

General Disclaimer

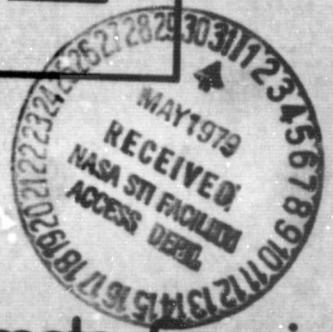
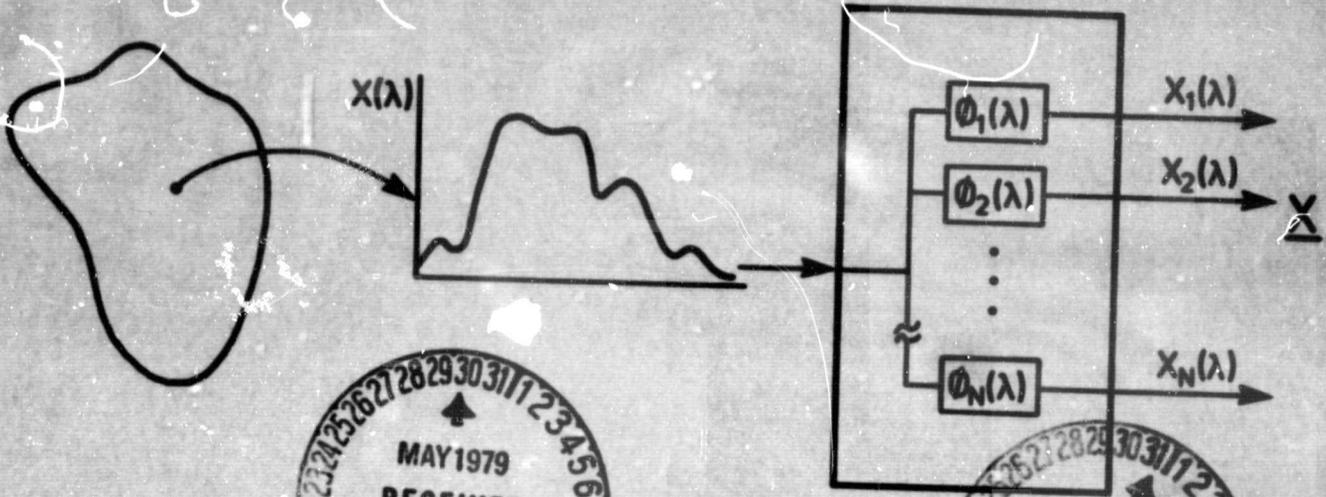
One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

"Data available under NASA sponsorship
in U.S. in order of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

The Analytical Design of Spectral Measurements for Multispectral Remote Sensor System

(E79-10200) THE ANALYTICAL DESIGN OF SPECTRAL MEASUREMENTS FOR MULTISPECTRAL REMOTE SENSOR SYSTEMS (Purdue Univ.) 272 p
HC A12/MP A01 CSCL 14B Unclas 00200
G3/43 N79-25449



Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA
1979

LARS Technical Report 122678
TR-EE 79-13

The Analytical Design of
Spectral Measurements for
Multispectral Remote Sensor Systems

by

D. J. Wiersma
D. A. Landgrebe

Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47906

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vii
GLOSSARY OF SYMBOLS	xiii
ABSTRACT	xvi
1. INTRODUCTION	1
1.1 The Pattern Recognition System	1
1.1.1 The Scene	3
1.1.2 The Sensor	6
1.1.3 The Processor	8
1.2 Previous Approaches to Sensor Design	10
1.3 Present Investigation	16
2. SCENE REPRESENTATION AND SPECTRAL PARAMETER DESIGN	18
2.1 The Analytical Procedure	18
2.2 The Stochastic Process	26
2.3 Representation of the Stochastic Process	29
2.4 Discrete Approximation	42
2.4.1 Spectral Sampling	43
2.4.2 Ensemble Sampling	49
2.4.3 Quantization	51
3. THE PARAMETERS AND OVERALL SYSTEM PERFORMANCE	55
3.1 Spectral Representation	57
3.2 Ancillary Data	64
3.3 Information Classes	70
3.4 Spatial Representation	76
3.5 Signal-to-Noise Ratio	82
4. EXPERIMENTAL SYSTEM AND RESULTS	89
4.1 Field Measurements Data Base	91
4.2 Spectral Parameter Evaluation System	93
4.3 Experimental Procedure	99

TABLE OF CONTENTS (cont.)

	Page
4.4 System Testing	106
4.4.1 Reconstruction	107
4.4.2 Choice of Weight Function	110
4.4.3 Evaluation of the Eigenvalue Algorithm	119
4.4.4 Sample Size	120
4.4.5 Results	125
4.5 Scene Understanding	156
4.5.1 The Dimensionality of the Observation Space	157
4.5.2 Feature Selection	160
4.5.3 Classification Performance as a Function of the Spectral Representation	163
4.5.4 Characteristics of the Eigenvectors	167
4.6 Suboptimum Sensor Design	170
4.6.1 Comparison with Suboptimal Systems .	171
4.6.2 Evaluation of Spectral Subintervals .	184
4.6.3 Proposed Sensor Design	185
5. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH	190
BIBLIOGRAPHY	197
APPENDICES	
Appendix A. A Stratified Posterior Classifi- cation Performance Estimator	205
Appendix B. Data Base Description	217
Appendix C. Correlation Sample Measurements on the Spectrum	225
Appendix D. Computer Program Listings	234

LIST OF TABLES

Table	Page
2.1 Eigenvalues for continuous and sampled covariance	48
4.1 Spectral band locations for two practical sensor designs	98
4.2 Comparison of the probability of correct classification using N terms in the weighted Karhunen-Loeve expansion among four choices of weight functions	119
4.3 Sample size assignments for data from Williams County, N.D. on August 4, 1977	121
4.4 Expected measurement error and proposed values for T for each of the data sets	159
4.5 Ranking of the first 10 optimum features on their ability to discriminate between classes.	162
4.6 Maximum probability of correct classification for the six data sets	167
4.7 Comparison of expected mean-square error (in relative units) for each of the six data sets using two suboptimal sensors and the optimal sensors consisting of the first 4, 6, and 10 eigenvectors	172
4.8 Correlation matrices for the four band suboptimal sensor number 1 using data taken over Williams County on June 29, 1977	183
4.9 Correlation matrices for the first four optimal basis functions using data taken over Williams County on June 29, 1977	184
4.10 Spectral band locations for the proposed sensor	188
4.11 Expected mean-square error (in relative units) and estimated probability of correct classification using the proposed sensor	189

LIST OF TABLES (cont.)

Appendix Table		Page
A1.	Test of error estimator	214
A2.	Theoretical bound of standard deviation for different dimensions	215
A3.	Experimental standard deviation of estimates .	216
B1.	ID information locations in data storage format	224

LIST OF FIGURES

Figure		Page
1.1	Pattern recognition system	2
1.2	Spectral response function for mature wheat collected on August 4, 1977 over Williams County, North Dakota	4
1.3	Sensor system block diagram	7
2.1	Sensor system model	20
2.2	Approximation of the spectral response function by a set of four basis functions . .	21
2.3	Flowchart of the design procedure	23
2.4	Realization of a stratum as the ensemble of spectral sample functions	28
2.5	Eigenvalues for the stochastic process example	39
2.6	Partitioning of the spectral interval into L subintervals	45
3.1	Two distributions which demonstrate a potential difficulty in using the best feature for representation to perform classification . .	60
3.2	Probability of correct classification as a function of expected mean-square representa- tion error	62
3.3	The effects of sample size on classification performance as a function of the number of features, a) true performance, b) Positive bias in \hat{P}_C due to testing on the design set, c) Negative bias in \hat{P}_C due to testing on the test set, d) estimate of \hat{P}_C when testing on the design set, e) \hat{P}_C for testing on the test set	69
3.4	An information tree for a typical stratum (Landgrebe, 1978)	71

LIST OF FIGURES (cont.)

Figure	Page
3.5 Distributions of four information classes . . .	73
3.6 Relationships between object size F and ground resolution element size Δ	78
3.7 Classification performance vs. spatial reso- lution using ECHO and per-point classifiers (Landgrebe et al., 1977)	81
3.8 The locus of eigenvalues for an N' dimensional signal in white Gaussian noise .	86
4.1 Spectral parameter design system	90
4.2 Sample output from SPOPTM	101
4.3 Sample output of class conditional statistics	104
4.4 Sample output of classification performance estimates	105
4.5 Reconstruction of a single spectral response function using from 1 to 8 terms in the expansion	108
4.6 Weight functions	111
4.7 First four optimal basis functions using weight function number 1 over Williams County, May 8, 1977 data	112
4.8 First four optimal basis functions using weight function number 2 over Williams County, May 8, 1977 data	115
4.9 First four optimal basis functions using weight function number 3 over Williams County, May 8, 1977 data	116
4.10 First four optimal basis functions using weight function number 4 over Williams County, May 8, 1977 data	117
4.11 Influence of sample size on expected mean- square error for Williams County, August 4, 1977, using weight function number 2	122

LIST OF FIGURES (cont.)

Figure	Page
4.12 Influence of sample size on the estimate of classification performance for Williams County, August 4, 1977 using weight function number 2	124
4.13 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Williams County, May 8, 1977, using weight function number 2	126
4.14 First twelve eigenvectors for Williams County, May 8, 1977, using weight function number 2	127
4.15 Estimate of probability of correct classification vs expected mean-square error for Williams County, May 8, 1977, using weight function number 2	130
4.16 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Williams County, June 29, 1977, using weight function number 2	131
4.17 First twelve eigenvectors for Williams County, June 29, 1977, using weight function number 2	132
4.18 Estimate of probability of correct classification vs expected mean-square error for Williams County, June 29, 1977, using weight function number 2	135
4.19 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Williams County, August 4, 1977, using weight function number 2	136
4.20 First twelve eigenvectors for Williams County, August 4, 1977, using weight function number 2	137
4.21 Estimate or probability of correct classification vs expected mean-square error for Williams County, August 4, 1977, using weight function number 2	140

LIST OF FIGURES (cont.)

Figure		Page
4.22	Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Finney County, September 28, 1976, using weight function number 2	141
4.23	First twelve eigenvectors for Finney County, September 28, 1976, using weight function number 2	142
4.24	Estimate of probability of correct classification vs expected mean-square error for Finney County, September 28, 1976, using weight function number 2	145
4.25	Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Finney County, May 3, 1977, using weight function number 2	146
4.26	First twelve eigenvectors for Finney County, May 3, 1977, using weight function number 2	147
4.27	Estimate of probability of correct classification vs expected mean-square error for Finney County, May 3, 1977, using weight function number 2	150
4.28	Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Finney County, June 26, 1977, using weight function number 2	151
4.29	First twelve eigenvectors for Finney County, June 26, 1977, using weight function number 2	152
4.30	Estimate of probability of correct classification vs expected mean-square error for Finney County, June 26, 1977, using weight function number 2	155
4.31	Estimate of probability of correct classification vs expected mean-square error for (a) Williams County and (b) Finney County, using weight function number 2	164

LIST OF FIGURES (cont.)

Figure	Page
4.32 Regions of the spectral interval which are not represented by a suboptimal sensor and which contribute heavily to the mean-square error	174
4.33 Comparisons of probability of correct classification for several sensors for Williams County, May 8, 1977	175
4.34 Comparisons of probability of correct classification for several sensors for Williams County, June 29, 1977	176
4.35 Comparisons of probability of correct classification for several sensors for Williams County, August 4, 1977	177
4.36 Comparisons of probability of correct classification for several sensors for Finney County, September 28, 1976	178
4.37 Comparisons of probability of correct classification for several sensors for Finney County, May 3, 1977	179
4.38 Comparisons of probability of correct classification for several sensors for Finney County, June 26, 1977	180
 Appendix	
Figure	
B1. Sample program for assembling sample functions using the GSPEC processor in EXOSYS	222
B2. Spectral parameter design system data storage format	223
C1. Correlation coefficients vs wavelength for Williams County, May 8, 1977	228
C2. Correlation coefficients vs wavelength for Williams County, June 29, 1977	229
C3. Correlation coefficients vs wavelength for Williams County, August 4, 1977	230

LIST OF FIGURES(cont.)

Appendix Figure		Page
C4.	Correlation coefficients vs wavelength for Finney County, September 28, 1976	231
C5.	Correlation coefficients vs wavelength for Finney County, May 3, 1977	232
C6.	Correlation coefficients vs wavelength for Finney County, June 26, 1977	233

GLOSSARY OF SYMBOLS

A	set of algorithms used in a processor
α	system parameter
β_n	sequence of sums of basis functions with coefficients
C_i	class i
δ_{ij}	Kronecker delta
$E\{ \}$	mathematical expectation
ϵ_r	mean-square representation error
ϵ_o	global performance criterion
ϵ_q	mean-square error due to quantization
ϵ_s	mean-square error due to spectral sampling
$\phi_i(\lambda)$	basis function, eigenvector
γ_i	eigenvalue
$K(\lambda, \xi)$	covariance function
K	covariance matrix
L	number of samples taken from the spectral interval
L_2	Hilbert space of square-integrable functions
L_{σ_2}	Hilbert space of σ -measurable functions
L_s	subset of Hilbert space of σ -measurable functions from a stratum

λ	wavelength parameter
Λ	spectral interval
$m(\lambda)$	mean function
N	number of terms or basis functions used in the expansion
N_s	number of sample functions used to estimate a parameter
P_i	a priori probability of class i
$p(x)$	probability density function
$\psi_i(\lambda)$	basis function
Q	number of quantization levels
S_i	stratum
S_o	collection or set of strata over which a sensor will be used
s	a point in the stratum
T	minimum acceptable expected mean-square representation error
$\theta_i(\lambda)$	sampling function
u_i	sample values from spectral sampling
U	vector of sampled values
$x(\lambda)$	spectral response function
x_i	measurement coefficient for a spectral measurement
ξ	wavelength parameter

W weight matrix
 $w(\lambda)$ weight function
 z output of pattern recognition system

ABSTRACT

The task of selecting the best set of spectral channels is vital to the design of multispectral remote sensor systems. It would be desirable to choose a sensor design such that the entire pattern recognition system performs in an optimal manner. In order to choose a design which will be optimal for the largest class of remote sensing problems, a method is developed which attempts to represent the spectral response function from a scene as accurately as possible. The performance of the overall recognition system, then, is studied relative to the accuracy of the spectral representation. The spectral representation is only one of a set of five interrelated parameter categories which also includes the spatial representation parameter, the signal-to-noise ratio, ancillary data, and information classes.

The spectral response functions observed from a stratum are modeled as a stochastic process with a Gaussian probability measure. The criterion for spectral representation is defined by the minimum expected mean-square error. The

sensor is modeled as a set of basis functions such that the output approximates the input by a linear combination of the basis functions suitably weighted by a sequence of coefficients. The optimum set of basis functions with respect to the mean-square error criterion is given by the solutions to the Karhunen-Loeve expansion. The development of the Karhunen-Loeve expansion was generalized to include a weight function such that each point in the spectral interval could be assigned a weight corresponding to its importance. The computation of the optimum set of basis functions is incorporated into an analytical procedure that seeks to design practical sensors, comparing their performance against the optimal design.

The five parameter categories are discussed with regard to their effect on the pattern recognition system performance. The usefulness of the graph of the recognition system performance as a function of spectral representation is introduced.

A software system is developed to test and evaluate this method using field measurements data taken from two locations on three different dates each. Four different weight functions are evaluated. The effect of sample size on the evaluation of a data set is demonstrated. For each stratum the first few eigenvectors are plotted, and the mean-square error and probability of correct classification are evaluated. The graphs of probability of correct

classification vs. expected mean-square error allow the study of the relationship between classification performance and spectral representation. One can also study the dimensionality of the observation space relative to representation and performance.

The procedure is demonstrated to be a valuable tool for the design of sensors for the limited collection of data. The value of the weighted Karhunen-Loeve expansion is demonstrated. The performance of several suboptimal sensors are compared with the optimal design. A proposed suboptimal sensor is designed which demonstrates superior performance in representation accuracy and classification accuracy over the other suboptimal sensors. It is shown that spectral sampling should be done using spectral channels which have a smaller bandwidth, particularly in the red part of the visible region, than are currently being used on operational sensor systems.

CHAPTER 1. INTRODUCTION

Earth observational remote sensing has emerged as a prominent technology in the last two decades. Important developments in sensor technology, computer systems, pattern recognition theory, and image processing techniques have brought the remote sensing state-of-the-art to the point where it is a powerful tool for studying earth resources. With the launching of the Landsat satellites and advanced automated processing of the image data, worldwide monitoring of the earth's surface for locating and utilizing natural resources is now a reality.

1.1 The Pattern Recognition System

A basic tool for remote sensing is pattern recognition. From a systems perspective the components of a pattern recognition system can be placed into three distinct blocks - the scene, the sensor, and the processor (Figure 1.1). The scene includes everything in front of the sensor. The information in the scene is contained in the spectral, spatial and temporal variations of the electromagnetic energy that is either reflected by or emitted from the earth's surface and passes through the atmosphere. The sensor measures the received electromagnetic energy and prepares the measurements

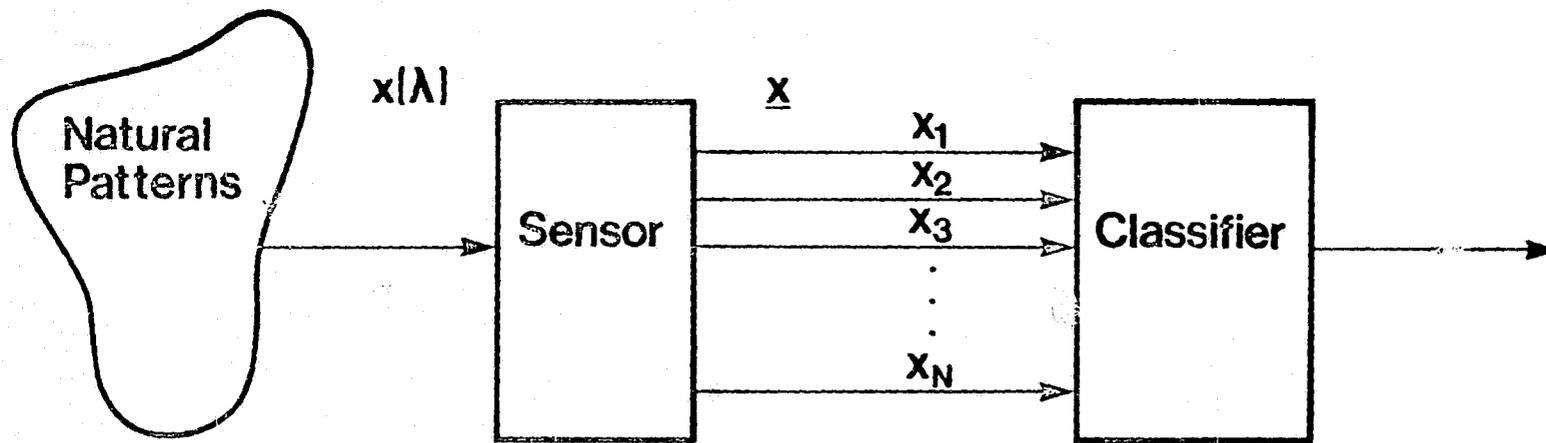


Figure 1.1 Pattern recognition system.

for transmission to the processor. The processor digitally implements a set of algorithms for classification and image processing as required by the user or analyst.

In this research we focus on the sensor subsystem and develop an analytical technique for selecting certain parameters for the sensor design. Because of the interrelationship with other parts of the pattern recognition system, the sensor design problem will be considered as a part of the integrated overall system design problem. That is, sensor design choices will be made on the basis of overall system performance. Therefore, we begin with a more detailed discussion of the parts of the system and how they interface with each other.

1.1.1 The Scene

A distinctive characteristic of the scene is that it is not under the control of the system designer or the analyst. In fact, the intent of remote sensing is to observe and learn as much about the scene as possible without modifying it or affecting it in any way.

For current earth observational remote sensing problems, the information bearing signal is the spectral response function $x(\lambda, r, s, t)$. The parameters of this function are the wavelength, λ , the spatial coordinates, r and s , and time, t . Historically, the desired information has been extracted primarily from the spectral variations (Holmes and MacDonald, 1969) to which we will limit ourselves, here,

although significant progress is being made in extracting information from spatial (Kettig and Landgrebe, 1976; Haralick et al, 1973; Wiersma and Landgrebe, 1976) and temporal variations (Swain, 1978). A typical spectral response function for green vegetation at a fixed location and time is shown in Figure 1.2. The interval of interest, Λ , typically includes the visible and infrared regions of the electromagnetic spectrum from 0.4 micrometers to 2.4 micrometers.

The scene is very dynamic and complex. Changes in sun angle, atmospheric conditions, climate, cover type, and a variety of other variables can produce significant changes in the spectral response. Instead of trying to account for each of the variables that affect the spectral response, we choose to model the scene as a stochastic process. The complete characterization of this process model is not known a priori. In order to obtain this knowledge, one observes the scene over a period of time, an area of space and an interval of the spectrum and estimates the parameters from the observations which are necessary to complete the characterization.

It is generally necessary to group the observations taken from the scene into classes. For purposes of classifying the data into distinct classes it is required that the class list have the following properties simultaneously (Landgrebe, 1978):

- Each class must be of interest to the user, i.e. of informational value

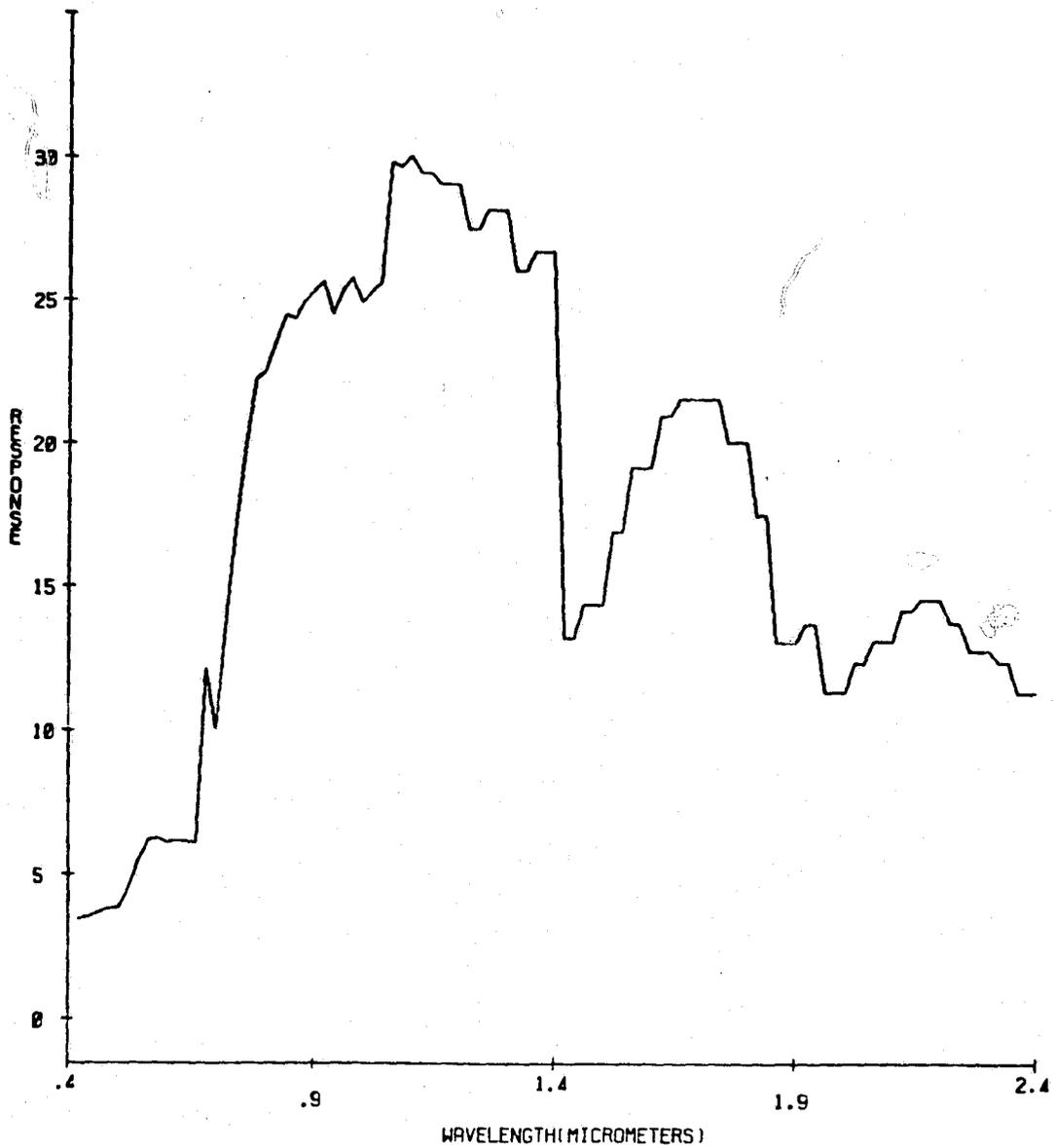


Figure 1.2 Spectral response function for mature wheat collected on August 4, 1977 over Williams County, North Dakota.

- The classes must be separable in terms of the features available
- The list must be exhaustive, i.e., there must be a class to which it is logical to assign each pixel in the scene

1.1.2 The Sensor

The function of the sensor is to transform the continuous parameter functions $x(\lambda)$ into a finite number, N , of measurement values (x_1, x_2, \dots, x_N) called features. Ideally, the sensor would be under the control of the system user who could then optimize the sensor and the processor for a specific application. However, in practice the sensor is a complex, expensive system which is designed infrequently. Control over the sensor, consequently is the responsibility of the system designer rather than the user. The system designer cannot optimize the sensor for a particular location, time, and application but must create a single instrument which must serve a broad spectrum of users and applications over a number of years.

A basic sensor is shown schematically in Figure 1.3. The system components can be placed into five blocks - the collector (typically a set of optics) which collects the electromagnetic energy over the spectrum of interest, the scanning mechanism which controls the pointing of the collector, a spectral dispersing device, the detectors which convert electromagnetic energy into electrical signals, and the signal processing unit. The sensor is mounted on

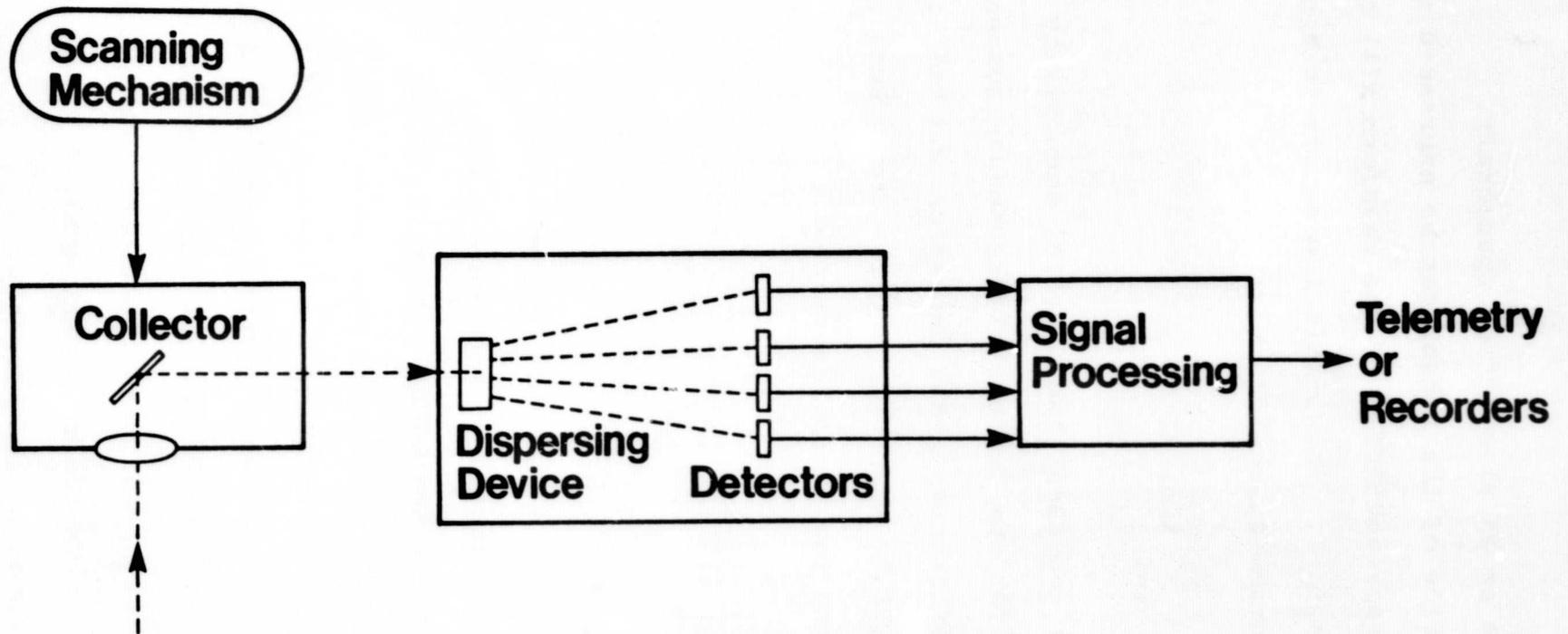


Figure 1.3 Sensor system block diagram.

a platform such as an aircraft or spacecraft.

The operation of the sensor can be expressed mathematically as the representation of the waveform $x(\lambda)$ by a set of functions $\{\phi_i(\lambda)\}$. The original waveform is approximated by the series expansion

$$x(\lambda) \approx \sum_{i=1}^N x_i \phi_i(\lambda) \quad (1.1)$$

where the $\phi_i(\lambda)$ represent the spectral sensitivity (as a function of λ) for one feature of the sensor system and the x_i are the coefficients in the expansion and the measurement values which will be used by the processor. Each x_i is obtained by using the linear functional

$$x_i = \int_{\Lambda} x(\lambda) \phi_i(\lambda) d\lambda \quad (1.2)$$

1.1.3 The Processor

The measurement values (x_1, x_2, \dots, x_N) from the sensor become the input to the processor which typically contains a digitally implemented classification algorithm. A comprehensive list of all the processors that have been implemented would require a monumental effort to compile. Texts such as Nilsson (1965), Fukunaga (1972), and Duda and Hart (1973) describe some basic classifiers which may be adapted for specific applications. The point, here, is that the system designer and the analyst have great flexibility in choosing the processor.

An important step in the design of a classifier is the training phase. After the classification algorithm has been selected, the parameters required by the algorithm must be determined in order to obtain good performance. The process of selecting these parameters is the training phase. To train the classifier a set of (presumably correctly) class-labeled samples from the scene are selected from which the necessary parameters can be computed.

During the design procedure it is important to specify the performance of the system which implies that a measure of the performance must be defined. The global performance criterion, ϵ_0 , is a function of many system parameters.

$$\epsilon_0 = f(\alpha) \quad (1.3)$$

The list of system parameters is indicated by α . This function is so complex that the straightforward analysis and subsequent optimization of the whole system with respect to ϵ_0 is not trivial. What seems more appropriate is to list the parameters in order of importance and investigate the effect of each on the performance.

Landgrebe (1978) has listed five general parameter categories which affect the system design. These are:

- Spectral representation
- Spatial representation
- Signal-to-noise ratio
- Ancillary data
- Information classes

It is important to recognize that these parameters are inter-related; hence, a change in one parameter may affect the value of one or more of the other four. In this research we will be concerned primarily with the spectral representation; however, the other four parameters will play a necessarily important role in the analysis.

1.2 Previous Approaches to Sensor Design

There have been basically three approaches to selecting spectral bands for multispectral scanner design. They are 1) in-depth studies of physical considerations, 2) empirical methods, and 3) simulators. All three of these approaches have contributed to our knowledge of the scene and to the design and development of present-day scanner systems.

Important physical considerations which have been studied are atmospheric effects and the interaction of light with various cover types. Atmospheric effects include scattering and absorption by water vapor, carbon dioxide, and ozone (Korb, 1969; Hulstrom, 1974). By evaluating the transmittance of the atmosphere over the spectral interval of interest, one can eliminate certain portions of the interval, since little or no energy will reach the sensor. Scattering effects are less pronounced but are important for consideration.

Studies have been done to investigate the interaction of electromagnetic radiation with plant leaves to determine

regions of the spectrum which will be useful for identifying vegetation and determining plant stress (Harnage, 1975; Gates et al, 1971). On a larger scale the interaction of light with a plant canopy has been studied which takes into consideration the effects of leaf size, plant size, and plant density (Colewell, 1974). Similar studies have been done with soils (May and Peterson, 1975; and Montgomery, 1976) and with water temperatures (Bartolucci, 1977). A typical procedure for these studies is to take measurements with a spectroradiometer on a restricted set of information classes over the entire spectrum. For a single observation a single spectral response function is recorded. The average response is taken over a small number of samples and conclusions are drawn from the average. It is important to note that over a collection of these spectral response functions, the functions vary significantly about the mean. Furthermore this variation is potentially information bearing. This information is lost if one considers only the mean response function.

The second approach is empirical in that a scanner with many spectral bands is constructed, and the selection of the bands is done experimentally. Examples of experimental scanners which have been constructed are the Michigan scanner (Hassel et al, 1974) and the MSDS scanner (Zaitzeff et al, 1971). The spectrum is sampled using on the order of 10-30 spectral channels (12 for Michigan scanner 24 for

MSDS) which are thought to be of interest. Data is collected using the scanner, and processing is performed to evaluate the channels over a variety of scenes. Some examples of empirical studies which have been done are Landgrebe et al (1977) with agricultural cover types, Coggeshall and Hoffer (1973) with forest covers, and Vincent and Thompson (1972) with geological applications. This empirical approach has the advantage of retaining the information in the variations about the mean since a large number of samples can be collected. However, the spectral sampling is crude and incomplete for representing the whole spectrum.

Simulators have been developed to generate typical spectra according to a scene model. The artificial spectral response functions can then be used to evaluate spectral bands. A system which has been set up to simulate multi-spectral data is described by Malila et al (1977). At this time there is not sufficient understanding of the scene to be able to develop and use accurate models.

One additional research effort due to Wiswell (1978) which differs from the previous approaches deserves mentioning. The purpose was to extract information from a scene using the entire spectral interval. The criterion of average mutual information was proposed which is a measure of the reduction in uncertainty about the scene after the observation has been made. This information theoretic technique was used to evaluate spectral bands on the basis

of maximum average mutual information. An autoregressive stochastic process model was used for the scene. Considerable effort was expended in developing and testing this model, the parameters of which were then used to compute average mutual information. While spectral bands were evaluated on the basis of the information criterion, the relationship between average mutual information and some global performance criterion such as classification accuracy was not demonstrated.

In this research it will be desirable to incorporate the positive features of past approaches and build on the knowledge that has been gained through them. We would like to extract information from the entire continuous spectral interval of interest rather than from the coarse sample of the interval provided by experimental scanners. A large collection of spectral response functions taken from field measurements of the scene will be utilized in order to take into consideration the variability of the data over the scene as well as the average values. A parametric stochastic model will be assumed which has been well studied. The complete characterization will be learned from observations of the real data.

An important consideration is the choice of criterion upon which the sensor system will be designed. The choice of a global performance criterion, such as probability of correct classification, seems attractive, since the overall

performance of the pattern recognition system is ultimately what we wish to optimize. Usually one would like to maximize the probability of correct classification; however, in most cases the integral involved is too complex to admit an analytic solution. Efforts have been made in this direction primarily drawing on results from the literature of feature selection. There have been three basic approaches along this line: 1) optimization of a separability measure, 2) discriminant analysis, and 3) principal component analysis.

Separability measures of the statistical distance between two class distributions are numerical quantities which are simpler to compute than classification performance and which provide bounds on the performance. The divergence (Marill and Green, 1963) and the Bhattacharyya distance (Kailath, 1967) are two well-known examples of separability measures. Wacker and Landgrebe (1971, Table 2.4.2) provide a listing of many of the separability measures that have been proposed in the literature. The approach is to select the set of spectral channels which is optimal with respect to the separability measure. Typically, a search procedure is used to arrive at the best choice of spectral channels (Tou and Heydorn, 1967; Whitsitt and Landgrebe, 1977; Kadota and Shepp, 1967; and Caprihan and deFigueiredo, 1970). Note that Kadota and Shepp (1967) and Caprihan and deFigueiredo (1970) are extracting information from continuous

functions. Since the separability measure provides at best only a bound on the classification performance, it cannot be guaranteed that the channels which optimize the separability measure will necessarily optimize the system performance.

In discriminant analysis one attempts to find some measure of the ratio of the between class separability to the within-class separability (Fukunaga, 1972; Foley and Sammon, 1975). The spectral channels are selected to maximize this ratio. One can observe intuitively that maximizing the ratio would improve the performance; however, it cannot be guaranteed that the chosen set is optimum with respect to the global performance criterion.

The method of principal components is a statistical procedure which reduces the number of variables to be analyzed to a manageable number (Anderson, 1962; Dempster, 1969). Principal vectors are found such that the variable in the first principal vector has maximum statistical variance, and so forth. A variation on principal component analysis which has found considerable application is that of canonical correlation (Dempster, 1969). It cannot be assured that once the principal vectors have been selected, the global performance criterion is optimized.

Although each of the feature selection procedures described above has been demonstrated to be practical in spite of the lack of a tight relationship to the global criterion, the approach that is proposed here will take

a different direction. Once the sensor has been designed, built, and placed into service, it should perform optimally over all possible scenes for any possible choice of processor. The difficulty with the methods for optimizing the global criterion is that one does not know the specific processor and set of classes of a problem at the time the sensor is designed. Optimizing the choice of basis functions with respect to a global criterion for a specific set of classes in general may yield poor results with the same basis functions on a different set of classes. Furthermore, the global performance criterion was described as being a complex function of many parameters (1.3). If we choose as our criterion some measure of the quality with which the output of the sensor represents the input, we can optimize the criterion while holding parameters from the other four categories fixed. The relationship between the spectral representation criterion and the global performance criterion can be evaluated for typical remote sensing problems.

1.3 Present Investigation

In Chapter 2 a procedure is developed to analytically select spectral channels for a sensor system. The collection of spectral response functions makes up the stochastic process. A representation technique based on the Karhunen-Loeve expansion which minimizes the criterion of mean-square representation error is developed. This technique is

generalized to include a priori weighting information which may be available. This weighting method has been termed the weighted Karhunen-Loeve expansion.

The Karhunen-Loeve expansion is attributed to Karhunen (1947) and Loeve (1963) and is used extensively in the stochastic process literature as a technique for representing stochastic processes (Davenport and Root, 1958; Wong, 1971). In the pattern recognition literature the Karhunen-Loeve expansion historically has been used as a feature selection technique (Watanabe, 1965; Chien and Fu, 1967; Fu, 1968; Fukunaga, 1970; Kittler and Young, 1973).

The parameters and their influence on the global performance criterion are discussed in Chapter 3. The principal parameter in this research is the spectral representation; hence, the relationship between the spectral representation parameter and the probability of correct classification is developed. The ancillary data, information classes, spatial representation, signal-to-noise ratio, and the interrelationships between these parameters are also discussed.

An experimental software system which implements the procedure that was developed in Chapter 2 is described and evaluated in Chapter 4. Results from tests of the system are presented and discussed.

In Chapter 5 some conclusions from the results are presented and suggestions are made for further work.

CHAPTER 2. SCENE REPRESENTATION AND SPECTRAL PARAMETER DESIGN

In this chapter an analytical procedure is developed to perform the spectral parameter design for a sensor system capable of operating as an integral part of any potential pattern recognition system. Due to the complexity of the scene, a stochastic process model is used to describe the scene. The theory necessary to support the procedure is developed for the case where the spectral response functions are square-integrable functions of the continuous parameter λ . Due to practical consideration for measuring real data in the field and performing computations on a digital computer, a discrete approximation is developed, and the potential error due to the approximation is discussed.

2.1 The Analytical Procedure

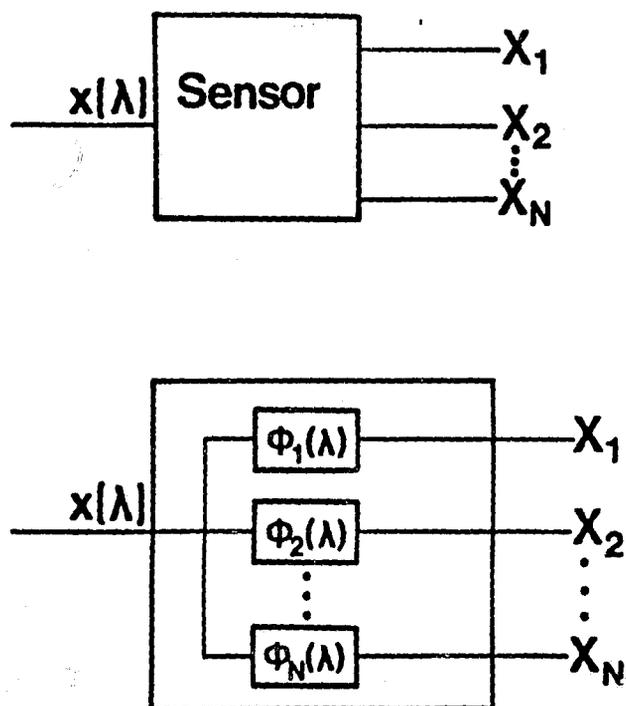
Consider a pattern recognition system where the scene which is being observed by the sensor is some portion of the earth's surface S_0 . It may be desirable to design a sensor such that S_0 is some subarea, for example, the land surfaces or a particular nation within its territorial boundaries. The area defined by the geographical boundaries of S_0 can be subdivided into areas called strata. We define

a stratum $S \subset S_0$ as the largest contiguous set of points $\{s \in S\}$ which can be classified to an acceptable accuracy with a single training of the pattern recognition algorithm.

The sensor model will be a set of basis functions $\{\psi_i(\lambda)\}$ on the interval Λ (Figure 2.1). These functions are essentially filters which have weighted passbands in differing portions of the spectral interval. The approximation of a function $x(\lambda)$ by a set of four rectangular basis functions is illustrated in Figure 2.2.

The processor for the pattern recognition system will be denoted by $P(A, z, \epsilon_0)$, where A represents the set of algorithms used in the processor, z is the output of the system, and ϵ_0 is the system performance criterion with respect to z . The set A may include feature selection and classification algorithms. The output z may be a map, a table or some other presentation of the desired information.

In order to define a remote sensing problem, the analyst decides upon an objective. Depending on the objective the analyst will specify the components of the pattern recognition system S' , $\{\psi_i(\lambda)\}$, and $P(A, z, \epsilon_0)$. Quite often the objective dictates which subset, S' , of S_0 will be used. As described in Chapter 1, the sensor $\{\psi_i(\lambda)\}$ is designed infrequently and once put into service remains fixed. The output z is based upon what information is desired from the scene to achieve the objective.



$$\hat{x}(\lambda) = X_1 \Phi_1(\lambda) + X_2 \Phi_2(\lambda) + \dots + X_N \Phi_N(\lambda)$$

Figure 2.1 Sensor system model.

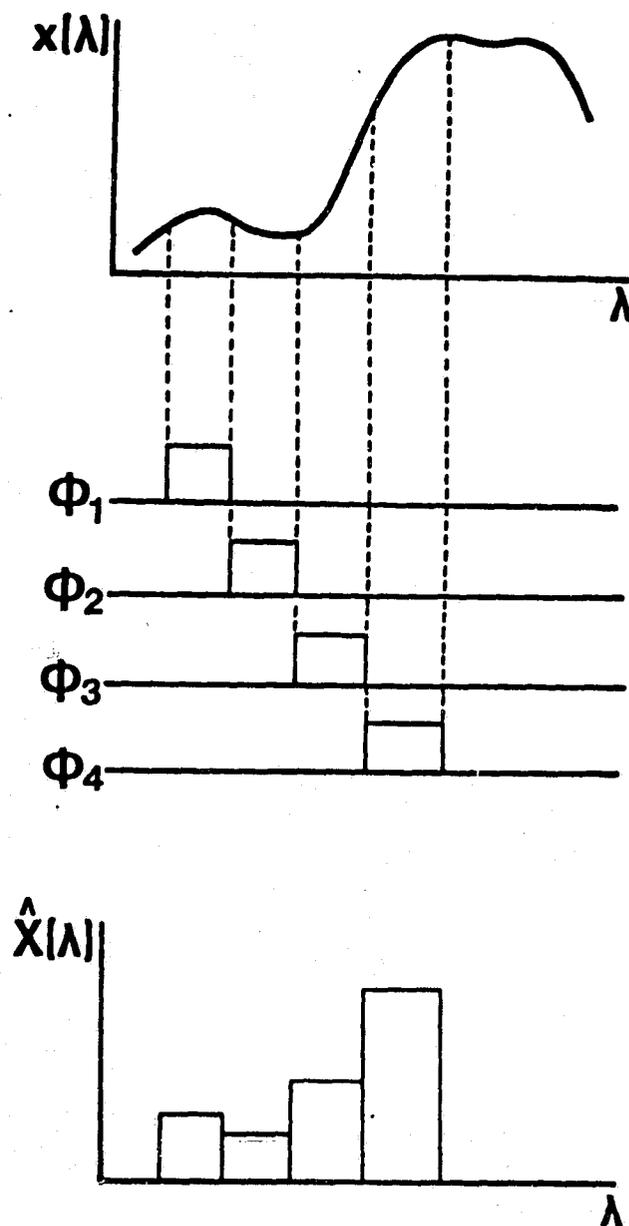


Figure 2.2 Approximation of the spectral response function by a set of four basis functions.

The ultimate goal before us is to select a sensor design $\{\psi_i(\lambda)\}$ which can be used on any subset S' of S_0 and will provide, as nearly as possible, optimal performance for any choice of processor $P(A, z, \epsilon_0)$.

To achieve the design goal an analytical procedure is set forth which incorporates the design of a theoretically optimal sensor against which the performance of candidate-practical sensors can be compared. The following procedure is proposed (see Figure 2.3):

1. Based on the intended use of the sensor system, the collection of strata comprising S_0 is specified. Because of the infinite number of possible strata in S_0 , only a finite number G of subsets $\{S_i\}$ which are representative of the entire collection S_0 will be used to evaluate the sensors.
2. An initial candidate sensor system is specified by defining a set of basis functions $\{\psi_i(\lambda)\}$. At appropriate steps in this procedure the set of basis functions may be modified to improve the performance.
3. In steps 4 through 7 each stratum S_i , $i=1, 2, \dots, G$ will be considered in sequence. If it is necessary at any stage to modify $\psi_i(\lambda)$, the sequence should be repeated to insure that the desired performance is obtained over all of S_0 .

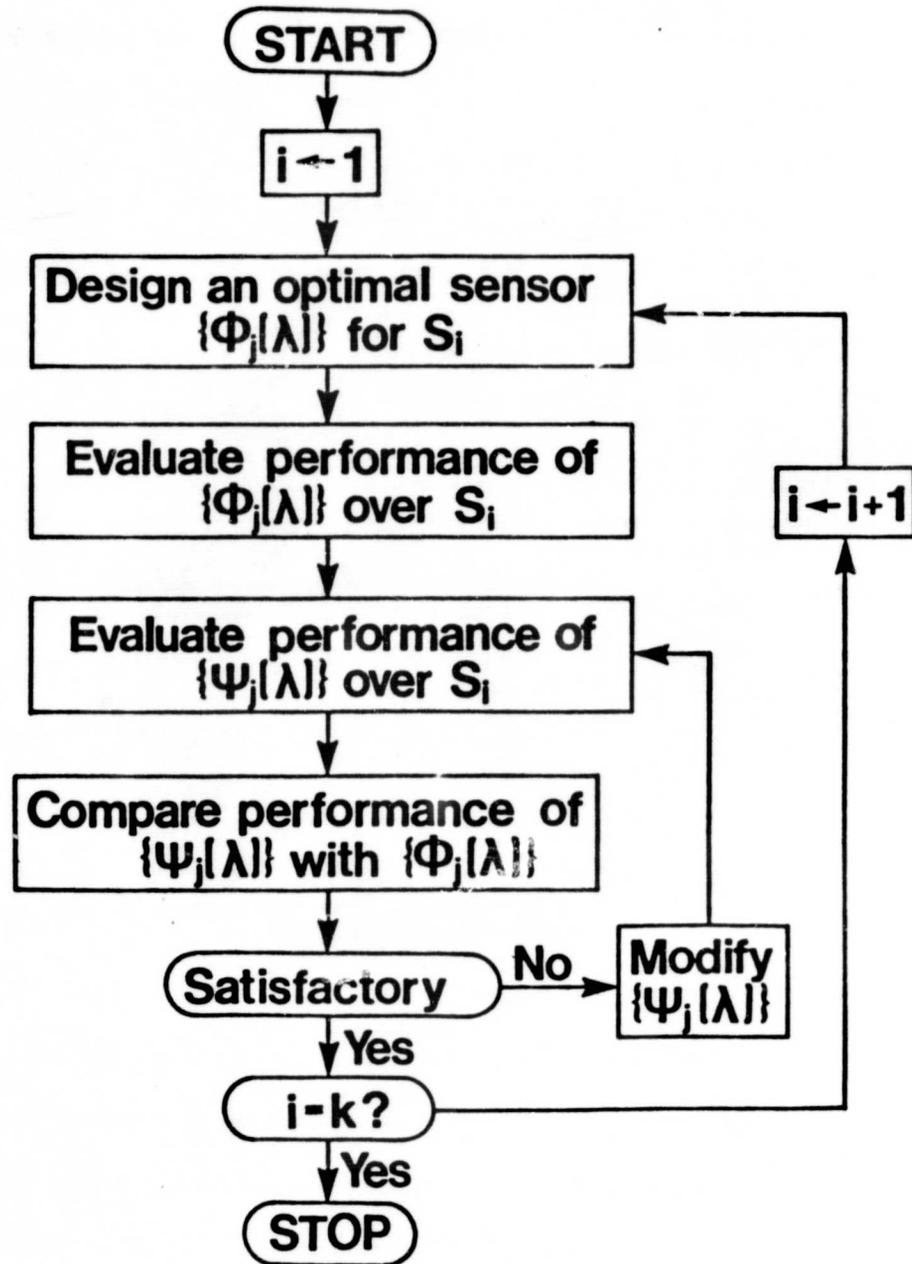


Figure 2.3 Flowchart of the design procedure.

4. For S_i design a sensor which will be optimum with respect to some criterion for any possible choice of processor. This optimal sensor $\{\psi_i(\lambda)\}$ will serve as a standard by which one can compare the performance of the candidate system.
5. Evaluate the performance of the optimal sensor based on the criterion of optimality over the stratum S_i .
6. Evaluate the performance of the candidate sensor $\{\psi_i(\lambda)\}$ over the same stratum using the criterion.
7. Compare the performance of the candidate sensor with the optimal design relative to the criterion. If the performance of the candidate sensor $\{\psi_i(\lambda)\}$ is very nearly the same as the optimum, then, the proposed sensor is adequate for the stratum under consideration. In this case the next stratum in the sequence is fetched and we return to step 4. In the event that all of the strata have been used the procedure halts. If the candidate system's performance is substantially below that of the optimum, it will be necessary to modify our choice of the set $\{\psi_i(\lambda)\}$ and return to step 6. The set of optimal basis functions $\{\phi_i(\lambda)\}$ can be used to provide an aid for modifying $\{\psi_i(\lambda)\}$.

The critical step in this procedure is the design of the optimum sensor, and will be the principal step to which this research will be addressed. The criterion for

optimality is an important quantity and must be dealt with carefully. An optimality criterion has the dual role of providing the measure of performance which will be optimized as well as providing a standard for comparison of suboptimal systems (Middleton, Sect. 2.3.4, 1960).

The optimal sensor system design will be optimum in the following sense. If one has the entire function $x(\lambda)$ at his disposal, a processor which is optimum with respect to a global performance criterion ϵ_0 may be designed. If $\hat{x}(\lambda)$ is the approximation by the sensor to the waveform $x(\lambda)$, then a fidelity criterion is defined by

$$\epsilon_r = \int_{\Lambda} f(x(\lambda) - \hat{x}(\lambda)) d\lambda \quad (2.1)$$

The condition for optimality requires that the original waveform be reconstructed with arbitrarily small ϵ_r .

There are several possible choices for the function $f(\cdot)$ in 2.1. It is desirable to choose a function for which there is a greater cost for large errors than for small errors. Since $x(\lambda)$ will be required to be a square-integrable function, a natural choice which satisfies the requirements for the cost of making an error is the function $f(x) = x^2$. Equation 2.1, then, becomes

$$\epsilon_r = \int_{\Lambda} [x(\lambda) - \hat{x}(\lambda)]^2 d\lambda \quad (2.1a)$$

2.2 The Stochastic Process

An experiment is defined as the observation of a point s in the stratum S . Each point $s \in S$ is mapped into a spectral response function $x(\lambda)$.

$$X: s \rightarrow x(\lambda) \quad (2.1)$$

The function $x(\lambda)$ is a real-valued function of the continuous parameter λ .

Let $\sigma(\lambda)$ be a non-decreasing function of bounded variation which is absolutely continuous. Construct a σ -measure on the interval Λ such that

$$\frac{d\sigma(\lambda)}{d\lambda} = w(\lambda) \quad (2.2)$$

We require that $x(\lambda)$ belong to the Hilbert space L_{σ_2} of all σ -measurable functions for which the Lebesgue-Stieltjes integral

$$\int_{\Lambda} [x(\lambda)]^2 d\sigma(\lambda) \quad (2.3)$$

exists. The inner product which generates the metric for this space is

$$(x, y) = \int_{\Lambda} x(\lambda)y(\lambda) d\sigma(\lambda) \quad (2.4)$$

(Akhiezer and Glazman, 1961). The norm is given by

$$\| X \| = (x, x)^{1/2} = \int_{\Lambda} [x(\lambda)]^2 d\sigma(\lambda)$$

(Kolmogorov and Fomin, 1957).

Consider the subset L_S of L_{σ_2} which consists of the set of all possible spectral response functions which may be mapped from points in the stratum. The members of this subset $\{x(\lambda), x \in L_S\}$ form an ensemble (Figure 2.4). This ensemble together with the probabilities of occurrence associated with the functions that belong to L_S specify a stochastic process (Papoulis, 1965; Doob, 1963; Gikhman and Skorokhod, 1969).

Crane et al (1972) and others have shown that for remote sensing applications this stochastic process may be assumed to have a Gaussian probability measure. The Gaussian assumption is attractive because its mathematics are well-studied and tractable and because of its robustness. Robustness implies that good estimates of the density function can be obtained with a relatively small number of training samples and that statistical procedures on the process yield good results even for some non-Gaussian processes (Lachenbruch et al, 1972). An important property of a Gaussian process is that every linear function of $x(\lambda) \in L_{\sigma_2}$ is a Gaussian random variable (Van Trees, 1968). Also, a Gaussian random variable is completely characterized by its first and second moments. The first moment or mean function of the process is denoted by

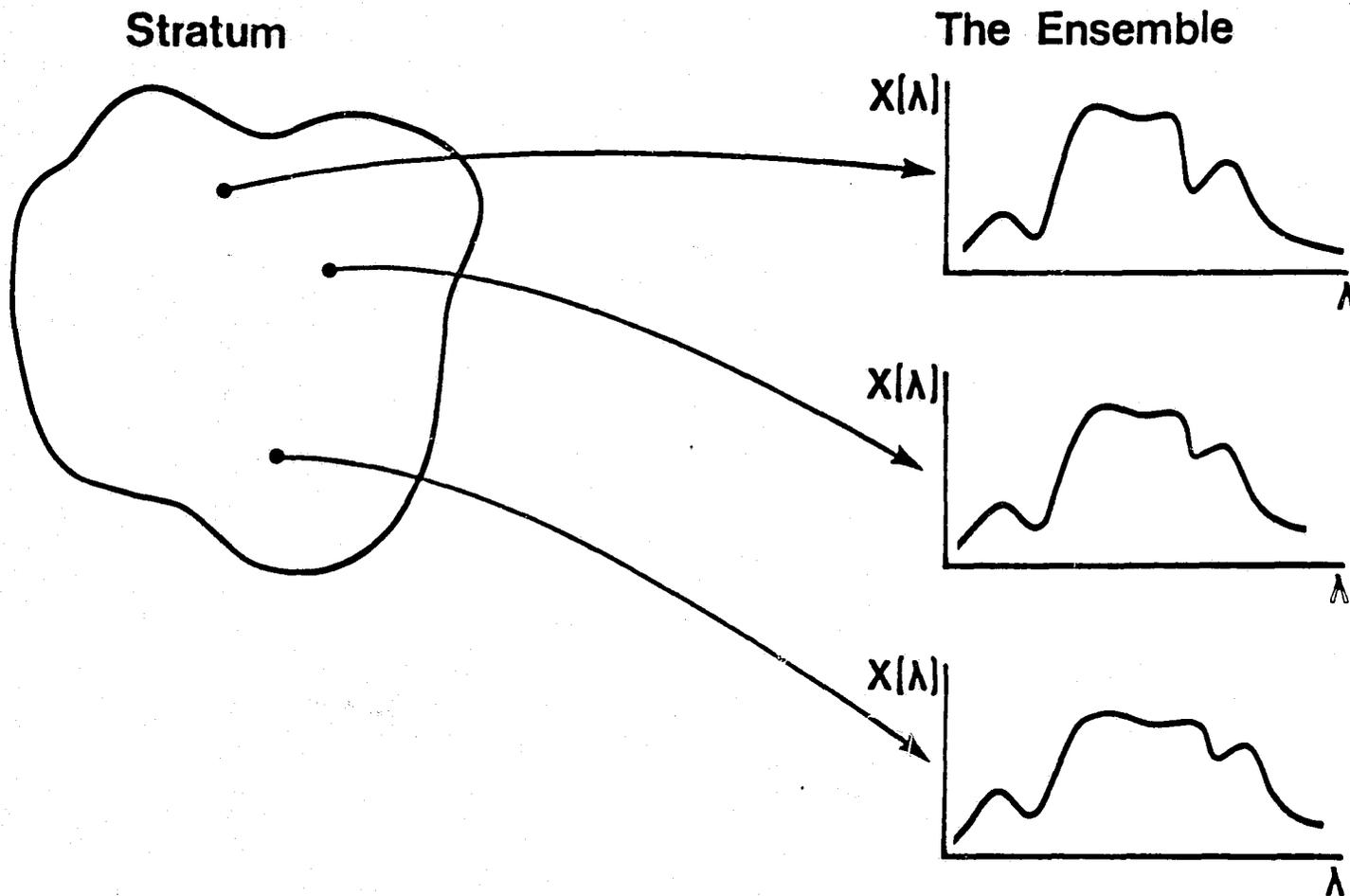


Figure 2.4 Realization of a stratum as the ensemble of spectral sample functions.

$$m(\lambda) = E\{x(\lambda)\} \quad (2.5)$$

and the second moment or covariance function is denoted by

$$K(\lambda, \xi) = E\{[x(\lambda) - m(\lambda)] [x(\xi) - m(\xi)]\} \quad (2.6)$$

where $E\{ \}$ denotes ensemble expectation. The covariance is assumed to be continuous.

2.3 Representation of the Stochastic Process

The criterion that has been proposed for designing the spectral representation parameter for the sensor system is based on the ability of the sensor to represent functions belonging to the stochastic process. Of the possible techniques for representation of stochastic processes (Wong, 1971), it would be desirable to choose a method which bears a close relationship to the physical model of the sensor. A well-studied technique is to represent the continuous parameter stochastic process $\{x(\lambda), \lambda \in L_S\}$ by a sequence of random variables which are the coefficients of a set of basis functions in a series expansion. The basis functions correspond to the basis functions described for the sensor model in Figure 2.1. That such a representation is possible without loss of information was shown by Bharucha and Kadota (1970).

Consider the linear Hilbert space L_{σ_2} and let $\{\phi_i(\lambda)\}$ be an infinite linearly independent set of functions

belonging to L_S . For an arbitrary function $x(\lambda) \in L_S$ we can associate the infinite series

$$x(\lambda) \approx x(\lambda) = \sum_{i=1}^{\infty} x_i \phi_i(\lambda) \quad (2.7)$$

Note that for the series expansion the continuous parameter function $x(\lambda)$ is transformed to a point in the Hilbert space L_{σ_2} whose coordinates are given by the vector of coefficients $[x_1, x_2, \dots]^T$

Without loss of generality the set $\{\phi_i(\lambda)\}$ will be taken to be orthonormal; that is,

$$\begin{aligned} (\phi_i(\lambda), \phi_j(\lambda)) &= \int_{\Lambda} \phi_i(\lambda) \phi_j(\lambda) d\sigma(\lambda) & (2.8) \\ &= \int_{\Lambda} \phi_i(\lambda) \phi_j(\lambda) w(\lambda) d\lambda \\ &= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \end{aligned}$$

If the set $\{\phi_i(\lambda)\}$ is not orthonormal to begin with, it can be orthonormalized by the Gram-Schmidt procedure (Courant and Hilbert, 1953). That such sets exist in Hilbert spaces has been demonstrated by the construction of sets such as complex sinusoids, Legendre polynomials, Tschebycheff polynomials and others.

The coefficients x_i are the Fourier coefficients defined by

$$\begin{aligned} x_i &= (x(\lambda), \phi_i(\lambda)) & (2.9) \\ &= \int_{\Lambda} x(\lambda) \phi_i(\lambda) d\sigma(\lambda) \\ &= \int_{\Lambda} x(\lambda) \phi_i(\lambda) w(\lambda) d\lambda \end{aligned}$$

For a given set of basis functions the set of coefficients which minimizes the mean-square error between each function and its approximation are the Fourier coefficients (Courant and Hilbert, 1953). Note that the set of coefficients $\{X_i\}$ can be treated as a vector $\underline{X} = [X_1, X_2, \dots]^T$. This vector representation of the function $x(\lambda)$ is a motivating factor in choosing this method of representing the stochastic process, since the vector representation provides an equivalent mathematical model to the physical sensor.

Since the Hilbert space L_{σ_2} has already been defined, the corresponding definitions for the inner product and the metric follow. It is possible, therefore, to talk about a set $\{\phi_i(\lambda)\}$ which is complete in L_{σ_2} and about the convergence of the sequence

$$\beta_n(\lambda) = \sum_{i=1}^n x_i \phi_i(\lambda) \quad (2.10)$$

to the function $x(\lambda)$. Convergence in E_S is convergence in the mean. If $\beta_n(\lambda)$ converges to $x(\lambda)$, then

$$x(\lambda) = \lim_{n \rightarrow \infty} \beta_n(\lambda) \quad (2.11)$$

where l.i.m. is defined as

$$\lim_{n \rightarrow \infty} \left[\int_{\Lambda} \left[x(\lambda) - \sum_{i=1}^n x_i \phi_i(\lambda) \right]^2 d\sigma(\lambda) \right] = 0$$

The problem of designing the optimal sensor becomes that of selecting the set of basis functions $\{\phi_i(\lambda)\}$ such that the series representation will be optimum with respect to the criterion. The criterion of minimum error in reconstructing a function is extended to the stochastic process where the expectation of the mean-square representation error is taken over the ensemble

$$E \{ \epsilon_r \} = E \left[\int_{\Lambda} [x(\lambda) - \hat{x}(\lambda)]^2 d\sigma(\lambda) \right] \quad (2.12)$$

We now propose a list of properties which would be desirable for the optimal design to have. Because it would be impractical to transmit an infinite or even a very large number of spectral channels over a data link to a processor as well as difficult for any processor to handle such volume, it is necessary that the representation of the signal space be characterized by a small number of dimensions. The series expansion provides a countable set;

hence, by truncating the series at some appropriate number of terms N a finite dimensional signal space is obtained. This finite dimensional space is only an approximation to the entire space L_S , but it is desired that the approximation be an adequate representation for L_S .

If we form the sequence $\{\beta_n\}$, where

$$\beta_n(\lambda) = \sum_{i=1}^n x_i \phi_i(\lambda) \quad (2.13)$$

it would be desirable that this sequence converge to $x(\lambda)$ in the mean-square sense. This convergence guarantees that the series can be made arbitrarily close to $x(\lambda)$ by increasing n . Another desirable property is that the convergence be rapid in the first few terms. One would expect that an increase in the number of terms in the expansion would reduce the representation error. It is desirable, though, that each additional term decrease the representation error by a maximum amount. A plot of the expected mean-square representation error as a function of the number of terms n would show a large decrease in the mean-square error for the first few terms with a considerably slower rate of convergence for higher order terms. If the expansion is truncated after N terms, the series

$$\beta_N(\lambda) = \sum_{i=1}^N x_i \phi_i(\lambda)$$

should represent the function $x(\lambda)$ with minimum expected mean-square representation error.

We would also like the representation in terms of the optimal basis functions to be complete in the following sense. Let T be the minimum acceptable expected mean-square representation error. If the representation by the series expansion for some finite number of terms is such that

$$E \{ \epsilon_r \} = E \left[\int_{\Lambda} [x(\lambda) - \beta_n(\lambda)]^2 d\sigma(\lambda) \right] < T \quad (2.14)$$

then the set of N basis functions will be complete in the N -dimensional subspace of L_S that has an expected error less than T .

The completeness of the set can be expressed in terms of the coefficients of the expansion. Squaring and integrating term-by-term the expression in 2.14 becomes

$$E \{ \epsilon_r \} = E \left[\int_{\Lambda} [x(\lambda)]^2 d\sigma(\lambda) \right] - \sum_{i=1}^{\infty} E \{ |x_i|^2 \}$$

since, $E \{ \epsilon_r \} \geq 0$

$$\sum_{i=1}^{\infty} E \{ |x_i|^2 \} \leq E \left[\int_{\Lambda} [x(\lambda)]^2 d\sigma(\lambda) \right] \quad (2.15)$$

Inequality 2.15 is Bessel's inequality and guarantees that the sum of the squares of the coefficients always converges. Furthermore, if there is equality, then Bessel's inequality becomes Parseval's equality

$$\sum_{i=1}^{\infty} E \{ |x_i|^2 \} = E \left[\int_{\Lambda} [x(\lambda)]^2 d\sigma(\lambda) \right] \quad (2.16)$$

and the set $\{\phi_i(\lambda)\}$ is said to be complete. The direction, here, is to find the complete set $\{\phi_i(\lambda)\}$ and use only those terms in $\{\phi_i(\lambda)\}$ which provide a good approximation ($E\{\epsilon_r\} < T$) in the mean-square sense.

To obtain the optimal set of basis functions $\{\phi_i(\lambda)\}$ some results from linear integral equation theory are required (Courant and Hilbert, 1953; Akhiezer and Glazman, 1961; Riesz and Sz.-Nagy, 1956; Lovitt, 1924; Tricomi, 1957; Ash, 1967).

The linear integral operator \mathbb{K} on L_S is defined by

$$\mathbb{K} x(\lambda) = \int_{\Lambda} k(\lambda, \xi) x(\xi) d\sigma(\xi) \quad (2.17)$$

where $k(\lambda, \xi)$ is the kernel of the operator. An operator is compact if for every bounded sequence of functions $\{X_n(\lambda)\}$, the sequence of functions $\mathbb{K} x_m(\lambda)$ has a convergent subsequence. A bounded operator is self-adjoint if

$$(\mathbb{K} x, y) = (x, \mathbb{K} y)$$

We now state a theorem and some consequences of that theorem which will determine the set of basis functions $\{\phi_i(\lambda)\}$

Theorem: If \mathbb{K} is compact and self-adjoint, then the solutions to the linear homogeneous integral equation

$$\gamma_i \phi_i(\lambda) = \mathbb{K} \phi_i(\lambda) \quad (2.18)$$

is a set of eigenfunctions $\{\phi_i(\lambda)\}$ with corresponding eigenvalues γ_i . The following statements can be made:

- The eigenvalues are real
- The eigenfunctions form a basis for the space L_S
- The eigenfunctions for distinct eigenvalues are orthogonal
- The series $\sum_{i=1}^{\infty} x_i \phi_i(\lambda)$ converges in mean-square to $x(\lambda)$.

The covariance function $K(\lambda, \xi)$ satisfies the necessary conditions on the kernel. Since the covariance kernel is Hilbert-Schmidt,

$$\iint_{\Lambda} |K(\lambda, \xi)|^2 d\sigma(\lambda) d\sigma(\xi) < \int_{\Lambda} E[x(\lambda)]^2 d\sigma(\lambda) < \infty \quad (2.19)$$

it can be shown that the operator K is compact (Weston, 1977). The covariance function is real and symmetric; hence, it is self-adjoint. If the covariance kernel is non-negative definite, the inequality

$$\iint_{\Lambda} K(\lambda, \xi) x(\lambda) x(\xi) d\sigma(\lambda) d\sigma(\xi) \geq 0 \quad (2.20)$$

is satisfied and the eigenvalues are non-negative. If the kernel is positive definite then the inequality is strict and the eigenvalues are non-zero and positive (Van Trees, 1968).

The proof of the theorem draws upon results that have been well-established in the literature of Hilbert Space theory (Ash, 1967).

The random variables x_i generated by the linear functionals are uncorrelated

$$\begin{aligned}
 E\{x_i x_j\} &= E\left[\int x(\lambda) \phi_i(\lambda) d\lambda \int x(\xi) \phi_j(\xi) d\xi \right] \quad (2.21) \\
 &= \int \phi_i(\lambda) \int K(\lambda, \xi) \phi_j(\xi) d\xi d\lambda \\
 &= \gamma_j \int \phi_i(\lambda) \phi_j(\lambda) d\lambda = \begin{cases} \gamma_i & i=j \\ 0 & i \neq j \end{cases}
 \end{aligned}$$

If $x(\lambda)$ is a Gaussian process, then the coefficients x_i are independent Gaussian random variables (Ash, 1967).

It is possible to order the set of eigenfunctions $\{\phi_i(\lambda)\}$ such that the sequence $\alpha_n(\lambda)$ for $n=N$, fixed, minimizes the expected mean-square representation error. To accomplish this ordering, the corresponding eigenvalues are ranked such that

$$\gamma_1 \geq \gamma_2 \geq \gamma_3 \geq \dots$$

The expected mean square error for N terms is (Brown, 1960)

$$\begin{aligned}
 E\{\epsilon_r\}_{n=N} &= E\left[\int_{\Lambda} \left[\sum_{i=N+1}^{\infty} x_i \phi_i(\lambda) \right]^2 d\sigma(\lambda) \right] \quad (2.22) \\
 &= \sum_{i=1}^{\infty} \{ |x_i|^2 \} \int_{\Lambda} \phi_i^2(\lambda) d\sigma(\lambda) \\
 &= \sum_{i=N+1}^{\infty} \gamma_i
 \end{aligned}$$

since from equation 2.21 $E\{|x_i|^2\} = \gamma_i$. The graph of the expected mean-square error as a function of the number of terms N in the expansion will show a sharp decrease in the error as the first terms are added. As an example consider the second-order stochastic process on the interval $[-1,1]$ with mean zero and covariance $K(\lambda, \xi) = \exp(-|\lambda - \xi|)$. The eigenvalues are given by Van Trees (1968, p. 188)

$$\gamma_i = \frac{2}{1 + b_i^2} \quad (2.23)$$

where the b_i are solutions to the equation

$$(\tan b_i + b_i) \left(\tan b_i - \frac{1}{b_i} \right) = 0 \quad (2.24)$$

A graph of the expected mean-square error for this process as a function of the number of terms (Figure 2.5) illustrates the desired rapid convergence property. It is important to note that for a fixed N the best set of N basis function from the set of all possible basis functions is the ordered set $\{\phi_i(\lambda)\}$, $i = 1, 2, \dots, N$.

In an effort to present an intuitive interpretation of the eigenvalues and eigenfunctions consider the first eigenvalue and its corresponding eigenfunction for a particular stochastic process. The first eigenvalue is found by choosing a function $\phi_1(\lambda)$ which maximizes the variance of the coefficient of that function. That is, the coefficient x_1 is given by the linear functional

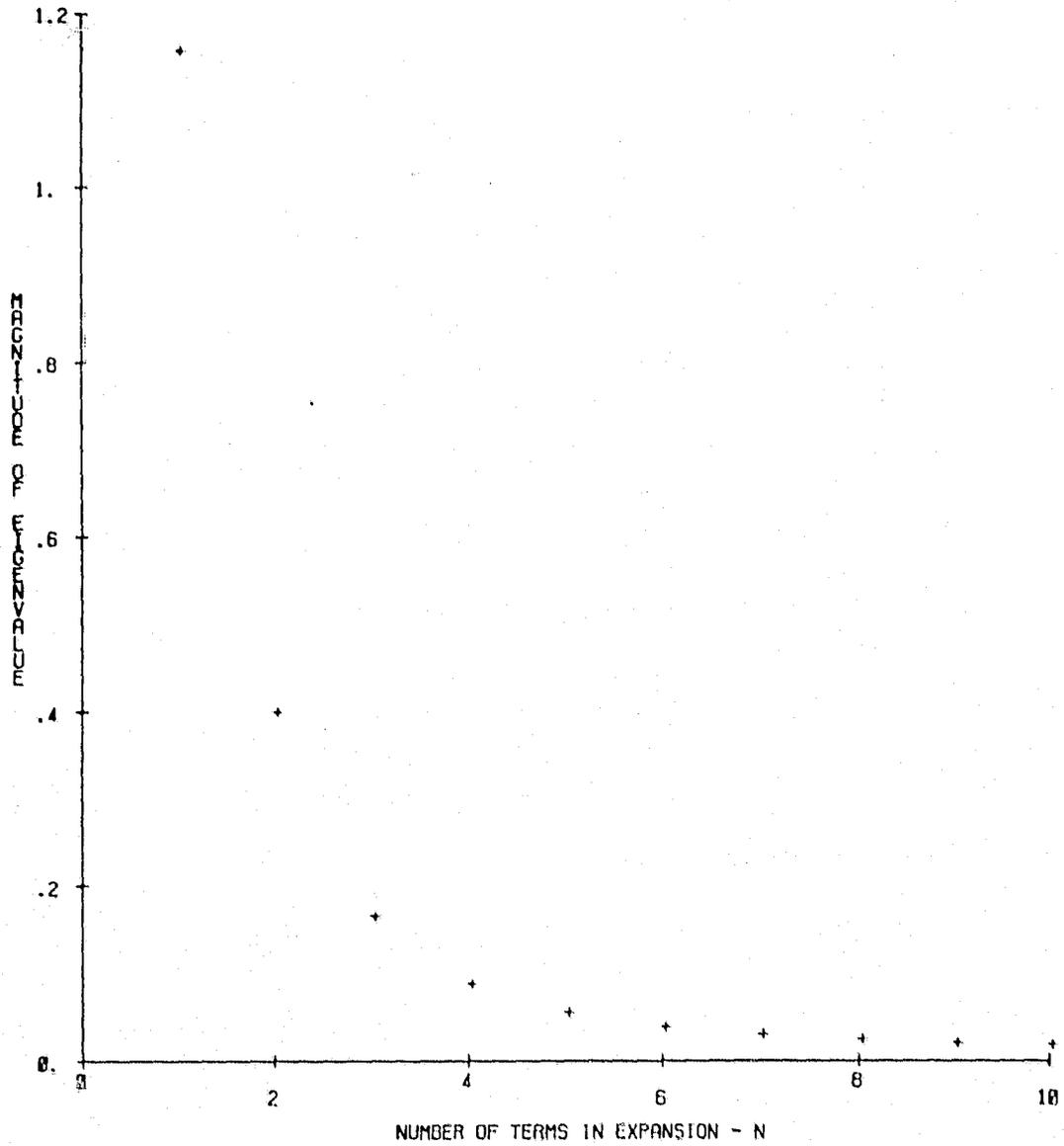


Figure 2.5 Eigenvalues for the stochastic process example.

$$x_1 = \int_{\Lambda} x(\lambda) \phi_1(\lambda) d\sigma(\lambda) \quad (2.25)$$

The variance of x_1 is equal to the first eigenvalue (2.21) which was chosen to be the largest of the set of eigenvalues. Since the variance is the largest for the coefficient x_1 , the uncertainty about the original function $x(\lambda)$ is reduced the most by using the first term. From a Shannon information theory point of view (Shannon, 1948) knowing the value of the coefficient x_1 provides the most information concerning the input signal that a single measurement can give. From the argument of being able to reconstruct the waveform, the coefficient x_1 gives the single most valuable measurement from which the input signal could be reconstructed.

The first eigenfunction can be used to identify portions of the spectral interval which may be more useful than others. If at a point λ on Λ the value of $\phi_1(\lambda)$ is close to zero, then the contribution to x_1 is not significant. On the other hand, if at a point λ , $\phi_1(\lambda)$, is significantly different from zero, then, the spectral response at that point may be of importance.

The second eigenvalue and eigenfunction attempt to find the second most useful portions of the spectral interval. The variance of the second coefficient x_2 is the second largest since the eigenvalue is the second

largest. The third eigenvalue and eigenfunction correspond to the third most useful and so forth. Therefore, there is an ordered sequence of eigenfunctions $\phi_1(\lambda)$, $\phi_2(\lambda)$, ... whose corresponding coefficients x_1, x_2, \dots provide a decreasing amount of information and a decreasing contribution to the reconstruction of the original function $x(\lambda)$. Based on the ranking, the eigenfunctions provide some intuitive indications concerning the importance of the points in the spectral interval.

A useful concept when discussing a signal set is the dimensionality of the signal space. The dimension of a signal space can be defined as the minimum number of basis functions required to completely reconstruct any function from the ensemble (Bennett, 1969). The orthogonal expansion which we have just derived provides an approximate method of determining the dimensionality of the observation space. If T is the value of expected mean-square error such that the approximate representation using only enough eigenfunctions to reduce $E\{\epsilon\}$ to a level below T , then the number of eigenfunctions is a reasonable approximation to the dimensionality of the signal space.

A general method has been developed for obtaining an optimal set of orthonormal basis functions such that if we choose an acceptable value of expected mean-square representation error $E\{\epsilon\}$, the series expansion can be truncated at some finite number N which will represent

any function in L_S with $E\{c_r\}$ less than the required value. Built into this derivation is the capability of adding a priori information which may be available concerning the spectrum. If we let $w(\lambda) = 1.0$ for all $\lambda \in \Lambda$, the series expansion is identical to the Karhunen-Loeve expansion derived in many texts (Davenport and Root, 1958; Van Trees, 1968; Middleton, 1960). When the weighting function is unity for all λ , the expansion will be referred to as the unweighted Karhunen-Loeve expansion.

Due to measurement difficulties in water absorption bands and differences in detector characteristics it has become apparent that the use of a weighting function different from the uniform one used above may be advantageous. The use of the weighted Karhunen-Loeve expansion has appeared only briefly in the literature (Kailath, 1971; Kailath, 1974). It is thought that the lack of wider use for the weighted Karhunen-Loeve expansion is due primarily to a lack of need for it until this time. The weighted Karhunen-Loeve expansion will be used extensively in the results to be presented later.

2.4 Discrete Approximation

It is proposed to solve the optimal sensor problem described in the previous section on a digital computer

using real data taken in the field. However, the solution must be approximated in order to take into consideration some practical constraints. First, the spectral response functions are not available as square-integrable functions on the dense set $\{\lambda \in \Lambda\}$. The functions are obtained in the field by sampling the spectral response with an instrument that uses very fine spectral windows. Secondly, the parameters of the process are not known a priori; hence, it is necessary to estimate the mean and covariance functions using a representative sample from the ensemble. Finally, because the data will be stored and processed digitally it is necessary to quantize the amplitude of the response at each of the spectral sample points. Each of these constraints can potentially contribute to the representation error for the process. In this section we want to consider the significance of the error due to spectral sampling, ensemble sampling and quantization.

2.4.1 Spectral Sampling

Up to this point the spectral response functions have been treated as functions of the continuous parameter λ . Suppose that the function $x(\lambda)$ is sampled at L intervals. Each spectral response function then becomes a vector $\underline{u} = [u_1, u_2, u_3, \dots, u_L]^T$. It would be desirable to use a large enough number of sample points such that the error introduced by sampling the spectral interval is not significant.

Note that once the sampling has been done the discrete equivalents to the solutions of the eigenvalue problem are used. The linear integral equation becomes

$$\Gamma \Phi = KW\Phi \quad (2.26)$$

where Φ is the $L \times L$ matrix of eigenvectors, K is the $L \times L$ covariance matrix given by

$$E\{[u_i - \bar{u}_i][u_j - \bar{u}_j]\} \quad (2.27)$$

$$\bar{u}_i = E\{u_i\}$$

Γ is the diagonal matrix of L eigenvalues, and W is the diagonal matrix of L weighting coefficients.

Because the actual covariance function is unknown, the loss of information or representation error from sampling cannot be evaluated. However, we can derive an expression that gives some insight into the effect of the error due to sampling. To evaluate the error due to sampling, the interval over the random process with mean $m(\lambda)$ and covariance $k(\lambda, \xi)$, the interval Λ is partitioned (Figure 2.6) into L equal intervals with $L+1$ end points λ_i .

Define a set of sampling functions by

$$\theta_i(\lambda) = \begin{cases} \frac{1}{\sqrt{\Delta\lambda}w(\lambda)}, & \lambda_{i-1} < \lambda < \lambda_i \\ 0, & \text{elsewhere} \end{cases} \quad \lambda \in \Lambda \quad (2.28)$$

where $\Delta\lambda = \lambda_i - \lambda_{i-1}$. The waveform $x(\lambda)$ can be approximated by

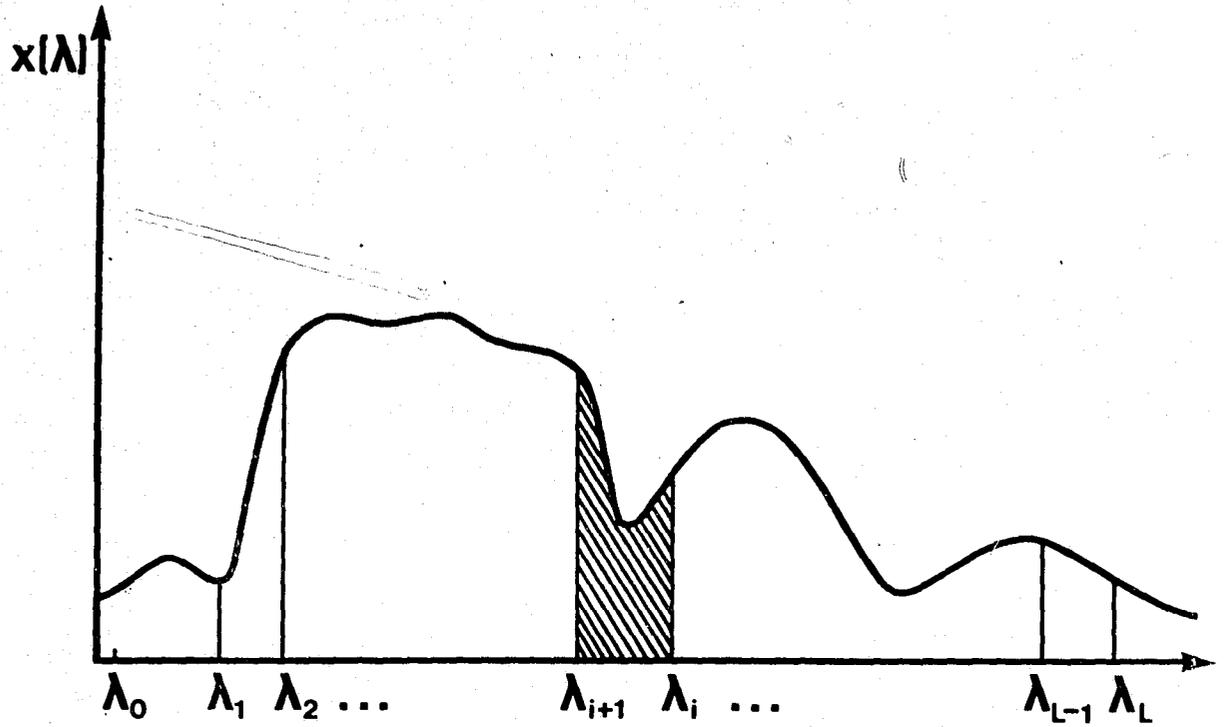


Figure 2.6 Partitioning of the spectral interval into L subintervals.

$$x_L(\lambda) = \sum_{i=1}^L x_i \theta_i(\lambda) \quad (2.29)$$

Using the expression for mean-square representation error,

$$\epsilon_S = \int_{\Lambda} [x(\lambda) - x_L(\lambda)]^2 d\sigma(\lambda) \quad (2.30)$$

the form for the expected error due to spectral sampling can be derived. The coefficient for the i th sampling function is given by

$$x_i = \int_{\Lambda} x(\lambda) \theta_i(\lambda) d\sigma(\lambda) \quad (2.31)$$

The expected mean square error due to sampling is

$$\begin{aligned} E\{\epsilon_S\} &= E \left[\int_{\Lambda} [x(\lambda) - x_L(\lambda)]^2 d\sigma(\lambda) \right] \quad (2.32) \\ &= \int_{\Lambda} K(\lambda, \lambda) d\sigma(\lambda) - E \left[\sum_{i=1}^L [x_i - m_i]^2 \right] + \int_{\Lambda} m^2(\lambda) d\sigma(\lambda) - \sum_{i=1}^L m_i^2 \end{aligned}$$

where $m(\lambda) = E\{x(\lambda)\}$ and $m_i = E\{x_i\}$. If the number of intervals L approaches infinity,

$$\lim_{L \rightarrow \infty} E \left[\sum_{i=1}^L (x_i - m_i)^2 \right] = \int_{\Lambda} K(\lambda, \lambda) d\sigma(\lambda) \quad (2.33)$$

and

$$\lim_{L \rightarrow \infty} \sum_{i=1}^L m_i^2 = \int_{\Lambda} m^2(\lambda) d\sigma(\lambda) \quad (2.34)$$

and the limit for the expected error is zero.

The discrete version of the Karhunen-Loeve expansion can be evaluated using the sampled spectral response functions. Let $y_i = x_i - m_i$, then

$$E\{y_i\} = 0 \quad (2.35)$$

$$E\{y_i y_j\} = k_{ij}$$

The discrete form of the integral equation is

$$\phi_L \Gamma_L = KW\phi_L \quad (2.36)$$

where Γ_L is the diagonal matrix of eigenvalues γ_{L_i} and ϕ_L is the matrix of eigenvectors ϕ_{L_i} for the L sampling intervals. Therefore,

$$E \sum_{i=1}^L (x_i - m_i)^2 = \sum_{i=1}^L \gamma_{L_i} \quad (2.37)$$

Hence, the expected error is the difference between the sum of the eigenvalues for the unsampled covariance function and the sum of the eigenvalues for the covariance matrix plus the difference between the integral of the mean function squared and the sum of the squares of the elements of the mean vector.

$$E\{\epsilon_S\} = \sum_{i=1}^{\infty} \gamma_i - \sum_{i=1}^L \gamma_{L_i} + \int_{\Lambda} m^2(\lambda) d\sigma(\lambda) - \sum_{i=1}^L m_i^2 \quad (2.38)$$

As an example consider the second-order zero-mean process described earlier with covariance $K(\lambda, \xi) = \exp(-|\lambda - \xi|)$.

Suppose the interval $[-1,1]$ is partitioned into 20 sub-intervals and the eigenvalues for the 20 dimensional system is computed. The expected mean square representation error due to sampling is:

$$E\{\epsilon_S\} |_{L=20} = \int_{-1}^1 e^{-|\lambda-\lambda|} d\lambda - \sum_{i=1}^L \gamma_{Li} \quad (2.39)$$

$$= 2.0 - \sum_{i=1}^L \gamma_{Li}$$

The first ten eigenvalues for the continuous covariance and the sampled covariance are listed in Table 2.1.

Table 2.1 Eigenvalues for continuous and sampled covariance.

EIGENVALUES		
	CONTINUOUS	SAMPLED
1	1.149	1.149
2	.391	.390
3	.157	.156
4	.080	.078
5	.047	.046
6	.031	.029
7	.022	.020
8	.016	.015
9	.012	.011
10	.010	.008

The expected mean-square error due to spectral sampling for 20 terms is 0.065. Depending on the form of the covariance function and the mean function, one can choose a

sufficient number of samples L to reduce the error to a negligible value.

2.4.2 Ensemble Sampling

Ideally, one would have available the complete ensemble from which the stochastic process could be accurately characterized. Unfortunately, a complete ensemble may require an infinite number of sample response functions since there are an infinite number of points in a stratum. A reasonable alternative is to select a representative sample from the ensemble from which the unknown parameters may be estimated. In sampling the ensemble one is concerned with the number of samples that are needed and how the sampling is done. By a 'representative' sample it is implied that the sample functions are taken from all typical observations in the stratum.

The number of samples required to adequately estimate the eigenvalues and eigenvectors can be evaluated in a straightforward manner. Using perturbation theory (Wilkinson, 1965), a first order approximation to the estimates of the eigenvalues and eigenvectors can be derived as functions of the covariance estimate

$$\hat{\gamma}_i \approx \phi_i^T \hat{K} \phi_i \quad (2.40)$$

$$\hat{\phi}_i \approx \phi_i + \sum_{\substack{j=1 \\ j \neq i}}^L \frac{\phi_i^T \hat{K} \phi_j}{(\gamma_i - \gamma_j)} \phi_j \quad (2.41)$$

where $\hat{\gamma}_i$ and $\hat{\phi}_i$ are estimates of the eigenvalues γ_i and eigenvectors ϕ_i respectively. These estimates are approximately unbiased (Fukunaga, 1972) since

$$E\{\hat{\gamma}_i\} \approx \{\phi_i^T \hat{K} \phi_i\} = \phi_i^T K \phi_i = \gamma_i \quad (2.42)$$

and

$$E\{\hat{\phi}_i\} \approx \phi_i + \sum_{\substack{j=1 \\ j \neq i}}^L \left[\frac{\phi_i^T E\{\hat{K}\} \phi_j}{(\gamma_i - \gamma_j)} \right] \phi_j = \phi_i \quad (2.43)$$

The variances of the estimates are expressed by

$$\text{Var}\{\hat{\gamma}_i\} = E\{(\hat{\gamma}_i - \gamma_i)^2\} \approx E\{(\phi_i^T \hat{K} \phi_i)^2 - \gamma_i^2\} \quad (2.44)$$

and

$$\text{Var}\{\hat{\phi}_i\} = E\{\|\hat{\phi}_i - \phi_i\|^2\} \approx \sum_{\substack{j=1 \\ j \neq i}}^L \frac{E(\phi_i^T \hat{K} \phi_j)^2}{(\gamma_i - \gamma_j)^2} \quad (2.45)$$

The term that must be evaluated is $E\{(\phi_i^T \hat{K} \phi_j)^2\}$. The derivation follows that of Fukunaga (1972) from which the result is shown to be

$$E\{(\phi_i^T \hat{K} \phi_j)^2\} = \frac{(N_S - 1) N_S}{(N_S - 1)^2} \gamma_i^2 \delta_{ij} + \frac{N_S}{(N_S - 1)^2} 2\gamma_i^2 \delta_{ij} \quad (2.46)$$

$$+ \frac{N_S}{(N_S - 1)^2} \gamma_i \gamma_j$$

Where N_S is the number of sample functions in the ensemble and δ_{ij} is the Kroncker delta. Now the variances can be written

$$\text{Var} [\gamma_i] = \frac{4 N_S - 1}{(N_S - 1)^2} \gamma_i^2 \quad (2.47)$$

and

$$\text{Var} [\hat{\phi}_i] = \sum_{\substack{j=1 \\ j \neq i}}^N \frac{N_S}{(N_S - 1)^2} \frac{\gamma_i \gamma_j}{(\gamma_i - \gamma_j)^2} \quad (2.48)$$

Note that the variance of the eigenvalue is proportional to the square of the eigenvalue. The variance will decrease as the number of samples is increased and asymptotically approaches zero as N_S approaches infinity. Since the expected mean-square error is a function of the eigenvalues the estimate of the error is also asymptotically unbiased.

The variance of the eigenvectors is very large when two eigenvalues are close together. If all of the eigenvalues are well separated the variance of the eigenvectors approaches zero as N_S approaches infinity.

2.4.3 Quantization

Quantization of the amplitude of each element in the output vector is necessary for subsequent data transmission, storage and digital processing. A Q-level quantizer divides the amplitude range into Q equally-spaced intervals. The

ith interval has end points x_i and x_{i+1} and output level y_i equal to $(x_{i+1} + x_i)/2.0$. The expected mean-square error due to quantization is given by Max (1960).

$$E\{\epsilon_q\} = \sum_{i=0}^{Q-1} \int_{x_i}^{x_{i+1}} (x - y_i)^2 p(x) dx \quad (2.49)$$

where $p(x)$ is the probability density function for the random amplitude x .

Suppose the interval corresponding to the amplitude range has length I and is divided into Q subintervals. Assume that the probability density function is very small outside I . The length of each subinterval L_I is the ratio I/Q . An upper bound to $E\{\epsilon_q\}$ can be found easily by noting that

$$(x - y_i)^2 \leq \left(\frac{1}{2} L_I\right)^2 = \frac{L_I^2}{4} \quad (2.50)$$

and

$$\sum_{i=1}^Q \int_{x_i}^{x_{i+1}} p(x) dx = 1 \quad (2.51)$$

Therefore,

$$E\{\epsilon_q\} < \frac{L_I^2}{4} \quad (2.52)$$

By keeping the length L_I reasonably small the quantization error will not be significant.

If the probability density function has a significant portion of the function outside the designated amplitude range the expected error may increase substantially. The quantizer will assign the value y_Q to all values of x greater than x_Q , and y_0 to all values of x less than x_1 . If x is outside the amplitude range saturation will occur. The mean-square error will increase significantly if this situation occurs.

The total expected mean-square representation error is a function of the errors due to truncation, spectral sampling, ensemble sampling, and quantization. It has been demonstrated that the error due to quantization is not significant. In fact the uncertainty in the measuring devices is considerably greater than the uncertainty due to quantization. Since the covariance and mean function are not known a priori it is not possible to evaluate the expected error due to spectral sampling. However, it was demonstrated that for a known case the number of samples required to reduce the error to a negligible value was not unreasonable. Therefore, in the experimental work it will be assumed that the expected error due to spectral sampling will be sufficiently smaller than the average error in the measuring system.

Since the estimates of the eigenvalues and eigenvectors are unbiased, it is expected that the corresponding error

due to the fact that only a finite set of sample functions was available would be small, especially if the number of sample functions was sufficiently large.

The principal source of error which will be considered will be the error due to truncation. Hence, provided that some care has been taken with regard to the number of spectral samples, number of sample functions, and the length of the quantization intervals, the approximation to the continuous case is not unreasonable.

CHAPTER 3. THE PARAMETERS AND OVERALL SYSTEM PERFORMANCE

The intent in choosing a particular sensor design is to optimize the expected performance of the pattern recognition system with respect to the global performance criterion ϵ_0 . The quantity ϵ_0 is a complicated function of a set by parameters α . By varying the parameters a search can be made to find the best combination to optimize ϵ_0 . The first step is to list the parameters. Five parameter categories were listed in Chapter 1:

- spectral representation
- spatial representation
- signal-to-noise ratio
- ancillary data
- information classes

The problem is to quantify these parameters categories such that an optimization procedure can be applied. As a preliminary step, it is proposed to consider each category individually and study the relationship between that category and the global criterion. The other parameters will be held constant while allowing the parameter under investigation to vary. It is also important to understand the inter-relationships between the parameters; a change in one

parameter may influence the performance criterion both directly and indirectly through another parameter.

The primary focus, here, is on the spectral representation parameter and its corresponding quantity, mean-square representation error, ϵ_r . In Chapter 2, an analysis procedure was developed in which an ordered sequence of basis functions allows the spectral response function to be represented with decreasing expected mean-square error. It remains to show the effect of the spectral parameter on the overall system performance. This chapter will first deal with the relationship between ϵ_o and ϵ_r , followed by a discussion of some research results relating other parameters to ϵ_o .

There are a variety of processors which can be used to evaluate a data set depending on the nature of the problem. Typical processors include separability computers, linear classifiers, quadratic classifiers, non-parametric classifiers and context classifiers. We will choose the maximum likelihood Gaussian classifier as an example of a quadratic classifier which will be used as representative processor for evaluation of the pattern recognition system. Let X be an observation from one of M classes C_i , $i=1,2,\dots,M$, with a priori probabilities P_i . The maximum likelihood decision rule can be stated as follows: Assign X to the class C_k if

$$P_k p(X|C_k) = \max_i \{P_i p(X|C_i)\} \quad (3.1)$$

where the $p(X|C_i)$ are the class conditional probability density functions. If Ω is the observation space, then this rule partitions Ω into the subspaces $\Omega_1, \Omega_2, \dots, \Omega_M$ corresponding to the classes C_1, C_2, \dots, C_M , respectively.

The probability of correct classification has found widespread use in the pattern recognition and remote sensing community, and will be used here as the system performance criterion. For a multivariate, multiclass pattern recognition problem the probability of correct classification is defined as

$$P_c = \int_{\Omega} \max_i \{P_i p(X|C_i)\} dx \quad (3.2)$$

where $p(X|C_i)$ is the conditional jointly Gaussian probability density function for class i .

3.1 Spectral Representation

The expected mean-square error, $E\{\epsilon_r\}$, has been used as a measure of the fidelity of the spectral representation. The Karhunen-Loeve expansion has been developed as a means of representing the spectral response functions in the ensemble by a finite series expansion such that $E\{\epsilon_r\}$ is minimized. We wish to study the relationship between the spectral representation and the performance of the overall pattern recognition system.

If the stochastic process is completely known, and if all the terms in the Karhunen-Loeve expansion are used in the representation, then a decision scheme exists which is optimal in the sense of maximizing the probability of correct classification. For the M-class pattern recognition problem an experiment is defined whose outcome is the vector X belonging to the set of all possible outcomes. A decision scheme is realized by partitioning the observation space into M regions such that if X belongs to Ω_i , then the decision, 'X belongs to class C_i ,' is made. Such a decision scheme can be arrived at by evaluating the a posteriori probabilities for each class. The a posteriori probability $(P(C_i|X))$ is the conditional probability that class C_i occurs given that the measurement value is equal to X . If the vector X is finite dimensional, then it is straightforward to evaluate the a posteriori probabilities and, using equation 3.1, to design a classifier which is optimal in the sense of maximizing the probability of correct classification (Anderson, 1958).

This approach has often been generalized to the case where the vectors are infinite dimensional and the outcomes are real functions $x(\lambda)$ on an interval (Grenander, 1950; Kadota, 1964, 1965; Van Trees, 1971). The procedure begins by representing observed sample function in the

ensemble by a finite vector $[x_1, x_2, \dots, x_N]^T$ of coefficients which are the coefficients in the Karhunen-Loeve expansion. The a posteriori probability for each class is constructed and the limit as N approaches infinity of the conditional probability $P(C_i | x_1, x_2, \dots, x_N)$ is taken. Bharucha (1969) has shown that this limit exists, and furthermore, that the resulting decision scheme optimally partitions the observation space such that the probability of correct classification is maximized.

We now consider the implications of the constraint that the number of terms in the expansion be finite has on the classification performance. If N features are used, it cannot be guaranteed that the first N features are the best for discriminating between M classes in a particular remote sensing problem (Foley and Sammon, 1975). A simple example has been used to demonstrate this fact. Suppose there are two features and it is desired to use only one feature to discriminate between the two classes. Let the classes be distributed as shown in Figure 3.1. Based on the criterion of minimum mean-square representation error the basis function ϕ_1 should be chosen. However, it is obvious that ϕ_2 is the better choice for discriminating between the classes. Hence, if say ten terms are used in the representation, it may be true that the 34th term, for example, is superior for discriminating between classes than some of the first 10 features.

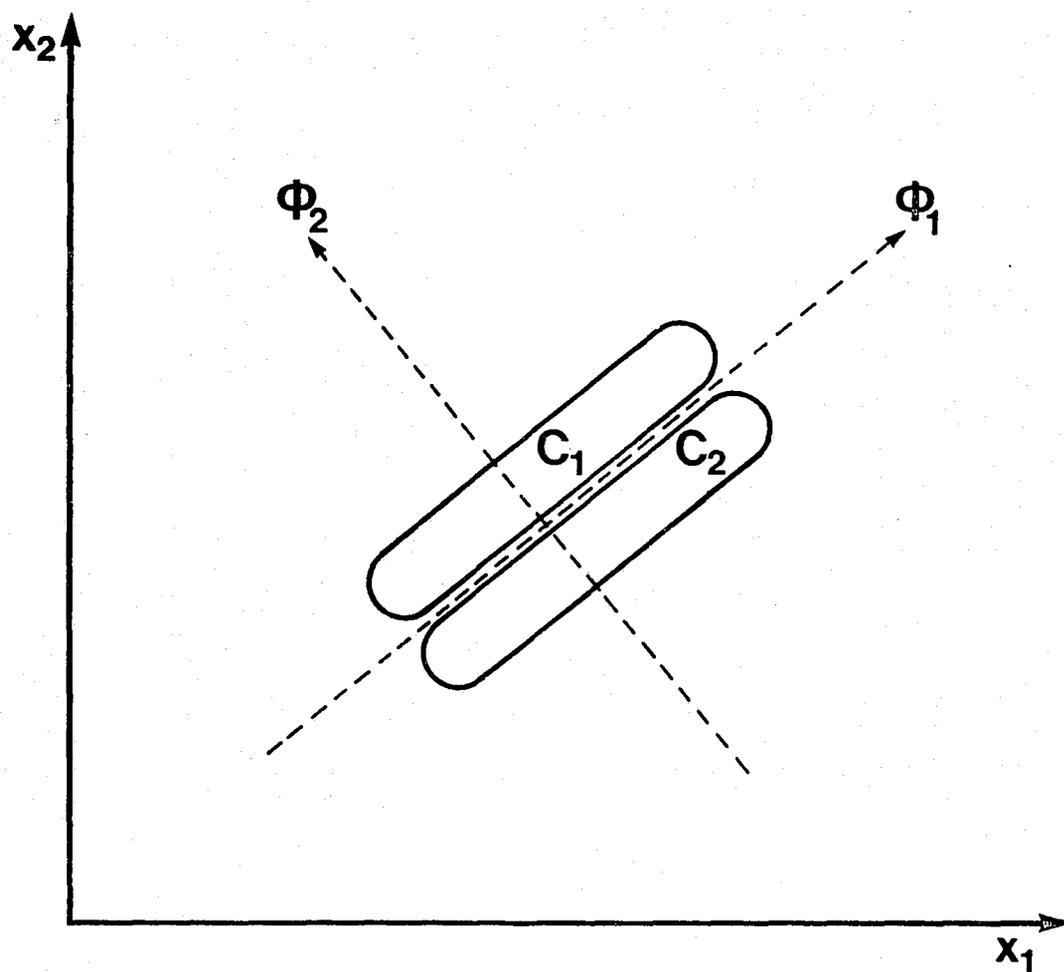


Figure 3.1 Two distributions which demonstrate a potential difficulty in using the best feature for representation to perform classification.

We wish to develop, here, the relationship between the expected mean square error for an optimal set of basis functions and the probability of correct classification. If the stochastic process is completely known, a decrease in the mean-square representation error does not result in a decrease in the probability of correct classification. The addition of a measurement or feature does not decrease the separability. If the added measurement contributes only noise, then the separability of the distributions is the same as without the added measurement. This monotonicity is implied in the convergence of the a posteriori probabilities as N approaches infinity.

The intent here is to have as small a value of $E\{\epsilon_r\}$ as possible or at least drive it well below the average measurement noise. Every decrease in $E\{\epsilon_r\}$ is known to not decrease P_c . Returning to the example, we would not choose only one feature if we could help it, but rather choose to keep both features since this would reduce $E\{\epsilon_r\}$ to zero for the two-dimensional case.

If the probability of correct classification is plotted as a function of the expected mean-square representation error as sketched in Figure 3.2, some important insights into the nature of the data can be gained. We know that as the expected error decreases the classification accuracy does not decrease. The monotonicity is indicated by the solid line in the figure. We wish to observe the behavior of the

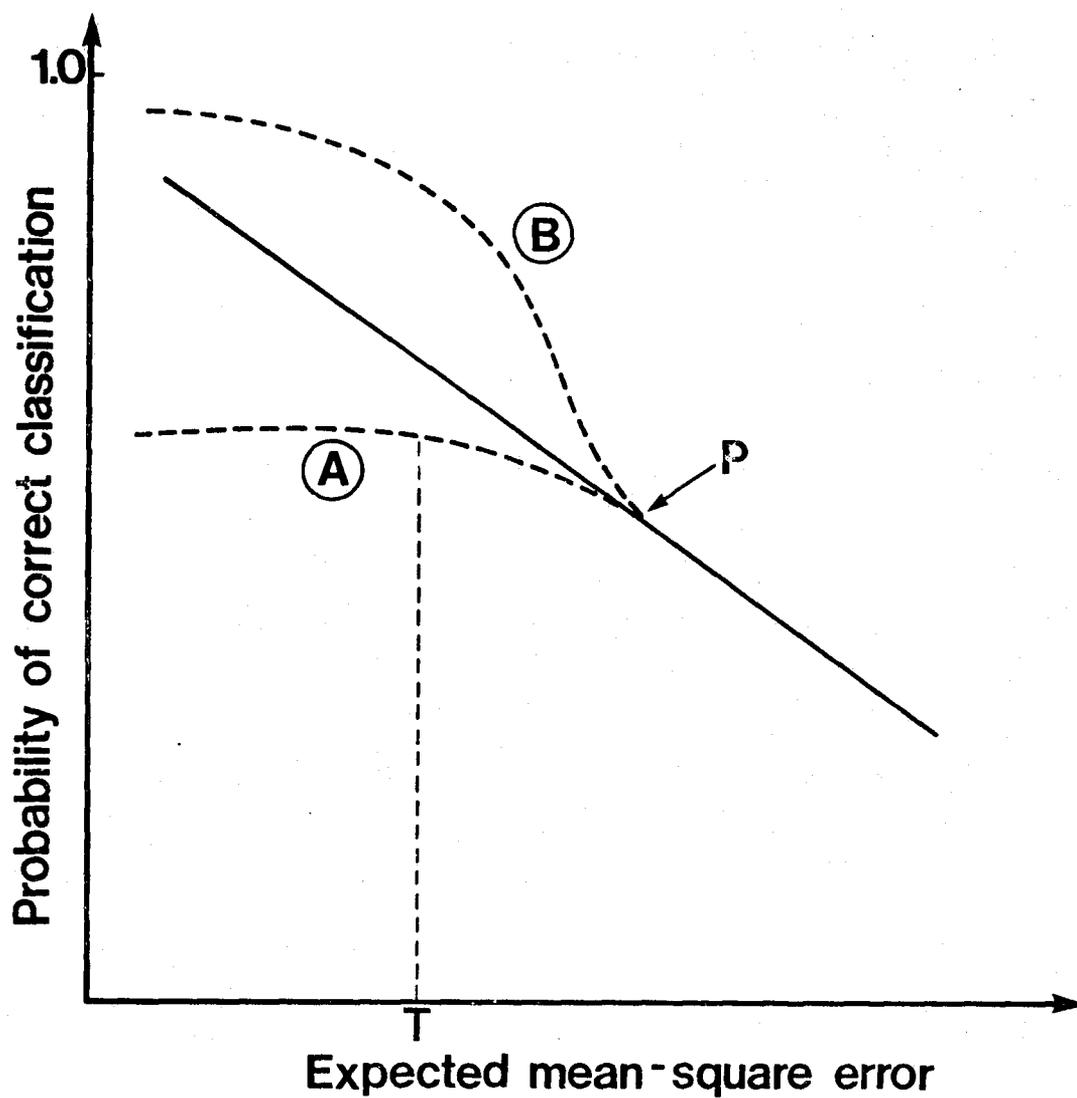


Figure 3.2 Probability of correct classification as a function of expected mean-square representation error.

relationship between the expected error and the classification performance as $E\{\epsilon_r\}$ becomes small. It may occur that at some point P, a large decrease in the expected error results in little or no change in the classification performance as indicated by the dashed line (A). In this case the number of terms required to represent the process with an error of T is sufficient for the information classes chosen. One may be able to evaluate the portions of the spectrum which are of most value based on the first few eigenvectors. Also, in this case one can estimate the maximum classification performance that can be achieved by noting the value of P_c that the graph is approaching as the expected error becomes small.

Suppose, however, that at point P a small decrease in the expected error results in a significant improvement in classification performance as indicated by the dashed line (B). In this case more terms are required to attain the maximum discrimination capability. Also the eigenvectors which correspond to the largest improvements in performance can be analyzed to determine which spectral regions are contributing the most.

Several times in this discussion the condition that the process be completely known was stated. If the process is not completely known, but must be estimated from a finite data set then the situation becomes different. The effect of a finite data set size is now discussed.

3.2 Ancillary Data

Ancillary data is information other than the spectral response functions themselves associated with a stratum which has bearing on the performance of the system over that stratum. An example of ancillary data which is important for this research is the design set. Sample response functions drawn from the ensemble are used to design the classifier. For a maximum likelihood Gaussian classifier the design procedure is to estimate the mean vectors and covariance matrices for each class from the design set.

For a fixed number of features or dimensions, it is well known that if the design set is used to test the classifier performance, the estimate of probability of correct classification \hat{P}_c will be optimistically biased (Fukunaga, 1972; Toussaint, 1974). That is, the estimate is better than the true performance. If a test set, consisting of sample functions from the ensemble different from those in the design set, is used, the performance estimate is inferior to the true performance. If the number of sample functions N_s is increased the estimates of classification performance both approach the true performance. If the number of sample functions N_s approaches infinity, the probability structure will become completely known and the true performance can be evaluated.

Now, consider the case where the number of sample functions is fixed while letting the number of features be a variable. If N_s is infinite, increasing the number of features from N to $N+1$ will either improve the performance or the performance will be the same for N and $N+1$ features. However, if N_s is finite, increasing the number of features may have an adverse affect on the performance estimate.

Three research results have been published which attempt to determine the relationship between the design set size and the number of features. One of the first attempts to quantify and explain this relationship was done by Allais (1964). The study involved the linear prediction problem which is closely associated with the linear two-class pattern recognition problem. Allais showed both analytically and experimentally that for a fixed N_s , increasing the number of measurements improved the performance for a while until a certain peak was reached, after which the performance deteriorated drastically.

A second research result reported by Hughes (1968) showed the same peaking for mean recognition accuracy as measurement complexity is increased. The mean recognition accuracy is the average over all discrete non-parametric probability structures of the correct recognition probability using the Bayes recognition rule. The measurement complexity is the total number of discrete values and

is equal to the product of the number of features and the number of quantization levels. Hughes argues that increasing the measurement complexity necessarily means that there are fewer samples, N_s , per measurement cell available to estimate the probabilities associated with each cell. Hence, when the classification accuracy is computed using these cell probabilities, the average classification accuracy will decrease as the number of features increases if the design set size is too small.

A third research result is due to Foley (1975) who studied two-class multivariate Gaussian pattern recognition problems with different means but identical covariances. An analytical expression was developed to determine what the ratio of design set size to feature size should be to obtain a good estimate of the performance of the classifier. A ratio of 3-to-1 was considered to be a good engineering rule-of-thumb for choosing the number of features for a given sample size.

These results have been somewhat controversial and often misinterpreted, especially the work by Hughes, and have frequently been discussed in the literature (Kanal and Chandrasekaran, 1971; Abend et al, 1969; Chandrasekaran, 1971; and Chandrasekaran and Jain, 1974, 1975).

The underlying cause of the influence of sample size is due to the statistical uncertainty that occurs in estimating the statistics for the classes. As sketched

in Figure 3.3 the positive bias in the performance estimate when testing on the design set increases with the number of features used. This bias is due to the cumulative effects of the uncertainty in estimating the statistical parameters (Chen, 1978). When the test set is used to evaluate the performance the bias decreases as more features are added. The end result is that a positive bias becomes significant at some point determined by the sample size for the estimate on the training set. A degradation in the performance occurs at the same point for the estimate based on the test set.

A concept which is brought out in much of the literature dealing with the relationship between feature size and design set size is that the more a priori knowledge about the underlying probability structure that is available the more features that can be used with a given data set size (Foley, 1972). Conversely, for a fixed number of features, added knowledge of the probability structure allows one to reduce the number of design set samples collected (Mogera and Cooper, 1977). As an example, the fact that the probability densities are assumed to be Gaussian implies that fewer sample functions are required to get good estimates of performance than if no parametric assumption was made.

Figure 3.3 The effects of sample size on classification performance as a function of the number of features, a) true performance, b) Positive bias in \hat{P}_C due to testing on the design set, c) Negative bias in \hat{P}_C due to testing on the test set, d) estimate of \hat{P}_C when testing on the design set, e) \hat{P}_C for testing on the test set.

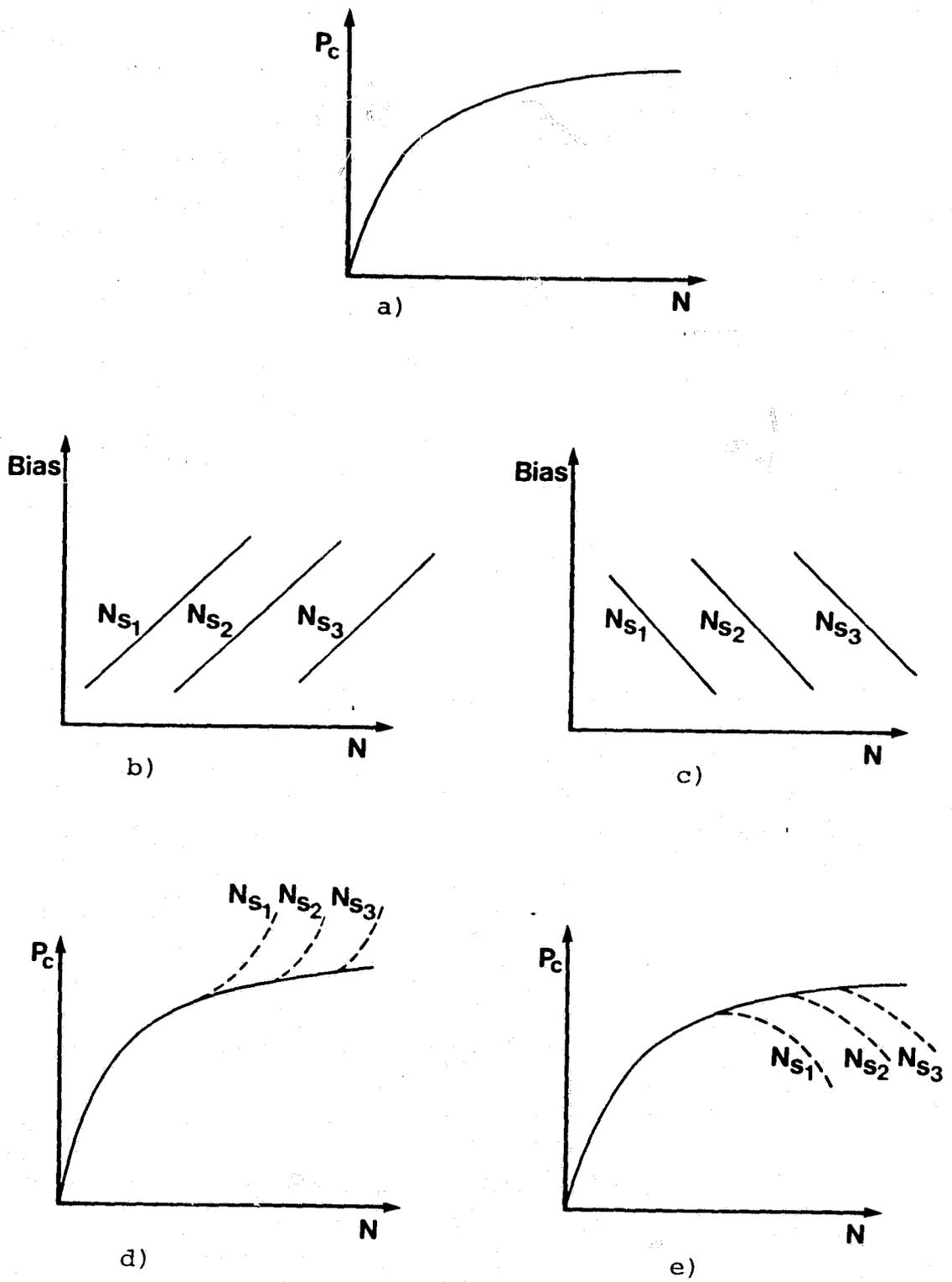


Figure 3.3

3.3 Information Classes

An information class refers to the label assigned to the sample points in the stratum. The labels are chosen to be meaningful in the context of the pattern recognition problem under consideration. Since we are looking for a sensor which will work well for a variety of pattern recognition problems, we consider the influence of the choice of information classes on the overall performance criterion for the pattern recognition system.

We first note that there is an intrinsic set of classes which is associated with each stratum. For example, in some strata the class list may consist of primarily vegetation classes; whereas, in other strata urban classes may be predominant. For each stratum a non-unique hierarchial tree structure may be constructed (Figure 3.4) (Landgrebe, 1978). To construct the information tree it is important to remember that the class list must be exhaustive; that is, every point in the stratum must be assigned to one of the classes. The choice of the class labels depends on the informational value that they have to the user. At the top of the tree the classes are easily separable using few features. As one selects class sets which are deeper in the tree structure, it becomes increasingly more difficult to discriminate between the classes.

An example using artificial data can be generated which demonstrates the effect of the choice of information

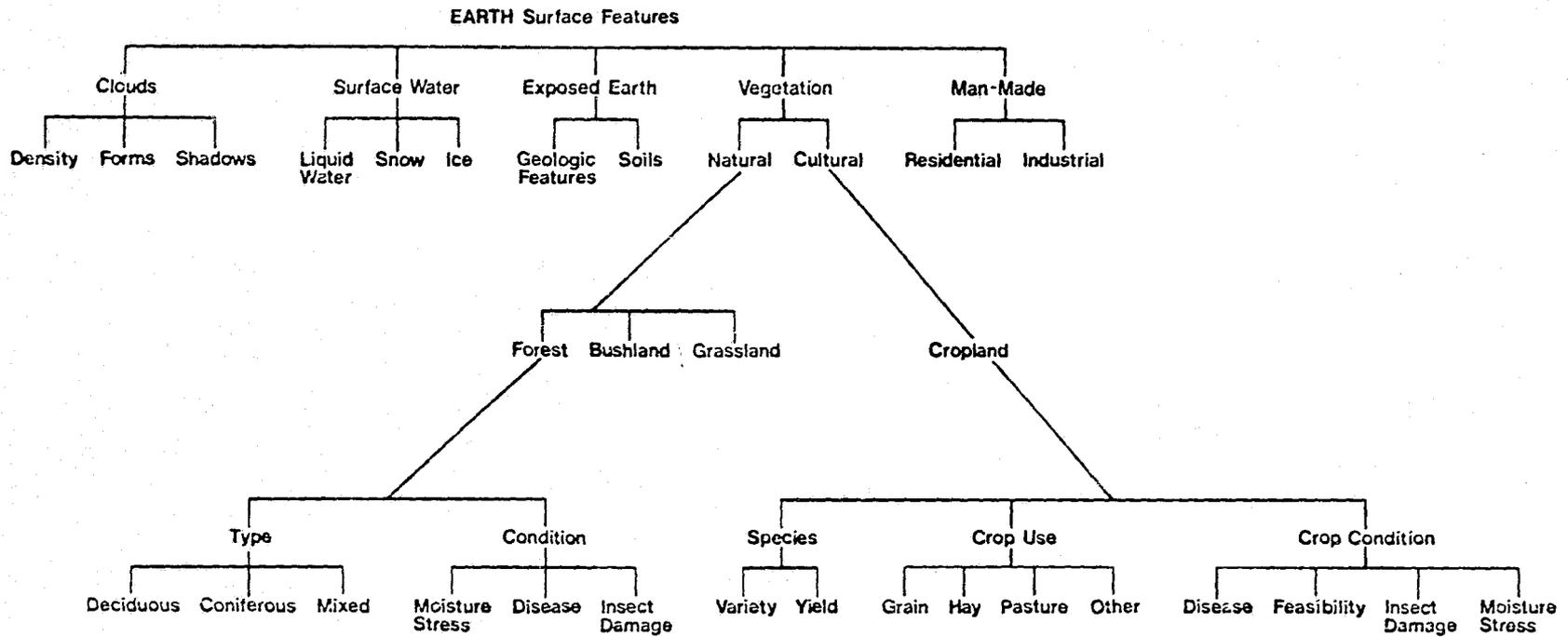
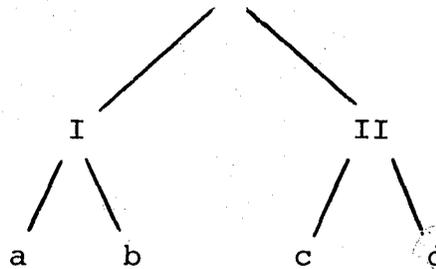


Figure 3.4 An information tree for a typical stratum (Landgrebe, 1978).

classes on the probability of correct classification. Assume that the data is two-dimensional and that a tree structure can be drawn as follows:



where I and II denote the first level classes and a, b, c, and d denote second level classes. Let the mean vectors and the covariance matrices for the four classes be

$$\text{class a} \quad \underline{M}_a = \begin{bmatrix} 10.0 \\ 11.0 \end{bmatrix} \quad K_a = \begin{bmatrix} 3 & 3/2 \\ 3/2 & 5 \end{bmatrix}$$

$$\text{class b} \quad \underline{M}_b = \begin{bmatrix} 10.5 \\ 11.0 \end{bmatrix} \quad K_b = \begin{bmatrix} 2 & 1/2 \\ 1/2 & 3/2 \end{bmatrix}$$

$$\text{class c} \quad \underline{M}_c = \begin{bmatrix} 15.0 \\ 9.0 \end{bmatrix} \quad K_c = \begin{bmatrix} 8 & 5/4 \\ 5/4 & 10 \end{bmatrix}$$

$$\text{class d} \quad \underline{M}_d = \begin{bmatrix} 14.5 \\ 9.5 \end{bmatrix} \quad K_d = \begin{bmatrix} 6 & 1 \\ 1 & 4 \end{bmatrix}$$

Plotting 20 random points from each of these distributions in two-dimensions gives some idea of the four distributions (Figure 3.5).

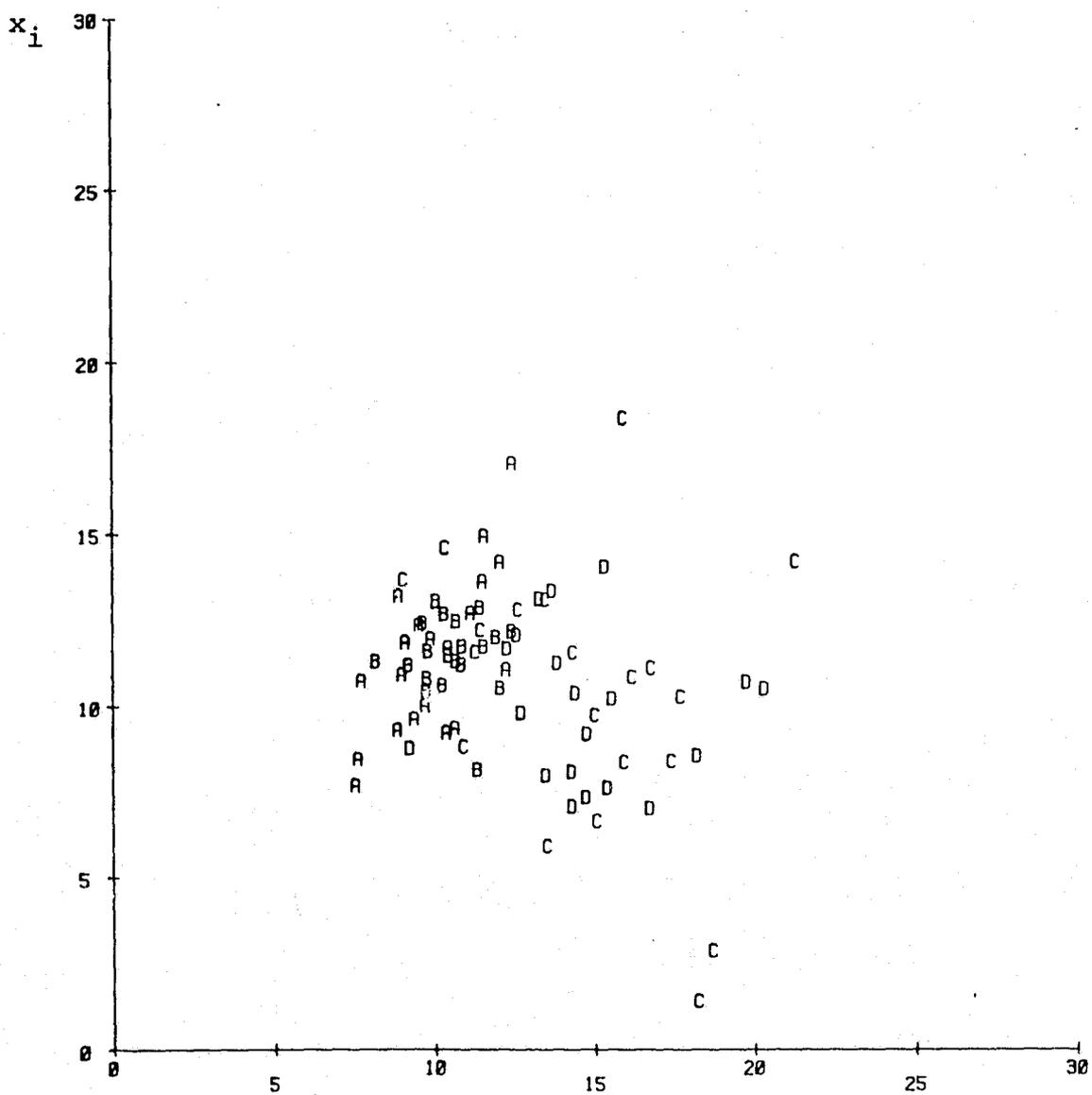


Figure 3.5 Distributions of four information classes.

The performance can be evaluated for the first level by combining the statistics assuming that the four classes are equally probable. A performance estimator was used to evaluate the probability of correct classification from the known statistics. The overall probability of correct classification at level one is 0.91 whereas the overall probability when attempting to discriminate between the four classes is 0.59. One can readily see from this example that the choice of classes will effect the overall performance criterion.

Recalling the graph of the classification performance as a function of expected mean-square error, a different set of information classes may alter the graph significantly. In general, information classes that are deeper in the information tree will require smaller representation error to achieve a specified classification performance. Kulkarni (1978) provides further discussion of the performance of a classifier as a function of the design set size, the measurement complexity, and the depth of the information tree.

One can also observe that the information classes present in a stratum influence the selection of the optimum set of basis functions $\{\phi_i(\lambda)\}$. Let each class have a Gaussian probability density with mean function $m_i(\lambda)$ and covariance function $K_i(\lambda, \epsilon)$, $i=1,2, \dots, M$. The covariance function for the stochastic

process can be written as a function of the class conditional mean and covariance functions.

$$K(\lambda, \xi) = E\{(x(\lambda) - m(\lambda))(x(\xi) - m(\xi))\}$$

$$\text{where } m(\lambda) = \sum_{i=1}^M P_i m_i(\lambda)$$

$$K(\lambda, \xi) = \sum_{k=1}^M P_k K_k(\lambda, \xi) + \sum_{k=1}^M P_k \left[m_k(\lambda) - \sum_{i=1}^M P_i m_i(\lambda) \right] \left[m_k(\xi) - \sum_{i=1}^M P_i m_i(\xi) \right]$$

For the special case where $M=2$, this equation reduces to

$$K(\lambda, \xi) = P_1 K_1(\lambda, \xi) + P_2 K_2(\lambda, \xi) + P_1 P_2 [m_1(\lambda) - m_2(\lambda)] [m_1(\xi) - m_2(\xi)]$$

Recall that $K(\lambda, \xi)$ is the kernel of the integral equation which is solved to obtain the optimum set of basis functions $\{\phi_i(\lambda)\}$. Hence, the information classes determine the values of the mean and covariance functions and their relationships and, subsequently, influence the selection of the basis functions. The solutions $\{\phi_i(\lambda)\}$ to the integral equation are ordered by the eigenvalues such that regions of the interval Λ which have large variance are weighted more heavily. A change in the spectral

classes such that the means are further apart, will cause an increase in the variance along the coordinates in which there was an increase in the distance between probability distributions.

3.4 Spatial Representation

The spatial representation parameter reflects the ability of the sensor to represent the spatial characteristics of objects in the scene. Spatial characteristics may include the size, orientation, and texture of objects as well as the distance and direction from other objects in the scene. In image-oriented pattern recognition systems the spatial representation parameter is paramount since the spatial characteristics are information bearing features; whereas, in numerically oriented systems the spatial representation is less important but significant.

The fundamental quantity for spatial representation is the ground resolution element size. The ground resolution element is the area of the earth's surface which is being observed by the sensor at a given instant of time. A physically realizable sensor system is constrained to observe an area of finite size. The area of the ground resolution element is determined by the sensor's instantaneous field of view (IFOV), altitude, velocity, and scan rate.

The size of the ground resolution element determines what information classes can be observed. If the size of

an object is denoted by F and the size of the ground resolution element is denoted by Δ , three relationships between F and Δ can be expressed: $\Delta > F$, $\Delta \approx F$, and $\Delta < F$ as shown in Figure 3.6. If the size of objects or fields is smaller than the resolution element size it is very difficult to identify them. If the object size and resolution element size are about the same, the performance is marginal, principally because the center of the object differs from the center of a ground resolution element a significant percentage of the time. Quite often the object will occupy space in small portions of two or more resolution elements. The resulting mixed elements may have spectral response functions which are not characteristic of either the object or the surrounding area.

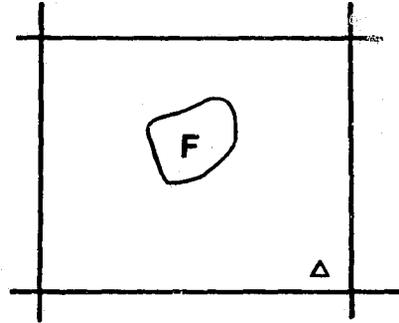
The best case is when the ground resolution element is much smaller than the field size. For crop inventory applications the field size determines the approximate resolution element size required to keep root mean square error of area estimates below a specified level (CITARS experiment; see Harnage and Landgrebe, 1975). Results of the CITARS experiment indicate that the number of resolution elements per field should be greater than forty to avoid the effects of boundary resolution elements.

Having a small ground resolution element also provides more sample functions per class. As discussed in a previous section more sample functions will provide

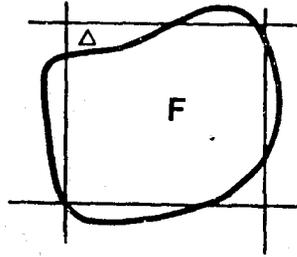
ORIGINAL PAGE IS
OF POOR QUALITY

Δ Ground resolution element
F Field or object in the scene

$\Delta \gg F$



$\Delta = F$



$\Delta \ll F$

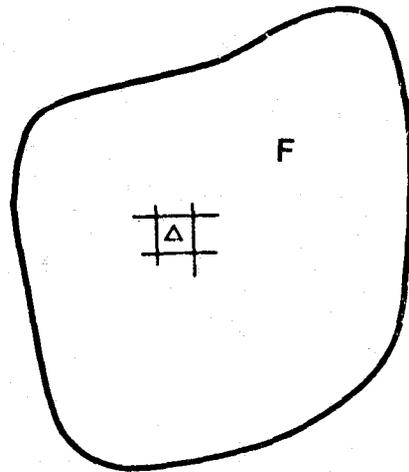


Figure 3.6 Relationships between object size F and ground resolution element size Δ .

better estimates of the statistics and allow more features to be used to represent the spectrum.

It might seem that the smaller the ground resolution element the better the performance; however, the signal-to-noise ratio deteriorates with decreasing resolution element size. The energy available to the sensor decreases as the area observed by the sensor at a given time decreases. The resulting decrease in signal-to-noise ratio tends to cause a degradation in the overall system performance. Mobasser (1978) has shown that an increase in the ground resolution element size corresponds to a significant improvement in the classification accuracy. It is assumed that the size of the fields or objects is sufficiently large as to not be a factor in these results. Also the spectral representation parameters, sample size, signal-to-noise ratio, and the set of information classes were held fixed.

In this discussion only per-point or per-element classifiers have been considered so far. Classifiers which incorporate spatial information to improve the performance have been developed. The ECHO classifier developed by Kettig and Landgrebe (1975) divides the scene into homogeneous objects. These objects are then classified on a per field basis. Since the decision rule decides to which class a field belongs on the basis of its mean vector and covariance matrix rather than the single

vector from a single point, a potentially faster and better classification can be made.

In the experiment of Landgrebe, Biehl and Simmons (1976), the ECHO and per element classifiers were compared for different ground resolution element sizes. The results are shown in Figure 3.7. Note that for the smaller resolution element sizes the spatial classifier is slightly better than the per-element. As the ground resolution element size increases the objects size become closer to the resolution element size and the ECHO classifier becomes essentially a per-element classifier. Also the per-element classifier improves as the resolution element size increases.

Another effort to utilize spatial information is to generate texture features (Haralick et al, 1973; and Wiersma and Landgrebe, 1976). The texture features are numerical quantities which loosely correspond to some intuitive properties of textures which humans can perceive. The spatial resolution in this case affects the textures which one can observe. A fine resolution has a more detailed texture as in the response variations due to the size and shapes of leaves. A coarse resolution is more sensitive to large scale textures such as the quilt-like patterns of agricultural fields.

The choice of the spatial representation parameter depends primarily on the choice of information classes.

C-2

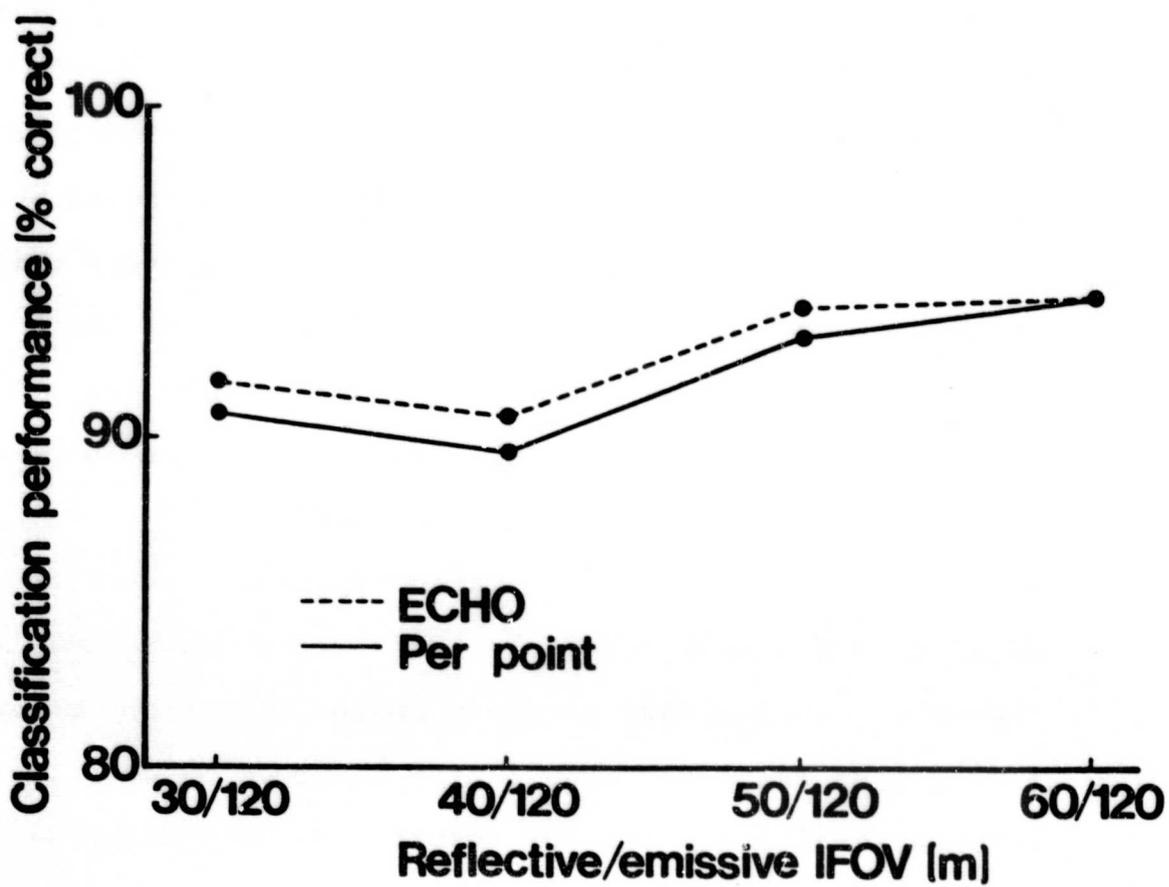


Figure 3.7 Classification performance vs. spatial resolution using ECHO and per-point classifiers (Landgrebe et al., 1977).

Tradeoffs may be required to achieve improved signal-to-noise ratios or larger sample sizes. The use of spatial classifiers is still in early development and quantitative results on the effects of spatial resolution are still limited.

3.5 Signal-to-Noise Ratio

For a given remote sensing problem the signal is the part of the received spectral response function which is information bearing, and the noise is that part which is non-information bearing. The performance of the pattern recognition system is dependent on the ratio of the signal to the noise (S/N). For remote sensing problems this parameter is difficult to quantify.

There are essentially three types of noise introduced into the pattern recognition system - scene noise, atmospheric noise, and hardware noise. The scene noise consists of the variations in the response which have no informational value for the remote sensing problem being studied. An example would be the variations in the response of the soil when an analyst is trying to discriminate between two crops growing in the soil. Hence, the choice of information classes will affect the signal-to-noise ratio.

The atmospheric noise includes variations in the absorption and scattering of the electromagnetic energy in the atmosphere. The visible regions of the spectrum

tend to suffer mostly from scattering in the atmosphere. The infrared portions are very susceptible to absorption particularly in certain bands known as water absorption bands (Korb, 1969).

The noise generated in the sensor system hardware comes from the thermal and shot noise introduced by the optics, the detectors, and the electronics. In addition quantization noise is added by the sensor (Billingsley, 1975).

Of interest here, is the effect of the noise on the overall performance of the system and in particular on the choice of the spectral parameters.

Intuitively one would expect the noise to be a limiting factor on the classification performance. Because of the randomness of the spectral response at the earth's surface, the probability distributions will overlap even if no atmospheric or hardware noise is added. Hence, in general there is some inherent classification performance which cannot be improved upon due to scene and atmospheric noise. However, noise introduced in the hardware can degrade this inherent performance.

Several research efforts have been directed at determining the effect of noise on the system performance. In each case the noise was modeled as additive white Gaussian noise. In an experiment reported by Ready et al (1971) pseudo-random noise was generated on a digital computer and added to multispectral data taken over an

agricultural scene. The classification performance was estimated for varying amounts of added noise power. The results showed that the overall classification performance decreased with an increase in the noise level. Also, it was shown that a class which was the most difficult to identify with low noise levels suffers the most degradation when noise is added.

In a similar experiment, using data taken by the MSDS scanner, Landgrebe et al (1976), also, demonstrated the performance degradation due to added noise. An interesting result in this experiment was that the degradation was significantly less when a spatial classifier was used.

In another research result Mobasseri et al (1978) studied the relationship between the spatial representation by the sensor and the signal-to-noise ratio. Noise was added to simulated multispectral data statistics, and it was concluded that the added noise reduced the class separabilities and degraded the classification accuracy.

The effect of additive white Gaussian noise on the Karhunen-Loeve expansion can be demonstrated quite easily. The covariance matrix for white noise with variance σ_n^2 in N dimensions is

$$K_n = \begin{bmatrix} \sigma_n^2 & 0 & \dots & 0 \\ 0 & \sigma_n^2 & & \vdots \\ & & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_n^2 \end{bmatrix}$$

A linear transformation on the noise such as the transformation determined by the Karhunen-Loeve expansion does not change K_n . For additive white Gaussian noise the signal covariance and noise covariance are additive

$$K = K_s + K_n$$

After the KL expansion the transformed covariance matrix is the diagonal matrix given by

$$K = \begin{bmatrix} \gamma_1 + \sigma_n^2 & 0 & \dots & 0 \\ 0 & \gamma_2 + \sigma_n^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \gamma_N + \sigma_n^2 \end{bmatrix}$$

If the signal is of dimension N' then the eigenvalues for the terms greater than N' are equal to σ_n^2 . The plot of the locus of the eigenvalues corresponding to the terms in the expansion is shown in Figure 3.8. The eigenvalues become constant at the value σ_n^2 for N greater than N' . The signal-to-noise ratio for each channel, then, is γ_i / σ_n^2 .

The weighted Karhunen-Loeve expansion can be used to good advantage when it is known that certain portions of the spectral interval have low S/N. By weighting those

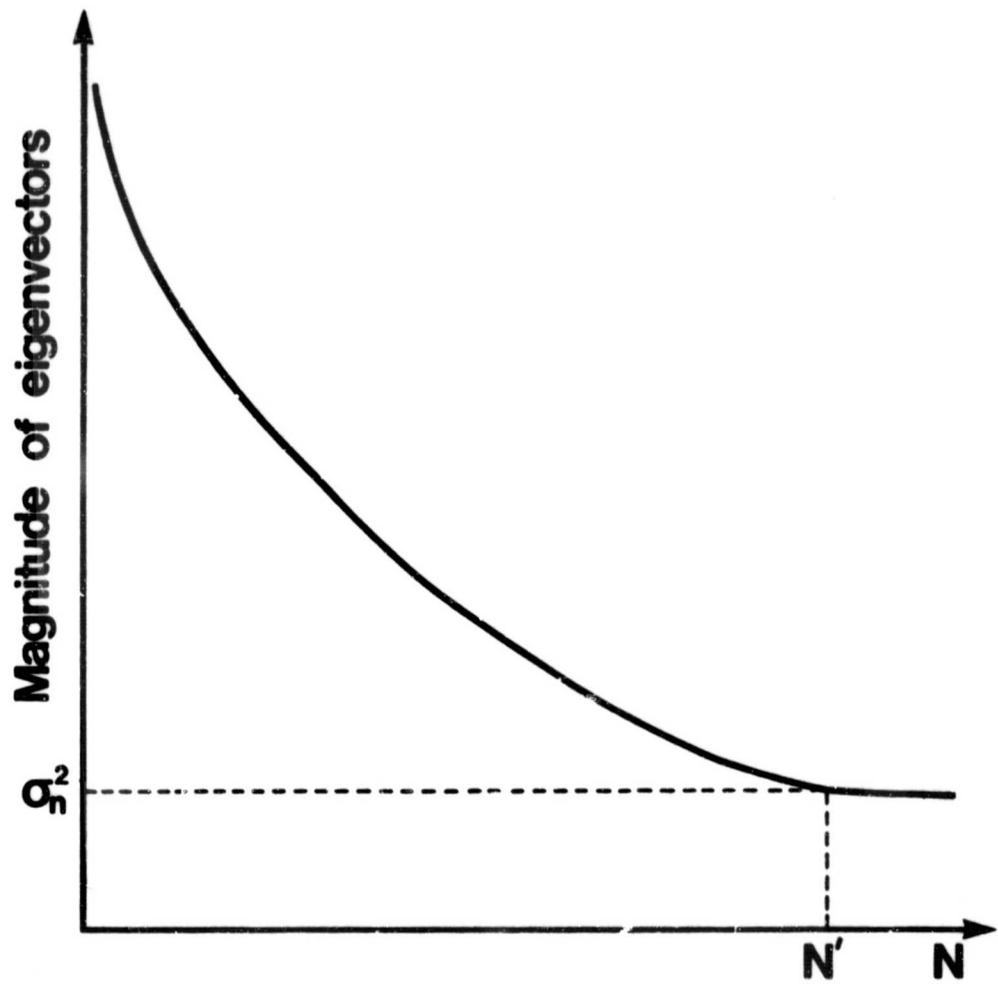


Figure 3.8 The locus of eigenvalues for an N' dimensional signal in white Gaussian noise.

portions with high S/N more heavily, the eigenvectors will tend to be more sensitive to the regions with high S/N. In effect basis functions which have significant components from regions with low weights will have smaller eigenvalues; hence, they will be ranked lower in the ordering of the basis functions.

In general noise from any source tends to make discrimination between information classes more difficult. The degree of the performance degradation depends upon the statistical separability of the classes. Improvements in the signal-to-noise ratios are most helpful when the separability is small.

It is important to realize that one cannot simply specify a high signal-to-noise ratio without considering the other parameters. Because of the law of conservation of energy, the amount of received energy in a fixed spectral band over a fixed surface area at a given time is determined. Therefore, in order to improve S/N, it is necessary to modify the spectral representation parameter, spatial representation parameter, or both.

We have listed one parameter from each of the five categories which is believed to be significant. It is important to note that a change in any one of the parameters--mean-square representation error ϵ_r , the size of the ground resolution element Δ , the signal-to-noise ratio, the number

of sample functions per class, or the set of information classes - frequently causes a change in the optimal value of one or more of the remaining parameters.

One can conceive of an experiment in which a data set is constructed which is large enough to include several values for each of the parameters. An algorithm could be devised to optimize ϵ_0 over the set of parameters with respect to a set of constraints which may be placed on a sensor system. At this time, however, a data set which would satisfy these requirements is not available.

As stated before the spectral parameter is of primary importance in this investigation. Due to the dependence on the other parameters the conditions on the other parameters must be stated. The size of the ground resolution element will be a constant for each data set. The same instrument will be used at the same altitude for all observations. Also, since the same instrument and calibration procedures are used, the noise due to the hardware will be constant. The noise due to atmospheric and scene variations, however, may change from stratum to stratum. The number of sample functions per class will vary, but in each case the number should be sufficiently large to obtain reliable results. The information classes will vary from location to location and for different dates of collection.

CHAPTER 4. EXPERIMENTAL SYSTEM AND RESULTS

A software system has been developed which implements the sensor design procedure described previously. The software package basically consists of an algorithm to compute eigenvalues and eigenvectors, an algorithm to transform the data, a suboptimal sensor simulator, and a method of estimating classification performance. A very necessary part of the experimental system is the field measurements data library consisting of spectra taken over typical agricultural scenes. A block diagram showing the essential parts of the sensor design system is displayed in Figure 4.1. This system has been implemented on the IBM 370/148 at the Laboratory for Application of Remote Sensing at Purdue University.

This chapter begins with a description of the field measurements data base and how it is accessed to provide spectral data for the sensor system design. The software required to compute the eigenvalues and eigenvectors for an ensemble, to perform linear transformations, to simulate suboptimum sensors, and to estimate classification performance is described. The experimental procedure which is used to test the software system is presented and results

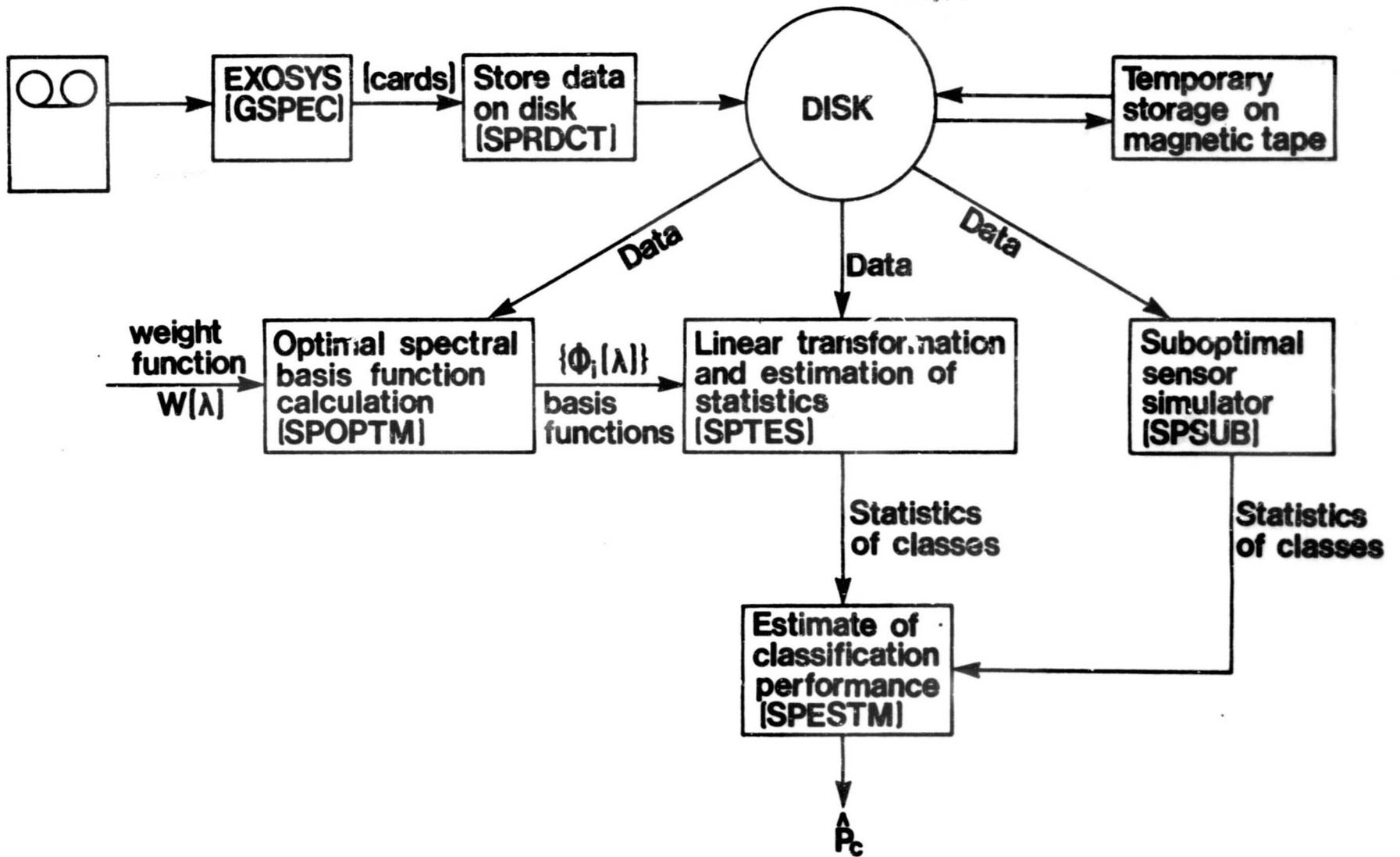


Figure 4.1 Spectral parameter design system.

from using the system are displayed. An important by-product of the sensor design procedure is an increase in the understanding of the scene. Knowledge of some important scene characteristics is extracted with the optimal design system and procedure. The procedure is used to develop a proposed sensor design which is compared against the optimal design for each stratum. A discussion of the overall pattern recognition system performance using the proposed sensor is given.

4.1 Field Measurements Data Base

The field measurements data base consists of spectral samples taken with very fine spectral resolution by the Field Spectrometer System (FSS) mounted in a helicopter. The spectral resolution was 0.02 micrometers for the interval from 0.4 to 2.4 micrometers. The spectra that will be used to test and evaluate the method developed here were collected over each of two sites at three different times of the year.

Field data were taken over Williams County, North Dakota on May 8, June 29, and August 4, 1977. The three principal information classes are SPRING WHEAT, FALLOW, which are fields plowed regularly to conserve moisture, and PASTURE. For the May 8, observation date the wheat was about 8 cm high so that the wheat field would be expected to have spectral characteristics very similar to

bare soil; hence, one would expect that it would be quite difficult to distinguish between the WHEAT and the FALLOW classes. The second date, June 29, provided data during the period of the growing season when the wheat is full grown and is typical of green vegetation. The final date, August 4, provided a data set containing fields with mature wheat. Some of the wheat fields were harvested by August 4; making it necessary to add the class HARVESTED WHEAT.

A second location in Finney County, Kansas was chosen as an example of similar classes in a different location. Three dates, September 28, 1976, May 3, 1977, and June 26, 1977, were chosen corresponding to the growth stages emerging, full canopy, and mature. Other crops in nearby fields, notably grain sorghum, are ripe on the fall date and emergent on the spring date. The information classes used for this data set are WINTER WHEAT, FALLOW, and OTHER CROPS.

The data sets are assembled and stored on disk in a format that is used by all routines that require access to the data. Details of the data set assembly along with the data storage format specification are described in Appendix B. Also, in the appendix complete information on each of the six data sets is listed.

4.2 Spectral Parameter Evaluation System

The key system elements in the spectral parameter evaluation system are the processors SPOPTM, which computes the optimal basis functions, SPTES which uses the basis functions to transform the data, and SPSUB which simulates suboptimal sensors (Figure 4.1).

The computation of the optimal set of basis functions for an ensemble is accomplished by solving the matrix equation

$$\phi \Gamma = KW \phi \quad (4.1)$$

to get the eigenvalues $\gamma_1, \gamma_2, \dots, \gamma_N$ and the eigenvectors $\phi_1, \phi_2, \dots, \phi_N$. The matrix ϕ is the matrix of eigenvectors, $\phi = [\phi_1, \phi_2, \dots, \phi_N]$ and Γ is the diagonal matrix of eigenvalues.

$$\Gamma = \begin{bmatrix} \gamma_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & & & & & \cdot \\ \cdot & & \gamma_2 & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & & \gamma_N \end{bmatrix}$$

The matrix W is a diagonal matrix of weight coefficients

$$W = \begin{bmatrix} w_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & & & & & \cdot \\ \cdot & & w_2 & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & & w_N \end{bmatrix}$$

K is the covariance matrix for the ensemble. Let the mean vector for the ensemble be $\underline{M} = [m_1, m_2, \dots, m_j]^T$, then

$$k_{ij} = E \{ (x_i - m_i) (x_j - m_j) \} \quad (4.2)$$

The unbiased estimate is

$$k_{ij} = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} (x_{ik} - m_i) (x_{jk} - m_j) \quad (4.3)$$

where N_s is the number of sample functions in the ensemble.

Note that in general the stochastic process is non-stationary. A zero-mean process is defined to be stationary in the wide sense if the covariance function depends only on the difference $|\lambda - \xi|$ (Papoulis, 1965). That is,

$$K(\lambda, \xi) = K(\lambda - \xi) \quad \lambda, \xi \in \Lambda$$

The covariance matrix of a stationary process has elements which are equal along the diagonals. The methods used to compute the covariance matrices and to compute eigenvalues and eigenvectors are valid for both stationary and non-stationary stochastic processes.

Let A be the matrix product of the covariance matrix K and the diagonal weighted matrix W .

$$A = KW \quad (4.4)$$

If the weighting matrix W is equal to the identity matrix I , then the kernel A is real symmetric and the solution to 4.1 can be found using a standard numerical algorithm known as the Jacobi method (Wilkinson, 1965, p 266). The Jacobi method uses a sequence of similarity transformations to reduce a real-symmetric matrix to a diagonal matrix. This method is very stable and provides all of the eigenvalues and eigenvectors with good precision.

However, if W is not the identity matrix, then, A is not symmetric. An algorithm which solves the eigenvalue problem for real general matrices was published by Grad and Brebner (1968). This algorithm, EIGENP, computes the eigenvalues by the QR double-step method and the eigenvectors by inverse iteration. Some comments on the application of the algorithm to the specific computer used here were published by Niessner (1972).

The complete algorithm package consists of the main subroutine EIGENP and four callable subroutines SCALE, HESQR, REALVE, and COMPVE. Subroutine SCALE scales the matrix so that the absolute sums of corresponding rows and columns are roughly equal. The scaled matrix is then normalized so that the Euclidean norm is equal to one. These two preliminary modifications are carried out to improve the accuracy of the computed results. In HESQR the scaled matrix is reduced to upper-Hessenberg form by Householder's method. The QR double-step iterative process

is performed on the Hessenberg matrix to reduce the matrix to diagonal form within the computational accuracy limits, where the elements along the diagonal are the eigenvalue. The inverse iteration process to find corresponding eigenvectors is carried out in REALVE for real eigenvalues and COMPVE for complex eigenvalues. Since it has been shown that the eigenvalues will be real for the application under consideration, there is no need to include COMPVE.

Both the EIGENP algorithm and the Jacobi method have been tested on the same covariance matrix using the identity matrix as the weight. The differences using the two methods were negligible even for matrices of order 100.

A necessary part of the Karhunen-Loeve expansion is the ordering of the eigenvalues and corresponding eigenvectors. Since the eigenvalues are not ordered in the eigenvalue algorithm a sorting routine was added to the system to perform this task.

The set of ordered eigenvectors $\{\phi_i(\lambda)\}$ will be used to perform a linear transformation on the original data vectors X . To perform the linear transformation the coefficients corresponding to each eigenvector are computed. Instead of the vector X , the waveforms are represented by the set of coefficients $\{x_i\}$ where

$$x_i = \int_{\Lambda} [x(\lambda) - m(\lambda)] \phi_i(\lambda) w(\lambda) d\lambda \quad (4.5)$$

or in terms of discrete vectors

$$x_i = \phi_i^T W[X - M] \quad (4.6)$$

This transformation on the field data is performed in the program SPTES.

The statistics for each information class are needed to evaluate the probability of correct classification. The data set, now represented by the transform coefficients, is partitioned into classes and the corresponding mean vectors and covariance matrices are computed in SPTES. The maximum likelihood estimates are used for the mean vectors and covariance matrices.

The routine SPSUB was developed to simulate several suboptimal sensors. A set of N basis functions $\{\psi_i(\lambda)\}$ is stored in memory where each function is approximated by a 100 element vector. As an example, a set of four vectors, $\psi_1(\lambda)$, $\psi_2(\lambda)$, $\psi_3(\lambda)$, $\psi_4(\lambda)$ was implemented where

$$\psi_i(\lambda) = \begin{cases} 1.0 & \lambda_i \leq \lambda \leq \lambda_{i+1} \\ 0.0 & \text{elsewhere} \end{cases} \quad (4.7)$$

The endpoints λ_i and λ_{i+1} are given under sensor number i in Table 4.1. The basis functions may be normalized by requiring

$$\int_{\Lambda} [\psi_i(\lambda)] w(\lambda) d\lambda = 1 \quad (4.8)$$

Each waveform in the ensemble is approximated by

$$x(\lambda) \approx \sum_{i=1}^4 x_i \psi_i(\lambda) \tag{4.9}$$

where

$$x_i = \int_{\Lambda} x(\lambda) \psi_i(\lambda) w(\lambda) d\lambda \tag{4.10}$$

For the normalized basis functions the expected mean-square representation error over the ensemble, is given by

$$E\{e_r\} = E \left[\int_{\Lambda} \left[x(\lambda) - \sum_{i=1}^4 x_i \psi_i \right]^2 w(\lambda) d\lambda \right] \tag{4.11}$$

A second sensor which has been considered for practical implementation and which has band edges given under sensor number two in Table 4.1 has, also, been included in the routine SPSUB

Table 4.1 Spectral band locations for two practical sensor designs.

Sensor Number 1		Sensor Number 2	
<u>Band</u>	<u>Wavelength</u>	<u>Band</u>	<u>Wavelength</u>
1	0.5 μm to 0.6 μm	1	0.45 μm to 0.52 μm
2	0.6 μm to 0.7 μm	2	0.52 μm to 0.60 μm
3	0.7 μm to 0.8 μm	3	0.63 μm to 0.69 μm
4	0.8 μm to 1.1 μm	4	0.76 μm to 0.90 μm
		5	1.55 μm to 1.75 μm
		6	2.08 μm to 2.35 μm

The output of both SPTES and SPSUB is a set of statistics from which it is desired to evaluate the global performance criterion of probability of correct classification. In pattern recognition terminology the estimation of the class conditional statistics is the training phase or design of the classifier. It now remains to use these training sets to compute the performance. A Monte Carlo technique has been developed to evaluate the probability of correct classification integral. The details of the technique and an evaluation of an algorithm, SPESTM, designed to implement the technique are covered in Appendix A. A sufficient number of representative spectral response functions to represent the stratum is necessary in order to obtain a good estimate of the statistics. Experience with the performance estimator algorithm has demonstrated that the algorithm is reasonably efficient in terms of execution time and accuracy.

4.3 Experimental Procedure

In this section some comments concerning the procedures for the operation of the spectral parameter design system are made. These procedures are followed in generating the results that are given in later sections.

A stratum is selected by choosing a location and collection date for which a set of field data has been

acquired and stored in the field measurements library. Sample spectral response functions are selected from the field data to represent the stratum. This selection is accomplished by specifying the tape that a particular data set is stored on and the date on which it was collected. Details concerning this procedure are covered in Appendix B. The deck of cards, containing the numerical values of the spectra, is read by SPRDCT which stores the response functions and some ID information onto a disk file. All of the analysis algorithms using the data require the data to be in the format described in the appendix.

The estimate of the covariance matrix of the ensemble and the solutions to the matrix equation which gives the eigenvalues and eigenvectors are computed by the routine SPOPTM. A weight function which is stored as a vector in a callable subroutine is selected in SPOPTM. A subroutine is used to sort the eigenvalues and corresponding eigenvectors such that the eigenvalues are in descending order of magnitude.

An example of the output listing for SPOPTM which lists the first 30 eigenvalues is shown in Figure 4.2. Corresponding to each eigenvalue estimate is an estimate of the variance of the eigenvalue, an estimate of the variance of the eigenvector, and the expected mean-square representation error for using the Karhunen-Loeve expansion.

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY
SAMPLE FUNCTION INFORMATION 17 JULY, 1978

```

EXP. NO.....140
NUMBER OF CLASSES.....3
CLASS.....WHEAT
NUMBER OF SAMPLE FUNCTIONS.....658
CLASS.....FALLOW
NUMBER OF SAMPLE FUNCTIONS.....211
CLASS.....UNKNOWN
NUMBER OF SAMPLE FUNCTIONS.....682
  
```

N	EIGENVALUE	VAR (GAM)	VAR (PHI)	MEAN-SQUARE
1	2671.2320	18423.0508	0.0003	716.056468
2	624.0196	1005.3904	0.0004	92.036884
3	38.0230	3.7328	0.0023	54.013897
4	17.3402	0.7763	0.3666	36.673708
5	16.6256	0.7137	0.3665	20.048093
6	5.3304	0.0734	0.0049	14.717731
7	2.8647	0.0212	0.0107	11.853076
8	1.8119	0.0085	0.0960	10.041203
9	1.6529	0.0071	0.1131	8.388302
10	1.4203	0.0052	0.0491	6.967997
11	0.8294	0.0018	0.0572	6.138592
12	0.6867	0.0012	2.4945	5.451897
13	0.6755	0.0012	2.5225	4.776416
14	0.6203	0.0010	0.2657	4.156083
15	0.5687	0.0008	0.1672	3.587406
16	0.4833	0.0006	0.0706	3.104095
17	0.3670	0.0003	0.1362	2.737054
18	0.3384	0.0003	0.1403	2.398652
19	0.2647	0.0002	0.7469	2.133988
20	0.2567	0.0002	0.7494	1.877260
21	0.2105	0.0001	0.1172	1.666809
22	0.1888	0.0001	0.1336	1.477964
23	0.1656	0.0001	0.1281	1.312349
24	0.1478	0.0001	0.1062	1.164518
25	0.1177	0.0000	0.2557	1.046809
26	0.1112	0.0000	0.2611	0.935630
27	0.0903	0.0000	0.6519	0.845328
28	0.0873	0.0000	0.6752	0.758043
29	0.0786	0.0000	0.1728	0.679479
30	0.0672	0.0000	0.3435	0.612256

Figure 4.2 Sample output from SPOPTM.

The variances are computed using the results derived in Chapter 2. It should be remembered that the eigenvalues and eigenvectors are estimated from random samples from a Gaussian stochastic process. The estimate of the first eigenvalue is 2671.23. This estimate is approximately unbiased and has a standard deviation of 135.7. Similarly the estimate of the norm of the difference between the true eigenvector and the estimate is approximately unbiased. The standard deviation is .02. It is interesting to note that the variance for the 12th and 13th eigenvectors is relatively large. Recalling that the expression for the variance is sensitive to eigenvalues which are close together, the large variances are not surprising. The mean-square error is computed using the eigenvalue estimates. ID information concerning the data set is included for reference. The eigenvectors are punched and stored in a card data file. A plotting routine is used to display the eigenvectors. Also, the eigenvectors will be used later to perform linear transformations on the data.

A crude approximation to the system measurement error, introduced in making the field measurements, is used to provide a comparison with the expected mean-square representation error. Measurement error was assumed to be 7% of the numerical response value. If $x(\lambda)$ is the true signal and $s(\lambda)$ is the measured signal including added noise, then, the measurement error is

$$\epsilon_m = \int_{\Lambda} [x(\lambda) - s(\lambda)]^2 d\lambda \quad (4.12)$$

In discrete form let $x_k - s_k = .07 x_k$ and

$$\epsilon_m = \sum_{k=1}^L (.07 x_k)^2 \quad (4.13)$$

The average ϵ_m over the ensemble is the estimate of the expected measurement error.

The linear transformation on the original data set using the computed eigenvectors is performed using SPTEs. The statistics for the first N terms or features are computed for each class and displayed on the printer (Figure 4.3). Also a card deck with the statistics stored on it is punched for use with the classification performance estimator.

The estimate of the probability of correct classification is obtained by SPESTM (see Appendix A). The statistics deck output of SPTEs is designed to be identical to the required input for SPESTM. The output of SPESTM includes the conditional probability of correct classification for each class and the overall probability of correct classification (Figure 4.4).

It is possible to evaluate the contribution of each feature to the separability of the classes. Feature selection is performed using the SEPARABILITY processor in

MEAN VECTOR

22.1997	17.3667	0.9930	-0.0932
---------	---------	--------	---------

COVARIANCE MATRIX

2576.8550			
-371.4817	716.0183		
-54.7175	3.4695	24.5298	
33.6030	-6.8187	-8.2161	15.4123

MEAN VECTOR

-16.1261	-17.6414	-1.4610	-0.2640
----------	----------	---------	---------

COVARIANCE MATRIX

2100.3152			
-370.0427	144.1309		
-111.1944	-5.3496	48.6596	
-49.1371	33.2773	-2.7137	19.3354

MEAN VECTOR

-16.4757	-11.2648	-0.4962	0.1673
----------	----------	---------	--------

COVARIANCE MATRIX

2112.1230			
-170.6249	168.9570		
50.6295	-32.0785	46.1451	
-14.4906	-1.3899	8.9443	18.5254

Figure 4.3 Sample output of class conditional statistics.

PROBABILITY OF CORRECT CLASSIFICATION FOR CLASS 1 = 0.8308

PROBABILITY OF CORRECT CLASSIFICATION FOR CLASS 2 = 0.8450

PROBABILITY OF CORRECT CLASSIFICATION FOR CLASS 3 = 0.5773

OVERALL PROBABILITY OF CORRECT RECOGNITION = 0.7509

Figure 4.4 Sample output of classification performance estimates.

LARSYS (Phillips, 1973) in which the divergence is computed using various combinations of N' features. The feature sets are ordered according to the average pairwise divergence. This feature selection technique allows one to find the best feature set quickly without having to try all of the possible combinations.

Two practical sensor designs are evaluated for comparison with the optimal design. The spectral bands used to simulate these sensors was presented in Table 4.1. The spectral bands are contained in SPSUB which uses them as a set of basis functions to represent the response function. A linear transformation is performed on the data, and the statistics for each class are computed. The average mean-square error for the suboptimal representation is computed and printed. The statistics are again punched on cards in a format suitable for SPESTM.

SPSUB can also be used to design a practical sensor. The program can be modified to include any choice of spectral bands desired.

4.4 System Testing

The system was exercised in an effort to determine its capabilities and limitations. The data sets taken over the two locations at different times were used in the tests. In particular it would be good to get some feel as to what would be a good choice for the weight function. Also the

number of samples that are required will be important when specifying what is desired in future data sets.

4.4.1 Reconstruction

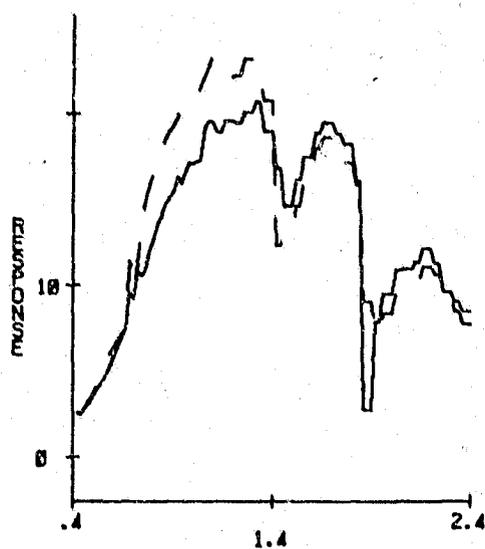
As a first test of the system we would like to demonstrate the capability of the first few terms in the Karhunen-Loeve expansion to reconstruct the original waveform. A sample spectral response function from an ensemble is selected and the coefficients in the expansion are computed. Using N' terms in the expansion the approximation to the original function is given by:

$$\hat{x}(\lambda) = \sum_{i=1}^{N'} x_i \phi_i(\lambda) + m(\lambda) \quad (4.14)$$

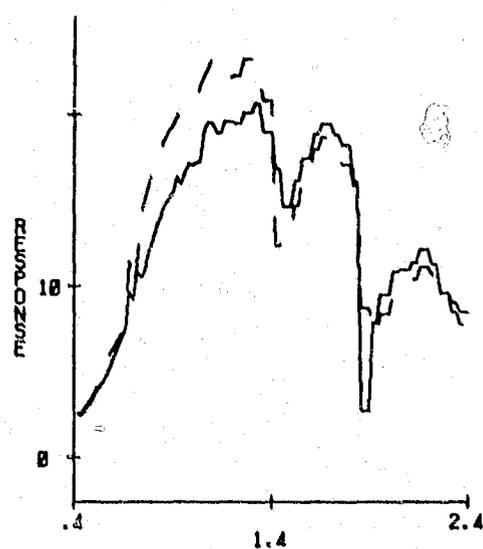
where $m(\lambda)$ is the mean function of the process. For this example a uniform weight function, $w(\lambda) = 1.0$ for all $\lambda \in \Lambda$, was used.

A sequence of graphs showing the original function $x(\lambda)$ as a solid line and the approximated function $\hat{x}(\lambda)$ as a dashed line is shown in Figures 4.5a to 4.5h. Only the first term in the expansion is used in Figure 4.5a; the first two terms are used in Figure 4.5b, and so forth. It is readily observed that after a few terms the approximation is very close to the original.

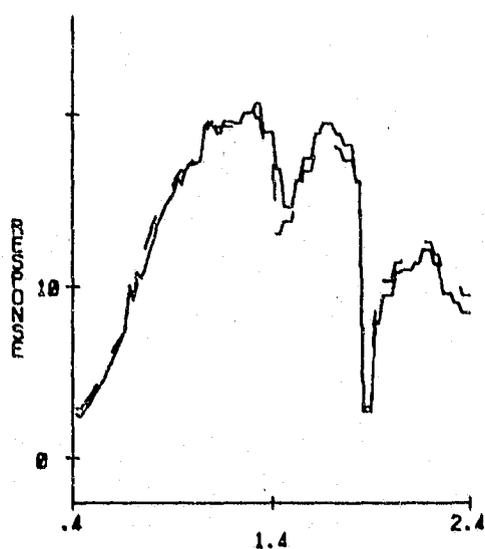
If the average mean-square error is computed directly using the equation



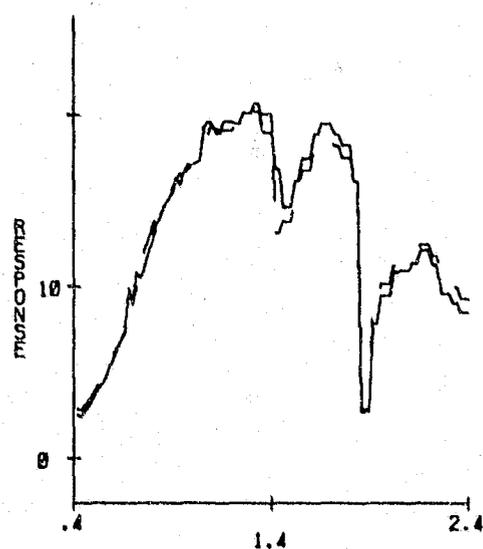
1 TERM



2 TERMS

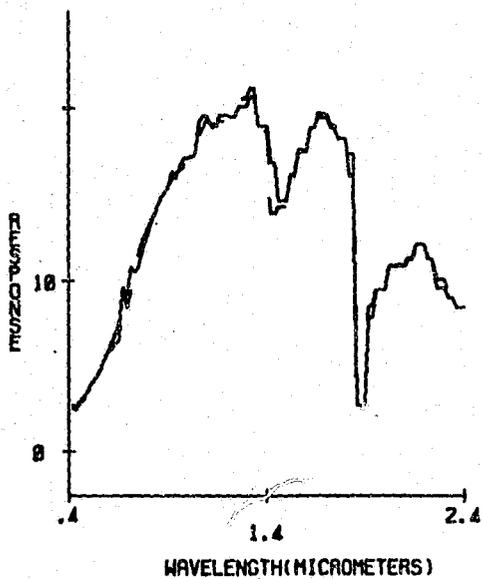


3 TERMS

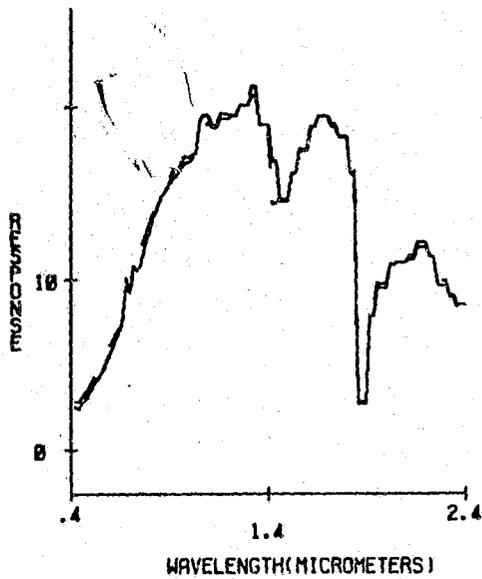


4 TERMS

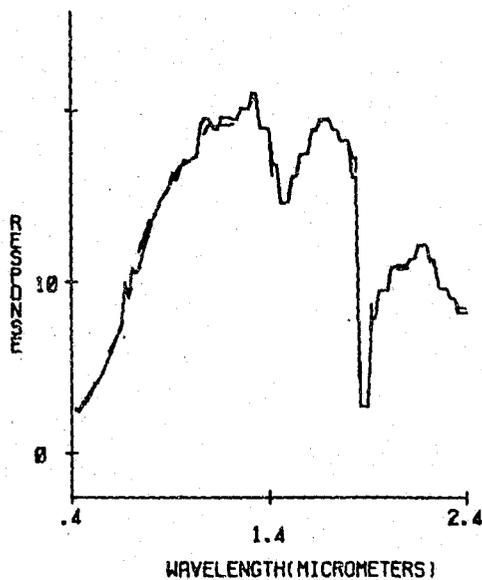
Figure 4.5 Reconstruction of a single spectral response function using from 1 to 8 terms in the expansion.



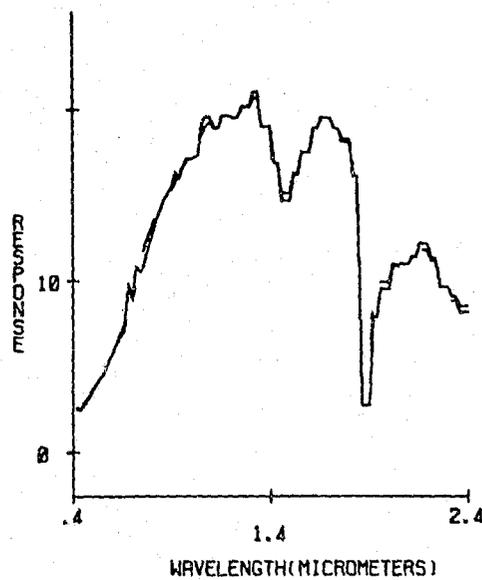
5 TERMS



6 TERMS



7 TERMS



8 TERMS

Figure 4.5 (cont.)

$$\epsilon = \int_{\lambda} [x(\lambda) - \hat{x}(\lambda)]^2 d\lambda \quad (4.15)$$

and averaging over the ensemble, the value of $E\{\epsilon\}$ is equal within numerical error to the value given by summing the eigenvalues

$$E\{\epsilon\} = \sum_{i=N'+1}^{\infty} \gamma_i \quad (4.16)$$

as predicted by equation 2.

4.4.2 Choice of Weight Function

An important part of the analysis procedure is the choice of the weight function $w(\lambda)$ to be used in the weighted Karhunen-Loeve expansion. Four different weight functions, which are displayed in Figure 4.6, were proposed and tested. Data taken over Williams County, North Dakota on May 8, 1977 was used to evaluate the different weight functions. Comparisons were made by evaluating the eigenvalues, eigenvectors and classification performances for each of the weight functions.

The motivation for the development of the weighted Karhunen-Loeve expansion is demonstrated by using the first weight function which has a weight of one assigned to all wavelengths on the spectral interval (Figure 4.6a). The first four eigenvectors for this weight function are graphed in Figure 4.7. It is noted that the first

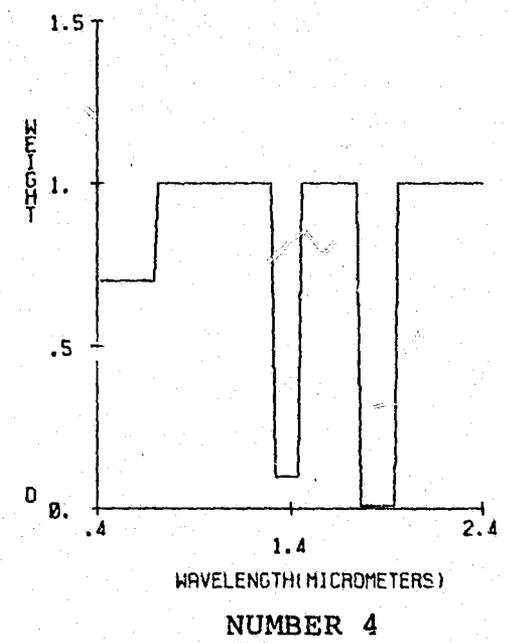
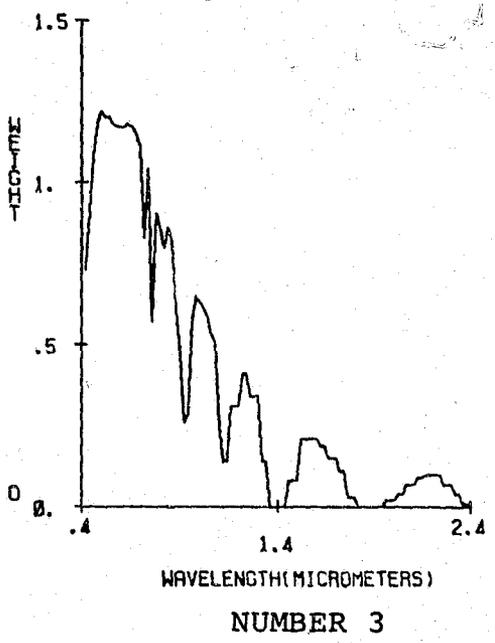
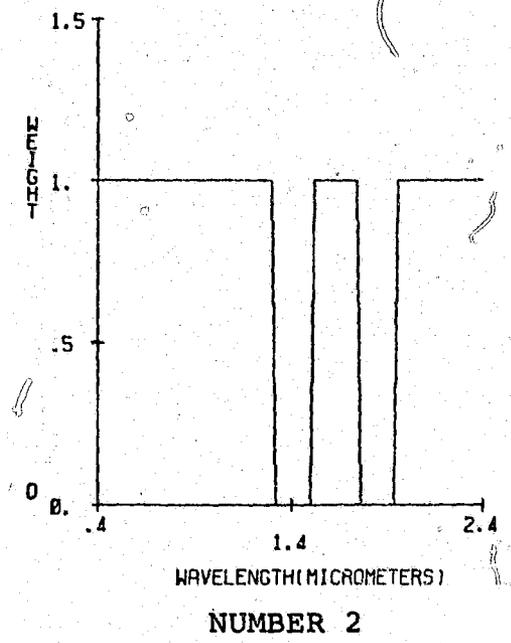
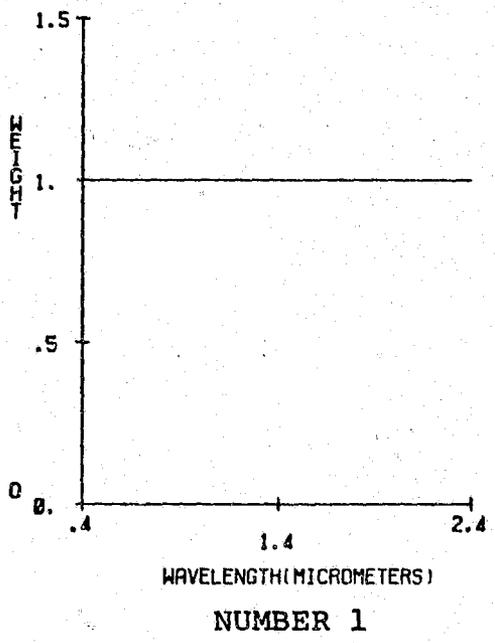
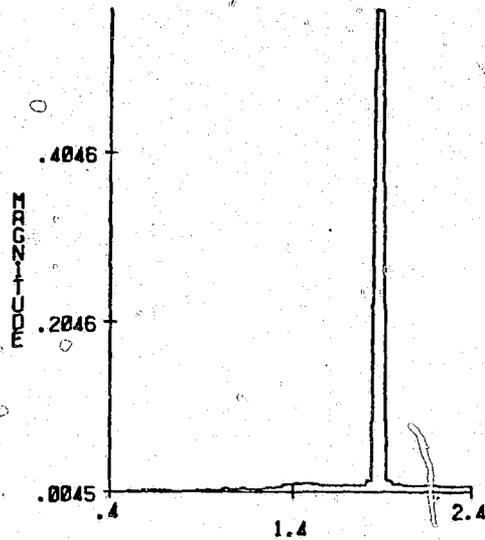
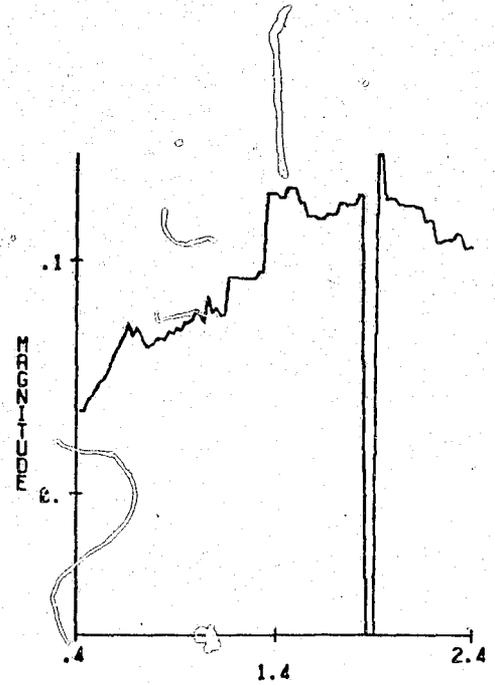


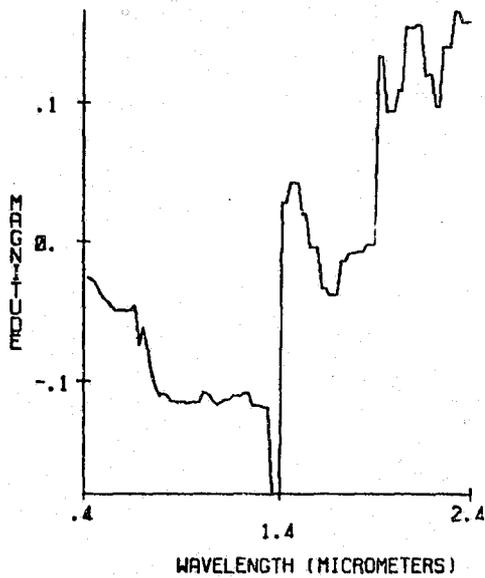
Figure 4.6 Weight functions.



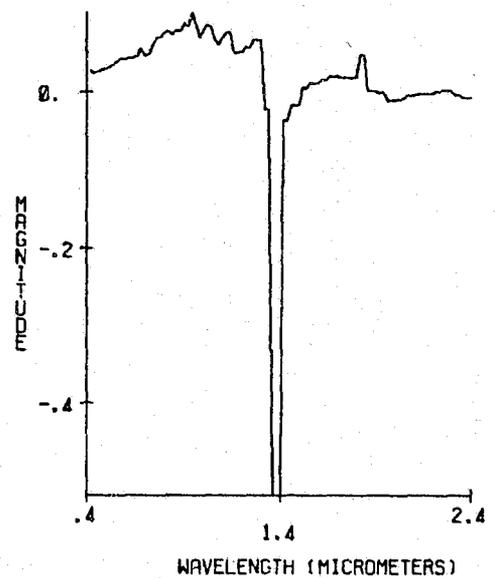
EIGENVECTOR 1



EIGENVECTOR 2



EIGENVECTOR 3



EIGENVECTOR 4

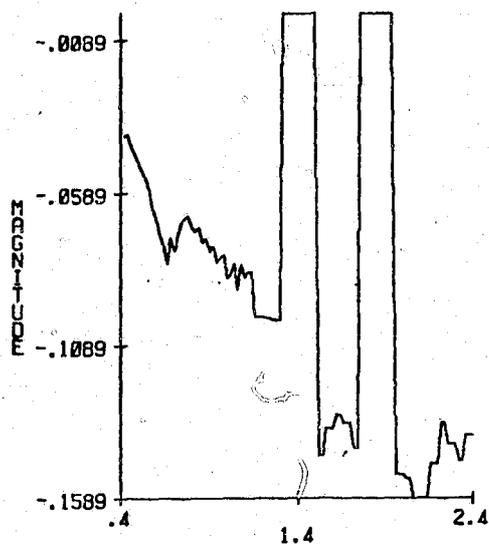
Figure 4.7 First four optimal basis functions using weight function number 1 over Williams County, May 8, 1977 data.

eigenvector is dominated by the variance in the signal for a narrow interval near 1.9 μm . A similar peak for the same interval occurs in the second eigenvector. In the fourth eigenvector a similar result occurs for a region near 1.4 μm . These two bands near 1.4 and 1.9 μm correspond to water absorption bands which severely attenuate the electromagnetic energy passing through the atmosphere at these wavelengths. The sample spectral response functions have large variations in these bands which causes the eigenvalue algorithm to select one or more eigenvectors which are sensitive almost entirely to the portion of the spectral interval corresponding to one of the water absorption bands. The source of these large variations is traced to the calibration procedure during which a division by a small number occurs, resulting in the noisy signals in the respective bands. The ability of eigenvectors 2 and 4 to aid discrimination between information classes is limited and real contributions to the performance for these eigenvectors are due to the small but finite sensitivity in the remainder of the spectrum.

The three remaining weight functions were chosen to minimize the effects of the water absorption bands. In the second weight function (Figure 4.6b), the weight is set equal to .001 for the intervals 1.32 to 1.50 μm and 1.76 to 1.94 μm and equal to one elsewhere. A more radical

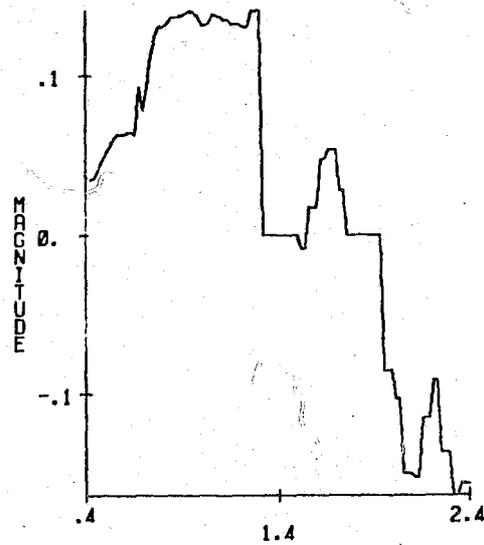
choice of weight function (Figure 4.6c) is based on the solar spectral irradiance at sea level (Handbook of Geophysics, 1961). The solar irradiance is strongest in the visible and decreases to very small values in the infrared. The two water absorption bands are accounted for as well as several other lesser molecular absorption bands. A criticism of this choice of weight function is that the reflectance from vegetation, for example, is very low in the visible while it is quite high in the infrared, which is the opposite of the solar irradiance curve. Hence, the third weight function, based on the irradiance curve will tend to give too much importance to the visible region and too little importance to the infrared regions, especially those between 1.5 and 1.7 μm and those between 2.2 and 2.4 μm . The fourth weight function was chosen to weight the low reflectance typical of the visible region lower. It has slightly higher weight values for the two water absorption bands and has a weight of 0.7 for the visible region. The first four eigenvectors for weight functions 2, 3, and 4 are shown in Figures 4.8, 4.9, and 4.10, respectively.

The expected value of the integral over Λ of the square of the response functions can be treated as a total received signal energy. This expected energy is equal to the sum of all of the eigenvalues, which is different for



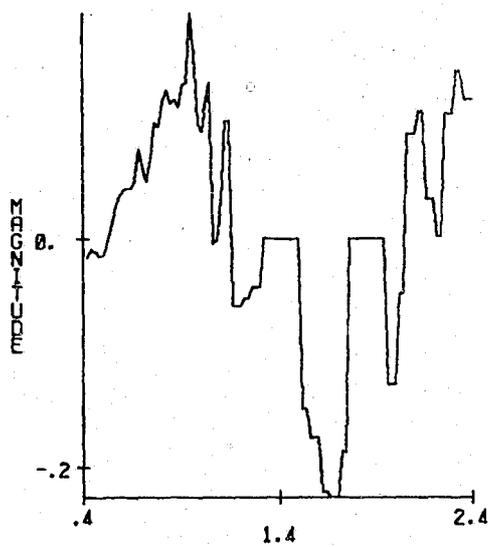
WAVELENGTH (MICROMETERS)

EIGENVECTOR 1



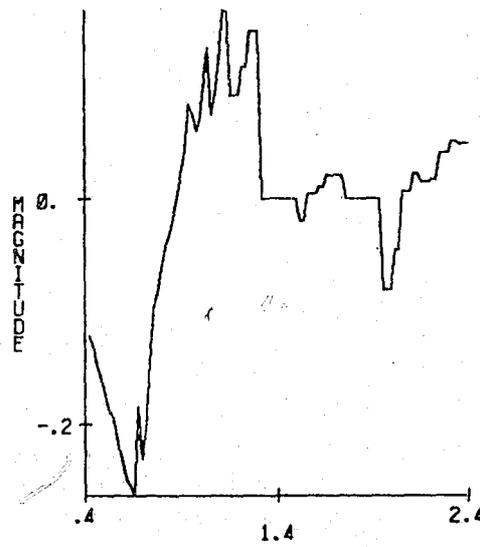
WAVELENGTH (MICROMETERS)

EIGENVECTOR 2



WAVELENGTH (MICROMETERS)

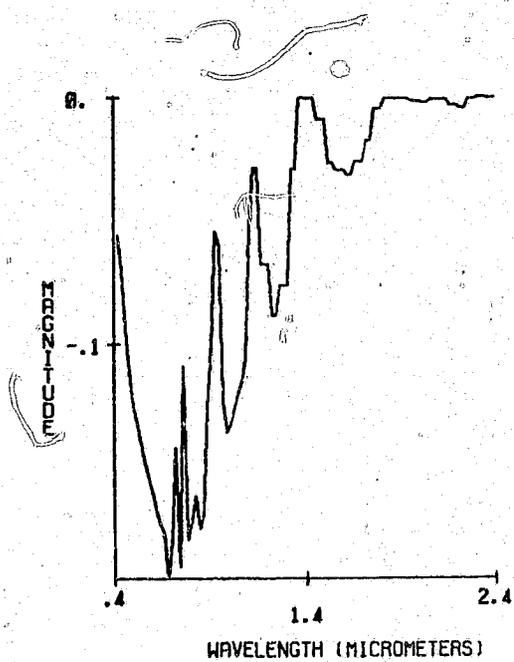
EIGENVECTOR 3



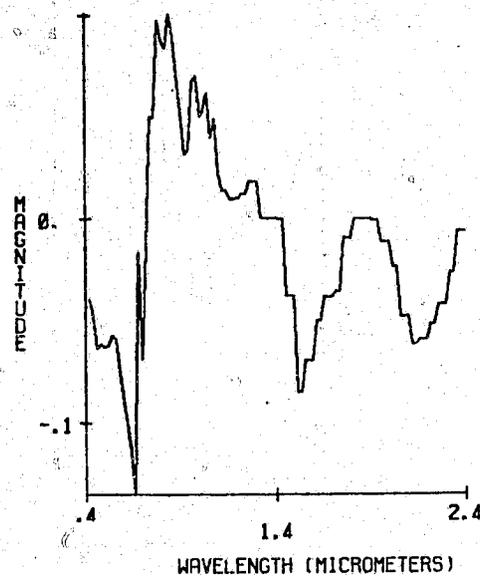
WAVELENGTH (MICROMETERS)

EIGENVECTOR 4

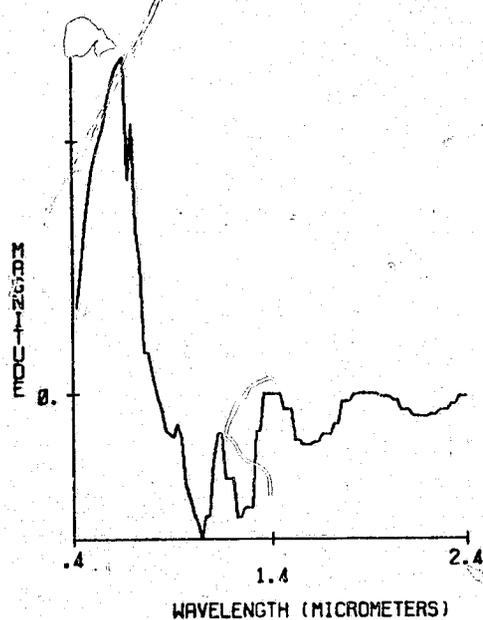
Figure 4.8 First four optimal basis functions using weight function number 2 over Williams County, May 8, 1977 data.



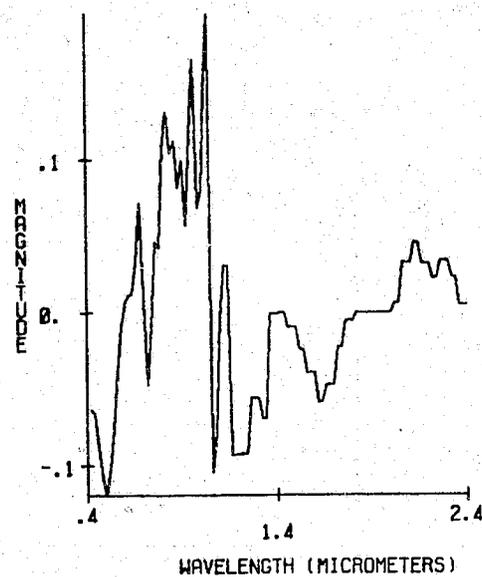
EIGENVECTOR 1



EIGENVECTOR 2



EIGENVECTOR 3



EIGENVECTOR 4

Figure 4.9 First four optimal basis functions using weight function number 3 over Williams County, May 8, 1977 data.

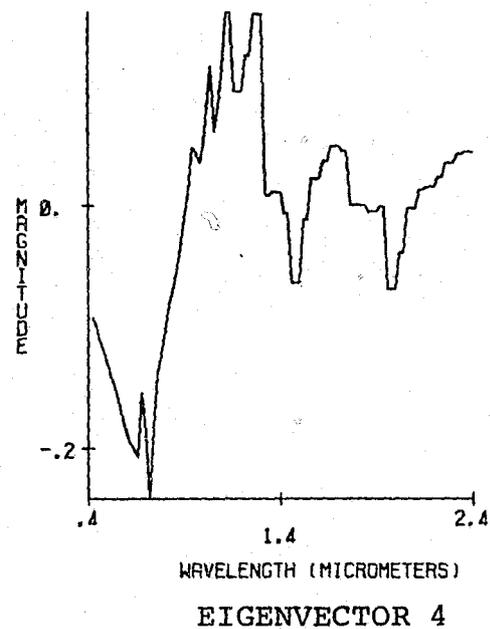
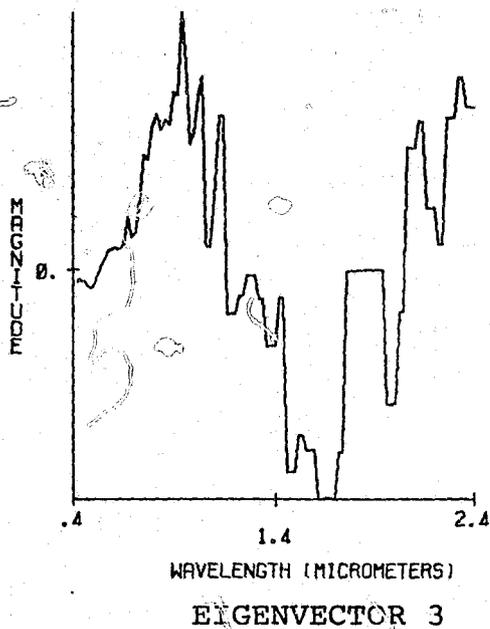
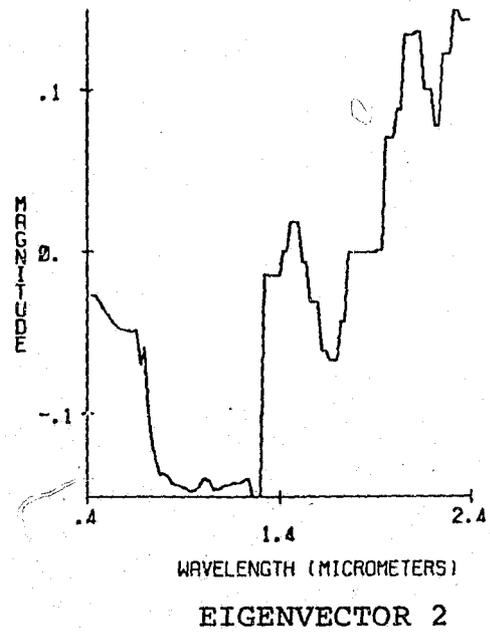
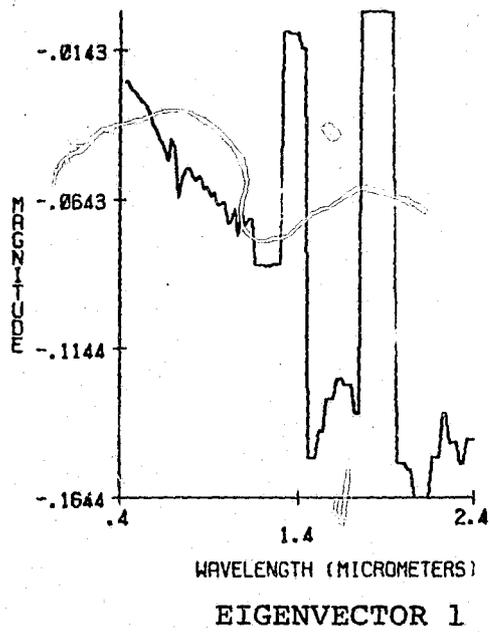


Figure 4.10 First four optimal basis functions using weight function number 4 over Williams County, May 8, 1977 data.

each weight function. The expected energies are 1985, 550, 177, and 589, respectively where the units are relative to the units of per cent response for the spectral response functions. Ideally, the weight function would reduce the energy which is noise and retain that which is signal. By using low weights in the water absorption bands, an improvement in overall signal-to-noise ratio has been gained. However, in the case of the third weight function, the reduction in energy may have been too much. The infrared regions are not represented significantly in any but perhaps the second eigenvector.

As a final comparison between weight functions, the classification performances are examined. In Table 4.2 the estimate of the probability of correct classification as a function of the number of terms in the expansion for each of the four weight functions. For ten terms it appears that the second and third weight functions are the better choices with the second weight function demonstrating a slight advantage in the first few terms. The conclusion drawn at this point is that the second weight function is the most reasonable choice and will be one used in the results that follow.

Table 4.2 Comparison of the probability of correct classification using N terms in the weighted Karhunen-Loeve expansion among four choices of weight functions.

N	Weight Function			
	1	2	3	4
1	.355	.467	.468	.488
2	.443	.729	.675	.730
3	.729	.819	.700	.799
4	.736	.833	.806	.817
5	.742	.851	.853	.822
6	.794	.882	.896	.834
7	.807	.894	.897	.851
8	.823	.943	.914	.860
9	.851	.943	.956	.889
10	.862	.949	.954	.931

4.4.3 Evaluation of the Eigenvalue Algorithm

The methods employed in the algorithm EIGENP have been well-studied (see Wilkinson, 1965) and are characterized by good numerical stability and accuracy even for covariance matrices which have a rank of 100. The accuracy of the algorithm depends largely on the particular machine on which the algorithm is implemented. The accuracy is proportional to the rank of the matrix, to the number of iterations required for the iterative procedures used, and to 2^{-t} where t is the number of significant digits in the mantissa of a binary floating point number. For the IBM 370 machine using double-precision the value of t is 56. Typically eigenvalues can be computed which are accurate to

six decimal places. The norm of the difference between a computed eigenvector and the true eigenvector is also on the order of 10^{-6} .

The accuracy of the computed eigenvalues and eigenvectors deteriorates slightly with the introduction of the weight matrix. Weight matrices containing small weights tend to cause underflow conditions to occur in the reduction to Hessenberg form.

Computation times for matrices of rank 100 are on the order of 10 minutes of CPU time. Hence, one is restricted somewhat in using this algorithm a large number of times.

4.4.4 Sample Size

The number of sample functions, used to represent the ensemble, influences both the estimates of the eigenvalues and eigenvectors and the estimate of the classification performance. The prediction of the general effects of the sample size have been described earlier; however, it would be desirable to demonstrate these effects in the context of the present problem for the purpose of deciding whether or not a sufficient number of samples were collected.

An experiment was performed using the data taken over Williams County on August 4, to demonstrate the effect of sample size. Subsets of the ensemble were used to simulate small data set sizes of 55, 110, and 294 sample functions.

The breakdown in the number of samples from the four information classes is shown in Table 4.3. The eigenvalues and eigenvectors were computed using 55, 110, 294, and 1444 samples, respectively. Several sample functions, which were used to compute the eigenvalues, were not used to evaluate the performance because they were from fields in which there was some uncertainty as to which cover type the functions belonged. The eigenvalues and eigenvectors for each case were computed using the second weight function, and the expected mean-square error was plotted as a function of the number of terms in the expansion in Figure 4.11. The effect of sample size on mean-square error is most detectable for the number of terms greater than ten. It is observed that the expected mean-square error increases with increasing sample size.

Table 4.3 Sample size assignments for data from Williams County, N.D. on August 4, 1977.

Class	NUMBER OF SAMPLES			
WHEAT	25	60	134	808
WHEAT HAR	5	10	22	34
FALLOW	15	25	76	330
PASTURE	<u>10</u>	<u>15</u>	<u>62</u>	<u>130</u>
Total	55	110	294	1,302

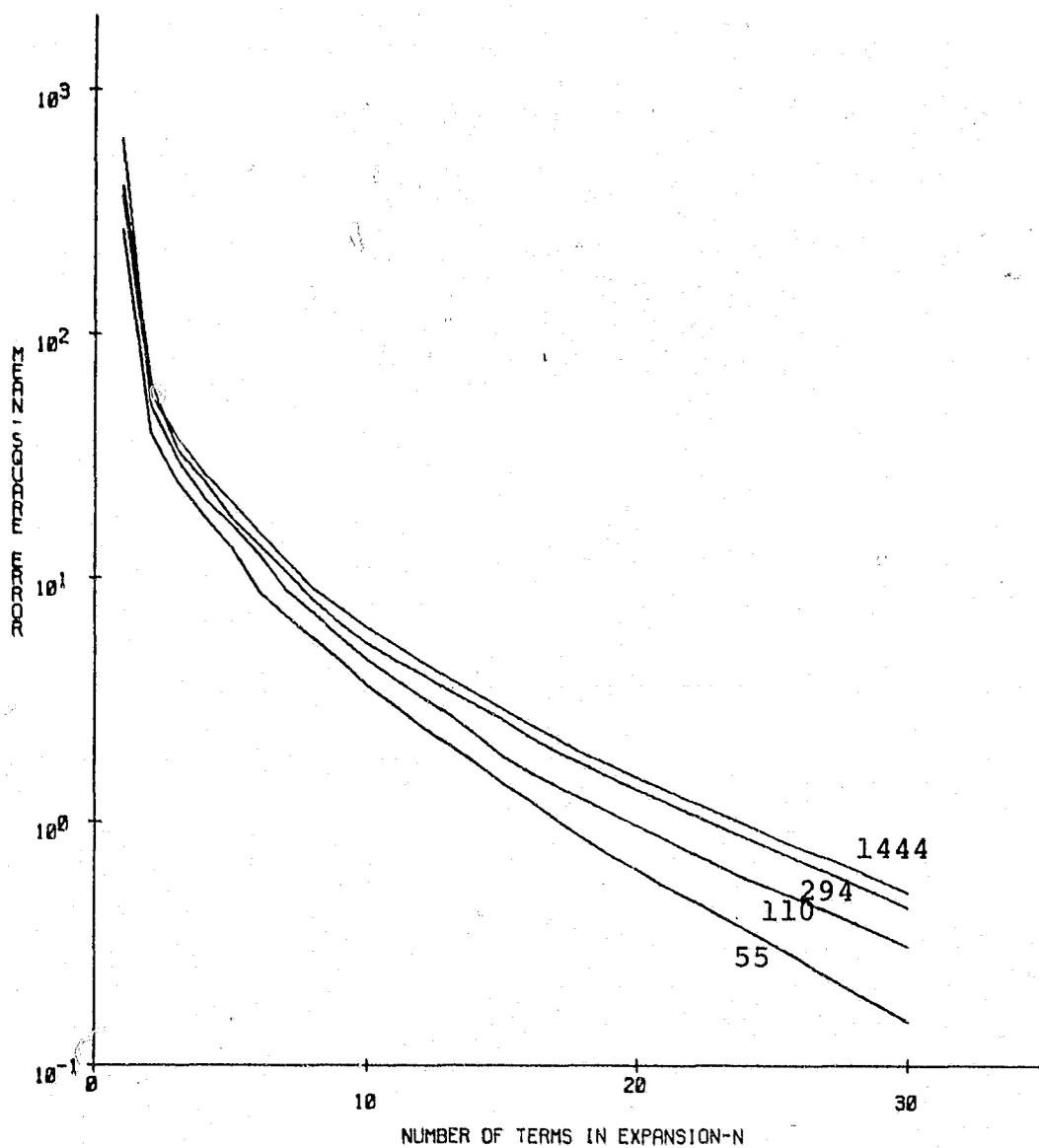


Figure 4.11 Influence of sample size on expected mean-square error for Williams County, August 4, 1977, using weight function number 2.

The classification performance was evaluated for each additional term in the sequence and a plot of \hat{P}_c as a function of the mean-square representation error was drawn for each case (Figure 4.12). The small sample function size has two effects on \hat{P}_c vs. $E\{\epsilon_r\}$ curve. First the smaller mean-square error causes the curve to be further to the left than it should be. Second, the small sample size causes the performance to be higher than it should be for a given expected mean-square error.

The question of whether or not the set of samples adequately represents a stratum is a difficult one. In particular the method of selecting which functions to include in the sample is not easy to determine. One reason is that relatively few sample functions are available and as in the case of this research one uses all the functions that are available. This experiment demonstrates the effects if we assume that the 1444 sample functions accurately represent the ensemble. Certain trends indicate that the number of samples available is adequate. The change in the expected mean-square error is quite small between the curves 297 and 1444 samples in Figure 4.11. Also, the performance as a function of representation error in Figure 4.12 is probably close to accurate for the largest sample size. In the following the ensemble will consist of all of the sample functions that are available which is on the order of 1000. It should be pointed out that

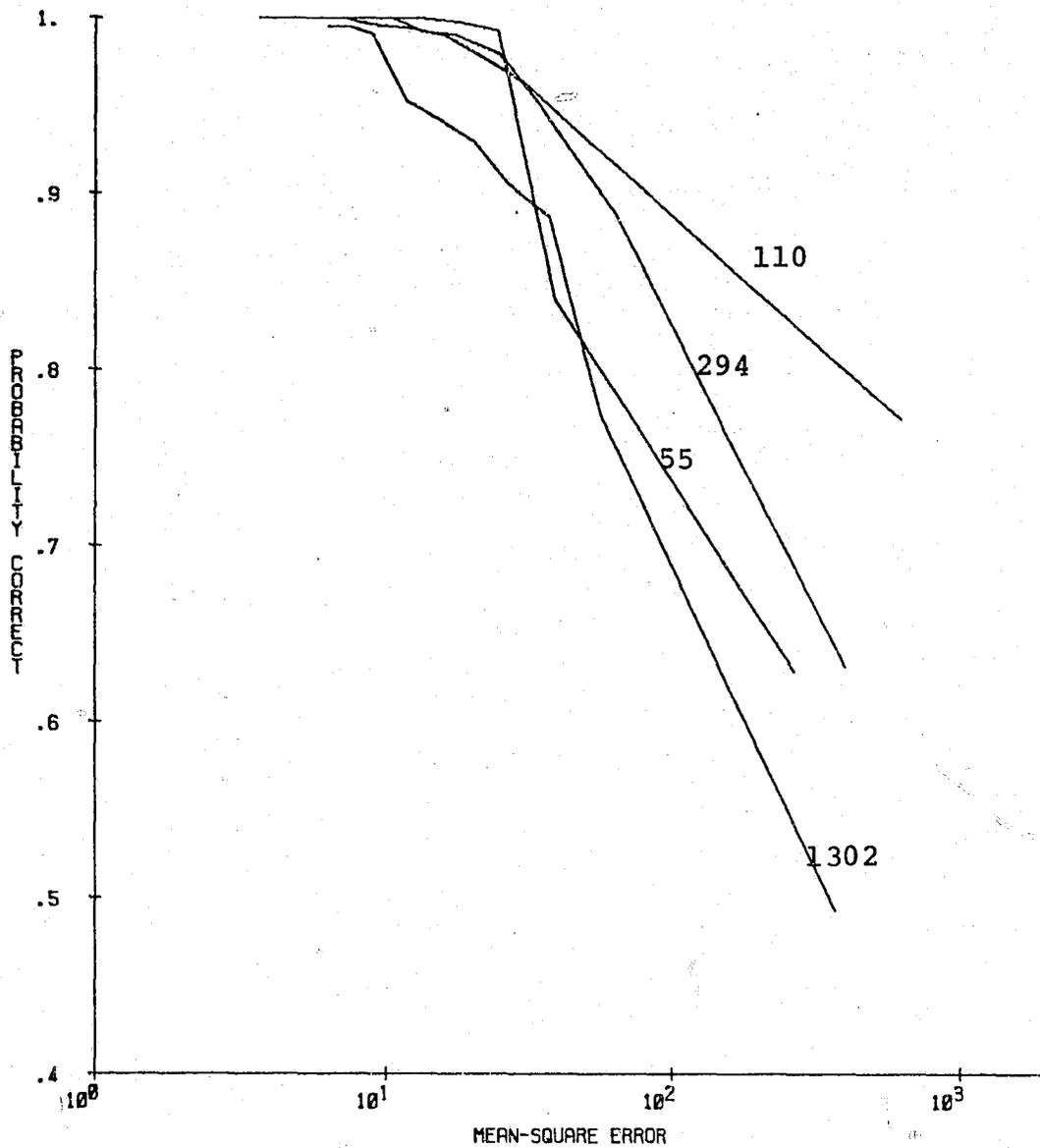


Figure 4.12 Influence of sample size on the estimate of classification performance for Williams County, August 4, 1977, using weight function number 2.

if there are a larger number of classes, a larger number of samples will be needed to obtain good classification performance estimates using high dimensionality.

4.4.5 Results

The analytical procedure for spectral parameter design of sensor systems was performed using the data collected on three dates over each of two locations. Results from using the experimental system are presented graphically in Figures 4.13 through 4.30. The three collection dates for Williams County, North Dakota, are presented first followed by the three data sets from Finney County, Kansas. Weight function number two was used for all cases.

For each data set the expected mean-square error is plotted as a function of the number of terms used in the Karhunen-Loeve expansion. A logarithmic scale is used for the mean-square error because of the large range of values. The units for the mean-square error are relative to the units on the spectral response function which are in terms of percent reflectance. Since

$$E \left[\int [x(\lambda)]^2 w(\lambda) d\lambda \right] = \sum_{i=1}^{\infty} \gamma_i, \quad (4.17)$$

the units of error are relative to the expected mean-square value of the response functions in the ensemble.

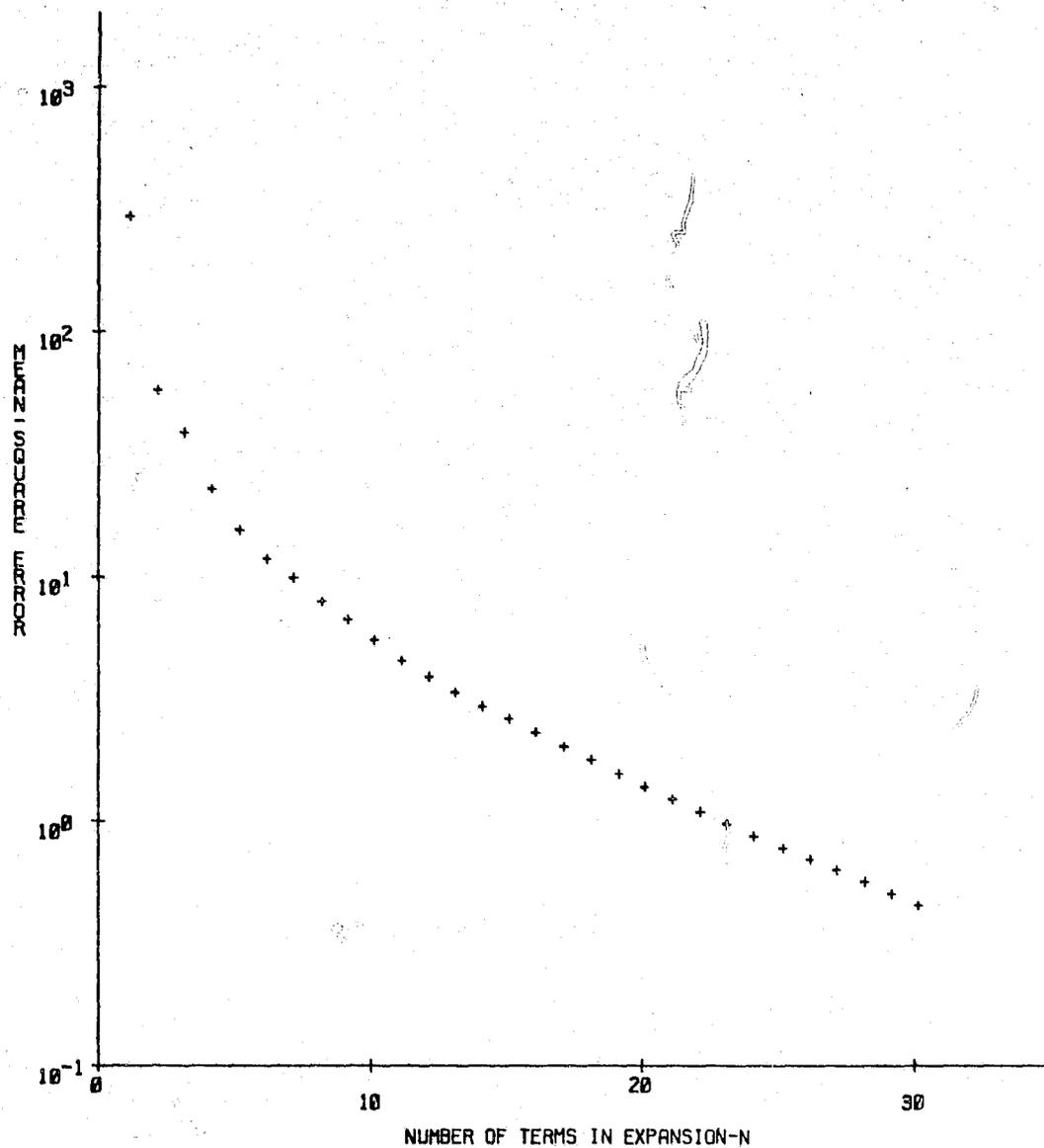
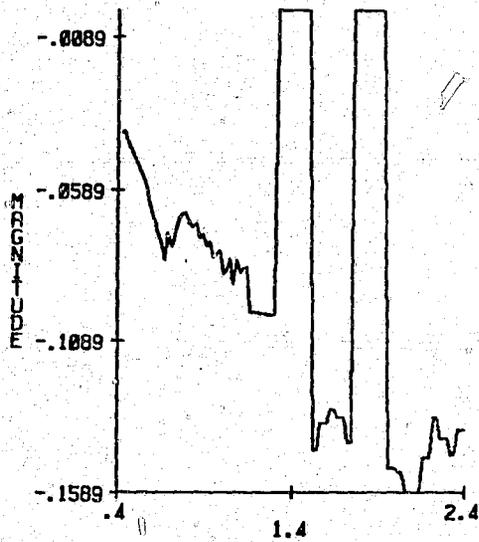
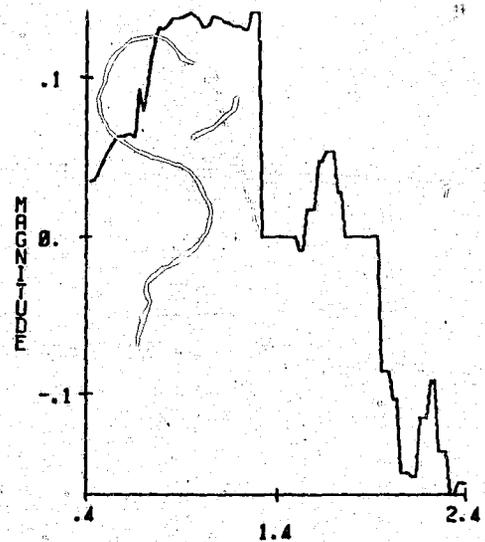


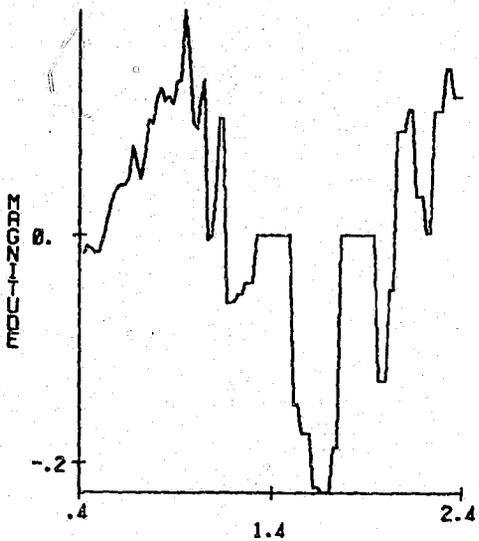
Figure 4.13 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Williams County, May 8, 1977, using weight function number 2.



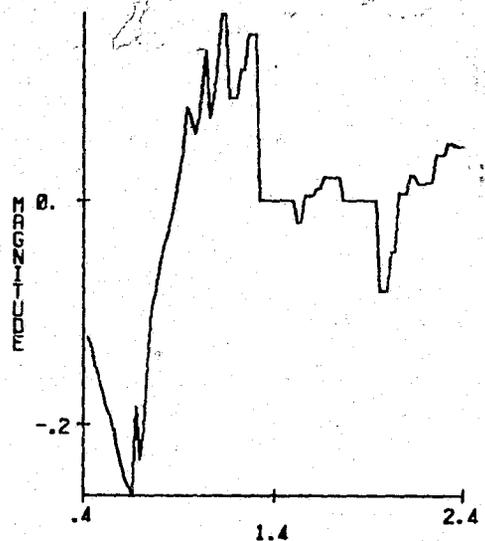
(5) WAVELENGTH (MICROMETERS)
EIGENVECTOR 1



(1) WAVELENGTH (MICROMETERS)
EIGENVECTOR 2



(3) WAVELENGTH (MICROMETERS)
EIGENVECTOR 3



(7) WAVELENGTH (MICROMETERS)
EIGENVECTOR 4

Figure 4.14 First twelve eigenvectors for Williams County, May 8, 1977, using weight function number 2.

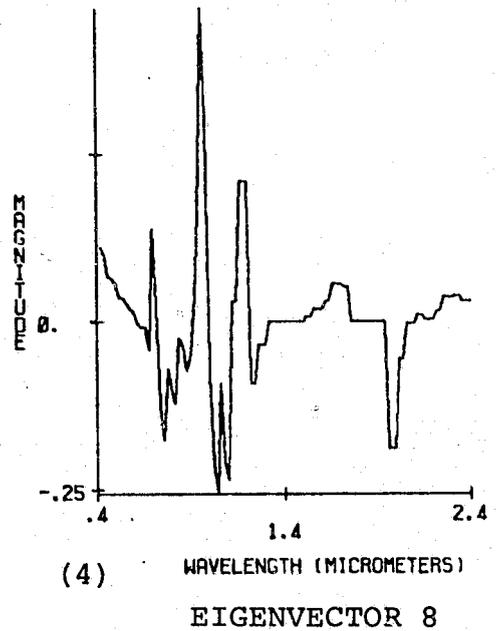
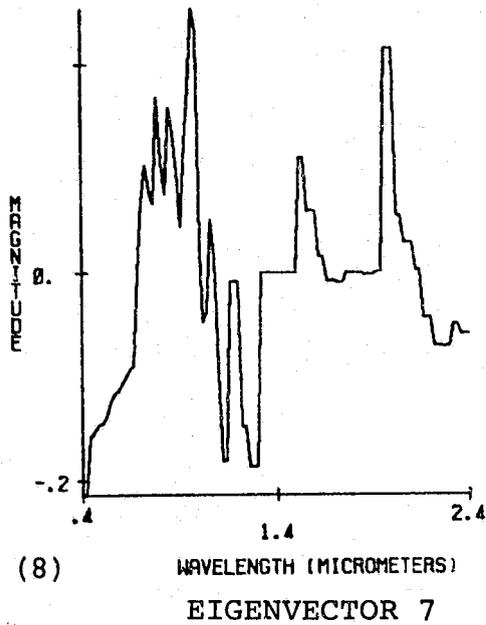
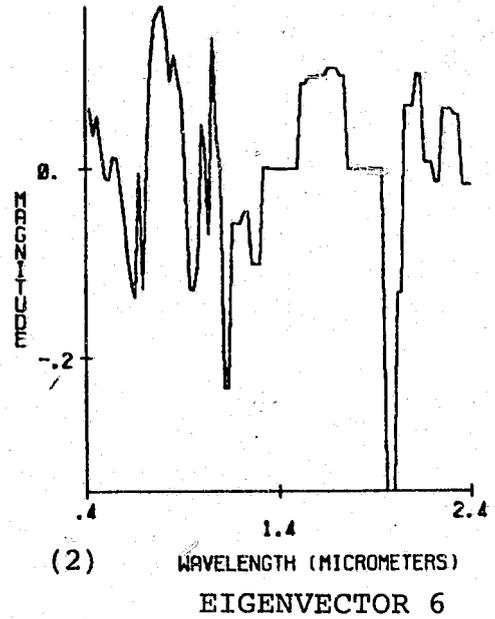
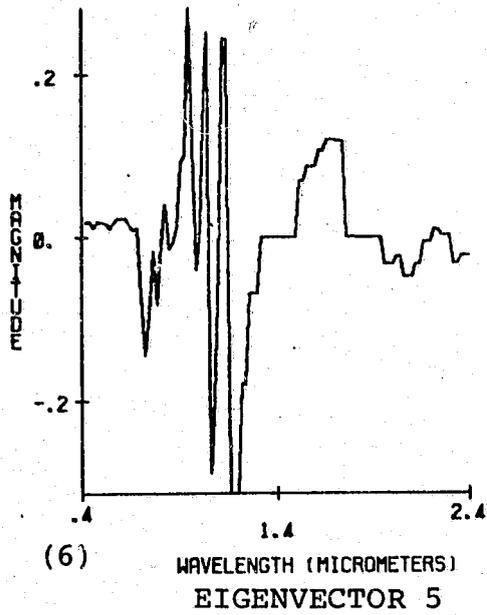
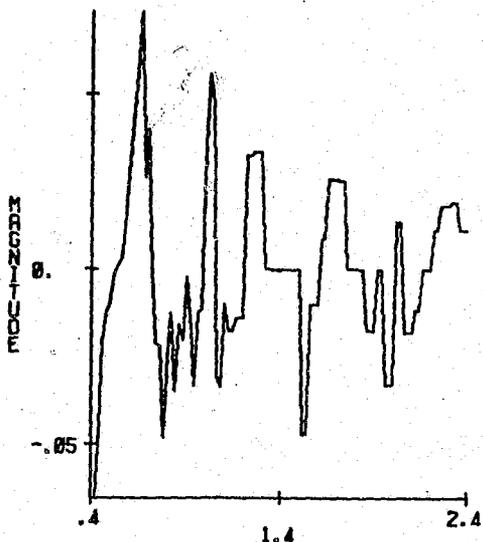
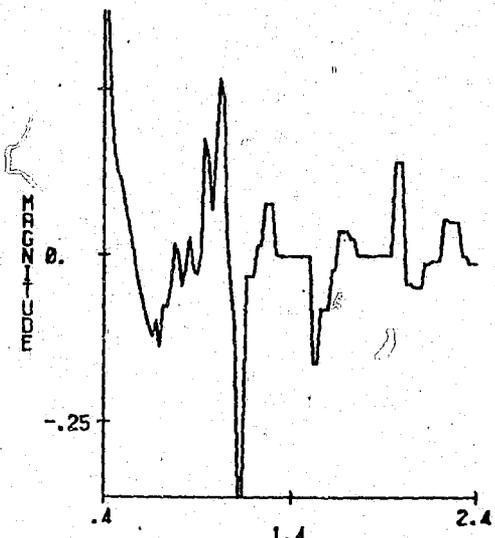


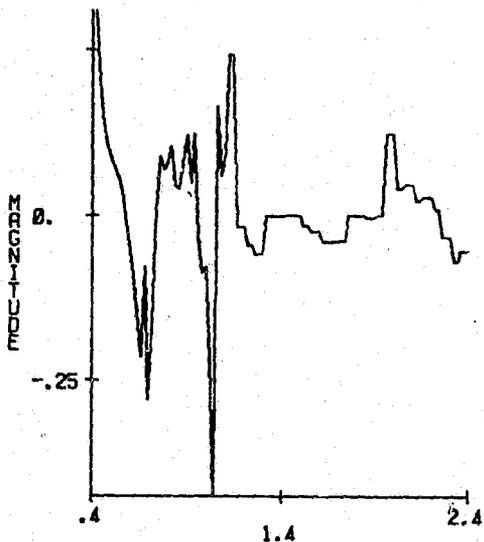
Figure 4.14 (cont.)



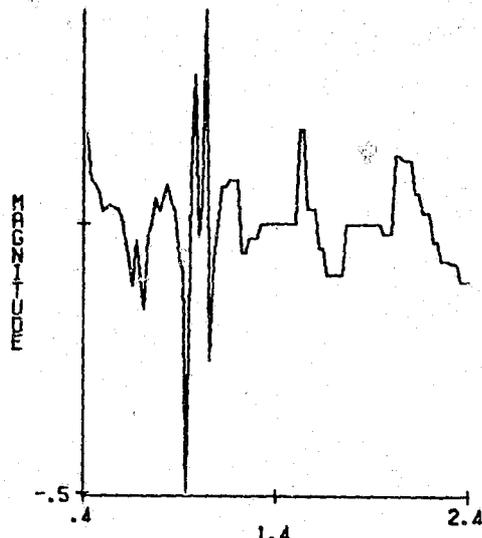
(10) WAVELENGTH (MICROMETERS)
EIGENVECTOR 9



(9) WAVELENGTH (MICROMETERS)
EIGENVECTOR 10



WAVELENGTH (MICROMETERS)
EIGENVECTOR 11



WAVELENGTH (MICROMETERS)
EIGENVECTOR 12

Figure 4.14 (cont.)

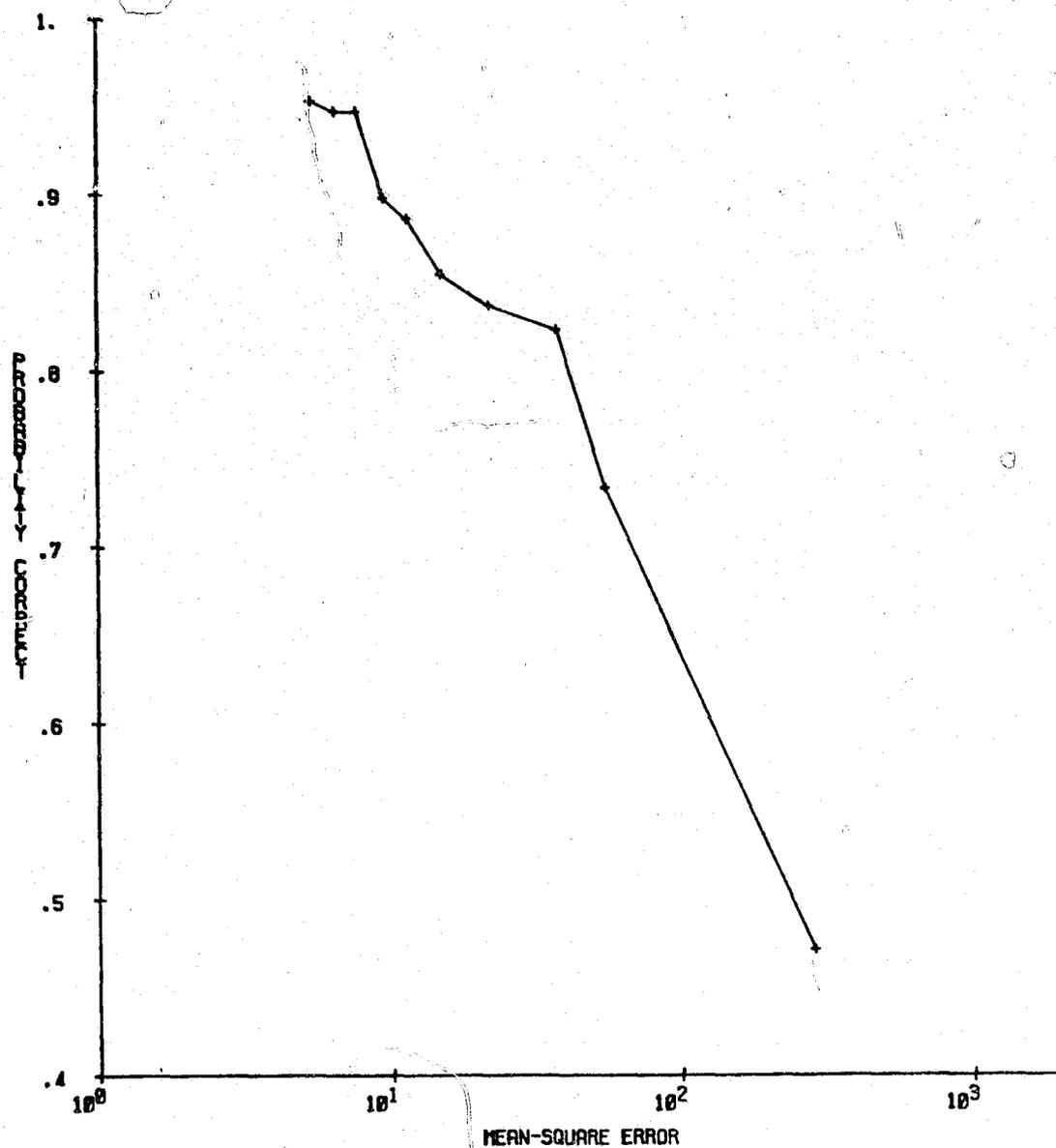


Figure 4.15 Estimate of probability of correct classification vs expected mean-square error for Williams County, May 8, 1977, using weight function number 2.

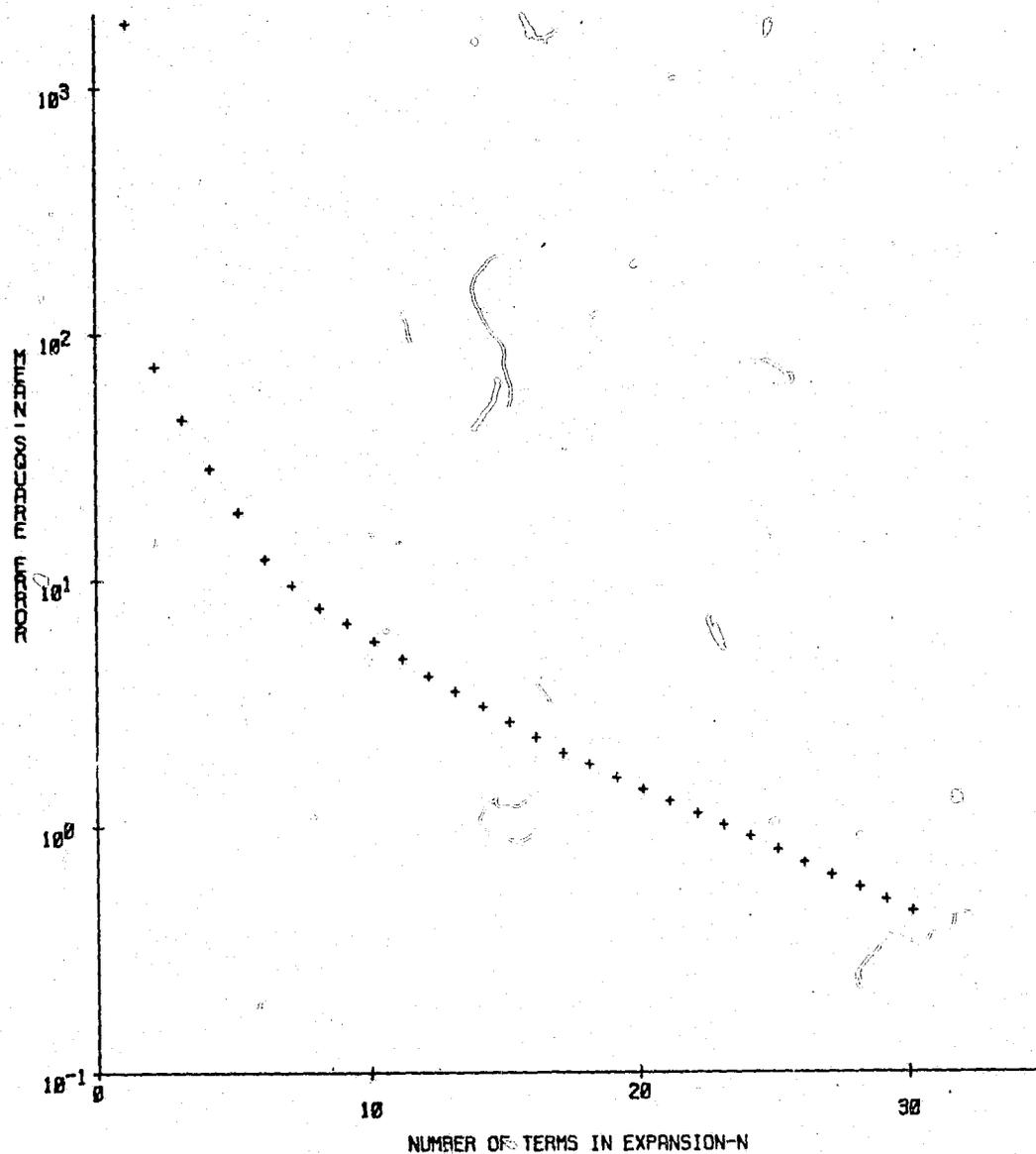
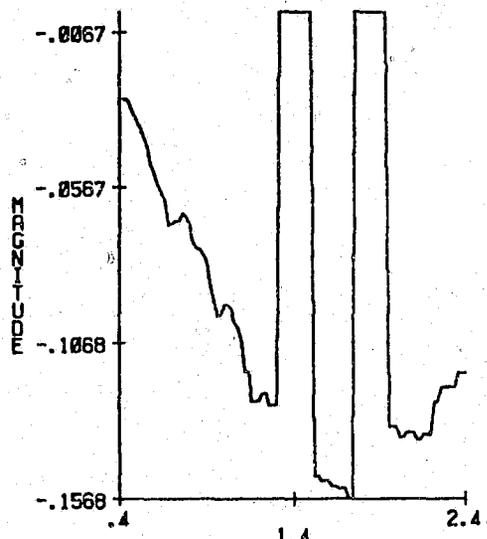
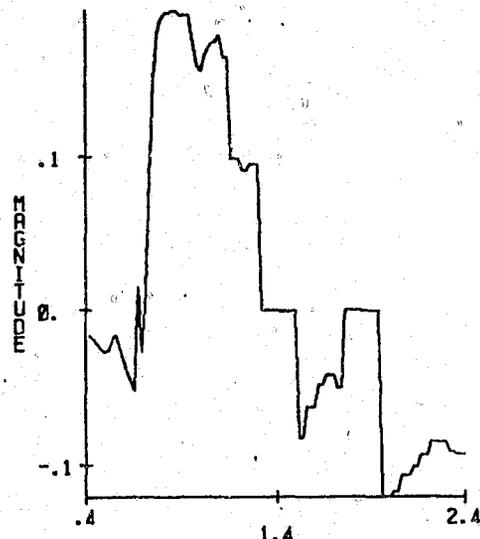


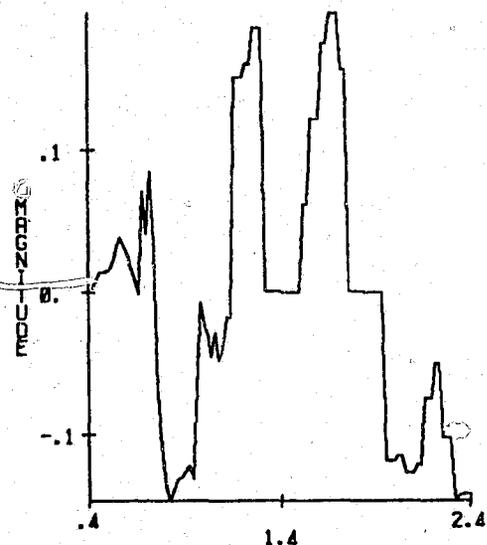
Figure 4.16 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Williams County, June 29, 1977, using weight function number 2.



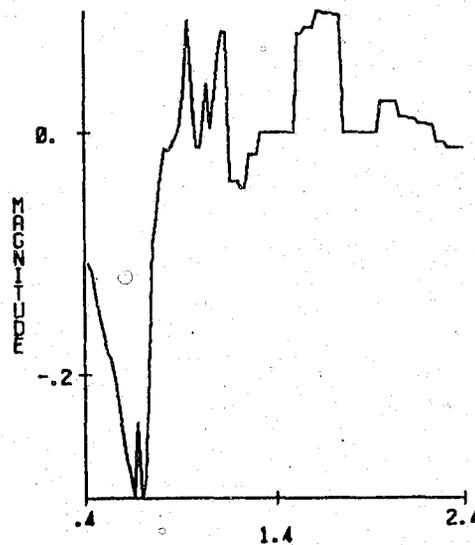
(3) WAVELENGTH (MICROMETERS)
EIGENVECTOR 1



(2) WAVELENGTH (MICROMETERS)
EIGENVECTOR 2

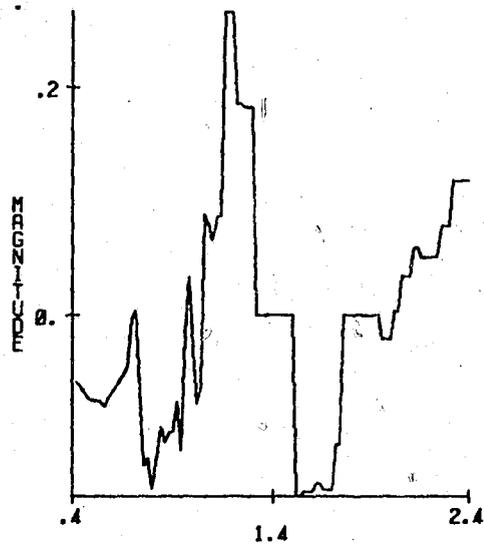


(1) WAVELENGTH (MICROMETERS)
EIGENVECTOR 3

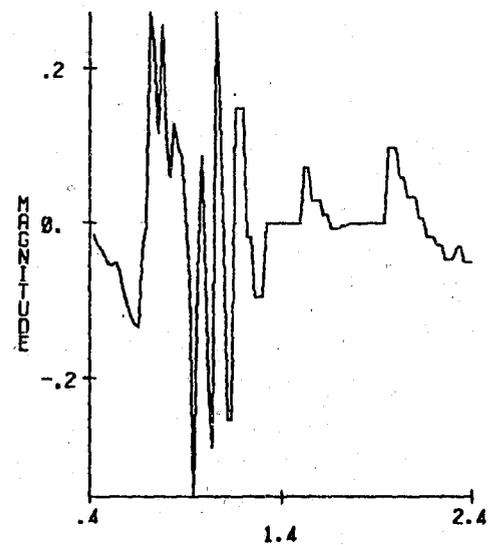


(4) WAVELENGTH (MICROMETERS)
EIGENVECTOR 4

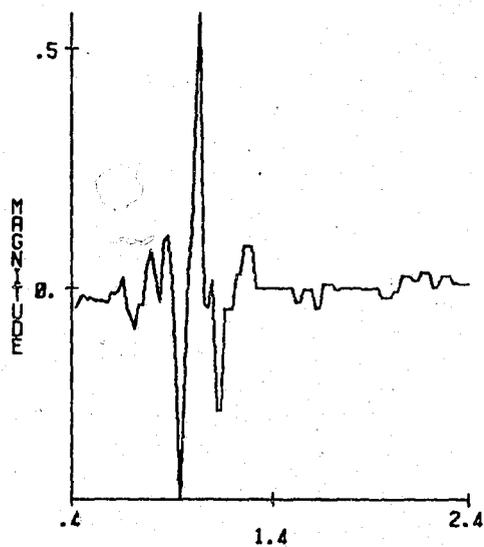
Figure 4.17 First twelve eigenvectors for Williams County, June 29, 1977, using weight function number 2.



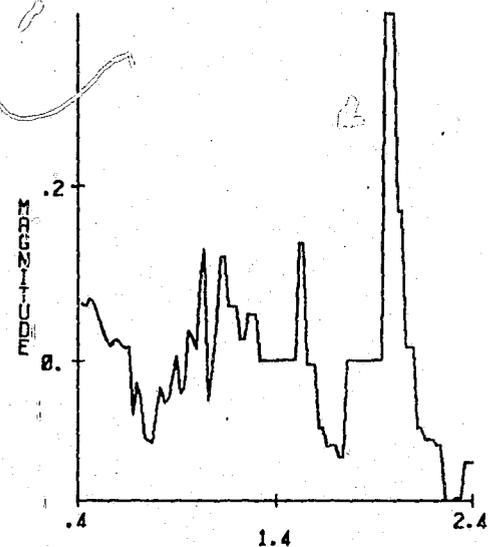
(5) WAVELENGTH (MICROMETERS)
EIGENVECTOR 5



(8) WAVELENGTH (MICROMETERS)
EIGENVECTOR 6



(7) WAVELENGTH (MICROMETERS)
EIGENVECTOR 7



(6) WAVELENGTH (MICROMETERS)
EIGENVECTOR 8

Figure 4.17 (cont.)

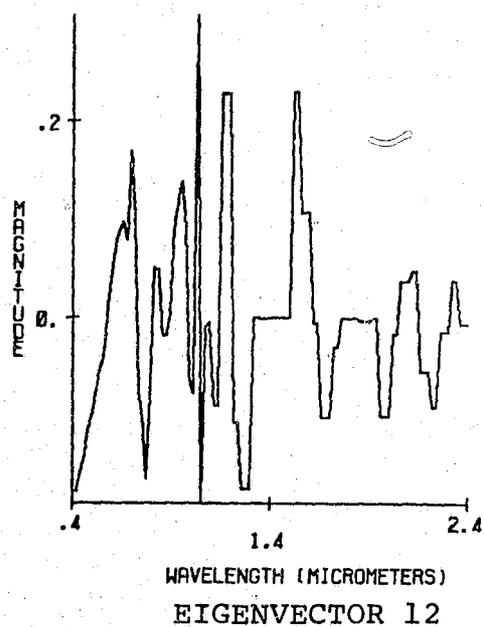
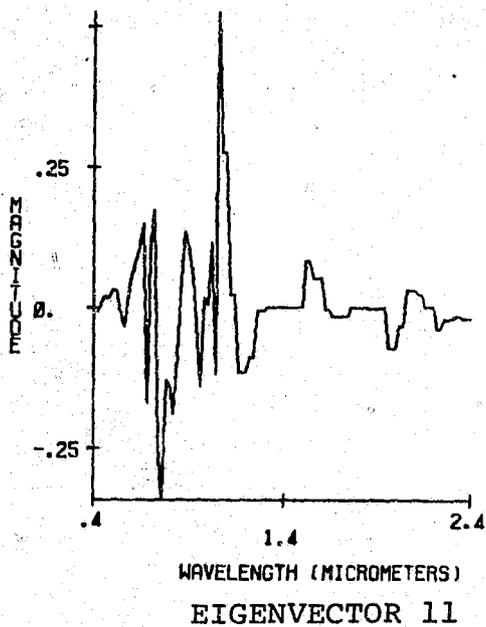
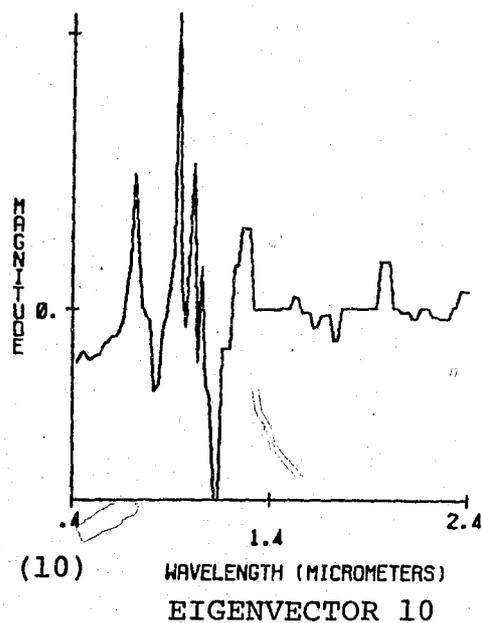
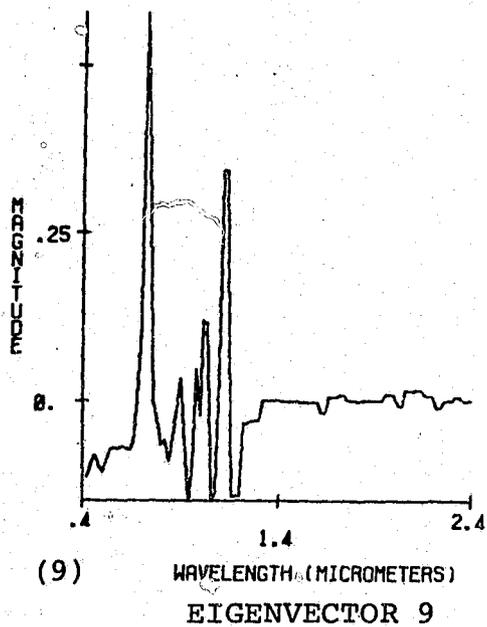


Figure 4.17 (cont.)

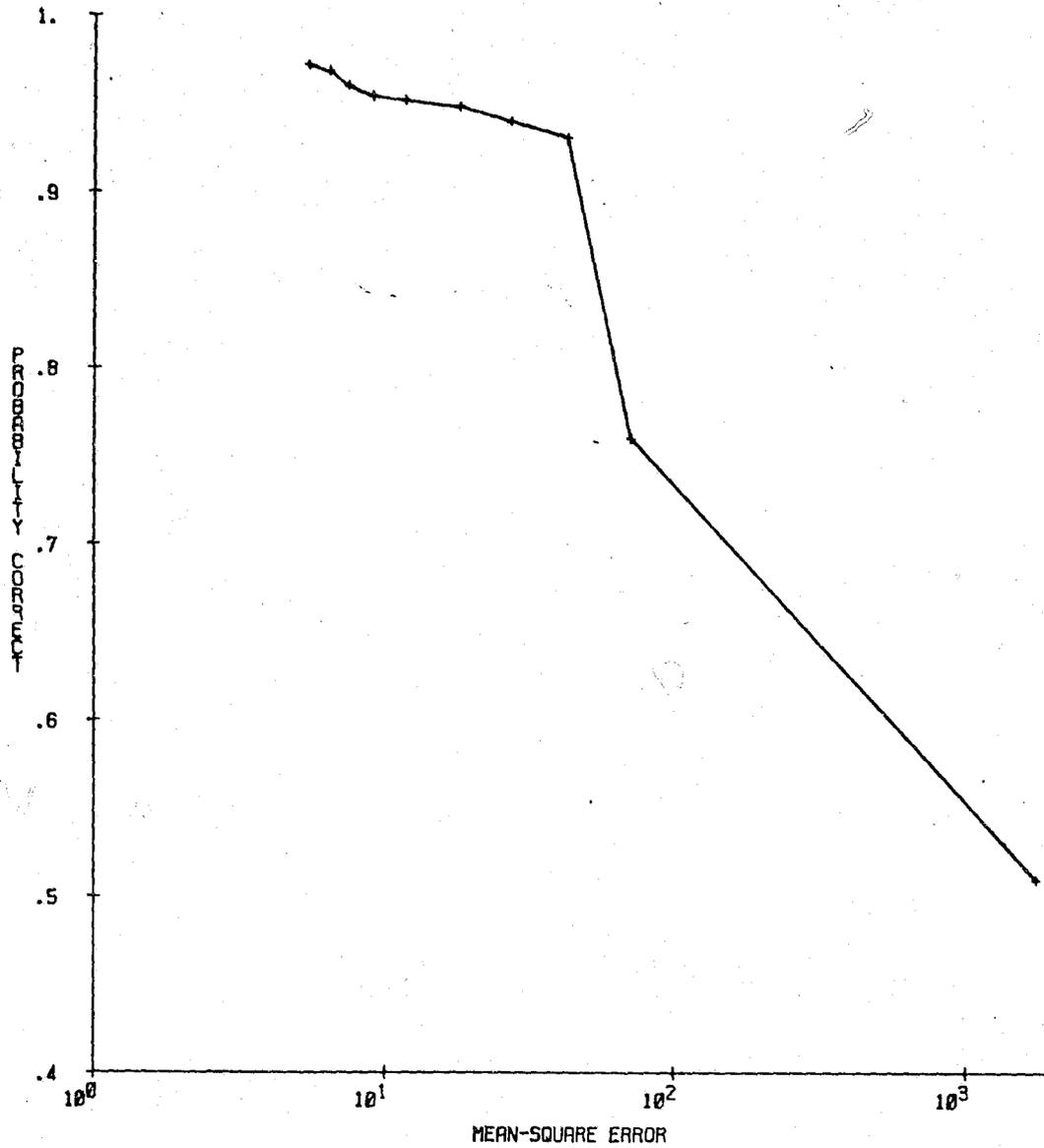


Figure 4.18 Estimate of probability of correct classification vs expected mean-square error for Williams County, June 29, 1977, using weight function number 2.

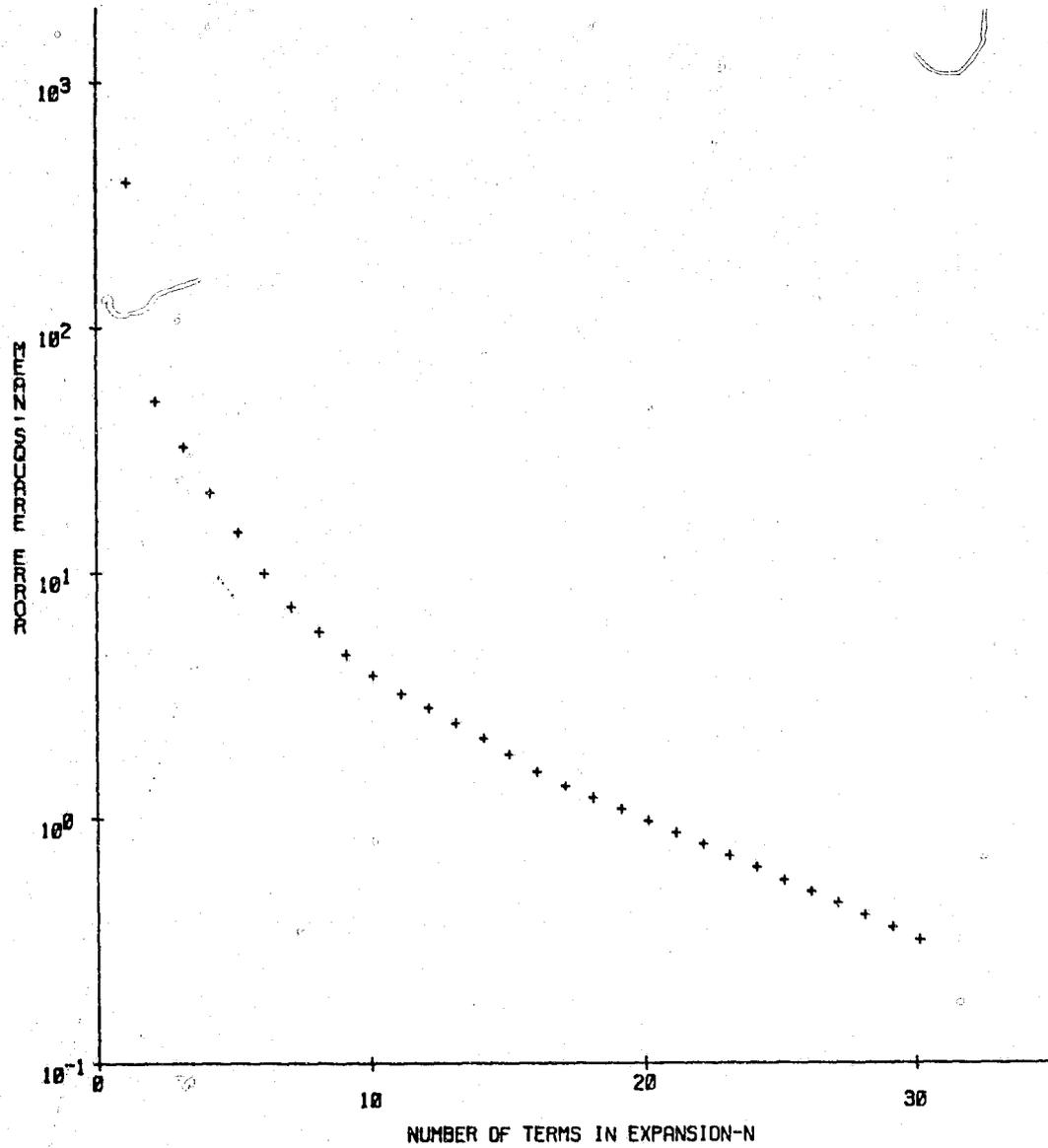


Figure 4.19 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Williams County, August 4, 1977, using weight function number 2.

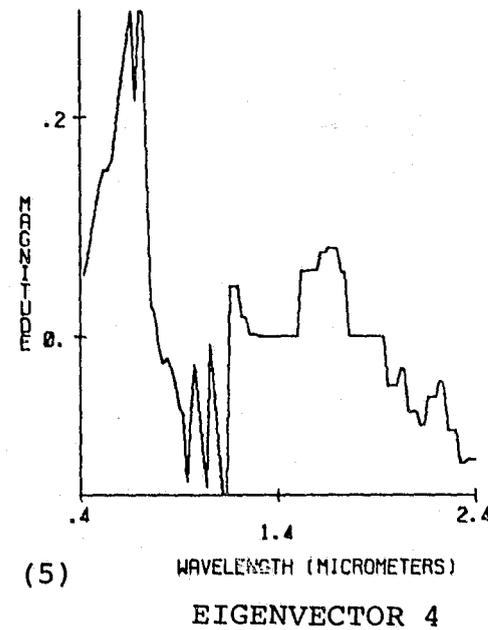
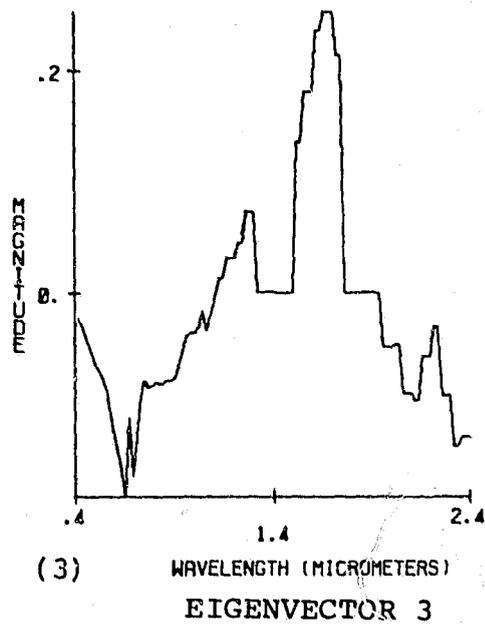
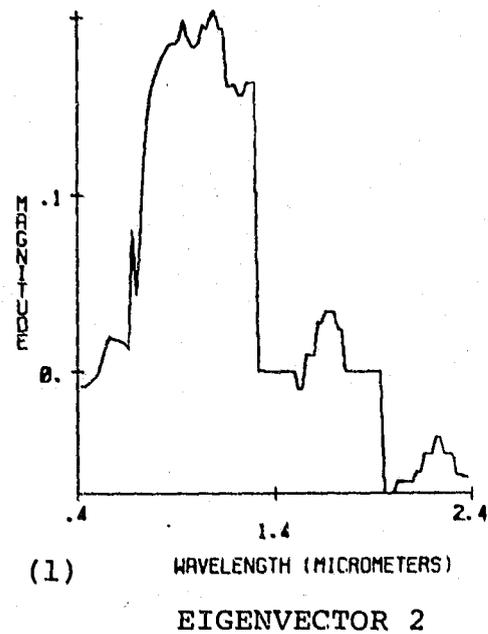
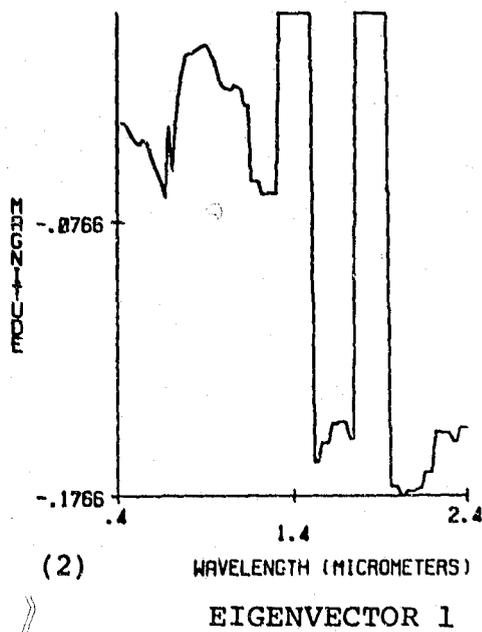
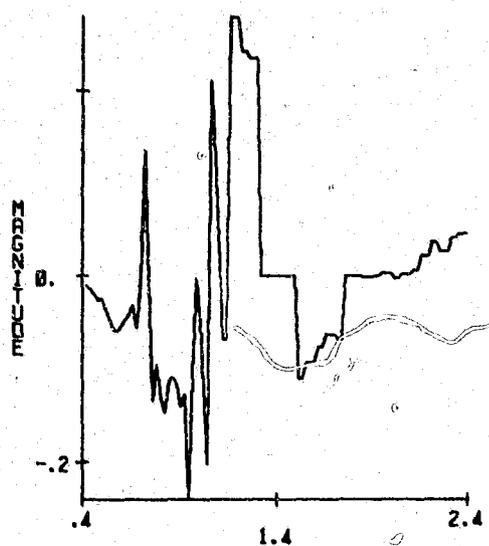
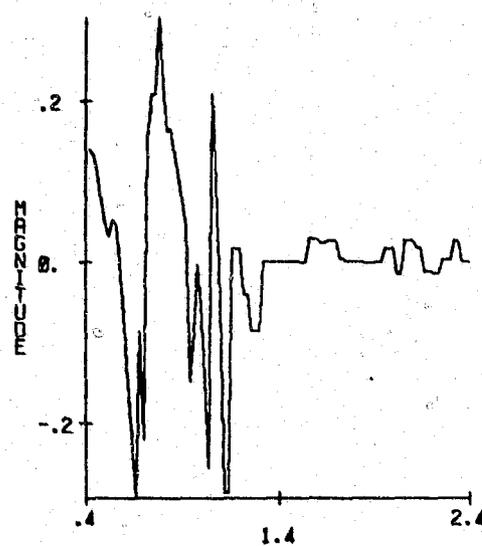


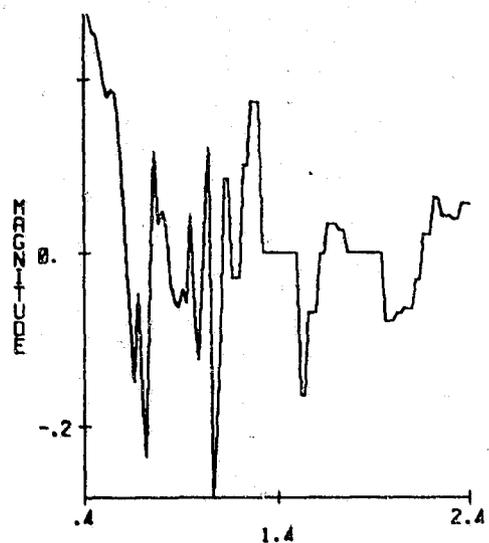
Figure 4.20 First twelve eigenvectors for Williams County, August 4, 1977, using weight function number 2.



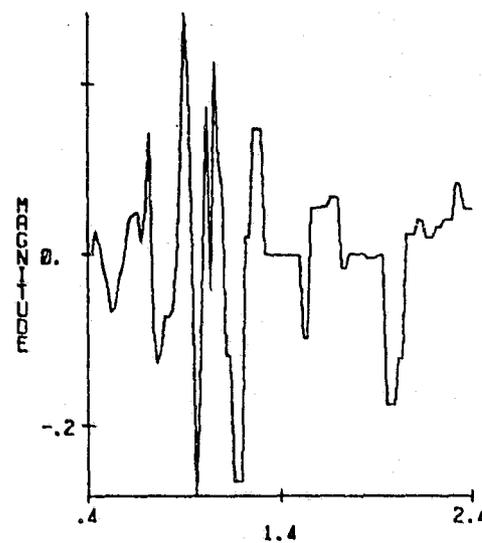
(8) WAVELENGTH (MICROMETERS)
EIGENVECTOR 5



(6) WAVELENGTH (MICROMETERS)
EIGENVECTOR 6



(4) WAVELENGTH (MICROMETERS)
EIGENVECTOR 7



(7) WAVELENGTH (MICROMETERS)
EIGENVECTOR 8

Figure 4.20 (cont.)

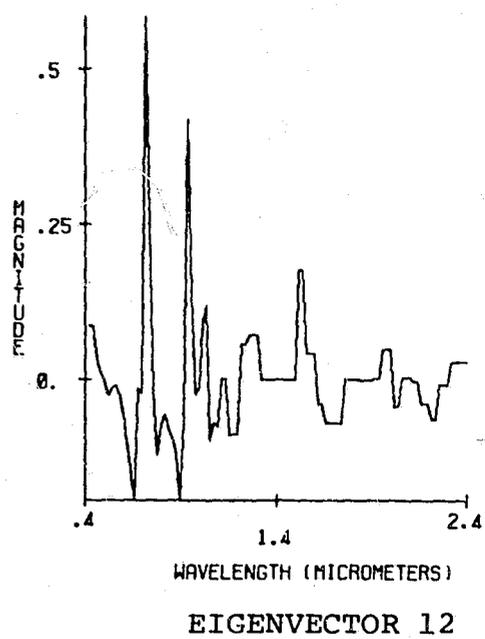
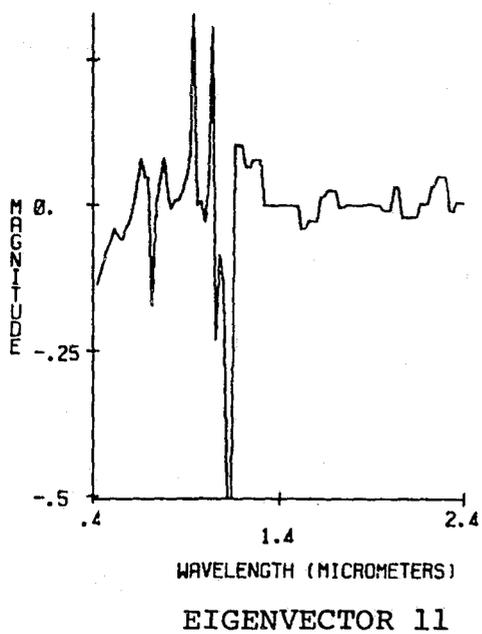
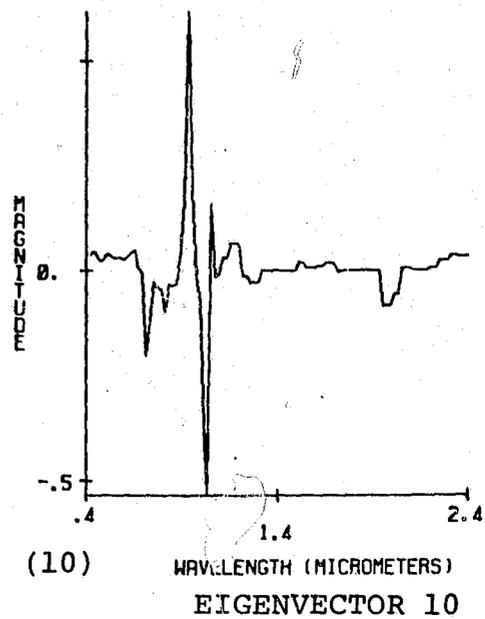
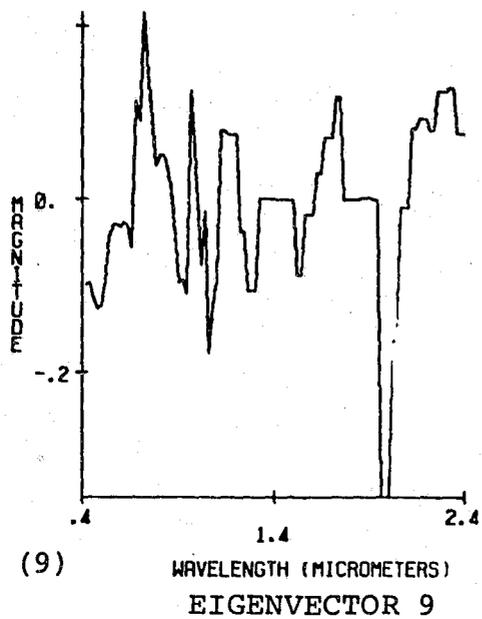


Figure 4.20 (cont.)

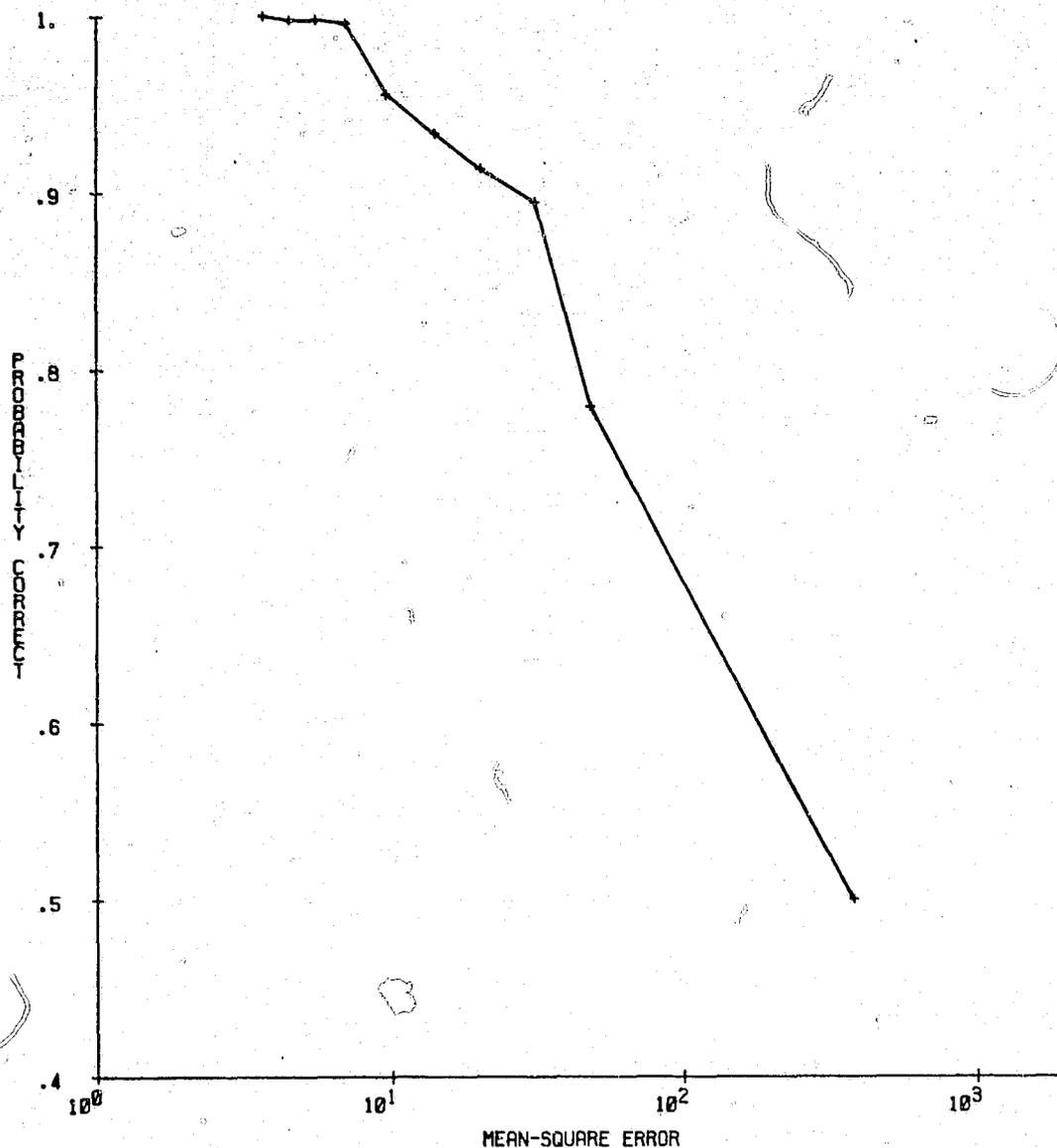


Figure 4.21 Estimate of probability of correct classification vs expected mean-square error for Williams County, August 4, 1977, using weight function number 2.

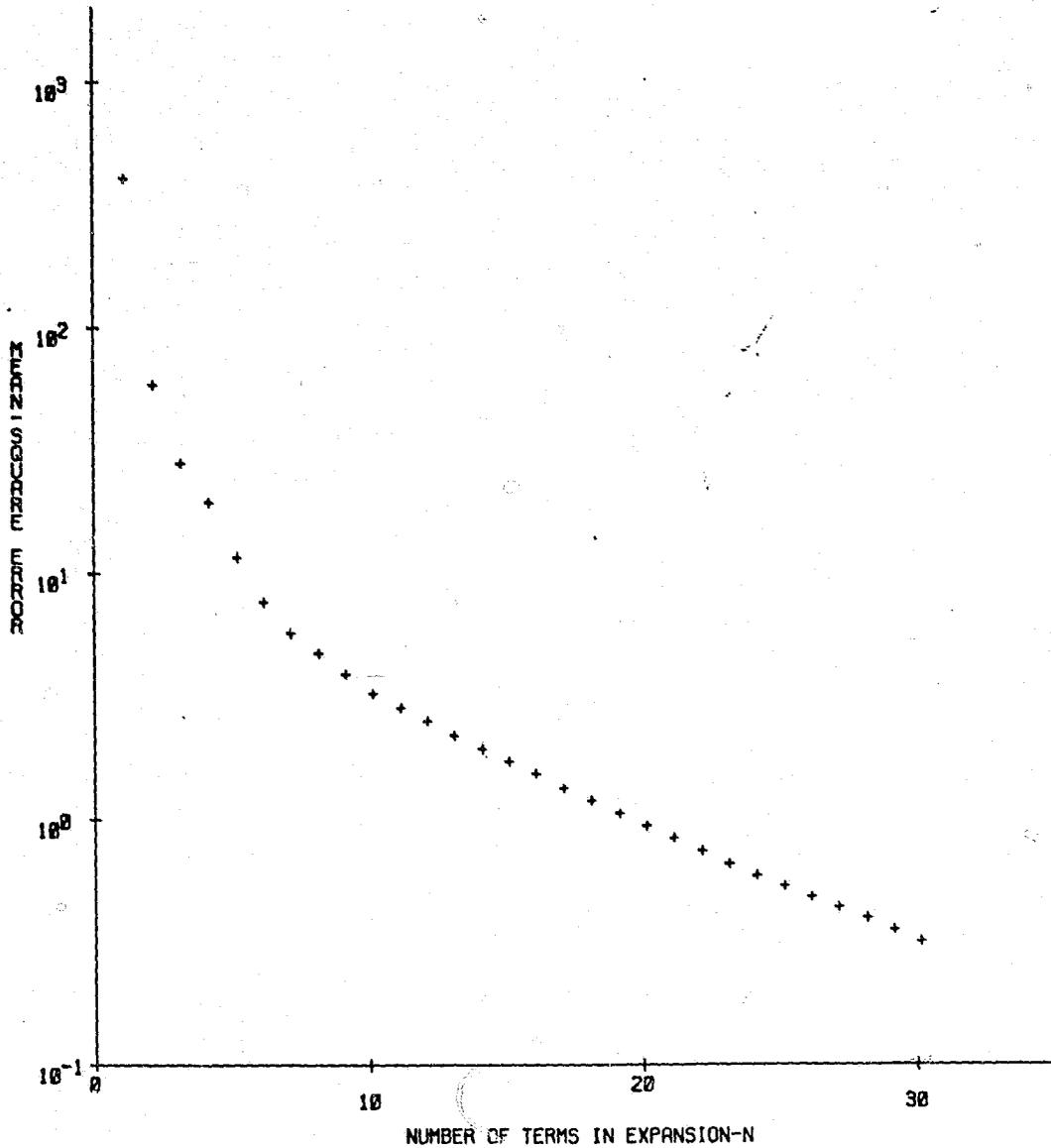
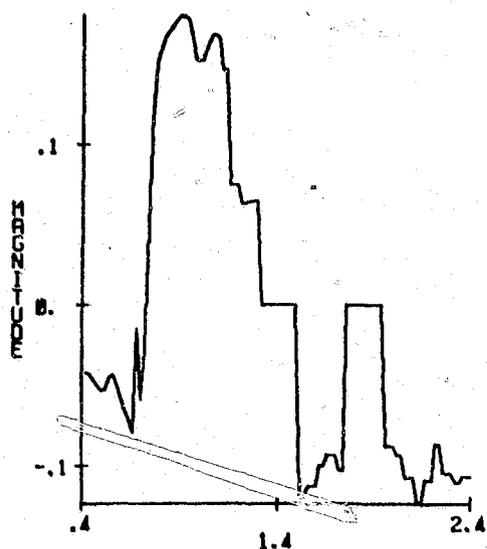
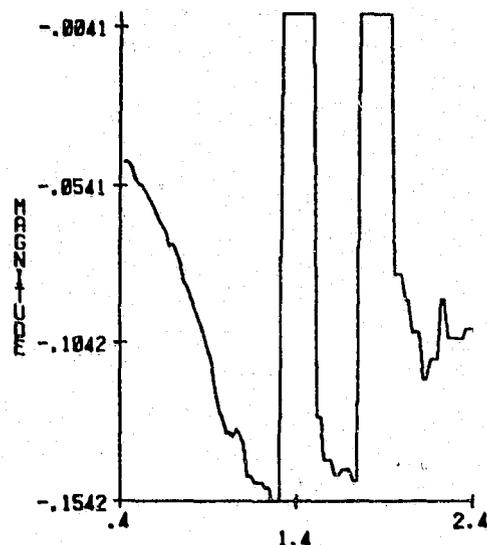


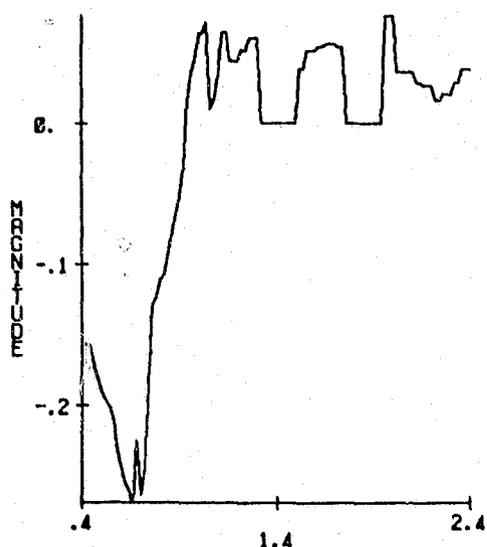
Figure 4.22 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Finney County, September 28, 1976, using weight function number 2.



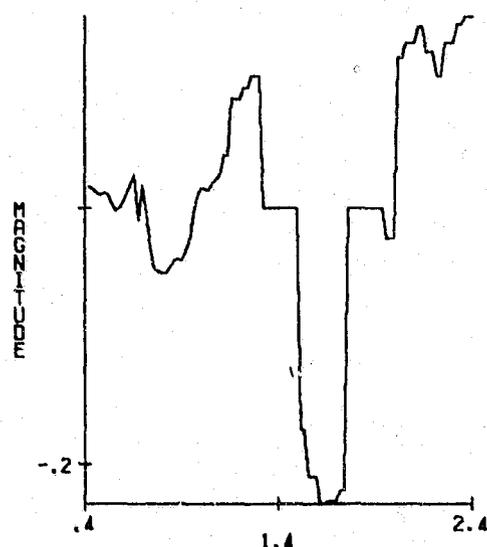
(1) WAVELENGTH (MICROMETERS)
EIGENVECTOR 1



(2) WAVELENGTH (MICROMETERS)
EIGENVECTOR 2

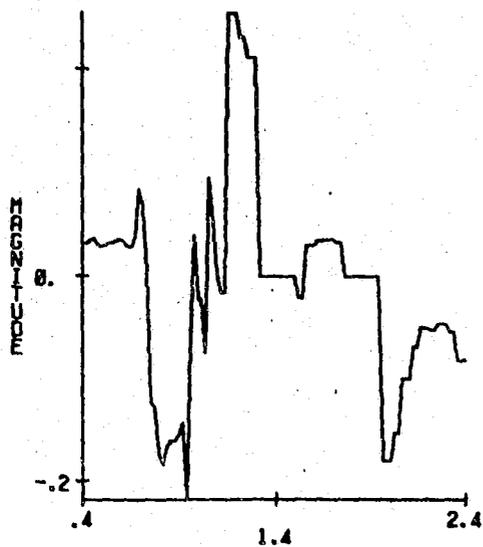


(4) WAVELENGTH (MICROMETERS)
EIGENVECTOR 3

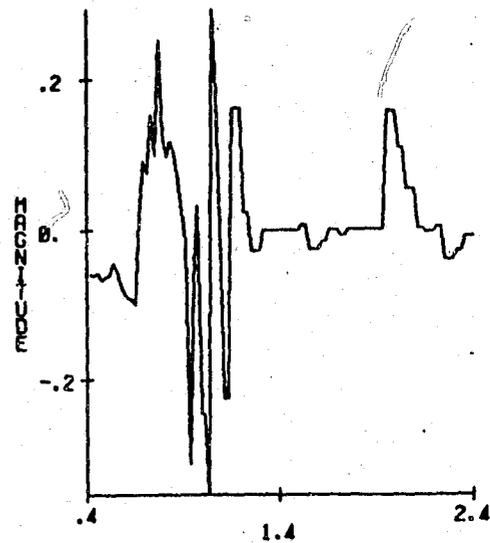


(3) WAVELENGTH (MICROMETERS)
EIGENVECTOR 4

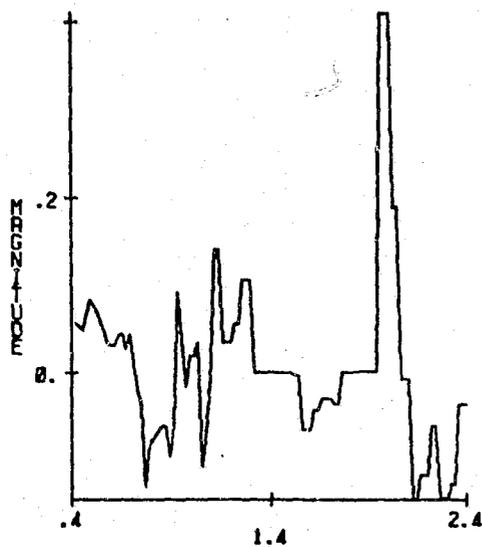
Figure 4.23 First twelve eigenvectors for Finney County, September 28, 1976, using weight function number 2.



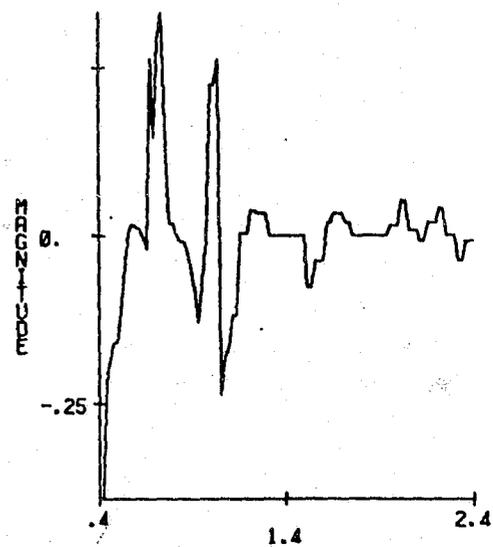
(6) WAVELENGTH (MICROMETERS)
EIGENVECTOR 5



(7) WAVELENGTH (MICROMETERS)
EIGENVECTOR 6



(5) WAVELENGTH (MICROMETERS)
EIGENVECTOR 7



(9) WAVELENGTH (MICROMETERS)
EIGENVECTOR 8

Figure 4.23 (cont.)

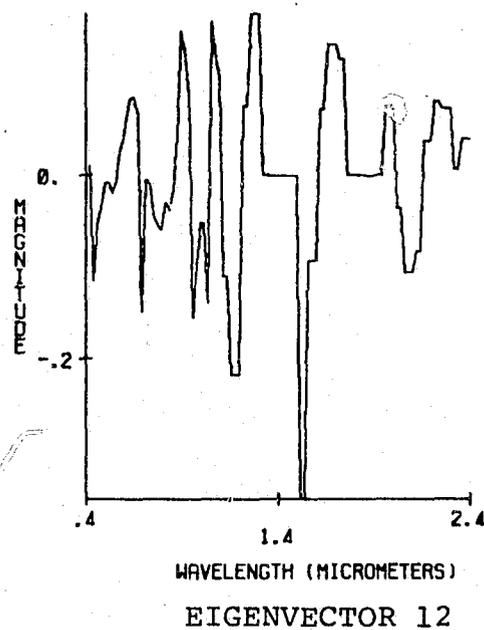
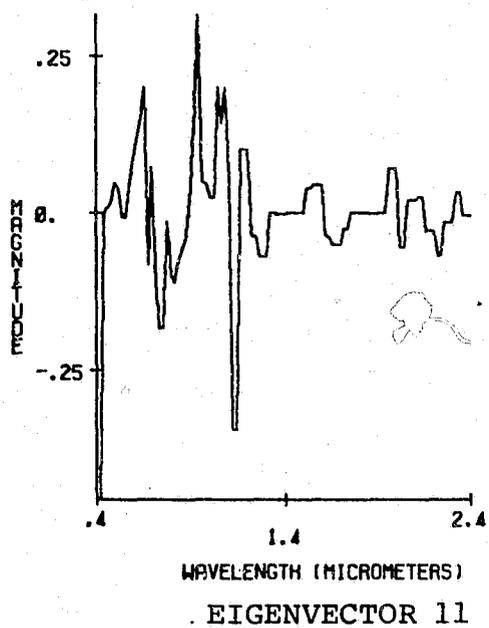
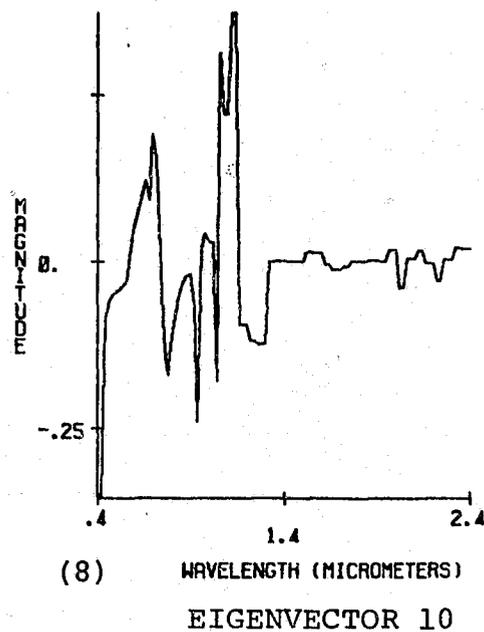
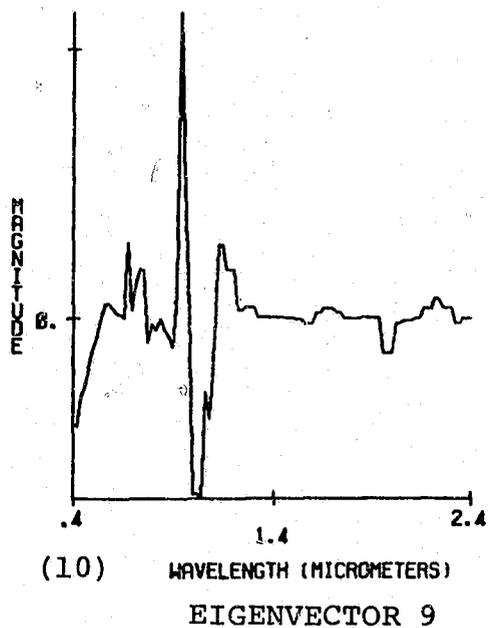


Figure 4.23 (cont.)

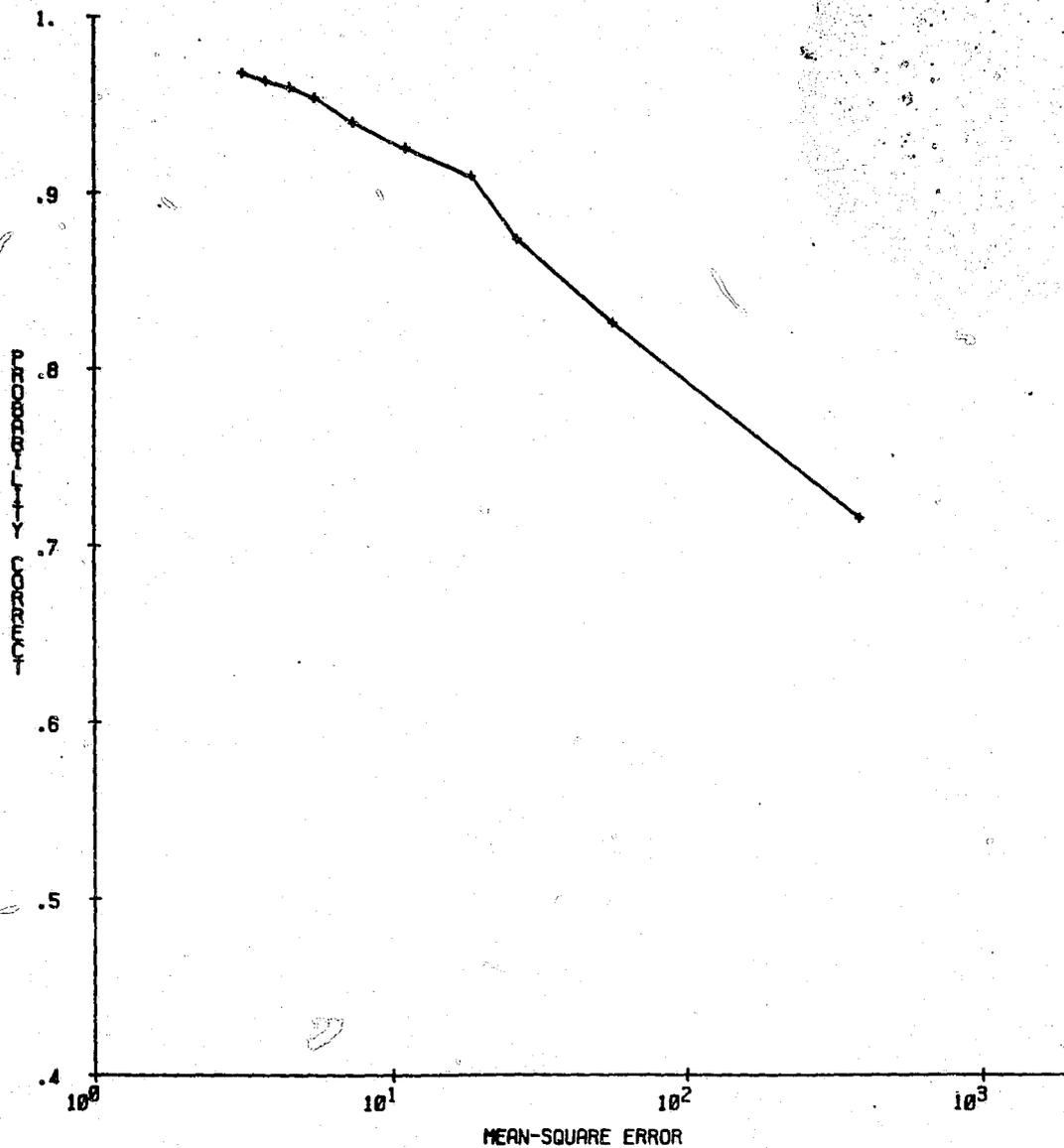


Figure 4.24 Estimate of probability of correct classification vs expected mean-square error for Finney County, September 28, 1976, using weight function number 2.

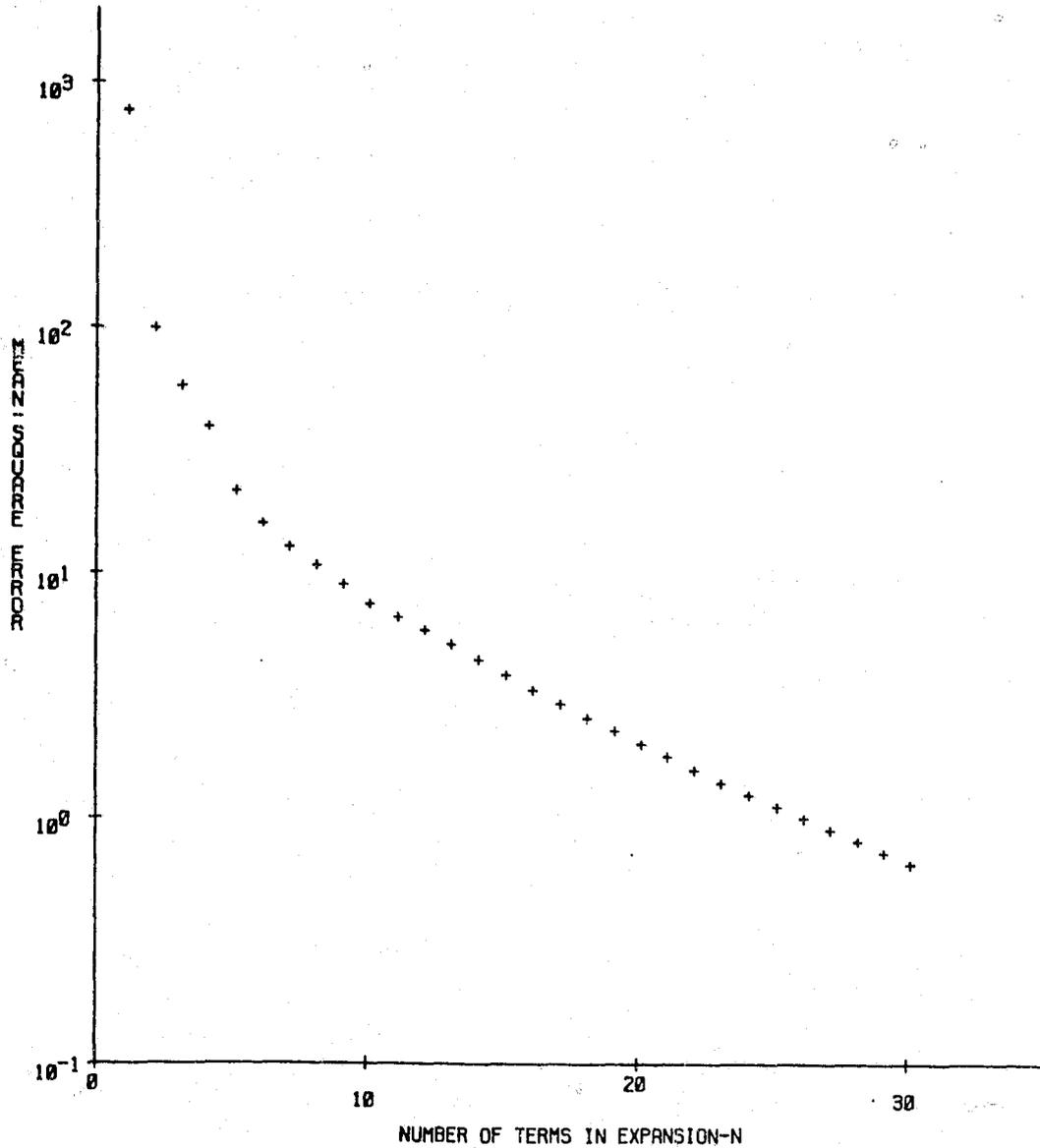


Figure 4.25 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Finney County, May 3, 1977, using weight function number 2.

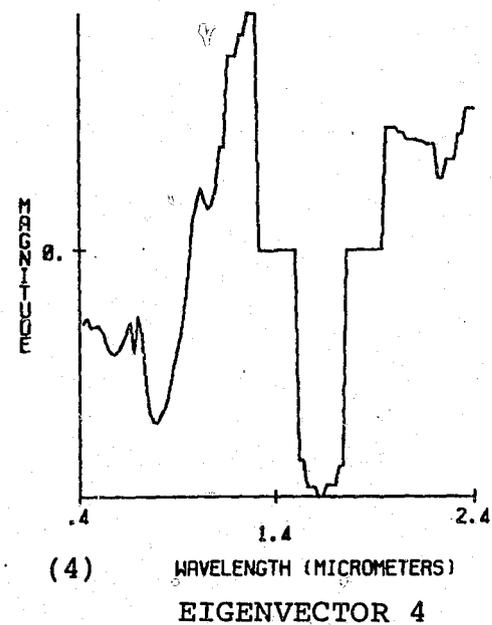
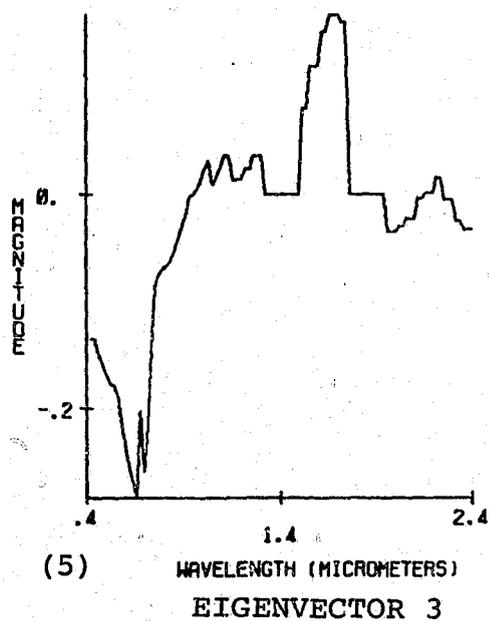
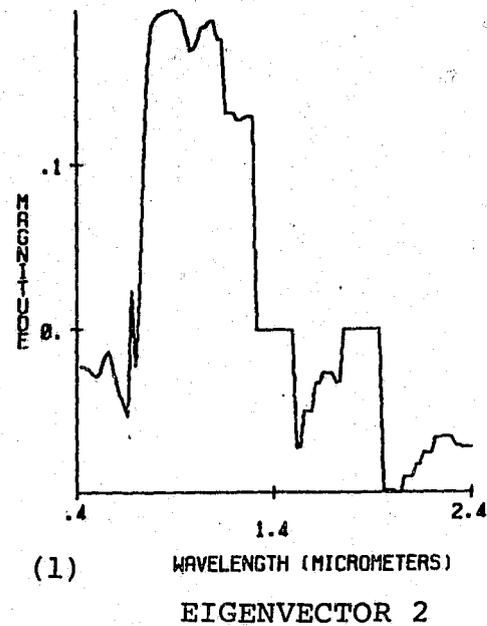
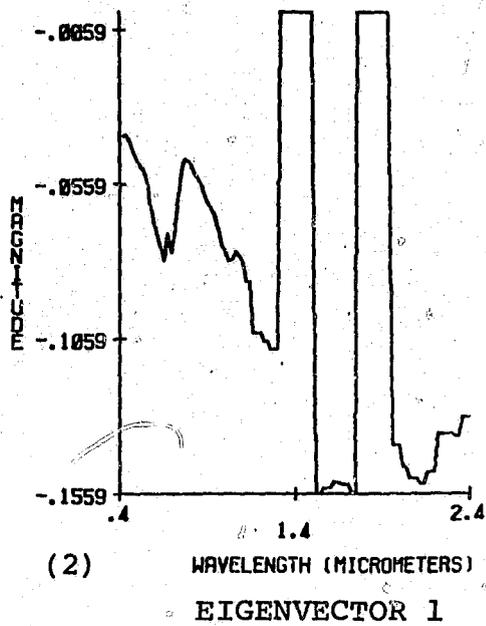
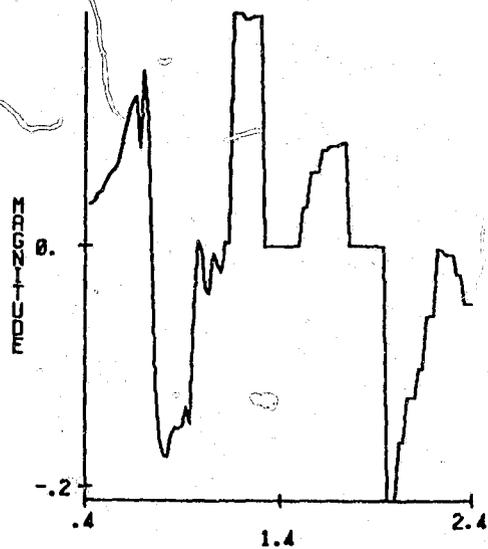
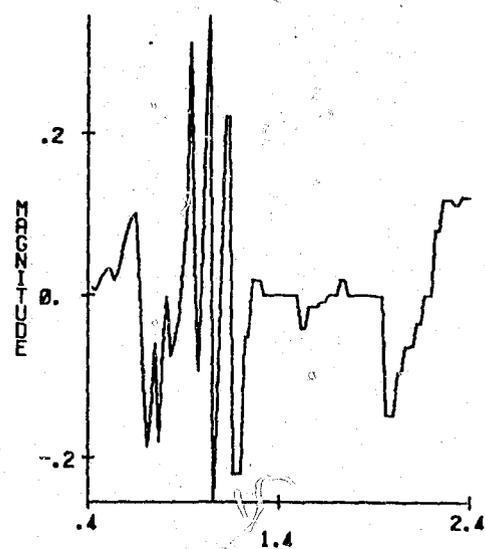


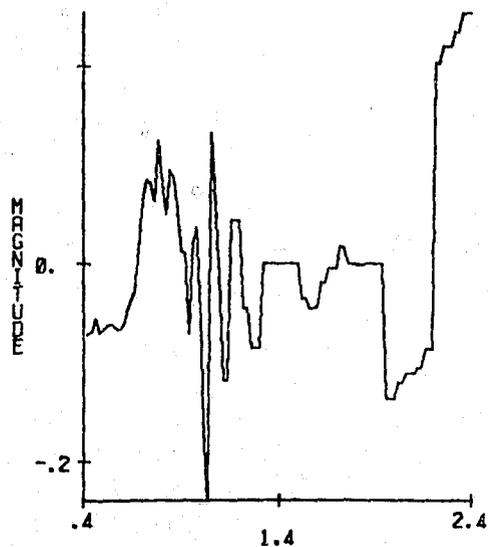
Figure 4.26 First twelve eigenvectors for Finney County, May 3, 1977, using weight function number 2.



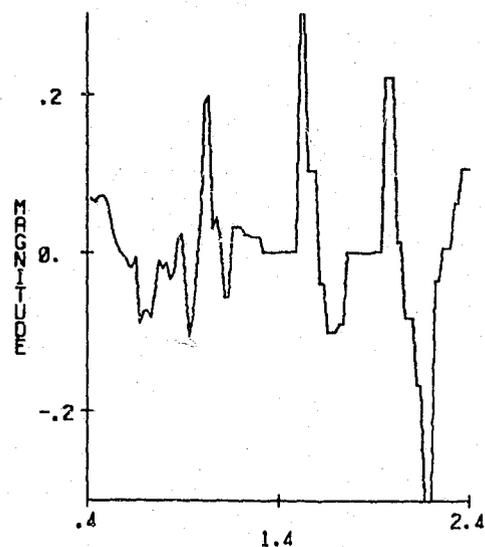
(3) WAVELENGTH (MICROMETERS)
EIGENVECTOR 5



(10) WAVELENGTH (MICROMETERS)
EIGENVECTOR 6

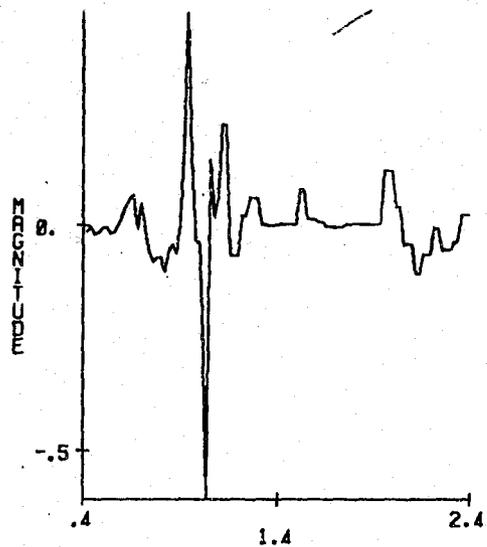


(8) WAVELENGTH (MICROMETERS)
EIGENVECTOR 7



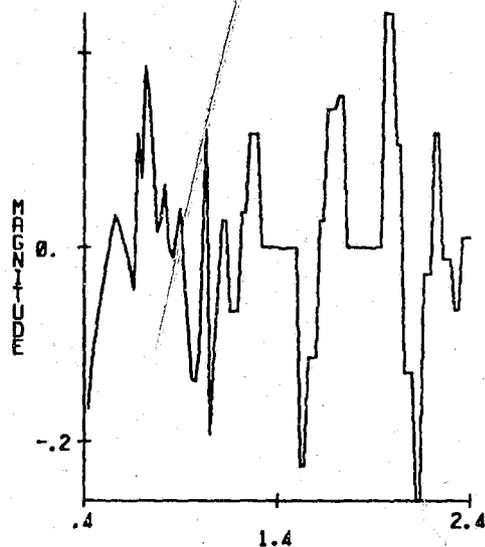
(7) WAVELENGTH (MICROMETERS)
EIGENVECTOR 8

Figure 4.26 (cont.)



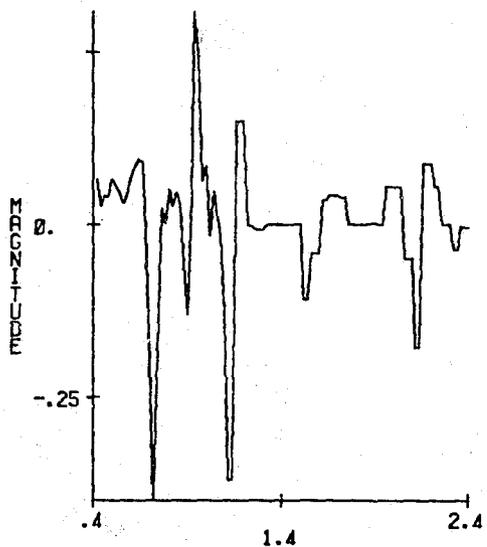
(9) WAVELENGTH (MICROMETERS)

EIGENVECTOR 9



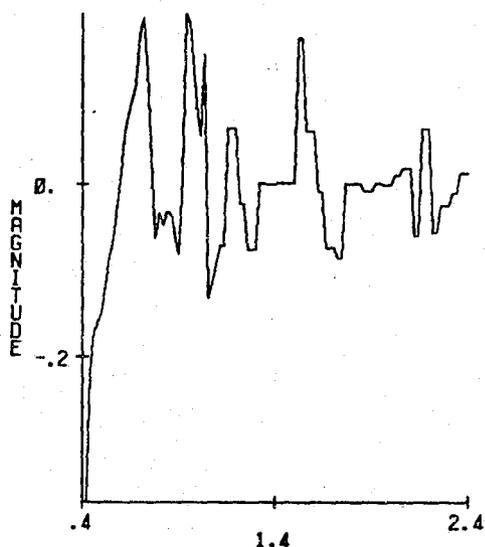
(6) WAVELENGTH (MICROMETERS)

EIGENVECTOR 10



WAVELENGTH (MICROMETERS)

EIGENVECTOR 11



WAVELENGTH (MICROMETERS)

EIGENVECTOR 12

Figure 4.26 (cont.)

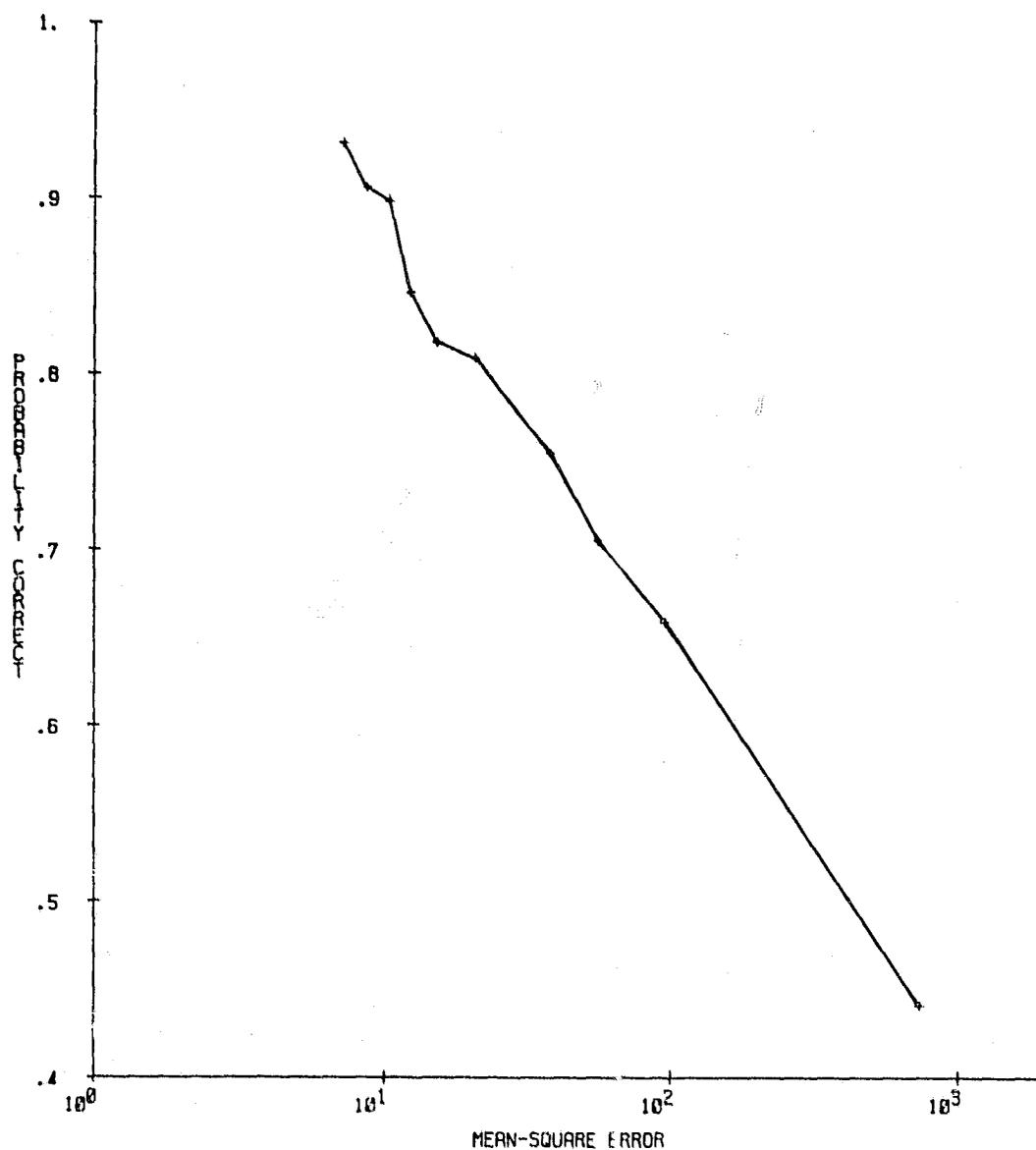


Figure 4.27 Estimate of probability of correct classification vs expected mean-square error for Finney County, May 3, 1977, using weight function number 2.

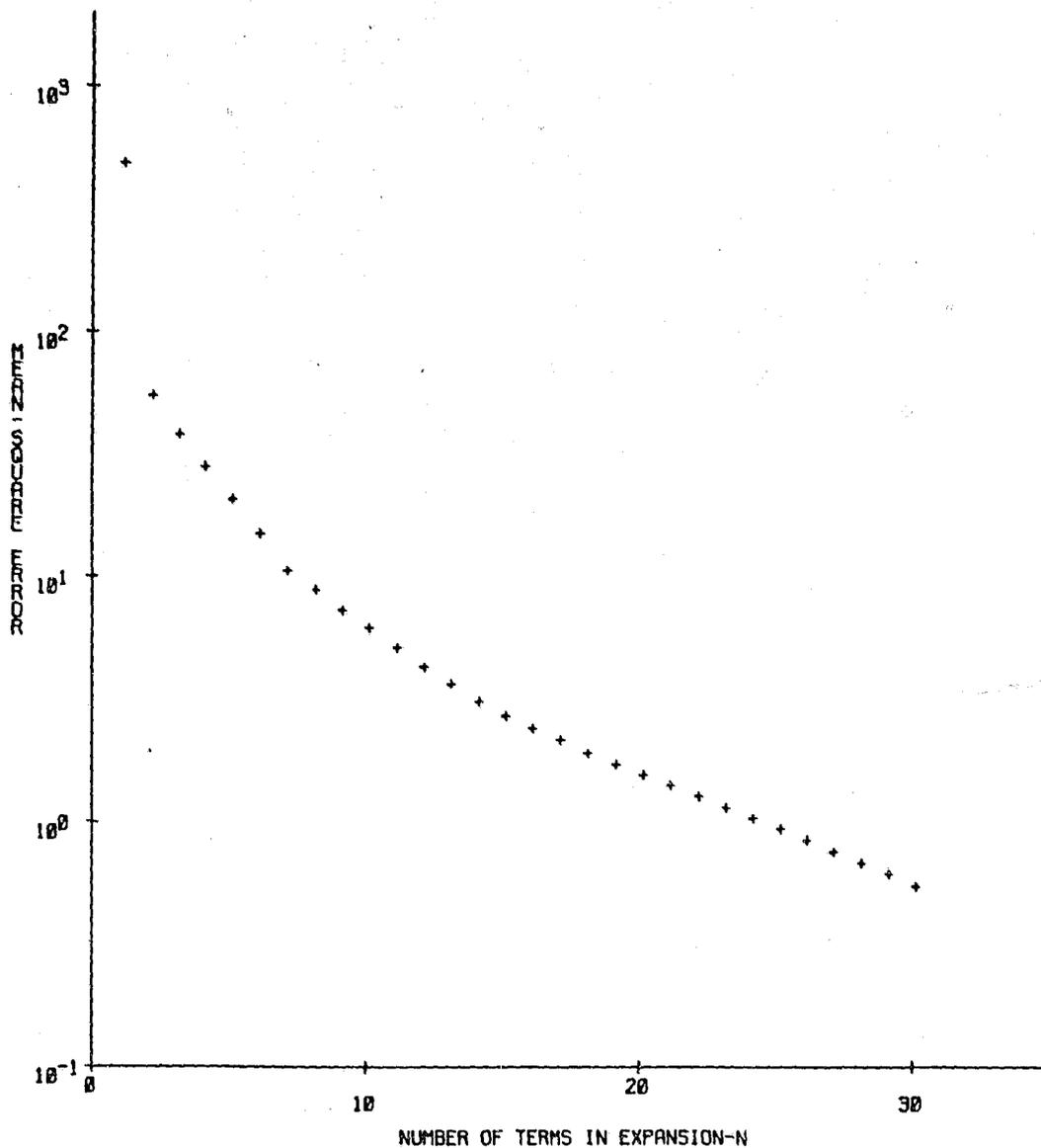
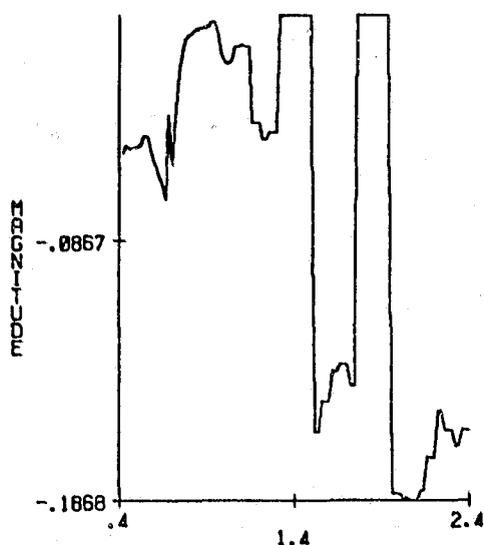
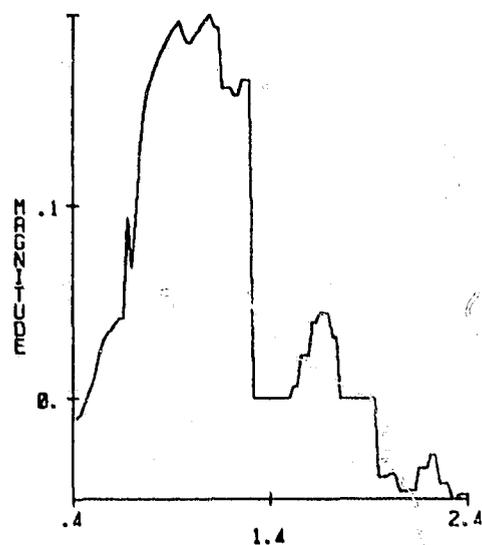


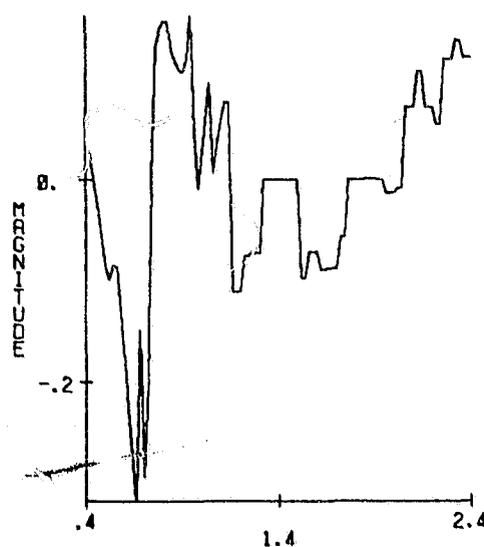
Figure 4.28 Expected mean-square error as a function of the number of terms in the Karhunen-Loeve expansion for Finney County, June 26, 1977, using weight function number 2.



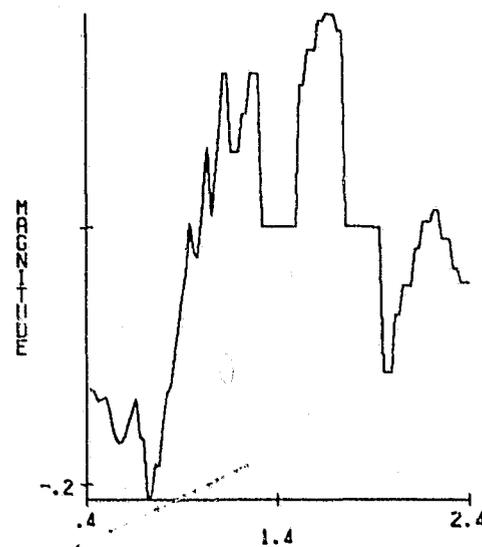
(1) WAVELENGTH (MICROMETERS)
EIGENVECTOR 1



(3) WAVELENGTH (MICROMETERS)
EIGENVECTOR 2



(2) WAVELENGTH (MICROMETERS)
EIGENVECTOR 3



(4) WAVELENGTH (MICROMETERS)
EIGENVECTOR 4

Figure 4.29 First twelve eigenvectors for Finney County, June 26, 1977, using weight function number 2.

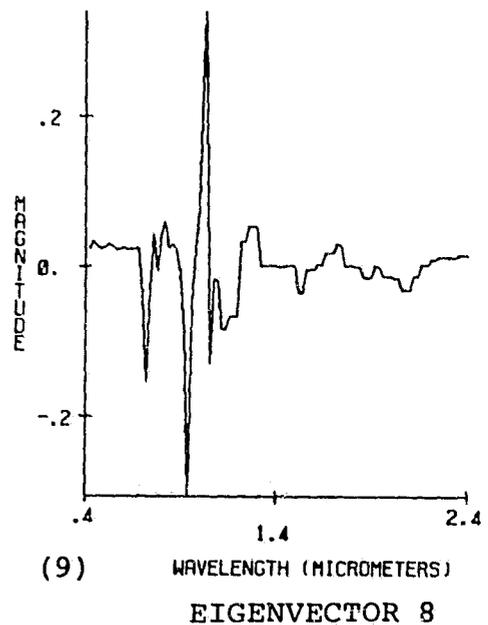
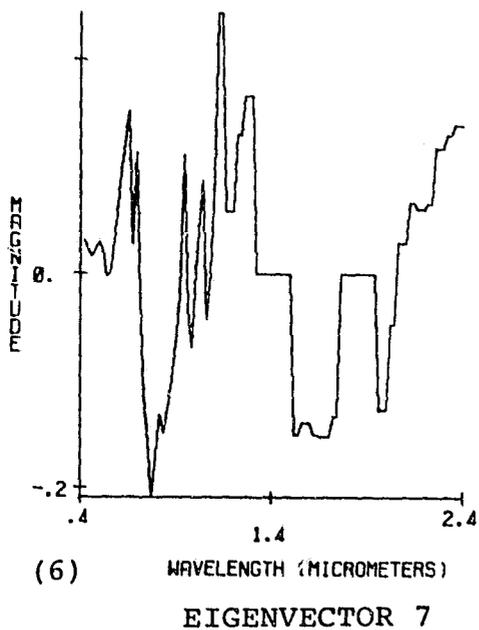
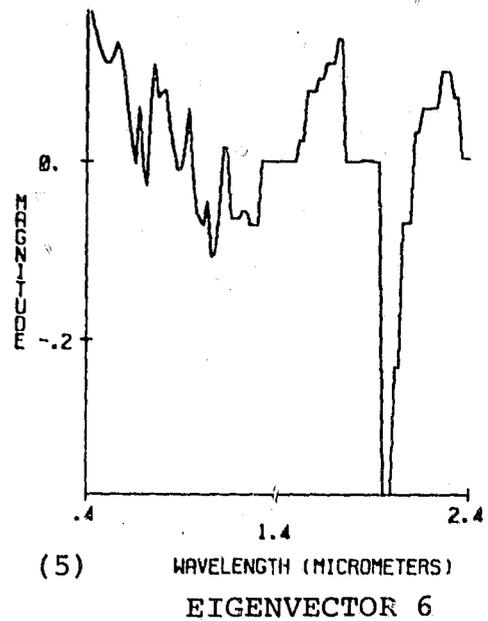
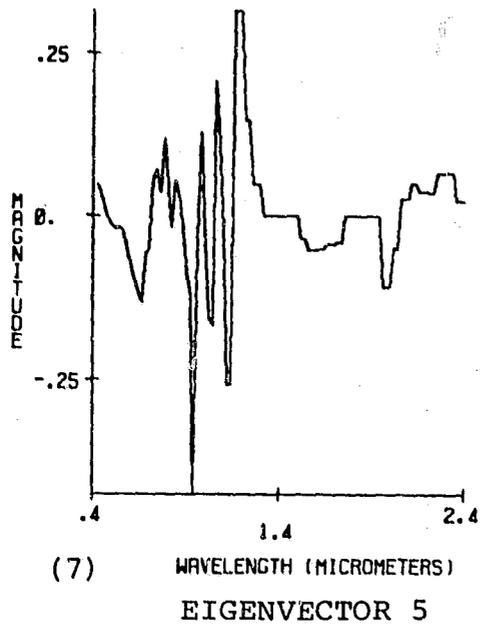


Figure 4.29 (cont.)

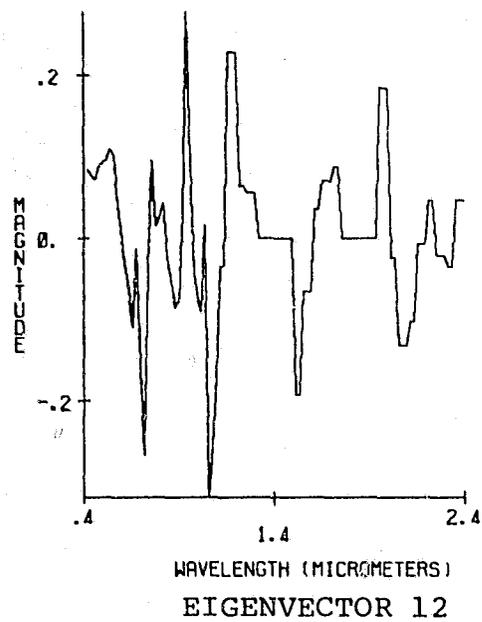
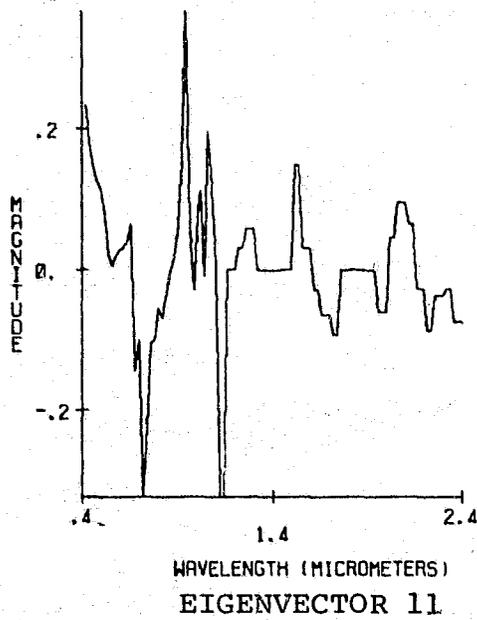
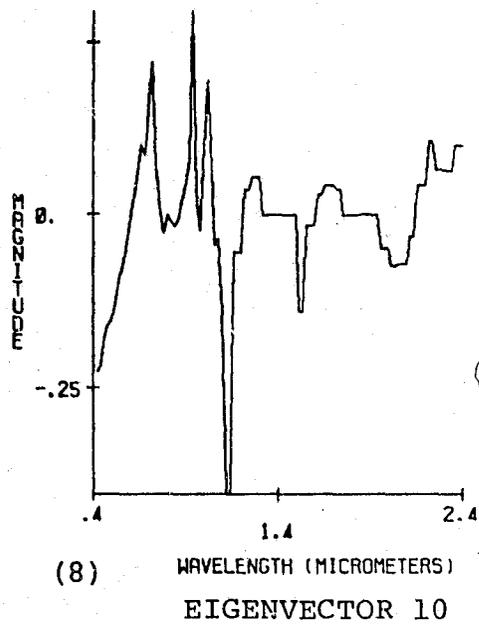
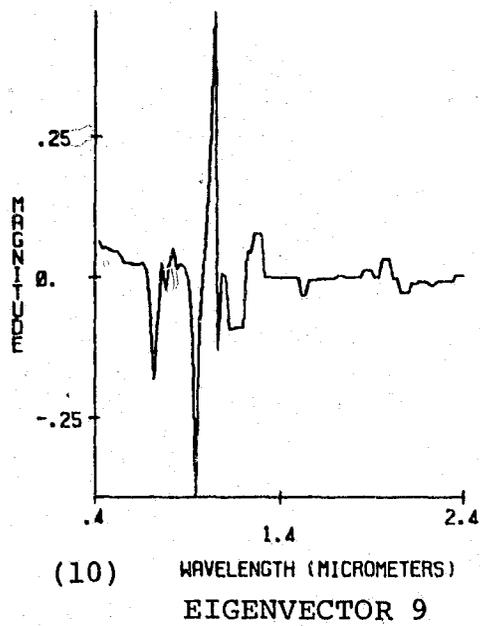


Figure 4.29 (cont.)

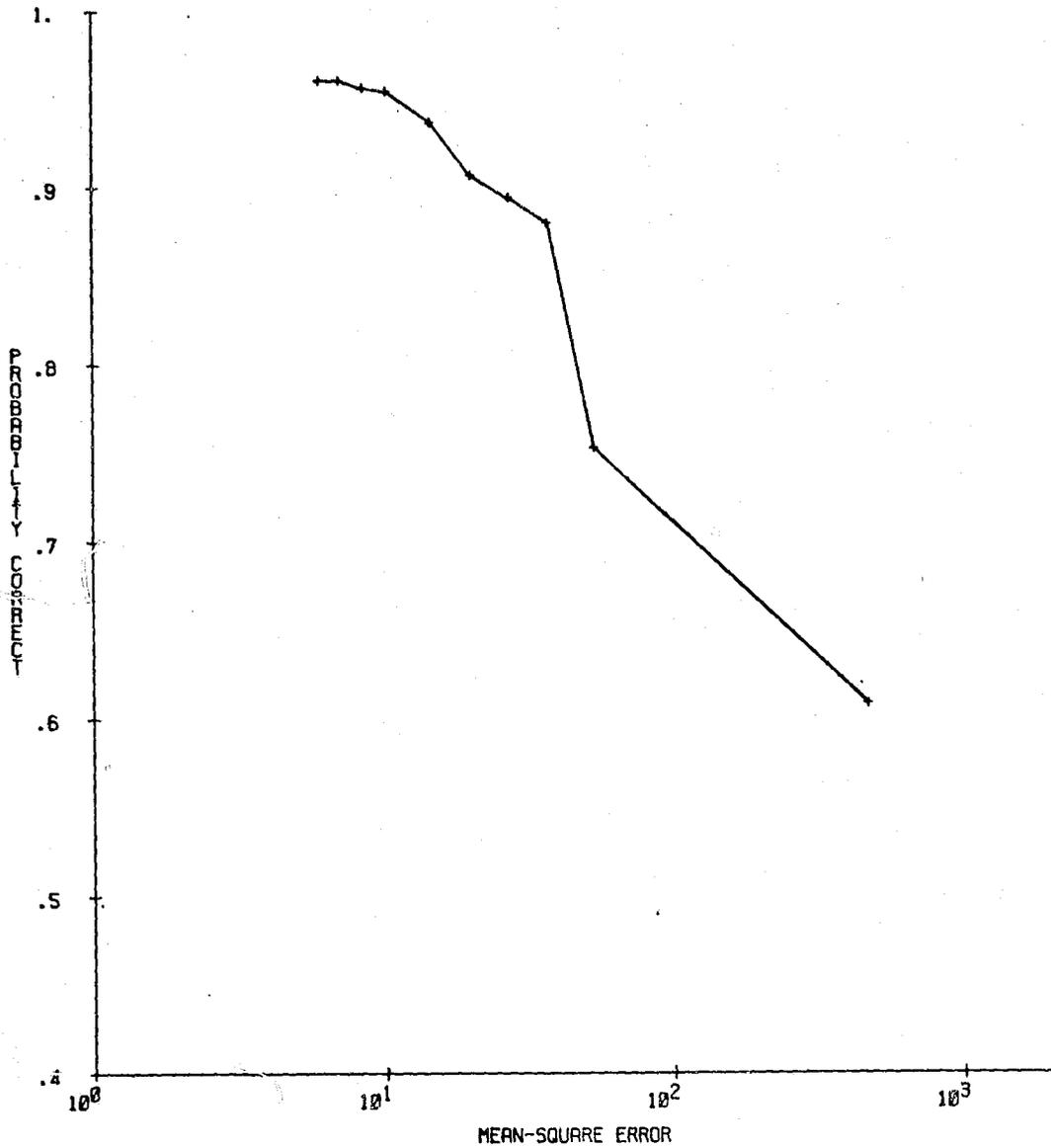


Figure 4.30 Estimate of probability of correct classification vs expected mean-square error for Finney County, June 26, 1977, using weight function number 2.

The first twelve weighted eigenvectors for each set of data are shown. Note that the graphs are of weighted eigenvectors; that is, $F_i(\lambda)$ where

$$F_i(\lambda) = \phi_i(\lambda) w(\lambda) \quad (4.18)$$

is plotted as a function of wavelength. The weighted eigenvectors will be used to determine effective ways of sampling the spectrum.

The important relationship between probability of correct classification and expected mean-square error is depicted in the graphs of \hat{P}_c vs $E\{\epsilon_r\}$ for each data set. Starting with the first eigenvector, the values of \hat{P}_c and $E\{\epsilon_r\}$ are plotted as the number of terms in the Karhunen-Loeve expansion is increased up to ten terms. Again a logarithmic scale is used for the mean-square error.

4.5 Scene Understanding

Although the primary thrust of this research was to arrive at an analytical approach to sensor design, it has beneficially resulted in some important contributions to scene understanding. Four important characteristics of the scene can be studied using the analysis procedure that has been developed, here - the dimensionality of the observation space, the determination of the important regions of the spectrum, the relationship between spectral

representation and classification performance, and the maximum achievable classification performance. These characteristics are evaluated over the limited number of data sets available.

4.5.1 The Dimensionality of the Observation Space

The dimensionality of the observation space is determined by the minimum number of basis functions required to reduce the expected mean-square representation error to a value below a specified level T . The problem becomes that of determining an appropriate value for T . Consider the expected measurement error discussed earlier. This measurement error is an attempt to quantify the capability of the field data gathering system to make accurate measurements. If the value of T is much less than the expected measurement error, then, one would expect that no real improvement in performance may be achieved by increasing the number of terms in the expansion.

The expected measurement error for each of the six data sets is listed in Table 4.4. Two choices for T will be considered. First, let the ratio of the expected measurement error to T_1 be ten-to-one. The number of terms required to reduce the expected mean-square representation error to less than T_1 is six in all but the first data set where only five terms are required (see Table 4.4). Six terms appears to be a very reasonable

Table 4.4 Expected measurement error and proposed values for T for each of the data sets.

Data Set	Expected Measurement Error	T_1	Number of Terms for T_1	T_2	Number of Terms for T_2	Number of terms $R < .99$
Williams Co. May 8, 1977	178.8	17.9	5	1.78	18	10
Williams Co. June 29, 1977	140.7	14.1	6	1.41	20	4
Williams Co. Aug. 4, 1977	127.4	12.7	6	1.27	17	5
Finney Co. Sept. 28, 1976	103.9	10.4	6	1.04	18	5
Finney Co. May 3, 1977	156.6	15.7	6	1.57	22	5
Finney Co. June 26, 1977	165.3	16.5	6	1.65	19	6

number considering both the representation accuracy and the data volume required to be transmitted to the processor. However, for the purposes of this work it is desirable to decrease T further to insure that as much information as possible is retained. Therefore, let the ratio of the expected measurement error to T_2 be one hundred-to-one. The number of terms required to reduce the expected measurement error to a value less than T_2 is approximately twenty.

A second criterion for determining how many terms in the expansion to use which has often been applied is to compute the ratio

$$R = \frac{\sum_{i=1}^N \gamma_i}{\sum_{i=1}^L \gamma_i} \quad (4.19)$$

where N is the number of terms in the expansion and L is the total number of terms available. If R is equal to 1.0 then the expected mean-square error for the process is zero. In general this occurs only when $N=L$, therefore, one must be content with choosing of value of R close to 1.0. Suppose that we choose $R=0.99$ and require that the number N be chosen such that the right-hand term in equation 4.19 is greater than R . This would guarantee that the expected representation error would be less than 1% of the total signal 'energy'. The last column in Table 4.4 lists the number of terms required to achieve this representation

accuracy. The first twenty eigenvectors were used in the analysis of the data which are presented in this research.

4.5.2 Feature Selection

It is desirable to evaluate the optimal set of basis functions to determine which features are contributing the most toward the discrimination between classes in a given problem. To evaluate the features it is proposed to rank them according to their ability to discriminate between classes. This ranking will achieve three purposes. First the ranking will indicate whether the order of the features based on expected mean-square error is relevant to the classification problem. Second by examining the eigenvectors of the most significant features, some information regarding the selection of the best set of features to use in the classifier is obtained. Finally, the relationship between the observed spectral response variations and the phenomena being observed on the earth's surface can be examined more closely, since the most significant variations which affect separability can now be determined.

For each data set, the information classes have been specified. The features in the optimal set will be evaluated based on the following criteria:

- Estimate of probability of correct classification for each feature.
- Computation of a separability measure (divergence) for combinations of features and ranking according to highest average separability.

- Estimate of probability of correct classification for combinations of features.

The rankings for the first ten optimal features are listed in Table 4.5. Note that the rankings are somewhat subjective because the importance of a particular feature may be different when used in combination with other features than when used alone. However, those features at the top of the lists are definitely superior to those at the bottom. For convenience the rankings are denoted by a number in parenthesis indicating the rank below each of the first 10 eigenvectors plotted in section 4.4.5.

In general, the ranking in Table 4.5 bears some similarities to the ranking based on expected mean-square error. For example, feature 1 is ranked first in two of the six data sets and second in two others while never being ranked below fifth. The low ranking for the May 8, Williams County data is not surprising since the first eigenvector is very similar to bare soil and the responses from both emerging wheat and fallow fields are similar to that characterized by bare soil. The first eigenvector would not be expected to be of much value for discriminating between the WHEAT and FALLOW classes. Feature 2 is also ranked high for all of the data sets. At the other end of the list features 9 and 10 are consistently at or near the bottom.

Table 4.5 Ranking of the first 10 optimal features on their ability to discriminate between classes.

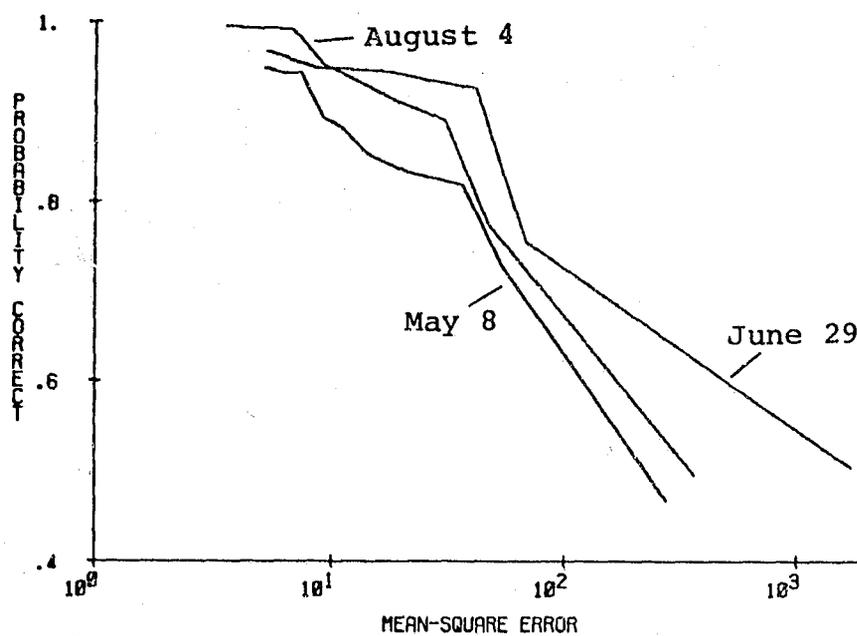
<u>Rank</u>	<u>May 8, 1977</u>	<u>June 29, 1977</u>	<u>Aug. 4, 1977</u>	<u>Sept. 28, 1976</u>	<u>May 3, 1977</u>	<u>June 26, 1977</u>
1	2	3	2	1	2	1
2	6	2	1	2	1	3
3	3	1	3	4	5	2
4	8	4	7	3	4	4
5	1	5	4	7	3	6
6	5	8	6	5	10	7
7	4	7	8	6	8	5
8	5	6	5	10	7	10
9	8	9	9	8	9	8
10	9	10	10	9	6	9

Only the first 10 features were ranked, however, since twenty features were available a check was made of the second ten for features which may be significant. The importance of these features was determined by estimating the probability of correct classification in combinations with other features as well as by themselves. For the Williams County data sets features 11 and 12 were important for the May 8 data and for the 29th data. For the Finney County data sets features 11 and 13 were important for the Sept. 28 date while features 15 and 14 were significant for the May 3rd and June 26th dates respectively.

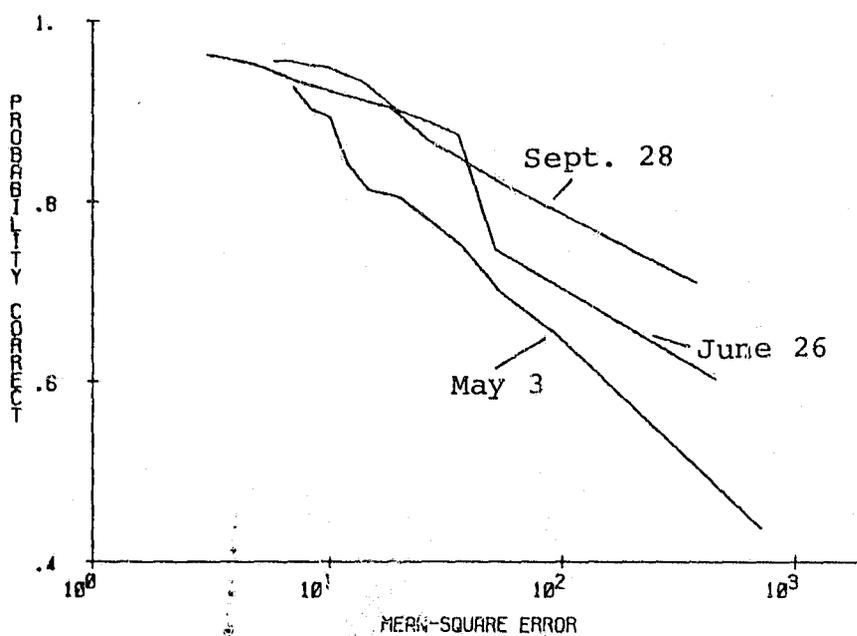
The evaluations of the spectral interval to select features for the classifier and to interpret observed phenomena will be discussed in the next sections.

4.5.3 Classification Performance as a Function of the Spectral Representation

The relationship between the overall pattern recognition system performance and the spectral representation parameter is graphically displayed by plotting the probability of correct classification, P_c , as a function of expected mean-square error, $E\{\epsilon_r\}$. These graphs are plotted again in Figure 4.31 with the three graphs for each location on the same coordinates. One can evaluate which terms contribute to the performance as well as to the representation.



(a)



(b)

Figure 4.31 Estimate of probability of correct classification vs expected mean-square error for (a) Williams County and (b) Finney County, using weight function number 2. (See also Figures 4.15, 4.18, 4.21, 4.24, 4.27, and 4.30.)

The graph for the May 8, 1977, Williams County data is typical. The value of \hat{P}_C increases steadily with decreasing $E\{\epsilon_r\}$. At the fourth term the graph begins to level off at a value of 0.83, indicating that \hat{P}_C is may be close to a maximum. However, at the eighth term the value of \hat{P}_C increases significantly for a corresponding small decrease in $E\{\epsilon_r\}$ before leveling off at about $\hat{P}_C = 0.95$. The June data set from Williams County has a similar graph with the final leveling off beginning at about the fifth term. Comparing these two data sets, a smaller mean-square error is required in the May data to achieve an equivalent classification performance. Hence fewer terms or dimensions are required to achieve a given level of performance.

The last data set from Williams County does not exhibit the early leveling off noted in the first two-sets. The performance improves steadily until it reaches approximately 1.0 at the seventh term.

The September 28, 1976 data set from Finney County has a steady increase in performance with decreasing mean-square error until the leveling occurs at about $\hat{P}_C = .96$. Note that the value of \hat{P}_C for the first term is the highest of the six graphs; hence, a lot of discriminating information is present in the first term. The graph associated with the May 3, 1977, Finney County data set is still increasing

at the tenth term, indicating that more terms are necessary to determine the maximum performance. The graph for the last set from Finney County is similar to the graphs for the first two data sets from Williams County.

The graph of P_c vs $E\{\epsilon_r\}$ can be used to determine the degree of representation accuracy required to achieve a specified level of performance. For the data for Williams County on June 29 a relatively high value of $E\{\epsilon_r\}$ is acceptable; whereas, for the May data from Finney County requires a more accurate representation.

For these curves there does not appear to be any trends based on location of the data sets. There does seem to be a trend as far as the time of the growing season at which the data was collected is concerned. The May dates in both locations tend to require more representation accuracy and tend to still be increasing in performance after using 10 terms.

The asymptotic properties can be used to estimate the value of the maximum achievable classification performance. To find the maximum performance let $E\{\epsilon_r\}$ approach zero and observe the value of \hat{P}_c . In most cases \hat{P}_c will be constant or increasing very slowly as $E\{\epsilon_r\}$ becomes small. The value of the constant to which \hat{P}_c is approaching is maximum value of the probability of correct classification. Table 4.6 lists the maximum probability of correct classification for each data set. Note that for the May 3,

Finney County, data, \hat{P}_c is still increasing so that the maximum value of \hat{P}_c is probably higher than that listed.

Table 4.6 Maximum probability of correct classification for the six data sets.

<u>Data Set</u>	<u>Approximate maximum probability of correct classification</u>
Williams Co., May 8, 1977	.95
Williams Co., June 29, 1977	.96
Williams Co., Aug. 4, 1977	1.00
Finney Co., Sept. 28, 1976	.96
Finney Co., May 3, 1977	.93
Finney Co., June 26, 1977	.95

4.5.4 Characteristics of the Eigenvectors

For the six data sets there are some general characteristics of the eigenvectors which can be readily observed. The contribution of the spectral response to the channel or feature which corresponds to the eigenvector is determined by the portions of the spectral interval where the eigenvector has a magnitude or sensitivity different from zero. This sensitivity is apparent from the linear functional which determines the coefficients

$$x_i = \int_{\lambda} x(\lambda) \phi_i(\lambda) w(\lambda) d\lambda \quad (4.20)$$

Therefore, a subinterval of Λ which has relatively large values for $\phi_i(\lambda)$ and $w(\lambda)$ will contribute significantly to the informational value of the coefficient x_i .

The eigenvectors provide some insights into the correlation between adjacent regions of the spectrum. Let the spectrum be sampled using very fine spectral bands. Let the measurements using these bands be denoted by u_i , $i = 1, 2, \dots, 100$. The correlation between any two of the measurements is given by

$$E\{u_i u_j\} = \sum_k \phi_{ik} \phi_{jk} \gamma_k \quad (4.21)$$

where ϕ_{ik} is the i th element of the k th eigenvector. If the correlation between two adjacent measurements u_i and u_{i+1} is high, then, the two measurements are not independent and they could be combined into a single measurement.

It is now possible by examining the eigenvectors to determine how narrow the spectral measurement bands should be in various parts of the spectrum. Eigenvectors which have high frequency variations in magnitude in a particular subinterval of the spectrum strongly indicate that it may be desirable to sample that subinterval using very narrow spectral bands.

Referring to the results presented in section 4.4.5, the first eigenvector typically has the characteristics of the weighted mean function of the ensemble. The second

eigenvector has a strong component between .72 and 1.3 μm and a second important component in the 1.95 to 2.4 μm band. For the Finney County data set taken on September 28, 1976 these two eigenvectors are reversed in order. The third and fourth eigenvectors in all of the data sets exhibit noticeable similarities. In the third eigenvectors Williams County data and the fourth eigenvectors for Finney County data there exists a significant component in the subinterval between 1.5 and 1.7 μm . The sensitivity in the visible region from .55 to .70 μm is strongest in the fourth eigenvectors for Williams County data and the third eigenvectors for the data from Finney County. These similarities over the different data sets are somewhat surprising and also encouraging in that these similarities indicate a strong possibility that a sensor can be built which will work very well over more than just a single data set.

As eigenvectors which are later in the sequence of optimum basis functions are examined, there is an increased occurrence of subintervals with high frequency variations in magnitude. It is of interest to note that several of these terms were important for classification performance. Examples of important eigenvectors which have high frequency variations are the sixth and eighth eigenvectors from the May 8, Williams County data and the seventh eigenvector from the August 4, Williams County data.

It was observed that the subintervals from 0.6 to 0.7 μm and from 0.9 to 1.1 μm have considerable high frequency variation. The 0.6 to 0.7 μm band has often been suggested as very important for identifying green vegetation. In particular, the chlorophyll absorption band centered at about 0.65 μm is present (Hoffer, 1978). Differences in the chlorophyll pigmentation are indicators of plant stress. Other pigments are also present in the visible part of the spectrum. Therefore, there is good evidence that narrow spectral bands in the region between 0.6 and 0.7 μm may be helpful. The spectral interval between 0.7 and 1.1 μm also possesses high frequency variations; however, some of these variations can be traced to water absorption bands occurring at 0.76, 0.93, and 1.12 μm . Furthermore tests using narrow spectral bands in this region did not improve the classification performance significantly over using a wide spectral band.

The significant sensitivity of important eigenvectors in the spectral bands from 1.5 to 1.7 μm and 1.96 to 2.4 μm clearly indicates that these bands should be included in the design. The importance of including these two bands was further substantiated by improved classification performance.

4.6 Suboptimal Sensor Design

The analytical procedure which has been developed and tested is particularly useful as a tool for the design of

practical sensor systems. Significant contributions to the design have been made through improved scene understanding; however, the primary purpose is to be able to design a practical sensor system by specifying a particular set of spectral bands $\{\psi_i(\lambda)\}$. The optimal set of basis functions generated by the procedure provides a standard against which any suboptimal practical sensors can be compared. In addition, the optimum basis functions $\{\psi_i(\lambda)\}$ provide information regarding the proper choice of spectral bands.

4.6.1 Comparison with Suboptimal Systems

An important use of the optimal design is to use it as a standard for comparing suboptimal systems. Two suboptimal sensors similar to existing or future practical scanner systems were simulated using the spectral bands listed in Table 4.1. The basis functions for these two sensors are given by

$$\psi_i(\lambda) = \begin{cases} 1.0, & \lambda_k \leq \lambda \leq \lambda_{k+1} \\ 0.0, & \text{elsewhere} \end{cases} \quad (4.22)$$

where the λ_k are the endpoints listed in Table 4.1.

The sensors are first compared on the basis of expected mean-square error. In Table 4.7 the mean-square error for the two suboptimal sensors is compared with the optimal sensor for all six data sets. The mean-square error for the optimal sensor is shown, using the first four, six, and ten eigenvectors. The units of the expected mean-square error are relative and are significant for comparison purposes only. The second weight function (Figure 4.6b) was used for all error computations. The large difference in mean-square error between the suboptimal and the optimal sensors is due to the fact that sensors one and two do not attempt to represent the entire spectral interval from 0.4 to 2.4 micrometers. Figure 4.32 illustrates how a large contribution to the mean-square for the suboptimal sensors results from the lack of spectral channels in large portions of the spectrum.

Comparison can also be made on the basis of overall pattern recognition system performance. For each data set information classes were selected. The performance criterion was the probability of correct classification. The performance of the two sensors is compared with the optimal sensor in Figures 4.33 through 4.38. Using ten eigenvectors in the representation of the ensemble, the best four features and the best six features as determined by feature selection were evaluated. The choice of four and six features was made because suboptimal sensors one and

Table 4.7 Comparison of expected mean-square error (in relative units) for each of the six data sets using two suboptimal sensors and the optimal sensors consisting of the first 4, 6, and 10 eigenvectors.

<u>Data Set</u>	<u>Sensor 1</u>	<u>Sensor 2</u>	<u>First Four</u>	<u>First Six</u>	<u>First Ten</u>
Williams Co. May 8, 1977	28570	17340	21.30	11.04	5.144
Williams Co. June 29, 1977	17320	16380	26.31	11.37	5.253
Williams Co. Aug. 4, 1977	18070	14010	19.76	9.315	3.539
Finney Co. Sept. 28, 1976	13360	11650	18.19	7.133	3.035
Finney Co. May 3, 1977	22110	16080	36.67	14.72	6.968
Finney Co. June 26, 1977	23210	17760	26.19	13.98	5.769

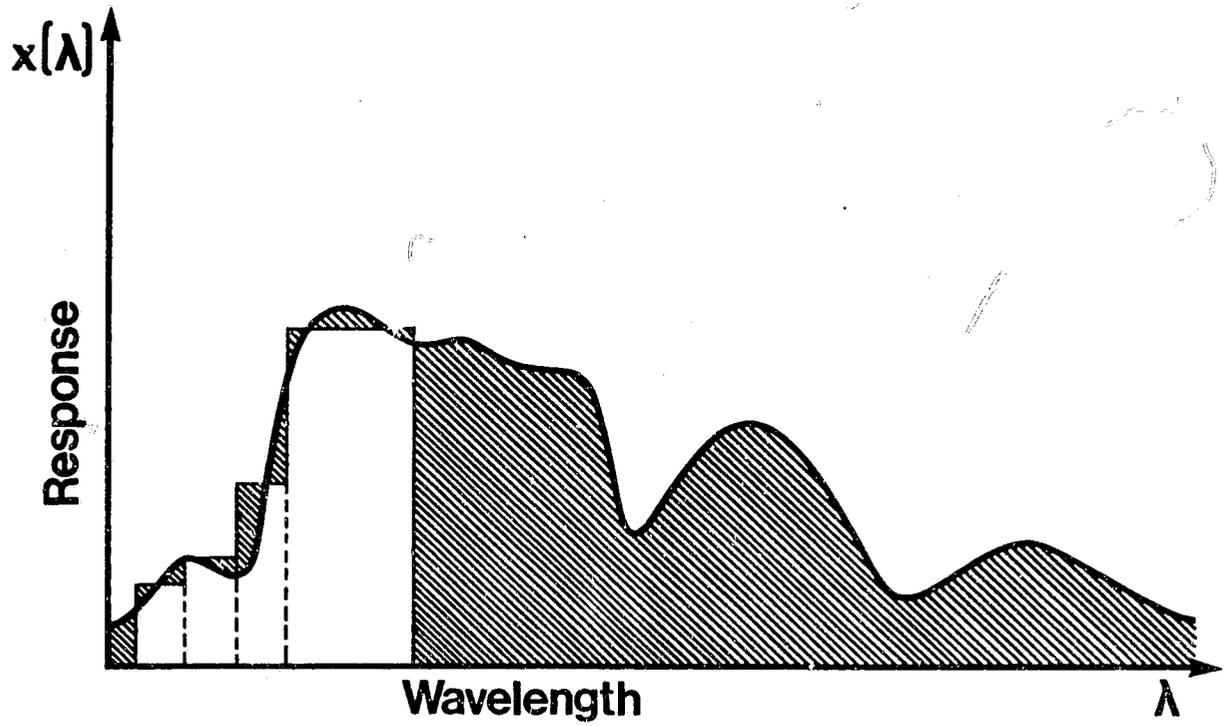


Figure 4.32 Regions of the spectral interval which are not represented by a suboptimal sensor and which contribute heavily to the mean-square error.

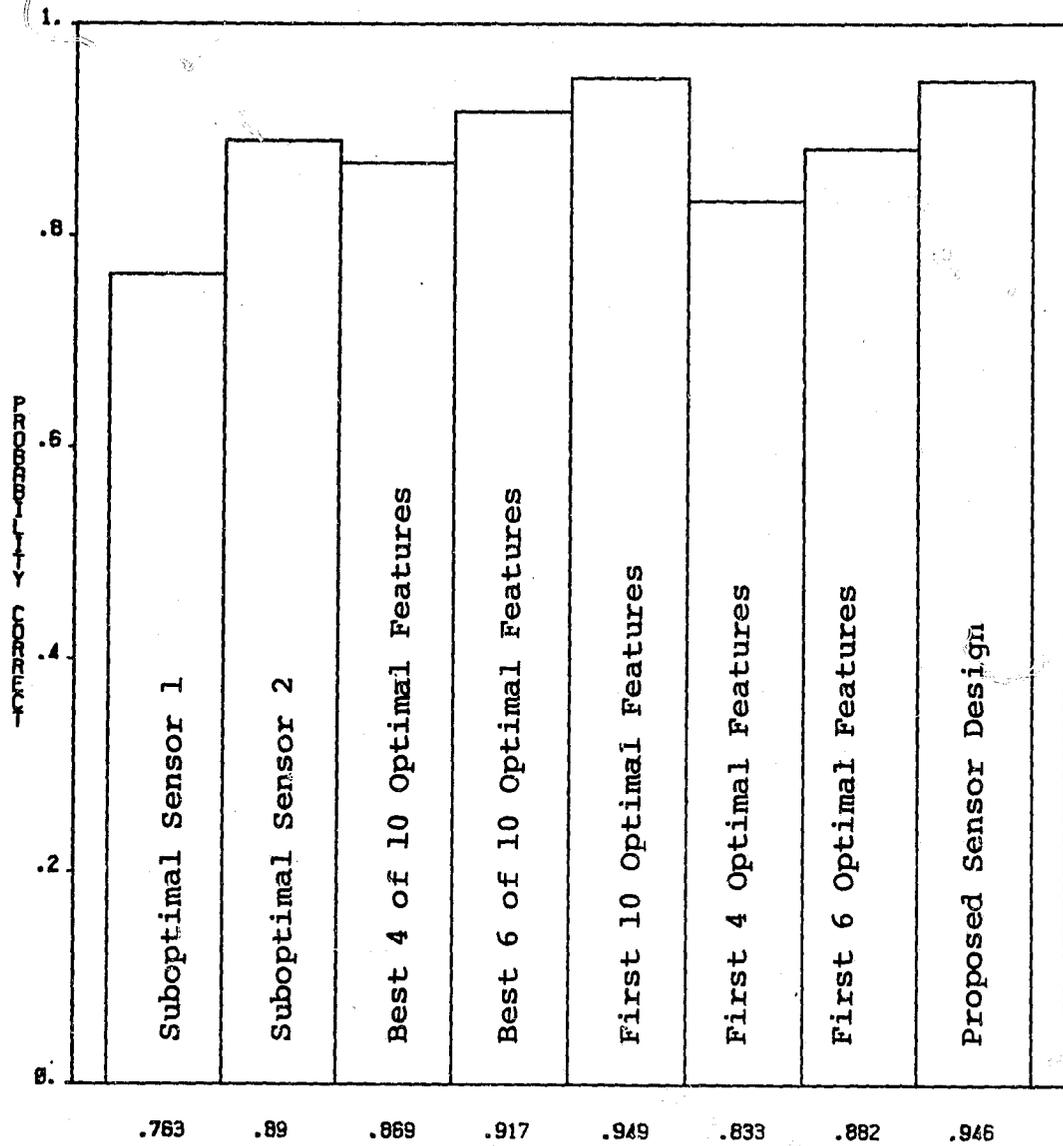


Figure 4.33 Comparisons of probability of correct classification for several sensors for Williams County, May 8, 1977.

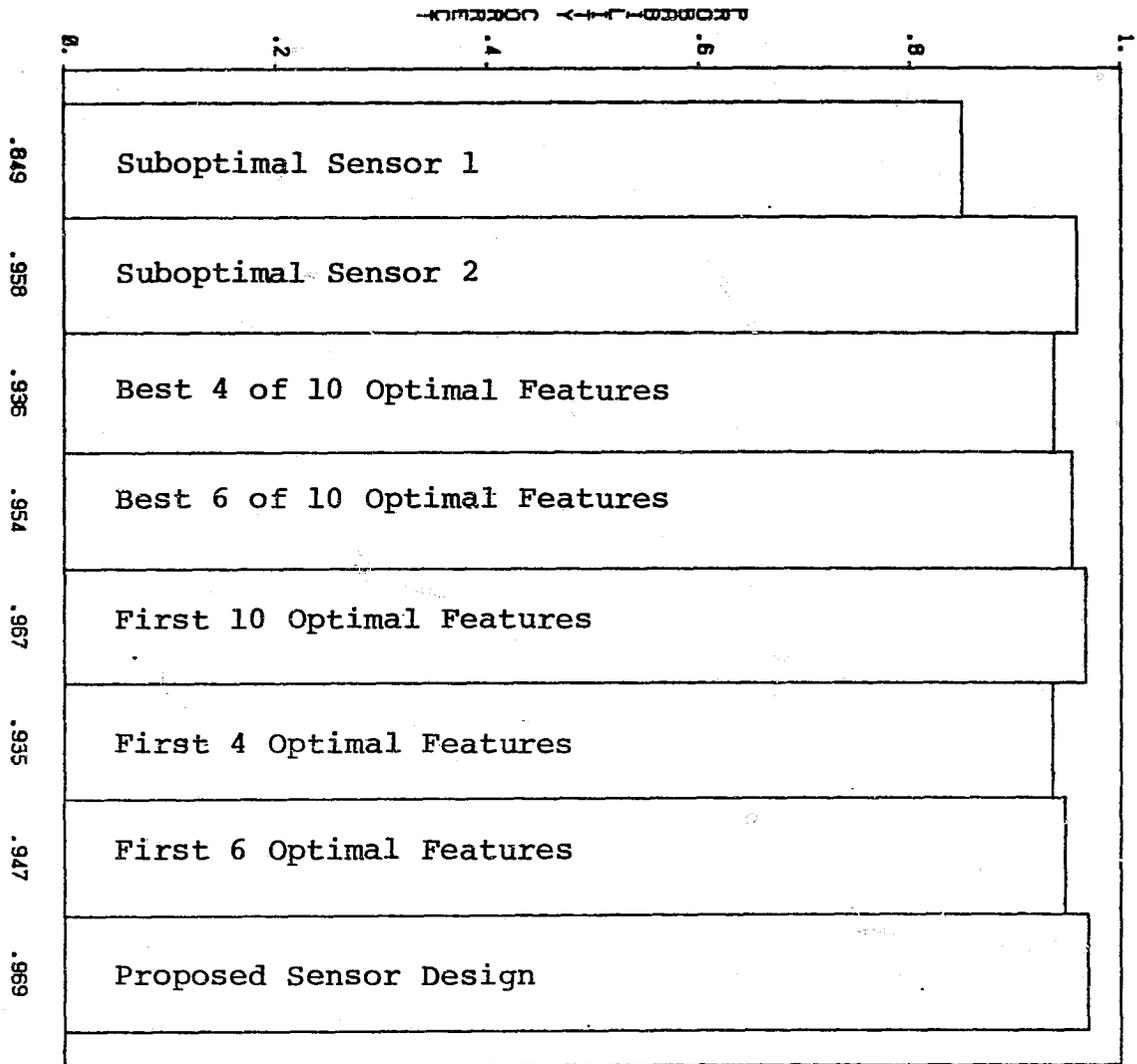


Figure 4.34 Comparisons of probability of correct classification for several sensors for Williams County, June 29, 1977.

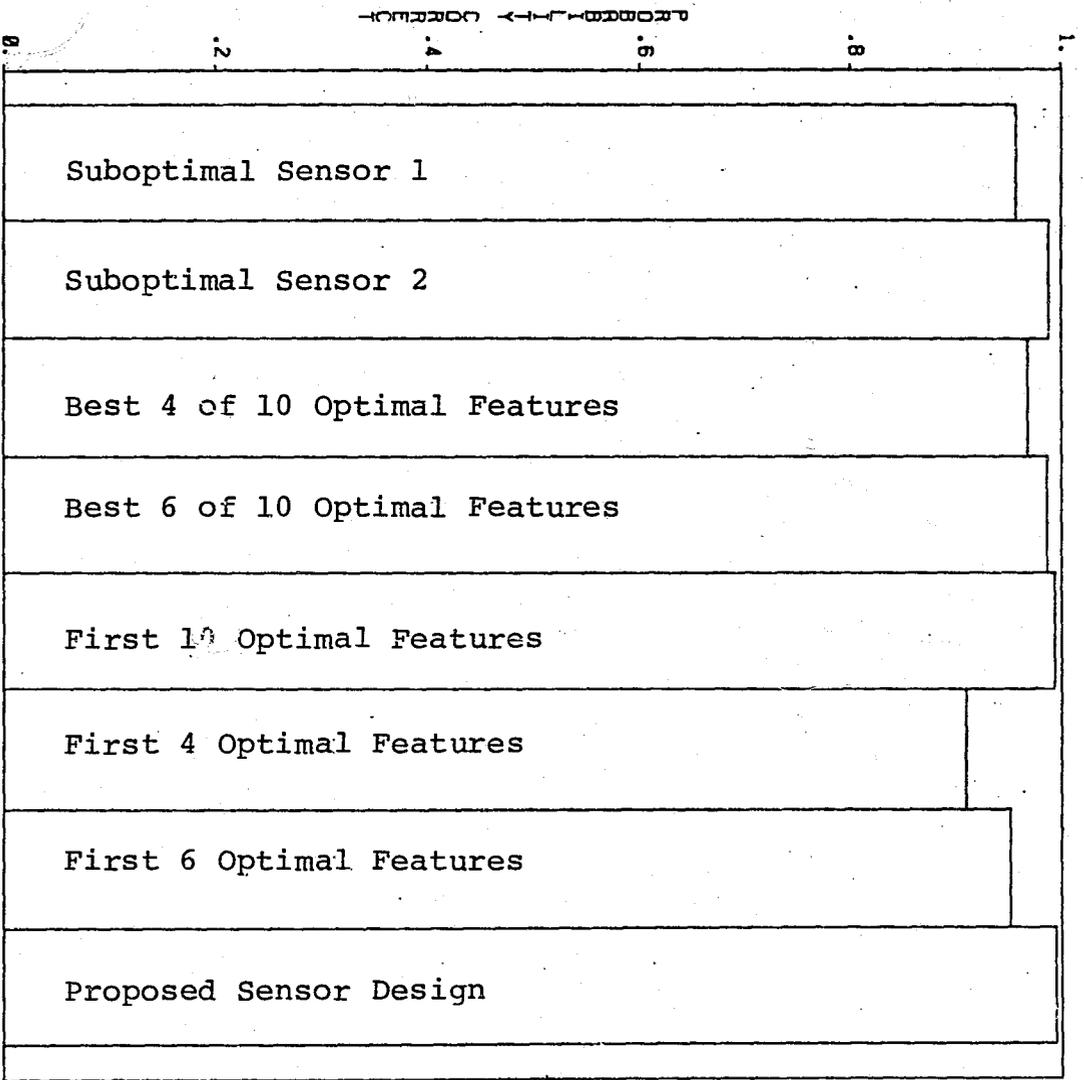


Figure 4.35 Comparisons of probability of correct classification for several sensors for Williams County, August 4, 1977.

D-3

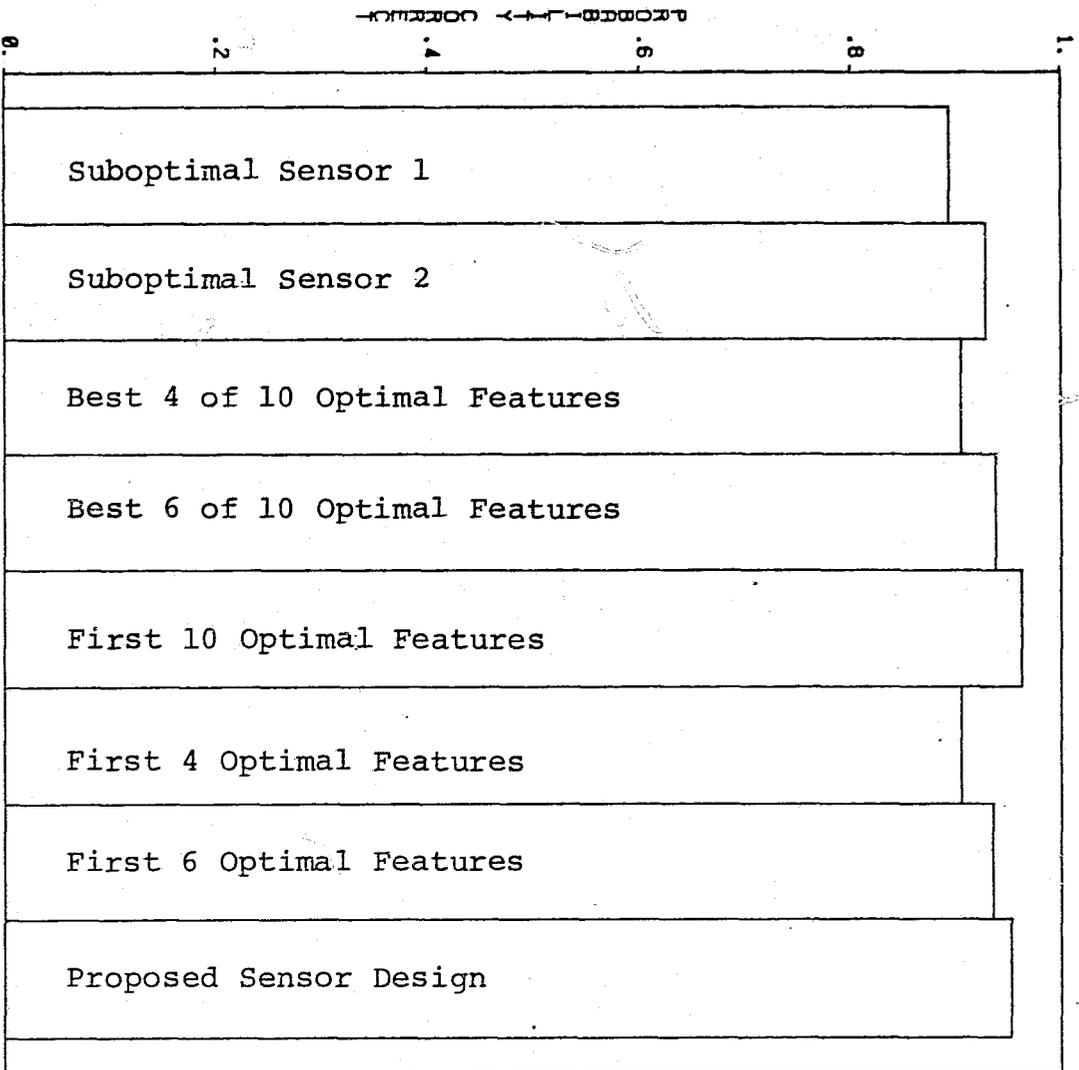


Figure 4.36 Comparisons of probability of correct classification for several sensors for Finney County, September 28, 1976.

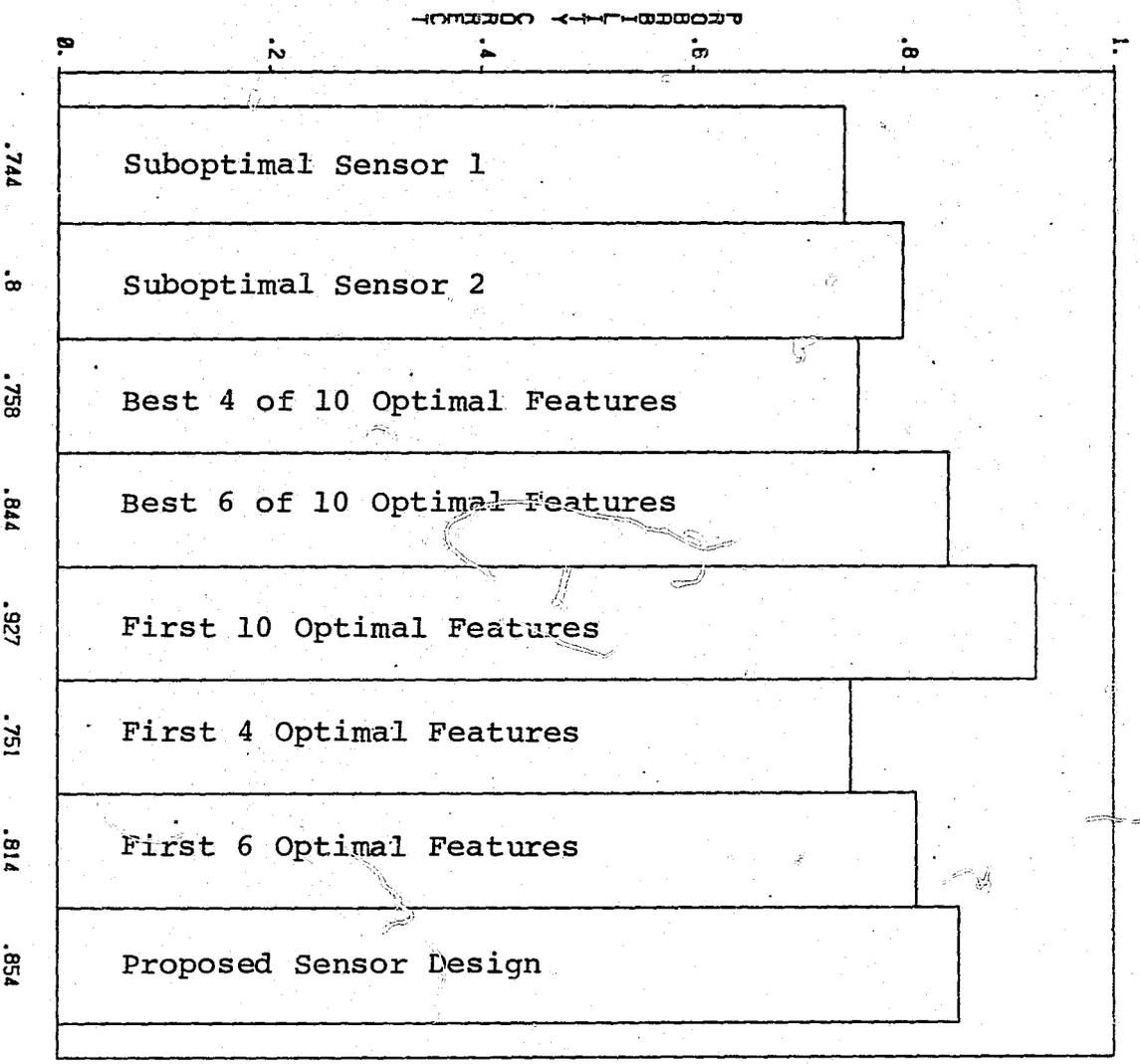


Figure 4.37 Comparisons of probability of correct classification for several sensors for Finney County, May 3, 1977.

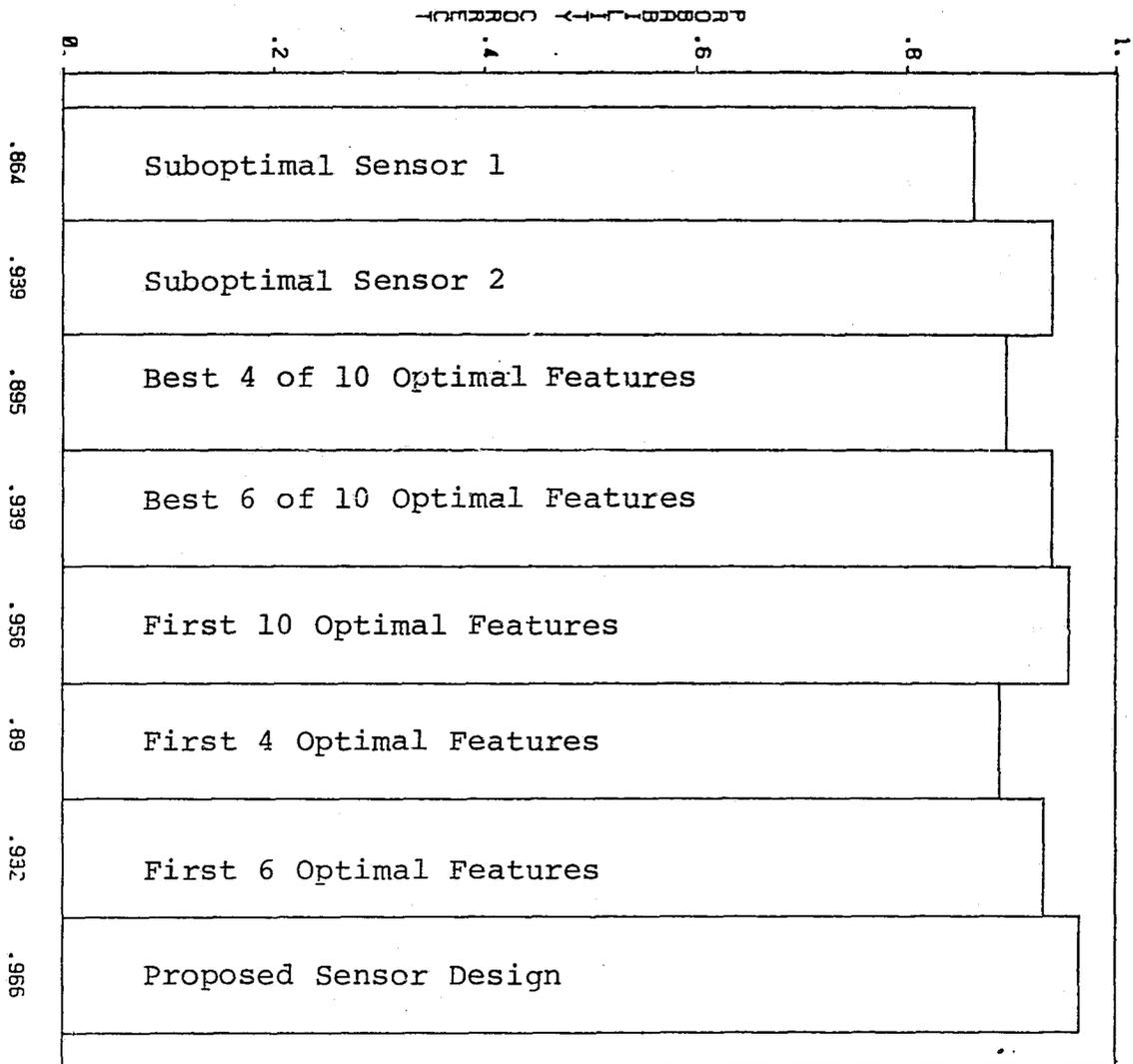


Figure 4.38 Comparisons of probability of correct classification for several sensors for Finney County, June 26, 1977.

two have four and six spectral channels respectively. The classification performances, using the first four and the first six eigenvectors as well as the first ten eigenvectors, are also provided for comparison. Several observations can be made from comparing the performances. In general considerable improvement in classification performance can be achieved over that of sensor number one. In several cases the estimate of the probability of correct classification for sensor 1 one was significantly less than any of the other combinations of channels presented. Sub-optimal sensor number two does quite well, however, even approaching in some cases the performance of the optimal sensor using the first ten eigenvectors.

For the chosen information classes, a very accurate representation of the original spectral response function is not required to obtain good performance. The information contained in the unused portion of the spectrum does not appear to be essential for the identification of these classes. However, for a set of information classes which are deeper in the information tree, a representation with smaller expected mean-square error may be necessary.

There is evidence that measurements made by the optimal set of basis functions are uncorrelated. A measure of the correlation between any two measurements is the correlation coefficient given by

$$\rho_{ij} = \frac{E\{(x_i - m_i)(x_j - m_j)\}}{\left[E\{(x_i - m_i)^2\} E\{(x_j - m_j)^2\} \right]^{1/2}} \quad (4.23)$$

Two measurements x_i and x_j , $i \neq j$ are said to be uncorrelated if $\rho_{ij} = 0$. The matrix of coefficients is called the correlation matrix. From the properties of the Karhunen-Loeve expansion the off-diagonal correlation coefficients in the correlation matrix for the stochastic process corresponding to a stratum are zero. Therefore, the measurements on the process are uncorrelated. However, the class conditional correlation matrices in general do not exhibit uncorrelated measurements (Bharucha and Kadota, 1969). In practice it was found that the the class conditional statistics are still relatively uncorrelated. As an example, the correlation matrices for the three classes from the data taken over Williams County, on June 29, 1977, the four band suboptimum sensor number 1 of Table 4.1 were computed. These matrices are listed in Table 4.8. The first two channels of the suboptimal sensor are highly correlated and the third and fourth channels are highly correlated. The correlation matrices for the first four optimal basis functions over the same data set are presented in Table 4.9. There is considerably less correlation between any pair of channels in the optimal sensor for any of the three classes. The fact that the measurements are uncorrelated implies that the redundancy

of information in the measurements is minimized, and maximum performance can be achieved with a minimum number of features.

Table 4.8 Correlation matrices for the four band suboptimal sensor number 1 using data taken over Williams County on June 29, 1977.

Class WHEAT

1.00				
0.99	1.00			
0.45	0.46	1.00		
0.22	0.24	0.96	1.00	

Class FALLOW

1.00				
0.99	1.00			
0.84	0.83	1.00		
0.68	0.67	0.95	1.00	

Class PASTURE

1.00				
0.99	1.00			
0.76	0.81	1.00		
0.68	0.73	0.99	1.00	

Table 4.9 Correlation matrices for the first four optimal basis functions using data taken over Williams County on June 29, 1977.

Class WHEAT

1.00				
-0.45	1.00			
-0.23	-0.35	1.00		
0.01	0.25	-0.15	1.00	

Class FALLOW

1.00				
-0.02	1.00			
0.24	0.30	1.00		
0.30	-0.30	0.02	1.00	

Class PASTURE

1.00				
-0.62	1.00			
-0.79	0.30	1.00		
-0.34	0.07	0.49	1.00	

4.6.2 Evaluation of Spectral Subintervals

In section 4.5.4 methods of evaluating the eigenvectors in order to determine how to select spectral channels for a practical sensor were discussed. Principally the eigenvectors are examined to identify regions which are contributing to the information content of the scene. The weight function effectively eliminated two subintervals which were shown to be of little value. The factors which are important for identifying important subintervals are the magnitudes of the eigenvectors in these subintervals and

their ranking with respect to representation error and with respect to classification performance.

Examination of the eigenvectors has revealed that the spectral bandwidths of current sensor systems may be too wide in certain subintervals of the spectrum. Two subintervals in particular appear to have significant high frequency variations to merit narrow spectral-sampling channel widths. One of these subintervals from 0.9 to 1.15 micrometers is known to have several minor molecular absorption bands which may be the cause of the increased high frequency variations. The subinterval from 0.6 to 0.9 μm also possesses significant variations in the magnitudes of the eigenvectors. This region is considered to be important for measurements on vegetation classes. Therefore, the proposed sensor design should reflect the importance of narrow sampling channels in the subinterval. Bandwidths as narrow as .02 μm may be required to achieve good performance. However, narrow spectral channels require more bands to cover the spectrum. The cost of adding more spectral channels which will cause greater data volume difficulties should also be considered during the design of the system.

4.6.3 Proposed Sensor Design

A proposed sensor is now designed using the techniques and knowledge that has been developed. It is desirable that

this sensor work well over all six data sets. The number of features or channels will be restricted to between six and ten. Only rectangular basis functions will be considered because they are orthogonal and simple to implement. The approach will be to pick a set of basis functions that will give a small expected mean-square error, and, then, compare the resulting classification performance with the optimum.

The selection of spectral channels for the proposed sensor was based upon manual examination of the eigenvectors and upon use of equation 4.21 to locate adjacent uncorrelated measurements of the spectrum. The eigenvectors over each band are studied with the intention of locating regions of the spectrum which need to be sampled with narrow spectral channels. The sampling measurements made by the field data collecting system are used to compute the correlation between measurements normalized to the respective variances. If two adjacent spectral measurements are uncorrelated, a good choice for the location of the edge of a rectangular basis function might be between the two measurements. Graphs of the correlation coefficients as a function of frequency for each data set are included in Appendix C. It should be pointed out that even though these groups indicate that two points are not correlated, there still may not be much improvement in performance as a result of locating the edge of a channel between the two points. The fact that the edges of two channels are uncorrelated does

not guarantee that the spectral channels themselves are uncorrelated. Furthermore, the examination of the eigenvectors and the computation of the correlations must be considered in light of the signal and noise properties across the spectrum. Therefore, the procedure is to design a proposed sensor using the principles discussed above and evaluate the system performance.

The proposed sensor design was developed using the May 8, Williams County, data. The spectral band locations are listed in Table 4.10 where the basis functions are, again, given by

$$\psi_i(\lambda) = \begin{cases} 1.0, & \lambda_k \leq \lambda \leq \lambda_{k+1} \\ 0.0, & \text{elsewhere} \end{cases}$$

The resulting design was tested on the remaining data sets and compared with the corresponding optimum sets of basis functions.

The performance of this sensor design was very good. The expected mean-square error for each data set is given in Table 4.11. The expected mean-square error is on the order of 1000 which is a factor of 10 less than either suboptimal sensor one or two. This value though high with respect to the optimal sensor is probably about as well as one can do with a small number of rectangular basis functions. The classification performance is listed in Table

4.11 and included in the bar graphs of Figures 4.33 to 4.38 for comparison with the other systems. Performance is significantly better in several cases to either of the suboptimal systems and very close to the 10 channel optimal system in a number of the data sets.

Table 4.10 Spectral band locations for the proposed sensor.

<u>Channel</u>	<u>Endpoints</u>
1	.42 μm - .54 μm
2	.56 μm - .66 μm
3	.68 μm - .70 μm
4	.72 μm - .90 μm
5	.92 μm - 1.00 μm
6	1.02 μm - 1.30 μm
7	1.52 μm - 1.74 μm
8	1.96 μm - 2.40 μm

Table 4.11 Expected mean-square error (in relative units) and estimated probability of correct classification using the proposed sensor.

<u>Data Set</u>	<u>E {ϵ}</u>	<u>\hat{P}_C</u>
Williams Co. May 8, 1977	939	.946
Williams Co. June 29, 1977	1700	.969
Williams Co. Aug. 4, 1977	1016	.995
Finney Co. Sept. 28, 1976	1068	.953
Finney Co May 3, 1977	1213	.854
Finney Co. June 26, 1977	1241	.966

CHAPTER 5. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The purpose of this research was to develop an analytical technique for selecting spectral channels as a part of the design of a multispectral scanner sensor system for remote sensing. The results and conclusions as a consequence of the development and implementation of this technique have been significant and are now summarized.

The spectral representation parameter is one of five suggested interrelated parameters which influence the overall pattern recognition system performance criterion. The quantity associated with the spectral representation parameter was defined by the expected mean-square error. The stochastic process, consisting of an ensemble of spectral response functions from a stratum, was represented by a series expansion in a set of basis functions suitably weighted by coefficients. By increasing the number of basis functions in the representation the expected mean-square representation error will decrease. The Karhunen-Loeve expansion was used to provide an ordered set of basis functions such that using the first N of them results in a minimum mean-square representation error over all possible choices of N basis functions.

The development of the Karhunen-Loeve expansion in Chapter 2 was generalized to include the possibility of using a weight function in order to weight the different portions of the spectrum relative to their importance. The motivation was the occurrence of strong but noisy spectral response variations in two regions of the spectrum corresponding to water absorption bands. Using the uniform weight function eigenvectors which were dominated by components in these bands were among the first five in the ordered sequence of optimal basis functions; however, their contribution to the overall performance of the system was very small. By using a weight function which was unity except in the water absorption bands where the weight was very small, the eigenvectors containing significant components from these bands were no longer in the top ten or twenty eigenvectors. A very noticeable improvement in classification performance on a term-by-term basis was noted with the inclusion of this weight function.

The analytical technique developed in this research has contributed to the understanding of the scene. The dimensionality of the observation space required to achieve sufficient representation accuracy to provide acceptable classification performance for the information classes was approximately six to eight. A more complex set of information classes may require more accurate representation which would necessitate using more basis

functions and increase the number of dimensions in the representation. The graph of the global performance criterion, which is typically the probability of correct classification, as a function of expected mean-square error is useful for studying the relationship between the spectral representation and the overall system performance. For the information classes selected the graph of \hat{P}_c versus $E\{\epsilon_r\}$ allows one to estimate the maximum probability of correct classification and to study which eigenvectors are contributing the most to the classification performance. Also, the shape of this curve indicates whether or not the selection of the basis functions with respect to the mean-square error criterion bears any relation to the contribution to classification performance. The largest contribution to improved performance occurred when the first few eigenvectors in the sequence were used. However, in several of the strata used in this work it was found that eigenvectors that were sixth or higher in the sequence of optimum basis functions made important contributions to the classification performance. In general, there is good correlation between the ranking of the basis functions on the basis of classification performance and the ranking on the basis of minimizing mean-square error.

An important aspect of understanding the scene is determining which portions of the spectral interval are

most useful. By examining which subintervals are being sampled from the eigenvectors which are most important for classification purposes, one can identify portions of the spectrum which are important and subintervals which are strongly correlated with other subintervals. The limited value of the subintervals corresponding to the water absorption bands near 1.4 and 1.9 micrometers was well-known and was verified in this research.

It was observed that the plots of the eigenvectors which were later in the sequence tended to have increased high frequency variations. Coupled with the indication that these later terms provide significant additional information for classification, it was concluded that some spectral regions may require a high spectral sampling rate. Bandwidth intervals of 0.02 μm may be required as compared to the 0.1 μm intervals used in the suboptimum sensor number one. Of particular importance was the indication of a need for fine sampling of the spectrum in the visible region corresponding to the chlorophyll absorption bands (0.55 - 0.70 μm).

The use of the weighted Karhunen-Loeve expansion was demonstrated to be a useful tool in the design of sensor systems. Two suboptimal systems which are similar to existing or planned operational sensor systems were compared with the optimal representation. For the

information classes used it was found that a very high representation accuracy was not necessary to obtain good performance. The practical sensors, which represented the spectral response functions very crudely, performed quite well compared to the system consisting of the set of optimal basis functions. However, there is a significant improvement in performance that can be achieved by a better representation in several of the cases.

A proposed sensor design was developed using the design procedure. The proposed sensor consisted of eight rectangular bands selected on the basis of the information provided by the procedure. The performance of the proposed design was superior in classification performance to two other practical sensor designs and very much superior in representation accuracy. For the information classes used the classification performance of the proposed sensor was very close to the maximum possible in most cases.

The conclusions drawn so far are based on a very limited collection of strata. To carry out the procedure such that the collection is representative of all possible strata that a given sensor may observe would require many more sets of data. Suppose, for example, that it is desired to use a sensor to map vegetation in the United States. Only wheat growing areas of the central plains are represented by the two locations used in this work.

The spectral variations that are peculiar to agricultural scenes in the midwestern cornbelt, the small farms of New England, the southern cotton belt, or the fresh produce growing regions of the far west are not represented. Furthermore, other useful areas which may be of interest such as urban areas, forest lands, deserts, mountainous regions and large bodies of water are not included in the representation. At present the available data is primarily taken over the great plains and the midwest. The helicopter-mounted sensor has proved to be an efficient method of gathering a sufficient amount of measurements in a short amount of time. The time-consuming effort that is needed is the collecting and correlating of ground truth information which will allow one to use various sets of information classes.

An important concept which has been alluded to but which requires further investigation is the design of methods for insuring that the ensembles assembled are representative. Specifically it would be desirable to be able to make some quantitative assessment as to whether or not the collection of spectral response functions are representative of the ensemble associated with a stratum and whether or not the set of strata are representative of all possible strata which the sensor may observe.

The five parameters were discussed at some length in Chapter 3. A considerable body of research results has been collected relating each of these parameters to classification performance and in some cases showing the interdependence of the parameters. However, at present only limited attempts to vary all of the parameters simultaneously to arrive at some optimal set have been reported. It is recommended that sets of data be assembled which would allow one to vary all of the parameters. The available knowledge should provide guidelines for the proper design of such a collection. Recommended variables are the mean-square representation error, the ground resolution element size, added white noise power, number of training samples, and the information trees which correspond to the spectral representation, spatial representation, S/N, ancillary data, and information class parameters respectively. Also, it would be desirable to have available several other classifiers including a spatial classifier to evaluate performance.

BIBLIOGRAPHY

- Abend, K., Harley, T. J., Jr., Chandrasekaran, B., and Hughes, G. F., (1969), "Comments on the mean accuracy of statistical recognizers," IEEE Trans. Inform. Theory, Vol. IT-15, pp 420-423.
- Akhiezer, N. I. and Glazman, I. M., (1961), Theory of Linear Operators in Hilbert Space, Vol. 1, Frederick Unger, New York.
- Allais, D. C., (1964), "The selection of measurements for prediction," Technical Report No. 6103-9 Stanford Electronics Laboratories, Stanford, California.
- Anderson, T. W., (1962), An Introduction to Multivariate Statistical Analysis, Wiley, New York.
- Ash, R. B., (1967), Information Theory, Wiley Interscience, New York.
- Bartolucci, L. A., Robinson, B. F., and Silva, L. F., (1977), "Field measurements of the spectral response of natural waters," Photogrammetric Engineering and Remote Sensing, Vol. 43, pp 595-598.
- Bennett, R. S., (1969), "The intrinsic dimensionality of signal collections," IEEE Trans. Inform. Theory, Vol. IT-15, pp 517-525.
- Bharucha, B. H., (1969), "A posteriori distributions and detection theory," Information and Control, Vol. 14, pp 98-132.
- Bharucha, B. H. and Kadota, T. T., (1970), "On the representation of continuous parameter processes by a sequence of random variables," IEEE Trans. Inform. Theory, Vol. IT-16, pp 139-141.
- Billingsley, F. C., (1975), "Noise consideration in digital image processing hardware," in Picture Processing and Digital Filtering, T. J. Hagan, Ed., Springer-Verlag, New York.

- Brown, J. L., Jr., (1960), "Mean-square truncation error in series expansions of random functions," J. SIAM, Vol. 8, pp 28-32.
- Caprihan, A., and deFigueiredo, R. J. P., (1970), "On the extraction of pattern features from continuous measurements," IEEE Trans. Syst. Man., Cybern., Vol. SMC-6, pp 110-115.
- Chandrasekaran, B., (1971), "Independence of measurements and the mean recognition accuracy," IEEE Trans. Inform. Theory, Vol. IT-17, pp 452-456.
- Chandrasekaran, B., and Jain, A. K., (1974), "Quantization complexity and independent measurements," IEEE Trans. Computers, Vol. C-23, pp 102-106.
- Chandrasekaran, B. and Jain, A. K., (1975), "Independence, measurement complexity, and classification performance," IEEE Trans. Syst., Man. and Cybern., Vol. SMC-5, pp 240-244.
- Chen, C. H., (1978), "Finite sample considerations in statistical pattern recognition," Proceedings IEEE Computer Society Conf. on Pattern Recognition and Image Processing, Chicago, IL., pp 188-192.
- Chien, Y. T., and Fu, K. S., (1967), "On the generalized Karhunen-Loeve expansion," IEEE Trans. Inform. Theory, Vol. IT-13, pp 518-520.
- Coggeshall, M. E., and Hoffer, R. M., (1973), "Basic forest cover mapping using digitized remote sensor data and automatic data processing techniques," LARS Information Note 030573, Laboratory for Applications of Remote Sensing, Purdue Univ., West Lafayette, Indiana.
- Colwell, J. E., (1974), "Vegetation canopy reflectance," Remote Sensing of Environment, Vol. 3, pp 175-183.
- Courant, R. and Hilbert, D., (1953), Methods of Mathematical Physics, Vol. 1, Interscience Publishers, Inc., New York.
- Crane, R. B., Malila, W. A., and Richardson, W., (1972), "Suitability of the normal density assumption for processing multispectral scanner data," IEEE Trans. Geosci. Elect. Vol. GE-10, pp 158-165.
- Davenport, W. B. and Root, W. L., (1958), An Introduction to the Theory of Random Signals and Noise. McGraw-Hill, New York.

- Dempster, A. P., (1969), Elements of Continuous Multivariate Analysis, Addison-Wesley, Reading, Mass.
- Doob, J. L., (1953), Stochastic Processes, Wiley, New York.
- Duda, R. O., and Hart, P. E., (1973), Pattern Classification and Scene Analysis, Wiley, New York.
- Foley, D. H., (1972), "Considerations of sample and feature size," *IEEE Trans. Inform. Theory*, Vol. IT-18, pp 618-626.
- Foley, D. H. and Sammon, J. W., Jr., (1975), "An optimal set of discriminant vectors," *IEEE Trans. Computers*, Vol. C-24, pp 281-289.
- Fu, K. S., (1968), "Feature selection and feature ordering," Chapter 2 in Sequential Methods in Pattern Recognition and Machine Learning, Academic Press, New York.
- Fukunaga, K., (1972), Introduction to Statistical Pattern Recognition, Academic Press, New York.
- Fukunaga, K. and Koontz, W. L. G., (1970), "Representation of random processes using the finite Karhunen-Loeve expansion," *Information and Control*, Vol. 16, pp 85-101.
- Gates, D. M., Keegen, H. J., Schleter, J. C., and Weidner, V. R., (1965), "Spectral properties of plants," *Applied Optics*, Vol. 4, pp 11-22.
- Gikhman, I. I. and Skorokhod, A. V., (1969), Introduction to the Theory of Random Processes, W. B. Saunders Company, Philadelphia, PA.
- Grad, J. and Brebner, M. A., (1968), "Eigenvalues and eigenvectors of a real general matrix, Algorithm 343," *Association for Computing Machinery, Comm.*, Vol. 11, No. 12, pp 820-826.
- Grenander, U., (1950), "Stochastic processes and statistical inference," *Arkiv for Matematik*, Vol. 17, pp 195-277.
- Handbook of Geophysics, (1961), United States Air Force, Air Research and Development Command, Air Research Division, Geophysics Research Directorate, MacMillan Co., New York.
- Haralick, R. M., Shanugam, K., and Dinstein, I., (1973), "Textural features for image classification," *IEEE Trans. Syst. Man. and Cybern.*, Vol. SMC-3, pp 610-621.

Harnage, J. L., and Landgrebe, D. A., (1975), "Landsat-D Thematic Mapper Technical Working Group - Final Report," NASA/Johnson Space Center Report No. JSC-90797, Houston, Texas.

Hassel, P. G., Jr., et al., (1974), "Michigan experimental multispectral mapping system, a description of the M7 airborne sensor and its performance," Report No. 190900-10-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan.

Hoffer, R. M., (1978), "Biological and physical considerations in applying computer-aided analysis techniques to remote sensor data," in Remote Sensing: The Quantitative Approach, P. H. Swain and S. M. Davis, Eds., McGraw-Hill, New York.

Holmes, R. A., and MacDonald, R. B., (1969), "The physical basis of system design for remote sensing in agriculture," Proceedings IEEE, Vol. 57, pp 629-639.

Hughes, G. F., (1968), "On the mean accuracy of statistical pattern recognizers," IEEE Trans. Inform. Theory, Vol. IT-14, pp 55-63.

Hulstrom, R. L., (1974), "Spectral measurements and analyses of atmospheric effects on remote sensor data," Proceedings of the Society of Photo-Optical Inst. Engineers, Vol. 51, Scanners and Imagery Sys. for Earth Obs. San Diego, Ca., pp 90-100.

Kadota, T. T., (1964), "Optimum reception of binary Gaussian signals," Bell System Tech. J., Vol. 43, pp 2767-2810.

Kadota, T. T., (1965), "Optimum reception of binary sure and Gaussian signals," Bell Sys. Tech. J., Vol. 44, pp 1621-1658.

Kadota, T. T., and Shepp, L. A., (1967), "On the best finite set of linear observables for discriminating two Gaussian signals," IEEE Trans. Inform. Theory, Vol. IT-13, pp 278-289.

Kailath, T., (1967), "The divergence and Bhattacharyya distance measures in signal selection," IEEE Trans. Comm. Tech., Vol. COM-15, pp 52-60.

Kailath, T., (1971), "RKHS approach to detection and estimation problems - part I: deterministic signals in Gaussian Noise," IEEE Trans. Inform. Theory, Vol. IT-17, pp 530-549.

- Kailath, T., (1974), "A view of three decades of linear filtering theory," IEEE Trans. Inform. Theory, Vol. IT-20, pp 146-181.
- Kanal, L. and Chandrasekaran, B., (1971), "On dimensionality and sample size in statistical pattern classification," Pattern Recognition, Vol. 3, pp 225-234.
- Karhunen, K., (1947), "Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung," Amer. Acad. Sci. Fennicade, Ser. A, I, Vol. 37, pp 3-79, (Transl. RAND Corp., Santa Monica, Calif., Rept. T-131, Aug., 1960).
- Kettig, R. L., and Landgrebe, D. A., (1976), "Classification of multispectral image data by extraction and classification of homogeneous objects," IEEE Trans. Geosci. Electron., Vol. GE-4, pp 19-26.
- Kittler, J., and Young, P. C., (1973), "A new approach to feature selection based on the Karhunen-Loeve expansion," Pattern Recognition, Vol. 5, pp 335-352.
- Kolmogorov, A. N., and Fomin, S. V., (1957), Elements of the Theory of Functions and Functional Analysis, Vol. 1, Metric and Normed Spaces, Graylock Press, Albany, New York.
- Korb, C. L., (1969), "Physical considerations in the channel selection and design specification of an airborne multispectral scanner," Purdue Centennial Year Symposium on Information Processing, Purdue University, West Lafayette, Ind., pp 646-657.
- Kulkarni, A. V., (1978), "On the mean accuracy of hierarchical classifiers," IEEE Trans. Comp., Vol. C-27, pp 771-776.
- Lachenbruch, P. A., Sneering, C., and Revo, L. T., (1972), "Robustness of the linear and quadratic discriminant functions to certain types of non-normality," Communications in Statistics, Vol. 1, pp 39-56.
- Landgrebe, D. A., (1978), "Useful information from multispectral image data: another look," in Remote Sensing: the Quantitative Approach, P. H. Swain and S. M. Davis, Eds., McGraw-Hill, New York.
- Landgrebe, D. A., Biehl, L. L., and Simmons, W. R., (1977), "An empirical study of scanner system parameters," IEEE Trans. Geosci. Elect., Vol. GE-15, pp 120-130.

- Loeve, M., (1963), Probability Theory, Princeton, N.J., Van Nostrand.
- Lovitt, W. V., (1924), Linear Integral Equations, McGraw-Hill, New York.
- Malila, W. A., Gleason, J. M., and Cicone, R. C., (1977), "Multispectral system analysis through modeling and simulation," Proceedings 11th International Symp. on Remote Sensing, Ann Arbor, Michigan, pp 1319-1328.
- Marill, T. and Green, D. M., (1963), "On the effectiveness of receptors in recognition systems," IEEE Trans. Inform. Theory, Vol. IT-9, pp 11-17.
- Max, J., (1960), "Quantizing for minimum distortion," IRE Trans. Inform. Theory, Vol. IT-6, pp 7-12.
- May, G. A., and Petersen, G. W., (1975), "Spectral signature selection for mapping unvegetated soils," Remote Sensing of Environment, Vol. 4, pp 211-220.
- Middleton, D., (1960), Introduction to Statistical Communication Theory, McGraw-Hill, New York.
- Mobasser, B.G., McGillem, C. D., and Anuta, P. E., (1978), "A parametric multiclass Bayes error estimator for the multispectral scanner spatial model performance evaluation," LARS Technical Report 061578, Laboratory for Application of Remote Sensing, Purdue University, West Lafayette, Ind.
- Montgomery, O. L., Baumgardner, M. F., and Weismiller, R. A., (1976), "An investigation of the relationship between spectral reflectance and the chemical, physical, and genetic characteristics of soils," LARS Inform. Note 082776, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Ind.
- Moore, D. S., Whitsitt, S. J., and Landgrebe, D. A., (1976), "Variance comparisons for unbiased estimators of probability of correct classification," IEEE Trans. Inform. Theory, Vol. IT-22, pp 102-105.
- Morgera, S. D., and Cooper, D. B., (1977), "Structured estimation: sample size reduction for adaptive pattern classification," IEEE Trans. Inform. Theory, Vol. IT-23, pp 728-741.

- Niessner, H., (1972), "Remark on algorithm 343," Association for Computing Machinery, Comm., Vol. 15, p 466.
- Nilsson, N. J., (1965), Learning Machines, McGraw-Hill, New York.
- Papoulis, A., (1965), Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York.
- Phillips, T. L., (1973), Ed., LARSYS Version III User's Manual, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Ind.
- Ready, P. J., Wintz, P. A., Whitsitt, S. J., and Landgrebe, D. A., (1971), "Effects of compression and random noise on multispectral data," Proceedings of the 7th International Symposium on Remote Sensing of the Environment, Ann Arbor, Michigan.
- Riesz, F., and Sz.-Nagy, B., (1956), Functional Analysis, Frederick Unger, New York.
- Shannon, C. E., (1948), "A mathematical theory of communication," Bell Syst. Tech. J., Vol. 27, p 379-423.
- Simmons, W. R., Wilkinson, S. R., Zurney, W. C., and Kast, J. L., (1972), EXOSYS, a system of LARS Exotech Model 20C field spectroradiometer," LARS Program abstract 5000, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Ind.
- Swain, P. H., (1978), "Bayesian classification in a time-varying environment," LARS Technical Report 030178, Laboratory for Applications or Remote Sensing, Purdue University, West Lafayette, Ind.
- Tou, J. T., and Heydorn, R. P., (1967), "Some approaches to optimum feature extraction," in Computer and Information Sciences - II, J. Tou, ed., Academic Press, New York.
- Toussaint, G. T., (1974), "Bibliography on estimation of misclassification," IEEE Trans. on Inform. Theory, Vol. IT-20, pp 472-479.
- Tricomi, F. G., (1957), Integral Equations, Interscience Publishers, New York.
- VanTrees, H. L., (1968), Detection, Estimation, and Modulation Theory - Part I, Wiley, New York.

- VanTrees, H. L., (1977), Detection, Estimation, and Modulation Theory - Part III, Wiley, New York.
- Vincent, R. K., and Thomson, F., (1972), "Spectral compositional imaging of silicate rocks," J. Geophys. Res., Vol. 77, pp 2465-2472.
- Wacker, A. G., and Landgrebe, D. A., (1971), "The minimum distance approach to classification," LARS Information Note 100771, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Ind.
- Watanabe, S., (1965), "Karhunen-Loeve expansion and factor analysis, theoretical remarks and applications," Trans. Fourth Prague Conf. Inform. Theory, Statist. Decision Functions, and Random Processes, Prague, pp 635-660.
- Weston, V. H., (1977), Class notes.
- Whitsitt, S. J., and Landgrebe, D. A., (1977), "Error estimation and separability measures in feature selection for multiclass pattern recognition," LARS Information Note 082377, Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind.
- Wiersma, D. J., and Landgrebe, D. A., (1976). "The use of spatial characteristics for the improvement of multispectral classification of remotely sensed data," Proceedings of the Symposium of Machine Processing of Remotely Sensed Data, West Lafayette, Ind.
- Wilkinson, J. H., (1965), The Algebraic Eigen-Value Problem, Oxford University Press, London.
- Wiswell, E. R., and Cooper, G. R., (1978), "Analytical techniques for the study of some parameters of multispectral scanner systems for remote sensing," LARS Technical Report No. 061778, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Ind.
- Wong, E., (1971), "Recent progress in stochastic processes - a survey," IEEE Trans. Inform. Theory, Vol. IT-19, pp 262-275.
- Zaitzeff, E. M., Korb, C. L., and Wilson, C. L., (1971), "MSDS: an experimental 24-channel multispectral scanner system," IEEE Trans. Geosci. Electron., Vol. GE-9, pp 114-120.

Appendix A

A Stratified Posterior Classification Performance
Estimator

Appendix A. A Stratified Posterior Classification Performance Estimator

A method was needed to estimate the classification performance for a maximum likelihood Gaussian classifier from a set of multiclass multivariate statistics. A Monte Carlo method may be used to evaluate the probability of correct classification integral. The method used here is based on the stratified posterior estimator developed by Whitsitt and Landgrebe (1977) (see also Moore, Whitsitt and Landgrebe, 1976).

Let X be an observation from one of M classes C_i , $i = 1, 2, 3, \dots, M$, with a priori probabilities P_i . The maximum likelihood decision rule can be stated as follows: Assign X to the class C_k if

$$P(C_k | X) = \max_i \{P(C_i | X)\}$$

where $P(C_i | X)$ is the conditional posterior probability for class C_i given the observation X . This rule partitions the observation space Ω into subregions $\Omega_1, \Omega_2, \dots, \Omega_M$, corresponding to the classes C_1, C_2, \dots, C_M , respectively. Define the indicator function

$$I_i(x) = \begin{cases} 1, & x \in \Omega_i \\ 0, & x \notin \Omega_i \end{cases}$$

The probability of correct classification integral is given by

$$P_c = \int_{\Omega} \sum_{i=1}^M P_i I_i(x) p(x|C_i) dx \quad (\text{A.1})$$

It is desirable to evaluate the probability of correct classification for each class as well as the overall probability. The performance probability for the i th class is

$$P_{C_i} = \int_{\Omega} I_i(x) p(x|C_i) dx \quad (\text{A.2})$$

This integral is equivalent to the integral of the conditional density function whose support is Ω_i . The overall performance, then, is

$$P_c = \sum_{i=1}^M P_i P_{C_i} \quad (\text{A.3})$$

From Bayes' rule

$$p(x|C_i) = \frac{P(C_i|x) p(x)}{P_i}$$

hence,

$$P_{C_i} = \int_{\Omega} I_i(x) \frac{P(C_i|x)}{P_i} p(x) dx$$

where $p(x)$ is the mixture density

$$p(x) = \sum_{j=1}^M P_j p(x|C_j)$$

Therefore,

$$P_{C_i} = \sum_{j=1}^M \frac{P_j}{P_i} \int_{\Omega} I_i(x) P(C_i|x) p(x|C_j) dx \quad (\text{A.4})$$

Define

$$Q_i(x) = I_i(x) P(C_i|x)$$

Then,

$$\int_{\Omega} Q_i(x) p(x|C_j) dx$$

is the conditional expected value of $Q_i(x)$ given that x comes from the class C_j . The estimate of this expected value is

$$\hat{Q}_i(x|C_j) = \frac{1}{N_j} \sum_{k=1}^{N_j} Q_i(x_k)$$

Therefore the estimate of the probability of correctly classifying observations belonging to class i is the unbiased estimate

$$\hat{P}_{C_i} = \sum_{j=1}^M \frac{P_j}{P_i} \frac{1}{N_j} \sum_{k=1}^{N_j} Q_i(x_k) \quad (\text{A.5})$$

(Whitsitt and Landgrebe, 1977). The stratification refers to the sampling scheme used to obtain the estimate. A stratified sampling scheme takes advantage of knowledge of the classes to which the samples belong, whereas random sampling does not use the class assignment information. N_i multivariate sample vectors are generated for class i from the given statistics. The maximum likelihood rule is used to determine the decision regions. Equation A.5 is evaluated for the N_i sample vectors from each class and the total probability of correct classification is computed from equation A.3.

From equation A.4 the term that must be evaluated is

$$P(C_i | x) = \frac{P_i p(x|C_i)}{\sum_k P_k p(x|C_k)} \quad (A.6)$$

To evaluate this probability compute $P_k p(x|C_k)$ for each class. Choose the largest value of $P_k p(x|C_k)$ which by the maximum likelihood decision rule will be $P_i p(x|C_i)$. The posterior probability is given by equation A.6.

The analysis so far can be applied to any probability measure. The remaining discussion will deal with the parametric case where the probability measure is Gaussian. That is

$$p(x|C_k) = \frac{1}{(2\pi)^{\frac{1}{2}} |K_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x-m_k)^T K_k^{-1} (x-m_k) \right\} \quad (A.7)$$

where m_k and K_k are the L-dimensional mean vector and covariance matrix respectively for class k.

To reduce the number of computations required the linear transformation

$$x = \phi_i \Gamma_i^{1/2} y + m_i \quad (\text{A.8})$$

is introduced where ϕ_i is the matrix of eigenvectors required to diagonalize the covariance matrix of class i, Γ_i is the diagonal matrix of eigenvalues and m_i is the mean vector for class i. Substituting equation A.8 into A.7,

$$p(x|C_k) = (2\pi)^{-L/2} |K_k|^{-L/2} \exp \left[-\frac{1}{2} [y^T \Gamma_i^{1/2} \phi_i^T K_k^{-1} \phi_i \Gamma_i^{1/2} y + 2y^T \Gamma_i^{1/2} \phi_i^T K_k^{-1} (m_j - m_k) + (m_j - m_k)^T K_k^{-1} (m_j - m_k)] \right]$$

In this form it is not necessary to perform the intermediate computational step of transforming the generated random vectors to get the desired statistics. It is only necessary to generate M sets of random vectors y with expected value the zero vector and covariance matrix I and to use them in expression A.8.

The random vectors are generated using a pseudo-random sequence of uniformly distributed random numbers.

The subroutine RANDU from the IBM 360 subroutine package generates the random numbers and transforms them by the inverse cumulative-distribution-function method to obtain zero mean, unit variance, independent Gaussian random numbers. These random numbers are used to fill the elements of the vector y . The y vectors have an expected value equal to the zero vector and a covariance matrix equal to the identity matrix.

Estimator Evaluation

Since $\frac{1}{N_j} \sum_{k=1}^{N_j} Q_i(x_k)$ is an unbiased estimate of

$\int_{\Omega} Q_i(x) p(x|C_i) dx$, the estimator

$$\hat{P}_C = \sum_{i=1}^M P_i \sum_{j=1}^M \frac{P_j}{P_i} \left(\frac{1}{N_j} \sum_{k=1}^{N_j} Q_i(x_k) \right) \quad (\text{A.9})$$

is an unbiased estimate of the probability of correct classification (Moore, Whitsitt, and Landgrebe, 1976).

The variance of the estimator can be shown to be smaller than the variance for a count estimator using stratified sampling (Moore, Whitsitt, and Landgrebe, 1976).

The variance for the stratified count estimator is

$$\sum_{i=1}^M \frac{P_i^2}{N_i} (P_{C_i} - P_{C_i}^2) \quad (\text{A.10})$$

If the probabilities of correct classification for each class are known then the variance of the stratified count estimator can be evaluated and used as a bound on the stratified posterior estimator.

A FORTRAN program SPESTM was written which accepts the mean vectors and covariance matrices for up to ten classes and up to 10 dimensions. These statistics are used to generate random vectors and estimate the classification performance for the classes specified by the distributions.

To test the method and the program a three-class problem was constructed. The mean vectors for the classes were

$$M_1 = [-1, -1, \dots, -1]^T$$

$$M_2 = [0, 0, \dots, 0]^T$$

$$M_3 = [1, 1, \dots, 1]^T$$

The covariance for each class was the identity matrix. The number of random vectors generated for each class was 1000. The exact classification accuracy as a function of the dimensionality can be evaluated for this case.

$$P_{c1} = 1 - \operatorname{erfc}(\sqrt{L/2})$$

$$P_{c2} = 1 - 2 \operatorname{erfc}(\sqrt{L/2})$$

$$P_{c3} = 1 - \operatorname{erfc}(\sqrt{L/2})$$

$$P_c = 1 - 4/3 \operatorname{erfc}(\sqrt{L/2})$$

$$\text{where erfc } (a) = \int_a^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

and L is the dimensionality. Table A1 contains the results of evaluating the class conditional performance and the overall performance for from one to ten dimensions.

A bound on the standard deviation of the estimator can be computed by calculating the standard deviation for the stratified count estimator. Table A2 lists the standard deviations for from one to ten dimensions for this experiment.

The actual variance was estimated by repeating the classification performance estimation 20 times using different starting points in the random number generator. The maximum difference between the estimate and the true value E_{\max} and the standard deviation from the true value were computed for from one to ten dimensions as shown in Table A3.

Based on the results presented in the tables, differences in estimation of overall performance of less than .005 ($\frac{1}{2}$ of 1%) will not be considered significant. The performance of the algorithm is demonstrated to be quite adequate for its intended use. The class conditional estimates are less reliable but are sufficient to observe trends in the performance due to the individual classes. The running time for this algorithm is quite reasonable, even for ten dimensions.

Table A1. Test of error estimator.

	P_{c_1}	P_{c_2}	P_{c_3}	P_c	\hat{P}_{c_1}	\hat{P}_{c_2}	\hat{P}_{c_3}	\hat{P}_c
1	0.6915	0.3829	0.6915	0.5886	0.6859	0.3793	0.7001	0.5884
2	0.7602	0.5205	0.7602	0.6803	0.7671	0.5116	0.7700	0.6829
3	0.8068	0.6135	0.8068	0.7423	0.8037	0.6202	0.8081	0.7440
4	0.8413	0.6827	0.8413	0.7885	0.4283	0.6852	0.8550	0.7895
5	0.8682	0.7364	0.8682	0.8243	0.8642	0.7425	0.8703	0.8256
6	0.8897	0.7793	0.8897	0.8529	0.8767	0.7939	0.8787	0.8498
7	0.9071	0.8141	0.9071	0.8761	0.8993	0.8242	0.9065	0.8766
8	0.9214	0.8427	0.9214	0.8951	0.9129	0.8472	0.9240	0.8947
9	0.9332	0.8664	0.9332	0.9109	0.9193	0.8809	0.9360	0.9120
10	0.9431	0.8862	0.9431	0.9241	0.9209	0.9012	0.9481	0.9234

Table A2. Theoretical bound of standard deviation for different dimensions.

L	σ
1	.00858
2	.00826
3	.00781
4	.00733
5	.00686
6	.00640
7	.00596
8	.00555
9	.00517
10	.00481

Table A3. Experimental standard deviation of estimates.

Dimensions	P_{c1}	P_{c2}	P_{c3}	P_c	
1	.016	.010	.017	.003	$\sigma_{E_{max}}$
	.033	.019	.049	.005	
2	.018	.010	.014	.002	$\sigma_{E_{max}}$
	.036	.018	.027	.005	
3	.016	.017	.017	.003	$\sigma_{E_{max}}$
	.046	.031	.055	.007	
4	.011	.016	.015	.003	$\sigma_{E_{max}}$
	.025	.029	.029	.005	
5	.015	.014	.012	.002	$\sigma_{E_{max}}$
	.031	.033	.026	.004	
6	.014	.014	.010	.003	$\sigma_{E_{max}}$
	.026	.023	.022	.006	
7	.009	.016	.012	.003	$\sigma_{E_{max}}$
	.027	.033	.027	.005	
8	.013	.013	.012	.003	$\sigma_{E_{max}}$
	.025	.036	.023	.006	
9	.013	.014	.012	.002	$\sigma_{E_{max}}$
	.026	.031	.021	.004	
10	.009	.012	.009	.002	$\sigma_{E_{max}}$
	.016	.024	.019	.005	

σ = standard deviation

E_{max} = maximum difference between estimate and true value
over 20 trials

Appendix B
Data Base Description

Appendix B. Data Base Description

The data sets used in this research are described and sufficient information to access this data is provided.

Data set number 1

Location: Williams County, North Dakota
Collection date: 8 May 1977

<u>CLASS</u>	<u>SAMPLE FUNCTIONS/CLASS</u>
SPRING WHEAT	664
SUMMER FALLOW	437
PASTURE	164

Field measurements library tape number: 4896
Comments: Wheat is just emerging (plant height ~8 cm).

Data set number 2

Location: Williams County, North Dakota
Collection date: 29 June 1977

<u>CLASS</u>	<u>SAMPLE FUNCTIONS/CLASS</u>
SPRING WHEAT	787
SUMMER FALLOW	291
PASTURE	161

Field measurements library tape number: 4897
Comments: Wheat is green and at full height. The mixture is below average.

Data set number 3

Location: Williams County, North Dakota
 Collection date: 4 August 1977

<u>CLASS</u>	<u>SAMPLE FUNCTIONS/CLASS</u>
SPRING WHEAT	931
SUMMER FALLOW	330
PASTURE	183

Field measurements library tape number: 4898
 Comments: Wheat is mature. In a few fields the wheat is harvested.

Data set number 4

Location: Finney County, Kansas
 Collection date: 28 September 1976

<u>CLASS</u>	<u>SAMPLE FUNCTIONS/CLASS</u>
WINTER WHEAT	141
SUMMER FALLOW	414
GRAIN SORGHUM	277

Field measurements library tape number: 4292
 Comments: Wheat is emergent while other crops are at mature stages.

Data set number 5

Location: Finney County, Kansas
 Collection date: 3 May 1977

<u>CLASS</u>	<u>SAMPLE FUNCTIONS/CLASS</u>
WINTER WHEAT	658
SUMMER FALLOW	211
OTHER CROPS	652

Field measurements library tape number: 4295
 Comments: Wheat is near full canopy and green. Other crops are emergent.

Data set number 6

Location: Finney County, Kansas
 Collection date: 26 June 1977

<u>CLASS</u>	<u>SAMPLE FUNCTIONS/CLASS</u>
WINTER WHEAT	677
SUMMER FALLOW	643
GRAIN SORGHUM	157

Field measurements library tape number: 4296
 Comments: Wheat is mature and ready for harvest.

Accessing the Data

A software package called EXOSYS (Simmons et al, 1972) was developed at LARS for handling field measurement data. Sampled spectral response functions are calibrated and stored on magnetic tape along with pertinent identification information. EXOSYS, also, provides access to the field measurement data through three processors - IDLIST, GSPEC, and DSEL. The IDLIST processor scans the tape and lists information from the identification record as required. One can use this information to select appropriate runs to represent the ensemble. The GSPEC processor creates a punched deck consisting of the 100 sampled values of the spectral response functions for all of the desired runs. The DSEL processor simulates rectangular spectral channels and uses data from the tape to evaluate the response in each channel for the ensemble.

The GSPEC processor is used to assemble the data sets. It is required to specify the library tape number, the

cover type or class, and the collection date to put together all of the sample functions on a particular date for a single class (see Figure B1). The sample functions are collected by class to facilitate the estimation of class dependent statistics. A deck of cards containing the sample functions for all of the classes is read by the routing SPRDCT and stored on disk in the format as shown in Figure B2. All programs which access the data sets expect the data to be in this format. Supplementary information such as the number of samples in each class and the name of each class are added from the terminal during the execution of SPRDCT (Table B1). Processing of the ensemble is accomplished with the data stored on the desk file; however, the data file may be stored on magnetic tape between processing sessions.

```
$TAPE 4896  
$GSPEC  
  GRAPH SPEC (SPRING WHEAT), DACO (770508)  
  LIST NOLIST  
  OPTIONS PUNCH, NOGRAPH  
$END  
$EXIT
```

Figure B1. Sample program for assembling sample functions using the GSPEC processor in EXOSYS.

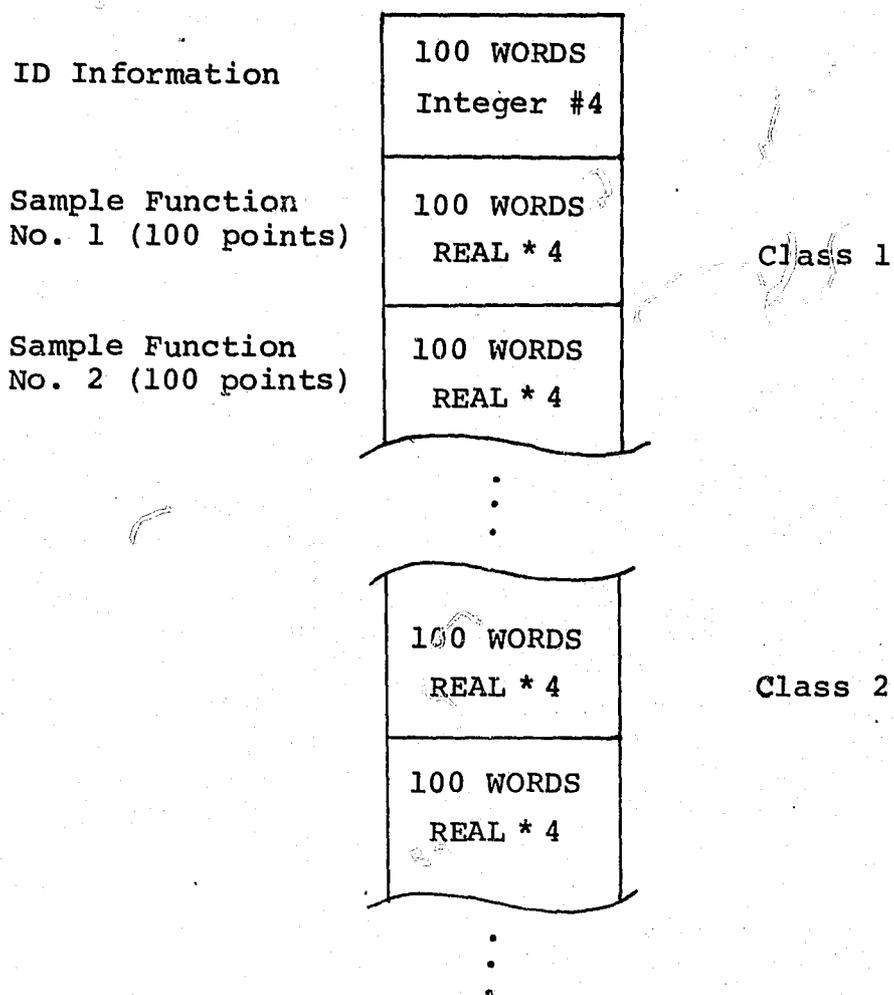


Figure B2. Spectral parameter design system data storage format.

Table B1. ID information locations in data storage format.

<u>WORD</u>	<u>ITEM</u>
1-15	Date data set was assembled
16	Experiment number
17	Number of classes
18	Number of sample points (=100)
21	Number of samples for class 1
22	Number of samples for class 2
23	Number of samples for class 3
24	Number of samples for class 4
25	Number of samples for class 5
26	Number of samples for class 6
27	Number of samples for class 7
30-39	Label for class 1
40-49	Label for class 2
50-59	Label for class 3
60-69	Label for class 4
70-79	Label for class 5
80-89	Label for class 6
90-99	Label for class 7

Appendix C

Correlation Between Sample Measurements
on the Spectrum

Appendix C. Correlation Between Sample Measurements on the Spectrum

Graphs of the correlation coefficients as a function of wavelength which were used to help locate spectral band edges are presented. Traditional methods of spectral analysis are not appropriate for this analysis since the calculation of the spectral density to obtain a sampling bandwidth assumes that the stochastic process is stationary and that the sampling rate will be uniform over the entire interval Λ . It is believed that it is necessary to sample some parts of the spectrum more frequently than others; hence, the correlation measure proposed here is used.

The measure of the correlation between two adjacent spectral samples u_i and u_{i+1} on the interval Λ is the correlation coefficient given by

$$\rho_{i,i+1} = \frac{E\{(u_i - \bar{u}_i)(u_{i+1} - \bar{u}_{i+1})\}}{\left[E\{(u_i - \bar{u}_i)^2\} E\{(u_{i+1} - \bar{u}_{i+1})^2\} \right]^{1/2}} \quad (C.1)$$

The correlation coefficient can be computed using the eigenvalues and eigenvectors. The matrix equation which was solved is given by

$$\phi \Gamma = K W \phi$$

or

$$K W = \phi \Gamma \phi^T \quad (C.2)$$

The covariance matrix K consists of the elements k_{ij} , where the k_{ij} are the correlations between the i^{th} and j^{th} spectral samples. Assuming W is the identity matrix, the correlation coefficient is equal to k_{ij} normalized by dividing by the square root of the product of the respective variances. Using equation C.2, two adjacent spectral channels have the correlation coefficient

$$\rho_{i,i+1} = \frac{\sum_k \phi_{ik} \phi_{i+1,k} \gamma_k}{\left[\left(\sum_k \phi_{ik} \phi_{ik} \gamma_k \right) \left(\sum_k \phi_{i+1,k} \phi_{i+1,k} \gamma_k \right) \right]^{1/2}} \quad (C.3)$$

where ϕ_{ik} is the i^{th} element of the k^{th} eigenvector.

The second weight function from Figure 4.6 is used. Since this weight function is unity everywhere on Λ except over the water absorption bands, equation C.3 is valid everywhere except over the absorption bands and on the edges of the absorption bands.

The graphs of the correlation coefficients as a function of wavelength for each of the six data sets are presented in Figures C.1 through C.6.

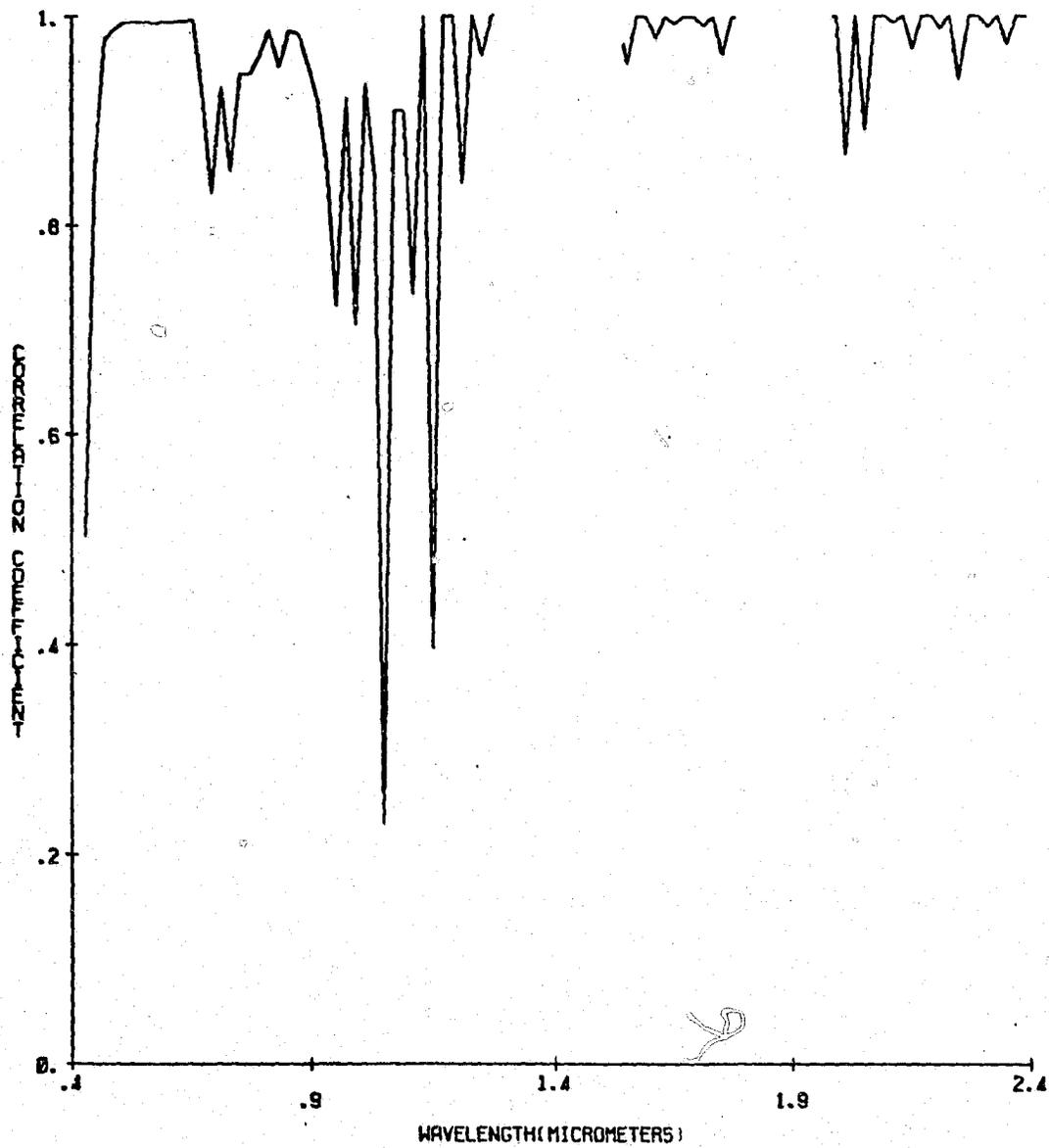


Figure C1. Correlation coefficients vs wavelength for Williams County, May 8, 1977.

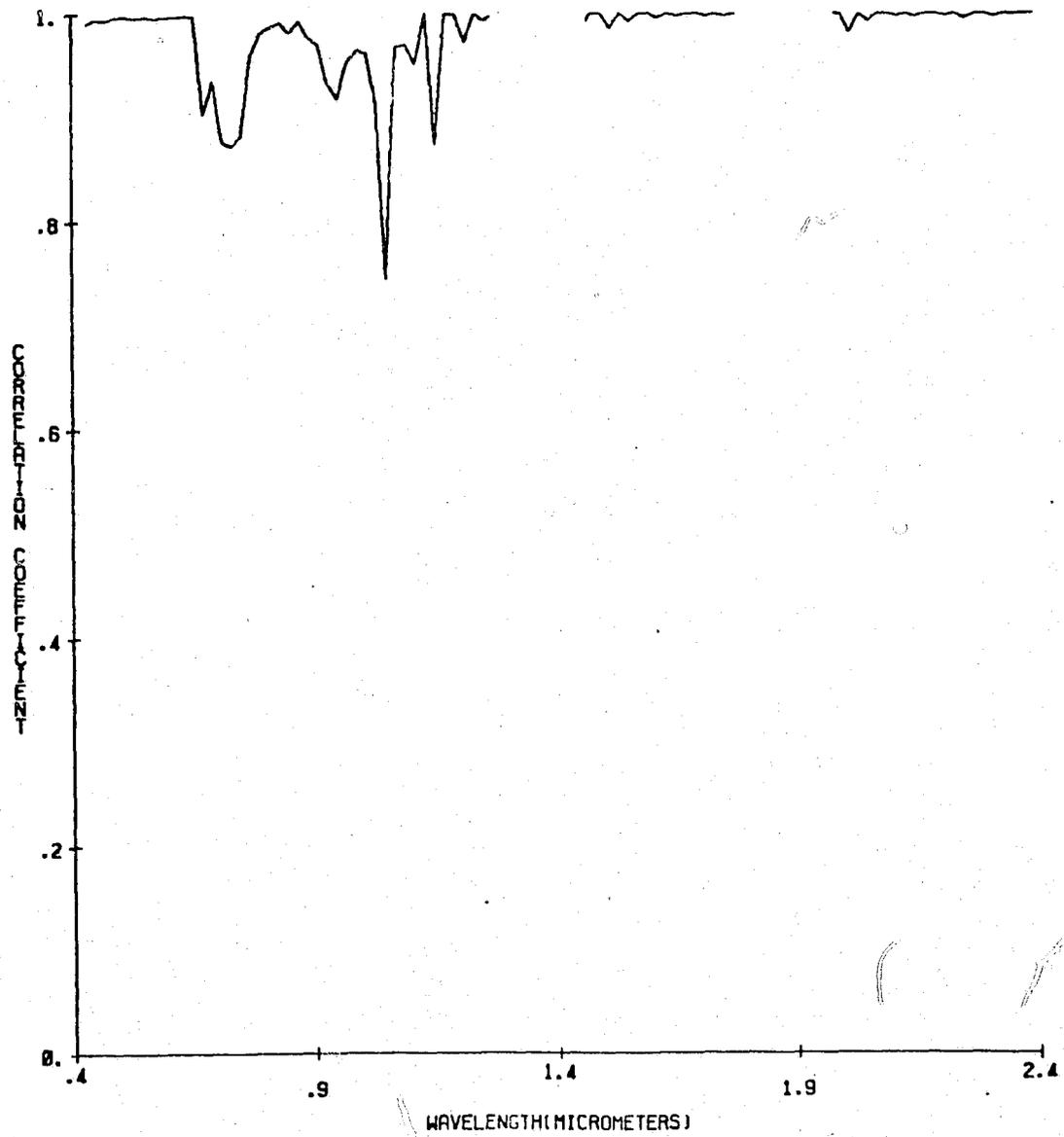


Figure C2. Correlation coefficients vs wavelength for William County, June 29, 1977.

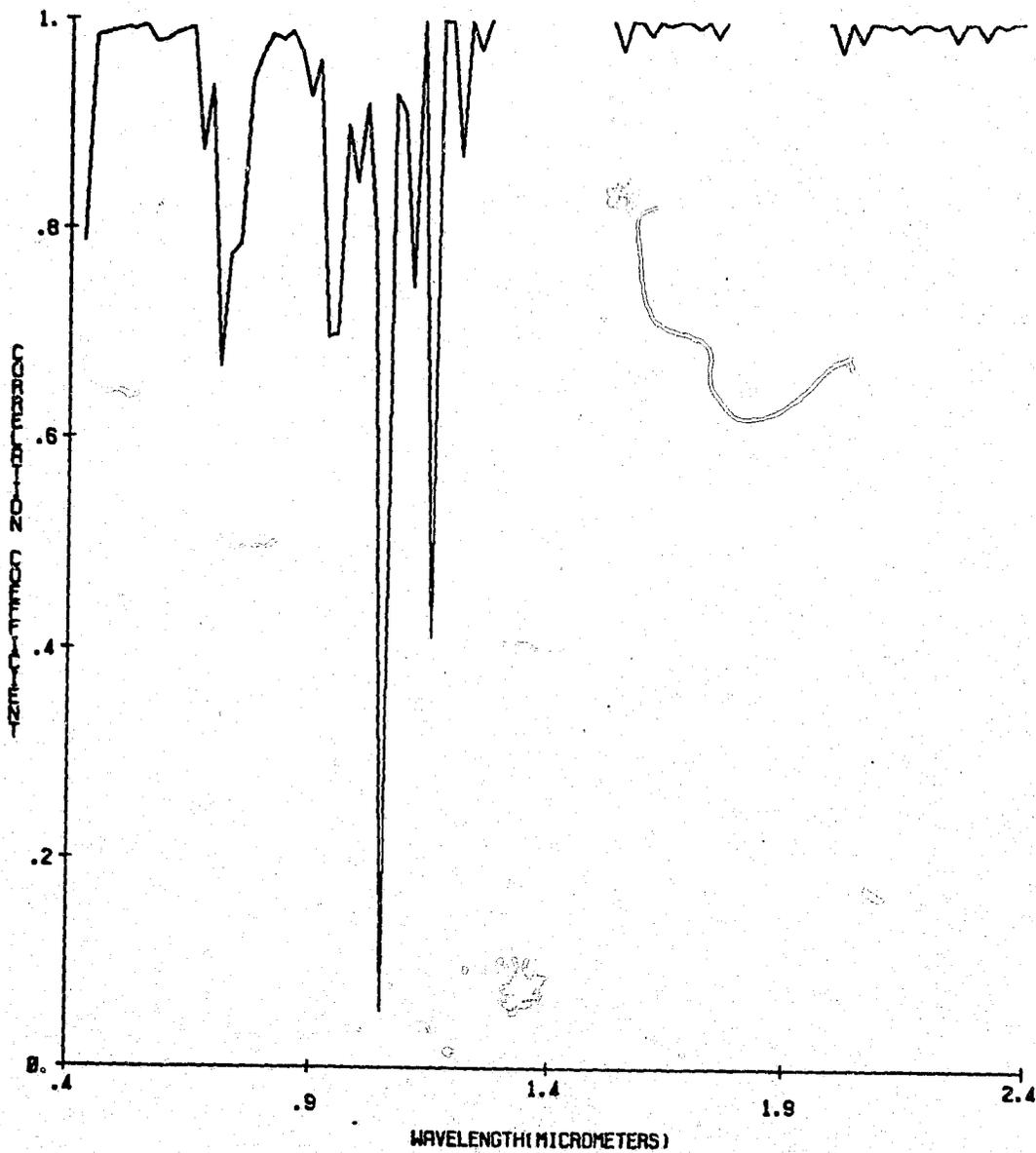


Figure C3. Correlation coefficients vs wavelength for Williams County, August 4, 1977.

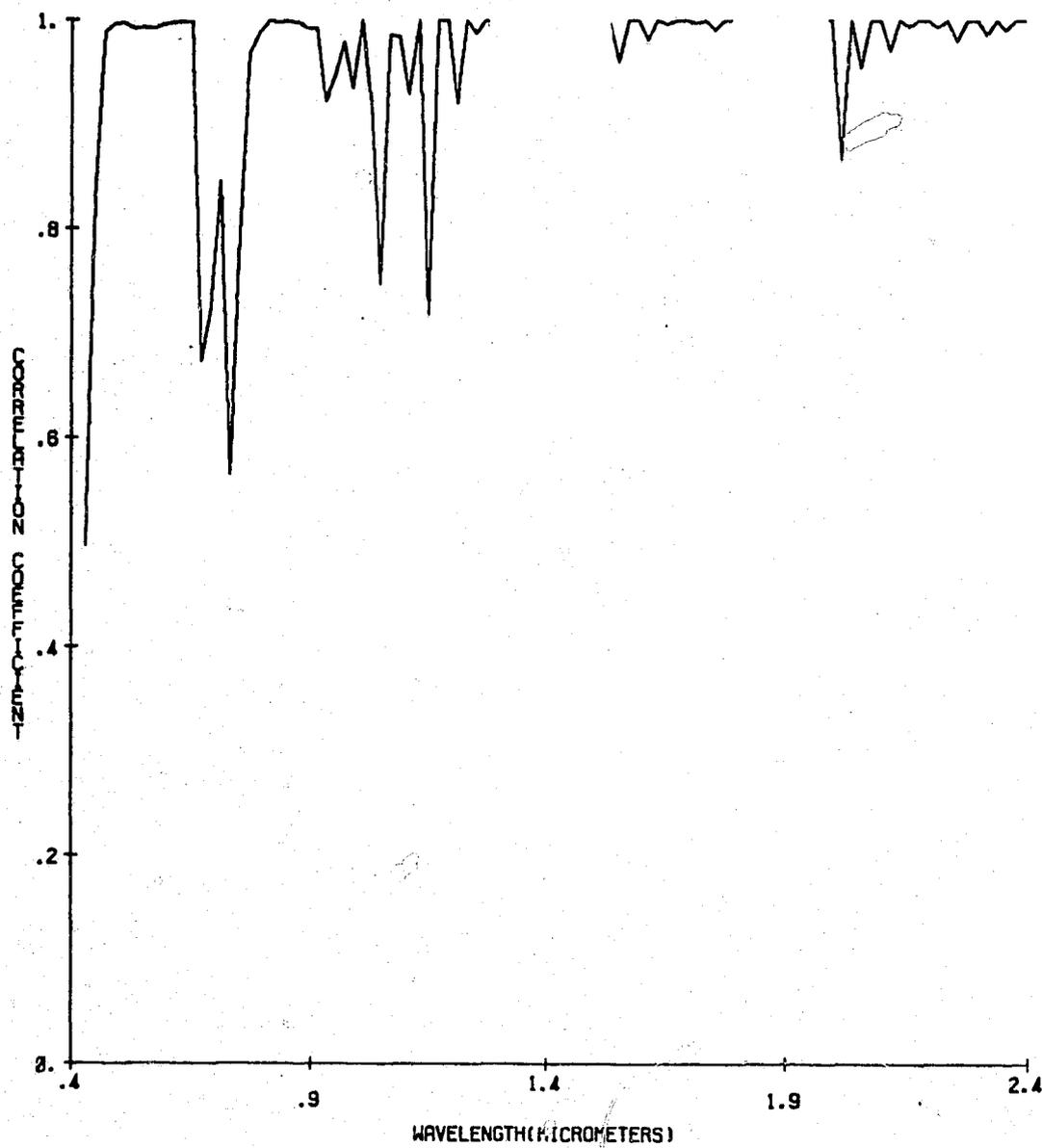


Figure C4. Correlation coefficients vs wavelength for Finney County, September 28, 1976.

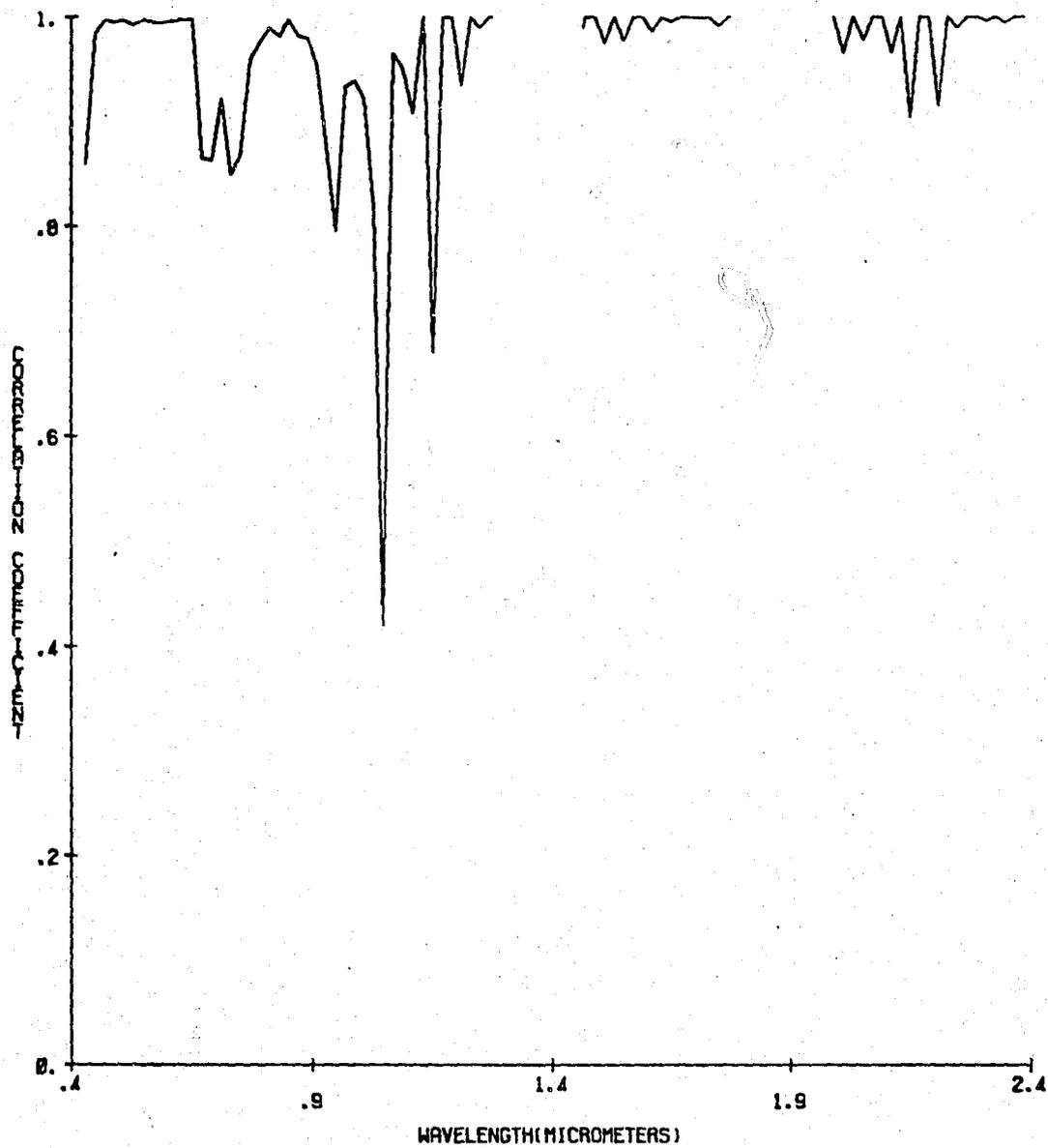


Figure C5. Correlation coefficients vs wavelength for Finney County, May 3, 1977.

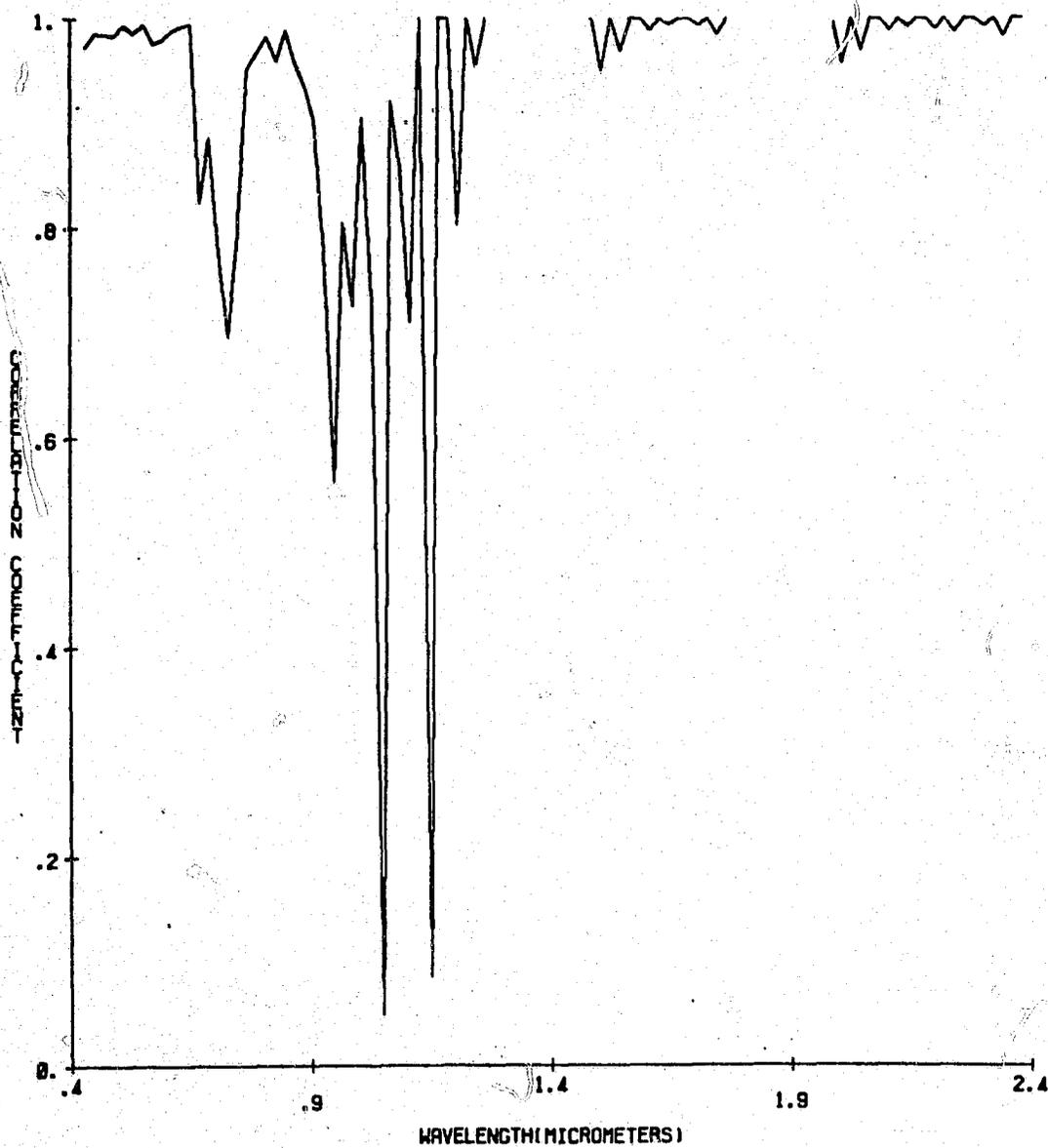


Figure C6. Correlation coefficients vs wavelength for Finney County, June 26, 1977.

Appendix D
Computer Program Listings

 SPRDCT

PURPOSE

EXOSYS DATA IN PUNCHED FORMAT IS READ AND STORED ON TAPE

REVISED

3 JULY, 1978

COMMON ID(100)

INTEGER*4 IN(5), DAY(3), TIME(3), NOGRPS, GROUPS(5)

INTEGER*4 ISAM, SINT, DATE(15), INFO(10,7)

REAL*4 WBCOEF(2,5), DATA(2500), X(100)

EQUIVALENCE(DATE(1), ID(1)), (INFO(1,1), ID(30)), (INFO(1,2), ID(40)),
 (INFO(1,3), ID(50)), (INFO(1,4), ID(60)), (INFO(1,5), ID(70)), (INFO(1,6),
 ID(80)), (INFO(1,7), ID(90))

REWIND 11

NT = 100

ID INFORMATION

WRITE(16,10)

FORMAT(5X, 'TYPE IN DATE', /, 1X, 15('/'))

READ(15,15) DATE

FORMAT(15A1)

EXP. NO., NUMBER OF CLASSES, AND NUMBER OF DIMENSIONS

WRITE(16,20)

FORMAT(5X, 'TYPE EXP. NO., CLASSES, AND DIMENSIONS', /, ' /// // ///')

READ(15,25) ID(16), ID(17), ID(18)

FORMAT(13, 2X, 12, 3X, 13)

EXPERIMENT INFORMATION

NCLS = ID(17)

DO 35 I=1, NCLS

WRITE(16,30) I

FORMAT(5X, 'TYPE CLASS INFO AND NO. SAMPLES FOR CLASS', I1, /, 1X, 10(
 ' / '), 4X, ' / / /')

READ(15,40) (INFO(L,I), L=1,10), ID(20+I)

FORMAT(10A1, 4X, I3)

WRITE(11) ID

CALL SPLBL

READ SAMPLE FUNCTIONS FROM EACH CLASS

DO 500 K=1, NCLS

WRITE(6,80)

FORMAT(5(/), 10X, 'SAMPLE FUNCTIONS')

NF = ID(20+K)

DO 200 JJ=1, NF

READ(5,1000) DAY, TIME, IN, NOGRPS, NOSAMS

FORMAT(3A4, 2X, 3(I2, 1X), /, 45X, 5A4, /, 20X, I1, 8X, I3, /)

READ(5,1100) (DATA(I), I=1, NOSAMS)

FORMAT(20A4)

WRITE(6,150) IN

FORMAT(10X, 20A4)

DO 160 I=1, NT

X(I) = DATA(I+1)

WRITE(11) X

CONTINUE

CONTINUE

END FILE 11

STOP

END


```

C
C COMPUTE COVARIANCE
C
10 WRITE(16,10)
   FORMAT(5X,'COVARIANCE BEING ESTIMATED (SPOPTN)')
   NCLS = ID(17)
   NFT = 0
   DO 20 I=1,NCLS
20   NFT = NFT + ID(20+I)
   CON = DFLOAT(NFT)/DFLOAT(NFT-1)
   DO 30 I=1,N
30   AM(I) = 0.0
   DO 35 I=1,NCT
35   COV(I) = 0.0
   DO 65 IJ=1,NFT
   READ(2) X
   IN = 0
   DO 50 I=1,N
   AM(I) = AM(I) + X(I)/DFLOAT(NFT)
   DO 50 J=1,I
   IN = IN + 1
   COV(IN) = COV(IN) + X(I)*X(J)/DFLOAT(NFT-1)
50 CONTINUE
65 CONTINUE
   IN = 0
   DO 60 I=1,N
   DO 60 J=1,I
   IN = IN + 1
   COV(IN) = COV(IN) - CON*AM(I)*AM(J)

60 CONTINUE
C
C WEIGHTING FUNCTION
C
   IN = 0
   DO 210 I=1,N
   DO 210 J=1,I
   IN = IN + 1
   ACOV(I,J) = COV(IN)
   ACOV(J,I) = COV(IN)
210 CONTINUE
C
   CALL SPWGT2(W)
C
   DO 250 I=1,N
   DO 250 J=1,N
250   ACOV(I,J) = ACOV(I,J)*W(J)
C
C COMPUTE TRACE OF COVARIANCE
C
   SUM = 0.0
   DO 80 I=1,N
80   SUM = SUM + ACOV(I,I)
C
C COMPUTE EIGENVALUES AND EIGENVECTORS
C
75 WRITE(16,75)
   FORMAT(5X,'EIGENVALUES AND EIGENVECTORS (EIGENP)')
   NM = N
   T = 56.
   CALL EIGENP(N,NM,ACOV,T,GAM,EVI,PHI,VECI,INDIC,W)
   CALL EISORT(N,GAM,PHI)

```

```

C
C
C PRINT EIGENVALUES AND MEAN-SQUARE ERROR
      C0 = FLOAT(NFT)/(FLOAT(NFT-1)*FLOAT(NFT-1))
      C1 = FLOAT(4*NFT-1)/(FLOAT(NFT-1)*FLOAT(NFT-1))
      WRITE(6,110)
110  FORMAT(5(/),5X,'N',5X,'EIGENVALUE',5X,'VAR(GAM)',5X,'VAR(PHI)',5X,
      •'MEAN-SQUARE ERROR')
      DO 150 I=1,30
      VARP = 0.0
      DO 120 J=1,100
      IF(J.EQ. I) GO TO '5
      VARP = VARP + C0*GAM(I)*GAM(J)/(GAM(I) - GAM(J))**2
115  CONTINUE
120  CONTINUE
      VARG = C1*GAM(I)*GAM(I)
      SUM = SUM - GAM(I)
      WRITE(6,145) I,GAM(I),VARG,VARP,SUM
145  FORMAT(4X,12,4X,F10.4,4X,F10.4,2X,F10.4,2X,F14.6)
150  CONTINUE
      DO 155 J=1,N
      DO 155 I=1,N
155  PHIP(I,J) = PHI(I,J)
      DO 180 J=1,20
      WRITE(7,160) (PHIP(I,J),I=1,N)
160  FORMAT(20A4)
180  CONTINUE
      STOP
      END

```

```

C -----
C
C SUBROUTINE FOR SORTING EIGENVALUES INTO DECENDING ORDER.
C -----
C
C SUBROUTINE EISORT(N,EVR,VECR)
C REAL*8 STORE,EVR(N),STOVEC(100),VECR(N,N)
C
C DO 85 I=1,N
C   DO 85 J=1,N
C     IF(EVR(I) - EVR(J))85,85,70
70  STORE = EVR(I)
C     EVR(I) = EVR(J)
C     EVR(J) = STORE
C   DO 80 K=1,N
C     STOVEC(K) = VECR(K,I)
C     VECR(K,I) = VECR(K,J)
C     VECR(K,J) = STOVEC(K)
C   CONTINUE
80  CONTINUE
85  RETURN
      END

```

```

C .....
C SUBROUTINE SPLBL PROVIDES A LABEL TO DESCRIBE THE DATA USED IN THE
C EXPERIMENT. INFORMATION INCLUDED IN THE LABEL CONSISTS OF THE DATE,
C EXPERIMENT NUMBER, NUMBER OF CLASSES, NUMBER OF SAMPLES IN EACH CLASS
C AND INFORMATION ON EACH CLASS.
C
C 7 JULY, 1977
C .....

```

```

SUBROUTINE SPLBL
COMMON ID(100)
INTEGER*4 DATE(15), INFO(10,5)
EQUIVALENCE(ID(1), DATE(1)), (INFO(1,1), ID(30)), (INFO(1,2), ID(40)),
• (INFO(1,3), ID(50)), (INFO(1,4), ID(60)), (INFO(1,5), ID(70))
WRITE(6,20)
WRITE(6,30)
WRITE(6,40) DATE
WRITE(6,50) ID(16)
WRITE(6,55) ID(17)
NCLS = ID(17)
DO 10 J=1,NCLS
WRITE(6,60) (INFO(I,J), I=1,10)
WRITE(6,80) ID(20+J)
10 CONTINUE
20 FORMAT(1H1,////15X, 'LABORATORY FOR APPLICATIONS OF REMOTE SENSING'
1)
30 FORMAT(29X, 'PURDUE UNIVERSITY')
40 FORMAT(12X, 'SAMPLE FUNCTION INFORMATION', 11X, 15A1/)
50 FORMAT(10X, 'EXP. NO.', 27('.'), 13)
55 FORMAT(10X, 'NUMBER OF CLASSES', 18('.'), 12)
60 FORMAT(10X, 'CLASS', 30('.'), 20A1)
80 FORMAT(10X, 'NUMBER OF SAMPLE FUNCTIONS', 9('.'), 13)
RETURN
END

```

```

C
C
C WEIGHTING FUNCTION NUMBER 1
C
SUBROUTINE SPWGT1(W)
REAL*4 W(100)
C
WRITE(6,15)
15 FORMAT(//5X, 'WEIGHTING FUNCTION NUMBER 1'//)
DO 20 I=1,100
W(I) = 1.0
20 CONTINUE
RETURN
END

```

C
C
C
C

WEIGHTING FUNCTION NUMBER 2

C

SUBROUTINE SPWGT2(W)
REAL*4 W(100)

15

WRITE(6,15)
FORMAT(//SX,'WEIGHTING FUNCTION NUMBER 2'//)

C

DO 20 I=1,100

W(I) = 1.0

20

CONTINUE

DO 30 I=46,55

30

W(I) = 0.001

DO 40 I=68,77

40

W(I) = 0.001

RETURN

END

C
C
C
C

WEIGHTING FUNCTION NUMBER 3

C

SUBROUTINE SPWGT3(W)
REAL*4 W(100)

15

WRITE(6,15)
FORMAT(//SX,'WEIGHTING FUNCTION NUMBER 3'//)

W(1) = .73

W(2) = .91

W(3) = 1.08

W(4) = 1.18

W(5) = 1.22

W(6) = 1.20

W(7) = 1.20

W(8) = 1.18

W(9) = 1.17

W(10) = 1.17

W(11) = 1.17

W(12) = 1.18

W(13) = 1.17

W(14) = 1.15

W(15) = 1.11

W(16) = .83

W(17) = 1.04

W(18) = .57

W(19) = .91

W(20) = .86

W(21) = .80

W(22) = .86

W(23) = .81

W(24) = .61

W(25) = .48

W(26) = .26

W(27) = .28

W(28) = .58

W(29) = .65

W(30) = .63

W(31) = .61
W(32) = .59
W(33) = .53
W(34) = .51
W(35) = .25
W(36) = .14
W(37) = .14
W(38) = .31
W(39) = .31
W(40) = .31
W(41) = .41
W(42) = .41
W(43) = .34
W(44) = .34
W(45) = .34
W(46) = .14
W(47) = .14
W(48) = .001
W(49) = .001
W(50) = .001
W(51) = .001
W(52) = .001
W(53) = .08
W(54) = .08
W(55) = .08
W(56) = .21
W(57) = .21
W(58) = .21
W(59) = .21
W(60) = .21
W(61) = .19
W(62) = .19
W(63) = .15
W(64) = .15
W(65) = .15
W(66) = .11
W(67) = .11
W(68) = .03
W(69) = .03
W(70) = .03
W(71) = .001
W(72) = .001
W(73) = .001
W(74) = .001
W(75) = .001
W(76) = .001
W(77) = .001
W(78) = .02
W(79) = .02
W(80) = .02
W(81) = .04
W(82) = .04
W(83) = .07
W(84) = .07
W(85) = .07
W(86) = .09
W(87) = .09
W(88) = .10
W(89) = .10
W(90) = .10
W(91) = .10
W(92) = .10
W(93) = .07
W(94) = .07
W(95) = .07
W(96) = .04
W(97) = .04
W(98) = .01
W(99) = .01
W(100) = .01
RETURN
END

```

C
C
C WEIGHT FUNCTION NUMBER 4
C
SUBROUTINE SPWGT4(W)
REAL*4 W(100)
C
WRITE(6,15)
15 FORMAT(/,5X,'WEIGHT FUNCTION NUMBER 4'//)
DO 30 I=1,15
30 W(I) = 0.7
DO 40 I=16,45
40 W(I) = 1.0
DO 50 I=46,52
50 W(I) = 0.1
DO 60 I=53,67
60 W(I) = 1.0
DO 70 I=68,77
70 W(I) = 0.01
DO 80 I=78,100
80 W(I) = 1.0
RETURN
END

```

```

C -----
C
C SPTES TRANSFORMS THE DATA USING THE OPTIMUM SET OF BASIS
C VECTORS, COMPUTES THE MEAN-SQUARE ERROR, AND COMPUTES THE
C STATISTICS FOR EACH CLASS.
C
C 6 FEBRUARY, 1978
C -----
C
COMMON ID(100)
REAL*4 P(10),PHI(100,20),X(100),Y(100),Z(100)
REAL*4 AM(100),AVE(20,10),COV(210,10)
C
C SELECT NUMBER OF TERMS
C
WRITE(16,10)
10 FORMAT(5X,'NUMBER OF TERMS?')
READ(15,15)NTERM
15 FORMAT(12)
REWIND 2
READ(2) ID
NCLS = ID(17)
N = ID(18)
NCT = NTERM*(NTERM + 1)/2
NFT = 0
DO 20 I=1,NCLS
20 NFT = NFT + ID(20+I)
DO 25 I=1,NCLS
25 P(I) = 1./FLOAT(NCLS)
CONTINUE
WRITE(7,28)NCLS,NTERM
28 FORMAT(12,3X,12)
WRITE(7,30)(P(I),I=1,NCLS)
30 FORMAT(10F6.4)

```

```

C
C COMPUTE MEAN FUNCTION
C
DO 300 I=1,N
300 AM(I) = 0.0
DO 320 K=1,NFT
READ(2) X
DO 320 I=1,N
AM(I) = AM(I) + X(I)/FLOAT(NFT)
320 CONTINUE
REWIND 2
READ(2) ID

C
C READ EIGENVECTORS
C
DO 40 J=1,NTERM
READ(5,35) (PHI(I,J),I=1,N)
35 FORMAT(20A4)
40 CONTINUE

C
C LOOP ON THE SAMPLE FUNCTIONS IN THE DATA SET
C
AVESQ = 0.0
DO 200 ICLS=1,NCLS
DO 50 I=1,NTERM
50 AVE(I,ICLS) = 0.0
DO 55 I=1,NCT
55 COV(I,ICLS) = 0.0

C
C
NF = ID(20+ICLS)
CON = FLOAT(NF)/FLOAT(NF-1)
DO 150 ISAM=1,NF

C
C READ SAMPLE POINTS FROM FUNCTION
C
READ(2) X

C
C TRANSFORM DATA USING BASIS FUNCTIONS
C
DO 70 J=1,NTERM
Y(J) = 0.0
DO 70 I=1,N
Y(J) = Y(J) + PHI(I,J)*(X(I) - AM(I))
70 CONTINUE

C
C COMPUTE SQUARED ERROR
C
DO 80 I=1,N
80 Z(I) = 0.0
DO 85 J=1,NTERM
DO 85 I=1,N
Z(I) = Z(I) + PHI(I,J)*Y(J)
85 CONTINUE
DO 88 I=1,N
88 Z(I) = Z(I) + AM(I)
XSQ = 0.0
ZSQ = 0.0
XZ = 0.0
TSQ = 0.0
DO 90 I=1,N
XSQ = XSQ + X(I)*X(I)
ZSQ = ZSQ + Z(I)*Z(I)
XZ = XZ + 2.0*X(I)*Z(I)
90 CONTINUE
ESQ = (XSQ - XZ + ZSQ)
AVESQ = AVESQ + ESQ

```



```

C
C READ STATS
C
DO 40 K=1,NCLS
READ(5,30) (AVE(I,K), I=1,N)
READ(5,30) (COV(I,K), I=1,NCT)
30  FORMAT(20A4)
40  CONTINUE
IF(IS)45,45,100
45  CONTINUE
C
C PUNCH STAT DECK
C
DO 50 K=1,NCLS
WRITE(7,70) (AVE(I,K), I=1,N)
DO 60 K=1,NCLS
60  WRITE(7,80) (COV(I,K), I=1,NCT)
70  FORMAT('MN',17A4)
80  FORMAT('CV',17A4)
90  STOP
C
C
100 WRITE(16,110)
110 FORMAT(5X,'TYPE NUMBER OF CHANNELS DESIRED (12)')
READ(15,115)NT
115  FORMAT(12)
DO 150 K=1,NT
WRITE(16,120)
120  FORMAT(5X,'SELECT CHANNELS (12)')
READ(15,125) IC(K)
125  FORMAT(12)
150  CONTINUE
C
WRITE(7,20)NCLS,NT
WRITE(7,25) (P(I), I=1,NCLS)
NCTP = NT*(NT+1)/2
C
C DO 200 ICLS=1,NCLS
C
DO 160 K=1,NT
ICP = IC(K)
160  AVE(K,ICLS) = AVE(ICP,ICLS)
WRITE(7,30) (AVE(I,ICLS), I=1,NT)
C
IN = 0
DO 165 I=1,N
DO 165 J=1,I
IN = IN + 1
COVTP(I,J) = COV(IN,ICLS)
COVTP(J,I) = COV(IN,ICLS)
165  CONTINUE
C
L = 0
DO 180 K=1,NT
L = L + 1
DO 180 KL=1,L
ICK = IC(K)
ICL = IC(KL)
COVTP(K,KL) = COVTP(ICK,ICL)
180  CONTINUE
C
IN = 0
DO 185 I=1,NT
DO 185 J=1,I
IN = IN + 1
COV(IN,ICLS) = COVTP(I,J)
185  CONTINUE
WRITE(7,30) (COV(I,ICLS), I=1,NCTP)
200  CONTINUE
C
STOP
END

```


C
C SET UP BASIS FUNCTIONS FOR THEMATIC MAPPER

C
100 CONTINUE
N = 6
NCT = 21
DO 110 I=2,6
110 PHI(I,1) = 1.0
DO 115 I=7,10
115 PHI(I,2) = 1.0
DO 120 I=11,15
120 PHI(I,3) = 1.0
DO 125 I=18,25
125 PHI(I,4) = 1.0
DO 130 I=58,68
130 PHI(I,5) = 1.0
DO 135 I=84,97
135 PHI(I,6) = 1.0
GO TO 300

C
C SET UP TEST BANDS

C
150 CONTINUE
N = 6
NCT = 21
PHI(10,1) = 1.0
PHI(30,2) = 1.0
PHI(52,3) = 1.0
PHI(64,4) = 1.0
PHI(74,5) = 1.0
PHI(92,6) = 1.0
GO TO 300

C
C SET UP PROPOSED SENSOR BASIS FUNCTIONS

C
200 CONTINUE
N = 8
NCT = 36
DO 210 I=1,7
210 PHI(I,1) = 1.0
DO 215 I=8,13
215 PHI(I,2) = 1.0
DO 220 I=14,15
220 PHI(I,3) = 1.0
DO 225 I=16,25
225 PHI(I,4) = 1.0
DO 230 I=28,30
230 PHI(I,5) = 1.0
DO 235 I=31,45
235 PHI(I,6) = 1.0
DO 240 I=56,67
240 PHI(I,7) = 1.0
DO 245 I=78,100
245 PHI(I,8) = 1.0
GO TO 300

C
C SET UP PROPOSED SENSOR NUMBER 2

C
250 CONTINUE
N = 8
NCT = 36
DO 255 I=1,13
255 PHI(I,1) = 1.0
DO 260 I=14,15
260 PHI(I,2) = 1.0
DO 265 I=16,26
265 PHI(I,3) = 1.0
DO 270 I=27,32
270 PHI(I,4) = 1.0

```

DO 275 I=33,35
275 PHI(I,5) = 1.0
DO 280 I=36,45
280 PHI(I,6) = 1.0
DO 285 I=56,67
285 PHI(I,7) = 1.0
DO 290 I=78,100
290 PHI(I,8) = 1.0
C
C
300 CONTINUE
C
C NORMALIZE BASIS FUNCTIONS
C
DO 430 L=1,N
SUM = 0.0
DO 410 I=1,100
SUM = SUM + ABS(PHI(I,L))
410 CONTINUE
DO 420 I=1,100
PHIN(I,L) = PHI(I,L)/SUM
420 CONTINUE
430 CONTINUE
C
C POSITION TAPE AND READ ID INFORMATION
C
REWIND 2
READ(2) ID

NCLS = ID(17)
NFT = 0
AVESQ = 0.0
DO 302 IL=1,NCLS
302 NFT = NFT + ID(20+IL)
NXP = ID(16)
WRITE(6,305)NXP,NCLS,N
305 FORMAT(1H1,5(/),20X,'SUBOPTIMUM SENSOR SIMULATION FOR EXP. NUMBER'
•,15,//20X,'NUMBER OF CLASSES =' ,13,//20X,'NUMBER OF DIMENSIONS =' ,
•13)
WRITE(6,308)SENSOR
308 FORMAT(//20X,'SENSOR ',11)
WRITE(7,310)NCLS,N
310 FORMAT(12,3X,12)
DO 312 IL=1,NCLS
NF = ID(20+IL)
312 P(IL) = 1./FLOAT(NCLS)
WRITE(7,314)(P(I),I=1,NCLS)
314 FORMAT(10F6.4)
CALL SPWGT2(W)
C
C LOOP ON THE SAMPLE FUNCTIONS IN THE DATA SET
C
DO 600 ICLS=1,NCLS
DO 315 I=1,N
315 AVE(I,ICLS) = 0.0
DO 320 I=1,NCT
320 COV(I,ICLS) = 0.0
C
C
NF = ID(20+ICLS)
CON = FLOAT(NF)/FLOAT(NF-1)
DO 500 ISAM=1,NF
C
C READ SAMPLE POINTS FROM FUNCTION
C
READ(2) X

```

```

C
C C   TRANSFORM DATA USING BASIS FUNCTION
C
      DO 330 J=1,N
      Y(J) = 0.0
      DO 330 I=1,100
      Y(J) = Y(J) + (PHIN(I,J)*X(I))
330   CONTINUE
C
C C   COMPUTE SQUARED ERROR
C
      DO 335 I=1,100
335   Z(I) = 0.0
      DO 340 J=1,N
      DO 340 I=1,100
      Z(I) = Z(I) + PHI(I,J)*Y(J)
340   CONTINUE
      XSQ = 0.0
      ZSQ = 0.0
      XZ = 0.0
      DO 345 I=1,100
      XSQ = XSQ + X(I)*X(I)*W(I)
      ZSQ = ZSQ + Z(I)*Z(I)*W(I)
      XZ = XZ + 2.0*X(I)*Z(I)*W(I)
345   CONTINUE
      ESQ = (XSQ - XZ + ZSQ)
      AVESQ = AVESQ + ESQ
C
C C   COMPUTE STATISTICS
C
      DO 350 I=1,N
      AVE(I,ICLS) = AVE(I,ICLS) + Y(I)/FLOAT(NF)
350   CONTINUE
      IN = 0
      DO 360 J=1,N
      DO 360 I=1,J
      IN = IN + 1
      COV(IN,ICLS) = COV(IN,ICLS) + Y(I)*Y(J)/FLOAT(NF-1)
360   CONTINUE
500   CONTINUE
C
C C   PRINT STATISTICS FOR THE CLASS
C
      IN = 0
      DO 510 J=1,N
      DO 510 I=1,J
      IN = IN + 1
      COV(IN,ICLS) = COV(IN,ICLS) - CON*AVE(I,ICLS)*AVE(J,ICLS)
510   CONTINUE
      WRITE(6,515)ICLS
515   FORMAT(5(/),10X,'STATISTICS FOR CLASS',14)
      CALL MCOVP(N,AVE(I,ICLS),COV(I,ICLS))
      WRITE(7,520) (AVE(I,ICLS),I=1,N)
520   FORMAT(20A4)
      WRITE(7,530) (COV(I,ICLS),I=1,NCT)
530   FORMAT(20A4)
C
C C
600   CONTINUE
      AVESQ = AVESQ/FLOAT(NFT)
      WRITE(6,610)AVESQ
610   FORMAT(///10X,'MEAN-SQUARE ERROR = ',E14.8)
      STOP
      END

```

```

C
C COMPUTATION OF AVERAGE ERROR OVER A DATA SET.
C
COMMON ID(100)
REAL*4 X(100),W(100)
C
REWIND 2
READ(2) ID
CALL SPLBL
CALL SPWGT4(W)
N = ID(18)
NCLS = ID(17)
DELTA = 0.07
AVE = 0.0
NFT = 0
DO 100 K=1,NCLS
NF = ID(20+K)
NFT = NFT + NF
DO 100 L=1,NF
READ(2) X
SUM = 0.0
DO 20 I=1,N
SUM = SUM + DELTA*DELTA*X(I)*X(I)*W(I)
20 CONTINUE
AVE = AVE + SUM
100 CONTINUE
AVE = AVE/FLOAT(NFT)
WRITE(6,120)AVE
120 FORMAT(5(/),5X,'AVERAGE MEASUREMENT ERROR =',E20.8)
C
STOP
END

```

```

C
C CONTROL PROGRAM FOR SPEST TO PROVIDE FOR VARIABLE DIMENSIONING.
C
C 24 JAN., 1978
C
REAL*4 PR(10),PHI(10,10,10),P(10),AM(10,10),COV(55,10)
C
C READ INPUT VARIABLES
C M - CLASSES: N - DIMENSIONS
C
READ(5,10)M,N
10 FORMAT(12,3X,12)
C
C P(I) - APRIORI PROBABILITIES
C
NCT = N*(N+1)/2
READ(5,15)(P(I),I=1,M)
15 FORMAT(10F6.4)

```

```

C
C MEANS AND COVARIANCES FOR EACH CLASS
C
      DO 30 I=1,M
      READ(5,20) (AM(J,I),J=1,N)
      READ(5,20) (COV(K,I),K=1,NCT)
20  FORMAT(20A4)
      CALL MCOVP(N,AM(1,I),COV(1,I))
30  CONTINUE
      CALL SPESTM(M,N,PHI,P,AM,COV,PR,PC)
      DO 40 I=1,M
40  WRITE(6,50) I,PR(I)
50  FORMAT(//10X,'PROBABILITY OF CORRECT CLASSIFICATION FOR CLASS',I3,
      *  = ',F6.4)
      WRITE(6,60)PC
60  FORMAT(///10X,'OVERALL PROBABILITY OF CORRECT RECOGNITION = ',F6.4
      *)
      STOP
      END

```

```

C -----
C
C SPESTM IS AN ESTIMATOR OF THE CLASSIFICATION PERFORMANCE FROM A
C GIVEN SET OF STATISTICS FROM M CLASSES. THE ESTIMATOR IS A
C STRATIFIED POSTERIOR ESTIMATOR (REF. WHITSITT AND LANDGREBE).
C THE PROBABILITY DISTRIBUTIONS ARE ASSUMED TO BE MULTIVARIATE
C GAUSSIAN
C

```

19 JANUARY, 1978

```

C -----
C
C SUBROUTINE SPESTM(M,N,PHI,P,AM,COV,PR,PC)
C REAL*4 QP(10),P(10),PR(10),AM(10,10),COV(55,10),COVT(55)
C REAL*4 GAM(10,10),PHI(N,N,M),DET(10),COVIN(55,10)
C REAL*4 Y(10),TE1(10),DEL(10),COVU(10,10)
C REAL*8 PX(10),BIG,DEN,SDET(10),BETA,Z0,Z1,Z2,Z3
C

```

LIST OF VARIABLES

```

C M = NUMBER OF CLASSES
C N = NUMBER OF DIMENSIONS
C P(I) = APRIORI PROBABILITIES OF CLASS I
C PR(I) = CLASS CONDITIONAL PERFORMANCE
C PC = OVERALL PERFORMANCE
C AM(J,I) = MEAN VECTOR OF CLASS I
C COV(J,I) = COVARIANCE MATRIX OF CLASS I (STORED IN UPPER TRIANGULAR
C FORM)
C

```

```

C -----
C
C IX = 947913
C NCT = N*(N+1)/2

```

```

C
C COMPUTE EIGENVALUES AND EIGENVECTORS FOR EACH MATRIX
C
      MV = 0
      EPS = 1.0E-6
      DO 100 IJ=1,M
      DO 55 I=1,NCT
55      COVT(I) = COV(I,IJ)
      CALL EIGEN(COVT,PHI(1,1,IJ),N,MV)
      L = 0
      DO 60 I =1,N
      L = L + 1
60      GAM(I,IJ) = COVT(L)
C
C COMPUTE DETERMINANT AND INVERSE OF EACH MATRIX
C
      DO 65 I=1,NCT
65      COVT(I) = COV(I,IJ)
      CALL SMINV(COVT,N,DET(IJ),MV,EPS,IER)
      IF(IER) 1000,70,1000
70      CONTINUE
      SDET(IJ) = SQRT(DET(IJ))
      DO 75 I=1,NCT
75      COVIN(I,IJ) = COVT(I)
100     CONTINUE
      MV = 0
      DO 105 I=1,M
105     QP(I) = 0.0
C
C LOOP ON CLASS ICL
C
      PC = 0.0
      DO 500 ICL=1,M
      AVEQ = 0.0
C
C LOOP ON THE NUMBER OF SAMPLES
C
      NS = 1000
      DO 300 IJ=1,NS
C
C GENERATE Y VECTOR FROM CLASS ICL
C
      DO 110 I=1,N
      CALL RANDU(IX,IY,XP)
      IX = IY
      CALL NDTRI(XP,Y(I),XD,IER)
110     CONTINUE
C
C COMPUTE CONDITIONAL PROBABILITIES FOR EACH CLASS
C
      DO 200 JCL=1,M
      IF(JCL .EQ. ICL) GO TO 180
      DO 130 I=1,N
      TE1(I) = 0.0
      DEL(I) = AM(I,ICL) - AM(I,JCL)
      DO 130 J=1,N
      TE1(I) = TE1(I) + SQRT(GAM(J,ICL))*Y(J)*PHI(I,J,ICL)
130     CONTINUE
      JJ = 0
      DO 140 I=1,N
      DO 140 J=1,I
      JJ = JJ + 1
      COVU(I,J) = COVIN(JJ,JCL)
      COVU(J,I) = COVIN(JJ,JCL)
140     CONTINUE

```

```

Z1 = 0.0
Z2 = 0.0
Z3 = 0.0
DO 150 I=1,N
DO 150 J=1,N
Z1 = Z1 - 0.5*TE1(I)*COVU(I,J)*TE1(J)
Z2 = Z2 - TE1(I)*COVU(I,J)*DEL(J)
Z3 = Z3 - 0.5*DEL(I)*COVU(I,J)*DEL(J)
150 CONTINUE
ZSUM = Z1 + Z2 + Z3
IF(ZSUM .LT. -100) GO TO 190
BETA = P(JCL)*1.0
PX(JCL) = BETA*DEXP(Z1+Z2+Z3)/SDET(JCL)
IF(PX(JCL) .EQ. 0.0) WRITE(16,919) ICL, JCL, ZSUM, SDET(JCL), PX(JCL)
170 CONTINUE
GO TO 200
180 CONTINUE
Z0 = 0.0
DO 185 I=1,N
Z0 = Z0 - 0.5*Y(I)*Y(I)
185 CONTINUE
IF(Z0 .LT. -100) GO TO 190
BETA = P(JCL)*1.0
PX(JCL) = BETA*DEXP(Z0)/SDET(JCL)
IF(PX(JCL) .EQ. 0.0) WRITE(16,919) ICL, JCL, Z0, SDET(JCL), PX(JCL)
GO TO 200
190 PX(JCL) = 0.0
200 CONTINUE
919 FORMAT(5X,2I5,3E12.4)
C
C
C CHOOSE THE LARGEST
C
BIG = -1000
DO 220 I=1,M
IF(PX(I) .GT. BIG) LOC = I
IF(PX(I) .GT. BIG) BIG = PX(I)
220 CONTINUE
DEN = 0.0
DO 230 I=1,M
DEN = DEN + PX(I)
230 CONTINUE
Q = BIG/DEN
C
C AVERAGE
C
QP(LOC) = QP(LOC) + P(ICL)*Q/P(LOC)
300 CONTINUE
500 CONTINUE
DO 510 ICL=1,M
PR(ICL) = QP(ICL)/FLOAT(NS)
PC = PC + P(ICL)*PR(ICL)
510 CONTINUE
RETURN
1000 WRITE(6,1100) IER
1100 FORMAT(10X, '***INVERSION ERROR(', I2, ')***')
STOP
END

```