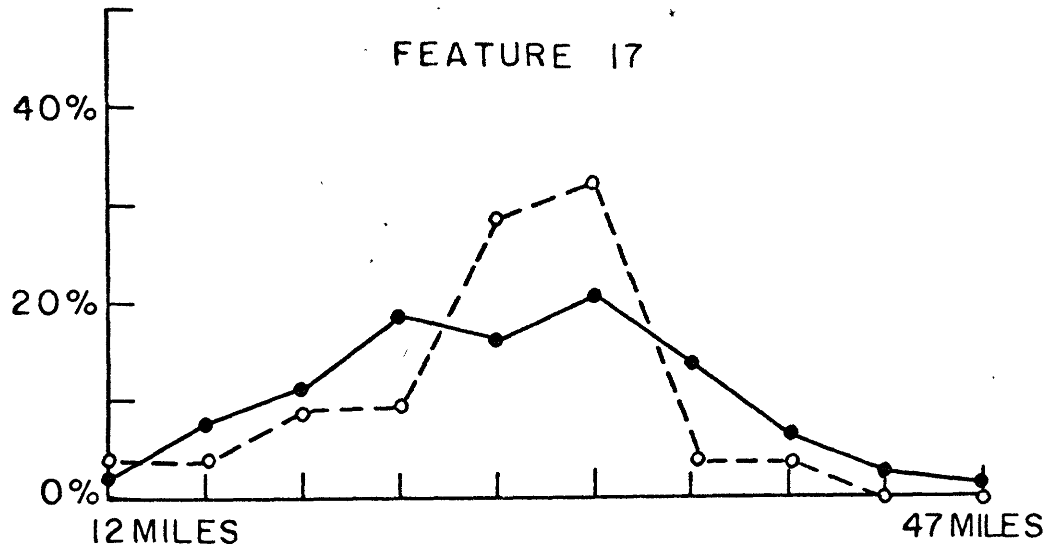

FEATURE 12
40%
20%
0%
0 MILES
23 MILES

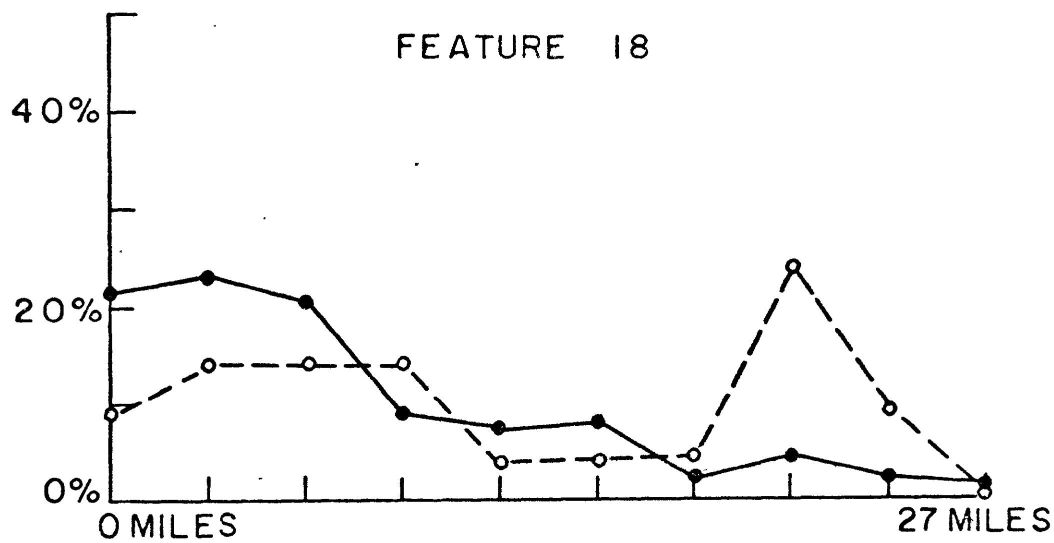

60%
40%
20%
0%
FEATURE 13
0 MILES
34 MILES

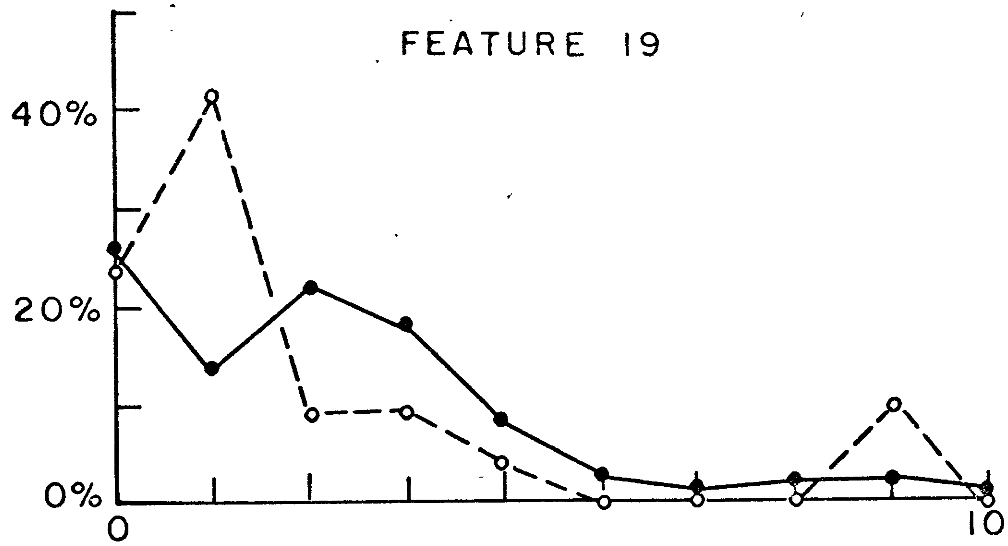

40%
20%
0%
FEATURE 14
0 MILES
30 MILES

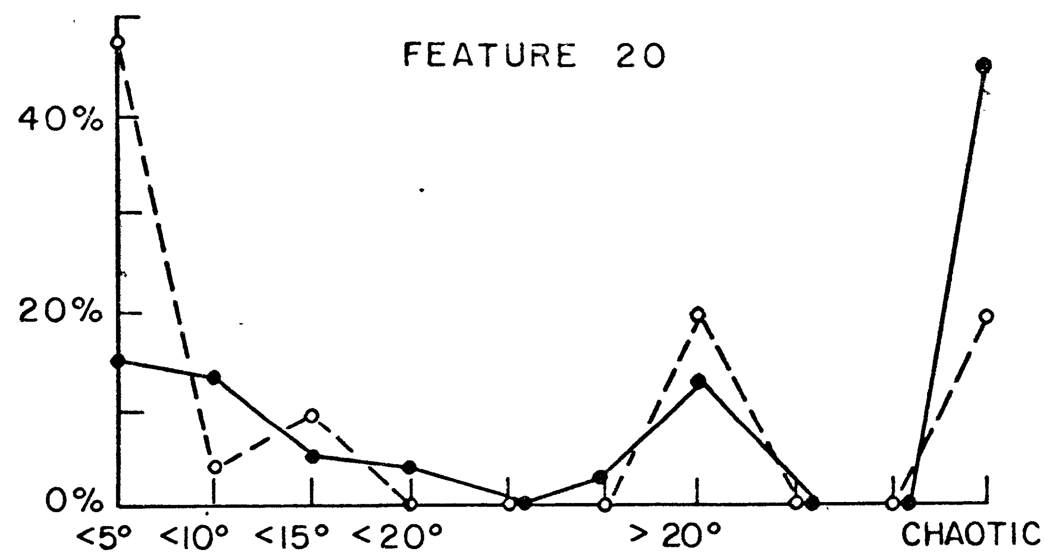

FEATURE 15
40%
20%
0%
0 MILES
27 MILES

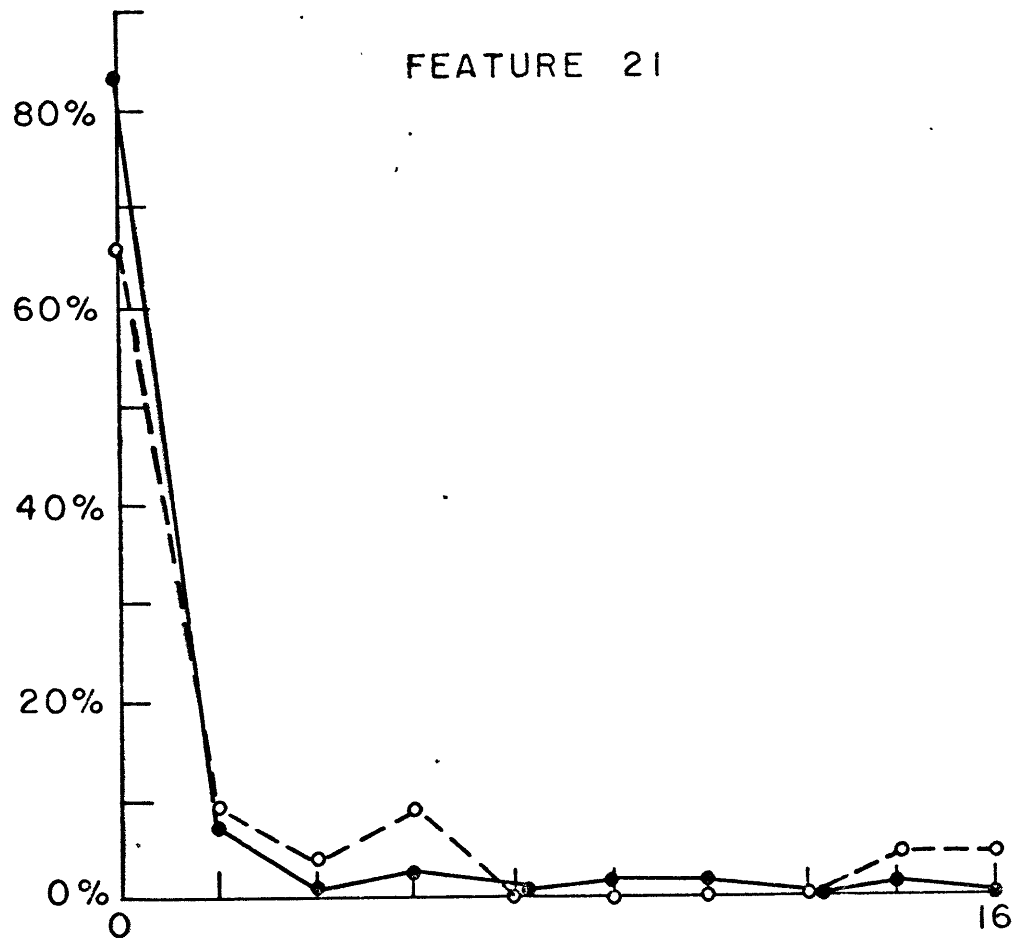

FEATURE 16
40%
20%
0%
0
≥10

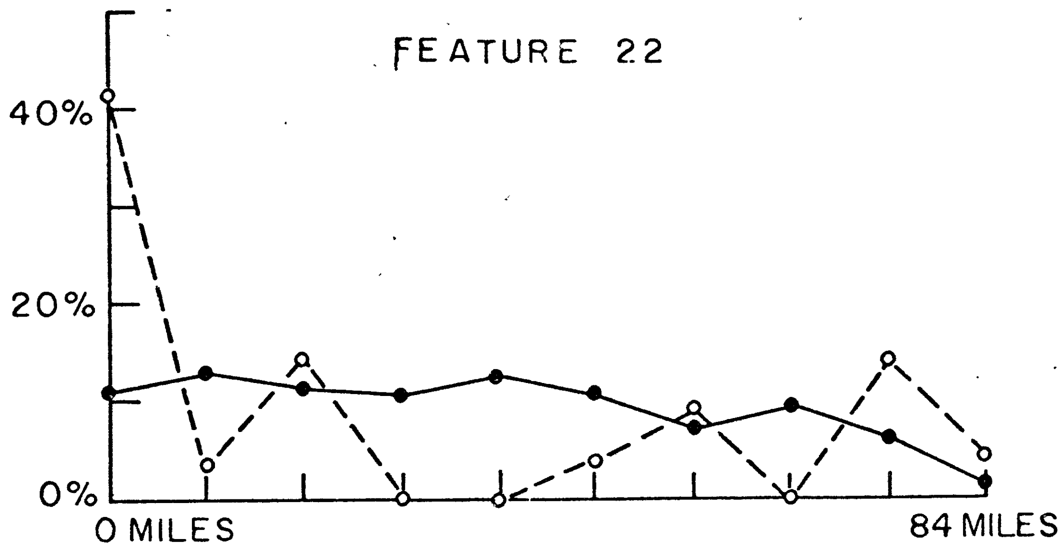

FEATURE 17
40%
20%
0%
12 MILES
47 MILES

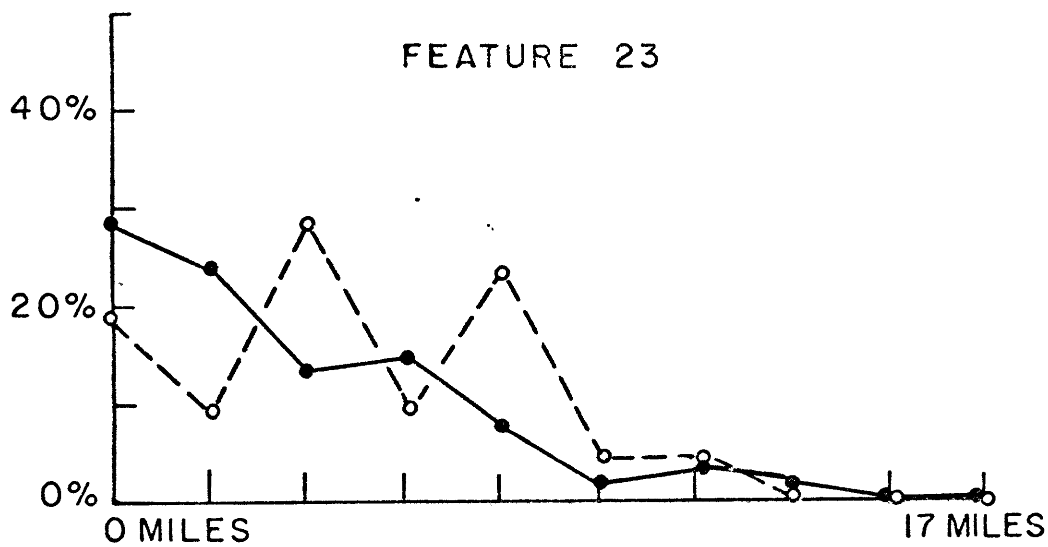

FEATURE 18
40%
20%
0%
0 MILES
27 MILES

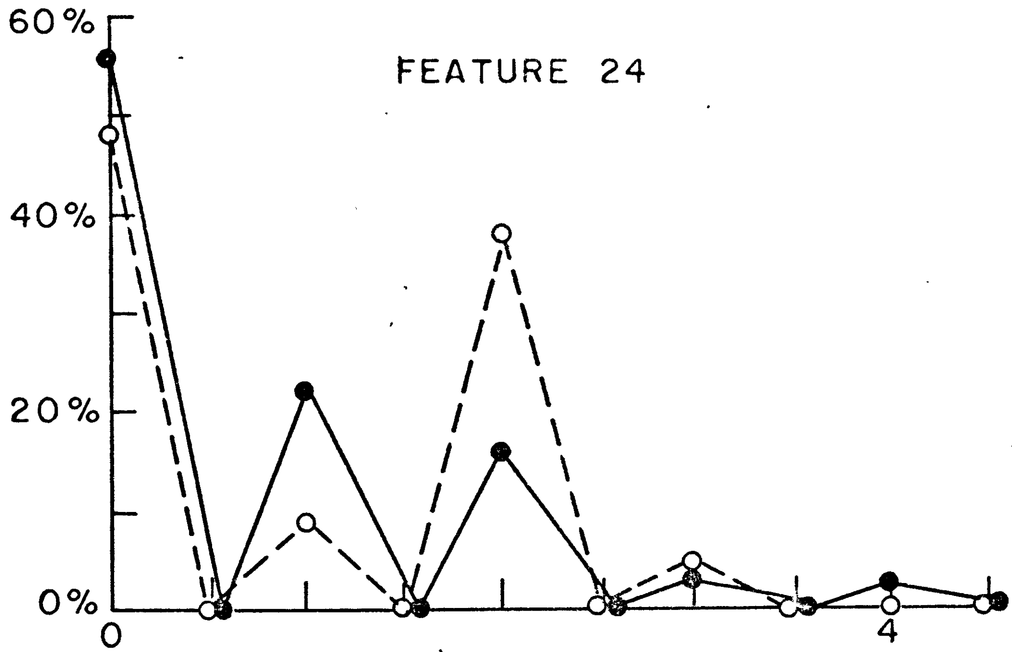

FEATURE 19
40%
20%
0%
0
10

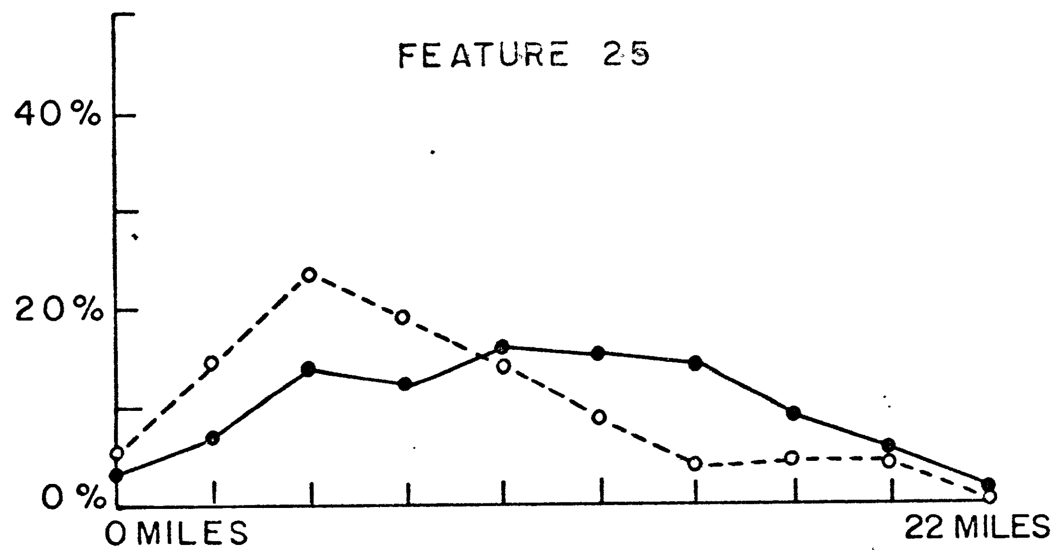

FEATURE 20
40%
20%
0%
<5°
<10°
<15°
<20°
> 20°
CHAOTIC

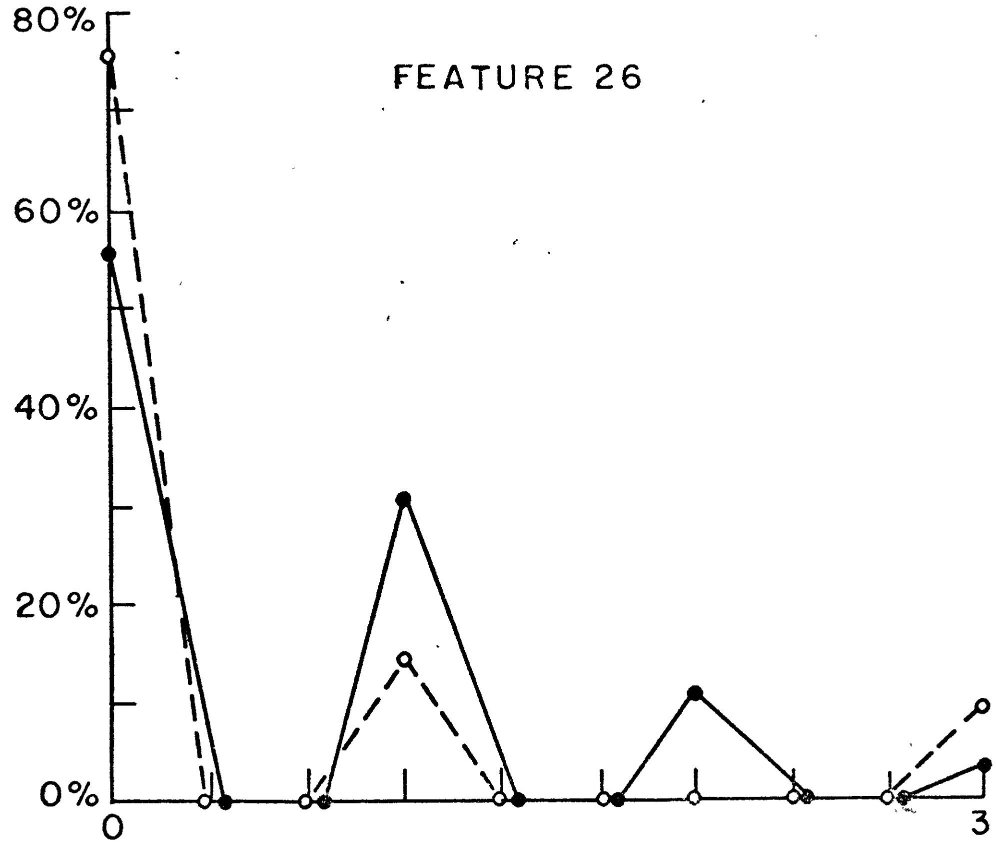

FEATURE 21
80%
60%
40%
20%
0%
0
16

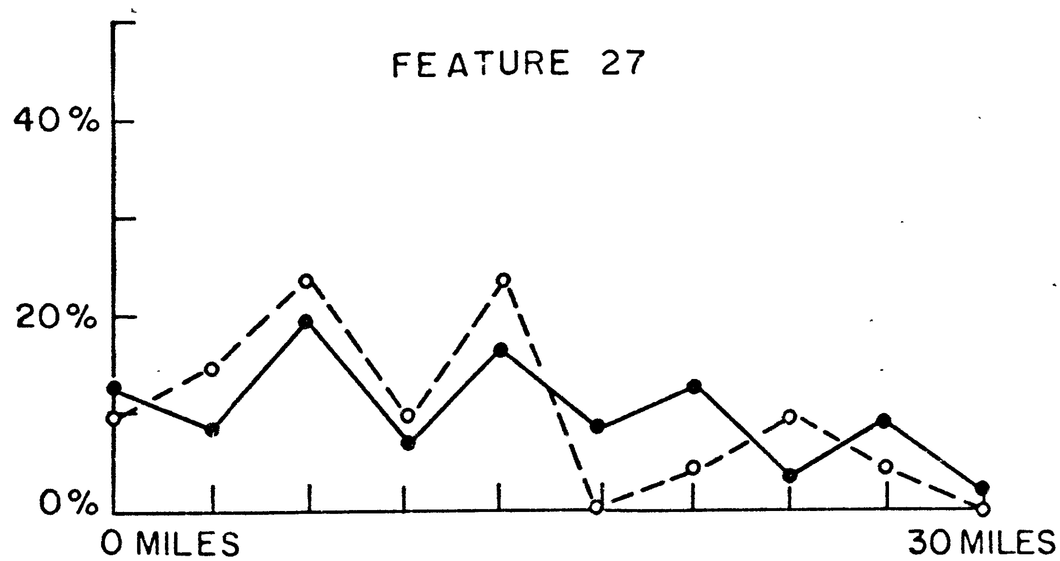

FEATURE 22
40%
20%
0%
0 MILES
84 MILES

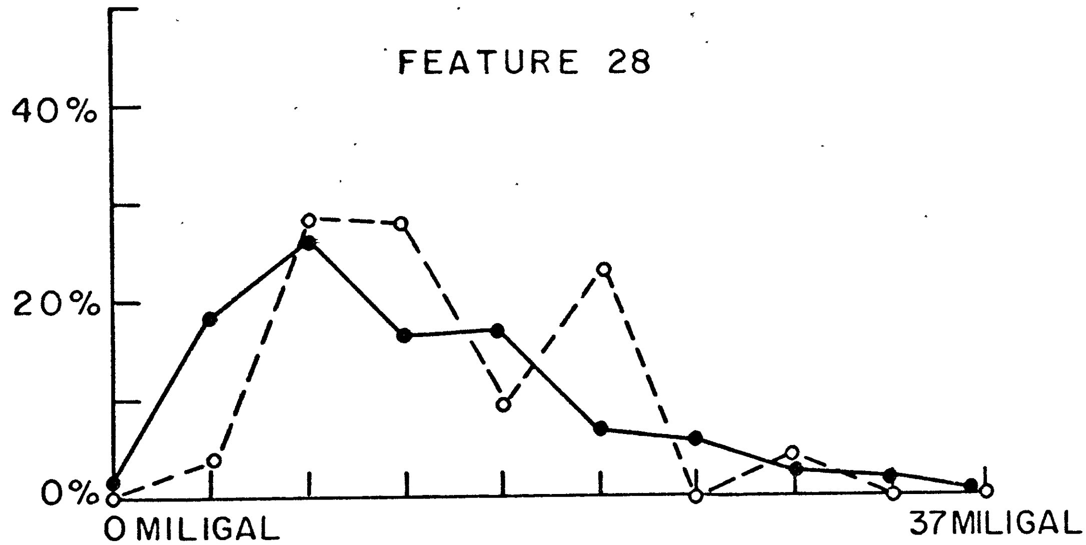

FEATURE 23
40%
20%
0%
0 MILES
17 MILES

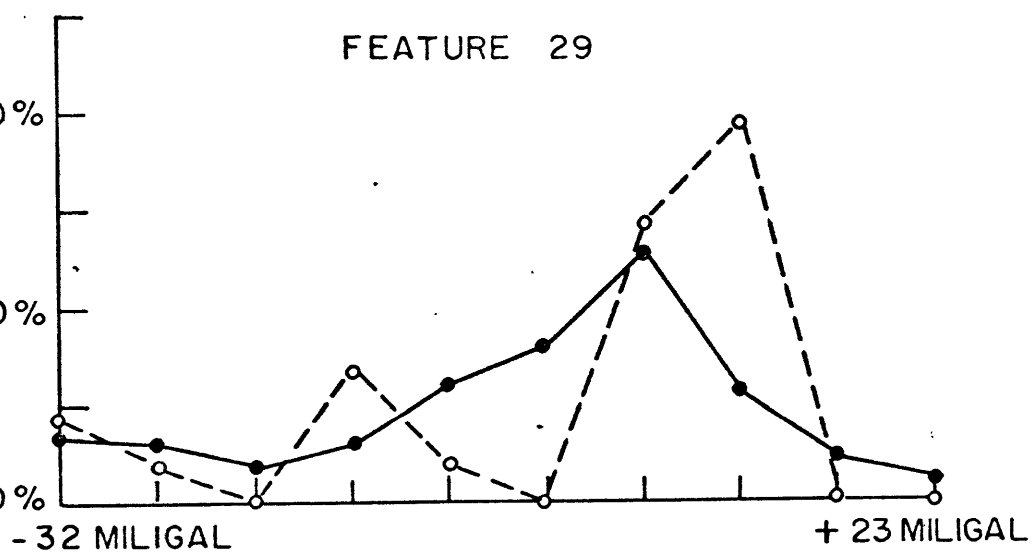

60%
FEATURE 24
40%
20%
0%
0
4

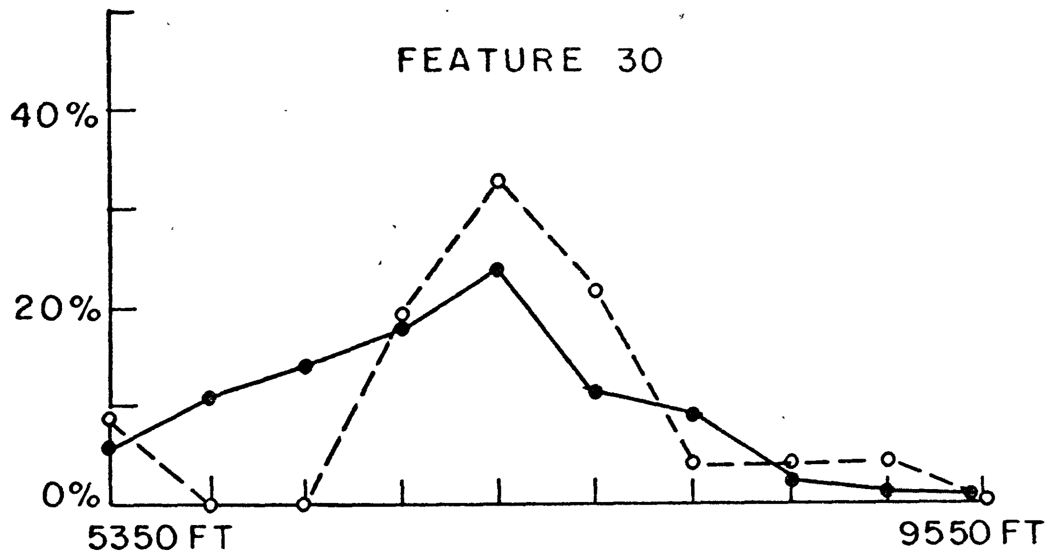

FEATURE 25
40%
20%
0%
0 MILES
22 MILES

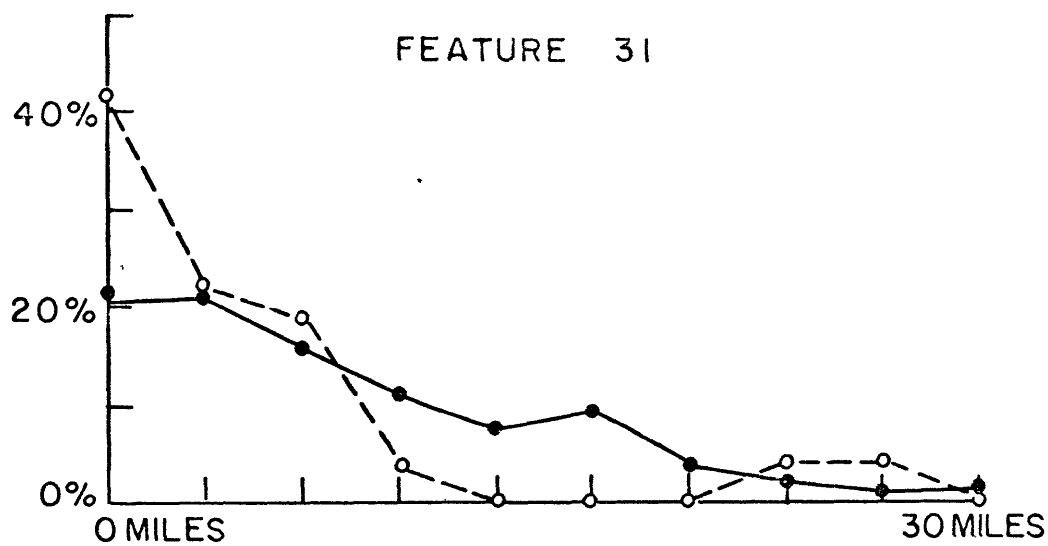

FEATURE 26
80%
60%
40%
20%
0%
0
3

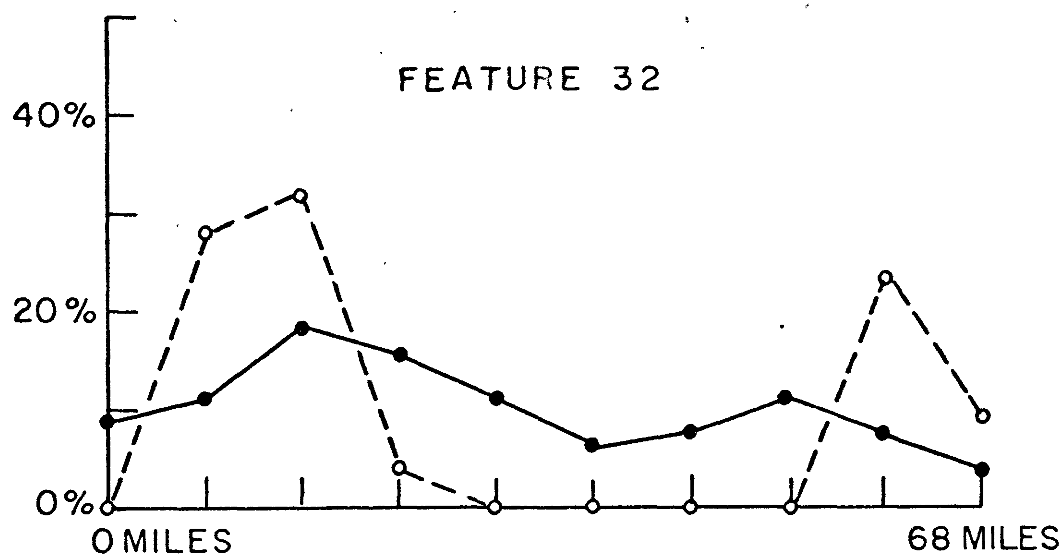

FEATURE 27
40%
20%
0%
0 MILES
30 MILES

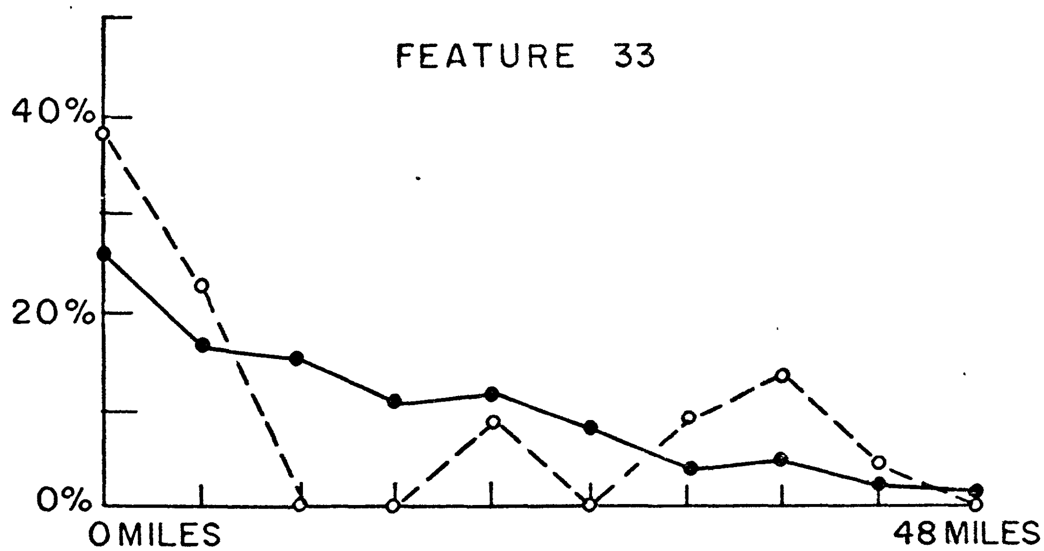

FEATURE 28
40%
20%
0%
0 MILIGAL
37 MILIGAL

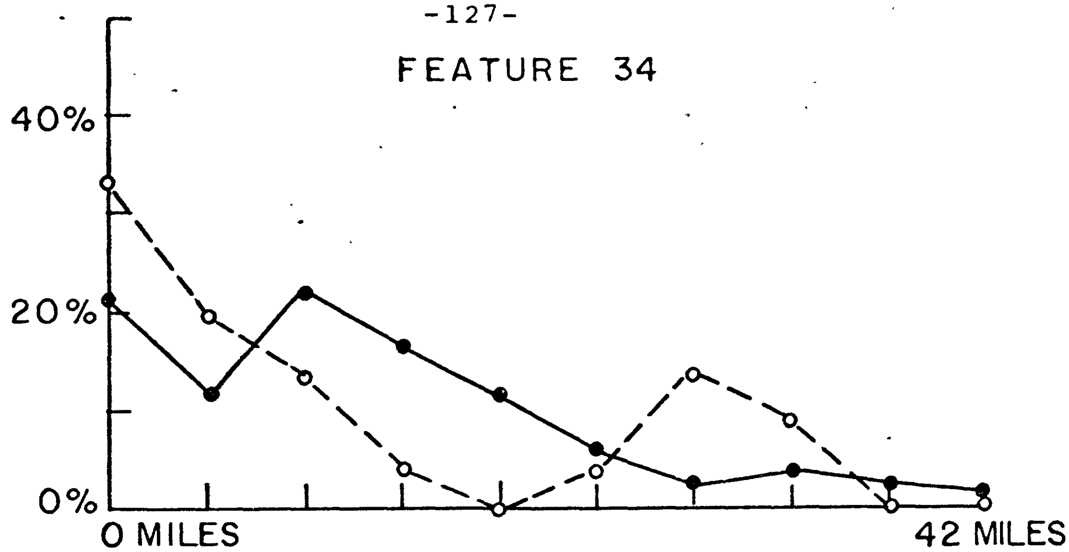

FEATURE 29
40%
20%
0%
-32 MILIGAL
+ 23 MILIGAL

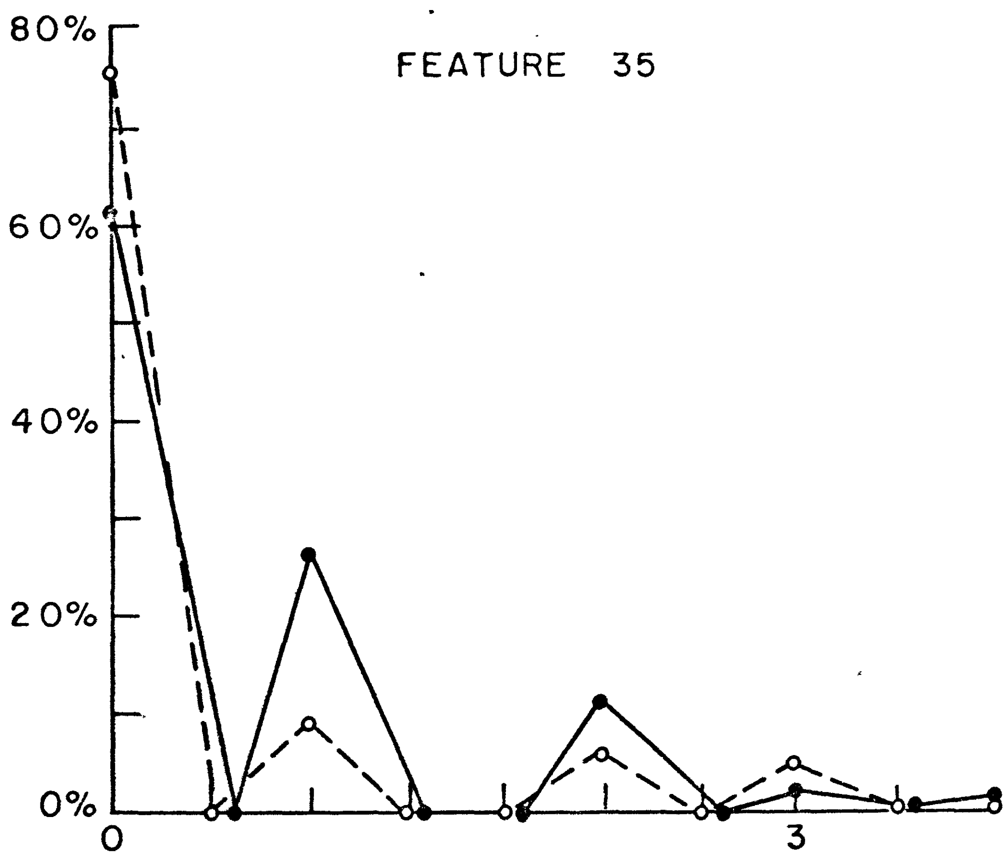

FEATURE 30
40%
20%
0%
5350 FT
9550 FT

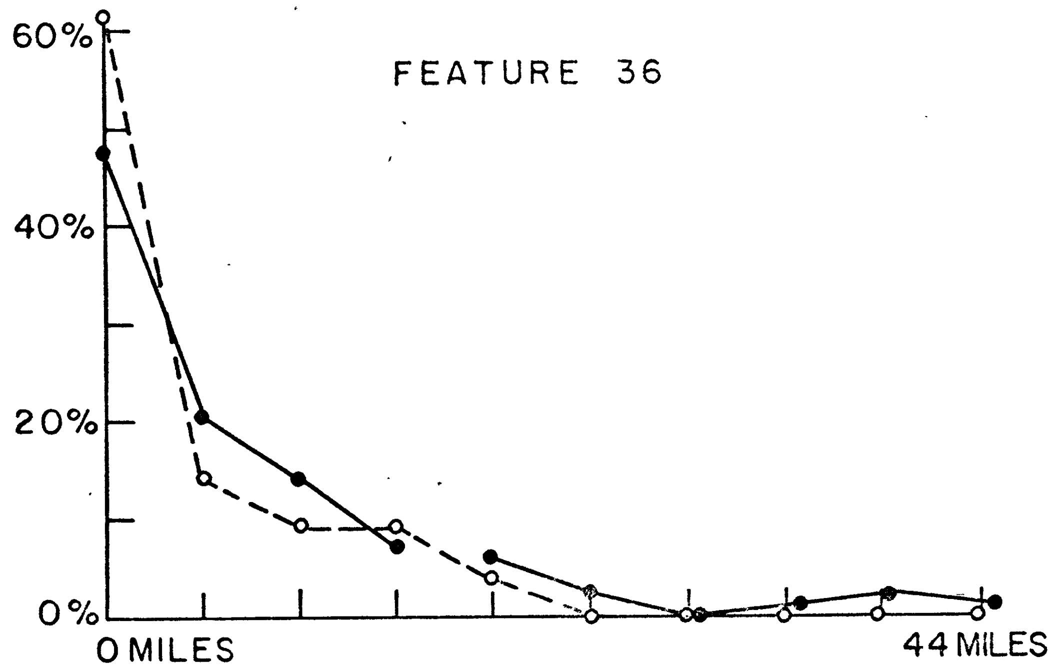

FEATURE 31
40%
20%
0%
0 MILES
30 MILES


FEATURE 32
40%
20%
0%
0 MILES
68 MILES


FEATURE 33
40%
20%
0%
0 MILES
48 MILES


FEATURE 34
40%
20%
0%
0 MILES
42 MILES


80%
FEATURE 35
60%
40%
20%
0%
0
3


FEATURE 36
60%
40%
20%
0%
0 MILES
44 MILES

CHAPTER 4

This chapter describes the four pattern recognition algorithms used here to quantify exploration evaluations.

Previous workers have made some efforts to suppliment subjective methods of prospecting and resource estimates with quantitative and semi-quantitative techniques. Organized lists of geologic features that are often associated with uranium deposits are a first step toward reproducible estimation of resource potential. With such lists as those proposed by Fisher (1974), McKay (1955), and Grutt (1972), various subjective predictions could at least proceed using the same criteria. Semi-quantitative prediction of uranium potential using a few, arbitrarily weighted geological characteristics, has been reported by Wier (1952). Harris (1966) has used factors hypothesized to be related to ore occurrence to estimate the probability that an area contains mineral resources of a certain worth. Collyer and Merriam (1973) used cluster analysis of subjectively weighted characteristics to group several tungsten mining areas into families. Such procedures often require more subjective input than the pattern recognition techniques used here, and sometimes are not easily tailored to suit available data. Application of more objective pattern recognition techniques to regional geology has been reported by Gelfand _et al_. (1976) for earthquake prediction problems.

Four algorithms are used here to recognize patterns in geological data that are indicative of uranium favorability. Three of these algorithms operate with binary-coded features, the fourth uses continuous valued features in a vector space treatment. Although one expects most rational decision processes to have much in common, these four algorithms have important differences and, accordingly, each has its idiosyncratic strengths and weaknesses.

4.1 Recognition with Individual Binary-Coded Features

The first algorithm, based on Bayes' rule, is perhaps the simplest. Each feature gives a yes/no answer about the geological patterns in a 2-class classification problem. Features are coded in binary form as discussed in Chapter 3, and with f features, the objects for recognition are described by f-dimensional data vectors with binary valued components,

$$X = (x_1, x_2, \ldots \ldots \ldots x_f).$$

A likelihood ratio is used to decide whether a data vector X represents a uranium-bearing or a barren area.

If $p_k = \text{Prob}(x_k = 1|U)$ and $q_k = \text{Prob}(x_k = 1|U^*)$, and if the features are conditionally independent, then $P(X|U)$ and $P(X|U^*)$ can be computed from the product of the probabilities of the components of X, the x_i:

$$P(X|U) = \prod_{i=1}^{f} p_i^{x_i}(1 - p_i)^{1-x_i}$$

and

$$P(X|U^*) = \prod_{i=1}^{f} q_i^{x_i}(1 - q_i)^{1-x_i}$$

Forming the likelihood ratio, $\frac{P(X|U)}{P(X|U^*)}$, yields the following discriminant function for classification:

$$G(X) = \sum_{i=1}^{f} [x_i \log\frac{p_i}{q_i} + (1-x_i)\log\frac{(1-p_i)}{(1-q_i)}] + \log\frac{P(U)}{P(U^*)}$$

which is of the form

$$G(X) = \sum_{i=1}^{f} w_i x_i + w_o \ .$$

Each binary feature is weighted according to how well it separates the two classes, U and U*, of a training population. The weight of the i-th feature is

$$w_i = \log \frac{p_i(1 - q_i)}{q_i(1 - p_i)} \ .$$

Features most often associated with U areas receive a positive weight; features characteristic of U* areas receive a negative weight (feature weights generated by this algorithm for the complete study area are given in Tables 3-1, 3-2, and 3-3). The weights of features occurring at each object are summed to give a classification rank; the more indicators of uranium an object has, the more positive its rank. The rank of each object is then compared to the threshold, w_o. If an object's rank exceeds the threshold

$$w_o = \sum_{i=1}^{f} \log \frac{1 - p_i}{1 - q_i} + \log \frac{P(U)}{P(U^*)}$$

that object is classed as U; otherwise, it is classed as U*.

Note that the weight given each feature depends on the significance of a "yes" answer for that feature. If $p_i = q_i$, the i-th feature gives no information about the state of nature, U or U*, and its weight is zero. The decision threshold, w_o, depends on the a priori probabilities of U and U*, and biases the decision in favor of whichever is more likely, here, U*. This decision algorithm is described more fully by Duda and Hart (1973).

The features used here are not independent, but are treated as though they were. Strongly correlated features effectively add two weighted votes for two pieces of information which, though perhaps geologically independent, occur often in combination with one another. Experiments were run with subsets of the full compliment of features to investigate the effect of this treatment. In general, the error rate of the classifier increases as features are removed from consideration. When one member from each pair of strongly correlated features is removed from recognition, the error rate of this classifier also increases. In general, it appears preferable with this classifier to include slightly redundant information rather than to discard features from recognition (see Chapter 6).

4.2 Combinations of Binary Features

A second approach to recognition attempts to seek out important contextual relationships between data. Features are binary coded as before, but all are weighted equally. An algorithm adapted from M.M. Bongard (Bongard *et al.*, 1966; Briggs and Press, 1977) looks for particular combinations of binary values of particular features that are especially characteristic of U or of U* objects.

Given sample populations of U and U* objects, the Bongard procedure searches all the different combinations of binary values of every different combination of 1, 2, and 3 features taken together to find some combinations of values that are particularly indicative of an object's membership in one class or another. These sets of characteristic feature values are called "traits." Traits may be formed of single features, F_i, with two possible binary values:

	F_i
Trait 1	1
Trait 2	0

or of one of the four possible combinations of values of two features, F_i and F_j:

	F_i	F_j
Trait 1	0	0
Trait 2	0	1
Trait 3	1	0
Trait 4	1	1

or of one of eight possible combinations of values of three features, F_i, F_j, F_k, in combination:

	F_i	F_j	F_k
Trait 1	0	0	0
Trait 2	0	0	1
. . .		. . . .	
Trait 8	1	1	1

Each of these possible traits may be characteristic of U or of U* objects if the particular combination of feature values occurs often in one group of objects and seldom in the other group of objects. For example, close proximity to an intrusive body and a thickening of the Morrison Formation may be characteristic of uranium-producing areas on the Colorado Plateau. It is required that traits characteristic of U objects occur at more than K_1 U objects and at fewer than K_1* U* objects in the U and U* training populations. Similarly, traits characteristic of U* objects must occur at more than K_2 U* objects and at fewer than K_2* objects of the U training sample. K_1, K_1*, K_2, and K_2* are adjusted to pick a manageable number of traits (usually 10 to 20) to characterize each group. There is no restriction on how often an individual feature need or may be a component of the traits selected. All traits are weighted equally; each U trait in an object's data vector contributes +1 vote for that object, each U* trait contributes a -1 vote for that object. Having found characteristic traits, the classifier then determines from each object's data vector the number of U and U* traits that occur at the object. The algebraic sum of the U and U* votes for an object is compared to a decision threshold (generally near zero) determined from recognition on the training sample. A net vote in excess of the threshold classifies an object as U.

The originators of this algorithm note the following relationships among the traits. Consider two traits of class C, T_A and T_B. If T_A and T_B should occur at exactly the same sets of objects of class C, then they are equivalent traits, and one is arbitrarily designated independent, the other dependent. If there is at least one object of class C at which T_A occurs but T_B does not, then T_A is independent of T_B. If the set of class C objects at which T_B occurs is a proper subset of the class C objects at which trait T_A occurs, then T_B is dependent on T_A. Recognition may be performed with independent traits only, or with all traits.

The Bongard algorithm is designed to pick out synergistic combinations of geologic features that may be more strongly indicative of class membership than the occurrence of the individual features used in the previous algorithm. No geologist could keep in mind the thousands of possible traits that this algorithm reviews ($2C_1^f$ + $4C_2^f$ + $8C_3^f$, where C_m^f is the number of combinations of f features taken m at a time - 9920 traits for 20 features, 34,280 traits for 30 features, etc.). Because this algorithm examines all singlet, doublet, and triplet traits, new insights into geological processes may emerge from interpretation of the particular combinations of features that are picked as traits. There is no provision for weighting the traits or features according to their hypothesized or apparent significance. The algorithm is free, however, to select individual features for use in traits as often as warrented subject to the K/K* conditions.

4.3 Unsupervised Hierarchical Clustering with Binary Features

A third approach to recognition involves unsupervised learning and clustering. An unsupervised learning technique is particularly useful in geological problems where the most natural grouping of objects into classes is sought. Here, for example, clustering experiments suggested that a grouping of Colorado Plateau U objects according to host bed ages was preferable to a grouping according to the size of deposits. Features are again taken in binary coded form, unweighted to form dendrograms of inter-object similarity.

The dendrogram is a tree-like diagram showing similarity relations among a group of objects. Individual objects form the leaves of the tree; all objects are joined in a single root of the tree. The intervening branches join one another at different levels between the leaves and root, graphically expressing the mutual similarity of objects on the branches. There is one and only one path joining any two objects. Two similar objects are connected by branches that join together in a node near the leaves; two relatively dissimilar objects are joined by a path that passes through the trunk of the tree rather near its root. The Hamming distance is used to measure inter-object dissimilarity. With f-dimensional binary data vectors, the Hamming distance simply measures the similarity of two objects as the number of features, m, with the same binary answer in each of the two objects's data vectors; the dissimilarity is then f-m. For example, the two data vectors V_1 = (1,0,1,1,0,1,0,0) and V_2 = (1,0,1,1,1,1,1,1) are separated by a Hamming distance of 3.

The measure of distance between clusters A and B is the maximum of the dissimilarities between objects of A and objects of B,

$$D(A,B) = \max(D(a,b))$$

where objects a are in cluster A, and objects b are in cluster B. Clusters are grown one object at a time to maintain a minimum cluster diameter; the diameter of a cluster is the maximum dissimilarity between any two objects within that cluster,

$$\mathrm{Dia}(A) = \max d(a,a')$$

Objects found most similar to known U objects are suggested as new targets for prospecting. The designation of new prospecting targets may proceed in two ways. Given a dendrogram with some known U objects as leaves, the classification of unfamiliar objects as U may proceed from the leaves of the dendrogram toward the root to include all the objects on those limbs that have higher-than-expected numbers of U objects above where the limb joins the body of the tree. Given a dendrogram, one may also consider using a nearest-neighbor or N-nearest neighbor rule for classification.

4.4 Minimum Distance Classifier

The fourth classification technique is a vector space minimum distance classifier (Young and Van Otterloo, to be published). Data are used in original form, not binary-coded, so that quantization noise is eliminated. Only features that are continuous measures with monomodal PDF's are used in recognition. Only monomodal features are considered so that the number of U objects in a cluster will not drop to a statistically unacceptable level.

Again, two classes of training objects are specified, U and U*. Data vectors for these samples are used to estimate means, variances, and covariances of the features for the clusters U and U*. The dissimilarity between an object and a cluster or class is taken as proportional to the distance from the object to the class mean. Because the original features are neither normalized to be compatible with one another, nor orthogonal, a coordinate system that is both scaled and rotated compared to the original input features is established. Distance, or dissimilarity, is measured in this new feature space, not in units of the individual features, but in terms of their standard deviations. The squared distance from an object x_i to the K-th class mean, μ_k, is given by:

$$D^2(x_i,\mu_k) = C(x_i - \mu_k)^t S^{-1}(x_i - \mu_k)$$

where C is a constant that may be dropped, and S^{-1} is the inverse covariance matrix of the features. An unfamiliar object is then classed as a member of the cluster to which it is closer in the transformed feature space.

This system makes use of a more formal definition of similarity/dissimilarity involving a distance metric in the transformed data space. The geological significance of the new coordinates, which are linear combinations of the original features, is obscured, however.

CHAPTER 5

This chapter presents the results of applying four pattern recognition algorithms to the three feature data bases discussed in Chapter 3. In these experiments, all available features were used in recognition and the class membership of each object was taken as known a priori in developing classifiers. Recognition experiments using the identities of all objects indicate whether or not the known deposits will cluster at all, and if so, how well the known deposits can be distinguished from barren areas given the data at hand. Classification results proved to be the best when the largest possible training samples were used and, with few exceptions, when all available features were used in recognition. These experiments correspond to a prospecting mode in which some deposits have been located in a metallogenic province and, based on their characteristics, one wishes to predict the locations of all similar deposits within the province.

5.1 Recognition with Individual Features

Casper Quadrangle:

All 21 points near known uranium deposits, and 303 points presumed barren were used to train the linear discriminant classifier. These points provided estimates of the probability of binary feature values for U and U* classes. Because there is little evidence that they mark significant uranium deposits, the 56 points near uranium prospects and occurrences were withheld from the training, but were evaluated by the classifier.

Points classed as U are indicated in Figure 5-1. Sixteen of 21 producing areas conform to a regional pattern and are classed as U. Two isolated U points in the northeast, and one in the south central part of the Quadrangle are not correctly recognized as U (rows 1 and 2, column 22, and row 13, column 15). These do not conform to the regional pattern which is set by the mining areas of greater areal extent, Shirley Basin, Gas Hills, and Crooks Gap. In all the recognition experiments for the Casper Quadrangle, these errors in classification persist. The regional pattern is perhaps too strongly influenced by the larger mining areas, yet with only 21 samples of U objects, one cannot consider dropping many of these points and still hope to retain a fair idea of the relevant feature distribution functions.

Nine prospects, one occurrence, and ten U* points are also classed as U; these areas are outlined in Figure 5-1. Most of these twenty non-producing objects are

FIGURE 5-1: Recognition results for the Casper Quadrangle - linear discriminant algorithm trained on 21 producing points. Areas classified favorable for uranium are within the solid outlines (see legend Figure 3-2).

CASPER, WYOMING QUADRANGLE - URANIUM DEPOSITS

4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	2	1	1	2	4	4	4	4	4	3	2	4	4	4	4	4	4	4	4	5	5
4	4	2	4	1	1	1	1	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	5
2	4	4	2	1	1	2	4	4	4	4	4	4	4	3	4	4	2	3	4	4	4	4	4	4	5
4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4
4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	4	4	4	4	4	4	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	1	4	4	3	4	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	2	1	4	4	3	4	4	4	4	4	4	4	4	2	4	3	4	4	4	4	4	4	1	4	1
4	3	1	1	3	2	2	4	4	4	4	4	4	4	2	4	4	4	4	3	3	4	4	1	1	4
3	4	2	4	2	2	4	4	4	4	4	4	4	4	1	3	3	4	2	4	4	4	4	4	1	4
4	3	4	3	2	2	4	4	4	4	4	4	4	4	4	4	3	2	3	4	4	4	4	4	4	4
4	4	3	4	3	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	4	4	3	4	4	4

FIGURE 5-1

adjacent to known producing areas, suggesting productive extensions of these mining areas. Three areas in the central part of the Quadrangle, however, are not near any known production. The southernmost of these (rows 12 and 13, column 19) is a highly faulted area of relatively thin sediments on the flanks of exposed Tertiary volcanic rocks that might have provided uranium source material. The other two areas (rows 5 and 8, column 18) are near the southern tip of the Wind River Basin, near oil and gas fields. Perhaps chemical conditions related to the occurrence of hydrocarbons might favor uranium deposition in these areas.

Colorado Plateau Jurassic/Cretaceous Deposits:

Fifty-eight cells with uranium production from Jurassic or Cretaceous host beds were contrasted with 450 remaining cells to train the linear discriminant. Preliminary feature selection and recognition experiments indicated that cells with uranium production from Triassic and Tertiary host beds, and vein and breccia pipe deposits introduced no significant contamination into the U* class. The classifier, trained on 508 cells, was then used to re-evaluate all 508 cells to suggest new prospects.

Cells classified favorable for uranium are indicated in Figure 5-2. Forty of 58 cells with production from Jurassic or Cretaceous strata are correctly classified as U. One hundred six, or 24% of the remaining 450 cells, however, are also classed as U; of these 106, 24% have significant uranium production from host beds other than Jurassic or Cretaceous strata. This result, to some extent, indicates a similarity in the genesis of uranium deposits within Cretaceous, Jurassic, and Triassic strata. More importantly, a weakness of the classifier becomes apparent in this application. Except for one cell in the northeast, the 146 cells classed U form a contiguous area despite the fact that producing cells are widely distributed over the Plateau, and that about 1/3 of the Jurassic/Cretaceous U cells are quite far from the largest single spatial cluster of U cells. The linear discriminant with binary features tends to ignore subtle differences between adjacent land areas and is, therefore, prone to error. This behavior is in strong contrast to the performance of the minimum distance classifier, discussed below. The performance of the linear discriminant on the Colorado Plateau also contrasts with the results for Wyoming discussed above. In the Casper Quadrangle, with very few training samples, this classi-

FIGURE 5-2: Recognition results for the Colorado Plateau - linear discriminant trained on 58 Jurassic/Cretaceous producing cells. Areas classified favorable for uranium are within the solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-2

fier provided sensible, rather conservative predictions. With three times as many U objects available for training on the Plateau, a more complex feature coding and recognition procedure seems to be favored. Though there is a high apparent error rate on U* objects on the Colorado Plateau, results from this classifier are not unreasonable, as the area classed as U includes that part of the Plateau where the Morrison Formation is thickest and most often productive.

Colorado Plateau Triassic Deposits:

Forty-five cells with uranium deposits in Triassic strata were contrasted with the 463 remaining Plateau cells to train the linear discriminant. Once again, inclusion of Jurassic, Cretaceous, and vein-type deposits in the U* class did not affect feature selection or recognition.

Cells classed as favorable for uranium deposits in Triassic strata are indicated in Figure 5-3. Twenty-eight of 45 U objects are recognized as U, along with 60 U* objects. Of these 60 cells, 21 have significant uranium production from other than Triassic host beds; 39 cells without uranium production are classed as U. Although the number of misclassifications is smaller than in the case of Jurassic/Cretaceous deposits,

FIGURE 5-3: Recognition results for the Colorado Plateau - linear discriminant trained on 45 Triassic producing cells. Areas classified favorable for uranium are within the solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-3

recognition results from this algorithm are similar to those obtained for training on Jurassic/Cretaceous deposits. The cells classed as U form a contiguous area which exludes many deposits in Triassic strata that are far from the north-central area of the Plateau, where many of these deposits are clustered.

5.2 Recognition with the Bongard Algorithm

The Bongard algorithm yields recognition results that have some properties of both the linear discriminant discussed above and the minimum distance classifier discussed below. In general, localized clusters of deposits still heavily weight recognition results, but the area classed as favorable for uranium extends away from these clusters to include deposits far from such a central cluster. The error rate of this classifier is high, however, because U* objects intervening between correctly recognized U objects are likely to have binary-valued features more similar to features of surrounding U objects than to features of more distant U* objects. These U* objects will have more characteristic U traits than U* traits, and thus are likely to be misclassified as U.

Casper Quadrangle:

Binary-valued features of twenty-one U points were contrasted with features of 303 points presumed barren to develop characteristic traits of U and U* objects. All points, including prospects and occurrences were then evaluated with these traits. Results of recognition are shown in Figure 5-4. As with the linear discriminant, the isolated mining areas in the south-central and north-east parts of the Quadrangle are not recognized as U. Sixteen of 21 U points are correctly classified; 8 prospects, 2 occurrences, and 26 presumably barren points are also classed as U. Most of these are extensions of known mining areas. Three areas distant from mining activity are suggested as prospects. Two of these areas overlap predictions of the linear discriminant (row 5, column 18 and row 8, column 18). A third area classified favorable for uranium consists of six points in the south Wind River Basin in an area of thick sediments and hydro-carbon accumulations (row 4, column 14 and adjacent points).

Colorado Plateau Jurassic/Cretaceous Deposits:

To produce the strongest contrast of traits for recognition and interpretation with the Bongard algorithm,

FIGURE 5-4: Recognition results for the Casper Quadrangle - Bongard algorithm trained on 21 producing points. Areas classified favorable for uranium are within the solid outlines (see legend Figure 3-2).

CASPER, WYOMING QUADRANGLE - URANIUM DEPOSITS

4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	2	1	1	2	4	4	4	4	4	3	2	4	4	4	4	4	4	4	4	5	5
4	4	2	4	1	1	1	1	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	5
2	4	4	2	1	1	2	4	4	4	4	4	4	4	3	4	4	2	3	4	4	4	4	4	4	5
4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4
4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	4	4	4	4	4	4	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	1	4	4	3	4	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	2	1	4	4	3	4	4	4	4	4	4	4	4	2	4	3	4	4	4	4	4	4	1	4	1
4	3	1	1	3	2	2	4	4	4	4	4	4	4	2	4	4	4	4	3	3	4	4	1	1	4
3	4	2	4	2	2	4	4	4	4	4	4	4	4	1	3	3	4	2	4	4	4	4	4	1	4
4	3	4	3	2	2	4	4	4	4	4	4	4	4	4	4	3	2	3	4	4	4	4	4	4	4
4	4	3	4	3	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	4	4	3	4	4	4

FIGURE 5-4

the binary features used in the previous algorithm were scanned to develop traits characteristic of the 55 cells with Jurassic/Cretaceous deposits and of 391 cells with no known uranium deposits; 59 cells with uranium production from other than Jurassic or Cretaceous host beds were not included in trait selection, but were later evaluated by the Bongard classifier.

Recognition with independent traits produced slightly fewer misclassifications of U* objects than did recognition with all traits. Results of this recognition are shown in Figure 5-5. Forty-four of 58 U objects are correctly recognized. Twenty cells with deposits in other host environments are recognized as U, and 80 of 391 barren cells are predicted to be favorable for uranium. These results are not unreasonable, but some large deposits have been discovered outside the favorable area; this suggests that the largest cluster of U objects has had an overly strong influence on recognition results. Because uranium deposits are rare geological objects, one might reasonably have more confidence in the predictions of a system that classified as U something less than the present 20% of U objects.

FIGURE 5-5: Recognition results for the Colorado Plateau - Bongard algorithm trained on 58 Jurassic/Cretaceous producing cells. Areas classified favorable for uranium are within the solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-5

Colorado Plateau Triassic Deposits:

To characterize Triassic deposits, 45 cells with uranium deposits in Triassic rocks were contrasted with 391 cells presumed barren. Known uranium deposits in other than Triassic host rock were withheld from the trait selection stage of classification.

The results of recognition using all traits is shown in Figure 5-6. Perhaps because Triassic deposits are more widely dispersed over the Plateau than Cretaceous deposits, the traits may have embodied more essential characteristics of these Triassic deposits, as they provided better discrimination between U and U* objects than in the Jurassic/Cretaceous case. Thirty-four of 45 U objects are correctly recognized. Eighteen other uranium producing cells and 48 non-producing cells are also classed as U. This error rate in U* objects is about half that of the Jurassic/Cretaceous case. Areas classed as U no longer form a contiguous area, but are split into two areas. In this case, the Bongard algorithm begins to approach the performance of an ideal classifier that could pick isolated productive zones out from surrounding barren areas. Interestingly, in both applications of the Bongard algorithm to the Colorado Plateau, producing cells of types different from the designated U objects

FIGURE 5-6: Recognition results for the Colorado Plateau - Bongard algorithm trained on 45 Triassic producing cells. Areas classified favorable for uranium are within the solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-6

are classed as U only when they lie between clusters of correctly recognized U objects. The Bongard algorithm tends to ignore producing cells different from those of the designated U class. In part this is due to differences in the features selected for recognition on Cretaceous and on Triassic deposits, and in part it reflects the fact that characteristic traits are developed to suit a particular type of deposit.

5.3 Recognition with the Minimum Distance Classifier

A treatment of recognition using a distance metric in a continuous-valued feature space provides the best classification results for the Colorado Plateau. In the Casper Quadrangle, however, its error rate is very high when all available features are used. When the number of samples of the smallest class is approximately equal to, or greater than, twice the number of features, the minimum distance classifier proved to be the best of the four recognition techniques used here. The Casper Quadrangle contrasts with the Colorado Plateau by proving an example of a feature space of high dimensionality that is under-populated with one class of objects (in this case, U objects).

Casper Quadrangle:

The small number of U objects, 21, is not sufficient to populate a 33-dimensional feature space (3 of 36 features are not suited to vector space treatment). The previous sections of this chapter suggest that at least 3 of these 21 U objects do not conform to a regional pattern of uranium deposits; this reduces the number of similar-appearing U objects to 18 or less. Performance is considerably improved when various subsets of the 33 features are used in recognition (see Chapter 6). The result of recognition with all features is shown in Figure 5-7 as an illustration of the danger of using a feature space of high dimensionality in recognition with too few training samples. With a very small number of training samples, performance was improved by use of binary-coded features that confine U and U* objects to the corners of a hypercube rather than continuous-valued features that disperse objects through the entire volume of a multidimensional feature space. When all available features are used in recognition, 15 of 21 U objects are correctly identified. Eight prospects and 8 occurrences are classed as U, and 83 of 303 U* objects are classed as U (Figure 5-7). The high error rate on U* objects, 27%, makes the predictions of this recognition analysis suspect.

FIGURE 5-7: Recognition results for the Casper Quadrangle - minimum distance classifier trained on 21 producing points. Areas classified favorable for uranium are within solid outlines (see legend Figure 3-2).

CASPER, WYOMING QUADRANGLE - URANIUM DEPOSITS

4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	2	1	1	2	4	4	4	4	4	3	2	4	4	4	4	4	4	4	4	5	5
4	4	2	4	1	1	1	1	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	5
2	4	4	2	1	1	2	4	4	4	4	4	4	4	3	4	4	2	3	4	4	4	4	4	4	5
4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4
4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	4	4	4	4	4	4	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	1	4	4	3	4	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	2	1	4	4	3	4	4	4	4	4	4	4	4	2	4	3	4	4	4	4	4	4	1	4	1
4	3	1	1	3	2	2	4	4	4	4	4	4	4	2	4	4	4	4	3	3	4	4	1	1	4
3	4	2	4	2	2	4	4	4	4	4	4	4	4	1	3	3	4	2	4	4	4	4	4	1	4
4	3	4	3	2	2	4	4	4	4	4	4	4	4	4	4	3	2	3	4	4	4	4	4	4	4
4	4	3	4	3	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	4	4	3	4	4	4

FIGURE 5-7

Colorado Plateau Jurassic/Cretaceous Deposits:

A two-class recognition problem was set up to classify Jurassic/Cretaceous deposits. Fifty-eight U objects were taken as a group and contrasted with the 450 remaining Plateau cells. Control recognition experiments with Triassic and other deposits withheld from training demonstrated that the introduction of these uranium-producing cells into the U* class did not adversely affect recognition.

The results of recognition with all available features is shown in Figure 5-8. Forty-seven of 58 U objects are correctly recognized; except for two isolated U objects which are not recognized, the classifier correctly recognizes some cells within every mining area on the Plateau without a large number of misclassifications of U* objects. Only 8 of the 391 non-producing cells are classed as U, all of them adjacent to producing cells. Two deposits in Triassic rocks are classed as U; all other objects are classed as U*. Although not all Jurassic/Cretaceous deposits conform to a regional pattern, the performance of this classifier is remarkable insofar as it can distinguish not only producing from barren cells, but also Jurassic/Cretaceous cells from neighboring cells with production from Triassic strata.

FIGURE 5-8: Recognition results for the Colorado Plateau - minimum distance classifier trained on 58 Jurassic/ Cretaceous producing cells. Areas classified favorable for uranium are within solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-8

Remarkably few U* objects are classified as U. This algorithm has combined features that are individually rather poor discriminants to form an effective classifier.

Colorado Plateau Triassic Deposits:

Forty-five cells with production from Triassic host rocks were opposed to all other cells in a two-class recognition problem. Inclusion of other types of producing cells in the U* class had no deleterious effects on recognition.

The results of recognition are shown in Figure 5-9. The number of Triassic cells available to form the U cluster in the present 17-dimensional feature space is only about 3/4 the size of the Jurassic/Cretaceous training population described above. This may explain why performance of the minimum distance classifier is slightly poorer here than in the Jurassic/Cretaceous case. From another point of view, this is a somewhat surprising result because the ratio of the number of U samples to the dimensionality of the feature space is higher in this recognition than in the Jurassic/Cretaceous case (2.6 and 1.8, respectively). Apparently the Triassic deposits are a more diverse family of objects

FIGURE 5-9: Recognition results from the Colorado Plateau - minimum distance classifier trained on 45 Triassic producing cells. Areas classified favorable for uranium are within solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-9

with respect to their feature space than are the Jurassic/Cretaceous. Another factor which could contribute to weakened recognition is a lack of redundant features which, though often considered a nuisance, appear useful in combined interpretation for bringing outliers into their proper cluster (see Chapter 6).

In Figure 5-9, 28 of 45 Triassic deposits are correctly recognized. As in the Jurassic/Cretaceous case, this classifier seldom misclassifies producing cells of other types as U objects; only 3 Jurassic/Cretaceous cells of the remaining 72 producing cells are designated U. In this case, the classifier correctly recognizes some cells in every area on the Plateau that has production from Triassic host rocks, except for 3 isolated producing cells. This widespread correct recognition is achieved without misclassification of U* cells intervening between areas of Triassic production. Only 10 non-producing cells are classed as U; all of these are adjacent to cells with production from Triassic strata (Figure 5-9).

5.4 Unsupervised Clustering with Binary Features

Clustering with binary features does not assume a priori any class identity for each object. Objects are grouped naturally into minimum diameter clusters to form a dendrogram. If any U* objects are very similar to U objects, they will be clustered with U objects, suggesting that these U* objects may hold undiscovered uranium deposits. With the dendrogram structure, it is never obvious where one should separate a branch from the larger tree and call it a cluster. Two reasonable, but ad hoc rules were followed in defining clusters.

With F features, two objects's feature vectors are expected to have F/2 identical components by chance; clusters were defined to include objects with at least 3/4 of their features identical. Also, a U cluster was not identified unless more than 3 U objects appeared in the cluster, and the fraction of U objects in the cluster exceeded three times the average frequency of occurrence of U objects in the population of all objects.

Casper Quadrangle:

The uranium-producing points within the Casper Quadrangle do not cluster together strongly enough to pass the tests outlined above. The clusters shown in Figure 5-10 were formed to include each of the 21 U points and all other points that had more than 2/3 of their binary feature values identical to one of these

FIGURE 5-10: Recognition results for the Casper Quadrangle-clustering algorithm. Clusters of points containing all known producing areas are shown by solid outlines (see legend Figure 3-2).

CASPER, WYOMING QUADRANGLE - URANIUM DEPOSITS

4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1	4	5	5	5
4	4	4	4	4	2	1	1	2	4	4	4	4	4	3	2	4	4	4	4	4	4	4	4	5	5
4	4	2	4	1	1	1	1	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	5
2	4	4	2	1	1	2	4	4	4	4	4	4	4	3	4	4	2	3	4	4	4	4	4	4	5
4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	2	2	2	4	4	4	4	4	4	4	4
4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	4	4	4	4	4	4	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	1	4	4	3	4	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	2	1	4	4	3	4	4	4	4	4	4	4	4	2	4	3	4	4	4	4	4	4	1	4	1
4	3	1	1	3	2	2	4	4	4	4	4	4	4	2	4	4	4	4	3	3	4	4	1	1	4
3	4	2	4	2	2	4	4	4	4	4	4	4	4	1	3	3	4	2	4	4	4	4	4	1	4
4	3	4	3	2	2	4	4	4	4	4	4	4	4	4	4	3	2	3	4	4	4	4	4	4	4
4	4	3	4	3	4	4	4	4	4	4	4	4	4	3	4	4	2	4	4	4	4	3	4	4	4

FIGURE 5-10

21 U points. Most predicted new deposits are extensions of existing mining areas.

Colorado Plateau Jurassic/Cretaceous Deposits:

Forty of 58 Jurassic/Cretaceous deposits fall in clusters conforming to the above rules. These clusters include 24 Triassic deposits and 48 non-producing cells. Cells classified as U form a single contiguous area shown in Figure 5-11.

Colorado Plateau Triassic Deposits:

Clustering with Triassic deposits results in two main clusters of cells shown in Figure 5-12. Thirty-four of 45 U objects are included in the clusters, along with 12 Jurassic/Cretaceous deposits and 34 non-producing cells.

<u>FIGURE 5-11</u>: Recognition results for the Colorado Plateau - clustering algorithm. Areas classified favorable for uranium in Jurassic/Cretaceous strata are within solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-11

FIGURE 5-12: Recognition results for the Colorado Plateau - clustering algorithm. Areas classified favorable for uranium in Triassic strata are within solid outlines (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 5-12

In all the foregoing recognition experiments with the four algorithms, all objects were classified as either U or U*. Performance in these recognition experiments is summarized in Table 5-1. Table 5-1 shows the numbers of objects of each type from each study area that are classified as U or U* by the four algorithms used here. For the Casper Quadrangle, recognition results are tabulated for four types of objects - known uranium resource areas; uranium prospects; uranium occurrences; and barren areas. For the Colorado Plateau, recognition results are tabulated for five types of objects - cells with uranium resources in Jurassic/Cretaceous strata; cells with uranium resources in Triassic strata; cells with uranium resources in Tertiary strata; cells with vein or breccia pipe deposits; and barren cells.

TABLE 5-1: SUMMARY OF PRINCIPAL RECOGNITION RESULTS

Classifier ↓ / Training: →	Casper Quadrangle	U	U*	Colorado Plateau Jurassic/Cretaceous	U	U*	Colorado Plateau Triassic	U	U*
Linear Decision Surface	Known U	16	5	J/C	40	18	J/C	20	38
	Prospects	9	21	TR	17	28	TR	28	17
	Occurrences	1	25	Tert	3	1	Tert	1	3
	Barren	10	293	Vein	1	9	Vein	0	10
				Barren	85	306	Barren	39	352
Bongard Algorithm	Known U	16	5	J/C	44	14	J/C	17	41
	Prospects	8	22	TR	16	29	TR	34	11
	Occurrences	2	24	Tert	4	0	Tert	0	4
	Barren	26	277	Vein	0	10	Vein	1	9
				Barren	80	311	Barren	48	343
Vector Space Distance Metric	Known U	15	6	J/C	47	11	J/C	3	55
	Prospects	8	22	TR	2	43	TR	28	17
	Occurrences	8	18	Tert	0	4	Tert	0	4
	Barren	83	220	Vein	0	10	Vein	0	10
				Barren	8	383	Barren	10	381
Clustering	Known U	21	N.A.	J/C	40	18	J/C	12	46
	Prospects	4	26	TR	24	21	TR	34	11
	Occurrences	1	25	Tert	0	4	Tert	0	4
	Barren	30	273	Vein	0	10	Vein	0	10
				Barren	48	343	Barren	34	357

J/C = Jurassic/Cretaceous; TR = Triassic; Tert = Tertiary; Vein = Vein, Breccia Pipe

CHAPTER 6

This chapter investigates changes in the performance of the linear discriminant and minimum distance classifier in several recognition experiments based on subsets of the features listed in Chapter 3. Experiments with various feature sets serve as tests of the stability of recognition and may in some cases lead to improved recognition results. The Bongard algorithm automatically selects subsets of features to form traits of U and U* objects; the clustering procedure used here provides no intrinsic feature ranks to guide the choice of feature subsets. These two algorithms are not considered here.

Good recognition results can often be obtained with a small number of features. Increasing the number of features used to classify a training sample can produce an under-populated feature space and unstable or misleading recognition results. Each new feature will add not only information (some of which may already be carried by previous features), but may also introduce additional noise into an already imperfect classification process. Performance may improve as the first few features are added to recognition and then may deteriorate as still more features are introduced. Human interpretors of geological data are not likely to disregard any information in making exploration decisions, however. One is reluctant, therefore, to cast features out from automated classification. The experiments presented here suggest that performance in combined interpretation will seldom deteriorate with the addition of features, so that all relevant data at hand may be used in recognition. With a training sample of reasonable size (about twice as many objects as features), the use of numerous features often appears to stabilize recognition results. When only a small number of training samples are available, simpler recognition algorithms or a smaller number of features may yield the best results.

In many pattern recognition problems, the incremental improvement in performance with each additional feature becomes smaller as the total number of features grows, if the features are of roughly equal significance. The worst possible performance is achieved when no features are used in classification and every object is assigned to that class which has the maximum a priori probability. In the present examples, the worst possible performance in this sense would be to class, for example, all 508 Colorado Plateau cells as U*, producing errors on all the known U objects. The error rate is then 100% for U objects and 0% for U* objects, but only approximately 15% on all 508 objects. When one, and then several, features are used in recognition, the fraction of the U population removed from the U* class will be larger than the fraction of U* objects incorrectly classed as U. Although the total number of errors may increase, the summed percentages of each class correctly recognized will also increase. Although there will then be two kinds of errors, more productive recognition will result.

Variations in the performance of a classifier should be judged in the context of some sort of penalty for wrong decisions and some reward for correct decisions. The reward for correct classification of U and U* objects is a minimization of exploration costs and the efficient, orderly exploitation of mineral resources. Because human prospectors often make mistakes and because U objects are the targets of exploration, the misidentification of a U* object seems implicitly less serious an error than misclassification of a U object.

Clearly the two types of errors have different penalties. When a U object is not correctly recognized, the opportunity to profit from a uranium deposit is lost, at least temporarily. The magnitude of the loss is related to the size of the deposit, to the cost of accessing mining and milling the ore, and to prevailing market prices for ore. Some gain is registered, however, because the expenses of land acquisition, detailed field exploration, and drilling are not incurred, and this unused exploration capability may be used elsewhere. When a U* object is misclassified, the loss consists of land acquisition costs, detailed exploration, and perhaps drilling costs until the futile search for a deposit is abandoned. There is also a less tangible loss from both types of errors that results from the incorrect disposition of capital and field crews and resultant loss of any time advantage in exploration to other groups that may be searching in more productive areas.

The penalty for either type of error is difficult to assess, and will be different for each misclassified object, depending on the difficulty of exploration and the expected payoff of a deposit. A logical framework for automated prospecting decisions could eventually include variously weighted penalties for incorrect decisions. It is not the aim of the present study to find guides for estimating these penalties; therefore, all errors in recognition are treated equally. The "best" recognition results are sought under this condition.

6.1 Casper Quadrangle

The Casper Quadrangle offers an interesting performance contrast between the linear discriminant with binary features and the minimum distance classifier with continuous-valued features.

The error rates for the minimum distance classifier using the first 1, 2, 4, 6, 12, and all vector space features ranked according to their information content are shown in Figure 6-1. As features are added through the first 12, recognition seems to be converging toward correct classification of about 95% of U* objects and 55% of U objects as the incremental change in recognition continually decreases. When all 32 features suitable for vector space treatment are used in recognition, however, recognition of U* objects deteriorates considerably. In the lower dimensional feature spaces, more than half of the 21 U objects apparently form a cluster that is tight enough to exclude all but a few U* objects. In 32-dimensional space, however, the U and U* objects are sufficiently dispersed and mixed so that the U* error rate rises dramatically. The danger of using too many features with too few training samples is obvious.

In geological problems such as this, where only a few samples may be available, it will be wise to check the

<u>FIGURE 6-1</u>: Error rates for several variants of recognition in the Casper Quadrangle. Axes represent fractions of the U and U* classes correctly recognized. MAP refers to the maximum a priori probability classification. Outcomes of recognition with the minimum distance classifier using 1, 2, 4, 6, 12 (1I, 2I, 4I, 6I, 12I), and all (A) features are linked by dotted lines. Outcomes of recognition with the linear discriminant using 1, 5, 10, 15, 20, 25, 30, and 36 features (1B . . . . 36B) are linked with solid lines.

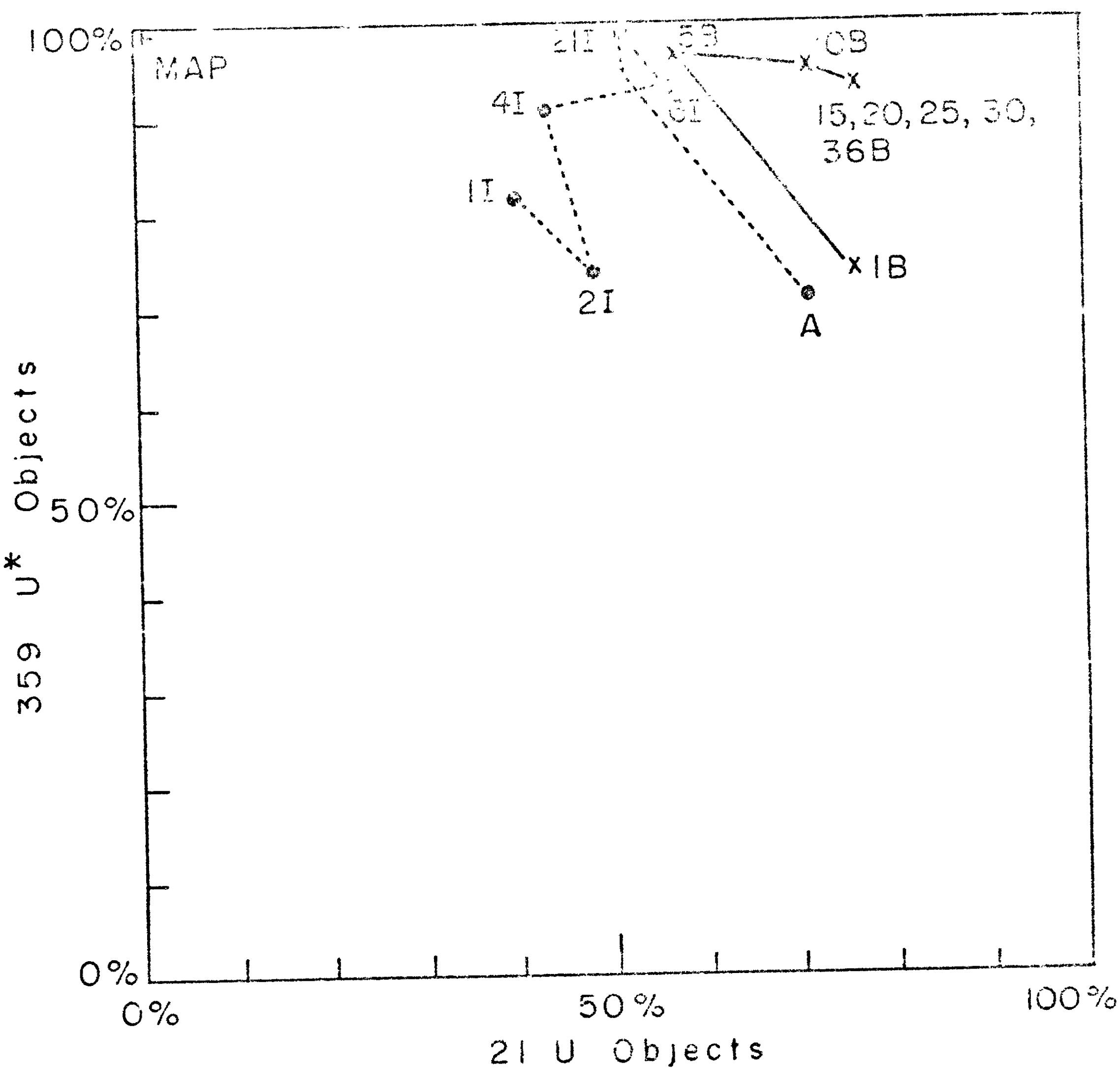

100%
MAP
2I
5B
10B
4I
8I
15, 20, 25, 30, 36B
1I
1B
2I
A
359 U* Objects
50%
0%
0%
50%
100%
21 U Objects

FIGURE 6-1

performance of a classifier with subsets of all available features. It may be the case, as here, that discarding some features will produce more useful recognition results. Because uranium deposits are known to be rare, one is more inclined to have confidence in predicted new prospects if new predictions amount to 3% rather than 30% of all objects originally presumed barren.

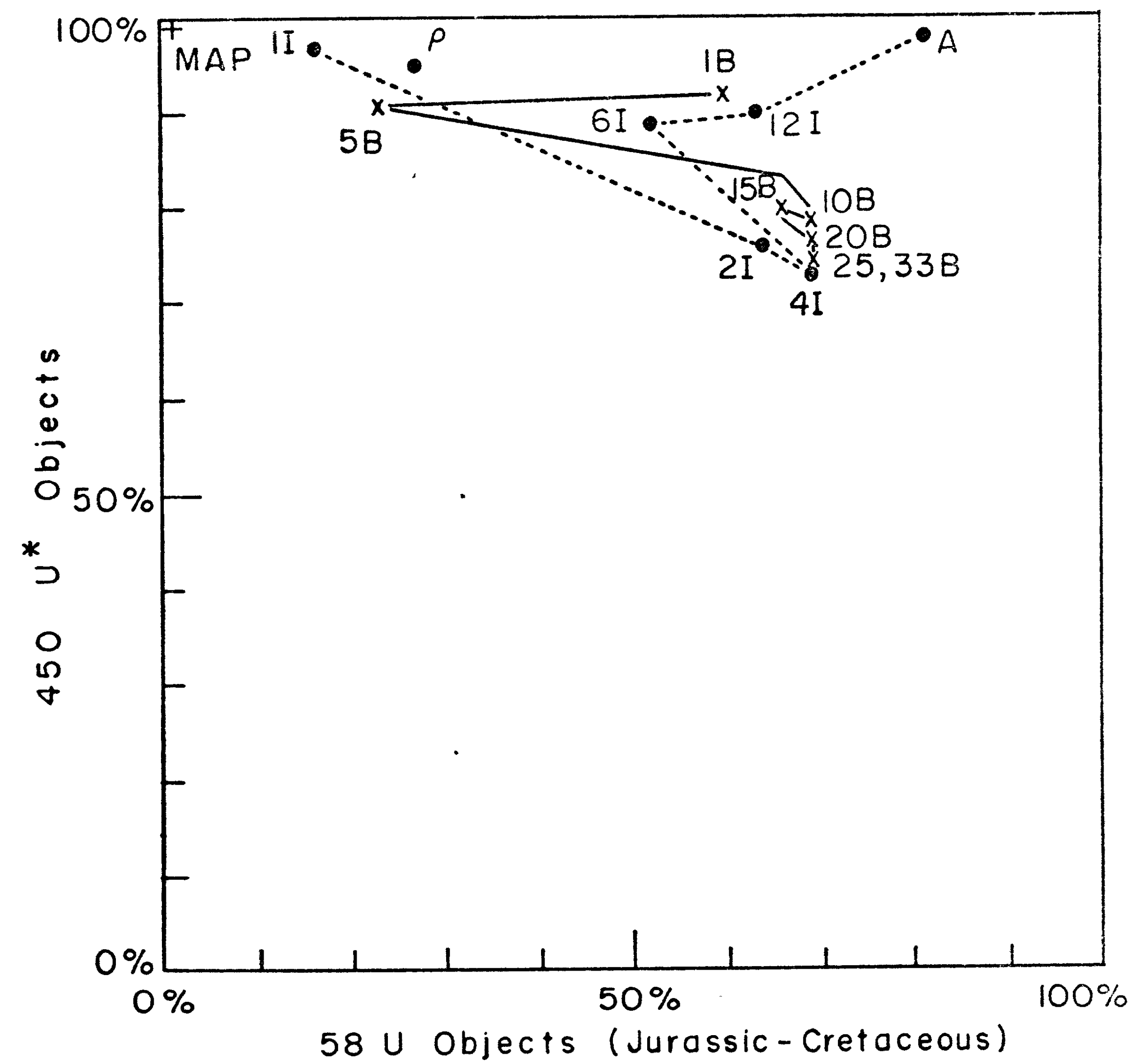
The linear discriminant with binary-coded features offers an alternative to discarding features in order to reduce the dimensionality of a feature space. When features are binary-coded, the entire volume of the 36-dimensional feature space is collapsed onto the vertices of a 36-dimensional hypercube. This severe quantization of features seems to inhibit mixing of the groups. Recognition with the linear discriminant using increasing numbers of features up to 15 very quickly converges to a correct recognition rate, of 76% of U objects and 93% of U* objects (Figure 6-1). More than doubling the number of features does not affect recognition. This simplified recognition algorithm that uses easily interpreted binary features may be most useful when a limited number of samples are available, although its performance is often poorer than the minimum distance classifier when numerous samples are available.

6.2 Colorado Plateau Jurassic/Cretaceous Deposits

Here there are 58 U objects and 450 U* objects for recognition with 32 features. In all experiments, the 58 U objects seem sufficient to support the 32 features. Although there are eight times as many objects in the U* class than in the U class, the U* class may not be excessively large. The paths of geological evolution that will lead an object to be barren are undoubtedly more numerous than those that will lead to an ore deposit. In exploration problems, barren objects should probably outnumber producing objects so that the variety of a geological province may be adequately represented in the training samples.

The trajectory of the linear discriminant's recognition through the U/U* error graph is shown in Figure 6-2. Recognition with fewer than 10 features is unstable, but with 10 or more features taken in order of decreasing weight from this classifier, recognition converges toward 70% U objects and 75% U* objects correctly classified. As features are added in order of decreasing information content to the minimum distance classifier, performance varies widely (Figure 6-2). The best performance of the minimum distance classifier is that using all features, indicating that most of the 58 U objects, though widely

<u>FIGURE 6-2</u>: Error rates for several variants of recognition on the Colorado Plateau. Axes measure fractions of the U and U* populations correctly recognized. MAP refers to the maximum a priori probability classification. Outcomes of recognition with the minimum distance classifier using 1, 2, 4, 6, 12, and all features (1I, 2I, 4I, 6I, 12I, A) are joined by dotted lines. Outcomes of recognition with the linear discriminant using 1, 5, 10, 15, 20, 25, and 32 features (1B . . . . . 32B) are joined by solid lines. The point labelled ρ is the outcome of recognition with the minimum distance classifier using all features with correlations less than 0.75.



100%
MAP
1I
ρ
A
1B
5B
6I
12I
15B
10B
20B
25,33B
2I
4I
50%
0%
0%
50%
100%
450 U* Objects
58 U Objects (Jurassic-Cretaceous)

FIGURE 6-2

scattered on the Plateau, do form a fairly compact cluster in feature space.

Another variant of recognition with the minimum distance classifier is also plotted in Figure 6-2. The feature set used consists of 23 features, all with mutual correlation coefficients, ρ, $-0.75 < \rho < 0.75$. When this set of more nearly independent features is used, correct recognition of U objects drops to about 25%. This is a dramatic deterioration of performance. Recognition with these 23 features is barely better than that using only one feature. Similar results were suggested in Chapter 5 for use of all features in the minimum distance classifier for the Casper Quadrangle (the magnitudes of mutual correlation coefficients for these Casper Quadrangle features were almost all less than 0.75). This result suggests the merit of retaining redundant, highly correlated features in recognition. If an object has an abnormal value for one feature, one or two partially redundant features may help return it to its proper class. For example, a U object may have an abnormal value for only one of 3 highly correlated features; a U* object, however, is more likely to have all 3 of these feature values far from the U class mean. The value of retaining highly correlated features, just as human interpretors do, is suggested here. When the dimensionality of a feature space is increased by the addition of redundant features, some outliers may be returned to their proper clusters.

Colorado Plateau Triassic Deposits:

Recognition results for both classifiers with subsets of features are shown in Figure 6-3. The 45 U objects are adequate to support the 17 available features in both the linear discriminant and minimum distance classifiers. Correct recognition with the linear discriminant converges toward about 60% of U and 85% of U* objects. Recognition with features taken in order of information rank converges somewhat less strongly toward about 60% of U and 95% of U* objects correctly recognized. The best performance is obtained with only two features in the minimum distance classifier, though the results with all 17 features are not substantially worse. Again, recognition with all features is superior to recognition using only features with mutual correlation coefficients, $-0.75 < \rho < 0.75$.

To summarize, it is dangerous to under-populate a feature space. In exploration problems, however, there will probably always be an abundance of locations presumed barren and rather few examples of productive areas.

FIGURE 6-3: Error rates for several variants of recognition on the Colorado Plateau. Axes measure the fractions of the U and U* populations correctly recognized. MAP refers to the maximum a priori probability classifier. Outcomes of recognition with the minimum distance classifier using 1, 2, 4, 6, 12, and all features (1I, 2I, 4I, 6I, 12I, A) are joined by dotted lines. Outcomes of recognition with the linear discriminant using 1, 5, 10, and 17 features (1B . . . . . 17B) are joined by solid lines. The point labelled ρ is the outcome of recognition with the minimum distance classifier using all features with correlations less than 0.75.

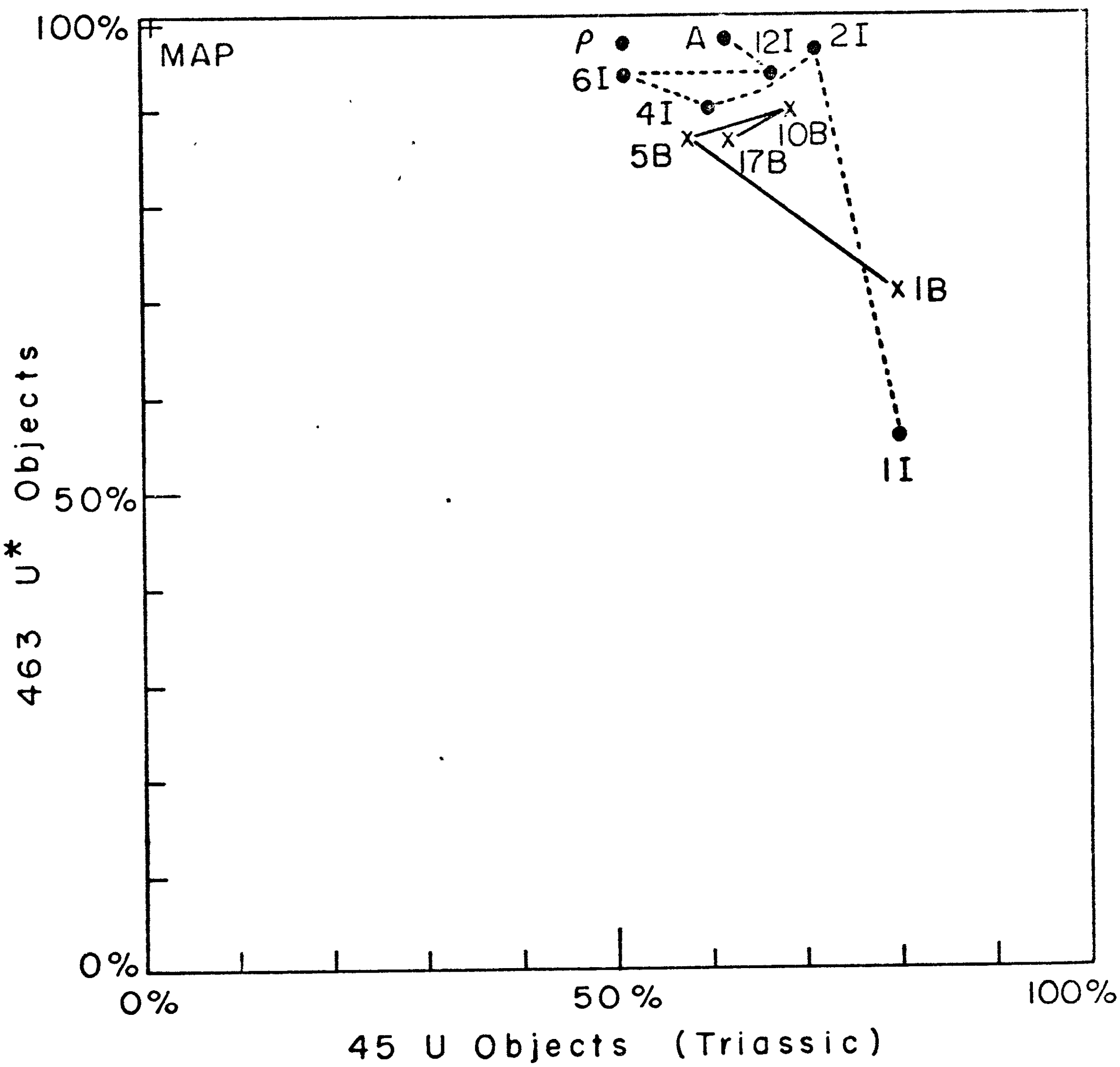

100%
MAP
ρ
A
12I
2I
6I
4I
10B
5B
17B
1B
1I
50%
0%
0%
50 %
100%
463 U* Objects
45 U Objects (Triassic)

FIGURE 6-3

The problem of sample size may be relieved in two ways. Simple recognition algorithms can be used (e.g. the linear discriminant or Bongard algorithms), with severely coded features. Partially redundant features can be included in recognition schemes to restore some outliers to the appropriate classes. In either case, the most unusual deposits may never be recognized. If the sample size is extremely small, less than about 10 objects, even the Bongard and linear discriminant algorithms may not be reliable, and sufficient statistics probably cannot be generated to support a vector space treatment. Some sort of template matching or nearest-neighbor algorithm may be useful in these cases - one could search in whatever feature space is available for unexplored areas that look identical or nearly identical to each of the known resource areas.

CHAPTER 7

This chapter presents recognition experiments that simulate prospecting with automated classification. Subregions within the Colorado Plateau are used to train classifiers; these classifiers then select prospecting targets in the remainder of the Plateau. These experiments serve as tests of previous recognition results and give some insight into the performance to be expected in prediction.

Within a training subregion, classifiers are developed as in previous chapters. Raw data are surveyed to find features that discriminate the U and U* objects within that region. Several variants of recognition are performed within the training region to "fine tune" a classifier for that area. The tuned classifier then picks prospecting targets within the part of the study area that was not used in training. If active mining areas within the prediction area are recognized by the classifier, this will lend confidence in the recognition procedure used. If the features developed from the smaller training district are similar to those given in Chapter 3 in terms of both the parameters that emerge as significant and the estimated PDF's on those parameters, this will substantiate the significance of the features given in Chapter 3. If, however, features and/or recognition results differ substantially from the previous results, an unstable, misleading classification system based on insufficient training is indicated at least for the partial training area, and possibly for the full study area.

We cannot now verify whether or not the U* objects classified in Chapter 5 to be favorable for uranium do, in fact, hold undiscovered uranium deposits. These experiments with a partial training area give some indication of the accuracy expected in prediction with the complete training areas. If the recognition appears stable, the more reasonable predictions of new U objects are probably those that emerge from training based on all known U objects.

7.1 Recognition/Prediction in the Casper Quadrangle

Recognition/prediction experiments for the Casper Quadrangle are not be presented in detail here. Recognition results are unsatisfactory in that they recognize as U only the U objects in the diminished training population. The linear discriminant algorithm that gave the best performance for the Quadrangle in Chapter 5 could not successfully recognize any of the three largest mining areas when they were not included in training.

The minimum distance classifier could not separate the U and U* groups without an unacceptably high error rate.

In the Casper Quadrangle, at least 4 of the 21 known U objects are never correctly recognized when all 21 of these U objects are used for training. Removing these from the training sample will not alter recognition. When the points within one of the three remaining mining areas are removed from training, estimates of feature PDF's are based on approximately 10 samples. These U objects that remain for training are too few to support the statistics for recognition. As they come from only two major mining areas, these training samples also seem to have too little variety to suggest the range of possible appearances of U objects within the Quadrangle.

7.2 Recognition/Prediction on the Colorado Plateau

Recognition and prediction with learning based on subsets of the 45 Triassic and of the 58 Jurassic/Cretaceous U objects on the Colorado Plateau have many characteristics in common. Examples of control experiments with the Jurassic/Cretaceous deposits are discussed here as these U objects are both more numerous and widespread on the Plateau.

For performance comparisons, the reader is referred to Figures 5-2, 5-5, and 5-8 which show recognition analyses with the linear discriminant, Bongard, and minimum distance algorithms, respectively, trained on Jurassic/Cretaceous deposits from the entire Colorado Plateau. Because the largest possible U and U* training populations were used in the analyses in Chapter 5, the numbers of objects correctly recognized in those experiments offer an upper limit to the performance that can be expected from these algorithms when they are trained on only a portion of the plateau.

Prediction Experiment 1:

Cells in the first 11 rows of the Colorado Plateau grid (north of 37-3/4°N) were used to train linear discriminant, Bongard, and minimum distance classifiers for simulated prediction in the southern part of the Plateau. Raw data were scanned to find features that would discriminate the 30 U objects from the remaining 162 U* objects in the north. Twenty of the 23 features found were among the 32 listed in Chapter 3 for recognition of Jurassic/Cretaceous deposits (in some of these features, the decision threshold was shifted by one decile); three new features arose because only the northern portion of the Plateau was used in training.

Prediction with the linear discriminant using binary-valued features picks 46 new prospects in the south that form a single contiguous area (Figure 7-1). Of these 46, 12, or 26% are correctly predicted Jurassic/ Cretaceous U objects, 3 have deposits in other host rocks, and 31 are believed to be barren. This prediction area is reasonable when compared to the linear discriminant's recognition on the whole Plateau (Figure 5-2). The predictions offer what a conservative geologist might, suggesting a southward extension of known mining areas with uranium favorability extended toward the east, following a slight southeasterly trend in the disposition of the known mining areas. No distant deposits in the extreme southeast are predicted (nor were these correctly recognized when this algorithm was trained on the entire Plateau).

The Bongard algorithm, using the same 23 binary-valued features selects prospects in the south that are substantially the same as those in Figure 7-1. The predicted favorable zone is again a single contiguous area with 16 U objects and 36 U* objects forming a southeasterly continuation of mining areas to the north. This area lies within the area recognized as uranium-favorable when the Bongard algorithm was trained on the entire Plateau (Figure 5-5). In prediction, however, the favorable zone extends southward only to row 18; no cells further south

FIGURE 7-1: Learning/prediction experiment with the linear discriminant trained to recognize Jurassic/Cretaceous uranium deposits in the northern 11 rows of cells. The area recognized favorable for uranium in the training area is within the solid outline. The area in the south that is predicted to be favorable for uranium is within the dotted outline (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 7-1

are recognized as U, although when the Bongard algorithm was trained on the entire Plateau, U objects as far south as row 23 were correctly recognized.

The minimum distance classifier exhibits somewhat different behavior in prediction. None of the objects in the prediction area south of row 11 are classed as U; all objects are more similar to the U* class. This suggests that the few U objects used in training form a relatively tight cluster in feature space. More examples of U objects are required to manifest the variety of U objects.

One alternative to giving up a search for uranium in the south is to examine those objects that are most similar to the U training cluster. The 50 objects nearest to the U cluster again form a southward and eastward extension of the known producing areas to the north. None of the deposits in the extreme south that were recognized with training on the whole Plateau are included in the 50 objects most similar to the U cluster.

An interesting recognition phenomenon appears here when various feature sets are used with the minimum distance classifier. When only a small training sample is available, the classifier performs better with fewer features. Figure 7-2 shows the number of U objects correctly predicted as a function of the total number of

FIGURE 7-2: Graph of the number of correct predictions of Jurassic/Cretaceous uranium-producing cells south of row 11 as a function of the total number of predictions through the first 50 prediction objects nearest the U training cluster. The minimum distance classifier was trained to recognize Jurassic/Cretaceous uranium deposits in the northern 11 rows of cells. Results of recognition with 5, 10, 15, and 23 features are plotted (see legend Figure 3-1).

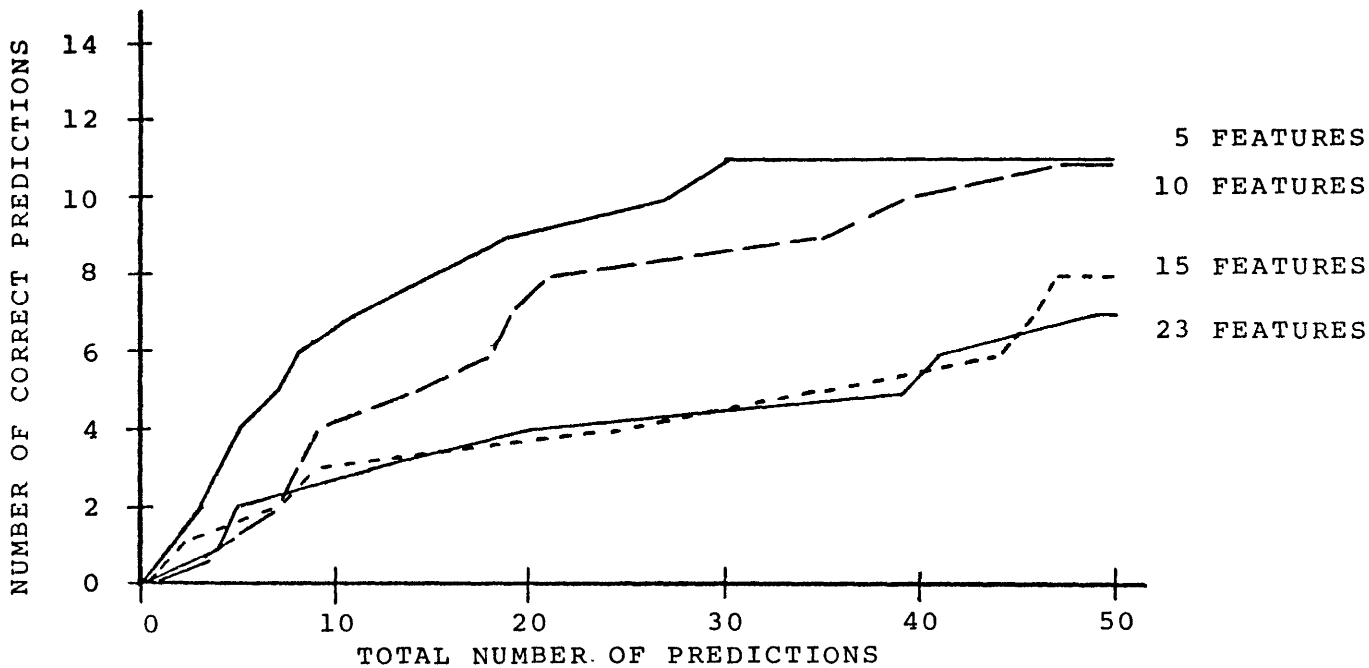

NUMBER OF CORRECT PREDICTIONS
14
12
10
8
6
4
2
0
0
10
20
30
40
50
TOTAL NUMBER OF PREDICTIONS
5 FEATURES
10 FEATURES
15 FEATURES
23 FEATURES

FIGURE 7-2

objects predicted to be favorable for uranium through the first 50 objects closest to the U cluster in 5, 10, 15, and 23-dimensional feature spaces. With all feature sets, the number of correct predictions increases logarithmically with the total number of predictions. One is naturally more inclined to accept as U those objects that, if not within the U cluster, are the closest objects to the U training cluster. Prediction with fewer features scores many more successes in the first few predictions then does prediction with more features. In Chapter 6 additional features tended to stabilize recognition results when the training sample was large and varied. Here the training sample is smaller and less varied so that additional features measured on an object are more likely to make that object an outlier from the U cluster and to mix it with U* objects in the feature space.

Prediction Experiment 2:

In a second set of prediction trials, information from the western 2/3 of the Plateau was used to pick exploration targets in the east. Cells in columns 1-17 formed a training set with 25 Jurassic/Cretaceous U objects and 320 U* objects.

Despite the fact that the number of U objects available for training was less than half of the 58 used to generate features for the whole Plateau, and the fact that 45 Triassic U objects were in the U* group, features selected for this recognition were very similar to those used to recognize Jurassic/Cretaceous deposits on the whole Plateau. Nineteen of 25 features were essentially unchanged from those given in Chapter 3. Six new features appeared for this limited training population.

The result of recognition with the linear discriminant using binary-valued features is shown in Figure 7-3. In the training area, a broad zone including 16 U objects and 46 U* objects is classified as U. In the prediction area, 49 objects are classified as U. All producing cells in and near the Uravan mineral belt are correctly classified as U. In all, 19 cells with production from Jurassic/Cretaceous host rocks are correctly recognized. The correct prediction rate here is at least 38%, and perhaps more if some cells now presumed barren do hold undiscovered deposits. The area predicted favorable for uranium combined with the area recognized as U in the training area form very nearly the same as that area classified as U when this classifier was trained on the whole Plateau, again indicating the stability of recognition. Very similar predictions emerge when the Bongard algorithm is used with these binary-valued features.

FIGURE 7-3: Recognition/prediction experiment with the linear discriminant trained to recognize Jurassic/Cretaceous uranium deposits in the western 17 columns of cells. Areas within the training area recognized favorable for uranium in the training area are within the solid outline. Areas predicted favorable for uranium are within the dotted outline (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 7-3

The performance of the minimum distance classifier is shown in Figure 7-4. The variant of recognition shown uses the first 10 of 25 available features, according to their information rank. Seventeen of 25 U objects in the training area are correctly recognized with few U* objects classified as U. In the prediction area, only 2 objects are within the U cluster; both of these are producing cells in the Uravan area. The 20 objects from the prediction area that are nearest to the U cluster are indicated in Figure 7-4. Ten of the first 10, and 15 of the first 20 have recorded production from Jurassic/Cretaceous host beds. Here, again, the number of U objects correctly predicted increases roughly logarithmically with the total number of objects predicted. A cutoff at the first 20 objects is arbitrary; more U objects could be predicted correctly, but only with the penalty of an increased error rate.

In a final test of prediction, the cells in a large cluster of 17 Jurassic/Cretaceous U objects around the Uravan-Gateway area (between rows 7 and 14 and columns 16 and 21) were withheld from training. When trained on the remaining 41 Jurassic/Cretaceous U objects, all three classifiers were able to correctly recognize all 17 of these U objects.

<u>FIGURE 7-4</u>: Recognition/prediction experiment with the minimum distance classifier trained to recognize uranium deposits in the western 17 columns of cells. Areas within the training area recognized favorable for uranium are within the solid outline. Areas predicted favorable for uranium are within the dotted outline. (see legend Figure 3-1).

COLORADO PLATEAU - URANIUM DEPOSITS

										4	4	4	4	4	1C	4	1C	1C	4								
										4	4	4	1T	1T	4	4	4	1C	4								
								4	4	4	4	4	4	4	1T	4	4	4	4	4	4						
								4	4	4	4	4	4	4	4	4	4	4	4	4	4						
						4	4	4	4	1C	4	4	4	4	4	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	4	4	4	1T	1C	4	4	4	4	4	4	4	4				
						4	4	4	4	2R	1R	2C	4	4	2C	2C	1C	4	4	4	4	4	4	4	4		
						4	4	1R	2R	2R	2R	4	1R	1C	2R	2C	2C	2C	4	4	4	4	4	4	4		
		4	1V	2V	4	4	4	1R	4	1C	1C	4	4	2R	2R	2C	2C	3C	4	4	4	4	4				
		4	4	4	4	4	4	2R	1C	4	1C	1R	4	4	2R	3R	2C	3C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	1R	2R	4	2C	2R	4	4	2R	2R	2C	2C	2C	4	2C	4	4				
4	4	4	4	4	4	4	4	4	1R	4	1C	2R	2R	2R	2C	1C	2C	4	4	4	4	2C	1V				
4	1V	1R	4	4	4	4	4	4	4	4	4	2R	4	4	2C	2C	1C	1C	4	4	2C	4	4	4	4		
2R	4	4	4	4	4	1V	4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4		
4	4	1R	4	4	4	4	4	4	4	4	2R	2R	2R	2R	4	2C	2C	4	4	1C	4	4	4	4	4	4	4
4	4	1V	4	4	4	4	1R	4	4	4	4	4	4	4	4	2C	2C	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	2C	4	2C	1C	4	4	4	4	4	4	4	4	4	4
1R	1V	4	4	4	2V	4	4	2R	4	4	4	4	1C	2C	4	4	4	4	4	4	4	4	4	4	4	4	1C
4	4	4	4	4	1V	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1C		
4	4	4	4	4	4	4	4	2R	1R	4	4	4	4	1V	4	4	4	2C	2C	4	4	4	4	4	4		
				4	4	4	4	4	2R	4	4	4	4	4	4	4	4	4	4	4	2C	3C	4	4	4		
				4	4	4	4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	2C	4	3C	4		
								4	4	4	4	4	4	2R	1R	4	4	4	4	4	4	4	4				
								4	4	4	4	4	4	4	1R	4	4	4	4	4	4	4	4				

FIGURE 7-4

In all cases where adequate training was available, prediction experiments classified as favorable for uranium areas that were substantially the same as those areas classified favorable when training on the entire Plateau was available. Features developed for the smaller training areas also show a close resemblance to features for the whole Plateau. Success rates in prediction ranged from 1 in 4 to 3 in 4 or better, depending on the number of predictions made. These success rates suggest that automated prediction of mineral resources is a very practical possibility.

CHAPTER 8

This chapter reviews the pattern recognition procedures applied to uranium prospecting and the results that emerge from combined interpretation of geological data for the Colorado Plateau and Casper Quadrangle. Other geological problems that might be investigated with pattern recognition techniques are suggested.

8.1 Feature Selection

Previous quantitative formulations of prospecting problems have had some subjective elements in the way that geological criteria were chosen, weighted, or combined for decision-making. In this study, the features used in recognition were selected algorithmically from a larger pool of candidate features. Construction of this pool of candidate features is the outstanding subjective aspect of the present procedures. There seems to be no way to avoid subjectivity at this most fundamental level of the prospecting problem.

One may minimize subjectively in feature selection, however, by proposing a large number of candidate features that thoroughly explore, if they do not exhaust, the types of features that might be constructed from a given data base. A single type of data may suggest a plurality of features, only one of which may finally be selected for use in recognition. For example, a map of faults in a province could be used to generate features related to fault density per unit area, strike of the faults, proximity to faults, proximity to faults of specific types, proximity to intersections of faults, and so on. The most effective features for recognition may be discovered only by examining a large number of candidates. A large pool of candidate features is easily winnowed by computer.

8.2 Feature Coding and Training Sample Size

The binary feature coding used in this study is an extreme type of coding, but is obviously useful in some situations (e.g. the Casper Quadrangle). Some precision in the descriptions of objects is lost in binary coding, but simple geological interpretations of features may emerge. With few training samples, one may be more accurate in estimating the statistics of binary-valued features than the statistics of probability density functions. Simple, easily implemented decision algorithms may be used with binary-valued features. Duda *et al*. (1977) have used less severely quantized features (effectively 4-place histograms) in diagnostic approaches to prospecting. When many training samples are available, exact values of feature data may be retained to give the most precise descriptions of objects for effective recognition.

Automated classification with continuous-valued features appears to be useful for combined interpretation when many samples of each class are available for training. In this study, classification with the minimum distance classifier proved to be effective when the number of samples of each class was approximately equal to or greater than twice the number of features. Because barren areas are probably more various than areas with a particular type of mineral deposit, the number of barren objects used in training should probably be greater than the number of producing areas, so that the full variety of these barren objects may be adequately represented.

When few samples are available to train an exploration classifier, pattern recognition techniques may still be applicable. Feature coding techniques that collapse feature spaces onto a finite number of points may be useful in unmixing classes of objects. These feature coding techniques lead naturally to simple decision algorithms that do not require, for example, estimates of covariance matrices or probability density functions. Parzen estimates might prove useful for providing estimates of feature probability density functions so that vector space techniques could be used with few training samples. With only one or a very few examples of mining areas, exact matches to the features of a mining area might be sought among barren objects. If no exact matches exist, a nearest-neighbor classification procedure might be useful, but with few samples it is difficult to determine how slight variations in feature values may affect the ore favorability of an object. With a small number of productive areas for training, the semantic network or diagnostic approaches to prospect evaluation might be the most reliable procedures, provided that models of ore deposition are sufficiently advanced to support the required software.

8.3 Recognition Algorithms

When pattern recognition is applicable to combined interpretation problems, a variety of algorithms are available for use. Different algorithms offer different models of the learning and decision-making process. The choice of algorithm depends on the problem at hand; if no one type of algorithm is a clear choice for recognition, the results from several different algorithms may compliment one another.

Clustering algorithms such as the one used here can be used with coded or numerical data. Though they do not automatically make decisions, unsupervised clustering algorithms are conceptually simple and manifest natural groupings of objects. Two types of clusterings might be particularly useful in geological problems. Clusters of minimum diameter can isolate a number of separate families of objects within a feature space. Chain-type clusterings that grow clusters by addition of the object nearest to a point already within the cluster might be useful for following through a feature space a spectrum of related objects that differ from one another by slight variations in several features.

Linear discriminants are conceptually simple and may be used with continuous-valued or quantized features. These algorithms divide a feature space into two regions, each associated with one of the classes in a 2-class problem. Classifiers that divide a feature space with several hyperplanes are more difficult to implement, but can be used in many-class problems. Non-linear decision surfaces can be used to segment a feature space, but these typically require more training samples than linear surfaces.

The Bongard algorithm using binary-valued features may be particularly useful in poorly understood problems. The characteristic traits that emerge from this algorithm are easily interpreted and may suggest new insights into complex, incompletely understood phenomena. Press and Briggs (1975) used this algorithm to suggest a model of Chandler Wobble excitation and decay.

Minimum distance classifiers such as the one used in this study use continuous-valued features that offer the precise descriptions of geological objects. The features used with these classifiers obscure neither the overall variety of objects within a class, nor the details of individual objects. These vector space techniques cannot readily accommodate non-numerical descriptive or qualitative geological data that may be important to exploration evaluations.

8.4 Summary of Results from the Colorado Plateau and Casper Quadrangle

Recognition analyses with binary-valued features in Chapter 5 suggest that large areas of the Colorado Plateau might be favorable for uranium in Jurassic, Cretaceous, or Triassic strata. The minimum distance classifier, however, recognizes few new productive areas. Most of the non-producing areas classified as favorable by the minimum distance classifier are adjacent to known resource areas. Because the performance of the minimum distance classifier proved to be superior to that of the other classifiers, one might infer that there are few new resource areas to be found on the Colorado Plateau (recall, though, that the data available here are insufficient to unequivocally determine uranium favorability). Fisher (1974) has suggested that exploration on the Colorado Plateau be directed toward location of new mining districts or clusters of deposits rather than isolated, individual deposits. If any such clusters exist, the minimum distance classifier suggests that they may be adjacent to known resource areas. Deposits dissimilar to those used to train the classifiers could not be recognized on the Colorado Plateau; unusual types of uranium deposits may exist on the periphery of the Plateau. The small number of new predictions supports Lieberman's (1976) conclusion that most major uranium deposits on the Colorado Plateau may have been discovered.

The Casper, Wyoming Quadrangle appears in all recognition experiments to be a less mature mining area than the Colorado Plateau. On the basis of more diverse data than are available for the Colorado Plateau, several areas far removed from active mining operations in the Quadrangle are recognized as favorable for uranium (Figures 5-1 and 5-4).

In addition to extensions of active mining areas, eleven points that form a north-south trending zone through the east-central portion of the Quadrangle were predicted to be favorable for uranium. This zone coincides with newly recognized anomalies in an airborne radiometric survey delivered to E.R.D.A. by subcontractors after the completion of this pattern recognition study. Eight of the eleven points picked by pattern recognition analysis coincide with significant radiometric anomalies. In recognition experiments, proximity to major radiometric anomalies proved to be the strongest single feature for the Casper Quadrangle. Geological information synthesized from other features was sufficient to override an absence of recognized radiometric anomalies in this zone and to pick this area out from the rest of the Quadrangle as favorable for uranium. Surface radiometric anomalies do not guarantee a productive zone at depth, but the coincidence of pattern recognition predictions with the most extensive anomalies not associated with established mining areas lends confidence in pattern recognition approaches to combined interpretation. The most favorable ground for prospecting in the Quadrangle may be near those areas where geological criteria from pattern recognition combined interpretations coincide with radiometric anomalies.

8.5 Other Geological Applications of Pattern Recognition

Computerized classification techniques may be the only practical way to manipulate and interpret large, complex geological data bases. When data are of unfamiliar interrelationships, quantified decision procedures may provide the most efficient paths toward establishing learned experience and expertise in interpretation.

LANDSAT photographs offer one example of large data bases that, for some applications, lack well-known interpretation rules (one LANDSAT frame has approximately 7.3 x 10^6 pixels). With the 4-band LANDSAT imagery now available, many problems are open to pattern recognition analysis. The generation of photographic satellites designed to succeed LANDSAT will procude many-band images that will segment the visible into more narrow wavelength bands and extend photographic coverage into the infrared. This increased number of wavelength bands over an extended spectral range will increase the number of features that can be considered in automated photographic analysis.

Limited recognition of surface rock type based on spectral signatures has been attempted with present 4-band images (Rowan _et al._, 1974). This work can undoubtedly be extended and refined with the many-band images soon to be available. Techniques for detection of hydrothermally altered rock (Rowan _et al._, 1977) and recognition of diagnostic spectral signatures of soils overlaying mineral deposits (Vincent, 1977) should follow similar growth. Where rock and soil features are not directly visible in photographs, anomalous patterns in seasonal variations of vegetation may be useful in geobotanical prospecting.

Geological structures might be analyzed for resource potential with pattern recognition techniques. Structures such as anticlines or basins may have features visible only in satellite imagery that might indicate favorable environments for hydrocarbon accumulation. Seismic sections through "bright spots" might be analyzed for subtle features indicative of the presence of hydrocarbons. Automated evaluations of these structures might supplement the interpretations of trained geologists.

Problems of a more inferential nature such as combined interpretation of geological data offer a different use of pattern recognition techniques. Pattern recognition techniques could provide quantitative rankings of favorability for ore in parcels of land. The following questions might be addressed:

1) Given a plurality of available tracts of land, which one(s) are most favorable to purchase or lease?
2) Given several parcels of land, which one(s) should be explored first and in what order should the totality be explored so as to maximize return?
3) Given several parcels of land, which should be disposed of to suit a limited exploration budget and/or the time constraints of leases?

As more detailed geological data are secured, data bases may be easily updated so that one may be directed with increasing resolution in the search for resources.

Pattern recognition analysis of diverse geological data might be most useful in producing quantitative, explicable, and reproducible analyses of geological problems. These analyses could serve as a common reference point for different interpretors and as a point of departure for more intuitive or speculative exploration decisions. The recognition framework offers an easily managed, well-organized data base into which new data may be readily integrated to produce revised, updated interpretations. Use of quantitative procedures for the combined interpretation of geological data may lead to reductions in the amounts of time, money, and manpower needed to locate ores, and may help to provide an uninterrupted supply of some minerals.

REFERENCES

Bongard, M., M. Vaintsveig, Sh. Guberman, M. Izvekova, and M. Smirnov, The use of self-learning programs in the detection of oil containing layers. Geol. Geofiz., 6, 96-105, 1966.

Boyle, B., The Decision to Grant Credit, Ph.D. thesis, M.I.T., Dept. of Elect. Eng., 1976.

Briggs, P., F. Press, and Sh. Guberman, Pattern recognition applied to earthquake epicenters in California and Nevada. Bull. Geol. Soc. Amer., 88, 161-163, 1977.

Collyer, P. and D. Merriam, An application of cluster analysis in mineral exploration. Math. Geol., 5, 213-223, 1973.

Duda, R. and P. Hart, Pattern Classification and Scene Analysis, Wiley-Interscience, New York, 482 pp., 1973.

Duda, R., P. Hart, and N. Nilsson, Subjective Bayesian methods for rule-based inference systems. Stanford Research Institute, Artificial Intelligence Center Technical Note #124, 22 pp., 1976.

Duda, R., P. Hart, N. Nilsson, and G. Sutherland, Semantic network representations in rule-based inference systems. Stanford Research Institute, Artificial Intelligence Center Technical Note #136, 31 pp., 1977.

Fisher, R., Exploration guides to new uranium districts and belts. Econ. Geol., 69, 362-376, 1974.

Gallager, R., Information Theory and Reliable Communication, John Wiley & Sons, N.Y., 588 pp., 1968.

Gelfand, I., Sh. Guberman, V. Keilis-Borok, L. Knopoff, F. Press, E. Ranzman, I. Rotwain, and A. Sadovsky, Pattern recognition applied to earthquake epicenters in California. Physics of Earth and Planetary Interiors, 11, 227-283, 1976.

Grutt, E., Prospecting criteria for sandstone-type uranium deposits. In: Uranium Prospecting Handbook, Inst. Mining and Metallurgy Trans. (London), pp. 47-76, 1972.

Harshman, E., Geology and uranium deposits, Shirley Basin area, Wyoming. U.S.G.S. Prof. Paper 745, 82 pp., 1972.

Harris, D., A probability model of mineral wealth. Trans. Soc. Mining Eng., p. 199-216, 1966.

Kelly, V., Influence of regional structure and tectonic history upon the origin and distribution of uranium on the Colorado Plateau. United Nations International Conference on the Geology and Uranium and Thorium, pp. 171-178, 1955.

Kerr, P., Uranium emplacement on the Colorado Plateau. Bull. Geol. Soc. Amer., 69, 1075-1112, 1958.

Lieberman, M., United States uranium resources - an analysis of historical data, Science, 192, 431-433, 1976.

McKay, E., Criteria for outlining areas favorable for uranium deposition in parts of the Colorado Plateau and Utah. U.S.G.S. Bull. 1009-J., pp. 263-282, 1955.

Melin, R., Description and origin of uranium deposits in Shirley Basin, Wyoming. Econ. Geol., 59, 835-849, 1964.

Press, F. and P.L. Briggs, Chandler Wobble, earthquakes, rotation, and geomagnetic changes. Nature, 256, 270-272, 1975.

Rackley, R., Environment of Wyoming Tertiary uranium deposits. A.A.P.G. Bull., 56, 755-774, 1972.

Rocky Mountain Association of Geologists, Geological Atlas of the Rocky Mountain Region, 331 pp., 1972.

Rowan, L., A. Goetz, and R. Ashley, Discrimination of hydrothermally altered and unaltered rocks in visible and near-infrared multispectral images. Geophysics, 42, 522-535, 1977.

Rowan, L., P. Wetlaufer, A. Goetz, F. Billingsly, and J. Stewart, Discrimination of rock types and detection of hydrothermally altered areas in south-central Nevada by use of computer-enhanced ERTS images. U.S.G.S. Prof. Paper 883, 35 pp., 1974.

Sharp, W. and A. Gibbons, Geology and uranium deposits of the southern part of the Powder River Basin, Wyoming. U.S.G.S. Bull. 1174-D, 60 pp., 1964.

Shoemaker, E., Structural features of the central Colorado Plateau and their relation to uranium deposits. United Nations International Conference on the Geology of Uranium and Thorium, pp. 155-168, 1955.

Vincent, R., Uranium exploration with computer processed LANDSAT data. Geophysics, 42, 536-541, 1977.

Wier, D., Geologic guides to prospecting for carnotite deposits on the Colorado Plateau. U.S.G.S. Bull. 988B, pp. 15-30, 1952.

Young, I.T. and P. Van Otterloo, Techniques of Pattern Classification. M.I.T. Class Notes 6.362, to be published.