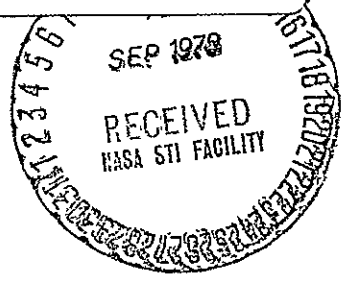
LARS Technical Report 061778

Analytical Techniques for the Study of Some Parameters of Multispectral Scanner Systems for Remote Sensing

E. R. Wiswell
G. R. Cooper



Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA
EE-TR 78-28 June 1978

STAR INFORMATION FORM

1 Report No	2 Government Accession No	3 Recipient's Catalog No
061778		

4 Title and Subtitle
Analytical Techniques for the Study of Some Parameters of Multispectral Scanner Systems for Remote Sensing

5 Report Date
June 17, 1978

6. Performing Organization Code

7 Author(s)
Eric R. Wiswell
George R. Cooper

8 Performing Organization Report No

10 Work Unit No

9 Performing Organization Name and Address
Laboratory for Applications of Remote Sensing
1220 Potter Drive
West Lafayette, Indiana 47906

11 Contract or Grant No
NAS9-15466

13 Type of Report and Period Covered

12 Sponsoring Agency Name and Address
NASA Johnson Space Flight Center
Houston, Texas 77058

14 Sponsoring Agency Code

15 Supplementary Notes

16 Abstract

This report presents analytical techniques for the study of some parameters of multispectral scanner systems for remote sensing. Selection of a subset of spectral bands, the effects of differing spectral bandwidths, and signal-to-noise characteristics of spectral bands are some specific parameters considered. The observations made by a multispectral scanner system are modeled as a spectral random process of the scene corrupted by an additive independent noise random process in wavelength.

The concept of average mutual information in the received spectral random process about the spectral scene is developed for use in the study. Techniques amenable to implementation on a digital computer are developed to make the required average mutual information calculations. The techniques developed for computation of average information require the identification of models for the spectral response process of scenes. Stochastic modeling techniques are adapted for use in the present context. Application of the developed techniques is demonstrated on empirical data for two spectral scene types. The first set of empirical data is from a wheat scene. The second data set is a scene consisting of several vegetation types. Average mutual information is then calculated for each spectral band of both scene types as a demonstration of its use for study of such parameters as spectral bandwidth, signal-to-noise properties, and selection of a subset of spectral bands.

It is concluded that average mutual information is a useful concept for the study of some parameters of multispectral scanner systems. This technique indicates, for example, that for the scene types studies, the infrared portion of the spectrum deserves increased attention in multispectral scanner system design. The model construction techniques are a flexible and systematic method for modeling the spectral response of a scene.

17 Key Words (Suggested by Author(s))	18 Distribution Statement
Multispectral Scanners Remote Sensing Information Theory Wavelength Band Selection	

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

*For sale by the National Technical Information Service Springfield Virginia 22161

LARS Technical Report 061778

ANALYTICAL TECHNIQUES FOR THE STUDY OF SOME PARAMETERS OF MULTISPECTRAL SCANNER SYSTEMS FOR REMOTE SENSING

by

E. R. Wiswell
G. R. Cooper

Published by the

Laboratory for Applications of Remote Sensing
Purdue University
Lafayette, Indiana 47907

EE-TR 78-28

This work was sponsored by the National Aeronautics and Space Administration under Contracts NAS9-14016, NAS9-14970 and NAS9-15466.

June 1978

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1. AR(n) -- autoregressive process of order n
2. ARC(n) -- autoregressive process of order n plus constant trend
3. IAR(n) -- integrated autoregressive process of order n
4. $IAR_2(n)$ -- integrated autoregressive process of the second kind of order n

Chapter I

Introduction

1. General Discussion

Although still a comparatively young technology, remote sensing of the environment has greatly extended man's perception of the world's resources and interaction of natural and unnatural influences. Remote sensing has grown from simple photography and photointerpretation to satellite borne sensors and sophisticated machine aided analysis. A critical portion of many modern remote sensing systems is a multispectral scanner. Multispectral scanner systems employ sensors to observe portions of the electromagnetic spectrum typically ranging from the visible region to the reflective infrared regions. The thermal (or emissive) portion of the spectrum also has important uses in remote sensing. Thus, investigation of multispectral scanner systems and parameters of multispectral scanner systems is an important and necessary endeavor.

Multispectral scanner systems are characterized by many parameters interacting in complicated ways. This research develops analytical techniques for the study of some of these parameters. Many of the parameters tend to be dependent on the scenes observed by the multispectral scanner systems. It is thought that consideration of scene dependent parameters in the current research provides a framework for considering very specialized scanner systems as well as more general scanner systems. An important example of a parameter that must be con-

sidered for multispectral scanner systems is what portions of the electromagnetic spectrum are to be observed. It is clear that this may tend to be a very scene dependent parameter. Indeed, under differing observation conditions, it may be necessary to observe different portions of the electromagnetic spectrum to obtain the desired information about a single specialized scene type. The development of analytical techniques to aid in the study of some of the parameters of multispectral scanner systems is the objective of this research.

2. Previous Work

Up to the present, there has been little analytical work aimed at general techniques for the study of parameters of multispectral scanner systems. Most reported work has tended to be ad hoc and empirical. This has produced detailed knowledge of various aspects of remote sensing problems, but has not produced studies of of multispectral scanner systems in an analytic context.

Studies such as those by Gates, Keegan, Schleter and Weidner [G3] and Sinclair, Hoffer and Schreiber [S4] are indicative of the type of detailed knowledge that has been gained about scenes that may be observed by multispectral scanner systems. Examples of studies that have been conducted for specific problems are those by Coggeshall and Hoffer [C3], and Kumar and Silva [K5]. These papers are not referenced so much for their contents, but rather as examples of the wide variety of studies that have been carried out in an effort to understand aspects of remote sensing problems.

An early study that considers a physical basis for remote sensor system design is described by Holmes and MacDonald [H4]. This paper

gives a good exposition of many of the physical considerations for multispectral scanner system design. A more recent study by Landgrebe, Biehl and Simmons [L2], [L4] considers in an empirical manner several important parameters in multispectral scanner systems. Some of these parameters are spatial resolution and spatial sampling characteristics, spectral sampling and bands, and signal-to-noise characteristics. These are important parameters and many conclusions can be drawn from empirical study. However, it is thought that the development of analytical techniques to study some of these and other parameters is now appropriate.

3. The Present Investigation

As previously mentioned, very little anaytical consideration of many multispectral scanner system parameters has been done. It is the intention of this research to develop analytical techniques to study some of these parameters. Although the developed techniques are applicable to a wide variety of scenes, this reseach uses vegetation scenes as a vehicle for consideration of the techniques.

Consider a single type of vegetation illuminated by the sun. If the reflected electromagnetic energy in the visible to reflective infrared wavelengths (approximately .4 to 3.0 micrometers, μm) is measured as a function of wavelength, λ, for several different observations of the same vegetation type, it is observed that the spectral response exhibits random variation about a mean value at each wavelength. That is, the observations tend to be stochastic in nature. Now, if the vegetation scene is observed remotely, say from a satellite or airborne platform, there are additional disturbances of the observations. These dis-

turbances may be due to atmospheric noise, random disturbance of the observation platform, or other sources. The point is that the multispectral scanner system receives electromagnetic energy from the scene that exhibits random variations corrupted by disturbances that also have random variation. Under these conditions, it is logical to suppose that certain spectral regions (or spectral bands) may be more useful than others for observing those features of the scene that may be of interest. It is also logical to conclude that by studying the effect of the disturbance on the observation of the scene, it may be possible to minimize the adverse effects. Thus, in view of the above comments, it would be useful to characterize analytically what information the observation conveys about the scene.

This is highly reminiscent of the classical problem in communication systems. A receiver (multispectral scanner) obtains a signal (the electromagnetic energy from the scene) that is corrupted in some manner (perhaps by random noise). It is then desired to introduce a quantitive measure of what the received signal conveys about the transmitted signal. This is the special scope of the subject of information theory. The birth of the field may be said to be, of course, in the work of Shannon [S1]. There are voluminous references in the field with major texts by Fano [F4] and Gallager [G1]. The relation of received signal to transmitted signal is described in information theory by the concept of average (mutual) information. Loosely speaking, the average information in the received signal about the transmited signal may be said to be the reduction in uncertainty about the transmitted signal that is obtained from the received signal.

Application of the concept of average information to the study of some parameters of multispectral scanner systems is pursued in some detail in this research. It is thought that this gives insight into the study of the relative utility of different spectral bands to be used in scanner systems for observation of spectral scenes. Further, utlization of information theoretic concepts can be used to study the effects of noise disturbances on the observation of spectral scenes. The development of the information theoretic concepts is the topic of Chapter II.

The computation of average information in spectral data received at the multispectral scanner about an observed spectral scene is not without difficulties. A method to circumvent the necessity of solving some rather intractable equations is described in Chapter II. Also, methods for digital computation of average information are developed in Chapter II. These computation procedures require that models for the spectral response of a scene be developed.

Chapter III contains the development of the techniques used to construct models for spectral scenes. Several approaches to construction of the models could be pursued. Most of the approaches are studied in terms of the system identification problem. Saridis has done extensive work on stochastic approximation methods for systems identification [S5], [S6], and [S7]. The stochastic approximation techniques have the advantages of being relatively easily implemented and having great generality. Maximum likelihood identification techniques have also been extensively used. The references by Kashyap and Rao [K3], and Kashyap [K6] give good dicussions of the maximum likelihood identification technique. The maximum likelihood techniques are used in Chapter III to

develop techniques for constructing models for spectral scenes. Concepts from the area generally known as time series analysis are also developed for use in construction of models for spectral scenes. Box and Jenkins [B1], Anderson [A4], and Kashyap and Rao [K3] are good references for time series analysis and its application to construction of dynamic models from empirical data. Also discussed is a Bayesian identification technique for models of spectral scenes. This technique is an adaptation of a method described by Kashyap and Rao [K3] to the present work.

A criterion for selecting one of several hypothesized models for a spectral scene is discussed in Chapter III. Further, once a candidate model for a spectral scene has been selected, the question of the validity of the model remains. Validation techniques for testing candidate models for spectral scenes are also discussed in Chapter III.

Chapter IV uses the model construction techniques developed in Chapter III on empirical data from actual scene types that may be observed by a multispectral scanner system for remote sensing of agricultural scenes. The empirical data consists of two vegetative scene types. To demonstrate the model construction technique on a scene of a single vegetation type, a wheat scene is considered. A set of empirical data made up of several vegetation types is used to demonstrate the model construction techniques on a more general vegetative scene. The empirical data sets are divided into several spectral bands for two reasons. First, most multispectral scanner systems tend to be designed around different sensors for different spectral bands. Second, it is thought that better models of the spectral scenes can be obtained by considering

several bands in the spectral region of interest than if only one model is constructed for the entire spectral region (approximately .45 to 2.4 μm for the present case) under consideration. Several models are hypothesized for each band and the parameters characterizing each are identified using the maximum likelihood technique. Candidate models are then selected using the selection criterion developed in Chaper III. Finally, candidate models are validated using the techniques developed in Chapter III.

In Chapter V a simple application of the computation of average information as developed in Chapter II is carried out using the models for the spectral response developed in Chapter IV. This application is intended to demonstrate how the average information computation can be used to select a subset of spectral bands. Also it is demonstrated that average information can be used to study such parameters as signal-to-noise properties in different spectral bands. The relation of average information to spectral bandwidth is implicit in the discussion.

Chapter VI is devoted to conclusions about the research and thoughts for extension of the research.

Chapter II

Information Theoretic Approach

1. Introduction

In this chapter, information theory concepts are developed for use in the study of spectral scenes. The basis for this study is the manner in which the spectral response is considered. The spectral response for several observations of the same variety of vegetation exhibits random variation about a mean spectral response at each wavelength. Thus a major consideration for representation of a spectral scene is the ability to adequately model its inherent randomness. Another consideration is analytical tractability. Hence, it is reasonable to consider the spectral response of a scene as a sample function of a portion of a random process in wavelength. That is, the spectral response is given by $s(\lambda)$, where $s(\lambda) \epsilon S$ and $\lambda \epsilon \left[\lambda_1, \lambda_2\right]$. The ensemble of sample functions for the spectral response is S and $\left[\lambda_1, \lambda_2\right]$ is the interval of wavelengths of interest. It is not necessary to assume that the spectral random process is stationary. In fact, it will be seen later that the spectral process will, in general, not be stationary. On the basis of empirical studies by Fu, Landgrebe and Phillips [F1], a reasonable assumption on the statistics of the spectral process can be made. The spectral random process will be assumed to be a gaussian process. The mean and variance of the process will be apparent when models of the spectral process are discussed in the next chapter. The gaussian assumption can also be justified from another point of view. When a multispectral scanner system views a scene, it receives a signal from many sources in the field of view. If it is assumed that the scanner system is observing many in-

dependent and identically distributed sources, then the central limit theorem can be invoked to justify the assumed gaussian statistics. The gaussian assumption is important in the analytical results of this research.

2. Definition of Average Mutual Informations

The signal received by a multispectral scanner is assumed to consist of the spectral signal for the scene, $s(\lambda)$, disturbed by a statistically independent additive random process (noise). This noise process consists of the disturbances in the spectral scene not attributable to the vegetation under observation and random disturbances in the channel between the scene and the multispectral scanner. In the present research these disturbances are all combined into one noise random process in wavelength, $n(\lambda)$. The noise is also assumed to be a gaussian random process. This assumption is made for the same reasons as for the signal process. It is further assumed that the noise process, $n(\lambda)$, is white. In the present context, white noise is a zero mean random process with autocovariance given by

$$E[n(\lambda)n(u)] = \frac{N_0}{2}\delta(\lambda - u), \quad \lambda_1 \leq \lambda, u \leq \lambda_2 \tag{2-1}$$

where

$E[\cdot]$ is the expectation operator,

$\delta(\cdot)$ is the Dirac delta function.

Thus the spectral process received by the multispectral scanner is

represented by

$$y(\lambda) = s(\lambda) + n(\lambda) \quad , \quad \lambda \in [\lambda_1, \lambda_2] \tag{2-2}$$

It is not necessary that the noise be white. However, for purposes of studying the bands of a multispectral scanner and their information content, this assumption is sufficient. It is still possible to allow the white noise to have a different spectral density level in each band. The more general problem of mutual information in time continuous processes with non-white noise is considered by Huang [H1], [H2].

The problem to be considered first is to define, and later calculate, the average (mutual) information in the process $y(\lambda)$ about the process $s(\lambda)$. First, however, it is necessary to state some basic and well-known results from information theory concerning the average information in one set of random varibles about another set of random variables.

The averge mutual information in a set of random variables $\underline{V} = \{v_1, v_2, \ldots, v_N\}$ about the set of random variables $\underline{U} \equiv \{u_1, u_2, \ldots, u_M\}$ is defined by [G1]

$$I(\underline{U},\underline{V}) = \int_{\underline{U}} \int_{\underline{V}} P_{\underline{UV}}(\underline{u},\underline{v}) \log \left[\frac{P_{\underline{UV}}(\underline{u},\underline{v})}{P_{\underline{U}}(\underline{u}) P_{\underline{V}}(\underline{v})} \right] d\underline{u} d\underline{v} \tag{2-3}$$

where

$P_{\underline{U},\underline{V}}(\underline{u},\underline{v})$ = joint density function of the sets $\underline{U}$ and $\underline{V}$

$P_{\underline{U}}(\underline{u})$ = density function of the set $\underline{U}$

$P_{\underline{V}}(\underline{v})$ = density function of the set $\underline{V}$

and $\int_{\underline{U}}$ and $\int_{\underline{V}}$ represent M-fold and N-fold integrals over all the possible values of the members of the sets $\underline{U}$ and $\underline{V}$.

Since the definition of average (mutual) information is known for random variables, an intuitively pleasing approach is to represent the spectral random processes in terms of random variables and thus apply the previously known definition. This is the approach of Shannon [S1] for the case of band limited time functions and an infinite observation interval. It has been shown rigorously by Gelfand and Yaglom [G2] that this approach leads to a valid definition of mutual information for time continuous processes under almost all conditions.

Suppose that there exists a set of random variables $S = \{s_1, s_2, \ldots\}$ that uniquely determines and is uniquely determined by $s(\lambda)$. Similarly suppose that there exists a set of random variables $Y = \{y_1, y_2, \ldots\}$ that uniquely determines and is uniquely determined by the portion of $y(\lambda)$ that contains $s(\lambda)$. Under the assumption of independence of $s(\lambda)$ and $n(\lambda)$, any portion of $y(\lambda)$ not represented by the set Y is irrelevant to the calculation of the average information in $y(\lambda)$ about $s(\lambda)$. Then it is reasonable to say that the average information in Y about S is the same as the average information in $y(\lambda)$ about $s(\lambda)$. Thus we make the following definition of the average mutual information in the process $y(\lambda)$ about the process $s(\lambda)$.

$$I(s(\lambda), y(\lambda)) \triangleq I(S,Y) \tag{2-4}$$

That the average (mutual) information can indeed be defined in such a manner merits some elaboration. A method to determine S and Y for given $s(\lambda)$ and $n(\lambda)$ is needed. First consider such a representation for $s(\lambda)$. Assume that the process $s(\lambda)$ has a finite mean square value. That is

$$E\left[s^2(\lambda)\right] < \infty, \qquad \lambda \epsilon \left[\lambda_1, \lambda_2\right] \tag{2-5}$$

Let $\Phi = \left[\phi_i(\lambda) \; ; \; i = 1, 2, \ldots\right]$ be a complete orthonormal set of functions for the class of square integrable functions on $\left[\lambda_1, \lambda_2\right]$. Then $s(\lambda)$ may be represented on the interval $\left[\lambda_1, \lambda_2\right]$ as

$$s(\lambda) = \underset{n\to\infty}{\text{l.i.m.}} \sum_{i=1}^{n} S_i \phi_i(\lambda) \quad , \quad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-6}$$

where l.i.m. is the limit in the mean defined by

$$\lim_{n\to\infty} E\left[\left(s(\lambda) - \sum_{i=1}^{n} s_i \phi_i(\lambda)\right)^2\right] = 0 \; , \qquad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-7}$$

The random variable s_i is defined by

$$s_i = \int_{\lambda_1}^{\lambda_2} s(\lambda)\phi_i(\lambda)\,d\lambda \tag{2-8}$$

The details for such a representation may be found in such texts as [G1] and [V1].

Similarly for another set of complete orthonormal functions $\theta_1(\lambda)$, $\theta_2(\lambda)$, ... on $[\lambda_1, \lambda_2]$ the received spectral process $y(\lambda)$ can be represented as

$$y(\lambda) = \underset{n\to\infty}{\text{l.i.m.}} \sum_{i=1}^{n} y_i\, \theta(\lambda) , \qquad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-9}$$

where

$$y_i = \int_{\lambda_1}^{\lambda_2} y(\lambda)\theta(\lambda)d\lambda \quad . \tag{2-10}$$

It is often convenient to choose the set $\{\theta_i(\lambda)\ ;\ i=1, \ldots\}$ to be the same as the set $\{\phi_i\ ;\ i=1, \ldots\}$. In particular, this is true for white noise disturbances.

Thus, if the sets $\{\phi_i(\lambda)\ ;\ i=1, \ldots\}$ and $\{\theta_i(\lambda)\ ;\ i=1, \ldots\}$ can be determined, we have sets of random variables $\{s_i\ ;\ i=1,2,\ldots\}$ and $\{y_i\ ;\ i=1,2,\ldots\}$ that uniquely determine and are uniquely determined by $s(\lambda)$ and $y(\lambda)$ respectively in the manner previously discussed. The complete orthonormal sets $\{\phi_i(\lambda)\ ;\ i=1,2,\ldots\}$ and $\{\theta_i(\lambda)\ ;\ i=1,2,\ldots\}$ will be determined later in a manner that is relevant to writing an expression for average information for the processes $s(\lambda)$ and $y(\lambda)$.

Hence, if we define

$$S_n = \{s_1, s_2, \ldots, s_n\} \tag{2-11}$$

and

$$Y_n = \{y_1, y_2, \ldots, y_n\} \tag{2-12}$$

we can write from the basic definition of average (mutual) information given in (2-3),

$$I(S_n,Y_n) = \int_{S_n}\int_{Y_n} P(S_n,Y_n)\log\left[\frac{p(S_n,Y_n)}{p(S_n)p(Y_n)}\right] dS_n dY_n \tag{2-13}$$

where $p(S_n,Y_n)$, $p(S_n)$, and $p(Y_n)$ are the appropriate joint probability density functions.

Since in general n will be countably infinite we write

$$I(S,Y) = \lim_{n\to\infty} I(S_n,Y_n) \quad . \tag{2-14}$$

Thus we have an expression for $I(s(\lambda),y(\lambda))$. The complete mathematical details are given by Huang [H1] and Gelfand and Yaglom [G2].

The problem now at hand is to translate this definition into a form that is useful for the current problem of determining the average information in the received spectral process $\overline{y(\lambda)}$ about the spectral scene $s(\lambda)$. In the next section this problem is pursued.

3. Calculation of Average Information

In this section, appropriate sets of basis functions $\{\phi_1(\lambda), \phi_2(\lambda), \ldots\}$ and $\{\theta_1(\lambda), \theta_2(\lambda), \ldots\}$ are defined. These basis functions will, at least in principle, yield a calculation technique for average information. First consider the representation for $s(\lambda)$,

$$s(\lambda) = \sum_{i=1}^{\infty} s_i \phi_i(\lambda) \quad , \lambda_1 \leq \lambda \leq \lambda_2 . \tag{2-15}$$

Note: For notational simplicity we have changed from

$$\text{l.i.m.}_{n\to\infty} \sum_{i=1}^{n} s_i \phi_i(\lambda)$$ to the above.

The covariance function of $s(\lambda)$ is defined as

$$K_s(\lambda,u) = E[s(\lambda)s(u)] \qquad \lambda,u \in [\lambda_1,\lambda_2] . \tag{2-16}$$

and it is straightforward to show that [P1, p. 431]

$$K_s^2(\lambda,u) \leq K_s(\lambda,\lambda)K_s(u,u) \quad . \tag{2-17}$$

Since we have restricted ourselves to processes with finite mean square value, it follows that

$$\int_{\lambda_1}^{\lambda_2}\int_{\lambda_1}^{\lambda_2} K_s^2(\lambda,u)\,d\lambda\,du \leq \left[\int_{\lambda_1}^{\lambda_2} E\left[s^2(\lambda)\right] d\lambda\right]^2 < \infty \tag{2-18}$$

That is, the processes under consideration also have square integrable covariance functions. Since we are dealing with a gaussian random process, a useful additional property that is to be required of the basis functions is that the terms s_i and s_j, $i \neq j$, be uncorrelated. That is,

$$E[s_i s_j] = a_j \, \delta_{ij} \tag{2-19}$$

where

$$\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases} \quad .$$

The usefulness of this requirement will be seen later.

These are the classic requirements for the representation of a random process in terms of a Karhunen-Loeve expansion [V1], [C1]. In this expansion, the basis functions $\phi_j(\lambda)$ are the eigenfunctions of the integral equation

$$a_j\phi_j(\lambda) = \int_{\lambda_1}^{\lambda_2} K_s(\lambda,u)\phi_j(u)du \ , \quad \lambda_1 \leq \lambda \leq \lambda_2 \quad . \tag{2-20}$$

The eigenvalues of the integral equation are the numbers $\{a_j, \ j = 1, 2, \ldots\}$. Of course, the eigenfunctions $\{\phi_j(\lambda), \ j = 1, 2, \ldots\}$ have the required orthonormality property

$$\int_{\lambda_1}^{\lambda_2} \phi_i(\lambda)\phi_j(\lambda)d\lambda = \delta_{ij} \quad . \tag{2-21}$$

Another property is that the sum of the eigenvalues is the average energy of the process $s(\lambda)$. That is,

$$E\left[\int_{\lambda_1}^{\lambda_2} s^2(\lambda)d\lambda\right] = \sum_{j=1}^{\infty} a_j \quad . \tag{2-22}$$

Since it is assumed that $s(\lambda)$ is from a gaussian random process, the uncorrelated random variables $\{s_i, \ i=1, 2, \ldots\}$ are also statistically independent. This property is used later.

Now consider the generation of appropriate basis functions $\{\theta_i(\lambda), \ i=1, 2, \ldots\}$ for representation of the received spectral process $y(\lambda)$. Since it is assumed that $y(\lambda)$ is of the form

$$y(\lambda) = s(\lambda) + n(\lambda) \quad , \quad \lambda_1 \leq \lambda \leq \lambda_2, \tag{2-23}$$

where $s(\lambda)$ and $n(\lambda)$ are statistically independent, we can write the covariance function of $y(\lambda)$ as

$$K_y(\lambda,u) = K_s(\lambda,u) + K_n(\lambda,u), \quad \lambda_1 \leq \lambda,u \leq \lambda_2. \tag{2-24}$$

Since it is assumed also that $n(\lambda)$ is gaussian white noise, its covariance function,

$$K_n(\lambda,u) = \frac{N_0}{2} \delta(\lambda-u) \tag{2-25}$$

is not square integrable. It is necessary to consider the ramifications of this for the selection of the basis functions $\{\theta_i(\lambda) \; ; \; i=1,2,\ldots\}$. Consider first the use of $K_n(\lambda,u)$ as a kernel for the integral equation (2-20).

$$a_j\theta_j(\lambda) = \int_{\lambda_1}^{\lambda_2} \frac{N_0}{2} \delta(\lambda-u)\theta_j(u)du, \quad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-26}$$

or

$$a_j\theta_j(\lambda) = \frac{N_0}{2} \theta_j(\lambda) \quad \lambda_1 \leq \lambda \leq \lambda_2 \quad . \tag{2-27}$$

The implication is that equation (2-26) is satisfied for any set of orthonormal basis functions with corresponding eigenvalues $a_j = \frac{N_0}{2}$. This result is a direct consequence of the fact that $K_n(\lambda,u)$ is a delta function. Hence, we may just as well use $\theta_j(\lambda) = \phi_j(\lambda)$ to represent the received process $y(\lambda)$. We need to make the following clarification here. From Mercer's theorem [V1, p. 181] we need a complete orthonormal

set $\{\phi_j(\lambda);\ j=1,2,\ldots\}$ to represent the covariance function of the white noise. If $K_s(\lambda,u)$ is not positive definite, the eigenfunctions $\{\phi_j(\lambda),\ j=1,\ldots\}$ will not form a complete orthonormal set [V1, p. 181]. In this case the eigenfunctions can be augmented by a sufficient additional number of orthogonal functions to form a complete orthonormal set. The importance of obtaining a complete orthonormal set is that such a set can be used to expand any deterministic square integrable function. The necessity for assuring ourselves that $\{\phi_j(\lambda);\ j=1,2,\ldots\}$ is complete will be clear later. We can be content with the assurance that even if $K_s(\lambda,u)$ is not positive definite, we can still obtain a set $\{\phi_j(\lambda);\ j=1,2,\ldots\}$ that is complete.

Thus, the integral equation that defines the eigenfunctions and eigenvalues for $y(\lambda)$ is

$$b_j\phi_j(\lambda) = \int_{\lambda_1}^{\lambda_2} K_y(\lambda,u)\phi_j(u)\,du$$

$$= \int_{\lambda_1}^{\lambda_2} \left[K_s(\lambda,u) + \frac{N_0}{2}\,\delta(\lambda-u)\right]\phi_j(u)\,du$$

$$= \frac{N_0}{2}\,\phi_j(\lambda) + \int_{\lambda_1}^{\lambda_2} K_s(\lambda,u)\phi_j(u)\,du, \qquad \lambda_1 \le \lambda \le \lambda_2 \qquad (2\text{-}28)$$

If we use

$$b_j = a_j + \frac{N_0}{2} \qquad (2\text{-}29)$$

we have the original integral equation (2-20) again. The implication is

that we may represent the received process with the eigenfunctions $\{\phi_j(\lambda);\ j=1,2,\ldots\}$ and the eigenvalues $\{a_i + \frac{N_0}{2},\ j=1,2,\ldots\}$. That is, we may write

$$y(\lambda) = \sum_{j=1}^{\infty} y_i \phi_i(\lambda), \quad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-30}$$

(l.i.m. is implicit here).

Hence, $y(\lambda)$ is represented by gaussian random variables $\{y_i;\ i=1,\ldots\}$ that are uncorrelated and thus statistically independent. The correlation of y_i and y_j, $i,j = 1,2,\ldots$ is given by

$$E\left[y_i y_j\right] = (a_j + \frac{N_0}{2})\,\delta_{ij} \quad . \tag{2-31}$$

The process $y(\lambda)$ uniquely determines $Y = \{y_1, y_2, \ldots\}$. By using the eigenfunctions $\{\phi_j(\lambda);\ j=1,\ldots\}$ we have a unique representation of that portion of $y(\lambda)$ in the signal space of $s(\lambda)$. By the independence property of average information, the portion of $y(\lambda)$ not in the signal space of $s(\lambda)$ is irrelevant to the average information in $y(\lambda)$ about $s(\lambda)$. Thus, we have achieved the goal of representing the processes $s(\lambda)$ and $y(\lambda)$ by uniquely defined sets of random variables $\{s_i,\ i=1,2,\ldots\}$ and $\{y_i,\ i=1,2,\ldots\}$. These sets of random variables are now used to write an appropriate expression for average information.

The covariance matrix for the random variables $\{s_i;\ i=1,2,\ldots\}$ is a diagonal matrix with the ith diagonal element given by $E\left[s_i^2\right] = a_i$. Similarly, the covariance matrix for the random variables $\{y_i;\ i=1,2,\ldots\}$ is diagonal with the ith diagonal element given by

$E\left[y_i^2\right] = a_i + \frac{N_0}{2}$. Also since we noted that the noise process $n(\lambda)$ could be represented by $\{\phi_i(\lambda);\ i=1,2,\ldots\}$, we can define a set of random varibles for the noise that have a diagonal covariance matrix with the ith diagonal element given by $E\left[n_i^2\right] = \frac{N_0}{2}$.

It is fairly easy to show that a set of gaussian random variables $\{y_i = s_i + n_i,\ i=1,2,\ldots\}$, have average information about the set of gaussian random variables $\{s_i,\ i=1,2,\ldots\}$ given by [S1, Theorem 16]

$$I(S,Y) = \frac{1}{2} \log \left[\frac{\det C_y}{\det C_n}\right] \tag{2-32}$$

where C_y and C_n are the covariance matrices of $\{y_i;\ i=1,2,\ldots\}$ and $\{n_i = \frac{N_0}{2};\ i=1,2,\ldots\}$ respectively. Since C_y and C_n are diagonal, we can write

$$\det C_y = \prod_{i=1}^{\infty} (a_j + \frac{N_0}{2}) \tag{2-33}$$

and

$$\det C_n = \prod_{i=1}^{\infty} \frac{N_0}{2} \quad . \tag{2-34}$$

Thus, we can write

$$\frac{\det C_y}{\det C_n} = \prod_{j=1}^{\infty} (1 + \frac{2}{N_0} a_j) \quad . \tag{2-35}$$

Hence, the average information can be written as:

$$I(S,Y) = \frac{1}{2} \sum_{j=1}^{\infty} \log\left[1 + \frac{2a_j}{N_0}\right] \quad . \tag{2-36}$$

The average information is now written in a form such that calculation may be carried out in principle. Furthermore, the average information can be approximated by using only the first n largest eigenvalues in the summation. This is reminiscent of the feature selection problem in pattern recognition. Thus average information might be useful as a feature selection criterion.

The present formulation of the average information offers insight but still is not in an easily calculable form. The next section is concerned with deriving a more useful formulation for the average (mutual) information. This form will also offer more insight into the idea of using average information to study parameters of multispectral scanners.

4. The Relation Between Average Information and the Wiener-Hopf Optimum Filter Problem

This section will show the relationship between the average information in the received process $y(\lambda)$ about the spectral process $s(\lambda)$ and the Wiener-Hopf optimum filter problem. As is well known, the Wiener-Hopf optimum filter gives the optimum (in the mean-square sense) linear estimate of a process that is corrupted by an independent, additive noise process. In terms of the spectral processes of immediate concern, we observe

$$y(\lambda) = s(\lambda) + n(\lambda) \quad , \quad \lambda_1 \leq \lambda \leq \lambda_2 \quad . \tag{2-37}$$

We then pass the spectral process $y(\lambda)$ through a linear filter to obtain

an optimal estimate $\hat{s}(\lambda)$ of $s(\lambda)$. This estimation technique may be described by the equation

$$\hat{s}(\lambda) = \int_{\lambda_1}^{\lambda} h(\lambda,u)y(u)du, \quad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-38}$$

where $h(\lambda,u)$ is the impulse response of the optimal filter such that $E\left[(s(\lambda) - \hat{s}(\lambda))^2\right]$ is minimized. It is straightforward to show [V1, p. 198-204] that under our assumptions concerning the spectral processes $y(\lambda)$ and $s(\lambda)$ the optimum filter must satisfy

$$\int_{\lambda_1}^{\lambda_2} K_s(u,v)h(\lambda,v)dv + \frac{N_0}{2} h(\lambda,u) = K_s(\lambda,u), \qquad \begin{matrix} \lambda_1 \leq \lambda \leq \lambda_2 \\ \lambda_1 < u < \lambda_2 \end{matrix} \tag{2-39}$$

where $K_s(\lambda,u)$ is the covariance function of $s(\lambda)$ and $\frac{N_0}{2}$ is the spectral level of the noise process. This above relation is a form of the famous Wiener-Hopf equation.

It is now possible to put $h(\lambda,u)$ in a form that is convenient to show its relation to average information. Since the eigenfunctions $\{\phi_i(\lambda), i=1,2,...\}$ form a complete orthonormal set, it is possible to write $h(\lambda,u)$ in a series expansion of the form

$$h(\lambda,u) = \sum_{i=1}^{\infty} h_i\phi_i(\lambda)\phi_i(u), \qquad \begin{matrix} \lambda_1 \leq \lambda \leq \lambda_2 \\ \lambda_1 < u < \lambda_2 \end{matrix} . \tag{2-40}$$

From Mercer's theorem [V1, p. 181] we can write

$$K_s(\lambda,u) = \sum_{i=1}^{\infty} a_i \phi_i(\lambda)\phi_i(u), \qquad \lambda_1 \leq \lambda,u \leq \lambda_2 \quad . \tag{2-41}$$

Substitute (2-40) and (2-41) into (2-39) to obtain

$$\frac{N_0}{2} \sum_{i=1}^{\infty} h_i \phi_i(\lambda)\phi_i(u) + \int_{\lambda_1}^{\lambda_2} \sum_{i=1}^{\infty} a_i \phi_i(u)\phi_i(v) \sum_{j=1}^{\infty} h_j \phi_j(\lambda)\phi_j(v)\,dv$$

$$= \sum_{i=1}^{\infty} a_i \phi_i(\lambda)\phi_i(u) \quad . \tag{2-42}$$

Using the orthonormality property of the eigenfunctions, the above equation can be rewritten as

$$\sum_{i=1}^{\infty} h_i\left(a_i + \frac{N_0}{2}\right) \phi_i(\lambda)\phi_i(u) = \sum_{i=1}^{\infty} a_i \phi_i(\lambda)\phi_i(u) \quad . \tag{2-43}$$

It is seen that if

$$h_i = \frac{a_i}{a_i + \frac{N_0}{2}} \qquad i=1,2,\ldots \tag{2-44}$$

then the equality of equation (2-43) is evident. Hence, expand $h(\lambda,u)$ in a series as

$$h(\lambda,u) = \sum_{i=1}^{\infty} \left[\frac{a_i}{a_i + \frac{N_0}{2}}\right] \phi_i(\lambda)\phi_i(u), \qquad \begin{matrix} \lambda_1 \leq \lambda \leq \lambda_2 \\ \lambda_1 < u < \lambda_2 \end{matrix} \quad . \tag{2-45}$$

This representation of $h(\lambda,u)$ is found to be useful in relating the Wiener filter impulse response to the average information in $y(\lambda)$ about

$s(\lambda)$.

We first manipulate some of the basic equations into a useful form. The equations to be considered are reproduced below for easy reference.

$$a_j \phi_j(\lambda) = \int_{\lambda_1}^{\lambda_2} K_s(\lambda,u)\phi_j(u)du \ , \quad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-20}$$

$$\int_{\lambda_1}^{\lambda_1+\ell} \phi_j^2(\lambda)d\lambda = 1 \tag{2-21}$$

$$I(S,Y) = \frac{1}{2} \sum_{j=1}^{\infty} \log\left[1 + \frac{2a_j}{N_0}\right] \ . \tag{2-36}$$

Note that in (2-21), $\lambda_2 = \lambda_1 + \ell$ since our major interest is in the spectral response interval $\ell = \lambda_2 - \lambda_1$. The first manipulation is the differentiation of I(S,Y) with respect to ℓ. This is

$$\frac{dI(S,Y)}{d\ell} = \frac{1}{2} \sum_{j=1}^{\infty} \frac{d}{d\ell}\left[\log(1 + \frac{2}{N_0} a_j)\right]$$

$$= \frac{1}{2} \sum_{j=1}^{\infty} (1 + \frac{2}{N_0} a_j)^{-1} \frac{2}{N_0} \frac{da_j}{d\ell}$$

or

$$\frac{dI(S,Y)}{d\ell} = \frac{1}{2} \sum_{j=1}^{\infty} \left[\frac{\frac{2}{N_0} \frac{da_j}{d\ell}}{1 + \frac{2}{N_0} a_j} \right] . \qquad (2\text{-}46)$$

From the above equation it is seen that an expression for $\frac{da_j}{d\ell}$ is needed. In order to obtain this derivative, first multiply both sides of (2-20) by $\phi_j(\lambda)$ and integrate over the interval $\left[\lambda_1, \lambda_1 + \ell\right]$. This gives

$$a_j \int_{\lambda_1}^{\lambda_1+\ell} \phi_j^2(\lambda)\, d\lambda = \int_{\lambda_1}^{\lambda_1+\ell} \left[\int_{\lambda_1}^{\lambda_1+\ell} K_s(\lambda,u)\phi_j(u)\, du \right] \phi_j(\lambda)\, d\lambda \qquad (2\text{-}47)$$

and using the normalization criterion (2-21) we obtain

$$a_j = \int_{\lambda_1}^{\lambda_1+\ell} \left[\int_{\lambda_1}^{\lambda_1+\ell} K_s(\lambda,u)\phi_j(u)\, du \right] \phi_j(\lambda)\, d\lambda \quad . \qquad (2\text{-}48)$$

Now take the derivative of (2-48) with respect to ℓ .

$$\frac{da_j}{d\ell} = \int_{\lambda_1}^{\lambda_1+\ell} \Bigg[K_s(\lambda,\lambda_1+\ell)\phi_j(\lambda_1+\ell) + \int_{\lambda_1}^{\lambda_1+\ell} K_s(\lambda,u) \frac{\partial \phi_j(u)}{\partial \ell}\, du \Bigg] \phi_j(\lambda)\, d\lambda$$

$$+ \int_{\lambda_1}^{\lambda_1+\ell} \Big[K_s(\lambda_1+\ell,u)\phi_j(\lambda_1+\ell) +$$

$$+\int_{\lambda_1}^{\lambda_1+\ell} K_s(\lambda,u)\,\frac{\partial\phi_j(\lambda)}{\partial\ell}\,d\lambda\Bigg]\phi_j(u)\,du \quad . \tag{2-49}$$

This can be simplified to give

$$\frac{da_j}{d\ell} = 2\phi_j(\lambda_1+\ell)\int_{\lambda_1}^{\lambda_1+\ell} K_s(\lambda_1+\ell,u)\phi_j(u)\,du$$

$$+\,2\int_{\lambda_1}^{\lambda_1+\ell}\left[\int_{\lambda_1}^{\lambda_1+\ell} K_s(\lambda,u)\phi_j(u)\,du\right]\frac{\partial\phi_j(\lambda)}{\partial\ell}\,d\lambda \quad . \tag{2-50}$$

We now make the observation that equation (2-20) can be written as

$$a_j\phi_j(\lambda_1+\ell) = \int_{\lambda_1}^{\lambda_1+\ell} K_s(\lambda_1+\ell,u)\phi_j(u)\,du \quad . \tag{2-51}$$

Using equations (2-51) and (2-20) in (2-50) we obtain

$$\frac{da_j}{d\ell} = 2a_j\phi_j^2(\lambda_1+\ell)$$

$$+\,2a_j\int_{\lambda_1}^{\lambda_1+\ell}\phi_j(\lambda)\,\frac{\partial\phi_j(\lambda)}{\partial\ell}\,d\lambda \quad . \tag{2-52}$$

A simplification of the integral expression on the right hand side of (2-52) is still needed. A useful expression may be obtained from the normalization expression (2-21). Differentiate (2-21) with respect to ℓ.

$$\frac{d}{d\ell}\int_{\lambda_1}^{\lambda_1+\ell} \phi_j^2(\lambda)d\lambda = 0 \tag{2-53}$$

or

$$0 = \phi_j^2(\lambda_1+\ell) + 2\int_{\lambda_1}^{\lambda_1+\ell} \phi_j(\lambda) \frac{\partial \phi_j(\lambda)}{\partial \ell} d\lambda \tag{2-54}$$

The integral term in (2-54) is the same as in (2-52). Hence, using (2-54) in (2-52) we obtain

$$\frac{da_j}{d\ell} = 2a_j\phi_j^2(\lambda_1+\ell) - a_j\phi_j^2(\lambda_1+\ell) \tag{2-55}$$

or

$$\frac{da_j}{d\ell} = a_j\phi_j^2(\lambda_1+\ell) \quad . \tag{2-56}$$

This is the expression for $\frac{da_j}{d\ell}$ that is needed in equation (2-46). Making the appropriate substitution we obtain

$$\frac{dI(S,Y)}{d\ell} = \frac{1}{2}\sum_{j=1}^{\infty}\left[\frac{\frac{2}{N_0}a_j\phi_j^2(\lambda_1+\ell)}{(1+\frac{2}{N_0}a_j)}\right] \quad . \tag{2-57}$$

Now compare this expression with the expression obtained for the Wiener optimal filter response (2-45). It is clear that the relation between (2-45) and (2-57) is

$$\frac{dI(S,Y)}{d\ell} = \frac{1}{2} \cdot h(\lambda_1+\ell, \lambda_1+\ell) \quad . \tag{2-58}$$

Thus it is seen that since ℓ varies from 0 to $\lambda_2 - \lambda_1$ we may make a sim-

plifying change of variables and write

$$I(S,Y) = \frac{1}{2}\int_{\lambda_1}^{\lambda_2} h(\lambda,\lambda)d\lambda \quad . \tag{2-59}$$

Thus we have a simple relationship between the average information in $y(\lambda)$ about $s(\lambda)$ and the Wiener optimum filter. The expression $h(\lambda,\lambda)$ is the weighting that should be given to $y(\lambda)$ at wavelength λ in order to obtain the optimum mean-square estimate of $s(\lambda)$ at wavelength λ. It is interesting that there is a relationship between the mean-square estimation of $s(\lambda)$ from the observation $y(\lambda)$ and the expression for average information as simple as the one given above.

There are, however, major drawbacks in the relations just derived. The first major problem is that the covariance function $K_s(\lambda,u)$ must be known in order to solve the Wiener-Hopf equation in an analytical manner. In general the covariance function of a spectral scene is not known in analytical form. An estimate of the covariance function can be made, but this estimate may not be in a useful analytical form and may not be positive definite. Furthermore, estimation errors may add to the difficulties of discerning a useful analytic form for $K_s(\lambda,u)$. The second major problem lies in actually solving the Wiener-Hopf equation even under the assumption that a functional form for $K_s(\lambda,u)$ is known. If the spectral processes are stationary and posess rational power spectral densities, then the Wiener-Hopf equation can be solved. However, stationarity cannot necessarily be assumed for the spectral processes. The solution of the Weiner-Hopf equation for nonstationary covariance functions is considerably more difficult.

5. Computation Considerations

The above considerations indicate that a more useful technique for determination of the Wiener filter impulse response is needed. A technique that is useful for this computation may be found by examining the relationship between Wiener filter theory and Kalman filter theory. These two theories are really different viewpoints of the same problem. Of the two, Kalman filtering offers much more computational capability. The relationship between Wiener and Kalman filter theory was shown by Kalman and Bucy [K1] and Kalman [K2].

Since the purpose of this section is to study computation techniques for average information, and computation is most easily carried out in discrete form, we shall first recast the previous expressions in a discrete formulation. That is, we shall study the problem in terms of discrete wavelengths rather than continuous wavelengths. Thus, the spectral process

$$y(\lambda) = s(\lambda)+n(\lambda) \qquad \lambda_1 \leq \lambda \leq \lambda_2$$

is written as

$$y(k) = s(k) + n(k) \qquad k \in [\lambda_1, \lambda_2] \tag{2-60}$$

where k is an integer corresponding to a discrete wavelength in the wavelength interval of interest. The white noise process $n(\lambda)$ then becomes a sequence, $n(k)$, of independent, identically distributed zero mean gaussian random variables with variance $\frac{N_0}{2}$. Next, consider the description of the spectral process $s(\lambda)$. Since we have assumed that $s(\lambda)$ is a gaussian random process, a reasonable model for $s(\lambda)$ is a

linear dynamic system driven by an independent gaussian random process. That is, the process $s(\lambda)$ is assumed to be described by a linear vector state variable form. The linear vector state variable form is written as:

$$\dot{\underline{s}}(\lambda) = \underline{A}\,\underline{s}(\lambda) + \underline{B}w(\lambda) \tag{2-61}$$

where

$$\underline{s}(\lambda) = \begin{bmatrix} s_1(\lambda) \\ s_2(\lambda) \\ \cdot \\ \cdot \\ \cdot \\ s_n(\lambda) \end{bmatrix}$$

and $s_1(\lambda)$, $s_2(\lambda)$, ..., $s_n(\lambda)$ are the state variables. More discussion of the state variables is given later.

$\underline{A}$ is an (n x n) matrix

$$\dot{\underline{s}}(\lambda) = \frac{d}{d\lambda}\underline{s}(\lambda) = \begin{bmatrix} \frac{d}{d\lambda} s_1(\lambda) \\ \frac{d}{d\lambda} s_2(\lambda) \\ \cdot \\ \cdot \\ \cdot \\ \frac{d}{d\lambda} s_n(\lambda) \end{bmatrix}$$

B is an n x 1 matrix

$w(\lambda)$ = an independent gaussian random process

n is order of the state variable system.

The concepts of state and state variable formulations may be reviewed in Schwartz and Friedland [S2, Chapt. 2]. Kalman [K2], and Schwartz and Friedland [S2, p. 125-127] also indicate the manner in which a system given by (2-61) may be written as a discrete-time dynamic system. By analogy the discrete form of the spectral process $s(\lambda)$ is given as:

$$\underline{s}(k+1) = \underline{\Phi}\underline{s}(k) + \underline{\Gamma}w(k) \qquad k\varepsilon\left[\lambda_1, \lambda_2\right] \tag{2-62}$$

where

$$\underline{s}(k+1) = \begin{bmatrix} S_1(k+1) \\ S_2(k+1) \\ \cdot \\ \cdot \\ \cdot \\ S_n(k+1) \end{bmatrix}$$

$\underline{\Phi}$ is an (n x n) matrix.

$\underline{\Gamma}$ is an (n x 1) vector.

$w(k)$ = a discrete inependent gaussian random process with zero mean and variance $V_w(k)\delta(k-j)$

k is an integer that corresponds to a discrete wavelength $\lambda\varepsilon\left[\lambda_1, \lambda_2\right]$. With this representation for $\underline{s}(k)$ we can write the discrete form for $y(\lambda)$ as

$$y(k) = \underline{H}^T \underline{s}(k) + n(k) \qquad k \in [\lambda_1, \lambda_2] \tag{2-63}$$

where

$\underline{H}^T \underline{s}(k) = s(k)$ is the relationship between the state variable formulation and the spectral response $s(\lambda)$ for the λ that corresponds to k.

This discrete form for representing the spectral processes is much more amenable to digital computation than the Wiener-Hopf form. It should be recalled that the solution of the Wiener-Hopf equation yields the optimum (in the minimum mean-square sense) filter for the estimate $\underline{\hat{s}}(\lambda)$ of $\underline{s}(\lambda)$. The estimate thus may be written

$$\underline{\hat{s}}(\lambda) = \int_{\lambda_1}^{\lambda} \underline{h}(\lambda,u) y(u) du, \qquad \lambda_1 \leq \lambda \leq \lambda_2 \quad . \tag{2-64}$$

The Kalman-Bucy estimation equations can be derived from (2-64). Furthermore, the same equations can be derived using other techniques. Thus the equivalence of the Wiener-Hopf techniques and the Kalman-Bucy techniques have been firmly established. The reader is referred to Sage and Melsa [S3, Chapter 7] for details. In addition, the complete Kalman filter algorithm is included in Appendix I for reference. In these derivations, it is a matter of course to obtain the equation for $\underline{h}(\lambda,\lambda)$ as

$$\underline{h}(\lambda,\lambda) = \underline{V}_{\underline{\tilde{s}}}(\lambda) \cdot \underline{H} \cdot \underline{R}^{-1}(\lambda) \tag{2-65}$$

where

$$\underline{\tilde{s}}(\lambda) = \underline{s}(\lambda) - \underline{\hat{s}}(\lambda) \tag{2-66}$$

$$\underset{\tilde{\underline{s}}}{\underline{V}}(\lambda) = \mathrm{Var}\left[\underline{\tilde{s}}(\lambda)\right] = \mathrm{Var}\left[\underline{s}(\lambda) - \underline{\hat{s}}(\lambda)\right] \tag{2-67}$$

and

$$\underline{R}(\lambda) = \mathrm{Var}\left[\underline{n}(\lambda)\right] \quad . \tag{2-68}$$

In the present research, $\underline{R}(\lambda)$ is a scalar. The Kalman filter algorithms provide a natural and efficient technique for the computation of the estimation error variance $\underset{\tilde{\underline{s}}}{\underline{V}}(\lambda)$. The algorithms given above are the same for the discrete case when the appropriate substitutions are made in the relevant variables. The results are given below

$$\underline{h}(k,k) = \underset{\tilde{\underline{s}}}{\underline{V}}(k) \cdot \underline{H} \cdot R^{-1}(k) \tag{2-69}$$

where

$$k \in \{\lambda: \lambda_1 \leq \lambda \leq \lambda_2\}$$

are integers corresponding to discrete wavelengths of interest.

$$\underline{\tilde{s}}(k) = \underline{s}(k) - \underline{\hat{s}}(k) \tag{2-70}$$

$$\underset{\tilde{\underline{s}}}{\underline{V}}(k) = \mathrm{Var}\left[\underline{\tilde{s}}(k)\right] = \mathrm{Var}\left[\underline{s}(k) - \underline{\hat{s}}(k)\right] \tag{2-71}$$

and

$$\underline{R}(k) = \mathrm{Var}[\underline{n}(k)] \quad . \tag{2-72}$$

It should be noted that $\underline{h}(k,k)$ is in vector form to conform with the state variable models used in the Kalman filter algorithm. The relationship between $\underline{h}(k,k)$ and $h(k,k)$ can be discerned from the state variable signal model. This topic will be covered in more detail in Chapter III which is concerned with modelling the spectral process.

Thus we are able to use Kalman filtering techniques to compute $h(k,k)$ and hence average information. The average information in $y(\lambda)$ about $s(\lambda)$ is given by

$$I(S,Y) = \frac{1}{2} \sum_{k\varepsilon[\lambda_1,\ \lambda_2]} h(k,k) \tag{2-73}$$

The Kalman filter computation technique has several advantages over the Wiener-Hopf approach. The most obvious advantage is the digital computer compatibility of the Kalman filter technique. The Wiener-Hopf equation is easily solved in only those cases for which the analytical form of $K_s(\lambda,u)$ is fairly simple. For other cases, solution of the Wiener-Hopf equation ranges from difficult to extremely difficult to solve. The second advantage is that it is not necessary to have explicit knowledge of the form of $K_s(\lambda,u)$ in order to use Kalman filtering techniques. This obviates the need for estimation of $K_s(\lambda,u)$. Another advantage is that $n(k)$ need not be "samples" of a white noise process. It is possible to use noise models that have a linear dynamic structure. Thus, noise models with nonwhite power spectral densities may be implemented.

The major remaining question is the method by which the signal model $\underline{s}(k+1)$ is obtained. Specifically, it is necessary to obtain the matrix $\underline{\Phi}$ and the vector $\underline{\Gamma}$. If the spectral process $s(\lambda)$ is stationary and has a rational power spectral density, then $\underline{\Phi}$ and $\underline{\Gamma}$ may be determined from knowledge of $K_s(\lambda,u)$ [V1, pp. 516-526]. In many physical situations $K_s(\lambda,u)$ is not known. The parameters $\underline{\Phi}$ and $\underline{\Gamma}$ must then be estimated from whatever empirical data is at hand. This is a problem which has been studied extensively in the area of system identification. The use of these techniques for modelling the spectral process $s(\lambda)$ (and hence $\underline{s}(\lambda)$) is the topic of concern for Chapter III. Thus we will leave this problem for later consideration.

6. Further consideration of the Relation Between Average Information and Optimum Mean-Square Filtering

In this section, the relationship of optimum mean-square filtering to average information is considered in a somewhat more direct manner than in the previous sections. The formulations are in the state variable viewpoint in order to be consistent with the final approach to average information computation in the previous section. We shall specifically be concerned with showing that the estimate $\underline{\hat{s}}(k)$ of $\underline{s}(k)$ given by

$$\underline{\hat{s}}(k) = E\left[\underline{s}_k / Y_k\right] \tag{2-74}$$

where

$$Y_k = \{y(k) : k\varepsilon[\lambda_1, \lambda_2]\} \tag{2-75}$$

= the observed spectral process.

is the estimate such that the average information in $\underline{\hat{s}}(k)$ about $\underline{s}(k)$, $I(\underline{s}(k), \underline{\hat{s}}(k))$ is maximized. Then since

$$s(k) = \underline{H}^T \underline{s}(k) \tag{2-76}$$

and

$$\hat{s}(k) = \underline{H}^T \underline{\hat{s}}(k) \tag{2-77}$$

we have the estimate $\hat{s}(k)$ of $s(k)$ that maximizes the average information $I(s(k), \hat{s}(k))$.

The estimate $\underline{\hat{s}}(k)$ is a natural result of the Kalman filtering technique. Thus, a by-product of the computation for $I(S,Y)$ is the estimate $\underline{\hat{s}}(k)$.

In order to demonstrate the above statements, it is first necessary to develop some intermediate results. We first show that the average information between the estimate $\underline{\hat{s}}(k)$ and the estimation error $\underline{\tilde{s}}(k) = \underline{s}(k) - \underline{\hat{s}}(k)$ is zero. That is,

$$I(\underline{\hat{s}}(k), \underline{\tilde{s}}(k)) = 0 \quad . \tag{2-78}$$

Now $I(\underline{\hat{s}}(k), \underline{\tilde{s}}(k)) = 0$ if and only if $\underline{\hat{s}}(k)$ and $\underline{\tilde{s}}(k)$ are statistically independent [G1, p. 24]. But, based on our initial definitions and assumptions, $\underline{\hat{s}}(k)$ and $\underline{\tilde{s}}(k) = \underline{s}(k) - \underline{\hat{s}}(k)$ are gaussian random variables. Hence, it is sufficient to observe that $\underline{\hat{s}}(k)$ and $\underline{\tilde{s}}(k)$ are uncorrelated.

This is easily seen from

$$E\left[\tilde{\underline{s}}(k)^T \hat{\underline{s}}(k)\right] = E\left[E\left[\tilde{\underline{s}}(k)^T / Y_k\right] \hat{\underline{s}}(k)\right] \tag{2-79}$$

However,

$$E\left[\tilde{\underline{s}}(k)^T / Y_k\right] = E\left[(\underline{s}(k)^T - \hat{\underline{s}}(k)^T) / Y_k\right]$$

$$= E\left[\underline{s}(k)^T / Y_k\right] - \hat{\underline{s}}(k)^T = 0 \quad .$$

Hence

$$E\left[\tilde{\underline{s}}(k)^T \hat{\underline{s}}(k)\right] = 0 \quad .$$

Thus $\tilde{\underline{s}}(k)$ and $\hat{\underline{s}}(k)$ are uncorrelated and hence independent. Therefore, it is clear that (2-78) is true.

Next, some entropy relations are needed. The relations to be shown are

$$H(\underline{s}(k)/\hat{\underline{s}}(k)) = H(\tilde{\underline{s}}(k)/\hat{\underline{s}}(k)) = H(\tilde{\underline{s}}(k)) \tag{2-80}$$

First consider the random variables $\underline{Z} = \underline{s}(k) - \hat{\underline{s}}(k)$ and $\underline{W} = \hat{\underline{s}}(k)$. The density function $P_{\underline{ZW}}(\underline{z},\underline{w})$ is to be determined in terms of the density function $P_{\underline{S\hat{S}}}(\underline{s}(k)\hat{\underline{s}}(k))$. It is easily shown [P1, p. 204] that

$$p(\underline{s}(k) - \hat{\underline{s}}(k), \hat{s}(k)) = P_{\underline{ZW}}(\underline{z},\underline{w}) =$$

$$= P_{\underline{S\hat{S}}}(\underline{z}+\underline{w},\underline{w}) = P_{\underline{S\hat{S}}}(s(k),\hat{\underline{S}}(k)) \quad . \tag{2-81}$$

Thus, since $\tilde{\underline{s}}(k) = \underline{s}(k) - \hat{\underline{s}}(k)$ we can write

$$p(\underline{s}(k),\underline{\hat{s}}(k)) = p(\underline{s}(k) - \underline{\hat{s}}(k), \underline{\hat{s}}(k))$$

$$= p(\underline{\tilde{s}}(k), \underline{\hat{s}}(k)) \quad . \tag{2-82}$$

So we have

$$p(\underline{s}(k)/\underline{\hat{s}}(k)) = p(\underline{\tilde{s}}(k)/\underline{\hat{s}}(k))$$

and hence

$$H(\underline{s}(k)/\underline{\hat{s}}(k)) = H(\underline{\tilde{s}}(k)/\underline{\hat{s}}(k)) \tag{2-83}$$

Now we have previously shown that

$$I(\underline{\hat{s}}(k), \underline{\tilde{s}}(k)) = 0 \tag{2-78}$$

Hence

$$0 = I(\underline{\hat{s}}(k), \underline{\tilde{s}}(k)) = H(\underline{\tilde{s}}(k)) - H(\underline{\tilde{s}}(k)/\underline{\hat{s}}(k))$$

or

$$H(\underline{\tilde{s}}(k)) = H(\underline{\tilde{s}}(k)/\underline{\hat{s}}(k)) \quad . \tag{2-84}$$

Now, we can write the average information $I(\underline{s}(k), \underline{\hat{s}}(k))$ as

$$I(\underline{s}(k), \underline{\hat{s}}(k)) = H(\underline{s}(k)) - H(\underline{s}(k)/\underline{\hat{s}}(k)) \quad . \tag{2-85}$$

But using (2-83) and (2-84) in (2-85) we have

$$I(\underline{s}(k), \underline{\hat{s}}(k)) = H(\underline{s}(k)) - H(\underline{\tilde{s}}(k)) \quad . \tag{2-86}$$

This is a very useful result. It shows that the average information in the estimate $\underline{\hat{s}}(k)$ of the state $\underline{s}(k)$ of the spectral process is directly

related to the entropy of the estimation error $\underline{\tilde{s}}(k) = \underline{s}(k) - \underline{\hat{s}}(k)$. Since for a given observation the state $\underline{s}(k)$ is fixed, it is clear that to maximize $I(\underline{s}(k), \underline{\hat{s}}(k))$ it is sufficient to minimize $H(\underline{\tilde{s}}(k))$.

Now since $\underline{\tilde{s}}(k)$ is gaussian, it is straightforward to compute the entropy $H(\underline{\tilde{s}}(k))$ as

$$H(\underline{\tilde{s}}(k)) = \frac{1}{2} \log(2\pi e)^n |\underline{P}_k| \tag{2-87}$$

where

$$|\underline{P}_k| = \det\left[E\left[\underline{\tilde{s}}(k)\underline{\tilde{s}}^T(k)\right]\right] .$$

Hence it is clear that to minimize $H(\underline{\tilde{s}}(k))$ it is sufficient to minimize $\underline{P}_k$. Tomita, Omatu and Soeda [T1] show directly that this is accomplished by the Kalman filter technique. It is sufficient for our purposes to note that we already have used the Kalman filter algorithm and it is known [S3, Chapter 7] that it gives the minimum error variance estimate for our case of assumed gaussian statistics.

Thus, it is seen that the Kalman filter algorithm produces an estimate $\hat{s}(k)$ of the spectral response $s(k)$ that is optimum in terms of average information. The optimum mean square estimate $\hat{s}(k)$ is thus a natural by-product that is consistent with the concept of using average information to study parameters of multispectral scanners.

In conclusion, this chapter develops the notion of average information in the received spectral process $y(\lambda)$ about the reflectance spectral process $s(\lambda)$. Furthermore, a technique for computation of average information has been developed. The relationship between optimum mean

square estimation and average information for the current problem is also shown.

Chapter III

Model Identification, Selection, and Validation Techniques

1. Introduction

This chapter is concerned with finding analytical models that adequately represent the spectral response process of scenes observed by multispectral scanners. A major requirement for the models is compatibility with the computational techniques discussed in the previous chapter. Specifically, we are interested in obtaining the necessary parameters to represent the models in the state variable forms discussed in Chapter II. In Chapter I, the division of the reflectance spectral response into bands is discussed. The technique for constructing models must, therefore, be sensitive to different characteristics of the spectral response process in different spectral bands. Hence, the techniques developed in this chapter are motivated by the above constraints.

Very useful techniques for model construction can be drawn from the subject area generally known as time series analysis. References for time series analysis are numerous with major works by Anderson [A4], Box and Jenkins [B1], and Kashyap and Rao [K3]. The reference to time is generally a misnomer in that time merely represents an indexing variable. We shall, of course, use wavelength as our indexing variable. We first discuss the models that are used in this research to represent spectral response process of scenes.

2. Models Used to Represent Spectral Scenes.

Recall that the spectral response of a scene is being considered as a portion of a realization of a stochastic process in wavelength. Thus, the form of the models used must reflect this consideraion. The models considered in this research are stochastic difference equations having a general form given by

$$S(k) = \sum_{j=1}^{m1} a_j S(k-j) + \sum_{j=1}^{m2} b_j \psi(k-j) + w(k) \qquad (3-1)$$

where

- $S(k)$ is the spectral response at the discrete wavelenth k. It is gaussian with mean and variance determined by the particular structure of (3-1).
- $w(k)$ is a zero mean independent gaussian disturbance with variance ρ.
- $\psi(k-j)$ is a deterministic trend term used to account for certain characteristics of the empirical data. An example is $\psi(k-1)=1.0$, which could be used to account for a nonzero mean in $w(k)$.
- a_j and b_j are unknown constant coefficients to be determined.
- m1 and m2 are constants that determine the dependence of $S(k)$ on preceeding values of the process.

Thus the dynamic nature of the spectral process $S(k)$ is expressed in terms of is own values at lower wavelengths, some possible deterministic

characteristics, and a gaussian independent disturbance. This model formulation though somewhat restricted is sufficient for the purposes of this research.

For completeness, we shall now introduce some asumptions on the model given by (3-1). Since (3-1) represents a linear system, it is fully described by its second order statistical properties. The coefficients a_j and b_j are said to be identifiable if they can be determined from a semi-infinite set of observations $\{S(k);\ 1 \leq k < \infty\}$ such that the difference equation (3-1) uniquely describes the second order properties of the observed process S(k). We shall now state some assumptions that are necessary and sufficient conditions for the identifiability of the coefficients a_j and b_j. The question of identifiability is covered in detail by Kashyap and Rao [K3, Chap. 4]. The assumptions that are used in this research are listed below.

Assumptions

1) w(k), k = 1,2,... is a sequence of zero mean identically distributed, independent gaussian random variables with variance ρ. w(k) is independent of S(k-j) for all $j \geq 1$.

2) Define the unit delay operator D by:

$$Dy(k) = y(k-1) \quad .$$

then (3-1) can be written as

$$S(k)\left[1 - \sum_{j=1}^{m1} a_j D^j\right] = \sum_{j=1}^{m2} b_j \psi(k-j) + w(k) \quad . \qquad (3\text{-}2)$$

The assumption is that all the zeros of the expression

$$A(D) = 1 - \sum_{j=1}^{m1} a_j D^j$$

lie outside the unit circle in the complex plane. This assumption gives assurance that (3-1) is asymptotically covariance stationary. The relation of this assumption to covariance stationarity is explored in detail by Box and Jenkins [B1, Chap. 3].

3) Suppose we have several trend terms denoted by $\psi_i(k), i=1,2,\ldots\ell$. Represent these trend terms by the vector

$$\underline{\psi}(k) = [\psi_1(k),\ldots,\psi_\ell(k)]^T .$$

Then the assumption is $\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \underline{\psi}(k)\underline{\psi}(k)^T$ exists and is positive definite. This assumption gives assurance that the coefficients of most trend terms can be identified. Note, however, that this assumption is invalid for the useful linear trend $\psi(k)=k$. Therefore, a weaker assumption that follows from the above may be useful.

4) The vector of trend terms obeys

$$\sum_{k=1}^{\infty} \sum_{i=1}^{\ell} (\alpha_i \psi_i(k))^2 = \infty \qquad (3\text{-}3)$$

for nonzero $\underline{\alpha} = [\alpha_1,\alpha_2,\ldots\alpha_\ell]^T$. The notation used in (3-3) means that the summation diverges. It is now demonstrated that (3-3) follows from assumption 3. Consider first

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \underline{\psi}(k)\underline{\psi}(k)^T =$$

$$= \lim_{N\to\infty} \begin{bmatrix} \frac{1}{N}\sum_{k=1}^{N} \psi_1(k)^2 & , \frac{1}{N}\sum_{k=1}^{N} \psi_1(k)\psi_2(k) & ,\ldots, & \frac{1}{N}\sum_{k=1}^{N} \psi_1(k)\psi_\ell(k) \\ & \frac{1}{N}\sum_{k=1}^{N} \psi_2(k)^2 & ,\ldots, & \frac{1}{N}\sum_{k=1}^{N} \psi_2(k)\psi_\ell(k) \\ & & & \\ & & & \frac{1}{N}\sum_{k=1}^{N} \psi_\ell(k)^2 \end{bmatrix}$$

Each of the diagonal terms in the above matrix is obviously positive. Hence, by the assumption and the above comment,

$$\prod_{i=1}^{\ell} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \psi_i(k)^2 > 0$$

and bounded. Thus for each i, there exists a $B_i > 0$, such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \psi_i^2(k) = B_i \quad .$$

Now consider $\sum_{k=1}^{\infty} \sum_{i=1}^{\ell} (\alpha_i \psi_i(k))^2$. There is at least one i such that $\alpha_i \neq 0$. Hence,

$$\sum_{k=1}^{\infty} \sum_{i=1}^{\ell} (\alpha_i \psi_i(k))^2 \geq \sum_{k=1}^{\infty} b^2 \psi_i(k)^2 \quad .$$

But since

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \psi_i(k)^2 = B_i$$

we have

$$\lim_{N\to\infty} \sum_{k=1}^{N} \psi_1(k)^2 = \infty \quad .$$

Hence,

$$b^2 \sum_{k=1}^{\infty} \psi_i(k)^2 = \infty \quad ,$$

and the desired relation is shown.

These assmptions will not be mentioned again unless specific need arises.

We shall now list some types of models that are used in the present research. The first type of model is known as the autoregressive model of order m1. It is defined by

$$S(k) = \sum_{j=1}^{m1} a_j \, S(k-j) + w(k) \quad . \tag{3-4}$$

A second type of model is found to be useful for the case of nonzero mean for w(k). This model, called the autoregressive plus constant trend model of order m1, is given by

$$S(k) = \sum_{j=1}^{m1} a_j \, S(k-j) + C + w(k) \quad . \tag{3-5}$$

Note that C corresponds to the coefficient of the trend term

$\psi(k-1) = 1.0$.

The third model type is useful in representing a nonstationary process [K3, Chap. 3]. This model is in the class of integrated autoregressive models of order m1. We shall have occasion to use two cases of this model. The first and more common case is given by

$$\nabla S(k) = \sum_{j=1}^{m1} a_j \nabla S(k-j) + w(k) \qquad (3-6)$$

where

$$\nabla S(k) = S(k) - S(k-1) \quad .$$

The second case of the integrated autoregressive models that will be used is given by

$$\nabla_2 S(k) = \sum_{j=1}^{m1} a_j \nabla_2 S(k-j) + w(k) \qquad (3-7)$$

where

$$\nabla_2 S(k) = S(k) - S(k-2) \quad .$$

These models are shown in the next chapter to give good fits to the spectral response processes under consideration. The models are also easily placed in state variable form. It is recalled from the previous chapter that this form is useful in the computational technique used to obtain the average information in the received spectral process about the spectral response process of the scene. An example for placing one of the above models (3-4) in state variable form may be useful.

Example

Assume that a spectral response is modeled as a third order autoregressive process.

$$S(k) = a_1 S(k-1) + a_2 S(k-2) + a_3 S(k-3) + w(k)$$

Define the state variables

$$S_1(k) = S(k-2)$$

$$S_2(k) = S(k-1)$$

$$S_3(k) = S(k) = a_1 S_3(k-1) + a_2 S_2(k-1) + a_3 S_1(k-1) + w(k)$$

Then we can write in vector and matrix form

$$\underline{S}(k) = \begin{bmatrix} S_1(k) \\ S_2(k) \\ S_3(k) \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a_3 & a_2 & a_1 \end{bmatrix} \quad \begin{bmatrix} S_1(k-1) \\ S_2(k-1) \\ S_3(k-1) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(k)$$

$$= \underline{\Phi}\, \underline{S}(k-1) + \underline{\Gamma} w(k)$$

and

$$S(k) = [0\ 0\ 1]\underline{S}(k) = \underline{H}^T \underline{S}(k) \quad .$$

We shall next consider techniques for estimating the unknown coefficients in (3-1).

3. Estimation Techniques

The estimation techniques that will be used for the model types discussed in the previous section will have two main properties. First,

the form of the estimators depends on the assumed gaussian nature of w(k). The second major property is that the estimation technique must be amenable to computational procedures. That is, it is desired that the estimation algorithm be in a recursive computational form.

Two related estimation techniques are discussed. The first technique is maximum likelihood estimation which does not depend on prior knowledge of the parameters. The second technique considered is Bayesian estimation in which prior knowledge about the parameters may be incorporated. The techniques will be shown to produce similar algorithms for computation.

We shall begin with some preliminary manipulations that are useful in discussing both estimation techniques. Equation (3-1) is recast in the following more compact form.

$$S(k) = \underline{Z}^T(k-1)\underline{\theta} + w(k) \tag{3-8}$$

where

$$\underline{Z}(k-1) = \begin{bmatrix} S(k-1) \\ \cdot \\ \cdot \\ \cdot \\ S(k-m1) \\ \psi(k-1) \\ \cdot \\ \cdot \\ \cdot \\ \psi(k-m2) \end{bmatrix}, \quad \underline{\theta} = \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_{m1} \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_{m2} \end{bmatrix}.$$

We shall assume that at wavelength k = N, we have accumulated the following data.

$$X(N) = \{\underbrace{S(N),\ldots,S(1)}_{X_1}, \underbrace{S(0),\ldots,S(-m1)}_{X_2}\} \qquad (3\text{-}9)$$

X_1 is the portion of the data set from which the estimates are constructed. X_2 is the portion of the data that initializes the dynamics of the spectral response process. Furthermore, let $\underline{\theta}$ be the true parameters of the model and let ρ be the variance of w(k).

A) Maximum Likelihood Estimation

The observations X(N) contain the only empirical data from which we can estimate the parameters $\underline{\theta}$ and ρ. The probability density function of the observations is needed to obtain the estimation scheme. This probability density can be written as

$$P(X(N)) = p(S(N),\ldots,S(-m1))$$

$$= p(S(N)/S(N-1),\ldots,S(-m1))\cdot p(S(N-1)/S(N-2),\ldots,S(-m1))$$

$$\cdot\ldots\cdot p(S(1)/S(0),\ldots,S(-m1))\cdot p(S(0),\ldots,S(-m1)) \qquad (3\text{-}10)$$

The likelihood principle states that the estimates of the parameters $\underline{\theta}$ and ρ are the values of the parameters that maximize the probability density function p(X(N)). This estimate is called the full information maximum likelihood estimate (FIML) by Kashyap and Rao [K3, Chap. 6]. However, the probability density function p(s(0),...s(:m1))

is usually either unknown or in a form that renders (3-10) too complicated to be useful. Thus, the density function $p(S(0),\ldots,S(-m1))$, which represents the effects of the initial observations, will be ignored in this estimation technique. If a large number of observations are available, it is reasonable to expect that these initial observations will not be of overriding influence. If the maximum likelihood estimation techniques are applied to the remaining terms of (3-10), we obtain the estimates called conditional maximum likelihood (CML) estimates by Kashyap and Rao [K3, Chap. 6]. The so-called CML estimation techniques are used in this research.

Due to the initial assumption of the normality of $w(k)$, we can write for each of the conditional probability densities in (3-10)

$$p(S(j)/S(j-1),\ldots,S(-m1)) = \frac{1}{\sqrt{2\pi\rho}} \exp\left[-\frac{1}{2}\frac{(S(j)-Z^T(j-1)\theta)^2}{\rho}\right] \qquad (3\text{-}11)$$

Hence, since we are now only concerned with the contributions of the conditional probability density functions in (3-10), we can write as our likelihood function

$$L(\underline{\theta},\rho,X(N)) = \prod_{k=1}^{N} p(S(k)|S(k-1),\ldots,S(-m1))$$

or

$$L(\underline{\theta},\rho,X(N)) = (2\pi\rho)^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\sum_{k=1}^{N}\frac{(S(k)-Z^T(k-1)\theta)^2}{\rho}\right] \qquad (3\text{-}12)$$

Since we are considering conditional maximum likelihood estimates, the

optimum estimates $\underline{\theta}^*$ and ρ^* of $\underline{\theta}$ and ρ are those values that maximize the likelihood function $L(\underline{\theta},\rho,X(N))$. Equivalently, the logarithm of the likelihood function may be maximized. That is, maximize

$$L_1(\underline{\theta},\rho,X(N)) = \log L(\underline{\theta},\rho,X(N)) =$$

$$= \frac{-N}{2}\log(2\pi\rho) - \frac{1}{2}\left[\sum_{k=1}^{N}\frac{(S(k)-\underline{Z}^T(k-1)\underline{\theta})^2}{\rho}\right] \qquad (3\text{-}13)$$

First obtain the estimate for $\underline{\theta}$ by taking the partial derivative of $L_1(\underline{\theta},\rho,X(N))$ with respect to $\underline{\theta}$ to obtain

$$\frac{\partial L_1(\underline{\theta},\rho,X(N))}{\partial\underline{\theta}} = \frac{\partial}{\partial\underline{\theta}}\left[\frac{-1}{2\rho}\sum_{k=1}^{N}(S(k)-\underline{Z}^T(k-1)\underline{\theta})(S(k)-\underline{Z}^T(k-1)\underline{\theta})\right]$$

$$= \frac{-1}{2\rho}\left[\sum_{k=1}^{N}S(k)\underline{Z}(k-1) - \sum_{k=1}^{N}\underline{Z}(k-1)\underline{Z}^T(k-1)\underline{\theta}\right] = \underline{0}$$

and equate to zero.

Thus the estimate becomes

$$\underline{\theta}^*(N) = \left[\sum_{k=1}^{N}\underline{Z}(k-1)\underline{Z}^T(k-1)\right]^{-1}\sum_{k=1}^{N}\underline{Z}(k-1)S(k). \qquad (3\text{-}14)$$

Next, differentiate (3-13) with respect to ρ to obtain the estimate ρ^*.

$$\frac{\partial L_1(\underline{\theta},\rho/X(N))}{\partial\rho} = \frac{-N}{2}\frac{1}{\rho} + \frac{1}{2\rho^2}\sum_{k=1}^{N}(S(k)-\underline{Z}^T(k-1)\underline{\theta})^2 = 0$$

or

$$\rho^*(N) = \frac{1}{N}\sum_{k=1}^{N}(S(k)-\underline{Z}^T(k-1)\underline{\theta})^2 \qquad (3\text{-}15)$$

Since we wish to estimate ρ from the observations above, it is necessary to solve the equations for $\frac{\partial L_1}{\partial \underline{\theta}}$ and $\frac{\partial L_1}{\partial \rho}$ jointly to obtain

$$\rho^*(N) = \frac{1}{N}\sum_{k=1}^{N}(S(k)-\underline{Z}^T(k-1)\underline{\theta}^*)^2 \qquad (3\text{-}16)$$

If the second partial derivatives of the likelihood function are taken with respect to $\underline{\theta}$ and ρ, it is seen that $\underline{\theta}^*$ and ρ^* satisfy the optimality criterion.

For our purposes, it is sufficient that the elements of the matrix in (3-14) be linearly independent in k to insure that the inverse exists. This is true of all the model structures used in this research. Kashyap and Rao [K3, Chap. 4] give a more detailed account of the condition under which the inverse in (3-14) exists. Kashyap and Rao [K3, Chap. 4] also demonstrate the asymptotic consistency in the mean square sense of the estimators $\theta^*(N)$ and $\rho^*(N)$.

Furthermore, Kashyap and Rao [K3, Chap. 4] show that the asymptotic mean and covariance of the estimate $\theta^*(N)$ are given by

$$E[\underline{\theta}^*(N)/\underline{\theta},\rho] = \underline{\theta} + O\left[\frac{1}{N^{1/2}}\right] \qquad (3\text{-}17)$$

where

$$\frac{O(x)}{x} \to K \neq 0 \text{ as } x \to 0,$$

and

$$\mathrm{Cov}[\underline{\theta}^*(N)/\underline{\theta},\rho] \approx \rho\left[E\left[\sum_{k=1}^{N}\underline{Z}(k-1)\underline{Z}^T(k-1)\right]\right]^{-1} + o(\tfrac{1}{N}) \qquad (3\text{-}18)$$

where

$$\frac{o(x)}{x} \to 0 \text{ as } x \to 0.$$

It is also shown by Kashyap and Rao [K3, Chap. 4] that the estimator has asymptotically minimum variance and is asymptotically equal to the Cramer-Rao lower bound on the estimation variance.

Next we demonstrate a recursive algorithm for the computation of the estimate $\underline{\theta}^*(N)$ in (3-14). The algorithm eliminates the necessity of computing the inverse matrix each time a new observation is made. This gives a large reduction in computational load.

Let us write

$$\underline{P}(N) = \left[\sum_{k=1}^{N}\underline{Z}(k-1)\underline{Z}^T(k-1)\right]^{-1} . \qquad (3\text{-}19)$$

Then

$$\underline{P}(N)^{-1} = \sum_{k=1}^{N-1}\underline{Z}(k-1)\underline{Z}^T(k-1) + \underline{Z}(N-1)\underline{Z}^T(N-1)$$

or

$$\underline{P}(N)^{-1} = \underline{P}(N-1)^{-1} + \underline{Z}(N-1)\underline{Z}^T(N-1) . \qquad (3\text{-}20)$$

Now apply the matrix inversion lemma (Sage and Melsa [S3, pp. 499-500] and given in Appendix II) to (3-18) to obtain

$$\underline{P}(N) = \underline{P}(N-1)-\underline{P}(N-1)\underline{Z}(N-1)\left[1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)\right]^{-1}\underline{Z}^T(N-1)\underline{P}(N-1)$$

or

$$\underline{P}(N) = \underline{P}(N-1) - \frac{\underline{P}(N-1)\underline{Z}(N-1)\underline{Z}^T(N-1)\underline{P}(N-1)}{1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)} \quad . \tag{3-21}$$

Thus (3-21) gives a recursive expression for $\underline{P}(N)$. This expression is useful in the derivation of a recursive algorithm for the estimate $\underline{\theta}^*(N)$. First write (3-14) in the form

$$\underline{\theta}^*(N) = \underline{P}(N)\sum_{k=1}^{N}\underline{Z}(k-1)S(k) \quad . \tag{3-22}$$

Then

$$\underline{\theta}^*(N) = \underline{P}(N)\left[\sum_{k=1}^{N-1}\underline{Z}(k-1)S(k) + \underline{Z}(N-1)S(N)\right]$$

$$= \underline{P}(N)\sum_{k=1}^{N-1}\underline{Z}(k-1)S(k) + \underline{P}(N)\underline{Z}(N-1)S(N) \quad .$$

But from the form of (3-22) we can write

$$\underline{\theta}^*(N) = \underline{P}(N)\underline{P}(N-1)^{-1}\underline{\theta}^*(N-1) + \underline{P}(N)\underline{Z}(N-1)S(N) \quad . \tag{3-23}$$

Now inserting the expression for $\underline{P}(N)$ from (3-21) into the first term on the right hand side of (3-23) we obtain

$$\underline{\theta}^*(N) = \left[\underline{P}(N-1) - \frac{\underline{P}(N-1)\underline{Z}(N-1)\underline{Z}^T(N-1)\underline{P}(N-1)}{1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)}\right]\underline{P}(N-1)^{-1}\underline{\theta}^*(N-1)$$

$$+ \underline{P}(N)\underline{Z}(N-1)S(N)$$

or simplifying

$$\underline{\theta}^*(N) = \underline{\theta}^*(N-1) - \frac{\underline{P}(N-1)\underline{Z}(N-1)\underline{Z}^T(N-1)\underline{\theta}^*(N-1)}{1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)} +$$

$$+ \underline{P}(N)\underline{Z}(N-1)S(N) \quad . \tag{3-24}$$

Consider the last expression on the right-hand side of (3-24). We can write from (3-21)

$$\underline{P}(N)\underline{Z}(N-1) = \underline{P}(N-1)\underline{Z}(N-1) - \frac{\underline{P}(N-1)\underline{Z}(N-1)\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)}{1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)}$$

or

$$\underline{P}(N)\underline{Z}(N-1) = \frac{\underline{P}(N-1)\underline{Z}(N-1)}{1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)} \quad . \tag{3-25}$$

substitute (3-25) into the last expression in (3-24) and we obtain

$$\underline{\theta}^*(N) = \underline{\theta}^*(N-1) + \frac{\underline{P}(N-1)\underline{Z}(N-1)}{1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)}\left[S(N) - \underline{Z}^T(N-1)\underline{\theta}^*(N-1)\right] \quad .$$

Now it is clear that the coefficient of the second term is identical to (3-25) so that we can finally write

$$\underline{\theta}^*(N) = \underline{\theta}^*(N-1) + \underline{P}(N)\underline{Z}(N-1)\left[S(N)-\underline{Z}^T(N-1)\underline{\theta}^*(N-1)\right] \quad . \qquad (3\text{-}26)$$

The above equation along with

$$\underline{P}(N) = \underline{P}(N-1)-\frac{\underline{P}(N-1)\underline{Z}(N-1)\underline{Z}^T(N-1)\underline{P}(N-1)}{1+\underline{Z}^T(N-1)\underline{P}(N-1)\underline{Z}(N-1)} \qquad (3\text{-}21)$$

constitute the recursive algorithm for the estimate $\theta^*(N)$. For initiation of the algorithm, one can use an arbitrary (but reasonable) vector $\underline{\theta}^*(0)$ and a positive-definite matrix $\underline{P}(0)$.

Thus we have shown the form of the maximuum likelihood estimators θ^* and ρ^*. In the next section, we shall consider an estimation technique that will allow use of prior knowledge of the parameters to be estimated.

B) Bayesian Estimation

In addition to the previously mentioned ability to consider a priori knowledge of the parameters, Bayesian estimation differs from maximum likelihood techniques in another aspect. This aspect is the incorporation of the loss function which is a measure of the consequence of the error in the estimate of the parameters. The optimum Bayesian estimate minimizes the expected value of the loss function (called the risk function). The loss function most commonly used for engineering work is the quadratic loss function

$$L(\underline{\theta},\underline{\theta}^*) = (\underline{\theta}-\underline{\theta}^*)^T(\underline{\theta}-\underline{\theta}^*)$$

It is well known [K3], [S3], [V1], [R1], [F2] that the optimum estimate of $\underline{\theta}$ under the quadratic loss criterion is given by

$$\underline{\theta}^*(N) = E[\underline{\theta}/X(N)] \quad . \tag{3-27}$$

That is, $\underline{\theta}^*(N)$ is the mean of the posterior density of $\underline{\theta}$ given the observations X(N). With a few preliminary assumptions, a form for the estimate $\underline{\theta}^*(N)$ is derived. These assumptions are:

b1) The prior distribution of $\underline{\theta}$ is normal with mean $\underline{\theta}_0$ and covariance matrix $\underline{P}_0\rho$.

b2) ρ is the variance of w(k).

b3) $p(\underline{\theta}/x_2) = p(\underline{\theta})$, which insures that $\underline{\theta}$ consists of coefficients of a difference equation that are independent of initial conditions X_2.

Using these assumptions we can derive the posterior density of $\underline{\theta}$ given the observations X(N).

Assertion

The posterior density $p(\underline{\theta}/X(N))$ is normal with mean

$$\underline{\theta}^*(N) = \underline{P}(N)\left[\sum_{k=1}^{N} \underline{Z}(k-1)S(k) + \underline{P}_0^{-1}\underline{\theta}_0\right] \tag{3-28}$$

and variance $\underline{P}(N)\rho$ where

$$P(N) = \left[\sum_{k=1}^{N} \underline{Z}(k-1)\underline{Z}(k-1)^T + \underline{P}_0^{-1}\right]^{-1} \tag{3-29}$$

Demonstration

The assertion is demonstrated by examination and manipulation of relevant density functions. We have by assumption

$$\underline{\theta} \sim N(\underline{\theta}_0, \underline{P}_0\rho)$$

and

$$w(k) \sim N(0,P)$$

for all k. The posterior density of $\underline{\theta}$ given X(N) is

$$P(\underline{\theta}/X(N)) = \frac{P(\underline{\theta},X(N))}{P(X(N))} = \frac{p(X(N)/\underline{\theta})p(\underline{\theta})}{p(X(N))} \quad . \tag{3-30}$$

Consider the first expression in the numerator of (3-30).

$$p(X(N)/\underline{\theta}) = p(\underbrace{S(N),\ldots S(1)}_{X_1}, \underbrace{S(0)\ldots S(-m1)}_{X_2}/\underline{\theta})$$

This can be written as

$$p(X(N)/\underline{\theta}) = p(S(N)/S(N-1),\ldots S(1)),X_2,\underline{\theta})\cdot$$

$$p(S(N-1)/S(N-2),\ldots S(1),X_2,\underline{\theta})\cdot$$

$$\cdot\ldots\cdot p(S(1)/X_2,\underline{\theta})\cdot p(\underline{\theta}/X_2)\cdot p(X_2) \quad .$$

But from (3-8) we can write

$$p(S(k)/S(k-1),\ldots,S(1),X_2,\underline{\theta}) =$$

$$= \frac{1}{\sqrt{2\pi\rho}} \exp\left[\frac{-1}{2\rho}(S(k)-\underline{Z}^T(k-1)\underline{\theta})^2\right] \quad .$$

The second term in the numerator of (3-30) is

$$p(\underline{\theta}) = \frac{1}{(2\pi\rho|\underline{P}_0|)^{1/2}} \exp\left[\frac{-1}{2\rho}(\underline{\theta}-\underline{\theta}_0)^T\underline{P}_0^{-1}(\underline{\theta}-\underline{\theta}_0)\right]$$

The denominator of (3-30) can be written as

$$p(X(N)) = p(X_1,X_2) = p(X_1/X_2)p(X_2)$$

Now it is noted that $p(X_1/X_2)$ does not depend on $\underline{\theta}$. It does depend on $\underline{\theta}_0$, $\underline{P}_0$, ρ, and $\underline{Z}^T(k-1)$. Therefore, $p(X_1/X_2)$ can only take on the significance of a normalization factor for the posterior density. This property will greatly simplify the necessary manipulations.

Hence, we can write the posterior density (3-30) from the above relations as

$$p(\underline{\theta}/X(N)) = \prod_{k=1}^{N}\left[\frac{1}{\sqrt{2\pi\rho}} \exp\left[\frac{-1}{2\rho}(S(k)-\underline{Z}^T(k-1)\underline{\theta})^2\right]\right]$$

$$\cdot\frac{1}{[2\pi\rho|\underline{P}_0|]^{1/2}} \exp\left[\frac{-1}{2\rho}(\underline{\theta}-\underline{\theta}_o)^T\underline{P}_0^{-1}(\underline{\theta}-\underline{\theta}_0)\right]$$

$$\cdot\frac{1}{p(X_1/X_2)} \quad .$$

Collecting all normalization factors as a single term we can write

$$p(\underline{\theta}/X(N)) = K_1 \exp\left[\frac{-1}{2\rho}\left[\sum_{k=1}^{N}(S(k)-\underline{Z}^T(k-1)\underline{\theta})^2 + (\underline{\theta}-\underline{\theta}_0)^T\underline{P}_o^{-1}(\underline{\theta}-\underline{\theta}_0)\right]\right]. \quad (3\text{-}31)$$

The exponent of (3-31) can be written as

$$\frac{-1}{2\rho}\left[\sum_{k=1}^{N}(S^2(k)-2\underline{Z}^T(k-1)\underline{\theta}S(k) + \underline{\theta}^T\underline{Z}(k-1)\underline{Z}^T(k-1)\underline{\theta}) + \underline{\theta}^T\underline{P}_0^{-1}\underline{\theta} - 2\underline{\theta}_0^T\underline{P}_0\underline{\theta} + \underline{\theta}_0^T\underline{P}_0^{-1}\underline{\theta}_0\right].$$

Rearrange and label some terms to give the exponent as

$$-\frac{1}{2\rho}\left[\underline{\theta}^T\underbrace{\left[\sum_{k=1}^{N}\underline{Z}(k-1)\underline{Z}^T(k-1)+\underline{P}_0^{-1}\right]}_{\underline{P}(N)^{-1}}\underline{\theta}\right.$$

$$- 2\left[\underbrace{\left[\left[\sum_{k=1}^{N}\underline{Z}^T(k-1)S(k)+\underline{\theta}_0^T\underline{P}_0^{-1}\right]\underline{P}(N)\right]}_{\underline{\theta}^*(N)^T}\underline{P}(N)^{-1}\underline{\theta}\right.$$

$$\left.\left. + \sum_{k=1}^{N}S^2(k) + \underline{\theta}_0^T\underline{P}_0^{-1}\underline{\theta}_0\right].\right.$$

From the expressions labeled as $\underline{P}(N)^{-1}$ and $\underline{\theta}^*(N)^T$ above, we see that in order to "complete the square," it is necessary to add and subtract the term

$$\left[\left[\sum_{k=1}^{N} \underline{Z}^T(k-1)S(k)+\underline{\theta}_0\underline{P}_0^{-1}\right]\underline{P}(n)\right]\underline{P}(N)^{-1}\left[\underline{P}(N)\left[\sum_{k=1}^{N} S(k)\underline{Z}(k-1)+\underline{P}_0^{-1}\underline{\theta}_0\right]\right] \quad .$$

$$\longleftarrow\ \underline{\theta}^*(N)\ \longrightarrow$$

When the operation of "completing the square" is carried out, the extraneous terms do not depend on $\underline{\theta}$. Therefore, they also contribute to the normalization factor previously mentioned. These manipulations give the exponent as

$$\frac{-1}{2\rho}\left[(\underline{\theta}-\underline{\theta}^*(N))^T\underline{P}(N)^{-1}(\underline{\theta}-\underline{\theta}^*(N))\right]$$

where

$$\underline{\theta}^*(N) = \underline{P}(N)\left[\sum_{k=1}^{N} S(k)\underline{Z}(k-1)+\underline{P}_0^{-1}\underline{\theta}_0\right] \qquad (3\text{-}28)$$

and

$$\underline{P}(N) = \left[\sum_{k=1}^{N}\underline{Z}(k-1)\underline{Z}^T(k-1)+\underline{P}_0^{-1}\right]^{-1} \quad . \qquad (3\text{-}29)$$

Thus the exponent is of Gaussian form. The normalizing factor K_1 is computed from $\underline{P}(N)$ by

$$K_1 = \frac{1}{\left[[2\pi]^{\bar{m}}\rho|\underline{P}(N)|\right]^{1/2}} \qquad \text{where } \bar{m} = m1 + m2 \ .$$

The determinant $|P(N)|$ can be computed as follows

$$\underline{P}(N)^{-1} = \sum_{k=1}^{N} \underline{Z}(k-1)Z(k-1)^T + \underline{P}_0^{-1}$$

$$= \underline{P}(N-1)^{-1} + \underline{Z}(N-1)\underline{Z}(N-1)^T$$

$$= P(N-1)^{-1}\left[\underline{I} + \underline{P}(N-1)\underline{Z}(N-1)\underline{Z}(N-1)^T\right] .$$

So

$$P(N) = \left[\underline{I} + \underline{P}(N-1)\underline{Z}(N-1)\underline{Z}(N-1)^T\right]^{-1}\underline{P}(N-1)$$

It is relatively straight forward to show that [F3, p.40]

$$|\underline{P}(N)| = \frac{|P(N-1)|}{\left[1 + \underline{Z}(N-1)^T\underline{P}(N-1)\underline{Z}(N-1)\right]} . \qquad (3\text{-}32)$$

We note that this is a recursive relation with

$$\underline{P}(1) = \left[\underline{I} + \underline{P}_0\ \underline{Z}(0)\underline{Z}(0)^T\right]^{-1}\underline{P}_0$$

and

$$|\underline{P}(1)| = \frac{|\underline{P}_0|}{1 + \underline{Z}(0)^T\underline{P}_0\underline{Z}(0)} .$$

Hence, we can write

$$|\underline{P}(N)| = \frac{|\underline{P}_0|}{\prod_{k=1}^{N}(1 + \underline{Z}(k-1)^T\underline{P}(k-1)\underline{Z}(k-1))} \qquad (3\text{-}33)$$

Hence, the normalization factor K_1 is given by

$$K_1 = \left[\frac{\prod_{k=1}^{N} (1 + \underline{Z}(k-1)^T \underline{P}(k-1)\underline{Z}(k-1))}{2\pi\rho|\underline{P}_0|} \right]^{1/2} \quad . \tag{3-34}$$

Thus the assertion is demonstrated.

The Bayesian estimate $\underline{\theta}^*(N)$ may be placed in a recursive form that is similar to that formulated for the conditional maximum likelihood (CML) estimator. In fact, the derivation of the recursive form of the Bayesian estimator is exactly the same as for the recursive form of the maximum likelihood estimator. The major difference in the two estimation schemes lies in the manner with which the initial values of the estimate $\underline{\theta}^*$ are chosen. In the Bayesian technique we can use our knowledge (or assumption) of the prior density function. Hence, logical initial values would be

$$\underline{\theta}^*(0) = \underline{\theta}_0 \tag{3-35}$$

and

$$\underline{P}(0) = \underline{P}_0 \tag{3-36}$$

These initial values along with equations (3-26) and (3-21) constitute the recursive Bayesian estimation scheme.

To summarize, we have two similar estimation schemes which give estimates of the parameter vector $\underline{\theta}$ based on the observation of the spectral process $S(k)$, k=1, ..., N. The technique presented thus far is clear. The hypothesized models determine which parameters are to be estimated. These parameters are then estimated by one of the above techniques.

The question now arises as to what criterion is to be used to select a model to represent a spectral process. Thus, we next consider a method for selecting a model from the hypothesized models for the spectral response.

4. A Model Selection Criterion

We begin the discussion of the model selection criterion by defining a class of models. Recall that we are dealing with models that can be written as

$$S(k) = \underline{Z}^T(k-1)\underline{\theta} + w(k) \qquad (3\text{-}8)$$

A class of models is defined by Kashyap and Rao [K3, p. 181] as the triple (f, H, Ω) where f is the stochastic difference equation (3-8), H is the range of values for $\underline{\theta}$ and Ω is the range of values for ρ. A member of the class (f, H, Ω) is written $(f, \underline{\theta}, \rho)$. The parameter H is defined such that every element of one of its members is nonzero. This means, for example, that AR(1) models are in a different class then AR(2) models. The classes are said to be nonoverlapping. Thus, given several nonoverlapping classes and the empirical data set $X = \{S(1),....S(N)\}$, the problem is to select the class which most likely produced the empirical data.

The decision rule for choosing a class of models selected is derived from statistical likelihood concepts. These methods are developed by Kashyap and Rao [K3, p. 183-188]. A different approach that produces essentially the same decision rule is developed by Akaike [A1], [A2], [A3]. The methods of Kashyap and Rao are followed here. The

selection criterion is designed to select from the hypothesized classes that class which gives the maximum likelihood of producing the given empirical observations. Suppose that the empirical data X came from a model $(f, \underline{\theta}_0, \rho_0)$ where $\underline{\phi}_0 = (\underline{\theta}_0, \rho)$ is unknown. Let the probability density function of X be given by $p(X, \underline{\phi}_0)$. Furthermore, assume that $p(X, \underline{\phi}_0)$ is a known function of X and $\underline{\phi}_0$. Then the log-likelihood function of $p(X, \underline{\phi}_0)$ is given in terms of X alone by the following theorem.

Theorem (Kashyap and Rao)

Let $\underline{\phi}^*$ be the maximum likelihood estimate of $\underline{\phi}_0$ based on the empirical observation set X. Then

$$E[\ell n(p(X, \underline{\phi}_0))/\underline{\phi}^*] =$$

$$= L + E[O(||\phi_0 - \phi^*||^3)/\phi^*] \tag{3-37}$$

where

$$L = \ell n\ p(X, \underline{\phi}^*) - n_\phi \tag{3-38}$$

and

$$n_\phi = \text{the dimension of } \phi_0.$$

The proof of this theorem is given in Kashyap and Rao [K3, p. 184] and will not be reproduced here. L is regarded as an approximation of $\ell n(p(X, \underline{\phi}_0))$, the log-likelihood of the class C with the empirical data X.

This theorem suggests the decision rule:

a) For every class of models C_i, $i = 1, \ldots \ell$, find the maximum likelihood estimate $\underline{\phi}_i^*$ of $\underline{\phi}_{oi}$ assuming $\underline{\phi}_{oi} \varepsilon H_i$ using the empirical data X. Compute the class of log-likelihood functions L_i as

$$L_i = \ell n(p(X, \underline{\phi}_i^*)) - n_i \qquad (3\text{-}39)$$

where n_i is the dimension of $\underline{\phi}_{oi} = (\underline{\theta}_{oi}, \rho_{oi})$.

b) Select the class which gives the maximum value of L_i among $\{L_i; i=1,\ldots,\ell\}$.

This decision rule is relatively easy to use and allows the simultaneous comparison of several classes of models.

Next, we simplify the form of L in (3-39) for the model types used in this research. The log-likelihood function is given by

$$\ell np(X, \underline{\phi}^*) = \ell np(S(N), \ldots S(m1+1)/S(m1), \ldots, S(1), \underline{\phi}^*)$$

$$+ \ell np(S(m1), \ldots S(1), \underline{\phi}^*) \quad . \qquad (3\text{-}40)$$

Consider the first term on the right side of (3-40). From (3-8) we can write

$$I = \ell np(S(N), \ldots S(m1+1)/S(m1), \ldots, S(1), \underline{\phi}^*)$$

$$= \ell n\left[\prod_{k=m1+1}^{N} p(S(k))/S(k-1), \ldots, S(k-m1), \underline{\phi}^*)\right]$$

$$= \ell n\left[\prod_{k=m1+1}^{N} \frac{1}{\sqrt{2\pi\rho^*}} \exp\left[\frac{-1}{2\rho^*}(S(k)-\underline{Z}^T(k-1)\underline{\theta}^*)^2\right]\right]$$

$$= -\frac{(N-m1)}{2}\ell n(2\pi\rho^*) - \frac{1}{2\rho^*}\sum_{k=m1+1}^{N}(S(k)-\underline{Z}^T(k-1)\underline{\theta}^*)^2 \quad . \tag{3-41}$$

But recall from (3-15) that the maximum likelihood estimate of ρ is given here by

$$\rho^* = \frac{1}{N-m1}\sum_{k=m1+1}^{N}(S(k)-\underline{Z}^T(k-1)\underline{\theta})^2 \quad . \tag{3-42}$$

Hence (3-41) can be written

$$I = \frac{-(N-m1)}{2}\ell n(2\pi\rho^*) - \frac{1}{2}(N-m1) \quad . \tag{3-43}$$

Now consider the second term on the right side of (3-40). Let us approximate $p(S(m-1),\ldots S(1),\underline{\phi}^*)$ by considering $S(m1),\ldots S(1)$ as independent Gaussian random variables with zero mean and variance ρ_S. Thus

$$II = \ell np(S(m1),\ldots,S(1),\underline{\phi}^*)$$

$$\approx \ell n\left[\prod_{k=1}^{m1} p(S(k))\right]$$

$$= \sum_{k=1}^{m1} \ell np(S(k))$$

$$= \frac{-m1}{2} \ell n(2\pi\rho_S) - \frac{1}{2\rho_S} \sum_{k=1}^{m1} s(k)^2 \quad . \tag{3-44}$$

So combining (3:43) and (3-44) we can write

$$L = \underbrace{\left[-\frac{N}{2}\ell n\rho^* - m1\right]}_{L_a} +$$

$$+ \left[-\frac{N}{2} \ell n 2\pi - \frac{m1}{2}\ell n\left[\frac{\rho_S}{\rho^*}\right] - \frac{N}{2} + \frac{1}{2}\left[3m1 - \frac{1}{\rho_S}\sum_{k=1}^{m1} s(k)^2\right]\right] =$$

$$\longleftarrow \qquad L_b \qquad \longrightarrow$$

$$= L_a + L_b \quad . \tag{3-45}$$

It is noted that since in general $m1 \ll N$, the term corresponding to L_b will not vary significantly from class to class when compared to the variation in the term corresponding to L_a. Thus, for comparing classes, it is sufficient to use the simplified form L_a in the decision rule. Let us now discuss the significance of the various terms in the expression

$$L_a = -\frac{N}{2}\ell n(\rho^*) - m1 \quad . \tag{3-46}$$

The $\ell n\rho^*$ term is a measure of the goodness of fit of the model with the estimated parameters to the empirical data. The influence of the number of empirical data points is reflected in N. Finally, the m1 term acts to oppose the selection of increasingly complex models. This is a quantitative method of incorporating the principle of parsimony in model

selction. Parsimony in model construction is a reflection of the intellectually appealing notion that a simple, but adequate, explanation of a physical process is superior to a more complex explanation. Kashyap and Rao [K3, p. 185, 214-216], Kashyap [K4], and Box and Jenkins [B1, pp. 17-18] further discuss the idea of parsimony.

Hence, we have a model selection criterion that is analyticaly tractable and applicable to all the model types used in this research. Further, this selection criterion has intuitively pleasing properties.

5. Validation of Models

The procedure for constructing models for the spectral processes from the empirical data is straight forward. First, the parameters for the hypothesized classes of models are estimated. Then the selection criterion developed in the previous section is used to choose the best fitting model for the spectral process. The question that remains is how well the selected model represents the empirical data. Specifically, if the model has a certain weakness, then knowledge of the weakness can be used to judge whether the model is appropriate for the empirical data. The selected model may also be judged on the basis of whether the initial assumptions used to formulate the class are valid. The study of these topics constitutes the subject of model validation. If the selected model fails these tests, then perhaps another class of models should be considered. If the selected model passes the validation tests, then it is said that the model is valid for representing the empirical data. However, it is possible that a class of models other than those hypothesized may give a better fit to the empirical data. Thus validation

of a selected model should not be considered as an absolutely definitive statement. Two approaches to validation are used.

The first set of validation tests check the initial assumptions used in constructing the model. The selected model is used with the empirical data to implement these tests. The assumptions to be tested are (see equation (3-8)):

1) the noise w(k) is zero mean with constant variance ρ,

2) w(k) is independent of w(j) for $k \neq j$ and S(k-j), $j \geq 1$,

3) periodicities in the empirical data are adequately modeled.

To conduct these tests, we will use the residual sequence obtained from the selected model and the empirical data. This residual sequence is

$$X(k) = S(k) - \underline{Z}^T(k-1)\underline{\theta}^*, \qquad k = 1,\ldots,N \quad . \tag{3-47}$$

Thus we are using the empirical data and the selected model to estimate the noise sequence w(k). The tests using the data generated by (3-47) are called residual tests.

The second set of validation tests determines whether some relevant staistical characteristics of simulated data generated by the model are adequately close to the statistical characteristics of the empirical data. We will be mainly concerned with two tests:

a) Comparison of correlograms

b) Comparison of periodograms

If the selected model is to be used to produce synthetic data, then it may be useful to affirm that the synthetic data have the same trend or other features as the empirical data. Since we are interested in the

model for other reasons (i.e., to aid in the computation of average information), we will be concerned only with the tests of the correlogram and periodograms.

Next we consider in detail the tests mentioned above. We begin with the residual tests.

A) Residual Tests

Test 1 - Zero Mean Test

The first test will be for the assumption that w(k), k = 1, ..., N, has zero mean. We can recast this in a hypothesis testing context. The hypotheses can be written

H_0: $X(k) = w(k)$

H_1: $X(k) = \theta + w(k)$

where w(k) is a sequence of independent identically distributed gaussian random variables with zero mean and variance ρ, $\rho > 0$. The alternate hypothesis H_1 has $\theta \neq 0$, $-\infty < \theta < \infty$. It is well known (see Roussas [R1, pp. 292-293]) that, in this case, the uniformly most powerful test for zero mean (as the null hypothesis) is given by

$$|t(x)| < \eta_0, \text{ accept } H_0 \tag{3-48}$$

$$|t(x)| \geq \eta_0, \text{ reject } H_0$$

where

$$t(x) = \frac{N \bar{x}}{\sqrt{\hat{\rho}}} \tag{3-49}$$

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x(k) \tag{3-50}$$

and

$$\hat{\rho} = \frac{1}{N-1} \sum_{k=1}^{N} (x(k) - \bar{x})^2 \quad . \tag{3-51}$$

t(x) is the Student's t-distribution with N-1 degrees of freedom and hence η_0 is chosen such that

$$P(|t(x)| < \eta_0 | H_0) = 1 - \alpha$$

where α is the level of significance, which is defined as the probability of rejecting H_0 when H_0 is true. The level of significance is chosen to reflect the degree of confidence we wish to place in the null hypothesis. Thus, we have an easily applied test for the zero mean assumption.

Test 2 - Serial Independence Test

This is a test of the assumption that the sequence w(k), k=1,...,N is serially independent. The test is discussed by Box and Jenkins [B1, pp. 289-293], Box and Pierce [B2], and Kashyap and Rao [K3, pp. 209-210]. The test is a goodness of fit test. Since it is a goodness of fit test, only the hypothesis

$$H_0: \quad x(k) = w(k)$$

is defined. The alternate hypothesis is the set of all other residual models. Note that H_0 is the same as the null hypothesis in test 1.

That is, H_0 is the class of zero mean white noise residual models of variance ρ. To implement the test, we define the covariance of the residual data at lag k as

$$\hat{R}_k = \frac{1}{N-k} \sum_{j=k+1}^{N} x(j)x(j-k) \quad . \tag{3-52}$$

The required test statistic is then

$$\eta(x) = \frac{(N-n1) \sum_{k=1}^{n1} (\hat{R}_k)^2}{(\hat{R}_0)^2} \tag{3-53}$$

where n1 is chosen to be 0.1N to 0.01N depending on the size of the empirical data set. If the residual data set is as defined by H_0, then $\eta(x)$ is (approximately) chi-square (χ^2) distributed with n1 degrees of freedom. This gives the decision rule

$$\eta(x) \begin{cases} < \eta_0, & \text{accept } H_0 \\ \geq \eta_0, & \text{reject } H_0 \end{cases} \tag{3-54}$$

where η_0 is computed from

$$p(\eta(x) \geq \eta_0 | H_0) = \alpha \quad . \tag{3-55}$$

and α is the level of significance of the test.

Thus, we have a test which examines the goodness of fit of the covariances taken as a whole. Furthermore, the test is easily implemented on the computer.

Test 3 - Cumulative Periodogram Test

The cumulative periodogram test is designed to check for the presence of nonrandom periodic components. In our model construction technique, it is assumed that deterministic periodic components in the empirical data will be accounted for by an appropriate deterministic trend term. Hence, we are interested in detecting any significant residual periodicities so that we can adjust our models if necessary.

This test is described by Bartlett [B3] based on arguments from stochastic random walk theory and by Box and Jenkins [B1, pp. 294-297] on the basis of similarity with the Kolmogorov-Smirnov tests for distribution functions (see Hoel [H3, pp. 324-327]).

We consider the equation

$$x(k) = \sum_{j=1}^{[N/2]} (\alpha_j \cos\omega_j k + \beta_j \sin\omega_j k) + w(t) \qquad (3\text{-}56)$$

where

$$\left[\frac{N}{2}\right] = \text{largest integer} \leq \frac{N}{2}$$

and

$$\omega_j = \frac{2\pi j}{N}, \quad j = 1, \ldots, \left[\frac{N}{2}\right] .$$

This equation obviously represents the possibility of periodic components in the residual data. It is noted that if the frequencies $\omega_j = \frac{2\pi j}{N}$ are considered, then the frequencies $\omega_{N-j} = 2\pi(1-j)/N$ are redundant if phase information is not considered. It will be seen that phase information is not considered here. Hence, the hypotheses under

consideration are

$$H_0: \quad x(k) = w(k)$$

$$H_1: \quad \text{at least one of the components} \tag{3-57}$$

$$\alpha_j, B_j \quad j = 1, \ldots, \left[\frac{N}{2}\right] \text{ is nonzero}$$

This test is performed differently than the previous two tests. First compute

$$g_k = \frac{\sum_{j=1}^{k} \gamma_j^2}{\sum_{j=1}^{\left[\frac{N}{2}\right]} \gamma_j^2}, \qquad k = 1, \ldots, \left[\frac{N}{2}\right] \tag{3-58}$$

where

$$\gamma_j^2 = \left[\frac{2}{N} \sum_{k=1}^{N} x(k) \cos \omega_j k\right]^2 + \left[\frac{2}{N} \sum_{k=1}^{N} x(k) \sin \omega_j k\right]^2 . \tag{3-59}$$

If $x(k) = w(k)$, a plot of g_k vs. k would be scattered about a straight line between the points (0,0) and (0.5, 1.0). The probability that the cumulative periodogram lies between lines parallel to the line between (0,0) and (.5, 1.0) at distances $\pm\lambda\sqrt{\left[\frac{N}{2}\right]}$ is given by (see Kashyap and Rao [K3, p. 208])

$$\sum_{j=-\infty}^{\infty} (-1)^j \exp\{-2\lambda^2 j^2\} .$$

The parameter λ is 1.36 for 0.95 probability and 1.63 for 0.99 probabil-

ity.

Now even if the model is correct, the residuals will not be exactly white since the parameters were estimated from a finite data set. However, the cumulative periodogram will tend to indicate those boundary crossings that are random and those that are due to gross model deficiencies. If the boundary crossing are due to model inaccuracies, the deviations will tend to be large and constitute a considerable portion of the cumulative periodogram. Conversely, if the deviations are small and occur over a small portion of the domain of the cumulative periodogram, they might be attributable to randomness in the residuals.

Hence, the decision procedure is

1) plot the cumulative periodogram and the boundaries.

2) if the plot is within the boundaries accept H_0.

3) if the plot crosses the boundaries either

   a) reject H_0

   or

   b) if the boundary crossings are not gross consider other characteristics (i.e., examine plots of simulated and empirical data) to determine whether to reject H_0.

Thus we have a fairly easily implemented test for nonrandom periodicities in the residual data.

This completes the descriptions of the residual tests used in this research. The tests are all easily implemented on a computer and are straight forward.

B) Tests of Statistical Characteristics

Test 4 - Correlogram Test

This test is designed to compare the autocovariance functions of the empirical data and synthetic data generated from the fitted model. The technique of comparison of correlorams is discussed by Kashyap and Rao [K3, pp. 211-212]. The estimate of the autocovariance functions that are used in this test is given by

$$\hat{R}(k) = \frac{1}{N-k} \sum_{j=k+1}^{N} (S(j) - \bar{s})(S(j-k) - \bar{s}) \qquad k << N \qquad (3\text{-}60)$$

where

$$\bar{s} = \frac{1}{N} \sum_{j=1}^{N} S(j) \quad .$$

This estimate is called the corellogram by Kashyap and Rao. Oppenheim and Schafer [O1, pp. 539-541] show that the estimate is unbiased and the variance is asymptotically zero (as $N \to \infty$). Hence, the estimate is consistent in the mean square sense.

An estimate of the correlogram for the fitted model can be obtained by computing the average of the estimate of $\hat{R}(k)$ given by (3-60) over several independent sets of synthetic data. That is, the estimate of the correlogram for the fitted model from J sets of synthetic data is given by

$$\hat{R}_M(k) = \frac{1}{J} \sum_{j=1}^{J} \hat{R}_j(k) \tag{3-61}$$

where

$\hat{R}_j(k)$ is the estimate given by (3-60) for the jth sequence of synthetic data.

The standard deviation of the estimate given by (3-61) is

$$\sigma_M(k) = \left[\frac{1}{J} \sum_{j=1}^{J} (\hat{R}_j(k) - \hat{R}_M(k))^2 \right]^{1/2} \tag{3-62}$$

The estimate of the correlogram for the empirical data is also computed from (3-60) and is denoted $\hat{R}(k)$.

Kashyap and Rao suggest that if the following relationship is satisfied, the correlogram of empirical data can be regarded as adequately fitting the correlogram of the synthetic data.

$$\hat{R}_M(k) - 2\sigma_M(k) \leq \hat{R}(k) \leq \hat{R}_M(k) + 2\sigma_M(k), \quad k=1, \ldots, N-1 \tag{3-63}$$

However the variance of the estimate (3-60) becomes large as k approaches N (see Jenkins and Watts [J1, p. 181] or Oppenheim and Schafer [O1, p. 540]). Therefore, the following relation is suggested.

$$\hat{R}_M(k) - 2\sigma_M(k) \leq \hat{R}(k) \leq \hat{R}_M(k) + 2\sigma_M(k), \quad k \ll N. \tag{3-64}$$

Usually k may be chosen to be approximately 0.1N. If the relation in (3-64) is satisfied, then the correlogram of the empirical data can be said to adequately fit the (estimated) theoretical correlogram of the simulated data.

Thus, we have a method of comparing the autocovariances of the empirical data and the synthetic data generated from the fitted model.

Test 5 - Periodogram Test

The periodogram test is a qualitative test to compare frequency components of empirical and synthesized data. The periodogram has been discussed as an estimate of the power spectrum by Oppenheim and Schaffer [O1, pp. 545-554]. However, as noted by Oppenheim and Schaffer, and Bartlett [B3, pp. 274-288], the estimate is biased and has a standard deviation of the same order as the mean of the estimate. Hence, without additional processing, the periodogram is not a particularly good estimate of the power spectrum. Therefore, it is not the intention of this test to produce an estimate of the power spectrum.

The periodogram is defined as

$$\gamma^2(k) = \left[\frac{2}{N}\sum_{j=1}^{N} S(j)\cos\omega_k j\right]^2 + \left[\frac{2}{N}\sum_{j=1}^{N} S(j)\sin\omega_k j\right]^2 \qquad (3\text{-}65)$$

where $\omega_k = \frac{2\pi k}{N}$

$S(j)$ is either the empirical spectral process or the synthetic spectral process generated from the fitted model.

A plot of the periodogram versus ω_k for the empirical and the synthetic data constitutes the test. If the periodogram of the synthetic data has relative peaks at approximately the same frequencies as the relative peaks of the empirical data, then the fit is said to be adequate. It must be said that this test is highly qualitative and will indicate only gross defects in modeling any periodicities in the empirical data.

These five tests constitute our validation criteria for a selected model. If a model successfully completes the tests, we will say that the model has been validated. Once again, it must be stated that some class of models other than those hypothesized may give a better fit to the empirical data. Hence, the model validation criteria are useful only in a relative sense.

6. Summary

In this chapter, we have developed some techniques for obtaining the state variable form models for use in the Kalman filter calculations discussed in Chapter II. These calculations are then used to determine average information in spectral bands. From these computations, an optimum (in terms of average information) subset of spectral bands is chosen.

The model construction technique developed in this chapter is listed below in a stepwise sequence.

1) Hypothesize several classes of models.

2) Identify (or estimate) the necessary parameters for each class of models from the empirical data using either maximum likelihood or Bayesian estimation techniques.

3) Select a class of models using the likelihood selection criterion.

4) Use the five validation tests to determine if the selected model adequately fits the empirical data.

The validation of a selected model is for a particular set of empirical data. The selected model is the best fitting model in a class of models. The likelihood selection criterion selects the most likely class of models to represent the empirical data. In terms of this research, the above may be interpreted as meaning that we do not state that the validated model is the model for, say, wheat. Instead, we have a class of models (i.e., second order autoregressive) that represents a wheat scene in a spectral band.

Thus, a very flexible technique for constructing models of spectral processes has been developed.

Chapter IV

Application of Modeling Techniques

1. Introduction

This chapter demonstrates the application of the modeling techniques developed in Chapter III to empirical spectral response data. The modeling techniques are shown for two empirical data sets for purposes of comparison and demonstration that the techniques are applicable to different spectral scene types.

The models developed are used in the next chapter to demonstrate the use of average information to choose subsets of spectral bands.

First, however, a description of the empirical data used in this research is given.

2. The Empirical Data

The data set for this research is chosen to exhibit several characteristics. First, the data must be representative of the types of scenes observed by the multispectral scanner systems. Second, the data set should be amenable to the techniques being pursued in this research. Third, the data should be relatively free of artifacts that may be introduced in the data collection process. Such data sets are available at the Purdue University Laboratory for Application of Remote Sensing (LARS). All of the empirical data used in this research is gathered with the Purdue/LARS Exotech 20C spectroradiometer [L1], and was col-

lected from some test sites in Williams County, North Dakota.

Two different sets of empirical data are used to demonstrate the techniques developed in this research. The first set consists of observations of wheat scenes and was selected primarily for two reasons. First, it is desired to demonstrate the techniques for a scene that consists of a single type of vegetation . It is thought that this will demonstrate the feasibility of using the developed techniques to analyze parameters of a multispectral scanner system for observing a particular scene type. Second, there is general interest in observing the world wheat crop. Reasons for observing the world wheat crop include estimating the world food supply and detecting pathological conditions.

The second set of empirical data consists of several vegetation scene types. Included in this second combined empirical data set are oats, barley, grass, alfalfa and fallow fields. Use of this second set of data provides insight in using the techniques developed in this research to analyze parameters of a multispectral scanner system for observing a more general scene. Also study of the second data set provides a comparison of results for two different data sets.

The data described above is available in the data library at the Laboratory for Applications of Remote Sensing at Purdue University. The specific data used in this study is stored on data tape 3990, and each observation is identified by its run number. The observations (run numbers) used for the wheat data are listed in Table IV-1. Similarly the observations used for the combined scene are listed in Table IV-2.

The empirical data is subjected to some initial processing to render it more useful for the current study. First, the data for the

Table IV-1. Wheat Scene Data Run Numbers

Run Number	Run Number	Run Number	Run Number	Run Number
75769000	75769800	75770600	75771400	75772200
75769100	75769900	75770700	75771500	75774900
75769300	75770000	75770800	75771700	75775000
75769400	75770200	75771000	75771800	75775100
75769500	75770300	75771100	75771900	75775200
75769600	75770400	75771200	75772000	75775300
75768700	75770500	75771300	75772100	75775400

Table IV-2. Combined Scene Data Run Numbers

Run Number	Run Number	Run Number	Run Number	Run Number
75768400	75768800	75774300	75775700	75776200
75768500	75768900	75774400	75775800	75776300
75768600	75774100	75774500	75775900	75776400
75768700	75774200	75774600	75776100	

wheat is averaged over the observations. That is, an ensemble average for each wavelength is taken. It is thought that the resultant average spectral response provides a relatively good data set to demonstrate the techniques of this research. The average spectral response for the wheat data is shown in Figure IV-1. Similarly, an average spectral response for the combined scene is taken. This average spectral response may be considered as an average vegetation scene for the purpose of this research. The average spectral response for the combined data is shown in Figure IV-2. It is noticed in both Figures IV-1 and IV-2 that there are two data drop-outs at approximately 1.34 - 1.45 micrometers (μm) and 1.82 - 1.96 μm. These two data drop-outs are due to atmospheric absorbtion of the incident and reflected electromagnetic energy. Thus, these two spectral bands are not useful for the current research.

In order that the study be carried out in a context that is relatively realistic for multispectral scanners, the spectral response of the two data sets is divided into spectral bands. The division is relatively arbitrary, but each spectral band must contain a sufficient number of data points to ensure fairly accurate parameter estimation for model identification as discussed in Chapter III. The spectral bands for the wheat data are shown in Table IV-3. It has been noted previously that the gaps between bands 7 and 8 and between bands 8 and 9 are due to atmospheric absorbtion of the incident and reflected spectral energy. Similarly, spectral bands for the combined scene are shown in Table IV-4. Thus, the data sets and the spectral bands are defined for this study. The next step is to identify models for the spectral bands de-

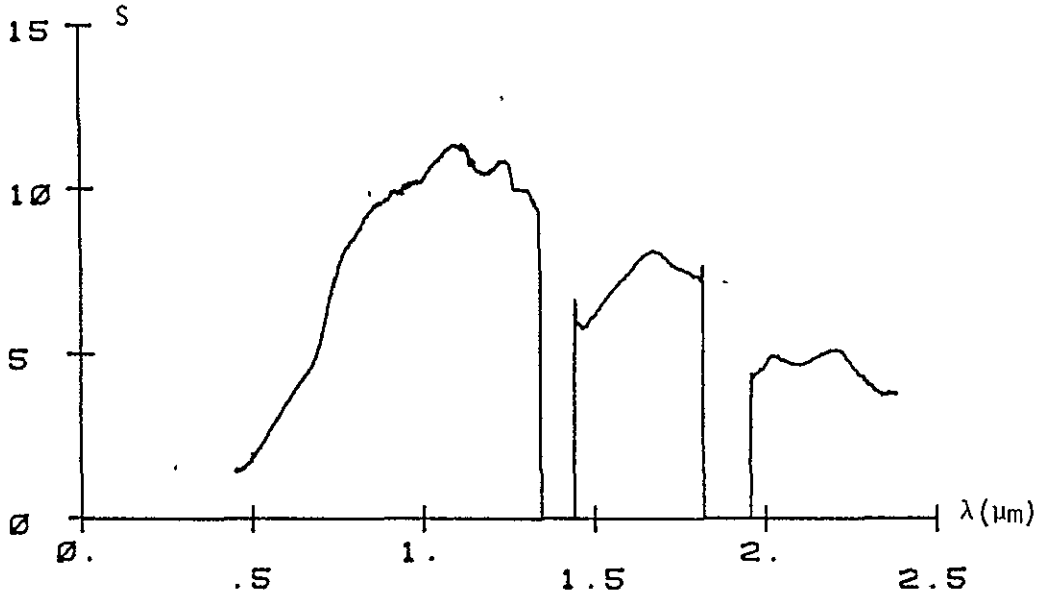


Fig. IV-1. Average Spectral Response--Wheat Scene

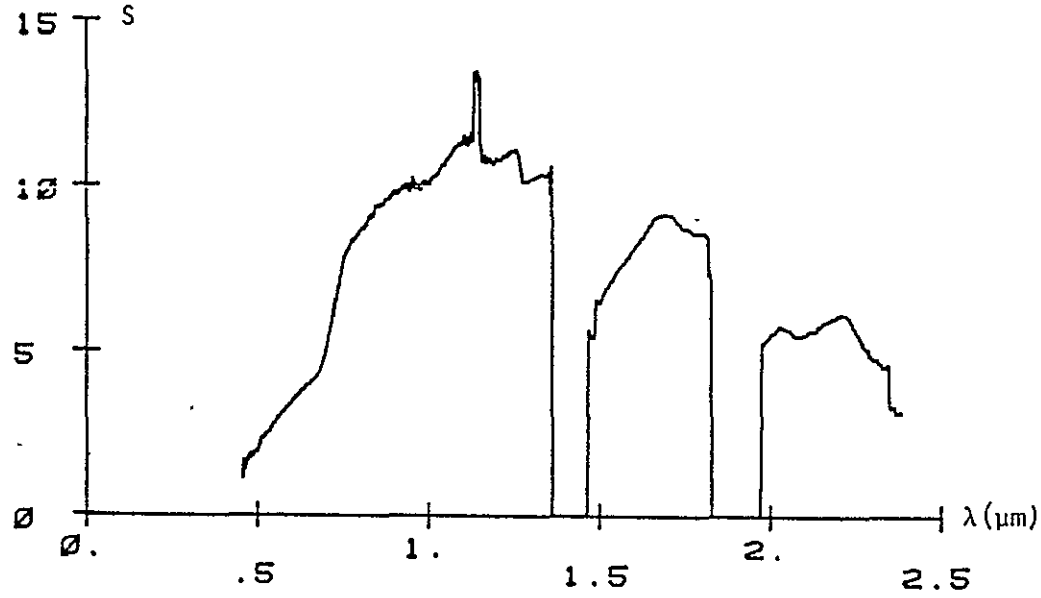


Fig. IV-2. Average Spectral Response--Combined Scene

Table IV-3. Spectral Bands for Wheat Scene

Band Number	Spectral Wavelength Interval
1	.4528 - .5380 μm
2	.5380 - .6239 μm
3	.6239 - .7097 μm
4	.7097 - .8517 μm
5	.8517 - .9910 μm
6	.9910 - 1.130 μm
7	1.130 - 1.344 μm
8	1.446 - 1.821 μm
9	1.959 - 2.386 μm

Table IV-4. Spectral Bands for Combined Scene

Band Number	Spectral Wavelength Interval
1	.4565 - .5402 μm
2	.5402 - .6246 μm
3	.6246 - .7097 μm
4	.7097 - .8481 μm
5	.8481 - .9850 μm
6	.9850 - 1.122 μm
7	1.122 - 1.307 μm
8	1.451 - 1.818 μm
9	1.967 - 2.386 μm

fined above.

3. Identified Models for the Empirical Data

This section discusses the models identified for the specral bands of the two empirical data sets described in the preceding section . The particular identification technique used is the maximum likelihood technique discussed in Chapter III. This technique obviates the need for assumptions about the prior density functions of the parameters. Instead, some arbitary (but reasonable) assumptions are made on the necessary parameters needed to initiate the estimation (identification) algorithm. A sample copy of a computer program that implements the identification algorithm in FORTRAN can be found in Appendix III.

The discussion of the models identified for the two empirical data sets is ordered by bands. It is thought that aside from being a logical method of proceeding, this will provide a simple comparison of corresponding spectral bands for the two empirical data sets.

A. Band 1

From Tables IV-3 and IV-4 it is seen that this spectral band is in the wavelength region of .4528 to .5402 μm for the wheat data and .4565 to .5402 μm for the combined scene data. The model types identified for the wheat scene are the autoregressive, autoregressive plus constant trend, and the integrated autoregressive models. Hypothesized models up to tenth order were identified. The results are tabulated in terms of order and selection criterion as defined by (3-46) in Table IV-5. On the basis of the selection criterion, it is clear that the sixth order autoregressive model is chosen. The residual variance is defined as the

variance of x(k) in equation (3-47) and can be considered as a measure of the goodness of fit of the model. The residual variance for the selected sixth order autoregressive model is 1.26×10^{-3}. Hence, the model is judged to be a good fit to the empirical data. The estimated coefficients for the sixth order autoregressive model are given in Table IV-6.

There was some difficulty in validation of a model for band 1 of the combined scene. Hence, the integrated autoregressivee model of the second kind discussed in Chapter III was identified in addition to the three model types identified for the wheat scene. Also higher order models were identified for the combined scene. It is thought that the more complicated models for band 1 of the combined scene are necessitated by the higher variability of the empirical data as seen in Figure IV-2. The identified hypothesized models are tabulated in terms of order and selection criterion in Table IV-7. The model with the highest selection criterion that also passes all the validation tests is the eleventh order integrated autoregressive model of the second kind. The other models with higher selection criterion values could not pass the serial independance test. Hence, we have an example of the ease with which the systematic approach to identification of hypothesized models developed in this research allows the examination of alternate models in the event of the inadequacy of a candidate model. The residual variance of the selected model is 1.46×10^{-3} which is evidence of a good fit to the empirical data for this model. The estimated coefficients for the eleventh order integrated autoregressive model of the second kind are given in Table IV-8.

Table IV-5. Identified Models for Band 1 of Wheat Scene

Order of Model	Selection Criterion Autoregressive	Autoregressive plus constant trend	Integrated Autoregressive
1	366.2	351.7	364.3
2	374.0	367.5	362.5
3	376.9	373.6	363.9
4	379.6	376.4	366.6
5	377.5	377.3	367.9
6	381.2	379.4	368.7
7	380.9	380.6	369.3
8	379.5	377.8	367.1
9	377.1	376.5	364.9
10	377.0	375.8	364.4

Table IV-6. Coefficients for Band 1 Selected Wheat Scene Model

Coefficient	Estimated Value
a_1	.20828
a_2	.16650
a_3	.16368
a_4	.16533
a_5	.14243
a_6	.16945

Table IV-7. Identified Models for Band 1 of Combined Scene

Order of Model	Selection Criterion			
	Autoregressive	Autoregressive plus constant trend	Integrated autoregressive first-kind	Integrated autoregressive second-kind
1	270.1	289.2	314.1	304.7
2	332.7	343.8	330.5	315.0
3	350.7	350.0	331.9	325.3
4	351.2	351.5	331.1	326.0
5	350.4	351.6	331.9	328.7
6	346.6	347.9	330.6	330.5
7	342.9	346.5	330.7	328.3
8	340.5	343.5	329.0	334.3
9	339.7	348.2	350.2	352.2
10	346.5	356.4	348.5	352.0
11	354.4	354.3	349.4	353.1
12	352.5	352.5	350.2	356.0
13	351.4	354.1	349.7	357.5
14	354.1	352.8	358.2	358.0
15	354.4	353.5	356.6	355.7
16	353.8			352.9
17	352.2			350.3
18	351.6			349.4
19	349.8			349.3
20	348.8			

Table IV-8. Coefficients for Band 1 Selected Combined Scene Model

Coefficient	Estimated Value
a_1	.081594
a_2	.037784
a_3	.13395
a_4	.089051
a_5	.084141
a_6	.10954
a_7	.10055
a_8	.075585
a_9	.12125
a_{10}	.033981
a_{11}	.064937

B) Band 2

This spectral band encompasses the wavelengths .5380 - .6239 μm for the wheat scene and .5402 - .6246 μm for the combined scene. The identified models for the wheat scene are the same types identified for band 1 of the wheat scene. It is thought that since the empirical data is fairly well behaved in this spectral band, the more complicated integrated autoregressive model of the second kind is not needed. The hypothesized models are tabulated in terms of order and selection criterion in Table IV-9. It is clear that on the basis of the selection criterion, the second order autoregressive model is to be chosen to represent band 2 of the wheat scene. The residual variance for this model is 6.19×10^{-5} indicating a good fit to the empirical data. The estimated coefficients for this model are given in Table IV-10.

The combined scene empirical data is also fairly smooth in this spectral band. Hence, only the three basic models are identified for this band. The hypothesized models are listed in terms of order and selection criterion in Table IV-11. It is clear from the table that the model selected is the second order autoregressive model. The residual variance for this model is 1.72×10^{-4} thus indicating a good fit to the empirical data. The estimated coefficients for this model are given in Table IV-12.

C) Band 3

The spectral intervals in band 3 are .6239 - .7097 μm for the wheat scene and .6246 - .7097 μm for the combined scene. The three basic models were identified for the wheat scene. The identified models are

Table IV-9. Identified Models for Band 2 of Wheat Scene

Order of Model	Selection Criterion Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	541.8	462.4	490.7
2	560.0	513.3	499.4
3	553.3	530.1	507.9
4	542.9	535.5	513.6
5	534.0	537.9	520.0
6	526.5	539.2	525.9
7	518.7	541.3	530.1
8	511.0	541.1	538.0
9	505.5	540.1	540.6
10	498.8	538.2	539.4

Table IV-10. Coefficients for Band 2 Selected Wheat Scene Model

Coefficients	Estimated Value
a_1	.50360
a_2	.50189

Table IV-11. Identified Models for Band 2 of Combined Scene

Order of Model	Selection Criterion Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	491.6	439.8	473.9
2	492.0	469.4	474.4
3	490.9	479.1	477.4
4	484.0	478.3	477.6
5	478.0	475.9	478.0
6	474.4	474.4	478.9
7	471.2	472.3	477.6
8	468.6	472.3	479.0
9	468.1	471.6	479.3
10	465.7	470.3	478.3
11	462.8	469.4	477.3
12	459.6	467.9	474.8
13	456.6	467.2	472.1
14	453.9	466.0	469.0
15	451.4	465.1	465.4

Table IV-12. Coefficients for Band 2 Selected Combined Scene Model

Coefficients	Estimated Value
a_1	.50475
a_2	.49893

listed in terms of order and selection criterion in Table IV-13. It is clear that according to the selection criterion the eleventh order integrated autoregressive model is to be chosen. The residual variance for this model is 6.00×10^{-5}, thus indicating a good fit to the empirical data. It is interesting that a higher order integrated autoregressive model gives a better fit than the best (and lower order) autoregressive models. This may be an indication of nonstationarity of the spectral process in this band. The estimated coefficients for the selected eleventh order integrated autoregressive model are given in Table IV-14.

The combined scene hypothesized models were the same as for the wheat scene. The identified models are listed according to order and selection criterion in Table IV-15. As seen from the selection criterion either the tenth or eleventh order integrated autoregressive model is to be chosen. The eleventh order model is chosen here since it has lower residual variance and has a sightly better selection criterion (if computed to more decimal places). The residual variance for the selected model is 1.15×10^{-4} which is indicative of a good fit between the model and the empirical data. The estimated coefficients for the selected eleventh order integrated autoregressive process are listed in Table IV-16.

D) Band 4

The spectral bands consist of the wavelength interval .7097 - .8517 μm for the wheat scene and .7097 - .8481 μm for the combined scene. The same three basic model types were hypothesized for this band of the wheat scene. The identified models are given in terms of order and

Table IV-13. Identified Models for Band 3 of Wheat Scene

Order of Model	Selection Criterion		
	Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	532.9	464.3	466.5
2	511.8	484.3	478.3
3	489.1	481.9	489.5
4	470.2	474.3	501.5
5	453.9	464.9	513.0
6	440.5	456.0	523.3
7	429.7	447.9	531.8
8	420.8	440.6	540.4
9	414.1	434.8	549.4
10	408.8	429.3	551.5
11			552.8
12			549.9
13			546.4
14			538.7
15			532.4

Table IV-14. Coefficients for Band 3 Selected Wheat Scene Model

Coefficient	Estimated Value
a_1	.098225
a_2	.099891
a_3	.099802
a_4	.10021
a_5	.098787
a_6	.099604
a_7	.099389
a_8	.10044
a_9	.10047
a_{10}	.099348
a_{11}	.10040

Table IV-15. Identified Models for Band 3 of Combined Scene

Order of Model	Selection Criterion		
	Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	489.3	436.8	452.7
2	462.8	447.0	460.9
3	440.5	440.6	470.3
4	421.7	431.0	480.8
5	406.9	420.9	488.8
6	396.0	412.7	493.2
7	388.6	405.9	499.8
8	384.0	401.0	505.1
9	381.4	397.5	504.6
10	380.3	394.7	505.9
11	380.2	392.7	505.9
12	380.6	391.6	504.6
13	381.5	391.0	502.6
14	383.3	391.6	500.5
15	386.3	392.5	497.3

Table IV-16. Coefficients for Band 3 Selected Combined Scene Model

Coefficient	Estimated Value
a_1	.099769
a_2	.098600
a_3	.10137
a_4	.10010
a_5	.10057
a_6	.097140
a_7	.10123
a_8	.10080
a_9	.097301
a_{10}	.098323
a_{11}	.10038

selection criterion in Table IV-17. The model chosen on the basis of the selection criterion is the first order autoregressive plus constant trend term. The constant trend term here accounts for a mean value in the driving noise which is not adequately modeled in the initial conditions taken from the empirical data. The residual variance for the first order autoregressive plus constant trend model is 4.43×10^{-4} thus indicating a good fit to the empirical data. The estimated coefficients for this model are given in Table IV-18.

The three basic model types were also identified for band 4 of the combined scene. The identified models are listed according to order and selection criterion in Table IV-19. It is clear that on the basis of the selection criteria, the first order autoregressive plus constant trend model is to be chosen. The model is a good fit to the empirical data as is exhibited by the 1.20×10^{-3} residual variance. The estimated coefficients for this model are listed in Table IV-20.

E) Band 5

The wavelength intervals for band 5 are .8517 - .9910 μm for the wheat scene and .8481 - .9850 μm for the combined scene. The three basic model types were identified for the wheat scene. The identified models are listed according to order and selection criterion in Table IV-21. It is clear from the value of the selection criterion that the first order autoregressive model is to be chosen. It is also noted that the other two model types have selection criterion very close to the one chosen. Hence, if there is any difficulty in validation of the selected model, two alternative model types are available. It is interesting to note that all three model types have a first order model

Table IV-17. Identified Models for Band 4 of Wheat Scene

Order of Model	Selection Criterion autoregressive	autoregressive plus constant trend	integrated autoregressive
1	469.4	483.3	438.1
2	439.2	471.3	450.0
3	417.3	454.3	459.3
4	403.8	442.1	468.0
5	396.7	433.6	473.6
6	394.7	428.7	477.4
7	395.3	425.4	480.2
8	393.0	423.6	481.1
9	400.2	422.9	481.8
10	403.6	423.4	482.0
11	406.9	423.8	480.8
12	410.2	425.6	477.0
13	411.9	425.8	474.5
14	412.7	426.7	472.4
15	412.5	426.1	469.1
16	411.4	426.5	466.9
17	410.2	424.9	464.1
18	408.8	423.4	461.2
19	407.7	421.5	458.1
20	406.7	419.4	455.1

Table IV-18. Coefficients for Band 4 Selected Wheat Scene Model

Coefficient	Estimated Value
a_1	.93593
c_1	.13969

Table IV-19. Identified Models for Band 4 of Combined Scene

Order of Model	Selection Criterion: Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	402.5	418.3	387.2
2	379.3	403.6	394.8
3	364.8	390.9	400.8
4	358.8	382.9	403.3
5	358.3	379.1	404.7
6	359.0	376.9	408.3
7	362.3	376.8	407.8
8	364.6	376.7	408.6
9	366.6	376.4	408.4
10	368.4	376.4	406.8
11	368.6	376.6	406.5
12	368.8	375.7	405.0
13	367.9	374.8	402.6
14	366.0	372.4	401.7
15	364.2	371.2	400.0

Table IV-20. Coefficients for Band 4 Selected Combined Scene Model

Coefficient	Estimated Value
a_1	.93382
c_1	.15938

Table IV-21. Identified Models for Band 5 of Wheat Scene

Order of Model	Selection Criterion: Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	339.9	339.7	338.4
2	334.1	333.6	336.3
3	330.6	330.0	334.3
4	328.4	327.7	332.1
5	327.4	327.3	331.7
6	326.3	325.2	329.7
7	326.5	325.7	329.2
8	326.2	325.0	328.4
9	324.8	323.9	329.3
10	322.6	321.7	327.0

Table IV-22. Coefficients for Band 5 Selected Wheat Scene Model

Coefficient	Estimated Value
a_1	1.00055

with the best selction criterion. This indicates that the empirical data in this band is fairly well behaved and that simple models of the spectral response suffice for purposes of representation . The residual variance of the first order autoregressive model is 2.80×10^{-3}, thus indicating a good fit to the empirical data. The estimated coefficient of the selected model is given in Table IV-22.

For the combined scene, the three basic models were identified. The identified models are tabulated according to order and selection criterion in Table IV-23. According to the selection criterion, the third order autoregressive model is to be chosen. The residual variance for this model is 3.86×10^{-3}, an indication of a good fit to the empirical data. The estimated coefficients for the selected third order autoregressive model are listed in Table IV-24.

F) Band 6

The spectral intervals included in band 6 are .9910 - 1.130 μm for the wheat scene and .9850 - 1.122 μm for the combined scene. The three basic models were identified for the wheat scene and are listed according to order and selction criterion in Table IV-25. It is seen from the table that the second order autoregressive plus constant trend model is to be chosen on the basis of the selection criterion. This model has a residual variance of 1.44×10^{-3} which is an indication of a good fit to the empirical data. The estimated coefficients for the selected model are given in Table IV-26.

The model types identified for band 6 of the combined scene are the same three basic types identified for band 6 of the wheat scene. The identified models are tabulated according to order and selection cri-

Table IV-23. Identified Models for Band 5 of Combined Scene

Order of Model	Selection Criterion Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	307.2	307.3	308.3
2	310.3	309.7	306.7
3	313.8	313.0	307.5
4	309.5	308.7	305.7
5	306.5	305.9	304.1
6	306.2	305.5	303.3
7	304.3	304.0	302.3
8	302.3	301.8	300.5
9	301.0	300.5	299.9
10	299.0	298.6	298.4

Table IV-24. Coefficients for Band 5 Selected Combined Scene Model

Coefficients	Estimated Value
a_1	.38814
a_2	.28995
a_3	.32304

Table IV-25. Identified Model's for Band 6 of Wheat Scene

Order of Model	Selection Criterion		
	Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	374.4	374.9	370.0
2	375.8	376.7	368.9
3	370.5	371.9	366.2
4	369.3	371.6	369.1
5	364.0	366.3	367.7
6	359.0	362.5	367.1
7	355.4	359.3	365.3
8	351.4	356.2	367.4
9	348.2	352.1	365.2
10	343.5	345.2	364.2

Table IV-26. Coefficients for Band 6 Selected Wheat Scene Model

Coefficient	Estimated Value
a_1	.51004
a_2	.47772
c^2	.14773

terion in Table IV-27. Based on the selection criterion, the first order autoregressive model is chosen. It has only a slighty better selection criterion than the first order autoregressive plus constant trend model. Thus a competitive alternate model is available. The residual variance of the selected model is 1.91×10^{-3} thus indicating a good fit to the empirical data. The estimated coefficient for the selected model is given in Table IV-28.

G) Band 7

Band 7 consists of the spectral intervals 1.130 - 1.344 μm for the wheat scene and 1.122 - 1.307 μm for the combined scene. The three basic model types identified for the preceeding spectral bands were also identified for band 7 of the wheat scene. The identified models are listed according to order and selection criterion in Table IV-29. According to the selection criterion, the eighth order integrated autoregressive model is to be chosen. However, the model with the best selection criterion that also passes all of the validation tests is the fifth order integrated autoregressive model. The residual variance of the selected fifth order integrated autoregressive model is 2.42×10^{-3} which indicates a good fit to the empirical data for band 7 of the wheat scene. The estimated coefficients for the selected model are given in Table IV-30.

The three basic models were identified for band 7 of the combined scene. The identified models for the combined scene are given according to order and selection criterion in Table IV-31. According to the selection criterion, the fifteenth order integrated autoregressive model is to be chosen. However, the fifteenth order integrated autoregressive

Table IV-27. Identified Models for Band 6 of Combined Scene

Order of Model	Selection Criterion: Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	355.9	355.8	353.9
2	347.8	346.7	353.1
3	338.5	337.5	351.8
4	332.2	332.5	351.3
5	329.1	328.5	349.2
6	326.2	327.7	348.7
7	325.7	328.2	347.8
8	352.6	326.4	347.5
9	323.7	325.4	345.4
10	322.5	323.4	345.4

Table IV-28. Coefficients for Band 6 Selected Combined Scene Model

Coefficients	Estimated Value
a_1	1.00091

Table IV-29. Identified Models for Band 7 of Wheat Scene

Order of Model	Selection Criterion		
	Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	456.6	454.3	457.3
2	437.3	435.6	457.4
3	429.0	427.0	460.2
4	425.0	424.8	461.6
5	425.5	424.2	464.8
6	425.9	424.2	462.5
7	425.2	423.3	464.1
8	424.8	423.1	469.3
9	427.1	427.5	469.1
10	429.5	427.2	467.2

Table IV-30. Coefficients for Band 7 Selected Wheat Scene Model

Coefficient	Estimated Value
a_1	.14802
a_2	.061559
a_3	.10349
a_4	.11204
a_5	.10538

Table IV-31. Identified Models for Band 7 of Combined Scene

Order of Model	Selection Criterion		
	Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	231.2	231.4	231.5
2	229.7	229.8	230.3
3	228.4	228.3	229.0
4	227.1	227.0	227.8
5	225.8	225.7	227.6
6	225.6	225.9	226.1
7	224.1	261.5	259.2
8	257.5	265.0	260.2
9	259.1	267.0	259.9
10	259.4	265.5	258.8
11	258.2	265.0	258.9
12	258.4	264.7	260.7
13	259.5	263.4	259.2
14	258.0	261.8	258.7
15	257.3	261.4	253.8

Table IV-32. Coefficients for Band 7 Selected Combined Scene Model

Coefficients	Estimated Value
a_1	.64063
a_2	.17721
a_3	.11417
a_4	.088031
a_5	.10856
a_6	-.058380
a_7	-.038789
a_8	-.061748
a_9	-.017016
c	.49218

model fails the correlogram validation test. Hence, the next alternate model is chosen according to Table IV-31. This is the ninth order autoregressive plus constant trend model. This model passes the validation tests and the residual variance is 3.27×10^{-2}. Hence the fit to the empirical data is good. The estimated coefficients for this model are given in Table IV-32.

H) Band 8

The spectral intervals included in band 8 are 1.446 - 1.821 μm for the wheat scene and 1.451 - 1.818 μm for the combined scene. The three basic models identified for the other spectral bands are also identified for band 8. The identified models for this spectral band of the wheat scene are tabulated according to order and selection criterion in Table IV-33. On the basis of the selection criterion, the ninth order integerated autoregressive model is to be selected. Unfortunately, this model does not pass the cumulative periodogram validation test. The model selected as an alternative is the eighth order integrated autoregressive model. This model passes all the validation tests. The residual variance for the selected model is 3.18×10^{-4} which indicates a good fit to the empirical data for this band of the wheat scene. Table IV-34 gives the estimated coefficients for the selected eighth order integraated autoregressive model.

The three basic models were also identified for band 8 of the combined scene and are listed according to order and selection criterion in Table IV-35. The eighth order integrated autoregressive model is chosen on the basis of the selection criterion. The residual variance

Table IV-33. Identified Models for Band 8 of Wheat Scene

Order of Model	Selection Criterion: Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	469.9	494.7	495.7
2	447.8	457.8	506.8
3	427.8	436.0	519.4
4	422.1	429.9	526.4
5	426.1	434.2	530.2
6	433.1	441.3	540.5
7	441.9	455.2	538.5
8	448.1	463.1	555.8
9	453.8	470.7	556.9
10	457.7	475.9	554.5
11	460.0	477.5	548.3
12	457.3	477.5	547.2
13	455.5	476.2	544.4
14	454.8	477.3	546.2
15	457.4	480.3	547.2

Table IV-34. Coefficients for Band 8 Selected Combined Scene Model

Coefficients	Estimated Value
a_1	.11977
a_2	.10839
a_3	.10895
a_4	.10525
a_5	.10025
a_6	.11203
a_7	.10458
a_8	.093896

Table IV-35. Identified Models for Band 8 of Combined Scene

Order of Model	Selection Criterion: Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	275.6	279.6	277.2
2	271.9	274.6	276.5
3	270.7	273.9	275.8
4	270.2	273.2	274.5
5	268.9	272.3	273.0
6	267.4	271.3	272.7
7	266.6	270.4	271.4
8	265.0	308.7	319.7
9	310.2	308.3	318.8
10	310.3	308.7	319.4
11	309.7	309.5	319.4
12	310.0	309.6	318.4
13	309.7	308.3	316.7
14	308.4	306.7	315.1
15	306.9	305.3	313.4

Table IV-36. Coefficients for Band 8 Selcted Combined Scene Model

Coefficients	Estimated Value
a_1	.074048
a_2	.054155
a_3	.10216
a_4	.058484
a_5	.080674
a_6	.090884
a_7	.083258
a_8	.11358

of the selected model is 9.58×10^{-3} thus indicating a good fit to the empirical data for band 8 of the combined scene. Table IV-36 gives the estimated coefficients for the selected model.

I) Band 9

The spectral intervals for this band are 1.959 - 2.386 μm for the wheat scene and 1.967 - 2.386 μm for the combined scene. The three basic model types were identified for this band of the wheat scene and are tabulated according to order and selection criterion in Table IV-37. According to the selection criterion, the sixth order integrated autoregressive model is selected. The residual variance for the selected model is 8.00×10^{-4} thus indicating a good fit to the empirical data for band 9 of the wheat scene. Table IV-38 gives the estimated coefficients for the selected model.

The three basic models were also identified for band 9 of the combined scene. Table IV-39 lists the identified models according to order and selction criterion. According to the selection criterion, the first order integrated autoregressive model should be chosen. However, this model is judged not to pass the qualitative periodogram validation test. Therefore, the alternative selected model is the first order autoregressive model. This model passes the validation tests. The residual variance for the first order autoregressive model is 1.18×10^{-2} thus indicating a good fit to the empirical data for band 9 of the combined scene. Table IV-40 gives the estimated coefficient for the selected first order autoregressive model.

We now have selected models for all the defined bands of both the wheat scene and the combined scene. Next we validate the selected

Table IV-37. Identified Models for Band 9 of Wheat Scene

Order of Model	Selection Criterion: Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	546.3	535.9	551.2
2	516.3	518.5	559.0
3	504.4	501.5	565.3
4	495.7	494.9	569.4
5	495.9	495.3	572.6
6	501.0	499.6	575.2
7	508.5	507.4	573.9
8	514.8	514.4	572.1
9	517.8	518.3	571.6
10	519.4	520.4	570.1

Table IV-38. Coefficients for Band 9 Selected Wheat Scene Model

Coefficient	Estimated Value
a_1	.10952
a_2	.096485
a_3	.10262
a_4	.11044
a_5	.11506
a_6	.11766

Table IV-39. Identified Models for Band 9 of Combined Scene

Order of Model	Selection Criterion: Autoregressive	Autoregressive plus constant trend	Integrated autoregressive
1	356.1	355.3	356.5
2	348.7	348.3	355.1
3	346.5	346.5	354.0
4	345.4	345.6	353.0
5	344.0	344.0	351.5
6	342.3	342.5	349.9
7	341.1	341.2	352.5
8	342.1	341.3	351.2
9	341.4	340.5	350.1
10	340.3	339.0	348.5

Table IV-40. Coefficients for Band 9 Selected Combined Scene Model

Coefficient	Estimated Value
a_1	.99759

models. Finally, these models are used for average information calculations in Chapter V.

4. Validation of Identified Models.

In this section validation of the identified models for each band of both scene types is carried out. The validaion techniques used here are those developed in Chapter III. We consider the validation of the selected models by bands as in the previous section of this chapter. The validation tests are implemented wih computer programs written in FORTRAN and included in Appendix III.

A) Band 1

We first consider validation of the selected sixth order autoregressive model for band 1 of the wheat scene.

1) Zero Mean Test

This test will be carried out for all bands of both scene types at a significance level of .01 so that there is a standard basis of comparison for all selected models. The value of n_0 at this level of significance for the number of samples in this research (approximately 115 to 160 samples) is $n_0 = 2.62$ [A5].

The value of the test statistic given by the selected model for the wheat scene is

$$|t(x)| = 1.431 \times 10^{-1} \tag{4.2}$$

Hence, the selected sixth order autoregressive model easily passes this test.

2) Serial Independence Test

This test is conducted on all bands of both scene types at a significance level of .01 so that there is a standard basis of comparison for all selected models. The critical value, η_0, is dependent on the value of n1 used for this test. In this study n1 = .1N (N is the number of empirical data points) for all models tested. The critical values are listed for several values of n1 in Table IV-41 [A5].

The test statistic for the selected sixth order autoregressive model for band 1 of the wheat scene is

$$\eta(x) = 1.515 \times 10^{1} \qquad (4\text{-}3)$$

for n1 = 10. It is clear from Table IV-41 that the selected model passes the serial independence test.

Table IV-41. Critical Values for Serial Independence Test

n1	Critical Value, η_0
9	21.6660
10	23.2093
11	24.7250
12	26.2170
13	27.6883
14	29.1413
15	30.5779

3) Cumulative Periodogram Test

This test is carried out on all bands of both scene types at a probability level of .99, thus giving a standard basis of comparison for all selected models. In all cumulative periodogram plots, the boundaries that determine the success of the test are, of course, the two parallel lines.

The selected model for band 1 of the wheat scene passes this test as is seen from Figure IV-3.

4) Correlogram Test

It is seen from Figure IV-4 that the selected sixth order autoregressive model passes this test.

In the correlogram plots, the test boundaries are shown as dashed lines.

5) Periodogram Test

As seen in Figure IV-5, the selected model for band 1 of the wheat scene may be judged to pass this qualitative test.

In the periodogram plots, the periodogram for the empirical data is plotted as a solid line and the periodogram for the synthetic data generated from the candidate model is plotted as a dashed line. For most cases, the two plots are so nearly coincident as to be indistinguishable.

Next we consider the validation of the selected eleventh order integrated autoregressive model of the second kind for band 1 of the combined scene.

6) Zero Mean Test

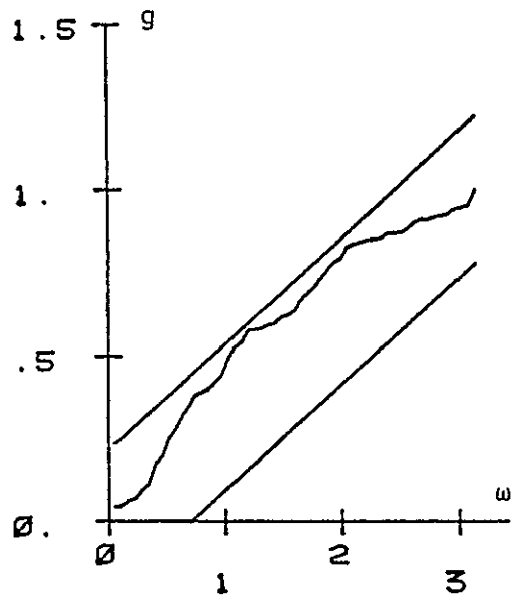


Fig. IV-3. Cumulative Periodogram, Band 1, Wheat Scene

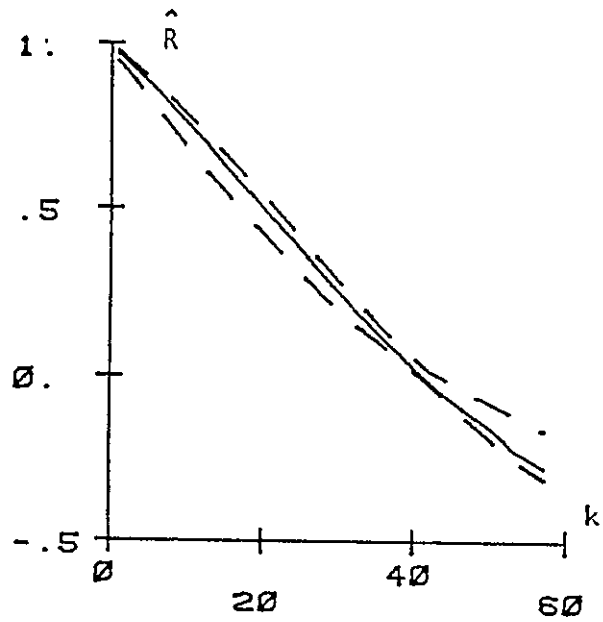


Fig. IV-4. Correlogram, Band 1, Wheat Scene

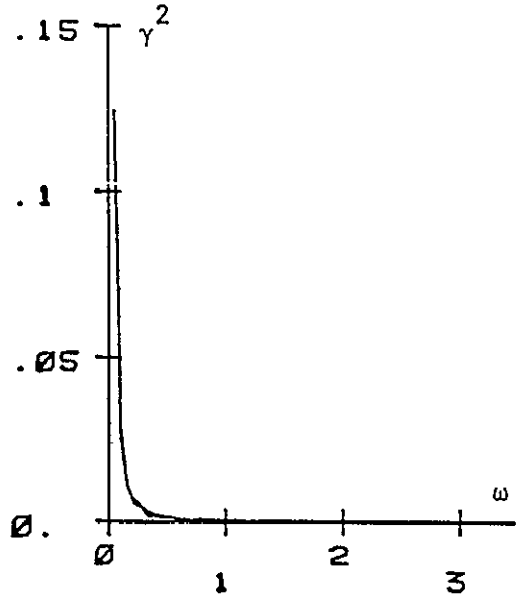


Fig. IV-5. Periodogram, Band 1, Wheat Scene

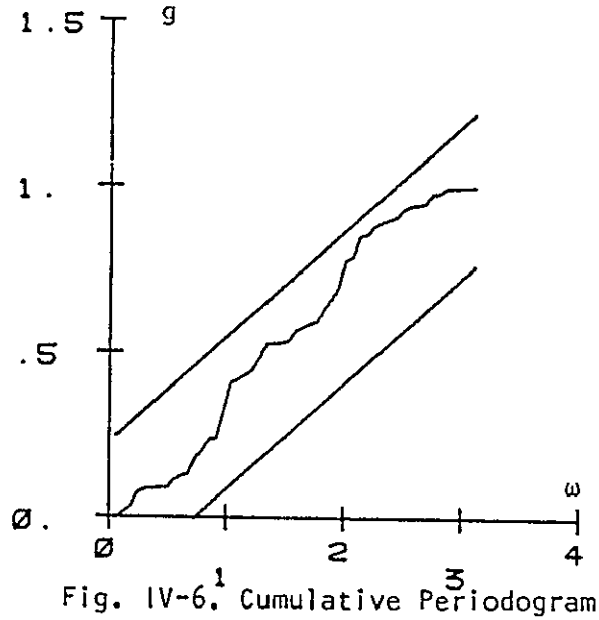


Fig. IV-6. Cumulative Periodogram, Band 1, Combined Scene

The test statistic for the selected model is

$$|t(x)| = 2.184 \times 10^{-1} \qquad (4-4)$$

Hence, the selected model easily passes this test.

7) Serial Independence Test

As noted previously, several of the identified models with higher selection criterion do not pass this test. As an example, the fifteenth order autoregressive model gives the test statistic

$$n(x) = 7.434 \times 10^{1} \qquad (4-5)$$

for n1 = 9. Thus the model clearly fails the test. Other models with higher selection criterion similarly failed this test. The selected model has the test statistic

$$n(x) = 2.009 \times 10^{1} \qquad (4-6)$$

for n1 = 10. Thus, the eleventh order integrated autoregressive model of the second kind passes this test.

8) Other Tests

Figures IV-6, IV-7, ad IV-8 show that the selected model for band 1 of the combined scene passes the cumulative periodogram, correlogram, and periodogram tests.

Hence, we have validated models for band 1 of both scene types.

B) Band 2

The validation of the selected second order autoregressive model for this band of the wheat scene is considered first.

1) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 1.955 \tag{4-6}$$

Thus the selected model passes this test.

2) Serial Independance Test

The test statistic for the second order autoregressive model is

$$_{n}(x) = 1.869 \times 10^{1} \tag{4-7}$$

for n1 = 11. Hence, the selected model passes the serial independance test.

3) Other Tests

It is seen from Figures IV-9, IV-10, and IV-11 that the selected model for band 2 of the wheat scene passes the cumulative periodogram, correlogram, and periodogram tests.

Next, we discuss the validation of the selected second order autoregressive model for band 2 of the combined scene.

4) Zero Mean Test

The test statistic for the selected model is

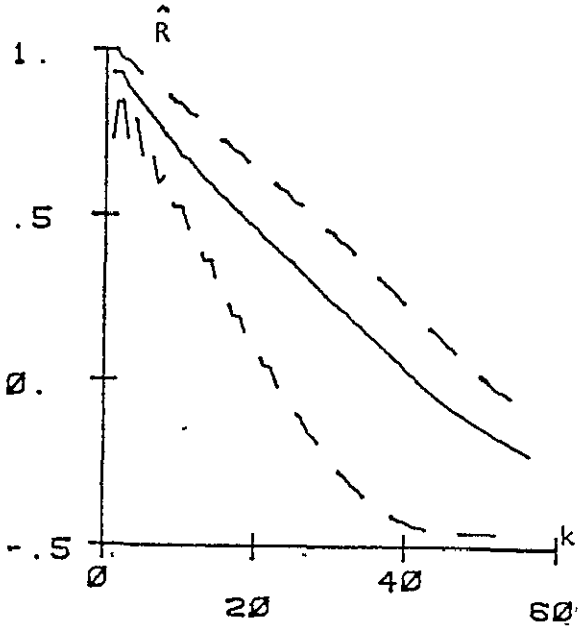


Fig. IV-7. Correlogram, Band 1, Combined Scene

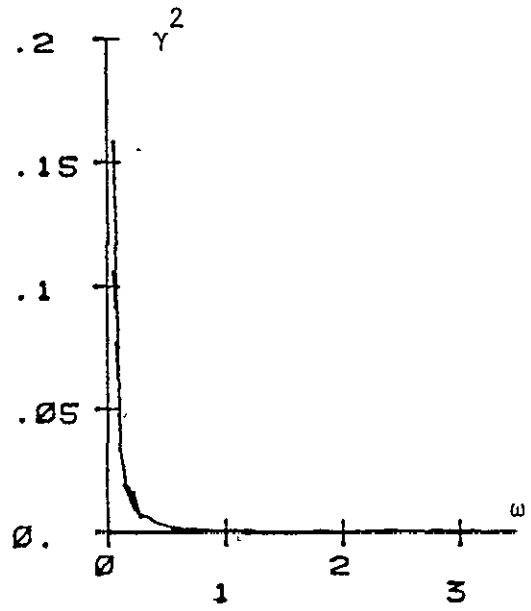


Fig. IV-8. Periodogram, Band 1, Combined Scene

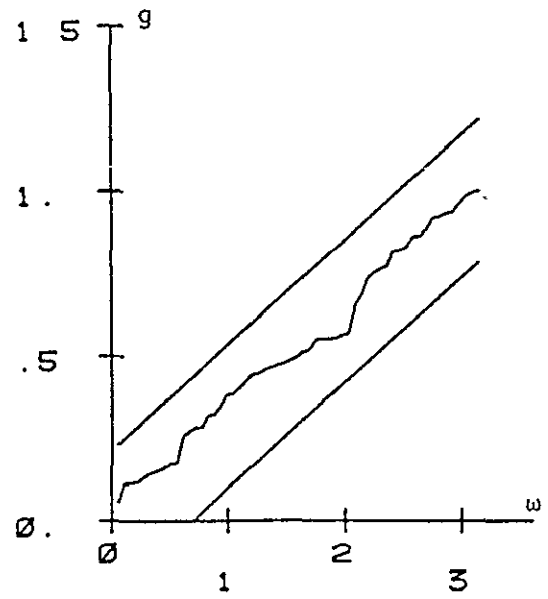


Fig. IV-9. Cumulative Periodogram, Band 2, Wheat Scene

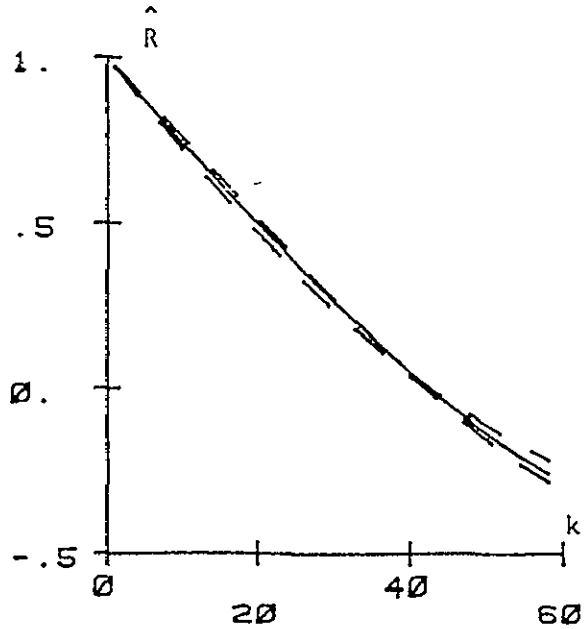


Fig. IV-10. Correlogram, Band 2, Wheat Scene

$$|t(x)| = 1.055 \qquad (4\text{-}8)$$

Hence, the selected model easily passes this test.

5) Serial Independance Test

The selected model gives the test statistic

$$\eta(x) = 1.365 \times 10^{1} \qquad (4\text{-}9)$$

for n1 = 11. Hence, the selected model also passes this test.

6) Other Tests

The selected model for band 2 of the combined scene passes the cumulative periodogram, correlogram, and periodogram tests as is seen from Figures IV-12, IV-13, and IV-14.

C) Band 3

Validation of the selected eleventh order integrated autoregressive model for band 3 of the wheat scene is considered first.

1) Zero Mean Test

The test statistics for the zero mean test is

$$|t(x)| = 3.616 \times 10^{-1} \quad . \qquad (4\text{-}10)$$

Hence, the selected model easily passes this test.

2) Serial Independence Test

The selected model gives the test statistic

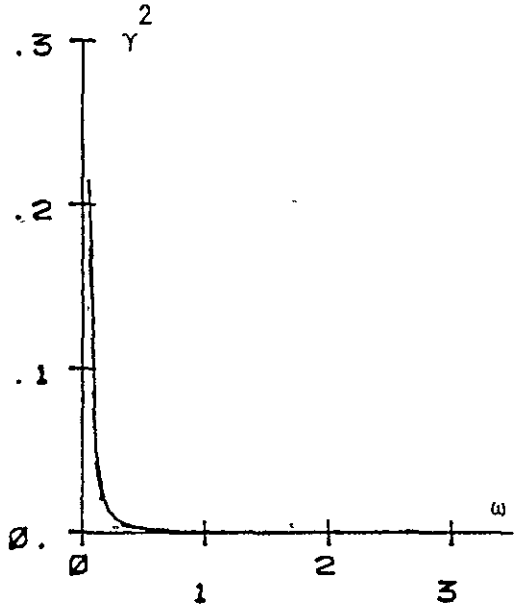


Fig. IV-11. Periodogram, Band 2, Wheat Scene

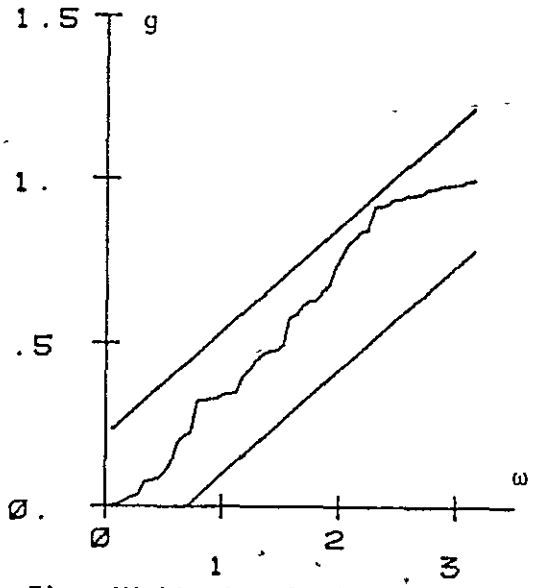


Fig. IV-12. Cumulative Periodogram, Band 2, Combined Scene

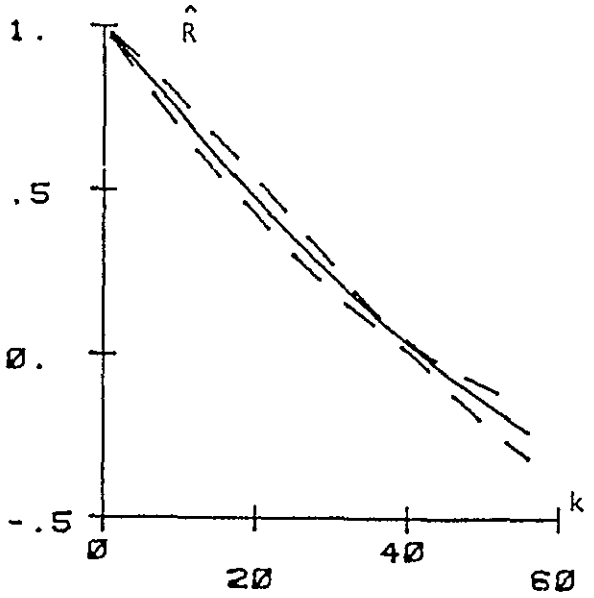


Fig. IV-13. Correlogram, Band 2, Combined Scene

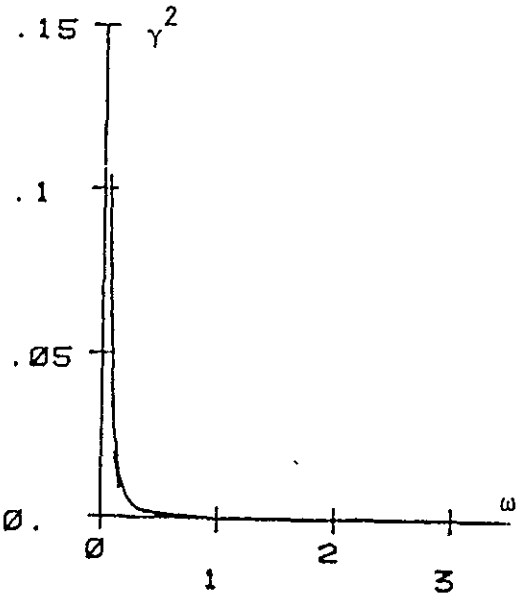


Fig. IV-14. Periodogram, Band 2, Combined Scene

$$\eta(x) = 1.153 \times 10^{1} \qquad (4\text{-}11)$$

for n1 = 10. Thus, the selected model easily passes this test.

3) Other Tests

Figures IV-15, IV-16, and IV-17 show that the selected model for band 3 of the wheat scene passes the cumulative periodogram, correlogram, and periodogram tests.

The selected eleventh order integrated autoregressive model for band 3 of the combined scene is considered next.

4) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 1.871 \times 10^{-1} \quad . \qquad (4\text{-}12)$$

Thus, the selected model easily passes this test.

5) Serial Independance Test

The selected model gives the test statistic

$$\eta(x) = 2.125 \times 10^{1} \qquad (4\text{-}13)$$

for n1 = 10. Thus, the selected model passes this test.

6) Other Tests

It is seen from Figures IV-18, IV-19, and IV-20 that the selected model for band 3 of the combined scene passes the cumulative periodogram, correlogram, and periodogram tests.

Hence, we have validated models for band 3 of both scene types.

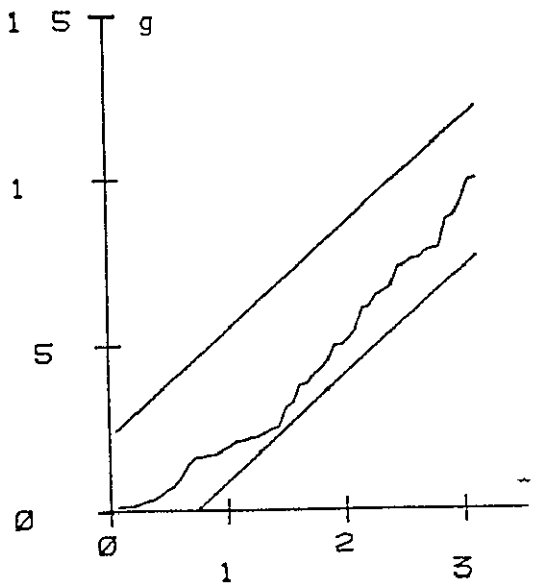


Fig. IV-15. Cumulative Periodogram, Band 3, Wheat Scene

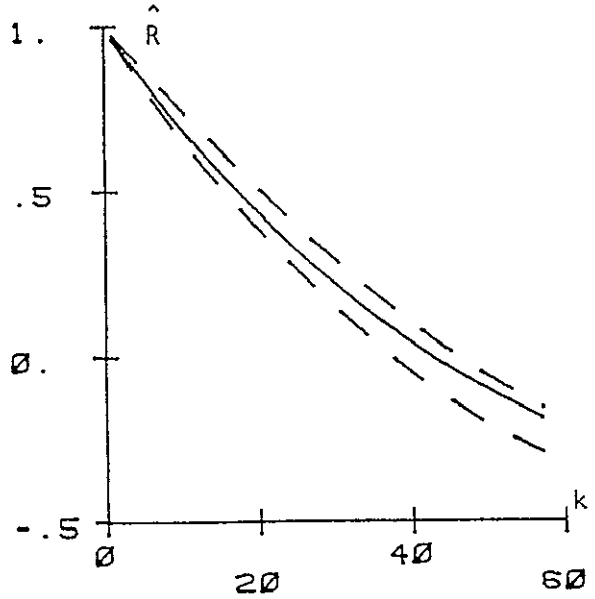


Fig. IV-16. Correlogram, Band 3, Wheat Scene

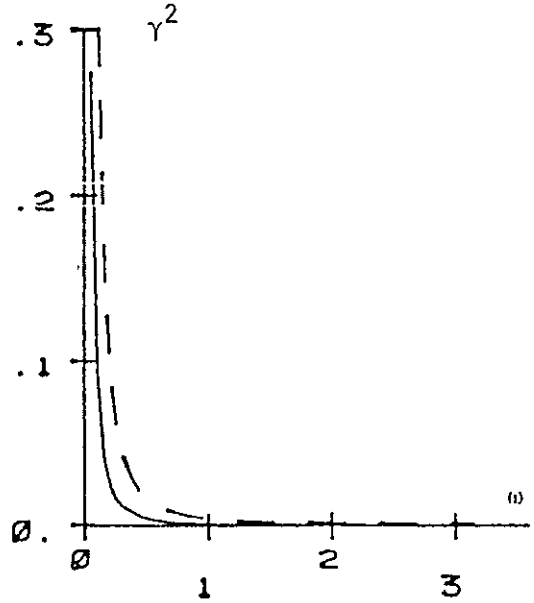


Fig. IV-17. Periodogram, Band 3, Wheat Scene

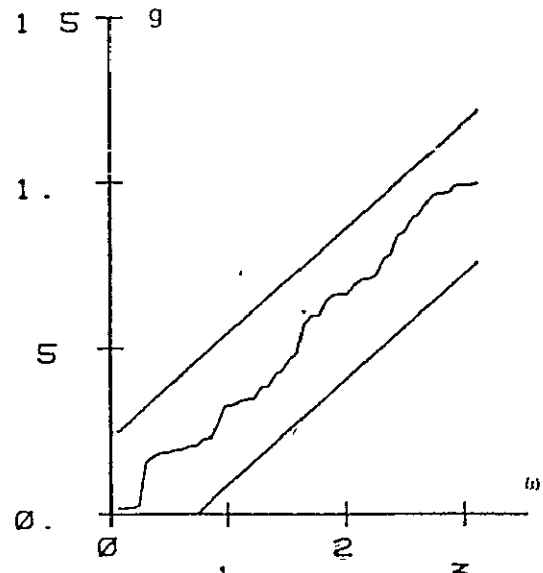


Fig. IV-18. Cumulative Periodogram, Band 3, Combined Scene

D) Band 4

Validation of the selected first order autoregressive plus constant trend model for band 4 of the wheat scene is considered first.

1) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 1.753 \times 10^{-1} \quad . \tag{4-14}$$

Hence, the selected model easily passes this test.

2) Serial Independence Test

The selected model gives the test statistic

$$n(x) = 2.00 \times 10^{1} \tag{4-15}$$

fo n1 = 12. Therefore, the selected model also passes this test.

3) Other Tests

Figures IV-21, IV-22, and IV-23 show that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

We next consider validation of the selected first order autoregressive plus constant trend model for band 4 of the combined scene.

4) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 1.598 \times 10^{-1} \quad . \tag{4-16}$$

Hence, the selected model easily passes this test.

5) Serial Independence Test

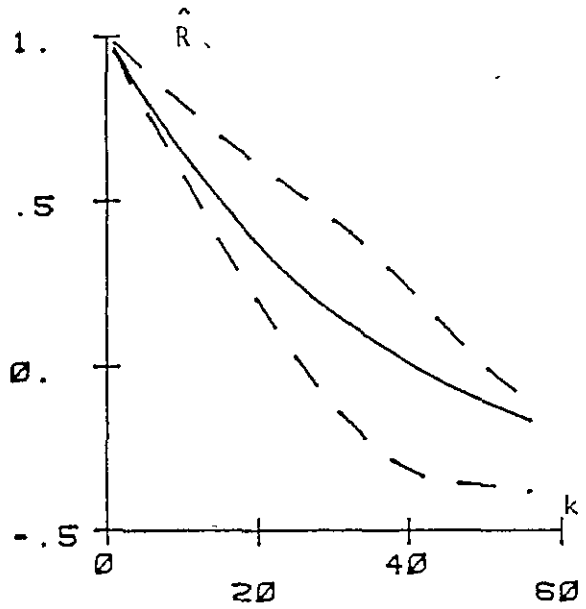


Fig. IV-19. Correlogram, Band 3, Combined Scene

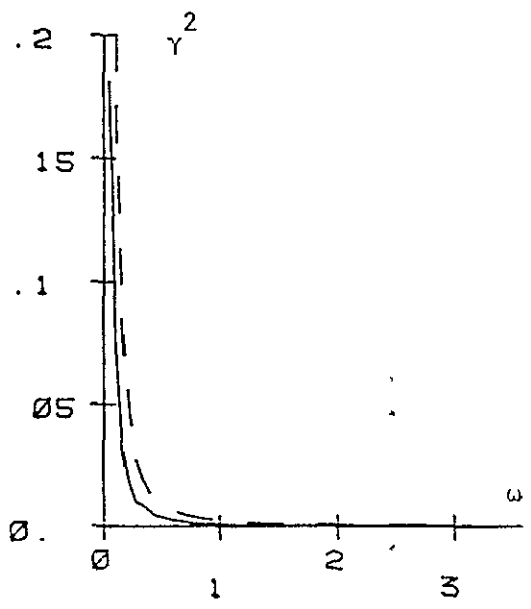


Fig. IV-20. Periodogram, Band 3, Combined Scene

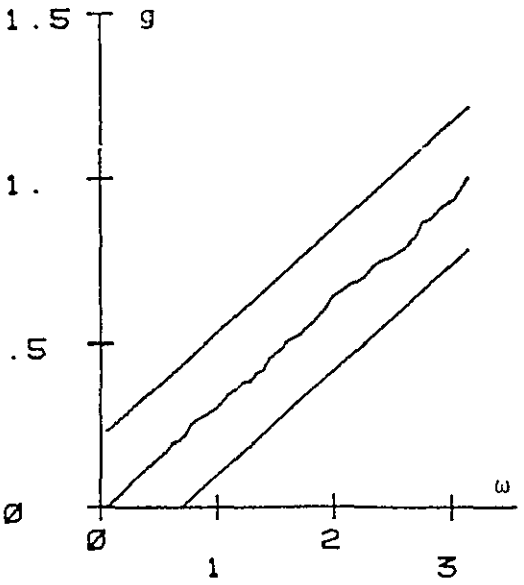


Fig. IV-21. Cumulative Periodogram, Band 4, Wheat Scene

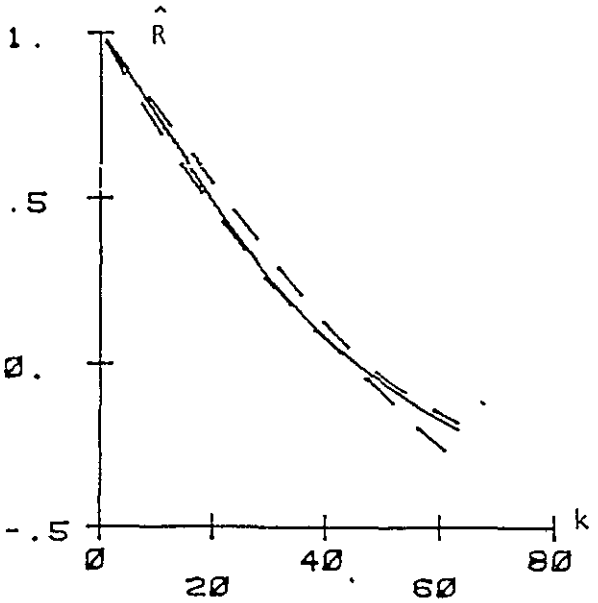


Fig. IV-22. Correlogram, Band 4, Wheat Scene

The selected model gives the test statistic

$$\eta(x) = 2.1250 \times 10^{1} \qquad (4\text{-}17)$$

for n1 = 12. Hence, the selected model also passes this test.

6) Other Tests

It is seen from Figures IV-24, IV-25, and IV-26 that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

Hence, we have validated models for both scene types of band 4.

E) Band 5

First, validation of the selected first order autoregressive model for band 5 of the wheat scene is considered.

1) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 2.495 \times 10^{-1} \quad . \qquad (4\text{-}18)$$

Hence, the selected model easily passes this test.

2) Serial Correlation Test

The selected model yields the test statistic

$$\eta(x) = 1.787 \times 10^{1} \qquad (4\text{-}19)$$

for n1 = 11. Therefore, the selected model pases this test.

3) Other Tests

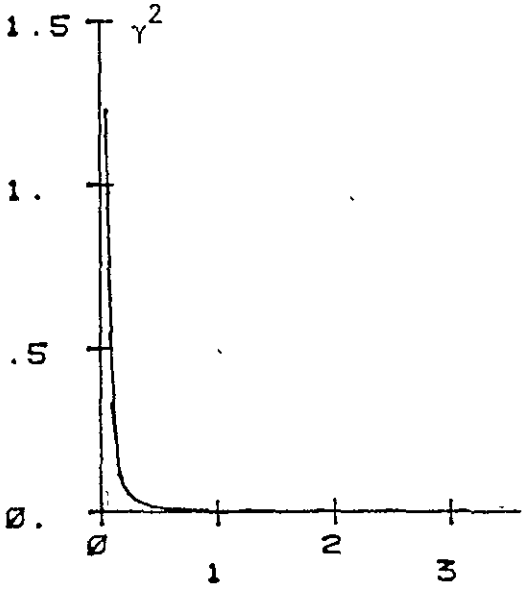


Fig. IV-23. Periodogram, Band 4, Wheat Scene

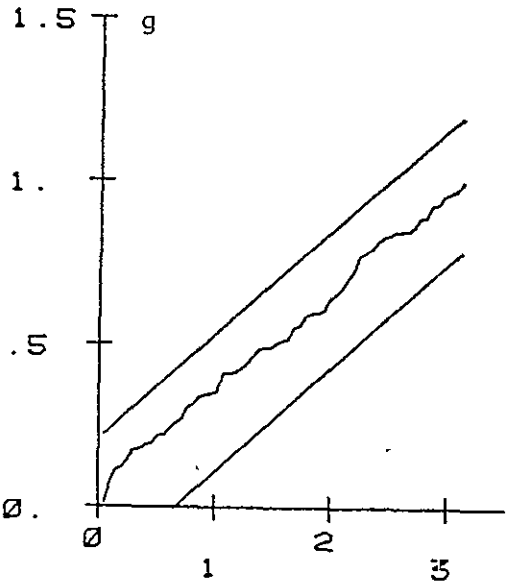


Fig. IV-24. Cumulative Periodogram, Band 4, Combined Scene

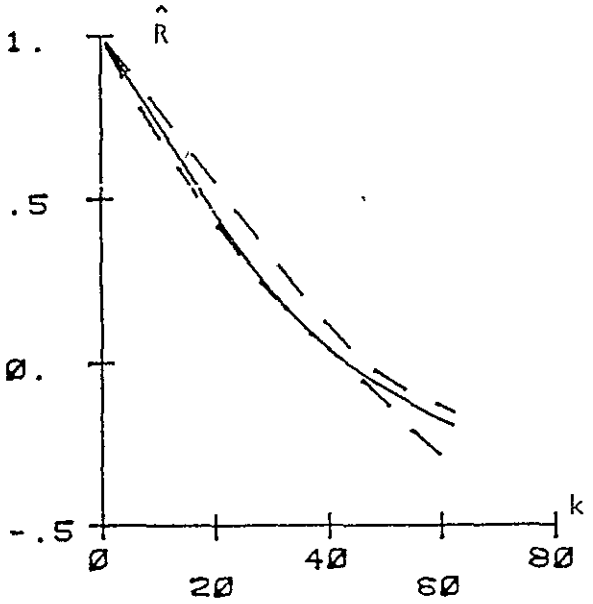


Fig. IV-25. Correlogram, Band 4, Combined Scene

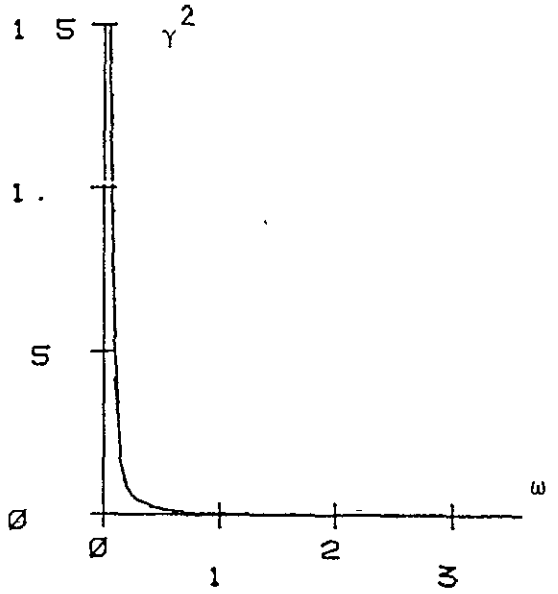


Fig. IV-26. Periodogram, Band 4, Combined Scene

Figures IV-27, IV-28, and IV-29 show that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

Next, validation of the selected third order autoregressive model for band 5 of the combined scene is considered.

4) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 8.622 \times 10^{-2} \quad . \tag{4-20}$$

Thus, the selectd model easily passes this test.

5) Serial Independence Test

The selected model gives the test statistic

$$n(x) = 1.628 \times 10^{1} \tag{4-21}$$

for n1 = 11. Hence, the selected model passes this test.

6) Other Tests

Figures IV-30, IV-31, and IV-32 show that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

Thus we have validated models for band 5 of both scene types.

F) Band 6

Validation of the selected second order autoregressive plus constant trend model for band 6 of the wheat scene is considered first.

1) Zero Mean Test

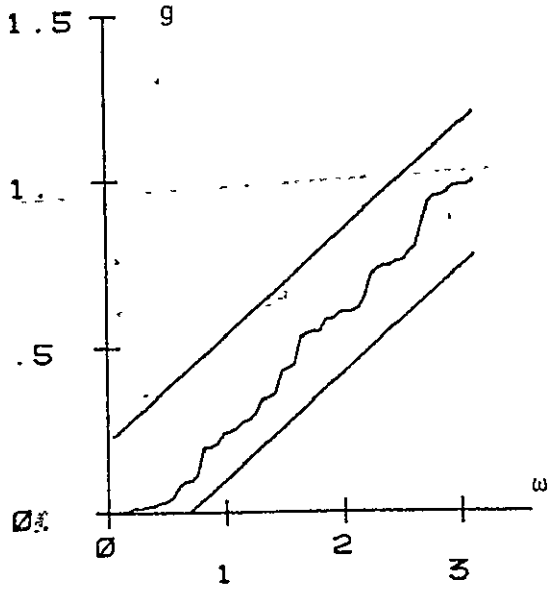


Fig. IV-27. Cumulative Periodogram, Band 5, Wheat Scene

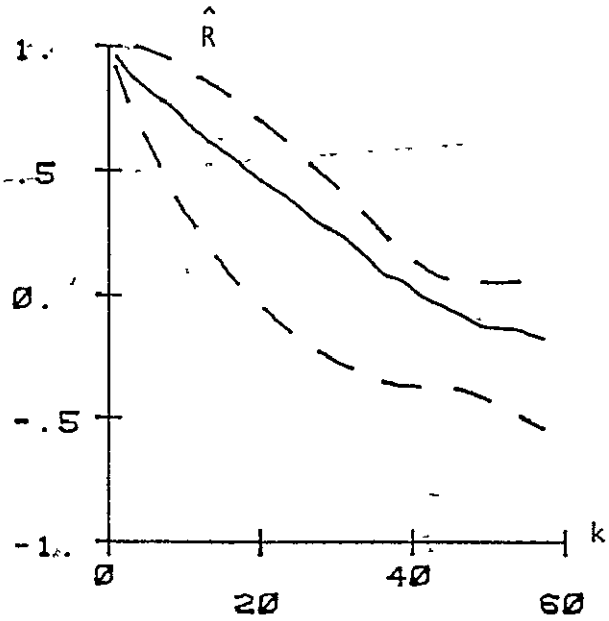


Fig. IV-28. Correlogram, Band 5, Wheat Scene

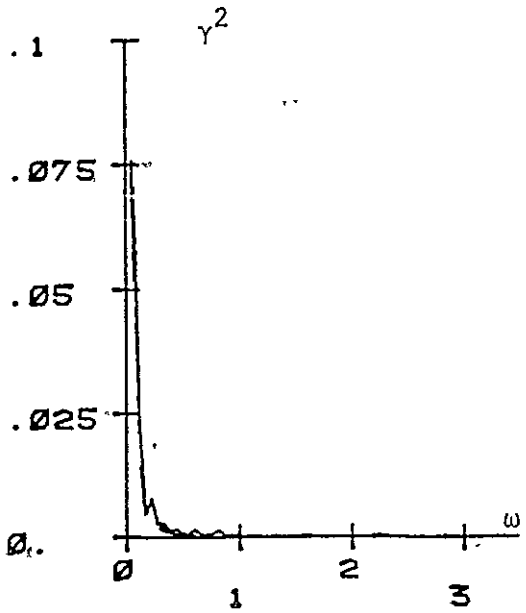


Fig. IV-29. Periodogram, Band 5, Wheat Scene

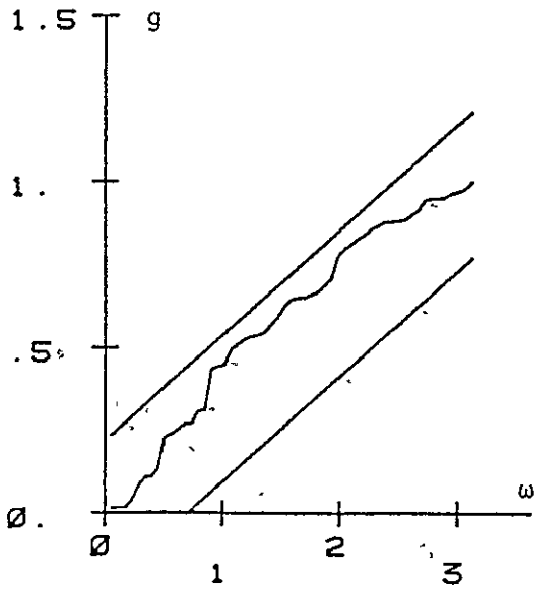


Fig. IV-30. Cumulative Periodogram, Band 5, Combined Scene

The test statistic for the selected model is

$$|t(x)| = 1.299 \times 10^{-1} \quad . \tag{4-22}$$

Thus, the selected model easly passes this test.

2) Serial Independence Test

The selected model gives the test statistic

$$\eta(x) = 2.345 \times 10^{1} \tag{4-23}$$

for n1 = 11. Hence, the selected model passes this test.

3) Other Tests

The selected model passes the cumulative periodogram, correlogram, and periodogram tests as is seen from Figures IV-33, IV-34, and IV-35.

Next, validation of the selected first order autoregressive model for band 6 of the combined scene is considered.

5) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 2.861 \times 10^{-1} \tag{4-24}$$

Hence, the selected model easly passes this test.

5) Serial Independence Test

The selected model yields the test statistic

$$\eta(x) = 8.322 \tag{4-25}$$

for n1 = 11. Hence, the selected model also easily passes this test.

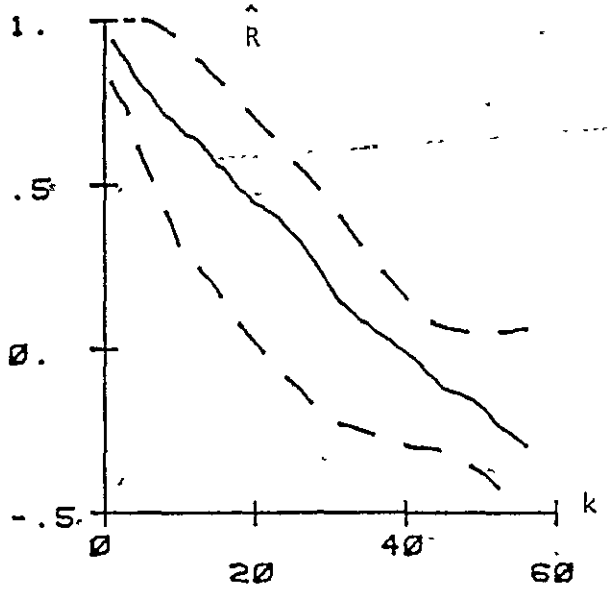
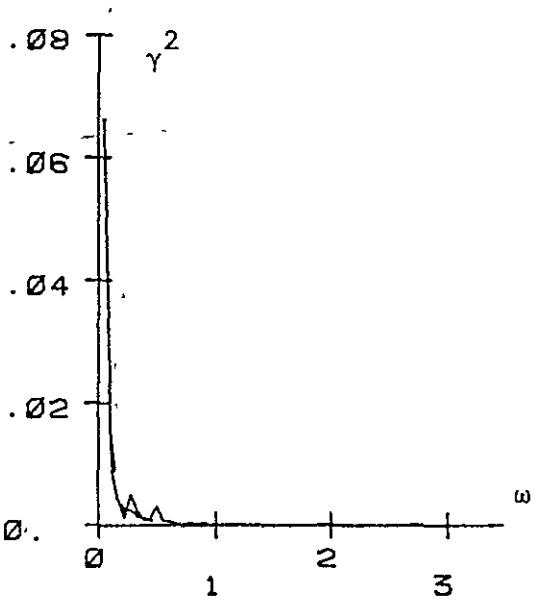


Fig. IV-31. Correlogram, Band 5, Combined Scene



Fiv. IV-32. Periodogram, Band 5, Combined Scene

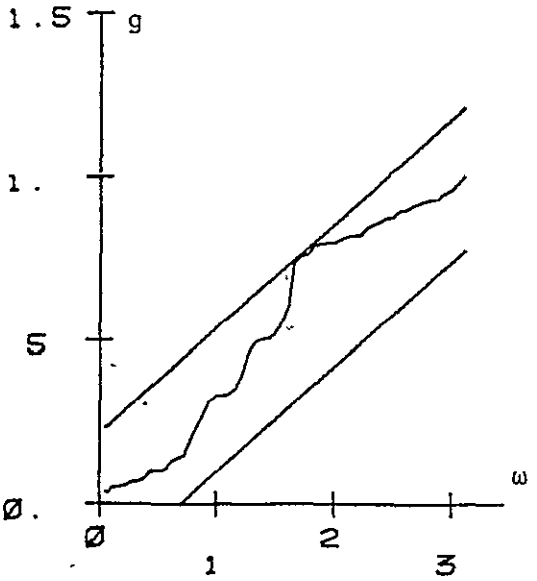


Fig. IV-33. Cumulative Periodogram, Band 6, Wheat Scene

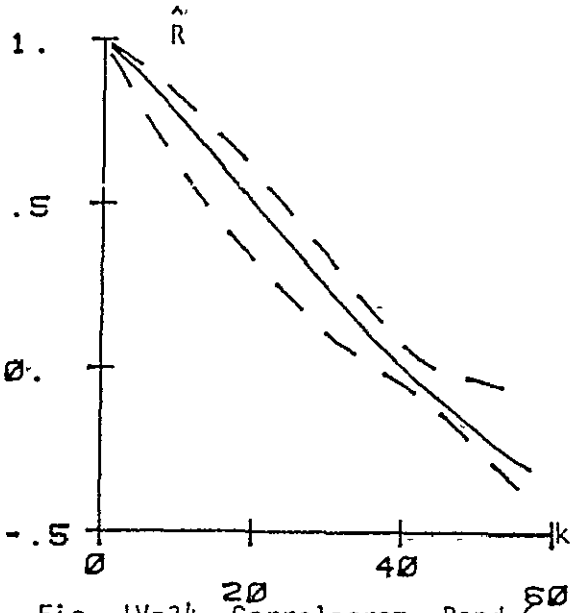


Fig. IV-34. Correlogram, Band 6, Wheat Scene

6) Other Tests

It is seen from Figures IV-36, IV-37, and IV-38 that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

Thus, we have validated models for band 6 of both scene types.

G) Band 7

Validation of the selected fifth order integrated autoregressive model for band 7 of the wheat scene.

1) Zero Mean Test

The selected model gives the test statistic

$$|t(x)| = 1.853 \tag{4-26}$$

Hence, the selected model easily passes this test.

2) Serial Independence Test

The test statistic for the selected model is

$$\eta(x) = 1.807 \times 10^{1} \tag{4-27}$$

for n1 = 14. Hence, this test is passed by the selected model.

3) Other Tests

Figures IV-39, IV-40, and IV-41 show that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

We next consider validation of the selected ninth order autoregressive plus constant trend model.

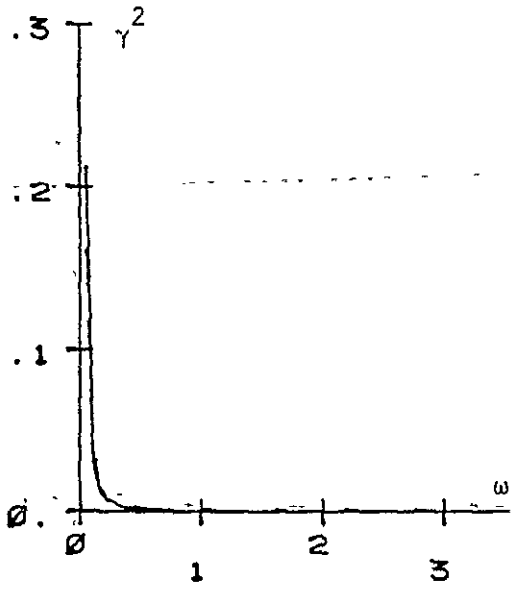
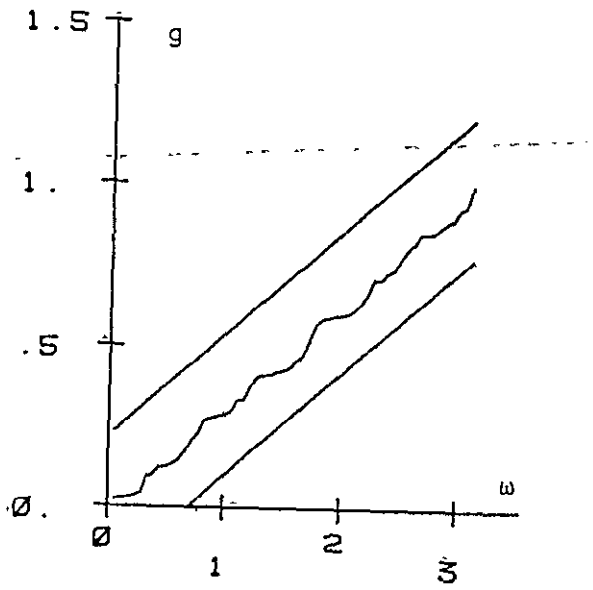


Fig. IV-35. Periodogram, Band 6, Wheat Scene



Fiv. IV-36. Cumulative Periodogram, Band 6, Combined Scene

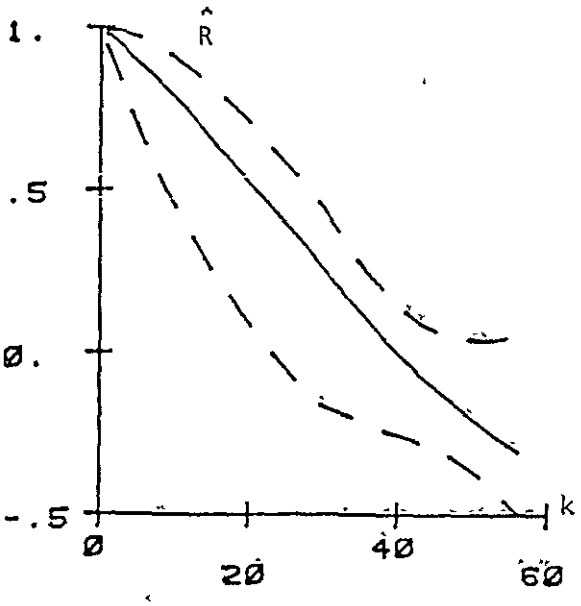


Fig. IV-37. Correlogram, Band 6, Combined Scene

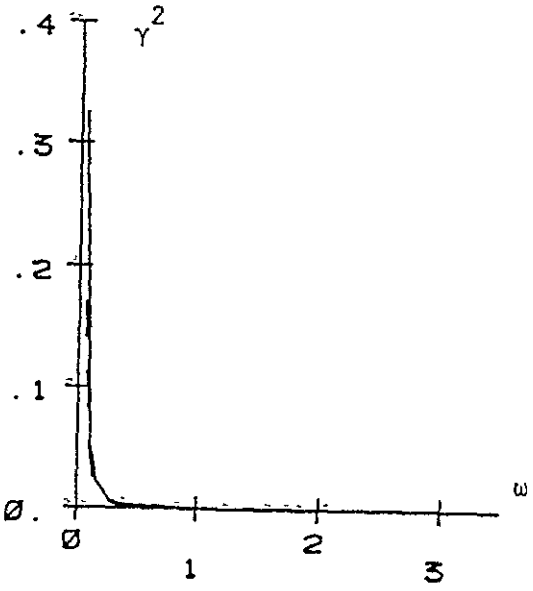


Fig. IV-38. Periodogram, Band 6, Combined Scene

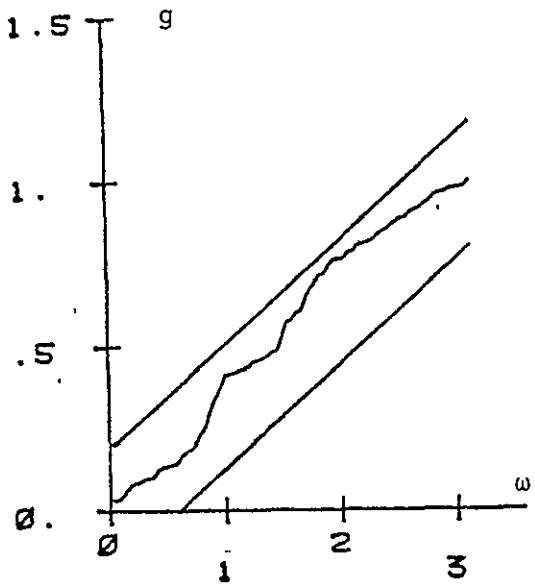


Fig. IV-39. Cumulative Periodogram, Band 7, Wheat Scene

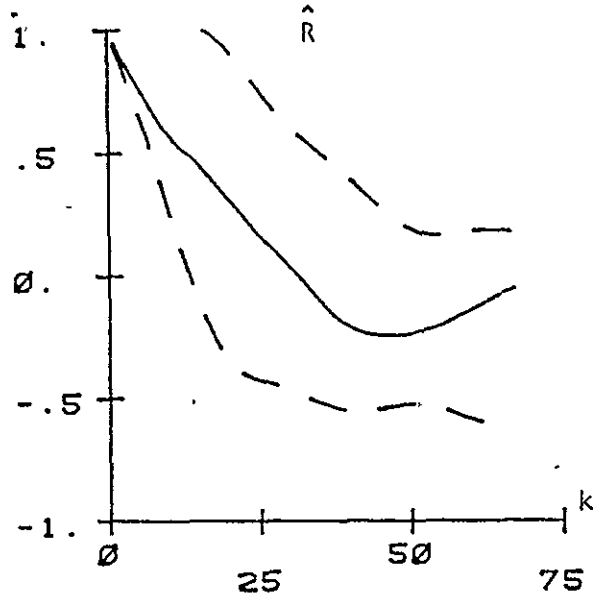


Fig. IV-40. Correlogram, Band 7, Wheat Scene

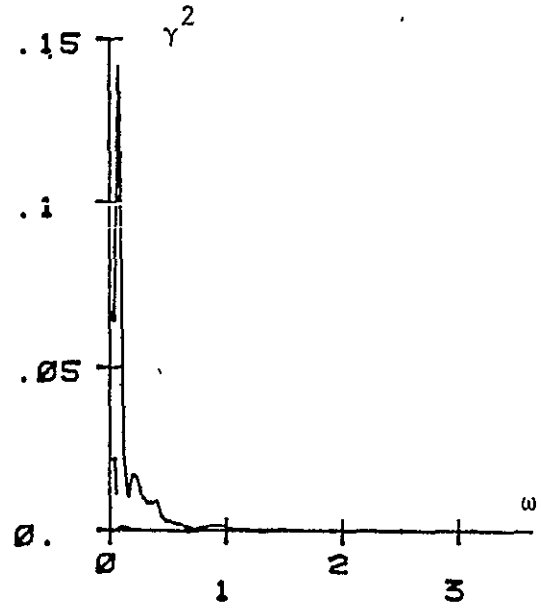
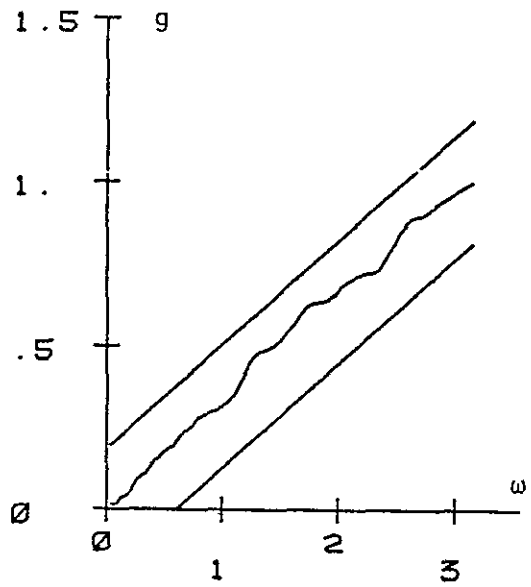


Fig. IV-41. Periodogram, Band 7, Wheat Scene



Fiv. IV-42. Cumulative Periodogram, Band 7, Combined Scene

4) Zero Mean Test

The selected model gives the test statistic

$$|t(x)| = 1.208 \times 10^{-1} \tag{4-28}$$

Thus, the selected model easily passes this test.

5) Serial Independence Test

The test statistic for the selected model is

$$_{n}(x) = 5.568 \tag{4-29}$$

for n1 = 15. Therefore, the selected model easily passes this test.

6) Other Tests

Figures IV-42, IV-43, and IV-44 show that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

Hence, the selected models are validated for band 7 of both scene types.

H) Band 8

Validation of the selected eighth order integrated autoregressive model for band 8 of the wheat scene is considered first.

1) Zero Mean Test

The test statistic for the selected model is

$$|t(x)| = 7.217 \times 10^{-1} \tag{4-30}$$

Thus, the selected model passes this test.

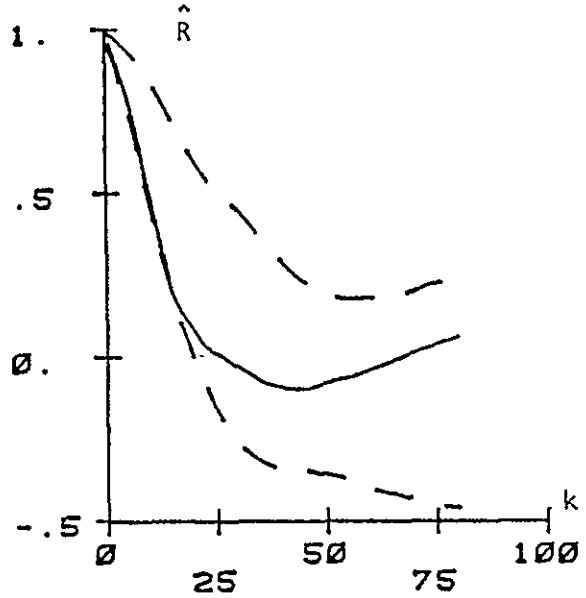


Fig. IV-43. Correlogram, Band 7, Combined Scene

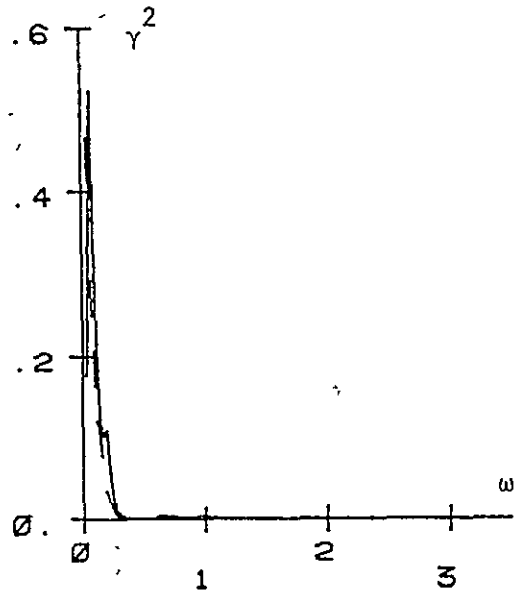


Fig. IV-44. Periodogram, Band 7, Combined Scene

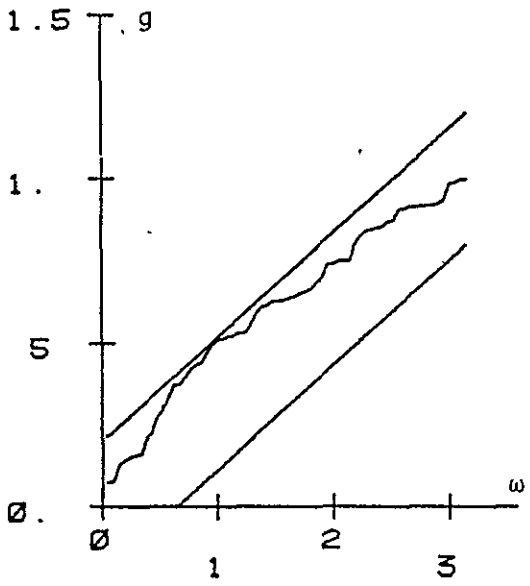


Fig. IV-45. Cumulative Periodogram, Band 8, Wheat Scene

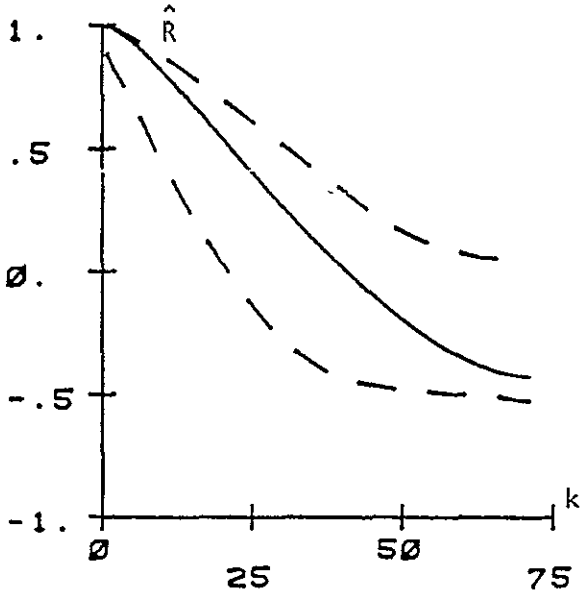


Fig. IV-46. Correlogram, Band 8, Wheat Scene

2) Serial Independence Test

The selected model gives the test statistics

$$n(x) = 2.516 \times 10^{1} \tag{4-31}$$

for n1 = 13. Hence, the selected model passes the test.

3) Other Tests

The selected model passes the cumulative periodogram, correlogram, and periodogram tests as is seen from Figures IV-45, IV-46, and IV-47.

Next, validation of the selected eighth order integrated autoregressive model or band 8 of the combined scene is considered.

4) Zero Mean Test

The selected model gives the test statistic

$$|t(x)| = 1.083 \tag{4-32}$$

Thus, the selected model easily passes this test.

5) Serial Independence Test

The test statistic for the selected model is

$$n(x) = 3.141 \tag{4-33}$$

for n1 = 13. Therefore, the selected model passes this test.

6) Other Tests

Figures IV-48, IV-49, and IV-50 show that the selected model passes the cumulative periodogram, correlogram, and periodogram tests.

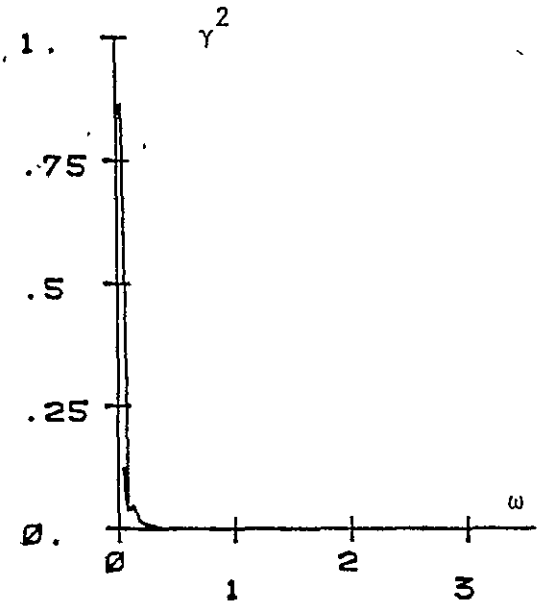


Fig. IV-47. Periodogram, Band 8, Wheat Scene

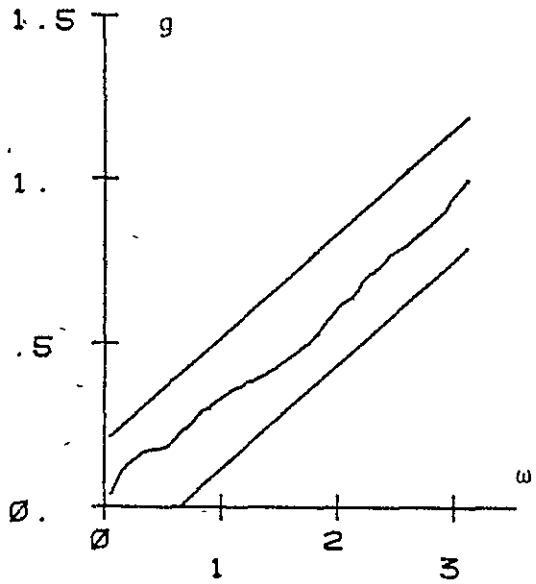


Fig. IV-48. Cumulative Periodogram, Band 8, Combined Scene

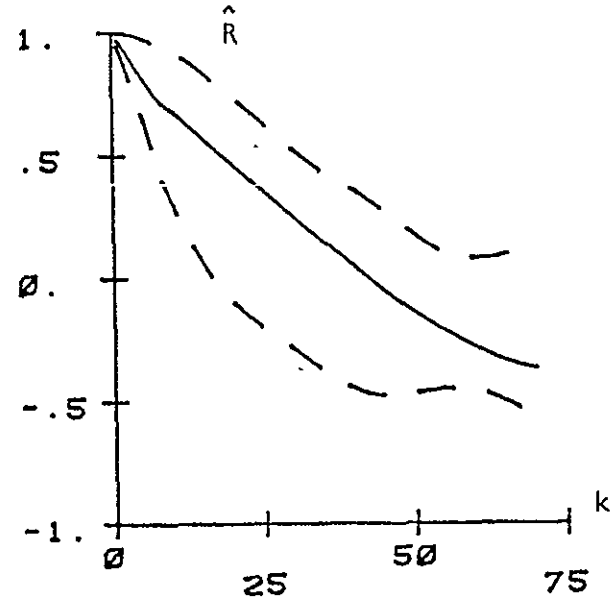


Fig. IV-49. Correlogram, Band 8, Combined Scene

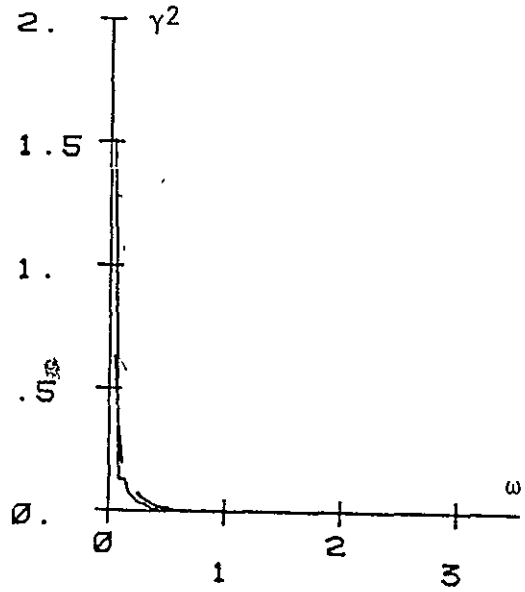


Fig. IV-50. Periodogram, Band 8, Combined Scene

Hence, we have validated models for band 8 of both scene types.

I) Band 9--

Validation of the selected sixth order integrated autoregressive model for band 9 of the wheat scene is considered first.

1) Zero Mean Test

The selected model gives the test statistic

$$|t(x)| = 8.142 \times 10^{-1} \qquad (4\text{-}34)$$

Thus, the selected model easily passes this test.

2) Serial Independence Test

The test statistic for the selected model is

$$\eta(x) = 9.915 \qquad (4\text{-}35)$$

for $n1 = 15$. Therefore, this test is passed by the selected model.

3) Other Tests

The selected model passes the cumulative periodogram, correlogram, and periodogram tests as is seen in Figures IV-51, IV-52, and IV-53.

Next, validation of the selected first order autoregressive model for band 9 of the combined scene is considered.

4) Zero Mean Test

The test statistic for the selected model is

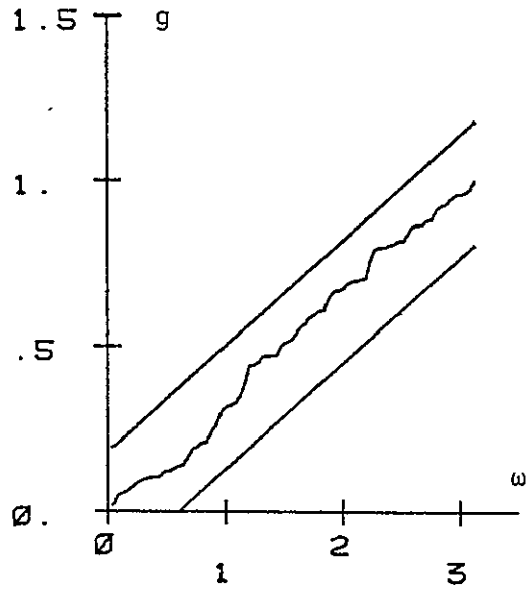


Fig. IV-51. Cumulative Periodogram, Band 9, Wheat Scene

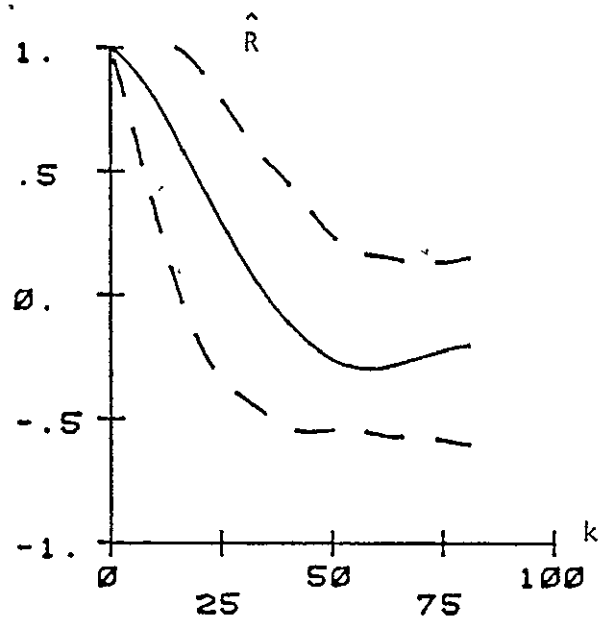


Fig. IV-52. Correlogram, Band 9, Wheat Scene

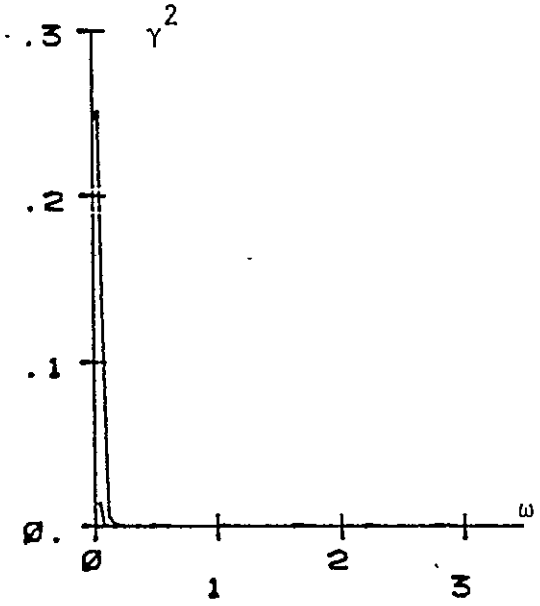


Fig. IV-53. Periodogram, Band 9, Wheat Scene

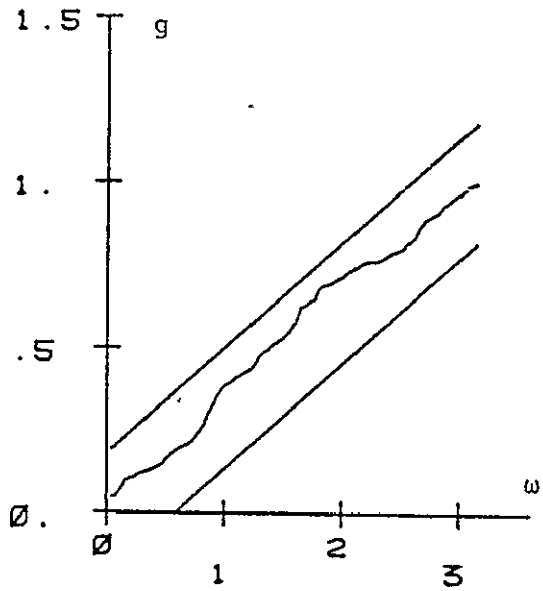


Fig. IV-54. Cumulative Periodogram, Band 9, Combined Scene

$$|t(x)| = 9.324 \times 10^{-1} \qquad (4\text{-}36)$$

Thus, the selected model easily passes this test.

5) Serial Independence Test

The selected model gives the test statistic

$$\eta(x) = 1.362 \times 10^{1} \qquad (4\text{-}37)$$

for n1 = 15. Therefore, the selected model passes this test.

6) Other Tests

Figures IV-54, IV-55, and IV-56 show that the selected model passes the cumulative periodogram, correlogram, and periodogram validation tests.

Hence, we have validated models for band 9 of both scene types.

5. Conclusion

Models of nine spectral bands for two empirical data sets have been identified, selected and validated. The validated models of the two scene types are given in Table IV-42 for easy reference.

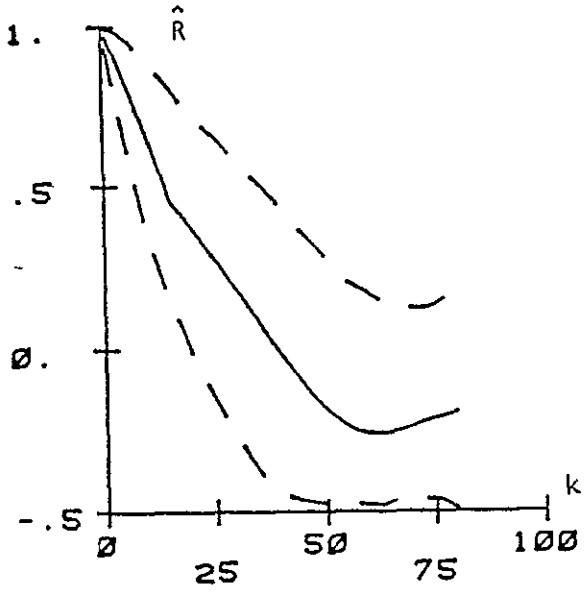


Fig. IV-55. Correlogram, Band 9, Combined Scene

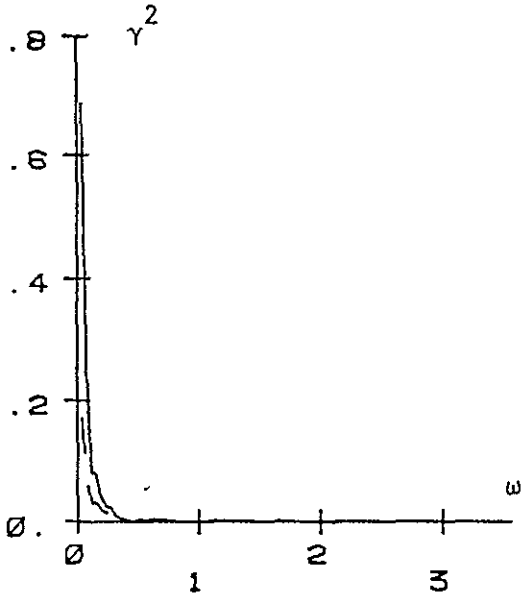


Fig. IV-56. Periodogram, Band 9, Combined Scene

Table IV-42. Validated Models

Band	Wheat Scene	Combined Scene
1	AR(6)	$IAR_2(11)$
2	AR(2)	AR(2)
3	IAR(11)	IAR(11)
4	ARC(1)	ARC(1)
5	AR(1)	AR(3)
6	ARC(2)	AR(1)
7	IAR(5)	ARC(9)
8	IAR(8)	IAR(8)
9	IAR(6)	AR(1)

Chapter V

An Application of Information Theoretic Techniques

1. Introduction

This chapter demonstrates an application of the information theoretic techniques developed in Chapter II to studying some parameters of multispectral scanner systems. In particular, the techniques are applied to the models constructed in Chapter IV for the two spectral scene types under consideration in this research. The average information criterion is used to select a subset of spectral bands. An attempt at estimation of classification accuracy for the hypothetical multispectral scanner is discussed.

2. Average Information Studies

The average information computation techniques developed in Chapter II are used to study average information in the received spectral process about the spectral response process of the scene under observation. Recall that we are representing the spectral process received by the multispectral scanner by

$$y(k) = S(k) + n(k) \quad , \quad k\epsilon[\lambda_1, \lambda_2] \tag{5-1}$$

where

S(k) is the spectral response process of the scene

and

n(k) is the disturbance or noise process.

The models constructed in Chapter IV are used for representing the spectral response process S(k) in each spectral band. As discussed in Chapter II, n(k) is assumed to be white noise with possibly different power spectral density levels in different spectral bands.

The first computation is average information in y(k) about S(k) as a function of spectral bandwidth for each spectral band of both scene types. The average information is computed for several values of the variance of the noise disturbance, σ_n^2. Since the noise disturbance is assumed to be of constant power spectral density level for each spectral band, considering several values of σ_n^2 has the effect of allowing the study of average information for several signal-to-noise ratio (SNR) conditions. Thus, the objective of studying the effects of spectral bandwidth and signal-to-noise ratios as parameters of multispectral scanners is achieved in these computations. The average information computations are made with the use of a computer program written in FORTRAN. A copy of the computer program is included in Appendix III for reference. The results of these computations are displayed graphically in Figures V-1 to V-18. It is noted that these figures have curves plotted with σ_n^2 as a running parameter. Also, the curves are plotted as a function of the number of points in the spectral interval. This has

the advantage of making the curves applicable to the same models for different spectral intervals.

Table V-1 gives the total average information for the defined spectral bands of the wheat scene for several values of the noise variance σ_n^2. Table V-2 gives similar data for the combined scene. When considering the results in Tables V-1 and V-2, it must be remembered that the spectral bands are of different spectral bandwidths. Hence, the average information in spectral bands of approximately the same spectral bandwidth may be more useful in selecting subsets of bands.

Thus the technique considered in this research is used to compute average information using the spectral models constructed for the defined spectral bands.

3. Selection of a Subset of Spectral Bands

We now demonstrate a simple application of using average information to select a subset of spectral bands for inclusion on a multispectral scanner. For the purpose of this demonstration we make the following assumptions. First assume that a subset of six of the defined spectral bands is derived. This is not an unusual number of spectral bands to be used in an application (i.e. scene classification). The second assumption concerns the amount of observation noise to include in each spectral band. For the purposes of this simple example, we assume that the variance of the observation noise is $\sigma_n^2 = 10^{-3}$ for all the defined spectral bands. This is clearly not a realistic assumption, but is sufficient for our simple demonstration. Third, it is assumed we are interested in ordering the preference of spectral bands on the basis of

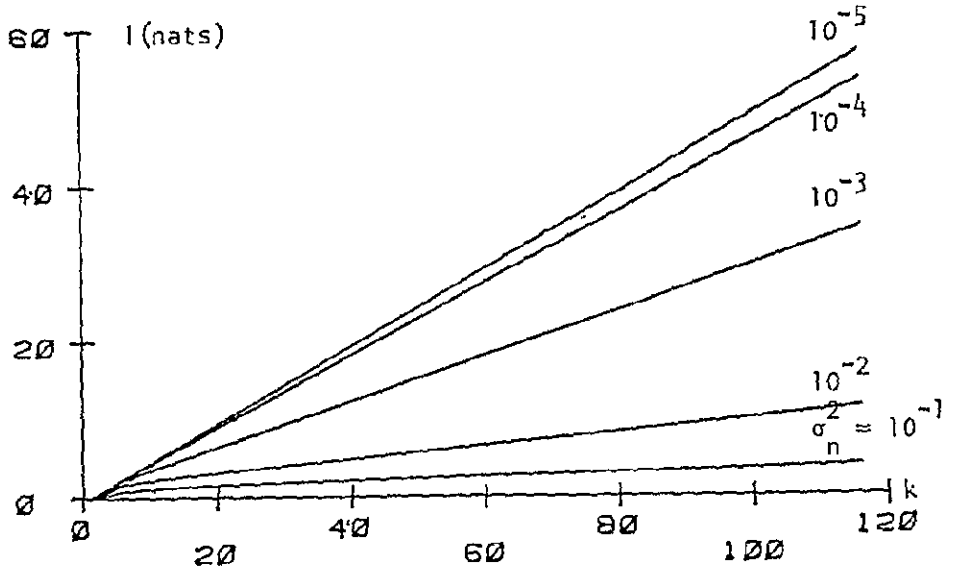


Fig. V-1. Average Information, Band 1, Wheat Scene

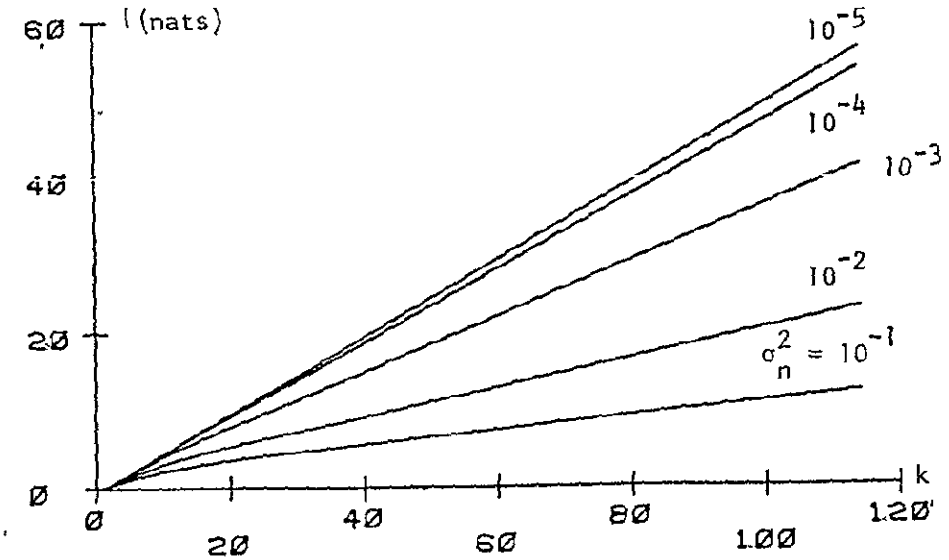


Fig. V-2. Average Information, Band 1, Combined Scene

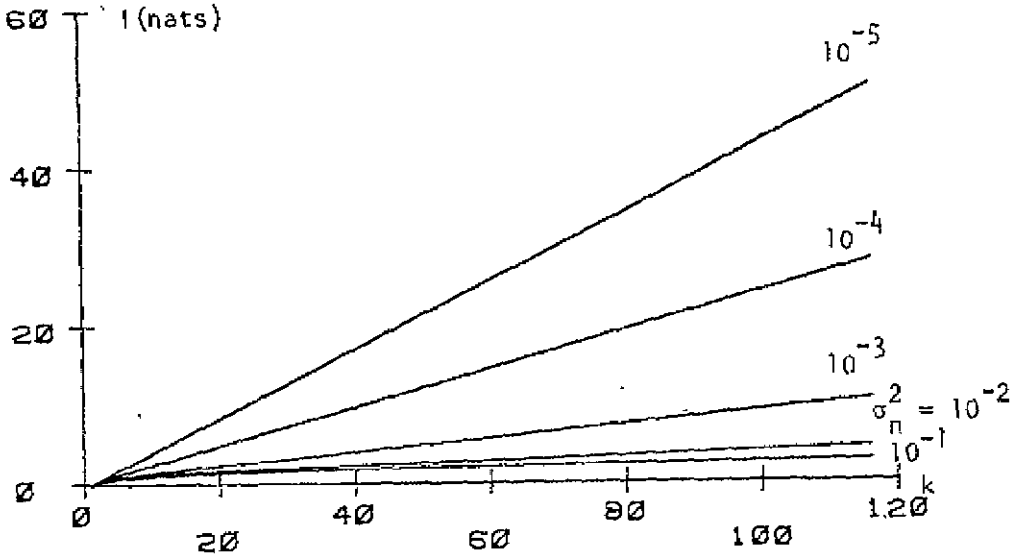


Fig. V-3. Average Information, Band 2, Wheat Scene

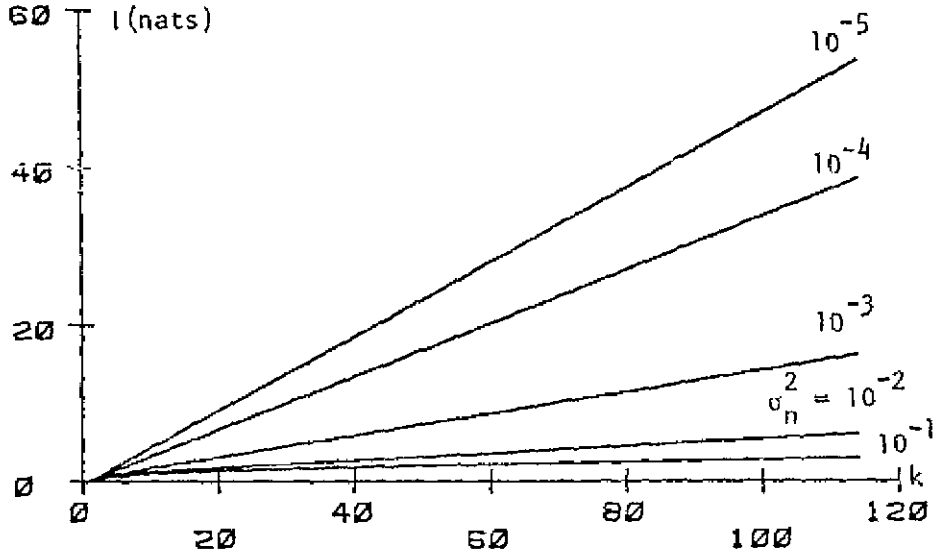


Fig. V-4. Average Information, Band 2, Combined Scene

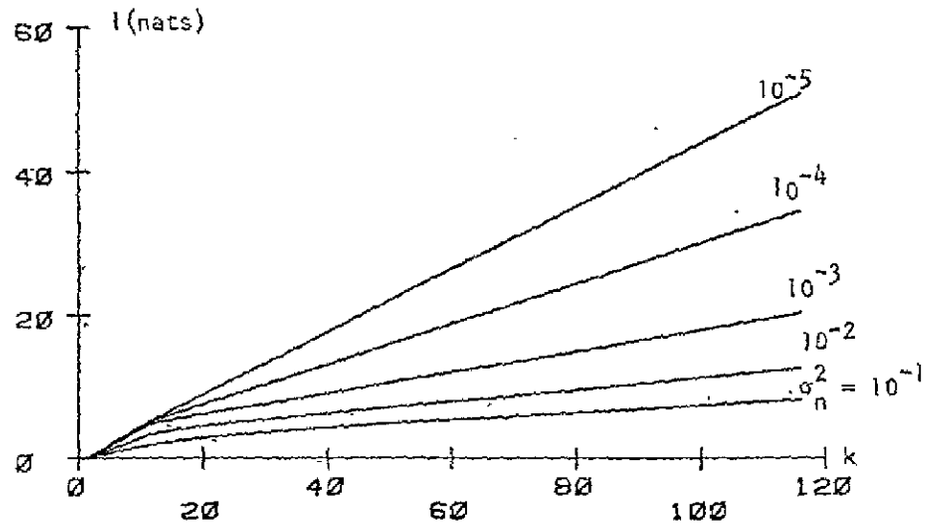


Fig. V-5. Average Information, Band 3, Wheat Scene

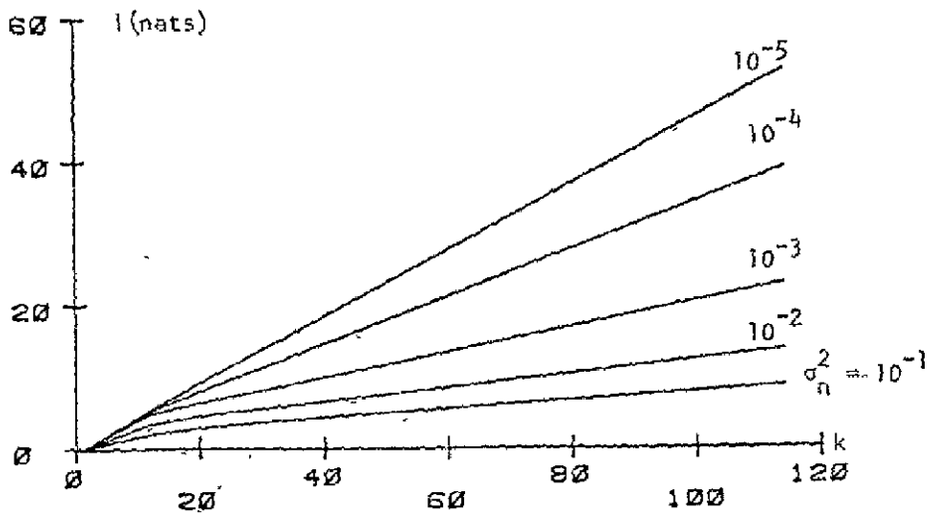


Fig. V-6. Average Information, Band 3, Combined Scene

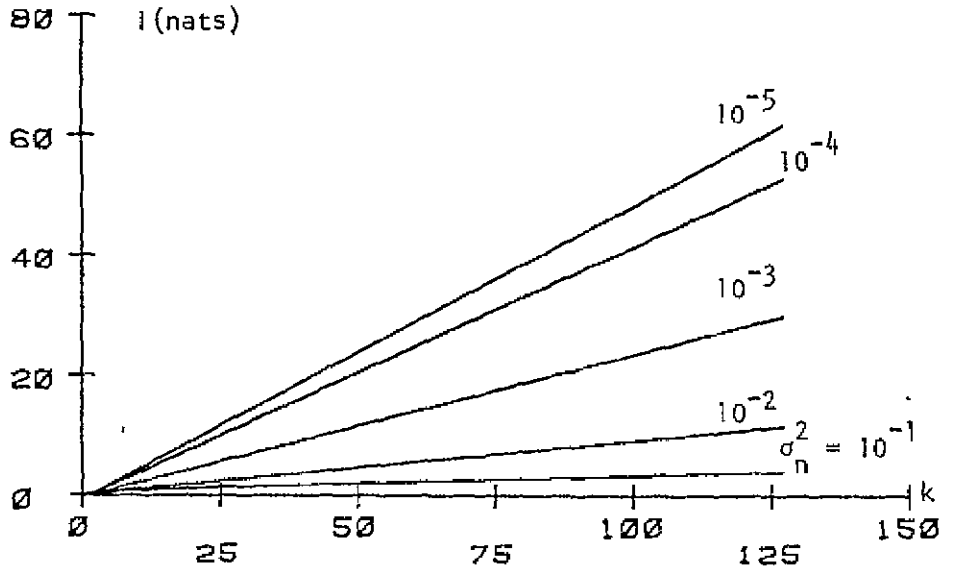


Fig. V-7. Average Information, Band 4, Wheat Scene

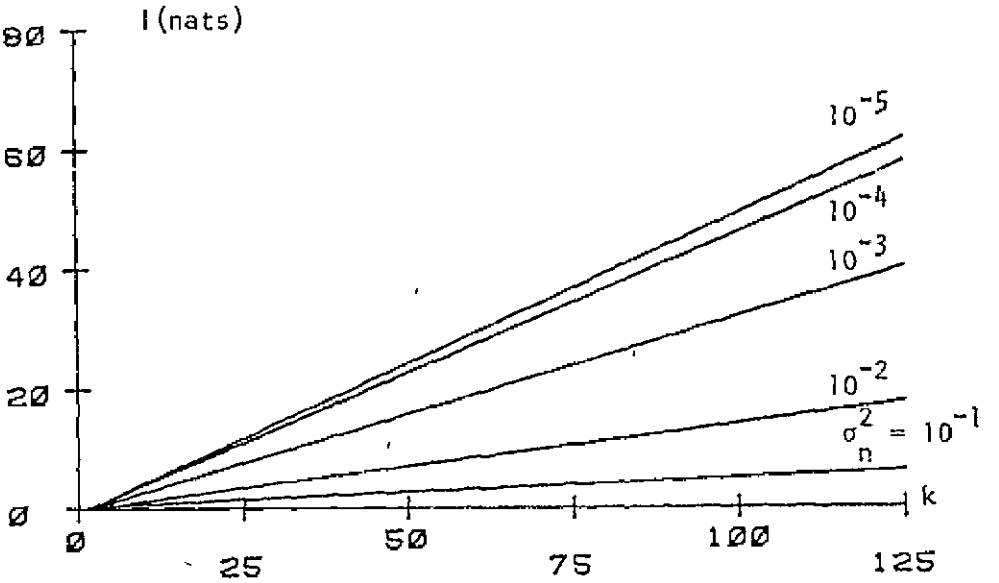


Fig. V-8. Average Information, Band 4, Combined Scene

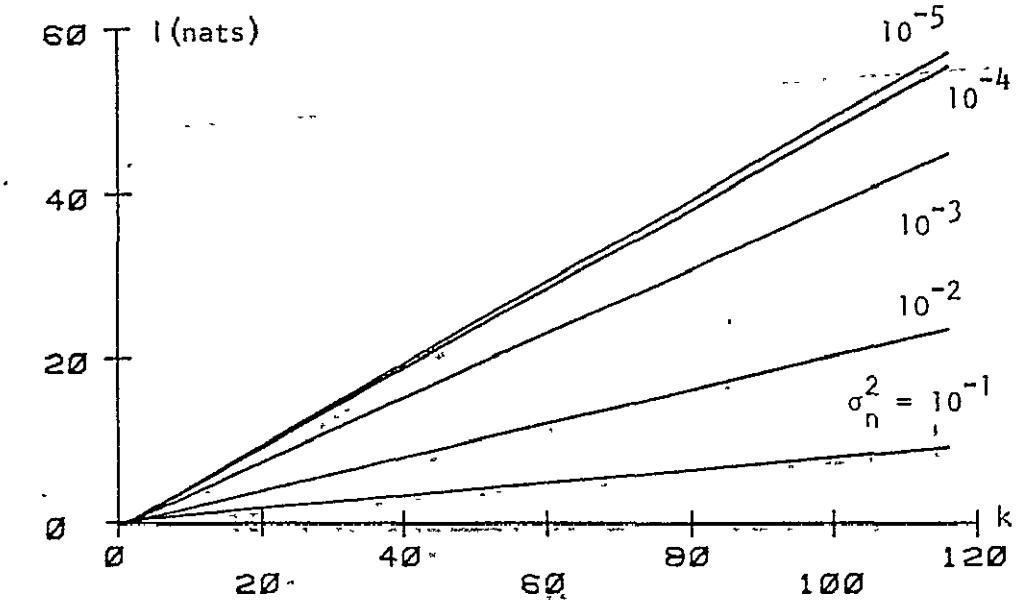


Fig. V-9. Average Information, Band 5, Wheat Scene

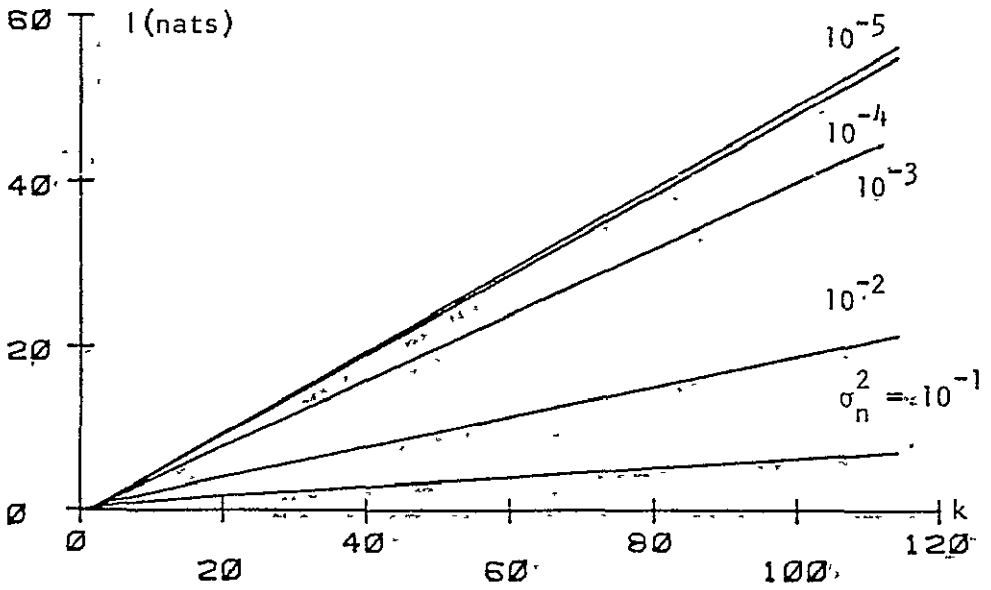


Fig. V-10. Average Information, Band 5, Combined Scene

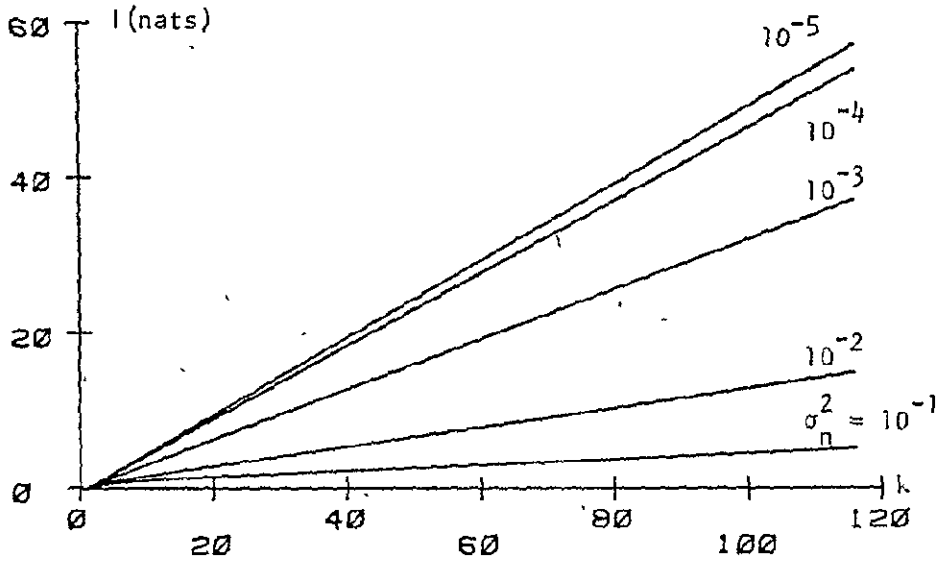


Fig. V-11. Average Information, Band 6, Combined Scene

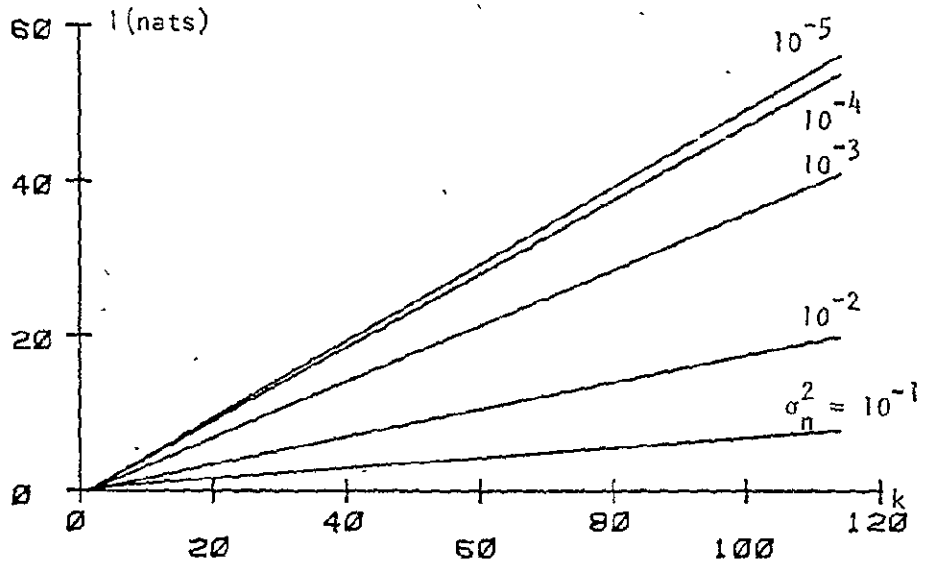


Fig. V-12. Average Information, Band 6, Combined Scene

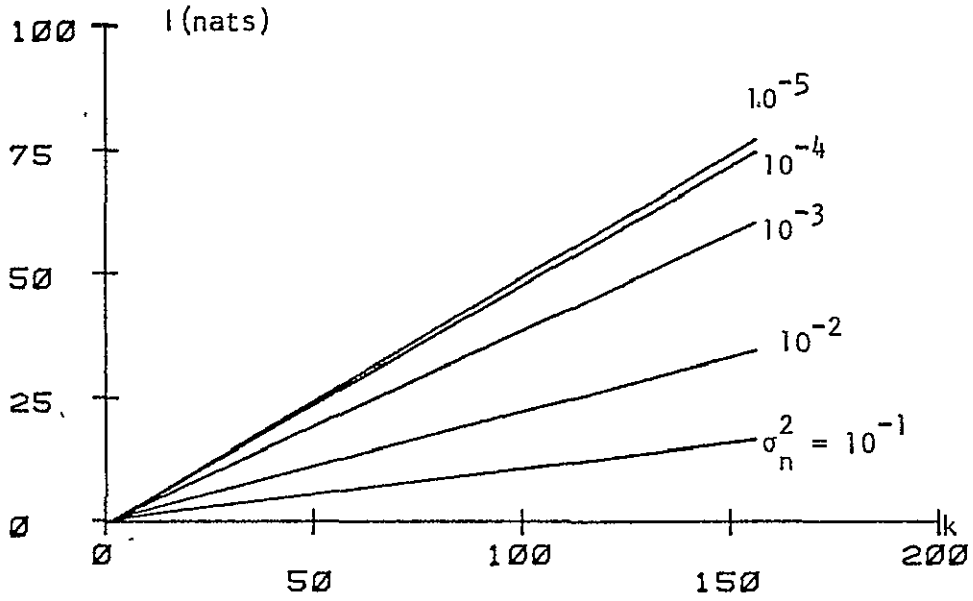


Fig. V-13. Average Information, Band 7, Wheat Scene

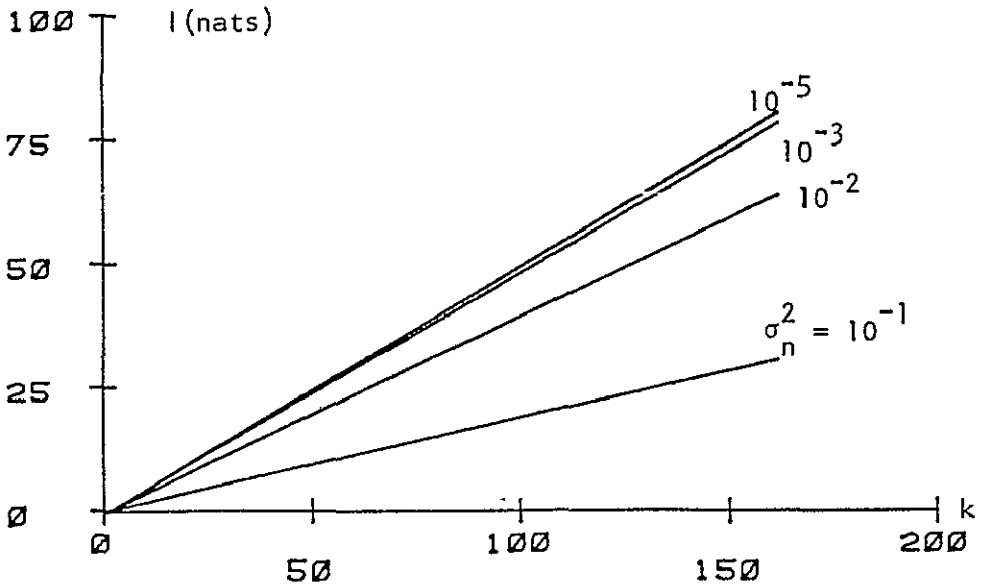


Fig. V-14. Average Information, Band 7, Combined Scene

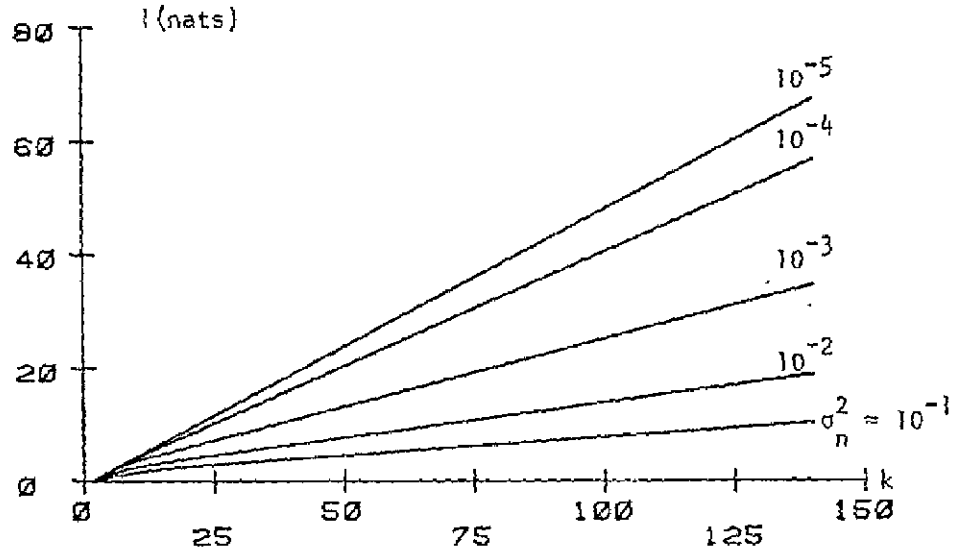


Fig. V-15. Average Information, Band 8, Wheat Scene

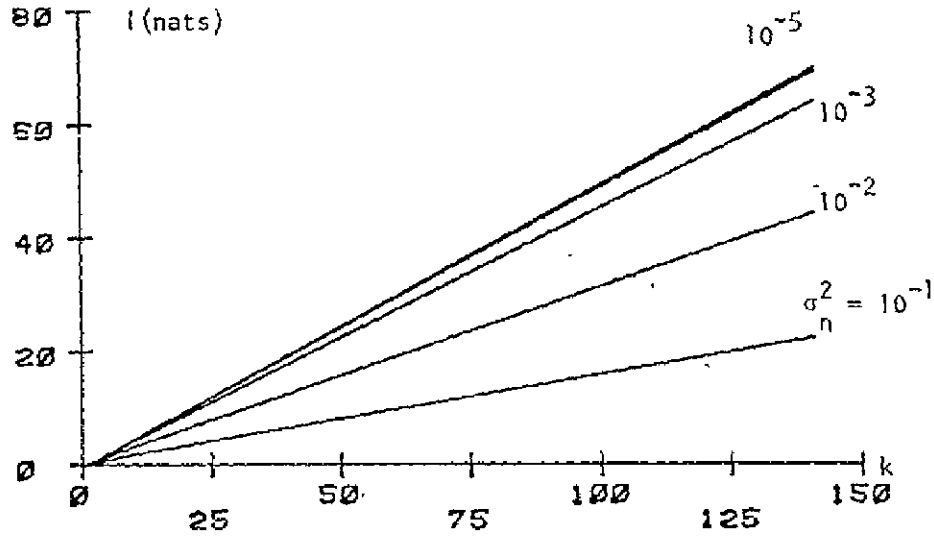


Fig. V-16. Average Information, Band 8, Combined Scene

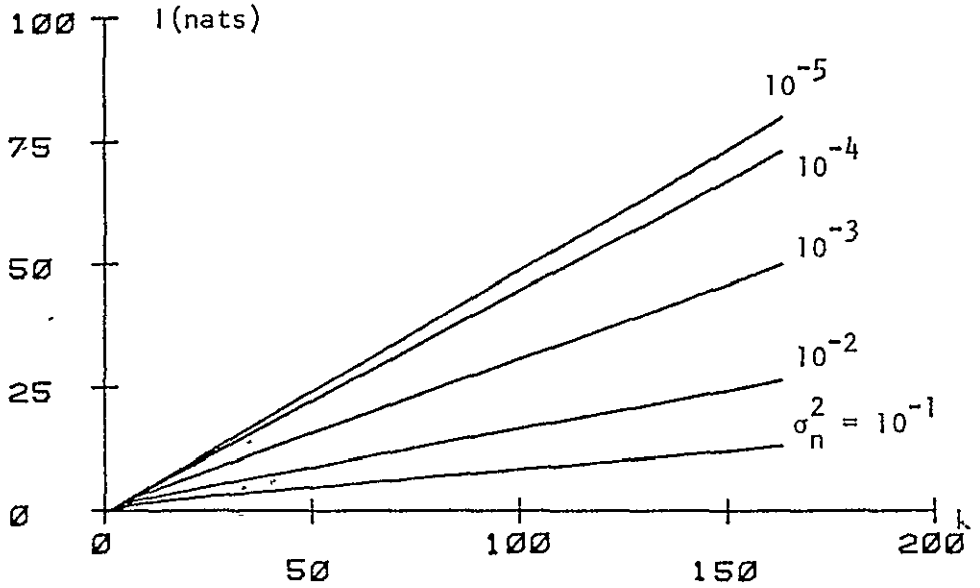


Fig. V-17. Average Information, Band 9, Wheat Scene

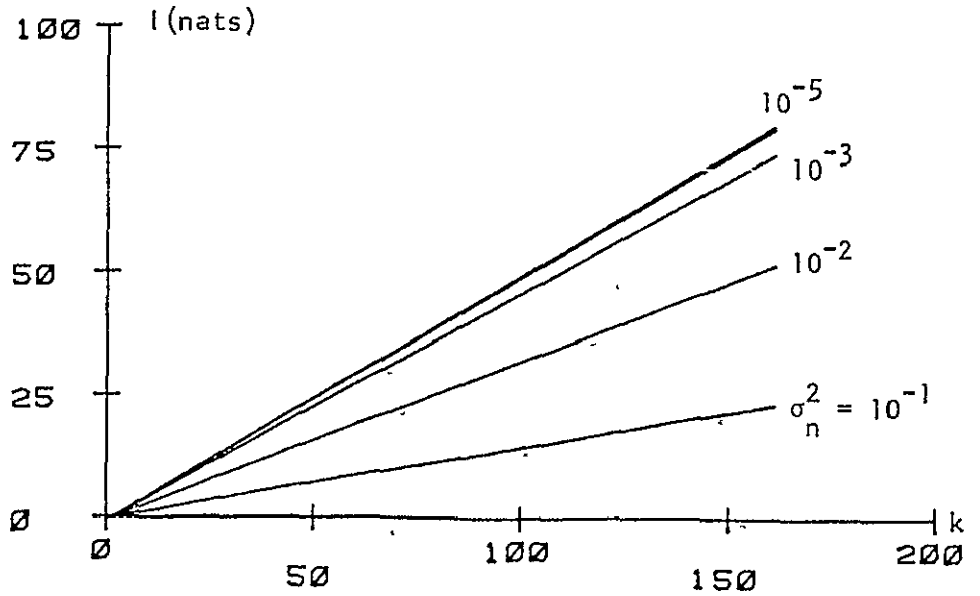


Fig. V-18. Average Information, Band 9, Combined Scene

the average information each band would have in an equal spectral bandwidth. This tends to ameliorate the effect of wider spectral bands having more average information due only to their larger spectral bandwidth. It is thought that this method of comparison will tend to select the subset of spectral bands with the highest amount of average information with each band competing on a more equal basis. Based on these assumptions, the spectral bands are ranked in order of preference in Table V-3.

Table V-1. Average Information for Wheat Scene Bands

Band	Noise Variance, σ_n^2				
	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
1	57.07	53.49	34.50	11.43	3.95
2	50.05	28.09	10.52	4.45	2.75
3	51.02	34.59	20.35	12.65	8.28
4	61.64	52.92	30.00	11.69	4.15
5	57.30	55.58	44.96	23.55	9.31
6	57.11	53.90	37.20	14.81	5.12
7	77.19	74.63	60.31	34.59	16.6
8	67.56	56.77	34.80	18.96	10.48
9	80.04	73.23	50.10	26.53	13.10

Note: The information values are given in nats here.

Table V-2. Average Information for Combined Scene Bands

Band	Noise Variance, σ_n^2				
	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
1	56.19	53.70	41.33	23.09	12.23
2	53.51	38.54	16.17	6.10	3.05
3	52.73	39.10	22.98	13.73	8.72
4	61.49	57.54	40.08	17.71	6.24
5	56.36	55.11	45.73	21.44	7.09
6	56.21	53.81	40.96	20.05	7.83
7	80.48	80.26	78.25	63.93	30.54
8	69.93	69.31	64.15	44.38	22.46
9	79.93	79.33	74.19	51.72	23.28

Table V-3. Order of Preference of Spectral Bands for the Wheat and Combined Scenes

Rank	Wheat Scene Band	Combined Scene Band
1	5	7
2	7	9
3	6	8
4	9	5
5	1	1
6	8	6
7	4	4
8	3	3
9	2	2

It is noted in Table V-3 that although the ordering is different, the six highest ranking bands are the same for both the wheat scene and the combined scene. Band 1 is in the visible region of the spectrum for both scene types (see Tables IV-3 and IV-4). The other five preferred bands are all in the infrared portion of the spectrum. Thus relative to our averge information criterion, the infrared portion of the spectrum is generally preferred to the visible portion of the spectrum since bands 2 and 3 are ranked lowest for both the wheat scene and the com-

bined scene. This tends to indicate that future multispectral scanners systems should have more spectral bands in the infrared portion of the spectrum. Indeed this is the case with the thematic mapper to be placed on LANDSAT-D (see, for example, reference [L2]).

Of course, if there were different levels of noise disturbance in different spectral bands, the order of preference could be entirely different. This research does, however, provide a systematic method for dealing with such circumstances. A more realistic application of this technique would require such an approach.

One of the major uses of data obtained with multispectral scanner systems is classification of the observed scenes. Thus it is felt that estimation of classification performance gives an important measure of the usefulness of a proposed subset of spectral bands. The estimation of classification error is an important and complicated topic in itself. Whitsitt and Landgrebe [W1] have recently spent considerable effort on this topic. A technique, used in this research, to estimate classification performance was developed by Lissack and Fu [L3]. This technique assumes a Bayesian classification technique and provides a computational technique for estimating classification performance. An attempt was made to use the Lissack-Fu technique to estimate the classification performance of the six selected spectral bands for the two scene types studied in this research. The empirical data, however, had the unfortunate property of producing covariance matrices that were singular to the numerical accuracy of the (IBM 370) computer system used in the research. Hence, a meaningful estimate of the classification performance was not possible with the present data set. Therefore, the clas-

sification performance characteristics of the selected spectral bands are left for future investigation.

4. Conclusions

This chapter has demonstrated the application of information theoretic techniques for the study of some parameters of multispectral scanner systems. First, average information in a received spectral band was calculated for several power spectral density levels of observation noise. This computation allowed the study of such parameters as spectral bandwidth and signal-to-noise effects on average information. Secondly, a simple demonstration of the use of average information as a technique for selection of a subset of spectral bands was given. Finally, an attempt at studying the classification performance of the selected subset of spectral bands indicated that a more detailed investigation of the problem was necessary. This was left for future investigation.

Chapter 6

Conclusions

1. Discussion

This thesis is devoted to development of techniques for analysis of some parameters of multispectral scanner systems. These techniques represent an initial effort to provide an analytical framework for what heretofore has been approached mainly in an empirical and ad hoc manner. The information theoretic techniques developed in Chapter II are sufficiently general that they can be used to explore many different practical questions in the study of parameters of multispectral scanner systems for remote sensing. The modeling techniques developed in Chapter III are applicable to almost any scene type of interest in remote sensing. Furthermore, models developed in such a manner could, of course, be used for other research on spectral scenes. Chapters IV and V are an extended study on empirical data using the techniques developed in Chapters II and III. Chapter IV demonstrates the advantages of a systematic approach to model construction by examination of several hypothesized models. Thus several alternative models are constructed for each spectral scene. Chapter V demonstrates that, for the empirical data studied, the infrared portion of the spectrum deserves increased attention in multispectral scanner system design.

2. Further Research

There are several aspects of this research that merit further exploration. Some of the more obvious topics are mentioned.

(1) It may be of considerable practical interest to extend the information theoretic results in Chapter II to the case of nonwhite observation noise. Though more complicated, an expression for average information can again be related to the optimum Wiener-Hopf filter impulse response [H1]. The state variable formulation of this problem would prove to be very useful for handling observation noise that could be described by a dynamic model. An application of this extension might be to study the effects of an extraneous spectral signal such as produced by bare soil surrounding a vegetation scene of actual interest.

(2) Another extension of this reseach might be to consider other models for spectral scenes. In particular, it may be fruitful to consider moving average models or combined autoregressive-moving average models [B1]. Such an extension may result in lower order models for spectral scenes. However, identification of such models is more complicated than the cases considered in this thesis [K3].

(3) Extension of the scalar models to the vector model case might be interesting. This could have an application in temporal studies of spectral scenes. That is, models of the spectral response of vegetation scenes and their change over the growing season would be very useful when considering multispectral scanner system design.

(4) An important extension of this research is the consideration of the relationship between the information theoretic methods for selecting a subset of spectral bands and the accuracy of scene classifi-

cation using these bands. A simple experiment of this nature was attempted in Chapter V and met with difficulties in estimation of some required covariance matrices. Nevertheless, it would be extremely useful to study the efficacy of the information theoretic band selction application in relation to the classification problem. It is expected that such things as types of models used for the spectral scenes and, indeed, the particular spectral scenes considered, would cause this to be a wide ranging and complicated study. Different classification tecniques might be expected to produce widely differing results when using a set of bands selected by the information theoretic criterion. Indeed, proper consideration of the many variables in such a study has been [W1] and should continue to be an area of fruitful research.

BIBLIOGRAPHY

BIBLIOGRAPHY

[A1] Akaike, H., "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, pp. 716-723, December 1974.

[A2] Akaike, H., "An Information Criterion For Time Series Model Fitting," Systems Identification: Advances and Case Studies, R. K. Mehra and D. G. Lainiotis, Editors, Academic Press, New York, 1976.

[A3] Akaike, H., "Autoregressive Model Fitting for Control," Annals of the Institute of Statistical Mathematics, Vol. 23, No. 2, pp. 163-180, 1971.

[A4] Anderson, T. W., The Statistical Analysis of Time Series, Wiley, New York, 1971.

[A5] Abramowitz, M. and Stegun, I., Handbook of Mathematical Functions, Dover, New York, 1965.

[A6] Akaike, H., "Fitting Autoregressive Models for Prediction," Annals of the Institute of Statistical Mathematics, Vol. 21, No. 2, pp. 243-247, 1969.

[B1] Box, G. E. P. and Jenkins, G. M., Time Series Analysis, Revised Edition, Holden-Day, San Francisco, 1976.

[B2] Box, G. E. P. and Pierce, D. A., "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," Journal of the American Statistical Association, Vol. 65, No. 332, pp. 1509-1526, 1970.

[B3] Bartlett, M. S., An Introduction to Stochastic Processes, University Press, Cambridge, 1956.

[C1] Courant, R. and Hilbert, D., Methods of Mathematical Physics, Vol. 1, Interscience Publishers, New York, 1953.

[C2] Cullen, C. G., Matrices and Linear Transformations, Addison-Wesley, Reading, MA, 1966.

[C3] Coggeshall, M. E. and Hoffer, R. M., "Basic Forest Cover Mapping Using Digitized Remote Sensor Data and ADP Techniques," LARS Information Note 030573, The Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN.

[F1] Fu, K. S., Landgrebe, D. A., and Phillips, T. L., "Information Processing of Remotely Sensed Data," Proceedings of the IEEE, Vol. 57, No. 4, pp. 639-653, April 1969.

[F2] Ferguson, T. S., Mathematical Statistics, Academic Press, New York, 1967.

[F3] Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972.

[F4] Fano, R. M., Transmission of Information, M.I.T. Press, Cambridge, 1961.

[G1] Gallager, R. G., Information Theory and Reliable Communication, Wiley, New York, 1968.

[G2] Gelfand, I. M. and Yaglom, A. M., "Calculation of the Amount of Information About A Random Function Contained In Another Such Function," American Mathematical Society Translation, Series 2, Vol. 12, pp. 199-250, 1959.

[G3] Gates, D. M., Keegan, H. J., Schleter, J. C., and Weidner, V. R., "Spectral Properties of Plants," Applied Optics, Vol. 4, No. 1, pp. 11-20, January 1965.

[H1] Huang, R. Y., An Information Theory for Time-Continuous Processes, Ph.D. Thesis, Syracuse University, 1962.

[H2] Huang, R. Y. and Jonson, R. A., "Information Transmission with Time-Continuous Random Processes," IEEE Transactions on Information Theory, Vol. IT-9, pp. 84-95, April 1963.

[H3] Holmes, R. A. and MacDonald, R. B., "The Physical Basis of System Design for Remote Sensing in Agriculture," Proceedings of the IEEE, Vol. 57, No. 4, pp. 629-639, April 1969.

[J1] Jenkins, G. M. and Watts, D. G., Spectral Analysis, Holden-Day, San Francisco, 1968.

[K1] Kalman, R. E. and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," Transactions ASME, Journal of Basic Engineering, Vol. 83D, pp. 95-108, March 1961.

[K2] Kalman, R. E., "A New Approach to the Linear Filtering and Prediction Problem," Transactions ASEM, Journal of Basic Engineering, Vol. 82D, pp. 34-45, March 1960.

[K3] Kashyap, R. L. and Rao, A. R., Dynamic Stochastic Models from Empirical Data, Academic Press, New York, 1976.

[K4] Kashyap, R. L., "A Bayesian Comparison of Different Classes of Dynamic Models Using Empirical Data," IEEE Transactions on Automatic Control, Vol. AC-22, No. 5, pp. 715-727, October 1977.

[K5] Kumar, R. and Silva, L. F., "Staistical Separability of Agricultural Cover Types In Subsets of One to Twelve Spectral Channels," LARS Information Note, The Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN.

[K6] Kashyap, R. L., "Maximum Likelihood Identification of Stochastic Linear Models," IEEE Transactions on Automatic Controls, Vol. AC-15, No. 1, pp. 25-34, February 1970.

[L1] Leamer, R. W., Myers, V. I., and Silva, L. F., "A Spectroradiometer for Field Use," Reviews of Scientific Instruments, Vol. 44, No. 5, pp. 611-614, May 1973.

[L2] Landgrebe, D., Biehl, L., and Simmons, W., "An Empirical Study of Scanner System Parameters," LARS Information Note 110976, The Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN.

[L3] Lissack, T. and Fu, K. S., "Error Estimation in Pattern Recognition via L^{α}-Distance Between Posterior Density Functions," IEEE Transactions on Information Theory, Vol. IT-22, No. 1, pp. 34-45, January 1976.

[L4] Landgrebe, D. A., Biehl, L. L., and Simmons, W. R., "An Empirical Study of Scanner System Paameters," IEEE Transactions on Geoscience Electronics, Vol. GE-15, No. 3, pp. 120-130, July 1977.

[M1] Meditch, J. S., Stochastic Optimal Linear Estimation and Control, McGraw-Hill, New York, 1969.

[O1] Oppenheim, A. V. and Schafer, R. W., Digital Signal Processing, Prentice-Hall, 1975.

[P1] Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965.

[R1] Roussas, G. G., A First Course in Mathematical Statistics, Addison-Wesley, Reading, MA, 1973.

[S1] Shannon, C. E., "A Mathematical Theory of Communication," Bell System Technical Journal, Vol. 27, pp. 379-423, July 1948 and Vol. 27, pp. 623-656, October 1948.

[S2] Schwartz, R. J. and Friedland, B., Linear Systems, McGraw-Hill, New York, 1965.

[S3] Sage, A. P. and Melsa, J. L., Estimation Theory with Applications to Communications and Control, McGraw-Hill, New York, 1971.

[S4] Sinclair, T. R., Hoffer, R. M., and Schreiber, M. M., "Reflectance and Internal Structure of Leaves from Several Crops During A Growing Season," Agronomy Journal, Vol. 63, pp. 864-868, November-December 1971.

[S5] Saridis, G. N., "Stochastic Approximation Methods for Identification and Control--A Survey," IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, pp. 798-809, December 1974.

[S6] Saridis, G. N. and Stein, G., "Stochastic Aproximation Algorithms for Linear Discrete-Time System Identification," IEEE Transactions on Automatic Control, Vol. AC-13, No. 5, pp. 515-523, October 1968.

[S7] Saridis, G. N., "Comparison of Six On-Line Identification Algorithms." Automatica, Vol. 10, pp. 69-79, January 1974.

[T1] Tomita, Y., Omatu, S., and Soeda, T., "An Applicatin of Information Theory to Filtering Problems," Information Sciences, Vol. 11, pp. 13-27, 1976.

[V1] VanTrees, H. L., Detection, Estimation, and Modulation Theory, Wiley, New York, 1968.

[W1] Whitsitt, S. J. and Landgrebe, D. A., "Error Estimation and Separability Measures in Feature Selection for Multiclass Pattern Recognition," LARS Information Note 082377, The Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

APPENDICES

Appendix I

Kalman Filter Algorithm

The Kalman filter algorithm is included as a reference for Chapter II. Derivation of the algorithm is fully explained by Sage and Melsa [S3] and Meditch [M1]. The algorithm is stated here in the manner of Sage and Melsa [S3, Chapter 7].

The system model is given as

$$\underline{x}(k+1) = \underline{\Phi}(k+1, k)\underline{x}(k) + \underline{\Gamma}(k)\underline{w}(k) \tag{1}$$

and

$$\underline{z}(k) = \underline{H}(k)\underline{x}(k) + \underline{v}(k) \tag{2}$$

where

$\underline{x}(k)$ is the state vector

$\underline{\Phi}(k+1,k)$ is the state transition matrix

$\underline{\Gamma}(k)$ is a matrix

$\underline{w}(k)$ is the driving noise vector

$\underline{z}(k)$ is the observation vector

$\underline{H}(k)$ is a matrix

$\underline{v}(k)$ is the observation noise vector .

The assumed prior statistics are given as

$$E[\underline{w}(k)] = 0 = E[\underline{v}(k)] \quad \text{all } k \tag{3}$$

$$E[\underline{x}(0)] = \underline{\mu}_x(0) \tag{4}$$

$$\text{cov}[\underline{w}(k), \underline{w}(j)] = \underline{V}_w(k)\delta(j-k) \tag{5}$$

$$\text{cov}[\underline{v}(k), \underline{v}(j)] = \underline{V}_v(k)\delta(j-k) \tag{6}$$

$$\text{cov}[\underline{w}(k), \underline{v}(j)] = \text{cov}[\underline{x}(0), \underline{w}(k)]$$

$$= \text{cov}[\underline{x}(0), \underline{v}(k)] = \underline{0} \tag{7}$$

where

$$\delta(j-k) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} . \tag{8}$$

These assumptions give the Kalman filter algorithm for the estimate, $\underline{\hat{x}}(k)$, of $\underline{x}(k)$. The estimate is

$$\underline{\hat{x}}(k) = \underline{\Phi}(k,k-1)\underline{\hat{x}}(k-1) + \underline{K}(k)\left[\underline{z}(k) - \underline{H}(k)\underline{\Phi}(k;k-1)\underline{\hat{x}}(k-1)\right] \tag{9}$$

where

$$\underline{K}(k) = \underline{V}_{\underline{x}}(k/k-1)\underline{H}^T(k)\left[\underline{H}(k)\underline{V}_{\underline{x}}(k/k-1)\underline{H}^T(k) + \underline{V}_v(k)\right]^{-1}$$

$$= \underline{V}_{\underline{x}}(k)\underline{H}^T(k)\underline{V}_v^{-1}(k) \tag{10}$$

$$\underset{\sim}{V}_{\underset{\sim}{x}}(k) = [\underset{\sim}{I} - \underset{\sim}{K}(k)\underset{\sim}{H}(k)]\underset{\sim}{V}_{\underset{\sim}{y}}(k/k-1) \tag{11}$$

$$\underset{\sim}{V}_{\underset{\sim}{x}}(k+1/k) = \underset{\sim}{\phi}(k+1,k)\underset{\sim}{V}_{\underset{\sim}{x}}(k)\underset{\sim}{\phi}^{T}(k+1,k)$$

$$+ \underset{\sim}{\Gamma}(k)\underset{\sim}{V}_{w}(k)\underset{\sim}{\Gamma}^{T}(k) \tag{12}$$

$$\underset{\sim}{\tilde{x}}(k) = \underset{\sim}{x}(k) - \underset{\sim}{\hat{x}}(k) \tag{13}$$

The ease with which this algorithm can be implemented on a digital computer is evident from the above equations. Sage and Melsa [S3, Chapter 7] give a good discussion of all the terms used above.

Appendix II

Matrix Inversion Lemma

The matrix inversion lemma is included here for reference since it is used in some of the derivations in Chapter III. The formulation and demonstration given here is the same as that given by Sage and Melsa [S5, p. 499-500].

Matrix Inversion Lemma

If for any N x N nonsingular matrix $\underline{A}$ and any two N x M matrices $\underline{B}$ and $\underline{C}$, the two matrices $(\underline{A} + \underline{B}\underline{C}^T)$ and $(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B})$ are nonsingular, then the matrix identity

$$(\underline{A} + \underline{B}\underline{C}^T)^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{B}(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B})^{-1}\underline{C}^T\underline{A}^{-1} \tag{1}$$

is valid.

Proof

Define

$$\underline{D} = \underline{A} + \underline{B}\underline{C}^T \quad . \tag{2}$$

Then since by assumption $\underline{D}$ is nonsingular, we can write

$$\underline{D}^{-1}\underline{D} = \underline{I} = \underline{D}^{-1}\underline{A} + \underline{D}^{-1}\underline{B}\underline{C}^T \quad . \tag{3}$$

Now postmultiply (3) by $\underline{A}^{-1}$ to obtain

$$\underline{A}^{-1} = \underline{D}^{-1} + \underline{D}^{-1}\underline{B}\underline{C}^T\underline{A}^{-1} \quad . \tag{4}$$

Next postmultiply each side of (4) by $\underline{B}$ to obtain

$$\underline{A}^{-1}\underline{B} = \underline{D}^{-1}\underline{B} + \underline{D}^{-1}\underline{B}\underline{C}^T\underline{A}^{-1}\underline{B}$$

$$= \underline{D}^{-1}\underline{B}(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B}) \quad . \tag{5}$$

Now by assumption $(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B})$ is nonsingular. Hence, we can write from (5)

$$\underline{D}^{-1}\underline{B} = \underline{A}^{-1}\underline{B}(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B})^{-1} \quad . \tag{6}$$

Postmultiply by $\underline{C}^T\underline{A}^{-1}$ to obtain

$$\underline{D}^{-1}\underline{B}\underline{C}^T\underline{A}^{-1} = \underline{A}^{-1}\underline{B}(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B})^{-1}\underline{C}^T\underline{A}^{-1} \tag{7}$$

But it is seen from (4) that we can write

$$\underline{A}^{-1} - \underline{D}^{-1} = \underline{D}^{-1}\underline{B}\underline{C}^T\underline{A}^{-1} \quad . \tag{8}$$

Hence using (8) in (7) we can write

$$\underline{A}^{-1} - \underline{D}^{-1} = \underline{A}^{-1}\underline{B}(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B})^{-1}\underline{C}^T\underline{A}^{-1} \quad . \tag{9}$$

and using (2) we can finally write

$$(\underline{A} + \underline{B}\underline{C}^T)^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{B}(\underline{I} + \underline{C}^T\underline{A}^{-1}\underline{B})^{-1}\underline{C}^T\underline{A}^{-1} \quad . \tag{10}$$

This is the desired result.

APPENDIX III

COMPUTER PROGRAMS

```
C
C.....PROGRAM TO IDENTIFY MODELS OF SPECTRAL BANDS BY MAXIMUM LIKELIHOOD
C.....TECHNIQUE.
C
      DIMENSION Y(1500),PSI(1500,1),ZZT(30,30),SP(30,30),S(30,30),ZZTS(3
     30,30),THETA(30),TEMP(30,30),THETAP(30),Z(30,1),SZ(30,1),CORR(30,1),
     AX(1500),D(1500)
C
C.....DATA INPUT
C
      READ(5,1) N
    1 FORMAT(I5)
      READ(5,2)(D(I),I=1,N)
    2 FORMAT(8F10.5)
C
C
C.....M1 MUST BE CHANGED TO CHANGE ORDER OF AR SYSTEM
C
      M1=1
      WRITE(16,71) M1
   71 FORMAT(I5)
      L1=0
C
C.....L1 IS THE NUMBER OF TREND TERMS
C
C.....INITIALIZATION OF THETA AND S(I,J)
C
      ID=2
      Y(1)=0.0
      DO 70 I=ID,N
      Y(I)=D(I)-D(I-1)
   70 CONTINUE
      M=M1+L1
      DO 10 I=1,M
      THETA(I)=.10
      DO 11 J=1,M
      S(I,J)=0.0
      ZZT(I,J)=0.0
      SP(I,J)=0.0
      ZZTS(I,J)=0.0
      TEMP(I,J)=0.0
   11 CONTINUE
      S(I,I)=1.0
   10 CONTINUE
      DO 12 I=1,M
      Z(I,1)=0.0
      CORR(I,1)=0.0
      SZ(I,1)=0.0
   12 CONTINUE
C
C.....INITIALIZATION COMPLETE
C
C.....COMPUTATION OF THETA ESTIMATE FOLLOWS
C
      MM=M+ID
      DO 13 K=MM,N
      DO 17 I=1,M1
      Z(I,1)=Y(K-I)
   17 CONTINUE
      IF(L1.EQ.0) GO TO 15
      Z(M,1)=1.0
   15 CONTINUE
      DO 18 I=1,M
      THETAP(I)=THETA(I)
      DO 19 J=1,M
      ZZT(I,J)=Z(I,1)*Z(J,1)
      SP(I,J)=S(I,J)
   19 CONTINUE
   18 CONTINUE
      DO 40 I=1,M
      SZ(I,1)=0.0
      DO 41 J=1,M
      IF(SP(I,J).EQ.0.0.OR.Z(J,1).EQ.0.0) GO TO 42
      SZ(I,1)=SZ(I,1)+SP(I,J)*Z(J,1)
      GO TO 41
   42 SZ(I,1)=SZ(I,1)+0.0
   41 CONTINUE
```

```
   40 CONTINUE
      TEMP=0.0
      DO 20 I=1,M
      ZTSZ=TEMP+Z(I,1)*SZ(I,1)
      TEMP=ZTSZ
   20 CONTINUE
      DENOM=1.0+ZTSZ
      DO 43 L=1,M
      DO 44 I=1,M
      ZZTS(I,L)=0.0
      DO 45 J=1,M
      IF(ZZT(I,J).EQ.0.0.OR.SP(J,L).EQ.0.0) GO TO 46
      ZZTS(I,L)=ZZTS(I,L)+ZZT(I,J)*SP(J,L)
      GO TO 45
   46 ZZTS(I,L)=ZZTS(I,L)+0.0
   45 CONTINUE
   44 CONTINUE
   43 CONTINUE
      DO 60 L=1,M
      DO 61 I=1,M
      TOP(I,L)=0.0
      DO 62 J=1,M
      IF(SP(I,J).EQ.0.0.OR.ZZTS(J,L).EQ.0.0) GO TO 63
      TOP(I,L)=TOP(I,L)+SP(I,J)*ZZTS(J,L)
      GO TO 62
   63 TOP(I,L)=TOP(I,L)+0.0
   62 CONTINUE
   61 CONTINUE
   60 CONTINUE
      DO 21 I=1,M
      DO 22 J=1,M
      S(I,J)=SP(I,J)-((TOP(I,J))/DENOM)
   22 CONTINUE
   21 CONTINUE
      TEMP=0.0
      DO 23 I=1,M
      THZ=TEMP+THETAP(I)*Z(I,1)
      TEMP=THZ
   23 CONTINUE
      COEFF=Y(K)-THZ
      DO 65 I=1,M
      CORR(I,1)=0.0
      DO 66 J=1,M
      IF(S(I,J).EQ.0.0.OR.Z(J,1).EQ.0.0) GO TO 67
      CORR(I,1)=CORR(I,1)+S(I,J)*Z(J,1)
      GO TO 66
   67 CORR(I,1)=CORR(I,1)+0.0
   66 CONTINUE
   65 CONTINUE
      DO 24 I=1,M
      THETA(I)=THETAP(I)+CORR(I,1)*COEFF
   24 CONTINUE
   13 CONTINUE
C
C.....COMPUTATION OF THETA ESTIMATE COMPLETE
C
C.....COMPUTATION OF RESIDUAL VARIANCE AND SELECTION CRITERION
C
      TEMP=0.0
      MM=M+1
      X(1)=0.0
      DO 25 K=MM,N
      X(K-M)=0.0
      THETAZ=0.0
      DO 26 I=1,M
      Z(I,1)=Y(K-I)
   26 CONTINUE
      IF(L1.EQ.0) GO TO 50
      Z(M,1)=1.0
   50 CONTINUE
      DO 52 I=1,M
      THETAZ=THETAZ+THETA(I)*Z(I,1)
   52 CONTINUE
      R=TEMP+((Y(K)-THETAZ)*(Y(K)-THETAZ))
      TEMP=R
      X(K-M)=Y(K)-THETAZ
   25 CONTINUE
      RHO=R/(FLOAT(N-M))
      CRIT=-((FLOAT(N)/2.0)*ALOG(RHO))-FLOAT(M)
C
C.....RESIDUAL VARIANCE AND CRITERION COMPUTATION COMPLETE
C
C
      WRITE(6,27)
   27 FORMAT(1H1,31HESTIMATED COEFFICIENTS ALPHA(I))
      DO 28 I=1,M
      WRITE(6,29) I,THETA(I)
   29 FORMAT(1H0,5HALPHA(,I2,3H)=,E16.8)
   28 CONTINUE
      WRITE(6,30) RHO
   30 FORMAT(1H1,25HRESIDUAL VARIANCE = ,E16.8)
      WRITE(6,31) CRIT
   31 FORMAT(1H0,22HSELECTION CRITERION = ,E16.8)
   75 CONTINUE
      IN=N-M
      WRITE(6,32) IN
   32 FORMAT(I5)
      WRITE(6,33)(X(I),I=1,IN)
   33 FORMAT(5F12.5)
      END
```

```
C
C.....PROGRAM TO PERFORM ZERO MEAN VALIDATION TEST.
C
      DIMENSION X(1500)
      READ(5,1) N
    1 FORMAT(I5)
      READ(5,2)(X(I),I=1,N)
    2 FORMAT(6F12.4)
      TEMP=0.0
      DO 3 I=1,N
      XM=TEMP+X(I)
      TEMP=XM
    3 CONTINUE
      XBAR=(1.0/FLOAT(N))*XM
      TEMP=0.0
      DO 4 I=1,N
      R=TEMP+(X(I)-XBAR)*(X(I)-XBAR)
      TEMP=R
    4 CONTINUE
      RHO=(1.0/FLOAT(N-1))*R
      ETA=SQRT(FLOAT(N)/RHO)*XBAR
      WRITE(10,5) XBAR
    5 FORMAT(7HXBAR = ,E10.8)
      WRITE(10,6) RHO
    6 FORMAT(6HRHO = ,E10.8)
      WRITE(10,7) ETA
    7 FORMAT(6HETA = ,E10.8)
      END
```



```
C
C.....PROGRAM TO PERFORM THE CUMULATIVE PERIODOGRAM VALIDATION TEST.
C
      DIMENSION X(1300),GAMMA(700),G(700),BU(700),BL(700)
      READ(5,1) N
    1 FORMAT(I5)
      READ(5,2)(X(I),I=1,N)
    2 FORMAT(6F12.4)
      FN=FLOAT(N)/2.0
      NP=IFIX(FN)
      DENOM=0.0
      DO 3 J=1,NP
      SUM1=0.0
      SUM2=0.0
      WJ=(6.283185307*FLOAT(J))/FLOAT(N)
      DO 4 L=1,N
      SUM1=SUM1+X(L)*COS(WJ*FLOAT(L))
      SUM2=SUM2+X(L)*SIN(WJ*FLOAT(L))
    4 CONTINUE
      GAMMA(J)=((2.0*SUM1)/FLOAT(N))**2+((2.0*SUM2)/FLOAT(N))**2
      DENOM=DENOM+GAMMA(J)
    3 CONTINUE
      DO 5 K=1,NP
      FNUM=0.0
      DO 6 J=1,K
      FNUM=FNUM+GAMMA(J)
    6 CONTINUE
      G(K)=FNUM/DENOM
    5 CONTINUE
      FLAM=1.63
      A=FLAM
      DO 7 K=1,NP
      BU(K)=(2.0*FLOAT(K)/FLOAT(N))+A/SQRT(FLOAT(NP))
      BL(K)=(2.0*FLOAT(K)/FLOAT(N))-A/SQRT(FLOAT(NP))
    7 CONTINUE
      WRITE(6,9)
    9 FORMAT(1H1,22HCUMULATIVE PERIODOGRAM)
   13 FORMAT(1H0)
      WRITE(7,20) N2
   20 FORMAT(I5)
      DO 8 K=1,N2
      WK=(6.283185307*FLOAT(K))/FLOAT(N)
      WRITE(6,10) WK,BL(K),G(K),BU(K)
   10 FORMAT(1H ,4(5X,E16.8))
      WRITE(7,21) WK,BL(K),G(K),BU(K)
   21 FORMAT(4F16.8)
      IF(G(K).LT.BL(K).OR.G(K).GT.BU(K)) GO TO 11
      GO TO 8
   11 WRITE(6,12) WK,K
   12 FORMAT(1H ,25HOUTSIDE BOUNDARY AT WK = ,E16.8,9H  OR K = ,I5)
    8 CONTINUE
      END
```

```
C
C.....PROGRAM TO PERFORM THE SERIAL INDEPENDANCE VALIDATION TEST.
C
      DIMENSION X(1440),R2(120),R(120)
      READ(5,1) N
    1 FORMAT(I5)
      READ(5,2)(X(I),I=1,N)
    2 FORMAT(5F12.4)
      FN=1.1*FLOAT(N)
      N1=IFIX(FN)
      R1=0.0
      DO 3 J=1,N
      R1=R1+X(J)**2
    3 CONTINUE
      R0=R1/FLOAT(N)
      DO 4 K=1,N1
      R2(K)=0.0
      K1=K+1
      DO 5 J=K1,N
      R2(K)=R2(K)+X(J)*X(J-K)
    5 CONTINUE
      R(K)=R2(K)/FLOAT(N-K)
    4 CONTINUE
      SUM=0.0
      DO 6 K=1,N1
      SUM=SUM+R(K)**2
    6 CONTINUE
      ETA=(FLOAT(N-N1)*SUM)/(R0**2)
      WRITE(15,7) N1,ETA
    7 FORMAT(5HN1 = ,I5,5X,6HETA = ,E16.8)
      END
```

```
C
C.....PROGRAM TO COMPUTE THE PERIODOGRAM.
C
      DIMENSION X(1500),GAMMA(700)
      READ(5,1) N
    1 FORMAT(I5)
      READ(5,2)(X(I),I=1,N)
    2 FORMAT(5F15.5)
      FN=FLOAT(N)/2.0
      N2=IFIX(FN)
      DO 3 J=1,N2
      SUM1=0.0
      SUM2=0.0
      WJ=(6.2831853072*FLOAT(J))/FLOAT(N)
      DO 4 L=1,N
      SUM1=SUM1+X(L)*COS(WJ*FLOAT(L))
      SUM2=SUM2+X(L)*SIN(WJ*FLOAT(L))
    4 CONTINUE
      GAMMA(J)=((2.0*SUM1)/FLOAT(N))**2+((2.0*SUM2)/FLOAT(N))**2
    3 CONTINUE
      WRITE(6,5)
    5 FORMAT(1H1,11HPERIODOGRAM)
      WRITE(6,6)
    6 FORMAT(1H0)
      WRITE(7,10) N2
   10 FORMAT(I5)
      DO 7 I=1,N2
      WI=(6.2831853072*FLOAT(I))/FLOAT(N)
      WRITE(6,8) I,WI,GAMMA(I)
    8 FORMAT(1H ,I5,2X,E16.8,2X,E16.8)
      WRITE(7,11) WI, GAMMA(I)
   11 FORMAT(2E16.8)
    7 CONTINUE
      END
```

```
C
C.....PROGRAM TO PERFORM THE COMPUTOGRAM VALIDATION TEST.
C
      DIMENSION ALPHA(2),Y(1500),R(2,100),DYBAR(25),W(100),R(25,100
     A0),SIG(1499),W1(100),Y(1500),DY(25,1499)
      READ(5,4) M1
    4 FORMAT(I5)
      READ(5,1)(ALPHA(I),I=1,M1)
    1 FORMAT(8F10.5)
      READ(5,2) N
    2 FORMAT(I5)
      READ(5,3)(Y(I),I=1,N)
    3 FORMAT(8F10.5)
      M=M1+1
      M2=M1+1
      N1=N-1
      MAVG=25
      DO 10 IM=1,MAVG
      DO 11 I=1,N
      DY(IM,I)=Y(I)
   11 CONTINUE
      IX=1175321+2*IM
      AM=0.0
      VARS=.011+33658
      S=SQRT(VARS)
      DO 12 K=M2,N
      CALL GAUSS(IX,S,AM,V)
      W(K)=V
      SUM=0.0
      DO 13 I=1,M1
      SUM=SUM+ALPHA(I)*DY(IM,K-I)
   13 CONTINUE
      DY(IM,K)=SUM+W(K)
   12 CONTINUE
      SUM1=0.0
      DO 14 I=1,N
      SUM1=SUM1+DY(IM,I)
   14 CONTINUE
      DYBAR(IM)=(1.0/FLOAT(N))*SUM1
      R1=0.0
      DO 51 I=1,N
      R1=R1+(DY(IM,I)-DYBAR(IM))*(DY(IM,I)-DYBAR(IM))
   51 CONTINUE
      R0=R1/FLOAT(N)
      DO 15 K=1,N1
      SUM2=0.0
      KN=N-K
      DO 16 J=1,KN
      SUM2=SUM2+(DY(IM,J)-DYBAR(IM))*(DY(IM,J+K)-DYBAR(IM))
   16 CONTINUE
      R(IM,K)=(1.0/FLOAT(N))*SUM2/R0
   15 CONTINUE
   10 CONTINUE
      SUM3=0.0
      DO 17 I=1,N
      SUM3=SUM3+Y(I)
   17 CONTINUE
      YBAR=(1.0/FLOAT(N))*SUM3
      RY1=0.0
      DO 52 I=1,N
      RY1=RY1+(Y(I)-YBAR)*(Y(I)-YBAR)
   52 CONTINUE
      RY0=RY1/FLOAT(N)
      DO 18 K=1,N1
      SUM4=0.0
      KN=N-K
      DO 19 J=1,KN
      SUM4=SUM4+(Y(J)-YBAR)*(Y(J+K)-YBAR)
   19 CONTINUE
      RY(K)=(1.0/FLOAT(N))*SUM4/RY0
   18 CONTINUE
      DO 20 K=1,N1
      SUM5=0.0
      DO 21 J=1,MAVG
      SUM5=SUM5+R(J,K)
   21 CONTINUE
      RM(K)=(1.0/FLOAT(MAVG))*SUM5
   20 CONTINUE
```

```
      DO 22 K=1,N1
      SUM6=0.0
      DO 23 J=1,MAVG
      SUM6=SUM6+(R(J,K)-RM(K))**2
   23 CONTINUE
      SIG(K)=SQRT((1.0/FLOAT(MAVG))*SUM6)
   22 CONTINUE
      IOUT=0
      WRITE(6,30)
   30 FORMAT(1H1,12HCORRELLOGRAM)
      WRITE(7,90) 1
   90 FORMAT(I5)
      DO 24 K=1,N1
      RU=RM(K)+(2.0*SIG(K))
      RL=RM(K)-(2.0*SIG(K))
      IF(RY(K).GT.RU) GO TO 25
      IF(RY(K).LT.RL) GO TO 27
      WRITE(6,40) K,RL,RY(K),RU
   40 FORMAT(I5,3(5X,F16.6))
      WRITE(7,91) K,RL,RY(K),RU
   91 FORMAT(I5,3F15.6)
      GO TO 24
   25 WRITE(6,26) K
   26 FORMAT(1H ,27HOUTSIDE UPPER BOUND AT K = ,I5)
      GO TO 29
   27 WRITE(6,28) K
   28 FORMAT(1H ,27HOUTSIDE LOWER BOUND AT K = ,I5)
   29 IOUT=IOUT+1
   24 CONTINUE
      IF(IOUT.EQ.0) GO TO 31
      GO TO 33
   31 WRITE(6,32)
   32 FORMAT(1H0,23HPASSES CORRELOGRAM TEST)
   33 CONTINUE
      DO 42 K=1,N
      SUM7=0.0
      DO 43 I=1,MAVG
      SUM7=SUM7+D(I,K)
   43 CONTINUE
      YAVG(K)=(1.0/FLOAT(MAVG))*SUM7
   42 CONTINUE
      WRITE(6,44) N
   44 FORMAT(I5)
      WRITE(6,41)(YAVG(K),K=1,N)
   41 FORMAT(6F10.5)
      END
```

```
C
C
C.....PROGRAM TO COMPUTE AVERAGE (MUTUAL) INFORMATION.
C
C
      DIMENSION ALPHA(20),A(40),PHI(20,20),GAMMA(20,20),GP(20,20),HT(20,
     A20),H(20,20),GPH(20,20),G(20,20),DIT(20,20),XAG(20,20),TI(20,20),
     AGT(20,20),VGT(20,20),T2(20,20),GPT(20,20),VGO(20,20),T3(20,20),T
     A(20,20),XXA(20,20),TT(20,20),VXA(20,20),VH(20,20),GA(20,20),GAIN
     A0(20),P1(20,20),P2(20,20),P3(20,20),PX(20,20),VDO(20,20),PHIT(20
     A(1400),VPO(20,20),PPHIT(20,20),XPHIT(20,20),XP(20,20),VX(20,20),\
     AT(20,20),HA(144),XX(144),HVH(20,20),INFO(1400),AVGINF(5,200)
      REAL INFORM,MUTUAL
      DO 70 I=1,20
      DO 71 J=1,20
      PHI(I,J)=0.0
      GAMMA(I,J)=0.0
      GP(I,J)=0.0
      HT(I,J)=0.0
      H(I,J)=0.0
      GPH(I,J)=0.0
      G(I,J)=0.0
      DIT(I,J)=0.0
      XAG(I,J)=0.0
      TI(I,J)=0.0
      GT(I,J)=0.0
      VGT(I,J)=0.0
      T2(I,J)=0.0
      GPT(I,J)=0.0
      VGO(I,J)=0.0
      T3(I,J)=0.0
      TT(I,J)=0.0
      VXA(I,J)=0.0
      VH(I,J)=0.0
      GA(I,J)=0.0
      GAIN(I,J)=0.0
      U(I,J)=0.0
      VXAP(I,J)=0.0
      P1(I,J)=0.0
      P2(I,J)=0.0
      P3(I,J)=0.0
      PX(I,J)=0.0
      VDO(I,J)=0.0
      PHIT(I,J)=0.0
      VXPHIT(I,J)=0.0
      PVP(I,J)=0.0
      VX(I,J)=0.0
      VT(I,J)=0.0
      HVH(I,J)=0.0
   71 CONTINUE
   70 CONTINUE
      DO 72 I=1,144
      AA(I)=0.0
      BA(I)=0.0
   72 CONTINUE
      READ(1,1) N
    1 FORMAT(I5)
C
C.....SYSTEM PARAMETERS FOR THE MODEL FOLLOW
C
      READ(2,2) M1
    2 FORMAT(I5)
      READ(2,3)(ALPHA(I),I=1,M1)
    3 FORMAT(5F10.5)
      A(1)=ALPHA(1)
      A(2)=1.0*ALPHA(2)
      DO 4 I=3,41
      A(I)=ALPHA(I)-ALPHA(I-2)
    4 CONTINUE
      A(M1+1)=-ALPHA(M1-1)
      A(M1+2)=-ALPHA(M1)
C
C.....M2 IS THE ORDER OF THE STATE EQUATIONS
C
      M2=M1+2
      M=M2-1
      IF(M.EQ.0) GO TO 60
      DO 5 I=1,M
```

```
      DO 6 J=1,M2
      PHI(I,J)=0.0
    6 CONTINUE
      PHI(I,I+1)=1.0
    5 CONTINUE
      M3=M2+1
      DO 7 J=1,M2
      PHI(M2,J)=A(M3-J)
    7 CONTINUE
      GO TO 21
   60 PHI(1,1)=ALPHA(1)
   61 CONTINUE
      DO 8 I=1,M2
      GAMMA(I,1)=0.0
    8 CONTINUE
      GAMMA(M2,1)=1.0
      HT(M2,1)=1.0
      IF(1.EQ.0) GO TO 62
      DO 9 I=1,4
      HT(I,1)=0.0
    9 CONTINUE
   62 CONTINUE
      VV=.0012*1.0640
      DO 10 I=1,12
      DO 11 J=1,M2
      VXO(I,J)=0.0
   11 CONTINUE
      VXO(I,I)=1.0
   10 CONTINUE
      VVO=.000001
      DO 200 NUT=1,5
      VVO=VVO*10.0
      WRITE(10,202) NUT
  202 FORMAT(I5)
      DO 12 I=1,N
      VV(I)=VVO
   12 CONTINUE
      VRO=VVO
      VR=VV(1)
      VVW=0.0
C
C.....END OF SYSTEM PARAMETERS
C
C
C.....INITIALIZATION OF FILTER PARAMETERS FOR COMPUTATION OF ENTROPY
C
      DO 13 I=1,M2
      GP(I,1)=GAMMA(I,1)*VVW**(1.0/VVO)
   13 CONTINUE
      CALL MATTRP(HT,20,2,M)
      CALL MATMUL(GP,M,20,20,20,PH)
      CALL MATSUM(PH,GPH,20,20,DIF)
      CALL MATTRP(DIF,20,20,TT)
      CALL MATMUL(VXO,DIF,20,20,20,DD)
      CALL MATMUL(DIF,VXO,20,20,20,T1)
      CALL MATTRP(GAMMA,20,2,GT)
      DO 40 J=1,M2
      VWGT(1,J)=VW*GT(1,J)
   40 CONTINUE
      CALL MATMUL(GAMMA,VWGT,20,20,20,T2)
      CALL MATTRP(GP,20,20,GPT)
      DO 41 I=1,M2
      DO 42 J=1,M2
      VGO(I,J)=VVO*GPT(I,J)
   42 CONTINUE
   41 CONTINUE
      CALL MATMUL(GP,VGO,20,20,20,T3)
      CALL MATADD(T1,T2,20,20,TT)
      CALL MATSUM(TT,T3,20,20,VXP)
      JJ=1
      DO 17 I=1,M2
      DO 18 J=JJ,M2
      VXAP(I,J)=VXP(I,J)
      VXAP(J,I)=VXAP(I,J)
   18 CONTINUE
      VXAP(I,I)=VXAP(I,I)
      JJ=JJ+1
   17 CONTINUE
```

ORIGINAL PAGE IS
OF POOR QUALITY

```
      CALL MATMUL(V,M2,MT,20,20,20,V1)
      CALL MATMUL(H,V1,M,20,20,1,HV1)
      CORR=1.0/(HV1(1,1)+V(1,1))
      CALL MATMUL(V1,M2,MT,20,20,20,G1)
      DO 14 I=1,M2
      GA1(I,1)=GA(I,1)*CORR
   14 CONTINUE
      DO 15 I=1,M2
      DO 16 J=1,M2
      U(I,J)=0.0
   16 CONTINUE
      U(I,I)=1.0
   15 CONTINUE
      CALL MATMUL(GA1N,M,20,20,20,P1)
      CALL MATSUB(U,P1,20,20,M2)
      CALL MATMUL(P2,VXAP,20,20,20,P3)
      JJ=1
      DO 19 I=1,M2
      DO 20 J=JJ,M2
      PX(I,J)=P3(I,J)
      PX(J,I)=P3(I,J)
   20 CONTINUE
      PX(I,I)=P3(I,I)
      JJ=JJ+1
   19 CONTINUE
      CALL MATTRP(PHI,2),21,PHIT)
      CALL MATMUL(VX0),PHI,20,20,20,VXPHIT)
      CALL MATMUL(PHI,VXPHIT,20,20,20,PVP)
      CALL MATADD(PVP,TC,20,20,VT)
      JJ=1
      DO 31 I=1,M2
      DO 32 J=JJ,M2
      VX(I,J)=VT(I,J)
      VX(J,I)=VT(I,J)
   32 CONTINUE
      VX(I,I)=VT(I,I)
      JJ=JJ+1
   31 CONTINUE
C
C.....END OF INITIALIZATION
C
C
C.....COMPUTATION OF VARIANCES FOLLOWS
C
      MUTINF=0.0
      DO 100 K=2,M
      VR=VV(K)
      DO 22 I=1,M2
      GP(I,1)=GAMMA(I,1)*VV(K)/(1.0/VV(K))
   22 CONTINUE
      CALL MATMUL(GP,M,20,20,20,GPH)
      CALL MATSUB(PHI,GPH,20,20,Q1)
      CALL MATTRP(Q1,20,20,Q1T)
      CALL MATMUL(PX,Q1T,20,20,20,VQ0)
      CALL MATMUL(Q1,VQ0,20,20,20,T1)
      CALL MATTRP(G,20,20,GPT)
      DO 23 I=1,M2
      DO 43 J=1,M2
      VGO(I,J)=VV(K)*GPT(I,J)
   43 CONTINUE
   23 CONTINUE
      CALL MATMUL(GP,VGO,20,20,20,T3)
      CALL MATADD(T1,T2,20,20,TT)
      CALL MATSUB(TT,T3,20,20,VXA)
      JJ=1
      DO 24 I=1,M2
      DO 25 J=JJ,M2
      VXAP(I,J)=VXA(I,J)
      VXAP(J,I)=VXA(I,J)
   25 CONTINUE
      VXAP(I,I)=VXA(I,I)
      JJ=JJ+1
   24 CONTINUE
      CALL MATMUL(VXAP,MT,20,20,20,VH)
      CALL MATMUL(H,VH,2,20,20,HVH)
      CORR=1.0/(HVH(1,1)+VV(K))
      CALL MATMUL(VXAP,MT,20,20,20,CA)
      DO 26 I=1,M2
```

C-3

```
      GAIN(I,I)=GA(I,I)*SQRT
   26 CONTINUE
      CALL MATMUL(GAIN,M,20,20,P1)
      CALL MATSUB(U,P1,20,20,P2)
      CALL MATMUL(P2,VXAP,20,20,20,P3)
      JJ=1
      DO 27 I=1,M2
      DO 28 J=JJ,M2
      PX(I,J)=P3(I,J)
      PX(J,I)=P3(I,J)
   28 CONTINUE
      PX(I,I)=P3(I,I)
      JJ=JJ+1
   27 CONTINUE
      IJ=0
      II=1
      DO 29 J=1,M2
      DO 30 I=1,II
      IJ=IJ+1
      AA(IJ)=P(I,J)
   30 CONTINUE
      II=II+1
   29 CONTINUE
C
C.....END OF VARIANCE COMPUTATION
C
C
C.....MUTUAL INFORMATION COMPUTATION FOLLOWS
C
C
      MUTINF=MUTINF+(.5*GAIN(M2,I))
      AVGINF(MJT,K)=MUTINF
C
C.....MUTUAL INFORMATION COMPLETE
C
  100 CONTINUE
  200 CONTINUE
      DO 201 K=2,N
      WRITE(6,52) K,AVGINF(1,K),AVGINF(2,K),AVGINF(3,K),AVGINF(4,K),AVGI
     NF(5,K)
   52 FORMAT(1H ,I5,5F12.5)
      WRITE(7,203) K,AVGINF(1,K),AVGINF(2,K),AVGINF(3,K),AVGINF(4,K),AVG
     INF(5,K)
  203 FORMAT(I5,5F12.5)
  201 CONTINUE
      END
```

```
      SUBROUTINE MATMUL(A,M,N,B,N2,N3,C)
      DIMENSION A(20,20),B(20,20),C(20,20)
      DO 1 J=1,N3
      DO 2 I=1,M
      C(I,J)=0.0
      DO 3 K=1,N2
      C(I,J)=C(I,J)+A(I,K)*B(K,J)
    3 CONTINUE
    2 CONTINUE
    1 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MATSUB(A,B,M,N2,C)
      DIMENSION A(20,20),B(20,20),C(20,20)
      DO 1 I=1,M
      DO 2 J=1,N2
      C(I,J)=A(I,J)-B(I,J)
    2 CONTINUE
    1 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MATADD(A,B,M,N2,C)
      DIMENSION A(20,20),B(20,20),C(20,20)
      DO 1 I=1,M
      DO 2 J=1,N2
      C(I,J)=A(I,J)+B(I,J)
    2 CONTINUE
    1 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MATTRP(A,M,N,B)
      DIMENSION A(20,20),B(20,20)
      DO 1 I=1,M
      DO 2 J=1,N
      B(J,I)=A(I,J)
    2 CONTINUE
    1 CONTINUE
      RETURN
      END
```

C-3