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
Made available under NASA sponsorship
In the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof.
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 bandwiths, 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.

**Key Words (Suggested by Author(s))**
- Multispectral Scanners
- Remote Sensing
- Information Theory
- Wavelength Band Selection
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

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>viii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1. General Discussion</td>
<td>1</td>
</tr>
<tr>
<td>2. Previous Work</td>
<td>2</td>
</tr>
<tr>
<td>3. The Present Investigation</td>
<td>3</td>
</tr>
<tr>
<td>CHAPTER II. Information Theoretic Approach</td>
<td>8</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>8</td>
</tr>
<tr>
<td>2. Definition of Average Mutual Information</td>
<td>9</td>
</tr>
<tr>
<td>3. Calculation of Average Information</td>
<td>14</td>
</tr>
<tr>
<td>4. The Relation Between Average Information and the Wiener-Hopf Optimum Filter Problem</td>
<td>21</td>
</tr>
<tr>
<td>5. Computation Considerations</td>
<td>29</td>
</tr>
<tr>
<td>6. Further Consideration of the Relation Between Average Information and Optimum Mean-Square Filtering</td>
<td>35</td>
</tr>
<tr>
<td>CHAPTER III. Model Identification, Selection and Validation Techniques</td>
<td>41</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>41</td>
</tr>
<tr>
<td>2. Models Used to Represent Spectral Scenes</td>
<td>42</td>
</tr>
<tr>
<td>3. Estimation Techniques</td>
<td>48</td>
</tr>
<tr>
<td>4. A Model Selection Criterion</td>
<td>65</td>
</tr>
<tr>
<td>5. Validation of Models</td>
<td>70</td>
</tr>
<tr>
<td>6. Summary</td>
<td>81</td>
</tr>
<tr>
<td>CHAPTER IV. Application of Modeling Techniques</td>
<td>83</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>83</td>
</tr>
<tr>
<td>2. The Empirical Data</td>
<td>83</td>
</tr>
<tr>
<td>3. Identified Models for the Empirical Data</td>
<td>89</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-1</td>
<td>85</td>
</tr>
<tr>
<td>IV-2</td>
<td>85</td>
</tr>
<tr>
<td>IV-3</td>
<td>88</td>
</tr>
<tr>
<td>IV-4</td>
<td>88</td>
</tr>
<tr>
<td>IV-5</td>
<td>91</td>
</tr>
<tr>
<td>IV-6</td>
<td>91</td>
</tr>
<tr>
<td>IV-7</td>
<td>92</td>
</tr>
<tr>
<td>IV-8</td>
<td>92</td>
</tr>
<tr>
<td>IV-9</td>
<td>94</td>
</tr>
<tr>
<td>IV-10</td>
<td>94</td>
</tr>
<tr>
<td>IV-11</td>
<td>95</td>
</tr>
<tr>
<td>IV-12</td>
<td>95</td>
</tr>
<tr>
<td>IV-13</td>
<td>97</td>
</tr>
<tr>
<td>IV-14</td>
<td>97</td>
</tr>
<tr>
<td>IV-15</td>
<td>98</td>
</tr>
<tr>
<td>IV-16</td>
<td>98</td>
</tr>
<tr>
<td>IV-17</td>
<td>100</td>
</tr>
<tr>
<td>IV-18</td>
<td>100</td>
</tr>
<tr>
<td>IV-19</td>
<td>101</td>
</tr>
<tr>
<td>IV-20</td>
<td>101</td>
</tr>
<tr>
<td>IV-21</td>
<td>102</td>
</tr>
<tr>
<td>IV-22</td>
<td>102</td>
</tr>
<tr>
<td>IV-23</td>
<td>104</td>
</tr>
<tr>
<td>IV-24</td>
<td>104</td>
</tr>
<tr>
<td>IV-25</td>
<td>105</td>
</tr>
<tr>
<td>IV-26</td>
<td>105</td>
</tr>
<tr>
<td>IV-27</td>
<td>107</td>
</tr>
<tr>
<td>IV-28</td>
<td>107</td>
</tr>
<tr>
<td>IV-29</td>
<td>108</td>
</tr>
<tr>
<td>IV-30</td>
<td>108</td>
</tr>
<tr>
<td>IV-31</td>
<td>109</td>
</tr>
<tr>
<td>IV-32</td>
<td>109</td>
</tr>
<tr>
<td>IV-33</td>
<td>111</td>
</tr>
<tr>
<td>IV-34</td>
<td>111</td>
</tr>
<tr>
<td>IV-35</td>
<td>112</td>
</tr>
<tr>
<td>IV-36</td>
<td>112</td>
</tr>
<tr>
<td>IV-37</td>
<td>114</td>
</tr>
<tr>
<td>IV-38</td>
<td>114</td>
</tr>
<tr>
<td>IV-39</td>
<td>115</td>
</tr>
<tr>
<td>IV-40</td>
<td>115</td>
</tr>
<tr>
<td>IV-41</td>
<td>117</td>
</tr>
<tr>
<td>IV-42</td>
<td>146</td>
</tr>
<tr>
<td>V-1</td>
<td>160</td>
</tr>
<tr>
<td>V-2</td>
<td>160</td>
</tr>
<tr>
<td>V-3</td>
<td>161</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-1</td>
<td>Average Spectral Response--Wheat Scene</td>
<td>87</td>
</tr>
<tr>
<td>IV-2</td>
<td>Average Spectral Response--Combined Scene</td>
<td>87</td>
</tr>
<tr>
<td>IV-3</td>
<td>Cumulative Periodogram, Band 1, Wheat Scene</td>
<td>119</td>
</tr>
<tr>
<td>IV-4</td>
<td>Correlogram, Band 1, Wheat Scene</td>
<td>119</td>
</tr>
<tr>
<td>IV-5</td>
<td>Periodogram, Band 1, Wheat Scene</td>
<td>119</td>
</tr>
<tr>
<td>IV-6</td>
<td>Cumulative Periodogram, Band 1, Combined Scene</td>
<td>119</td>
</tr>
<tr>
<td>IV-7</td>
<td>Correlogram, Band 1, Combined Scene</td>
<td>122</td>
</tr>
<tr>
<td>IV-8</td>
<td>Periodogram, Band 1, Combined Scene</td>
<td>122</td>
</tr>
<tr>
<td>IV-9</td>
<td>Cumulative Periodogram, Band 2, Wheat Scene</td>
<td>122</td>
</tr>
<tr>
<td>IV-10</td>
<td>Correlogram, Band 2, Wheat Scene</td>
<td>122</td>
</tr>
<tr>
<td>IV-11</td>
<td>Periodogram, Band 2, Wheat Scene</td>
<td>124</td>
</tr>
<tr>
<td>IV-12</td>
<td>Cumulative Periodogram, Band 2, Combined Scene</td>
<td>124</td>
</tr>
<tr>
<td>IV-13</td>
<td>Correlogram, Band 2, Combined Scene</td>
<td>124</td>
</tr>
<tr>
<td>IV-14</td>
<td>Periodogram, Band 2, Combined Scene</td>
<td>124</td>
</tr>
<tr>
<td>IV-15</td>
<td>Cumulative Periodogram, Band 3, Wheat Scene</td>
<td>126</td>
</tr>
<tr>
<td>IV-16</td>
<td>Correlogram, Band 3, Wheat Scene</td>
<td>126</td>
</tr>
<tr>
<td>IV-17</td>
<td>Periodogram, Band 3, Wheat Scene</td>
<td>126</td>
</tr>
<tr>
<td>IV-18</td>
<td>Cumulative Periodogram, Band 3, Combined Scene</td>
<td>126</td>
</tr>
<tr>
<td>IV-19</td>
<td>Correlogram, Band 3, Combined Scene</td>
<td>128</td>
</tr>
<tr>
<td>IV-20</td>
<td>Periodogram, Band 3, Combined Scene</td>
<td>128</td>
</tr>
<tr>
<td>IV-21</td>
<td>Cumulative Periodogram, Band 4, Wheat Scene</td>
<td>128</td>
</tr>
<tr>
<td>IV-22</td>
<td>Correlogram, Band 4, Wheat Scene</td>
<td>128</td>
</tr>
<tr>
<td>IV-23</td>
<td>Periodogram, Band 4, Wheat Scene</td>
<td>130</td>
</tr>
<tr>
<td>IV-24</td>
<td>Cumulative Periodogram, Band 4, Combined Scene</td>
<td>130</td>
</tr>
<tr>
<td>IV-25</td>
<td>Correlogram, Band 4, Combined Scene</td>
<td>130</td>
</tr>
<tr>
<td>IV-26</td>
<td>Periodogram, Band 4, Combined Scene</td>
<td>130</td>
</tr>
<tr>
<td>IV-27</td>
<td>Cumulative Periodogram, Band 5, Wheat Scene</td>
<td>132</td>
</tr>
<tr>
<td>IV-28</td>
<td>Correlogram, Band 5, Wheat Scene</td>
<td>132</td>
</tr>
<tr>
<td>IV-29</td>
<td>Periodogram, Band 5, Wheat Scene</td>
<td>132</td>
</tr>
<tr>
<td>IV-30</td>
<td>Cumulative Periodogram, Band 5, Combined Scene</td>
<td>132</td>
</tr>
<tr>
<td>IV-31</td>
<td>Correlogram, Band 5, Combined Scene</td>
<td>134</td>
</tr>
<tr>
<td>IV-32</td>
<td>Periodogram, Band 5, Combined Scene</td>
<td>134</td>
</tr>
<tr>
<td>IV-33</td>
<td>Cumulative Periodogram, Band 6, Wheat Scene</td>
<td>134</td>
</tr>
<tr>
<td>IV-34</td>
<td>Correlogram, Band 6, Wheat Scene</td>
<td>134</td>
</tr>
<tr>
<td>IV-35</td>
<td>Periodogram, Band 6, Wheat Scene</td>
<td>136</td>
</tr>
<tr>
<td>IV-36</td>
<td>Cumulative Periodogram, Band 6, Combined Scene</td>
<td>136</td>
</tr>
<tr>
<td>IV-37</td>
<td>Correlogram, Band 6, Combined Scene</td>
<td>136</td>
</tr>
<tr>
<td>IV-38</td>
<td>Periodogram, Band 6, Combined Scene</td>
<td>136</td>
</tr>
<tr>
<td>IV-39</td>
<td>Cumulative Periodogram, Band 7, Wheat Scene</td>
<td>137</td>
</tr>
<tr>
<td>IV-40</td>
<td>Correlogram, Band 7, Wheat Scene</td>
<td>137</td>
</tr>
<tr>
<td>IV-41</td>
<td>Periodogram, Band 7, Wheat Scene</td>
<td>137</td>
</tr>
<tr>
<td>IV-42</td>
<td>Cumulative Periodogram, Band 7, Combined Scene</td>
<td>137</td>
</tr>
<tr>
<td>IV-43</td>
<td>Correlogram, Band 7, Combined Scene</td>
<td>139</td>
</tr>
<tr>
<td>IV-44</td>
<td>Periodogram, Band 7, Combined Scene</td>
<td>139</td>
</tr>
<tr>
<td>IV-45</td>
<td>Cumulative Periodogram, Band 8, Wheat Scene</td>
<td>139</td>
</tr>
<tr>
<td>IV-46</td>
<td>Correlogram, Band 8, Wheat Scene</td>
<td>139</td>
</tr>
<tr>
<td>IV-47</td>
<td>Periodogram, Band 8, Wheat Scene</td>
<td>141</td>
</tr>
<tr>
<td>IV-48</td>
<td>Cumulative Periodogram, Band 8, Combined Scene</td>
<td>141</td>
</tr>
</tbody>
</table>
IV-49 Correlogram, Band 8, Combined Scene....................................... 141
IV-50 Periodogram, Band 8, Combined Scene......................................... 141
IV-51 Cumulative Periodogram, Band 9, Wheat Scene.............................. 143
IV-52 Correlogram, Band 9, Wheat Scene.............................................. 143
IV-53 Periodogram, Band 9, Wheat Scene................................................ 143
IV-54 Cumulative Periodogram, Band 9, Combined Scene.......................... 143
IV-55 Correlogram, Band 9, Combined Scene........................................... 145
IV-56 Periodogram, Band 9, Combined Scene.......................................... 145
V-1 Average Information, Band 1, Wheat Scene........................................ 150
V-2 Average Information, Band 1, Combined Scene................................... 150
V-3 Average Information, Band 2, Wheat Scene........................................ 151
V-4 Average Information, Band 2, Combined Scene.................................. 151
V-5 Average Information, Band 3, Wheat Scene........................................ 152
V-6 Average Information, Band 3, Combined Scene.................................. 152
V-7 Average Information, Band 4, Wheat Scene........................................ 153
V-8 Average Information, Band 4, Combined Scene.................................. 153
V-9 Average Information, Band 5, Wheat Scene........................................ 154
V-10 Average Information, Band 5, Combined Scene................................ 154
V-11 Average Information, Band 6, Wheat Scene........................................ 155
V-12 Average Information, Band 6, Combined Scene................................ 155
V-13 Average Information, Band 7, Wheat Scene........................................ 156
V-14 Average Information, Band 7, Combined Scene................................ 156
V-15 Average Information, Band 8, Wheat Scene........................................ 157
V-16 Average Information, Band 8, Combined Scene................................ 157
V-17 Average Information, Band 9, Wheat Scene........................................ 158
V-18 Average Information, Band 9, Combined Scene................................ 158
<table>
<thead>
<tr>
<th></th>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>AR(n)</td>
<td>autoregressive process of order n</td>
</tr>
<tr>
<td>2.</td>
<td>ARC(n)</td>
<td>autoregressive process of order n plus constant trend</td>
</tr>
<tr>
<td>3.</td>
<td>IAR(n)</td>
<td>integrated autoregressive process of order n</td>
</tr>
<tr>
<td>4.</td>
<td>IAR₂(n)</td>
<td>integrated autoregressive process of the second kind of order n</td>
</tr>
</tbody>
</table>
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 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 analytical 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 research 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 reflectiv infrared wavelengths (approximately .4 to 3.0 micrometers, \( \mu m \)) is measured as a function of wavelength, \( \lambda \), 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 ob-
servation platform, or other sources. The point is that the multispec-
tral scanner system receives electromagnetic energy from the scene that
exhibits random variations corrupted by disturbances that also have ran-
dom 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 in-
terest. 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 obser-
vation conveys about the scene.

This is highly reminiscent of the classical problem in communica-
tion 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 quantitative
measure of what the received signal conveys about the transmitted sig-
nal. 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 [51]. 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 informa-
tion in the received signal about the transmitted signal may be said to
be the reduction in uncertainty about the transmitted signal that is ob-
tained 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, utilization 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 discussions 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 Chapter 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 \( s(\lambda) \in S \) and \( \lambda \in [\lambda_1, \lambda_2] \). The ensemble of sample functions for the spectral response is \( S \) and \( [\lambda_1, \lambda_2] \) 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

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

(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
\[ y(\lambda) = s(\lambda) + n(\lambda), \quad \lambda \in [\lambda_1, \lambda_2] \] (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 variables about another set of random variables.

The average mutual information in a set of random variables \( \{v_1, v_2, \ldots, v_N\} \) about the set of random variables \( \{u_1, u_2, \ldots, u_M\} \) is defined by [G1]

\[
I(\underline{U}, \underline{V}) = \int_{\underline{U}} \int_{\underline{V}} P_{UV}(u,v) \log \left( \frac{P_{UV}(u,v)}{P_U(u)P_V(v)} \right) \, du \, dv \quad (2-3)
\]

where

\[ P_{U,V}(u,v) = \text{joint density function of the sets } U \text{ and } V \]
\[ P_U(u) = \text{density function of the set } U \]

\[ P_V(v) = \text{density function of the set } V \]

and \( \int_U \) and \( \int_V \) represent M-fold and N-fold integrals over all the possible values of the members of the sets \( U \) and \( V \).

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 \( s(\lambda) \).
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

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

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 \([\lambda_1, \lambda_2]\). Then \( s(\lambda) \) may be represented on the interval \([\lambda_1, \lambda_2]\) as

\[
s(\lambda) = \lim_{n \to \infty} \sum_{i=1}^{n} S_i \phi_i(\lambda), \quad \lambda_1 \leq \lambda \leq \lambda_2 \tag{2-6}
\]

where \( \text{l.i.m.} \) is the limit in the mean defined by

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

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
\( \theta_1(\lambda), \theta_2(\lambda), \ldots \) on \([\lambda_1, \lambda_2]\) the received spectral process \( y(\lambda) \) can
be represented as

\[
y(\lambda) = \lim_{n \to \infty} \sum_{i=1}^{n} y_{i} \theta(\lambda), \quad \lambda_1 \leq \lambda \leq \lambda_2
\]  

(2-9)

where

\[
y_{i} = \int_{\lambda_1}^{\lambda_2} y(\lambda) \theta(\lambda) d\lambda.
\]  

(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 com-
plete 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 expres-
sion for average information for the processes \( s(\lambda) \) and \( y(\lambda) \).

Hence, if we define

\[
S_n = \{s_1, s_2, \ldots, s_n\}
\]  

(2-11)

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

$$I(S, Y) = \int \int p(S_n, Y_n) \log \left( \frac{p(S_n, Y_n)}{p(S_n)p(Y_n)} \right) dS_n dY_n$$

(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

$$I(S, Y) = \lim_{n \to \infty} I(S, Y)$$

(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 $y(\lambda)$ 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 $\{\epsilon_1(\lambda), \epsilon_2(\lambda), \ldots\}$ are defined. These basis functions will, at least in principle, yield a calculation technique for average information. First consider the representation for $s(\lambda)$,
The covariance function of \( s(\lambda) \) is defined as

\[
K_s(\lambda, u) = E[s(\lambda)s(u)] \quad \lambda, u \in [\lambda_1, \lambda_2].
\]

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

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

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

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

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

\[
\delta_{ij} = \begin{cases} 1, & \text{i=1} \\ 0, & \text{i=j} \\ 0, & \text{i \neq j} \end{cases}
\]
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

$$a_j \phi_j(\lambda) = \int_{\lambda_1}^{\lambda_2} K_s(\lambda, u) \phi_j(u) du, \quad \lambda_1 \leq \lambda \leq \lambda_2.$$  \hspace{1cm} (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

$$\int_{\lambda_1}^{\lambda_2} \phi_i(\lambda) \phi_j(\lambda) d\lambda = \delta_{ij}.$$  \hspace{1cm} (2-21)

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

$$\mathbb{E} \left[ \int_{\lambda_1}^{\lambda_2} s^2(\lambda) d\lambda \right] = \sum_{j=1}^{\infty} a_j.$$  \hspace{1cm} (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

$$\{\phi_i(\lambda), i=1, 2, \ldots\}$$  for representation of the received spectral process $y(\lambda)$. Since it is assumed that $y(\lambda)$ is of the form
\begin{equation}
y(\lambda) = s(\lambda) + n(\lambda), \quad \lambda_1 \leq \lambda \leq \lambda_2, \tag{2-23}
\end{equation}

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

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

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

\begin{equation}
K_n(\lambda;u) = \frac{N_0}{2} \delta(\lambda-u) \tag{2-25}
\end{equation}

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

\begin{equation}
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}
\end{equation}

or

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

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 \( \theta_j(\lambda) = \delta_j(\lambda) \) 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=1,2,\ldots\} will not form a complete orthonormal set \[\text{V1, p. 181}\]. 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

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

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

\[
= \frac{N_0}{2} \phi_j(\lambda) + \int_{\lambda_1}^{\lambda_2} K_S(\lambda; u) \phi_j(u) \, du, \quad \lambda_1 \leq \lambda \leq \lambda_2 \quad (2-28)
\]

If we use

\[
b_j = a_j + \frac{N_0}{2} \quad (2-29)
\]

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

\[
y(\lambda) = \sum_{j=1}^{\infty} y_i \phi_i(\lambda), \quad \lambda_1 \leq \lambda \leq \lambda_2
\]  

(1.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

\[
E[y_i y_j] = (a_j + \frac{N_0}{2}) \delta_{ij} \tag{2-31}
\]

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 \( i \)th diagonal element given by \( E[s_i^2] = a_i \). Similarly, the covariance matrix for the random variables \( \{ y_i; i=1,2,\ldots \} \) is diagonal with the \( i \)th 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 variables for the noise that have a diagonal covariance matrix with the \( i \)th 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]

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

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

\[
\det C_n = \prod_{i=1}^{\infty} \frac{N_0}{2} \tag{2-34}
\]

Thus, we can write

\[
\frac{\det C_y}{\det C_n} = \prod_{j=1}^{\infty} \left( 1 + \frac{2}{N_0} a_j \right) \tag{2-35}
\]

Hence, the average information can be written as:
\[ I(S, Y) = \frac{1}{2} \sum_{j=1}^{n} \log \left[ 1 + \frac{2a_j}{N_0} \right] \]  \hspace{1cm} (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

\[ y(\lambda) = s(\lambda) + n(\lambda) \hspace{1cm}, \hspace{1cm} \lambda_1 \leq \lambda \leq \lambda_2 \]  \hspace{1cm} (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

\[ \hat{s}(\lambda) = \int_{\lambda_1}^{\lambda_2} h(\lambda, u) y(u) \, du, \quad \lambda_1 \leq \lambda \leq \lambda_2 \]  

(2-38)

where $h(\lambda, u)$ is the impulse response of the optimal filter such that

\[ E[(s(\lambda) - \hat{s}(\lambda))^2] \]

is minimized. It is straightforward to show [VI, p. 198-204] that under our assumptions concerning the spectral processes $y(\lambda)$ and $s(\lambda)$ the optimum filter must satisfy

\[ \int_{\lambda_1}^{\lambda_2} K_s(u, \nu) h(\lambda, \nu) \, d\nu + \frac{N_0}{2} h(\lambda, u) = K_s(\lambda, u), \quad \lambda_1 \leq \lambda \leq \lambda_2 \]

(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 $\{\phi_i(\lambda), i=1,2,\ldots\}$ form a complete orthonormal set, it is possible to write $h(\lambda, u)$ in a series expansion of the form

\[ h(\lambda, u) = \sum_{i=1}^{\infty} h_i \phi_i(\lambda) \phi_i(u), \quad \lambda_1 \leq \lambda \leq \lambda_2 \]

(2-40)

\[ \lambda_1 \leq u \leq \lambda_2 \]

From Mercer's theorem [VI, p. 181] we can write
\[ K_s(\lambda, u) = \sum_{i=1}^{\infty} a_i \phi_i(\lambda) \phi_i(u), \quad \lambda_1 \leq \lambda, u \leq \lambda_2. \quad (2-41) \]

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

\[
\frac{N_0}{2} \sum_{i=1}^{\infty} h_i \phi_i(\lambda) \phi_i(u) + \lambda_1 \sum_{i=1}^{\infty} a_i \phi_i(\lambda) \phi_i(u) \sum_{j=1}^{\infty} h_j \phi_j(\lambda) \phi_j(v) dv \\
= \sum_{i=1}^{\infty} a_i \phi_i(\lambda) \phi_i(u) \quad . \quad (2-42) 
\]

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

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

It is seen that if

\[
h_i = \frac{a_i}{a_i + \frac{N_0}{2}} \quad i=1, 2, \ldots \quad (2-44) 
\]

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

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

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

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

\[ \lambda_1 + \lambda \int_{\lambda_1}^{\lambda_2} \phi_j (\lambda) d\lambda = 1 \]  (2-21)

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

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

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

\[ = \frac{1}{2} \sum_{j=1}^{\infty} \left( 1 + \frac{2}{N_0} a_j \right)^{-1} \frac{2}{N_0} \frac{da_j}{d\lambda} \]

or
\[
\frac{d\mathbb{S}(S,Y)}{dz} = \frac{1}{2} \sum_{j=1}^{\infty} \left[ \frac{\frac{2}{N_0} \frac{d\zeta_j}{dz}}{1 + \frac{2}{N_0} \zeta_j} \right] .
\] (2-46)

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

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

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

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

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

\[
\frac{d\zeta_j}{dz} = \int_{\lambda_1}^{\lambda_1 + 2\lambda} \left[ K_s(\lambda, \lambda_1 + \lambda) \phi_j(\lambda_1 + \lambda) +
\right.
\]
\[
+ \int_{\lambda_1}^{\lambda_1 + 2\lambda} K_s(\lambda, u) \frac{\partial \phi_j(u)}{\partial z} du \left] \phi_j(\lambda) d\lambda
\]
\[
+ \int_{\lambda_1}^{\lambda_1 + 2\lambda} \left[ K_s(\lambda_1 + 2\lambda, u) \phi_j(\lambda_1 + \lambda) +
\right.
\]
\[
\int_{\lambda_1}^{\lambda_1 + 2\lambda} K_s(\lambda_1 + 2\lambda, u) \phi_j(\lambda_1 + \lambda) +
\]
\[
\int_{\lambda_1}^{\lambda_1 + 2\lambda} \left[ K_s(\lambda_1 + 2\lambda, u) \phi_j(\lambda_1 + \lambda) +
\right.
\]
\[
\phi_j(u) = \int_{\lambda_1}^{\lambda_1+\lambda} K_s(\lambda, u) \frac{\partial \phi_j(\lambda)}{\partial \lambda} d\lambda \phi_j(u) du \quad (2-49)
\]

This can be simplified to give

\[
\frac{da_j}{d\lambda} = 2\phi_j(\lambda_1 + \lambda) \int_{\lambda_1}^{\lambda_1+\lambda} K_s(\lambda, u) \phi_j(u) du 
\]

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

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

\[
a_j \phi_j(\lambda_1 + \lambda) = \int_{\lambda_1}^{\lambda_1+\lambda} K_s(\lambda_1, u) \phi_j(u) du \quad (2-51)
\]

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

\[
\frac{da_j}{d\lambda} = 2a_j \phi_j^2(\lambda_1 + \lambda) 
\]

\[
+ 2a_j \int_{\lambda_1}^{\lambda_1+\lambda} \phi_j(\lambda) \frac{\partial \phi_j(\lambda)}{\partial \lambda} d\lambda 
\quad (2-52)
\]

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 \( \lambda \).
\[
\frac{d}{d\lambda} \int_{\lambda_1}^{\lambda_1+\xi} \phi_j^{2}(\lambda) d\lambda = 0 
\]  
(2-53)

or

\[
0 = \phi_j^{2}(\lambda_1+\xi) + 2\int_{\lambda_1}^{\lambda_1+\xi} \phi_j(\lambda) \frac{\partial \phi_j(\lambda)}{\partial \xi} d\lambda 
\]  
(2-54)

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

\[
\frac{d a_j}{d \xi} = 2a_j \phi_j^{2}(\lambda_1+\xi) - a_j \phi_j^{2}(\lambda_1+\xi) 
\]  
(2-55)

or

\[
\frac{d a_j}{d \xi} = a_j \phi_j^{2}(\lambda_1+\xi) 
\]  
(2-56)

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

\[
\frac{dI(S_x Y)}{d \xi} = \sum_{j=1}^{N_0} \left[ \frac{2a_j \phi_j^{2}(\lambda_1+\xi)}{(1 + \frac{2}{N_0} a_j)} \right] . 
\]  
(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

\[
\frac{dI(S_x Y)}{d \xi} = \frac{1}{2} \cdot h(\lambda_1+\xi, \lambda_1+\xi) 
\]  
(2-58)

Thus it is seen that since \( \xi \) varies from 0 to \( \lambda_2 - \lambda_1 \) we may make a sim-
plifying change of variables and write

\[ I(S, Y) = \frac{1}{2} \int_{\lambda_1}^{\lambda_2} h(\lambda, \lambda) d\lambda \]  

(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 \( \lambda \) in order to obtain the optimum mean-square estimate of \( s(\lambda) \) at wavelength \( \lambda \). It is interesting that there is a relationship between the mean-square estimation of \( s(\chi) \) from the observation \( y(\chi) \) 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 \( K_s(\lambda, \lambda) \) 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 \( K_s(\lambda, \lambda) \). The second major problem lies in actually solving the Wiener-Hopf equation even under the assumption that a functional form for \( K_s(\lambda, \lambda) \) is known. If the spectral processes are stationary and possess 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 tech-
nique 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 technique
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

\[ y(\lambda) = s(\lambda) + n(\lambda) \quad \lambda_1 \leq \lambda \leq \lambda_2 \]

is written as

\[ y(k) = s(k) + n(k) \quad k \in [\lambda_1, \lambda_2] \quad (2-60) \]

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 \( \frac{N_0}{\mathcal{Z}} \). 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:

$$s(\lambda) = A \ s(\lambda) + Bw(\lambda)$$  \hspace{1cm} (2-61)

where

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

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

$A$ is an $(n \times n)$ matrix

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

$B$ is an $n \times 1$ matrix
\( w(\lambda) = \text{an independent gaussian random process} \)

\( n \) 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:

\[
s(k+1) = \Phi s(k) + \Gamma w(k) \quad k \in [\lambda_1, \lambda_2]
\]

where

\[
\begin{bmatrix}
S_1(k+1) \\
S_2(k+1) \\
\vdots \\
S_n(k+1)
\end{bmatrix}
= \begin{bmatrix}
\phi \\
\vdots \\
\end{bmatrix}
\]

\( \Phi \) is an \((n \times n)\) matrix.

\( \Gamma \) is an \((n \times 1)\) vector.

\( w(k) = \text{a discrete independent gaussian random process with zero mean and variance \( V_w(k) \delta(k-j) \)} \)

\( k \) is an integer that corresponds to a discrete wavelength \( \lambda \in [\lambda_1, \lambda_2] \). With this representation for \( s(k) \) we can write the discrete form for \( y(\lambda) \) as
\( y(k) = H^T s(k) + n(k) \quad k \in [\lambda_1, \lambda_2] \) \hspace{1cm} (2-63)

where
\[
H^T s(k) = s(k) \text{ is the relationship between the state variable formulation and the spectral response } s(\lambda) \text{ for the } \lambda \text{ that corresponds to } k.
\]

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 \( \hat{s}(\lambda) \) of \( s(\lambda) \). The estimate thus may be written
\[
\hat{s}(\lambda) = \int_{\lambda_1}^{\lambda_2} h(\lambda, u) y(u) \, du, \quad \lambda_1 \leq \lambda \leq \lambda_2 \hspace{1cm} (2-64)
\]

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 [53, 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 \( h(\lambda, \lambda) \) as
\[
h(\lambda, \lambda) = \frac{V}{\hat{s}}(\lambda) \cdot H \cdot R^{-1}(\lambda) \hspace{1cm} (2-65)
\]

where
\( \bar{s}(\lambda) = s(\lambda) - \bar{s}(\lambda) \) \hspace{1cm} (2-66)

\[ \frac{V}{\bar{s}}(\lambda) = \text{Var}\left[\bar{s}(\lambda)\right] = \text{Var}\left[s(\lambda) - \bar{s}(\lambda)\right] \] \hspace{1cm} (2-67)

and

\[ \bar{R}(\lambda) = \text{Var}[n(\lambda)] \] \hspace{1cm} (2-68)

In the present research, \( \bar{R}(\lambda) \) is a scalar. The Kalman filter algorithms provide a natural and efficient technique for the computation of the estimation error variance \( \frac{V}{s}(\lambda) \). 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

\[ h(k,k) = \frac{V}{s}(k) \cdot H \cdot R^{-1}(k) \] \hspace{1cm} (2-69)

where

\[ k \in \mathbb{Z} : \lambda_1 \leq \lambda \leq \lambda_2 \]

are integers corresponding to discrete wavelengths of interest.

\[ \bar{s}(k) = s(k) - \bar{s}(k) \] \hspace{1cm} (2-70)

\[ \frac{V}{\bar{s}}(k) = \text{Var}[\bar{s}(k)] = \text{Var}[s(k) - \bar{s}(k)] \] \hspace{1cm} (2-71)

and
\[ \mathbf{R}(k) = \text{Var}[\mathbf{n}(k)] \]  \hspace{1cm} (2-72)

It should be noted that \( \mathbf{h}(k,k) \) is in vector form to conform with the state variable models used in the Kalman filter algorithm. The relationship between \( \mathbf{h}(k,k) \) and \( \mathbf{h}(k,k) \) 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 \( \mathbf{h}(k,k) \) and hence average information. The average information in \( \gamma(\lambda) \) about \( s(\lambda) \) is given by

\[ I(S,Y) = \frac{1}{2} \text{ke} \left[ \lambda_1, \lambda_2 \right] \sum h(k,k) \]  \hspace{1cm} (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 \( s(k+1) \) is obtained. Specifically, it is necessary to obtain the matrix \( \Phi \) and the vector \( \Gamma \). If the spectral process \( s(\lambda) \) is stationary and has a rational power spectral density, then \( \Phi \) and \( \Gamma \) may be determined from knowledge of \( K_s(\lambda,u) \) [VI, pp. 516-526]. In many physical situations \( K_s(\lambda,u) \) is not known. The parameters \( \Phi \) and \( \Gamma \) 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 \( s(\lambda) \)) 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 \( \hat{s}(k) \) of \( s(k) \) given by

\[
\hat{s}(k) = E \left[ s_k/Y_k \right]
\]  

(2-74)

where
\[ Y_k = \{ y(k) : k \in \left[ \lambda_1, \lambda_2 \right] \} \]  

= the observed spectral process.

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

\[ s(k) = H^T \hat{s}(k) \quad (2-76) \]

and

\[ \hat{s}(k) = H^T \hat{s}(k) \quad (2-77) \]

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

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

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 \( \hat{s}(k) \) and the estimation error \( \hat{s}(k) = s(k) - \hat{s}(k) \) is zero. That is,

\[ I(\hat{s}(k), \hat{s}(k)) = 0 \quad (2-78) \]

Now \( I(\hat{s}(k), \hat{s}(k)) = 0 \) if and only if \( \hat{s}(k) \) and \( \hat{s}(k) \) are statistically independent [61, p. 24]. But, based on our initial definitions and assumptions, \( \hat{s}(k) \) and \( \hat{s}(k) = s(k) - \hat{s}(k) \) are gaussian random variables. Hence, it is sufficient to observe that \( \hat{s}(k) \) and \( \hat{s}(k) \) are uncorrelated.
This is easily seen from

\[ E\left\{ \bar{s}(k)^T \bar{s}(k) \right\} = E\left[ E\left[ \frac{s(k)^T \bar{y}_k}{y_k} \right] \bar{s}(k) \right] \quad (2-79) \]

However,

\[ E\left\{ \bar{s}(k)^T \bar{y}_k \right\} = E\left[ \frac{s(k)^T - \bar{s}(k)^T}{y_k} \right] \]

\[ = E\left[ \frac{s(k)^T \bar{y}_k}{y_k} \right] - \bar{s}(k)^T = 0 \]

Hence

\[ E\left\{ \bar{s}(k)^T \bar{s}(k) \right\} = 0 \]

Thus \( \bar{s}(k) \) and \( \bar{s}(k) \) 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

\[ H(s(k)/\bar{s}(k)) = H(\bar{s}(k)/\bar{s}(k)) = H(\bar{s}(k)) \quad (2-80) \]

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

\[ p(s(k) - \bar{s}(k), \bar{s}(k)) = P_{ZW}(z_\bar{w}) = \]

\[ = P_{Z\bar{W}}(z + w, \bar{w}) = P_{Z\bar{W}}(s(k) + \bar{s}(k)) \]

Thus, since \( \bar{s}(k) = s(k) - \bar{s}(k) \) we can write
\[ p(s(k), \bar{s}(k)) = p(s(k) - \bar{s}(k), \bar{s}(k)) \]
\[ = p(\bar{s}(k), \bar{s}(k)) \] \hfill (2-82)

So we have
\[ p(s(k)/\bar{s}(k)) = p(\bar{s}(k)/\bar{s}(k)) \]
and hence
\[ H(s(k)/\bar{s}(k)) = H(\bar{s}(k)/\bar{s}(k)) \] \hfill (2-83)

Now we have previously shown that
\[ I(\bar{s}(k), \bar{s}(k)) = 0 \] \hfill (2-78)

Hence
\[ 0 = I(\bar{s}(k), \bar{s}(k)) = H(\bar{s}(k)) - H(s(k)/\bar{s}(k)) \]
or
\[ H(\bar{s}(k)) = H(s(k)/\bar{s}(k)) \] \hfill (2-84)

Now, we can write the average information \( I(s(k), \bar{s}(k)) \) as
\[ I(s(k), \bar{s}(k)) = H(s(k)) - H(s(k)/\bar{s}(k)) \] \hfill (2-85)

But using (2-83) and (2-84) in (2-85) we have
\[ I(s(k), \bar{s}(k)) = H(s(k)) - H(s'(k)) \] \hfill (2-86)

This is a very useful result. It shows that the average information in the estimate \( \bar{s}(k) \) of the state \( s(k) \) of the spectral process is directly
related to the entropy of the estimation error \( \hat{s}(k) - \bar{s}(k) \). Since for a given observation the state \( s(k) \) is fixed, it is clear that to maximize \( I(s(k), \hat{s}(k)) \) it is sufficient to minimize \( H(\hat{s}(k)) \).

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

\[
H(\hat{s}(k)) = \frac{1}{2} \log(2\pi e)^n |P_k|
\]  

(2-87)

where

\[
|P_k| = \det \left[ \mathbb{E} \left[ \hat{s}(k)\hat{s}^T(k) \right] \right].
\]

Hence it is clear that to minimize \( H(\hat{s}(k)) \) it is sufficient to minimize \( P_k \). 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 \( Y(x) \) 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.

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 consideration. The models considered in this research are stochastic difference equations having a general form given by

$$S(k) = \sum_{j=1}^{m_1} a_j S(k-j) + \sum_{j=1}^{m_2} b_j \psi(k-j) + w(k)$$  \hspace{1cm} (3-1)

where

- $S(k)$ is the spectral response at the discrete wavelength $k$. It is gaussian with mean and variance determined by the particular structure of (3-1).
- $\psi(k)$ is a zero mean independent gaussian disturbance with variance $\sigma$.
- $\psi(k-j)$ 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.
- $m_1$ and $m_2$ are constants that determine the dependence of $S(k)$ on preceding values of the process.

Thus the dynamic nature of the spectral process $S(k)$ is expressed in terms of its 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 assumptions 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) $w(k), k = 1, 2, \ldots$ is a sequence of zero mean identically distributed, independent gaussian random variables with variance $\sigma^2$. $w(k)$ is independent of $S(k-j)$ for all $j \geq 1$.

2) Define the unit delay operator $D$ by:

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

then (3-1) can be written as

$$S(k) \left[ 1 - \sum_{j=1}^{m_1} a_j b_j \right] = \sum_{j=1}^{m_2} b_j \phi(k-j) + w(k). \quad (3-2)$$

The assumption is that all the zeros of the expression
\[ A(D) = 1 - \sum_{j=1}^{m} a_j D^j \]

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 \( \psi_i(k) \), \( i = 1, 2, \ldots \). Represent these trend terms by the vector

\[ \mathbf{\psi}(k) = [\psi_1(k), \ldots, \psi_m(k)]^T. \]

Then the assumption is \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbf{\psi}(k)\mathbf{\psi}(k)^T \) 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

\[ \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} (\alpha_i \psi_i(k))^2 = \infty \quad (3-3) \]

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

\[ \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbf{\gamma}(k)\mathbf{\gamma}(k)^T = \]
Each of the diagonal terms in the above matrix is obviously positive.
Hence, by the assumption and the above comment,

\[ \prod_{i=1}^{\mathcal{I}} \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \psi_i(k)^2 > 0 \]

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

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

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

\[ \sum_{k=1}^{\infty} \sum_{i=1}^{\mathcal{I}} (a_{i,k} \psi_i(k))^2 \geq \sum_{k=1}^{\infty} b_i^2 \psi_i(k)^2. \]

But since
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \psi_i(k)^2 = B_i
\]

we have

\[
\lim_{N \to \infty} \sum_{k=1}^{N} \psi_i(k)^2 = \infty.
\]

Hence,

\[
b^2 \sum_{k=1}^{\infty} \psi_i(k)^2 = \infty,
\]

and the desired relation is shown.

These assumptions 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 \(m_1\). It is defined by

\[
S(k) = \sum_{j=1}^{m_1} a_j S(k-j) + w(k) \quad (3-4)
\]

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 \(m_1\), is given by

\[
S(k) = \sum_{j=1}^{m_1} a_j S(k-j) + C + w(k) \quad (3-5)
\]

Note that \(C\) corresponds to the coefficient of the trend term.
ψ(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

\[ \psi_S(k) = \sum_{j=1}^{m1} a_j \psi_S(k-j) + w(k) \quad (3-6) \]

where

\[ \psi_S(k) = S(k) - S(k-1) \]

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

\[ \psi_2S(k) = \sum_{j=1}^{m1} a_j \psi_2S(k-j) + w(k) \quad (3-7) \]

where

\[ \psi_2S(k) = S(k) - S(k-2) \]

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

\[
\begin{bmatrix}
S_1(k) \\
S_2(k) \\
S_3(k)
\end{bmatrix} =
\begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
a_3 & a_2 & a_1
\end{bmatrix}
\begin{bmatrix}
S_1(k-1) \\
S_2(k-1) \\
S_3(k-1)
\end{bmatrix}
+ \begin{bmatrix}
0 \\
0 \\
w(k)
\end{bmatrix}
\]

\[ = \Phi S(k-1) + rw(k) \]

and

\[ S(k) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} S(k) = H^T S(k) \]

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.

\[
S(k) = z^T(k-1)\hat{a} + w(k) \quad (3-8)
\]

where

\[
S(k-1) \quad \ldots \quad \begin{bmatrix} S(k-1) \\ \vdots \\ S(k-m_1) \\ \psi(k-1) \\ \vdots \\ \psi(k-m_2) \end{bmatrix} \quad \begin{bmatrix} a_1 \\ \vdots \\ \vdots \\ a_{m_1} \\ b_1 \\ \vdots \\ b_{m_2} \end{bmatrix}
\]

\[
z(k-1) = \begin{bmatrix} a_1 \\ \vdots \\ a_{m_1} \\ b_1 \\ \vdots \\ b_{m_2} \end{bmatrix}
\]

\[
\hat{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_{m_1} \end{bmatrix}
\]
We shall assume that at wavelength \( k = N \), we have accumulated the following data.

\[ X(N) = \{ S(N), \ldots, S(1), \ldots, S(0), \ldots, s(-m1) \} \]

\( X_1 \) 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 \( \theta \) be the true parameters of the model and let \( \rho \) 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 \( \theta \) and \( \rho \). The probability density function of the observations is needed to obtain the estimation scheme. This probability density can be written as

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

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

\[ \ldots \cdot p(S(1)/S(0), \ldots, S(-m1)) \cdot p(S(0), \ldots, S(-m1)) \]

(3-10)

The likelihood principle states that the estimates of the parameters \( \theta \) and \( \rho \) 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 \( p(s(0), \ldots, s(-m1)) \)
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)

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

(3-11)

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

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

or

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

(3-12)

Since we are considering conditional maximum likelihood estimates, the
optimum estimates $\theta^*$ and $\rho^*$ of $\theta$ and $\rho$ are those values that maximize the likelihood function $L(\theta, \rho, X(N))$. Equivalently, the logarithm of the likelihood function may be maximized. That is, maximize

$$L(\theta, \rho, X(N)) = \log L(\theta, \rho, X(N)) =$$

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

(3-13)

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

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

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

and equate to zero.

Thus the estimate becomes

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

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

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

or
\[ \rho^*(N) = \frac{1}{N} \sum_{k=1}^{N} (s(k) - \bar{s}(k-1)) \bar{z}^T(k-1) \bar{z} \]  

(3-15)

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

\[ \rho^*(N) = \frac{1}{N} \sum_{k=1}^{N} (s(k) - \bar{s}(k-1)) \bar{z}^T(k-1) \bar{z} \]  

(3-16)

If the second partial derivatives of the likelihood function are taken with respect to \( \theta \) and \( \rho \), it is seen that \( \theta^* \) and \( \rho^* \) 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 conditions 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

\[ E[\theta^*(N)/\theta, 0] = \theta + O \left[ \frac{1}{N^{1/2}} \right] \]  

(3-17)

where

\[ \frac{O(x)}{x} \rightarrow K \neq 0 \text{ as } x \rightarrow 0, \]

and
\[
\text{Cov}[\hat{\theta}(N)/\theta, p] = \rho \left[ E \left[ \sum_{k=1}^{N} Z(k-1)Z^T(k-1) \right] \right]^{-1} + o(1) \quad (3-18)
\]

where

\[
\frac{o(x)}{x} \to 0 \quad \text{as} \; x \to 0.
\]

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 \( \hat{\theta}(N) \) 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

\[
P(N) = \left[ \sum_{k=1}^{N} Z(k-1)Z^T(k-1) \right]^{-1} \quad (3-19)
\]

Then

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

or

\[
P(N)^{-1} = P(N-1)^{-1} + Z(N-1)Z^T(N-1) \quad (3-20)
\]

Now apply the matrix inversion lemma (Sage and Melsa [S3, pp. 499-500] and given in Appendix II) to (3-18) to obtain
\[ P(N) = P(N-1) - P(N-1)Z(N-1) \left[ 1 + Z^T(N-1)P(N-1)Z(N-1) \right]^{-1} Z^T(N-1)P(N-1) \]

or

\[ P(N) = P(N-1) - \frac{P(N-1)Z(N-1)Z^T(N-1)P(N-1)}{1 + Z^T(N-1)P(N-1)Z(N-1)}. \]  

(3-21)

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

\[ \theta^*(N) = P(N) \sum_{k=1}^{N} Z(k-1)S(k). \]  

(3-22)

Then

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

\[ = P(N) \sum_{k=1}^{N-1} Z(k-1)S(k) + P(N)Z(N-1)S(N). \]

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

\[ \theta^*(N) = P(N)P(N-1)^{-1} \theta^*(N-1) + P(N)Z(N-1)S(N). \]  

(3-23)

Now inserting the expression for \( P(N) \) from (3-21) into the first term on the right hand side of (3-23) we obtain
\[
\theta^*(N) = \left[ P(N-1) - \frac{P(N-1)Z(N-1)Z^T(N-1)P(N-1)}{1 + Z^T(N-1)P(N-1)Z(N-1)} \right] P(N-1)^{-1} \theta^*(N-1)
\]

\[
+ P(N)Z(N-1)S(N)
\]

or simplifying

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

+ P(N)Z(N-1)S(N)
\]

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

\[
P(N)Z(N-1) = P(N-1)Z(N-1) - \frac{P(N-1)Z(N-1)Z^T(N-1)P(N-1)Z(N-1)}{1 + Z^T(N-1)P(N-1)Z(N-1)}
\]

or

\[
P(N)Z(N-1) = \frac{P(N-1)Z(N-1)}{1 + Z^T(N-1)P(N-1)Z(N-1)}
\]

(3-25)

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

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

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

\[ \tilde{P}(N) = \tilde{P}(N-1) - \tilde{P}(N-1)_{Z(N-1)}^T \tilde{Z}(N-1) \tilde{P}(N-1) \]

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

Thus we have shown the form of the maximum likelihood estimators \( \hat{\theta} \) and \( \tilde{P} \). 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

\[ L(\hat{\theta}, \hat{\theta}^*) = (\hat{\theta} - \hat{\theta}^*)^T (\hat{\theta} - \hat{\theta}^*) \]

It is well known [K3], [S3], [V1], [R1], [F2] that the optimum estimate of \( \hat{\theta} \) under the quadratic loss criterion is given by
That is, \( \hat{\theta}^*(N) \) is the mean of the posterior density of \( \theta \) given the observations \( X(N) \). With a few preliminary assumptions, a form for the estimate \( \hat{\theta}^*(N) \) is derived. These assumptions are:

b1) The prior distribution of \( \theta \) is normal with mean \( \theta_0 \) and covariance matrix \( P_0 \).

b2) \( \rho \) is the variance of \( w(k) \).

b3) \( p(\theta/X_2) = p(\theta) \), which insures that \( \theta \) 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 \( \theta \) given the observations \( X(N) \).

**Assertion**

The posterior density \( p(\theta/X(N)) \) is normal with mean

\[
\hat{\theta}^*(N) = P(N) \left[ \sum_{k=1}^{N} Z(k-1) S(k) + P_0^{-1} \theta_0 \right]
\]

and variance \( P(N) \rho \) where

\[
P(N) = \left[ \sum_{k=1}^{N} Z(k-1) Z(k-1)^T + P_0^{-1} \right]^{-1}
\]
Demonstration

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

\[ \theta \sim N(\theta_0, P_0^\theta) \]

and

\[ w(k) \sim N(0, \sigma) \]

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

\[
\frac{p(\theta/X(N))}{p(X(N))} = \frac{p(X(N)/\theta)p(\theta)}{p(X(N))} \quad (3-30)
\]

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

\[
p(X(N)/\theta) = p(S(N), \ldots S(1), S(0), \ldots S(-m)/\theta)
\]

This can be written as

\[
p(X(N)/\theta) = p(S(N)/S(N-1), \ldots S(1), x_2, \theta)
\]

\[
p(S(N-1)/S(N-2), \ldots S(1), x_2, \theta)
\]

\[
\ldots p(S(1)/x_2, \theta) \cdot p(\theta/x_2) \cdot p(x_2)
\]

But from (3-8) we can write

\[
p(S(k)/S(k-1), \ldots S(1), x_2, \theta) =
\]
The second term in the numerator of (3-30) is

\[
p(\theta) = \frac{1}{(2\pi \rho |P_0|)^{1/2}} \exp \left\{ -\frac{1}{2\rho} (\theta - \theta_0)^T P_0^{-1} (\theta - \theta_0) \right\}
\]

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 \( \theta \). It does depend on \( \theta_0, P_0, \rho, \) and \( T(k-1) \). 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

\[
p(\theta/X(N)) = \prod_{k=1}^{N} \left[ \frac{1}{\sqrt{2\pi \rho}} \exp \left\{ -\frac{1}{2\rho} (S(k) - Z^T(k-1) \theta)^2 \right\} \right] \cdot \frac{1}{(2\pi \rho |P_0|)^{1/2}} \exp \left\{ -\frac{1}{2\rho} (\theta - \theta_0)^T P_0^{-1} (\theta - \theta_0) \right\}
\]

\[
\cdot \frac{1}{p(X_1/X_2)}
\]

Collecting all normalization factors as a single term we can write
\[ p(\theta/X(N)) = K_4 \exp \left\{ \frac{-1}{2\rho} \left[ \sum_{k=1}^{N} (S(k) - Z^T(k-1)\theta) \right]^2 \right\} + (\theta - \theta_0)^T P^{-1} (\theta - \theta_0) \]  

(3-31)

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

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

\[ + \theta^T P^{-1} \theta - 2\theta^T P_0 \theta + \theta^T P^{-1} \theta_0 \]  

Rearrange and label some terms to give the exponent as

\[ \frac{-1}{2\rho} \left[ \theta^T \left[ \sum_{k=1}^{N} Z(k-1)Z^T(k-1) + \theta^T P_0^{-1}\theta \right] \theta \right. \]

\[ - 2 \left[ \sum_{k=1}^{N} Z(k-1)S(k) + \theta^T P_0^{-1}\theta_0 \right] P(N)^{-1} \theta \]

\[ + \sum_{k=1}^{N} S^2(k) + \theta^T P_0^{-1} \theta_0 \]  

From the expressions labeled as \( P(N)^{-1} \) and \( \theta^*(N)^T \) above, we see that in order to "complete the square," it is necessary to add and subtract the term
When the operation of "completing the square" is carried out, the extraneous terms do not depend on \( \theta \). Therefore, they also contribute to the normalization factor previously mentioned. These manipulations give the exponent as

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

where

\[
\theta^*(N) = \bar{P}(N) \left[ \sum_{k=1}^{N} S(k) Z(k-1) + P_0^{-1} \Theta_0 \right]
\]

and

\[
P(N) = \left[ \sum_{k=1}^{N} Z(k-1) Z^T(k-1) + P_0^{-1} \right]^{-1}.
\]

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

\[
K_1 = \frac{1}{\left( 2\pi \right)^{\frac{m}{2}} |P(N)|^{1/2}} \text{ where } m = m_1 + m_2.
\]

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

\[
P(N)^{-1} = \sum_{k=1}^{N} Z(k-1) Z(k-1)^T + P_0^{-1}
\]
= P(N-1)^{-1} + Z(N-1)Z(N-1)^T

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

So

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

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

\[ |P(N)| = \frac{|P(N-1)|}{1 + Z(N-1)^TP(N-1)Z(N-1)} \]  (3-32)

We note that this is a recursive relation with

\[ P(1) = \left[ I + P_0 Z(0)Z(0)^T \right]^{-1}P_0 \]

and

\[ |P(1)| = \frac{|P_0|}{1 + Z(0)^TP_0Z(0)} \]

Hence, we can write

\[ |P(N)| = \frac{|P_0|}{\prod_{k=1}^{N} (1 + Z(k-1)^TP(k-1)Z(k-1))} \]  (3-33)

Hence, the normalization factor \( K_1 \) is given by
Thus the assertion is demonstrated.

The Bayesian estimate $\hat{\theta}^*(N)$ 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 $\hat{\theta}^*$ are chosen. In the Bayesian technique we can use our knowledge (or assumption) of the prior density function. Hence, logical initial values would be

$$\hat{\theta}^*(0) = \theta_0$$

(3-35)

and

$$P(0) = P_0.$$  

(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 $\hat{\theta}$ 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

\[ S(k) = Z^T(k-1)\theta + w(k) \]  \hspace{1cm} (3-8)

A class of models is defined by Kashyap and Rao [K3, p. 181] as the triple \((f, H, \varnothing)\) where \(f\) is the stochastic difference equation (3-8), \(H\) is the range of values for \(\varnothing\) and \(\varnothing\) is the range of values for \(\rho\). A member of the class \((f, H, \varnothing)\) is written \((f, \varnothing, \rho)\). 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 than AR(2) models. The classes are said to be nonoverlapping. Thus, given several nonoverlapping classes and the empirical data set \(X = \{S(1), \ldots, S(N)\}\), 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 \((f, \Theta_0, \rho_0)\) where \( \Theta_0 = (\theta_0, \rho) \) is unknown. Let the probability density function of \( X \) be given by \( p(X, \theta_0) \). Furthermore, assume that \( p(X, \theta_0) \) is a known function of \( X \) and \( \theta_0 \). Then the log-likelihood function of \( p(X, \theta_0) \) is given in terms of \( X \) alone by the following theorem.

Theorem (Kashyap and Rao)

Let \( \hat{\theta} \) be the maximum likelihood estimate of \( \theta_0 \) based on the empirical observation set \( X \). Then

\[
E[\ln(p(X, \theta_0)) / \hat{\theta}] = \\
= L + E[0(||\theta_0 - \hat{\theta}||^3) / \hat{\theta}]
\]

where

\[
L = \ln p(X, \hat{\theta}) - n \phi
\]

(3-38)

\( n \phi \) = 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 \( \ln(p(X, \theta_0)) \), 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 \( \hat{\phi}_i \) of \( \phi_0 \) assuming \( \phi_0 \in H_i \) using the empirical data \( X \). Compute the class of log-likelihood functions \( L_i \) as

\[
L_i = \ln(p(X, \hat{\phi}_i^*)) - n_i
\]  

(3-39)

where \( n_i \) is the dimension of \( \hat{\phi}_i = (\hat{\theta}_0, \rho_0) \).

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

\[
\ln p(X, \hat{\phi}^*) = \ln p(S(N), \ldots, S(m+1)/S(m), \ldots, S(1), \hat{\phi}^*)
\]

\[
+ \ln p(S(m), \ldots, S(1), \hat{\phi}^*) \quad \text{.} 
\]  

(3-40)

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

\[
I = \ln p(S(N), \ldots, S(m+1)/S(m), \ldots, S(1), \hat{\phi}^*)
\]

\[
= \ln \left[ \prod_{k=m+1}^{N} p(S(k)/S(k-1), \ldots, S(k-m1), \hat{\phi}^*) \right] 
\]
\[ \ln \left[ \prod_{k=m_1+1}^{N} \frac{1}{2\pi \rho^*} \exp \left( -\frac{1}{2\rho^*} (S(k) - \text{Z}^T(k-1) \alpha^*)^2 \right) \right] \]

\[ = - \frac{(N-m_1)}{2} \ln(2\pi \rho^*) - \frac{1}{2\rho^*} \sum_{k=m_1+1}^{N} (S(k) - \text{Z}^T(k-1) \alpha^*)^2 \]  \quad (3-41)

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

\[ \rho^* = \frac{1}{N-m_1} \sum_{k=m_1+1}^{N} (S(k) - \text{Z}^T(k-1) \alpha^*)^2 \]  \quad (3-42)

Hence (3-41) can be written

\[ I = -\frac{(N-m_1)}{2} \ln(2\pi \rho^*) - \frac{1}{2} (N-m_1) \]  \quad (3-43)

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

\[ \ln p(S(m), \ldots, S(1), \alpha^*) \]

\[ = \ln \left[ \prod_{k=1}^{m_1} p(S(k)) \right] \]

\[ = \sum_{k=1}^{m_1} \ln p(S(k)) \]
\[-\frac{m_1}{2} \ln(2\pi \rho S) - \frac{1}{\rho S} \sum_{k=1}^{m_1} s(k)^2 \]  

(3-44)

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

\[
L = \left[ N \frac{1}{2} n_p^* - m_1 \right] L_a - 
\]

\[
+ \left[ N - \frac{1}{2} \ln 2\pi - \frac{N}{2} \ln \left( \frac{p_S}{\rho} \right) \right] - \frac{N}{2} + \frac{1}{2} \left[ 3m_1 - 1 \right] \sum_{k=1}^{m_1} s(k)^2 ]
\]

\[
= L_a + L_b .
\]

(3-45)

It is noted that since in general \(m_1 << N\), 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

\[
L_a = N \frac{1}{2} \ln(p^*) - m_1 .
\]

(3-46)

The \(\ln p^*\) 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 \(m_1\) term acts to oppose the selection of increasingly complex models. This is a quantitative method of incorporating the principle of parsimony in model
selection. 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 analytically 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 straightforward. 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 \( \rho \),

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

\[
X(k) = S(k) - Z^T(k-1)\bar{e}^*, \quad k = 1, \ldots, N. \tag{3-47}
\]

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 statistical 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 \( w(k), k = 1, \ldots, N, \) 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 \( \sigma^2, \sigma > 0 \). 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

\[
|t(x)| < n_0, \text{ accept } H_0 \tag{3-48}
\]

\[
|t(x)| \geq n_0, \text{ reject } H_0
\]

where

\[
t(x) = \frac{N \bar{x}}{\sqrt{\hat{\sigma}}}
\]

(3-49)
\[ \overline{x} = \frac{1}{N} \sum_{k=1}^{N} x(k) \]  
\[ \rho = \frac{1}{N-1} \sum_{k=1}^{N} (x(k) - \overline{x})^2 \]  

(3-50)  

(3-51)

t(x) is the Student's t-distribution with N-1 degrees of freedom and hence \( n_0 \) is chosen such that

\[ P(|t(x)| < n_0 | H_0) = 1 - \alpha \]

where \( \alpha \) 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

\[ H_0: x(k) = w(k) \]

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 \( \sigma^2 \). To implement the test, we define the covariance of the residual data at lag \( k \) as

\[
\hat{R}_k = \frac{1}{N-k} \sum_{j=k+1}^{N} x(j)x(j-k) .
\]  

(3-52)

The required test statistic is then

\[
\eta(x) = \frac{(N-n_1) \sum_{k=1}^{n_1} (\hat{R}_k)^2}{\langle \hat{R}_0 \rangle^2}
\]

(3-53)

where \( n_1 \) 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 \( (x^2) \) distributed with \( n_1 \) degrees of freedom. This gives the decision rule

\[
\eta(x) \begin{cases} 
< n_0, & \text{accept } H_0 \\
\geq n_0, & \text{reject } H_0
\end{cases}
\]

(3-54)

where \( n_0 \) is computed from

\[
p(\eta(x) \geq n_0 | H_0) = \alpha .
\]

(3-55)

and \( \alpha \) 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

\[ x(k) = \sum_{j=1}^{[N/2]} (a_j \cos \omega_j k + b_j \sin \omega_j k) + w(t) \quad (3-56) \]

where

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

and

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

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:

\[ H_0: \quad x(k) = w(k) \]

\[ H_1: \quad \text{at least one of the components} \quad \alpha_j, \beta_j \quad j = 1, \ldots, \left\lfloor \frac{N}{2} \right\rfloor \quad \text{is nonzero} \quad (3-57) \]

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

\[ g_k = \frac{\sum_{j=1}^{N/2} \gamma_j^2}{\sum_{j=1}^{N/2} \gamma_j^2} \quad k = 1, \ldots, \left\lfloor \frac{N}{2} \right\rfloor \quad (3-58) \]

where

\[ \gamma_j^2 = \left[ \frac{2}{N} \sum_{k=1}^{N} x(k) \cos \omega_{jk} \right]^2 + \left[ \frac{2}{N} \sum_{k=1}^{N} x(k) \sin \omega_{jk} \right]^2 \quad (3-59) \]

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 \((0.5,1.0)\) at distances \( \pm \chi \sqrt{\frac{N}{2}} \) is given by (see Kashyap and Rao [K3, p. 208])

\[ \sum_{j=-\infty}^{\infty} (-1)^j \exp\{-2\chi^2 j^2 \} \]

The parameter \( \chi \) is 1.36 for 0.95 probability and 1.63 for 0.99 probabil-
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 straightforward.
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 correlograms 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

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

where

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

This estimate is called the correlogram 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 $R(k)$ 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
\[ \hat{R}_M(k) = \frac{1}{J} \sum_{j=1}^{J} \hat{R}_j(k) \]  

(3-61)

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

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

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

(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.

\[ \hat{R}_M(k) - 2\sigma_M(k) \leq \hat{R}(k) \leq \hat{R}_M(k) + 2\sigma_M(k), \quad k=1, \ldots, N-1 \]  

(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 [01, p. 540]). Therefore, the following relation is suggested.

\[ \hat{R}_M(k) - 2\sigma_M(k) \leq \hat{R}(k) \leq \hat{R}_M(k) + 2\sigma_M(k), \quad k << N. \]  

(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 [91, pp. 545-554]. However, as noted by Oppenheim and Schaffer, and Bartlett [83, 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

$$\gamma^2(k) = \left\{ \frac{2}{N} \sum_{j=1}^{N} S(j) \cos \omega_k j \right\}^2 + \left\{ \frac{2}{N} \sum_{j=1}^{N} S(j) \sin \omega_k j \right\}^2$$

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 $$\omega_k$$ 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 obser-
vations of wheat scenes and was selected primarily for two reasons.
First, it is desired to demonstrate the techniques for a scene that con-
sists 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 ob-
serving a more general scene. Also study of the second data set pro-
vides 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

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

Table IV-2. Combined Scene Data Run Numbers

<table>
<thead>
<tr>
<th>Run Number</th>
<th>Run Number</th>
<th>Run Number</th>
<th>Run Number</th>
<th>Run Number</th>
<th>Run Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>75768400</td>
<td>75768800</td>
<td>75774300</td>
<td>75775700</td>
<td>75776200</td>
<td></td>
</tr>
<tr>
<td>75768500</td>
<td>75768900</td>
<td>75774400</td>
<td>75775800</td>
<td>75776300</td>
<td></td>
</tr>
<tr>
<td>75768600</td>
<td>75774100</td>
<td>75774500</td>
<td>75775900</td>
<td>75776400</td>
<td></td>
</tr>
<tr>
<td>75768700</td>
<td>75774200</td>
<td>75774600</td>
<td>75776100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 absorption 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 absorption 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

<table>
<thead>
<tr>
<th>Band Number</th>
<th>Spectral Wavelength Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.4528 - .5380 μm</td>
</tr>
<tr>
<td>2</td>
<td>.5380 - .6239 μm</td>
</tr>
<tr>
<td>3</td>
<td>.6239 - .7097 μm</td>
</tr>
<tr>
<td>4</td>
<td>.7097 - .8517 μm</td>
</tr>
<tr>
<td>5</td>
<td>.8517 - .9910 μm</td>
</tr>
<tr>
<td>6</td>
<td>.9910 - 1.130 μm</td>
</tr>
<tr>
<td>7</td>
<td>1.130 - 1.344 μm</td>
</tr>
<tr>
<td>8</td>
<td>1.446 - 1.821 μm</td>
</tr>
<tr>
<td>9</td>
<td>1.959 - 2.386 μm</td>
</tr>
</tbody>
</table>

Table IV-4. Spectral Bands for Combined Scene

<table>
<thead>
<tr>
<th>Band Number</th>
<th>Spectral Wavelength Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.4565 - .5402 μm</td>
</tr>
<tr>
<td>2</td>
<td>.5402 - .6246 μm</td>
</tr>
<tr>
<td>3</td>
<td>.6246 - .7097 μm</td>
</tr>
<tr>
<td>4</td>
<td>.7097 - .8481 μm</td>
</tr>
<tr>
<td>5</td>
<td>.8481 - .9850 μm</td>
</tr>
<tr>
<td>6</td>
<td>.9850 - 1.122 μm</td>
</tr>
<tr>
<td>7</td>
<td>1.122 - 1.307 μm</td>
</tr>
<tr>
<td>8</td>
<td>1.451 - 1.818 μm</td>
</tr>
<tr>
<td>9</td>
<td>1.967 - 2.386 μm</td>
</tr>
</tbody>
</table>
3. **Identified Models for the Empirical Data**

This section discusses the models identified for the spectral 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 arbitrary (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 \( \mu \)m for the wheat data and .4565 to .5402 \( \mu \)m 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 $1.26 \times 10^{-3}$. 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 autoregressive 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 independence 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 $1.46 \times 10^{-3}$ 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

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

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.20828</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.16650</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.16368</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.16533</td>
</tr>
<tr>
<td>$a_5$</td>
<td>0.14243</td>
</tr>
<tr>
<td>$a_6$</td>
<td>0.16945</td>
</tr>
</tbody>
</table>
Table IV-7. Identified Models for Band 1 of Combined Scene

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

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.031594</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.037784</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.13395</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.089051</td>
</tr>
<tr>
<td>$a_5$</td>
<td>0.084141</td>
</tr>
<tr>
<td>$a_6$</td>
<td>0.10954</td>
</tr>
<tr>
<td>$a_7$</td>
<td>0.10055</td>
</tr>
<tr>
<td>$a_8$</td>
<td>0.075585</td>
</tr>
<tr>
<td>$a_9$</td>
<td>0.12125</td>
</tr>
<tr>
<td>$a_{10}$</td>
<td>0.033981</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.064937</td>
</tr>
</tbody>
</table>
B) Band 2

This spectral band encompasses the wavelengths .5380 - .6239 \(\mu\)m for the wheat scene and .5402 - .6246 \(\mu\)m 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 \(6.19 \times 10^{-5}\) 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 \(1.72 \times 10^{-4}\) 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 \(\mu\)m for the wheat scene and .6246 - .7097 \(\mu\)m 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

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>541.8</td>
<td>462.4</td>
<td>490.7</td>
</tr>
<tr>
<td>2</td>
<td>560.0</td>
<td>513.3</td>
<td>499.4</td>
</tr>
<tr>
<td>3</td>
<td>553.3</td>
<td>530.1</td>
<td>507.9</td>
</tr>
<tr>
<td>4</td>
<td>542.9</td>
<td>535.5</td>
<td>513.6</td>
</tr>
<tr>
<td>5</td>
<td>534.0</td>
<td>537.9</td>
<td>520.0</td>
</tr>
<tr>
<td>6</td>
<td>526.5</td>
<td>539.2</td>
<td>525.9</td>
</tr>
<tr>
<td>7</td>
<td>518.7</td>
<td>541.1</td>
<td>530.1</td>
</tr>
<tr>
<td>8</td>
<td>511.0</td>
<td>541.1</td>
<td>538.0</td>
</tr>
<tr>
<td>9</td>
<td>505.5</td>
<td>540.1</td>
<td>540.6</td>
</tr>
<tr>
<td>10</td>
<td>498.8</td>
<td>538.2</td>
<td>539.4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>.50360</td>
</tr>
<tr>
<td>$a_2$</td>
<td>.50189</td>
</tr>
</tbody>
</table>
Table IV-11. Identified Models for Band 2 of Combined Scene

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Selection Criterion</th>
<th>Autoregressive</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>491.6</td>
<td>439.8</td>
<td>473.9</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>492.0</td>
<td>469.4</td>
<td>474.4</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>490.9</td>
<td>479.9</td>
<td>477.4</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>484.0</td>
<td>478.3</td>
<td>477.0</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>478.0</td>
<td>475.9</td>
<td>478.0</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>474.4</td>
<td>474.4</td>
<td>478.9</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>471.2</td>
<td>472.3</td>
<td>477.6</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>468.6</td>
<td>472.3</td>
<td>479.0</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>468.1</td>
<td>471.6</td>
<td>479.3</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>465.7</td>
<td>470.3</td>
<td>478.3</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>462.8</td>
<td>469.4</td>
<td>477.3</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>459.6</td>
<td>467.9</td>
<td>474.8</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>456.6</td>
<td>467.2</td>
<td>472.1</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>453.9</td>
<td>466.0</td>
<td>469.0</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>451.4</td>
<td>465.1</td>
<td>465.4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>.50475</td>
</tr>
<tr>
<td>$a_2$</td>
<td>.49893</td>
</tr>
</tbody>
</table>
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 $6.00 \times 10^{-5}$, 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 slightly better selection criterion (if computed to more decimal places). The residual variance for the selected model is $1.15 \times 10^{-4}$, 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 $.7097 - .8517 \mu m$ for the wheat scene and $.7097 - .8481 \mu m$ 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

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>532.9</td>
<td>464.3</td>
<td>465.5</td>
</tr>
<tr>
<td>2</td>
<td>511.8</td>
<td>484.3</td>
<td>478.3</td>
</tr>
<tr>
<td>3</td>
<td>489.1</td>
<td>481.9</td>
<td>489.5</td>
</tr>
<tr>
<td>4</td>
<td>470.2</td>
<td>474.3</td>
<td>501.5</td>
</tr>
<tr>
<td>5</td>
<td>453.9</td>
<td>464.9</td>
<td>513.0</td>
</tr>
<tr>
<td>6</td>
<td>440.5</td>
<td>456.0</td>
<td>523.3</td>
</tr>
<tr>
<td>7</td>
<td>429.7</td>
<td>447.9</td>
<td>531.8</td>
</tr>
<tr>
<td>8</td>
<td>420.8</td>
<td>440.6</td>
<td>540.4</td>
</tr>
<tr>
<td>9</td>
<td>414.1</td>
<td>434.8</td>
<td>549.4</td>
</tr>
<tr>
<td>10</td>
<td>408.8</td>
<td>429.3</td>
<td>551.5</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>552.8</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>549.9</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td>546.4</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>538.7</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td>532.4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>.098225</td>
</tr>
<tr>
<td>a₂</td>
<td>.099391</td>
</tr>
<tr>
<td>a₃</td>
<td>.099802</td>
</tr>
<tr>
<td>a₄</td>
<td>.10021</td>
</tr>
<tr>
<td>a₅</td>
<td>.098787</td>
</tr>
<tr>
<td>a₆</td>
<td>.099604</td>
</tr>
<tr>
<td>a₇</td>
<td>.099389</td>
</tr>
<tr>
<td>a₈</td>
<td>.10044</td>
</tr>
<tr>
<td>a₉</td>
<td>.10047</td>
</tr>
<tr>
<td>a₁₀</td>
<td>.099348</td>
</tr>
<tr>
<td>a₁₁</td>
<td>.10040</td>
</tr>
</tbody>
</table>
Table IV-15. Identified Models for Band 3 of Combined Scene

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

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_1</td>
<td>.099769</td>
</tr>
<tr>
<td>a_2</td>
<td>.098600</td>
</tr>
<tr>
<td>a_3</td>
<td>.10137</td>
</tr>
<tr>
<td>a_4</td>
<td>.10010</td>
</tr>
<tr>
<td>a_5</td>
<td>.10057</td>
</tr>
<tr>
<td>a_6</td>
<td>.097140</td>
</tr>
<tr>
<td>a_7</td>
<td>.10123</td>
</tr>
<tr>
<td>a_8</td>
<td>.10080</td>
</tr>
<tr>
<td>a_9</td>
<td>.097301</td>
</tr>
<tr>
<td>a_10</td>
<td>.098323</td>
</tr>
<tr>
<td>a_11</td>
<td>.10038</td>
</tr>
</tbody>
</table>
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 $4.43 \times 10^{-4}$ 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 $1.20 \times 10^{-3}$ 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 $\mu$m for the wheat scene and .8481 - .9850 $\mu$m 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

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Selection Criterion</th>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>autoregressive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>plus constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>trend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>469.4</td>
<td>433.3</td>
<td>438.1</td>
</tr>
<tr>
<td>2</td>
<td>439.2</td>
<td>471.3</td>
<td>450.0</td>
</tr>
<tr>
<td>3</td>
<td>417.3</td>
<td>454.3</td>
<td>459.3</td>
</tr>
<tr>
<td>4</td>
<td>403.8</td>
<td>442.1</td>
<td>463.0</td>
</tr>
<tr>
<td>5</td>
<td>396.7</td>
<td>433.6</td>
<td>473.6</td>
</tr>
<tr>
<td>6</td>
<td>394.7</td>
<td>428.7</td>
<td>477.4</td>
</tr>
<tr>
<td>7</td>
<td>395.3</td>
<td>425.4</td>
<td>480.2</td>
</tr>
<tr>
<td>8</td>
<td>393.0</td>
<td>423.6</td>
<td>481.1</td>
</tr>
<tr>
<td>9</td>
<td>400.2</td>
<td>422.9</td>
<td>481.8</td>
</tr>
<tr>
<td>10</td>
<td>403.6</td>
<td>423.4</td>
<td>482.0</td>
</tr>
<tr>
<td>11</td>
<td>406.9</td>
<td>423.8</td>
<td>480.8</td>
</tr>
<tr>
<td>12</td>
<td>410.2</td>
<td>425.6</td>
<td>477.0</td>
</tr>
<tr>
<td>13</td>
<td>411.9</td>
<td>425.8</td>
<td>474.5</td>
</tr>
<tr>
<td>14</td>
<td>412.7</td>
<td>426.7</td>
<td>472.4</td>
</tr>
<tr>
<td>15</td>
<td>412.5</td>
<td>426.1</td>
<td>469.1</td>
</tr>
<tr>
<td>16</td>
<td>411.4</td>
<td>426.5</td>
<td>466.9</td>
</tr>
<tr>
<td>17</td>
<td>410.2</td>
<td>424.9</td>
<td>464.1</td>
</tr>
<tr>
<td>18</td>
<td>408.3</td>
<td>423.4</td>
<td>461.2</td>
</tr>
<tr>
<td>19</td>
<td>407.7</td>
<td>421.5</td>
<td>458.1</td>
</tr>
<tr>
<td>20</td>
<td>406.7</td>
<td>419.4</td>
<td>455.1</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>.93593</td>
</tr>
<tr>
<td>$c$</td>
<td>.13969</td>
</tr>
</tbody>
</table>
Table IV-19. Identified Models for Band 4 of Combined Scene

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>402.5</td>
<td>418.3</td>
<td>387.2</td>
</tr>
<tr>
<td>2</td>
<td>379.3</td>
<td>403.6</td>
<td>394.8</td>
</tr>
<tr>
<td>3</td>
<td>364.8</td>
<td>390.9</td>
<td>400.8</td>
</tr>
<tr>
<td>4</td>
<td>358.8</td>
<td>382.9</td>
<td>403.3</td>
</tr>
<tr>
<td>5</td>
<td>358.3</td>
<td>379.1</td>
<td>404.7</td>
</tr>
<tr>
<td>6</td>
<td>359.0</td>
<td>376.9</td>
<td>408.3</td>
</tr>
<tr>
<td>7</td>
<td>362.3</td>
<td>376.8</td>
<td>407.8</td>
</tr>
<tr>
<td>8</td>
<td>364.6</td>
<td>376.3</td>
<td>408.6</td>
</tr>
<tr>
<td>9</td>
<td>366.6</td>
<td>376.4</td>
<td>408.4</td>
</tr>
<tr>
<td>10</td>
<td>368.4</td>
<td>376.4</td>
<td>404.8</td>
</tr>
<tr>
<td>11</td>
<td>368.6</td>
<td>376.6</td>
<td>406.5</td>
</tr>
<tr>
<td>12</td>
<td>368.8</td>
<td>375.7</td>
<td>405.0</td>
</tr>
<tr>
<td>13</td>
<td>367.9</td>
<td>374.8</td>
<td>402.6</td>
</tr>
<tr>
<td>14</td>
<td>366.0</td>
<td>372.4</td>
<td>401.7</td>
</tr>
<tr>
<td>15</td>
<td>364.2</td>
<td>371.2</td>
<td>400.0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_{1}</td>
<td>.93382</td>
</tr>
<tr>
<td>c</td>
<td>.15938</td>
</tr>
</tbody>
</table>
Table IV-21. Identified Models for Band 5 of Wheat Scene

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Selection Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Autoregressive plus constant</td>
<td>Integrated autoregressive trend</td>
</tr>
<tr>
<td>1</td>
<td>339.9</td>
<td>339.7</td>
</tr>
<tr>
<td>2</td>
<td>734.1</td>
<td>733.6</td>
</tr>
<tr>
<td>3</td>
<td>330.6</td>
<td>330.0</td>
</tr>
<tr>
<td>4</td>
<td>328.4</td>
<td>327.7</td>
</tr>
<tr>
<td>5</td>
<td>327.4</td>
<td>327.3</td>
</tr>
<tr>
<td>6</td>
<td>326.3</td>
<td>325.2</td>
</tr>
<tr>
<td>7</td>
<td>326.5</td>
<td>325.7</td>
</tr>
<tr>
<td>8</td>
<td>326.2</td>
<td>325.0</td>
</tr>
<tr>
<td>9</td>
<td>324.8</td>
<td>322.9</td>
</tr>
<tr>
<td>10</td>
<td>322.6</td>
<td>321.7</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>1.00055</td>
</tr>
</tbody>
</table>
with the best selection 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 $2.80 \times 10^{-3}$, 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 $3.86 \times 10^{-3}$, an indication of a good fit to the empir­
ical data. The estimated coefficients for the selected third order au­
toregressive model are listed in Table IV-24.

F) **Band 6**

The spectral intervals included in band 6 are .9910 - 1.130 um for
the wheat scene and .9850 - 1.122 um for the combined scene. The three
basic models were identified for the wheat scene and are listed accord­
ing to order and selection 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 $1.44 \times 10^{-3}$ 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 5 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

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive Model plus constant trend</th>
<th>Autoregressive plus constant</th>
<th>Integrated autoregressive trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>307.2</td>
<td>307.3</td>
<td>308.3</td>
</tr>
<tr>
<td>2</td>
<td>310.3</td>
<td>309.7</td>
<td>306.7</td>
</tr>
<tr>
<td>3</td>
<td>313.8</td>
<td>313.0</td>
<td>307.5</td>
</tr>
<tr>
<td>4</td>
<td>309.5</td>
<td>308.7</td>
<td>305.7</td>
</tr>
<tr>
<td>5</td>
<td>306.5</td>
<td>305.9</td>
<td>304.1</td>
</tr>
<tr>
<td>6</td>
<td>306.2</td>
<td>305.3</td>
<td>303.3</td>
</tr>
<tr>
<td>7</td>
<td>304.3</td>
<td>304.0</td>
<td>302.3</td>
</tr>
<tr>
<td>8</td>
<td>302.3</td>
<td>301.8</td>
<td>300.5</td>
</tr>
<tr>
<td>9</td>
<td>301.0</td>
<td>300.5</td>
<td>299.9</td>
</tr>
<tr>
<td>10</td>
<td>299.0</td>
<td>298.6</td>
<td>298.4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>.38814</td>
</tr>
<tr>
<td>$a_2$</td>
<td>.28995</td>
</tr>
<tr>
<td>$a_3$</td>
<td>.32304</td>
</tr>
</tbody>
</table>
Table IV-25. Identified Models for Band 6 of Wheat Scene

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

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1)</td>
<td>.51004</td>
</tr>
<tr>
<td>(a_2)</td>
<td>.47772</td>
</tr>
<tr>
<td>(c)</td>
<td>.14773</td>
</tr>
</tbody>
</table>
tern in Table IV-27. Based on the selection criterion, the first order autoregressive model is chosen. It has only a slightly 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 $1.91 \times 10^{-3}$ 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 preceding 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 $2.42 \times 10^{-3}$ 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 Selection Criterion

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>355.9</td>
<td>355.8</td>
<td>353.9</td>
</tr>
<tr>
<td>2</td>
<td>347.8</td>
<td>346.7</td>
<td>353.1</td>
</tr>
<tr>
<td>3</td>
<td>338.5</td>
<td>337.5</td>
<td>351.8</td>
</tr>
<tr>
<td>4</td>
<td>332.2</td>
<td>332.5</td>
<td>351.3</td>
</tr>
<tr>
<td>5</td>
<td>329.1</td>
<td>328.5</td>
<td>349.2</td>
</tr>
<tr>
<td>6</td>
<td>326.2</td>
<td>327.7</td>
<td>348.7</td>
</tr>
<tr>
<td>7</td>
<td>325.7</td>
<td>328.2</td>
<td>347.8</td>
</tr>
<tr>
<td>8</td>
<td>352.6</td>
<td>326.4</td>
<td>347.5</td>
</tr>
<tr>
<td>9</td>
<td>323.7</td>
<td>325.4</td>
<td>345.4</td>
</tr>
<tr>
<td>10</td>
<td>322.5</td>
<td>323.4</td>
<td>345.4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>1.000091</td>
</tr>
</tbody>
</table>
Table IV-29. Identified Models for Band 7 of Wheat Scene

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

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.14802</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.061559</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.10349</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.11204</td>
</tr>
<tr>
<td>$a_5$</td>
<td>0.10538</td>
</tr>
</tbody>
</table>
Table IV-31. Identified Models for Band 7 of Combined Scene Selection Criterion

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

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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_4$</td>
<td>.64063</td>
</tr>
<tr>
<td>$a_2$</td>
<td>.17721</td>
</tr>
<tr>
<td>$a_3$</td>
<td>.11417</td>
</tr>
<tr>
<td>$-a_4$</td>
<td>.088031</td>
</tr>
<tr>
<td>$a_5$</td>
<td>.10856</td>
</tr>
<tr>
<td>$a_6$</td>
<td>-.058380</td>
</tr>
<tr>
<td>$a_7$</td>
<td>-.038789</td>
</tr>
<tr>
<td>$a_8$</td>
<td>-.061748</td>
</tr>
<tr>
<td>$a_9$</td>
<td>-.017016</td>
</tr>
<tr>
<td>$c$</td>
<td>.49218</td>
</tr>
</tbody>
</table>
model fails the correlogram validation test. Hence, the next alternate
model is chosen according to Table IV-31. This is the ninth order au-
toregressive plus constant trend model. This model passes the valida-
tion tests and the residual variance is $3.27 \times 10^{-2}$. 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 in-
tegrated 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 autore-
gressive model. This model passes all the validation tests. The resi-
dual variance for the selected model is $3.18 \times 10^{-4}$ 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 in-
tegrated autoregressive model.

The three basic models were also identified for band 8 of the com-
bined 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

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Selection Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Autoregressive</td>
</tr>
<tr>
<td></td>
<td>plus constant</td>
</tr>
<tr>
<td></td>
<td>trend</td>
</tr>
<tr>
<td></td>
<td>Integrated</td>
</tr>
<tr>
<td></td>
<td>autoregressive</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>469.9</td>
</tr>
<tr>
<td>2</td>
<td>447.8</td>
</tr>
<tr>
<td>3</td>
<td>427.8</td>
</tr>
<tr>
<td>4</td>
<td>422.1</td>
</tr>
<tr>
<td>5</td>
<td>426.1</td>
</tr>
<tr>
<td>6</td>
<td>433.1</td>
</tr>
<tr>
<td>7</td>
<td>441.9</td>
</tr>
<tr>
<td>8</td>
<td>448.1</td>
</tr>
<tr>
<td>9</td>
<td>453.8</td>
</tr>
<tr>
<td>10</td>
<td>457.7</td>
</tr>
<tr>
<td>11</td>
<td>460.0</td>
</tr>
<tr>
<td>12</td>
<td>457.3</td>
</tr>
<tr>
<td>13</td>
<td>455.5</td>
</tr>
<tr>
<td>14</td>
<td>454.8</td>
</tr>
<tr>
<td>15</td>
<td>457.4</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.11977</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.10839</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.10895</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.10525</td>
</tr>
<tr>
<td>$a_5$</td>
<td>0.10025</td>
</tr>
<tr>
<td>$a_6$</td>
<td>0.11203</td>
</tr>
<tr>
<td>$a_7$</td>
<td>0.10453</td>
</tr>
<tr>
<td>$a_8$</td>
<td>0.093896</td>
</tr>
</tbody>
</table>
Table IV-35. Identified Models for Band 8 of Combined Scene

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Selection Criterion</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive trend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>275.6</td>
<td>279.6</td>
<td>277.2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>271.9</td>
<td>274.6</td>
<td>276.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>270.7</td>
<td>273.9</td>
<td>275.8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>270.2</td>
<td>273.2</td>
<td>274.5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>268.9</td>
<td>272.7</td>
<td>273.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>267.4</td>
<td>271.3</td>
<td>272.7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>266.6</td>
<td>270.4</td>
<td>271.4</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>265.0</td>
<td>308.7</td>
<td>319.7</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>310.2</td>
<td>303.3</td>
<td>318.8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>310.3</td>
<td>308.7</td>
<td>319.4</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>309.7</td>
<td>309.5</td>
<td>319.4</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>310.0</td>
<td>309.6</td>
<td>318.4</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>309.7</td>
<td>308.3</td>
<td>316.7</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>308.4</td>
<td>306.7</td>
<td>315.1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>306.9</td>
<td>305.3</td>
<td>313.4</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>.074048</td>
</tr>
<tr>
<td>a₂</td>
<td>.054155</td>
</tr>
<tr>
<td>a₃</td>
<td>.10216</td>
</tr>
<tr>
<td>a₄</td>
<td>.058484</td>
</tr>
<tr>
<td>a₅</td>
<td>.080674</td>
</tr>
<tr>
<td>a₆</td>
<td>.090884</td>
</tr>
<tr>
<td>a₇</td>
<td>.083258</td>
</tr>
<tr>
<td>a₈</td>
<td>.11358</td>
</tr>
</tbody>
</table>
of the selected model is $9.58 \times 10^{-3}$ 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.

1) Band 9

The spectral intervals for this band are 1.959 - 2.386 $\mu$m for the wheat scene and 1.967 - 2.386 $\mu$m 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 $8.00 \times 10^{-4}$ 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 selection 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 $1.18 \times 10^{-2}$ 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

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>546.3</td>
<td>535.9</td>
<td>551.2</td>
</tr>
<tr>
<td>2</td>
<td>516.3</td>
<td>518.5</td>
<td>559.0</td>
</tr>
<tr>
<td>3</td>
<td>504.4</td>
<td>501.5</td>
<td>565.3</td>
</tr>
<tr>
<td>4</td>
<td>495.7</td>
<td>494.9</td>
<td>569.4</td>
</tr>
<tr>
<td>5</td>
<td>495.9</td>
<td>495.3</td>
<td>572.6</td>
</tr>
<tr>
<td>6</td>
<td>501.0</td>
<td>499.6</td>
<td>575.2</td>
</tr>
<tr>
<td>7</td>
<td>508.5</td>
<td>507.4</td>
<td>577.9</td>
</tr>
<tr>
<td>8</td>
<td>514.2</td>
<td>514.4</td>
<td>572.1</td>
</tr>
<tr>
<td>9</td>
<td>517.8</td>
<td>518.3</td>
<td>571.6</td>
</tr>
<tr>
<td>10</td>
<td>519.4</td>
<td>520.4</td>
<td>570.1</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>.10952</td>
</tr>
<tr>
<td>$a_2$</td>
<td>.096485</td>
</tr>
<tr>
<td>$a_3$</td>
<td>.10262</td>
</tr>
<tr>
<td>$a_4$</td>
<td>.11044</td>
</tr>
<tr>
<td>$a_5$</td>
<td>.11506</td>
</tr>
<tr>
<td>$a_6$</td>
<td>.11766</td>
</tr>
</tbody>
</table>
### Table IV-39. Identified Models for Band 9 of Combined Scene Selection Criterion

<table>
<thead>
<tr>
<th>Order of Model</th>
<th>Autoregressive</th>
<th>Autoregressive plus constant trend</th>
<th>Integrated autoregressive trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>356.1</td>
<td>355.3</td>
<td>356.5</td>
</tr>
<tr>
<td>2</td>
<td>348.7</td>
<td>348.3</td>
<td>355.1</td>
</tr>
<tr>
<td>3</td>
<td>346.5</td>
<td>346.5</td>
<td>354.0</td>
</tr>
<tr>
<td>4</td>
<td>345.4</td>
<td>345.6</td>
<td>353.0</td>
</tr>
<tr>
<td>5</td>
<td>344.0</td>
<td>344.0</td>
<td>351.5</td>
</tr>
<tr>
<td>6</td>
<td>342.3</td>
<td>342.5</td>
<td>349.9</td>
</tr>
<tr>
<td>7</td>
<td>341.1</td>
<td>341.2</td>
<td>352.5</td>
</tr>
<tr>
<td>8</td>
<td>342.1</td>
<td>341.3</td>
<td>351.2</td>
</tr>
<tr>
<td>9</td>
<td>341.4</td>
<td>340.5</td>
<td>350.1</td>
</tr>
<tr>
<td>10</td>
<td>340.3</td>
<td>339.0</td>
<td>348.5</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>.99759</td>
</tr>
</tbody>
</table>
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 validation 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 with 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

\[
|t(x)| = 1.431 \times 10^{-1}
\]  

(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, $n_0$, is dependent on the value of $n_1$ used for this test. In this study $n_1 = \frac{1}{10} N$ (N is the number of empirical data points) for all models tested. The critical values are listed for several values of $n_1$ in Table IV-41 [A5].

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

$$n(x) = 1.515 \times 10^1$$  \hspace{1cm} (4-3)

for $n_1 = 10$. 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

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>Critical Value, $n_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>21.6660</td>
</tr>
<tr>
<td>10</td>
<td>23.2093</td>
</tr>
<tr>
<td>11</td>
<td>24.7250</td>
</tr>
<tr>
<td>12</td>
<td>26.2170</td>
</tr>
<tr>
<td>13</td>
<td>27.6883</td>
</tr>
<tr>
<td>14</td>
<td>29.1413</td>
</tr>
<tr>
<td>15</td>
<td>30.5779</td>
</tr>
</tbody>
</table>
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

\[ |t(x)| = 2.184 \times 10^{-1} \]  

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

\[ n(x) = 7.434 \times 10^1 \]  

(4-5)

for \( n_1 = 9 \). Thus the model clearly fails the test. Other models with higher selection criterion similarly failed this test. The selected model has the test statistic

\[ n(x) = 2.009 \times 10^1 \]  

(4-6)

for \( n_1 = 10 \). 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

$$|t(x)| = 1.955$$ \hspace{1cm} (4-6)

Thus the selected model passes this test.

2) Serial Independance Test

The test statistic for the second order autoregressive model is

$$n(x) = 1.869 \times 10^{1}$$ \hspace{1cm} (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
\[ |t(x)| = 1.055 \quad (4-8) \]

Hence, the selected model easily passes this test.

5) Serial Independance Test

The selected model gives the test statistic

\[ \eta(x) = 1.365 \times 10^1 \quad (4-9) \]

for \( n_1 = 11 \). 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

\[ |t(x)| = 3.616 \times 10^{-1} \quad (4-10) \]

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
\[ n(x) = 1.153 \times 10^1 \quad (4-11) \]

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

\[ |t(x)| = 1.871 \times 10^{-1} \quad (4-12) \]

Thus, the selected model easily passes this test.

5) Serial Indepandance Test

The selected model gives the test statistic

\[ n(x) = 2.125 \times 10^1 \quad (4-13) \]

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

\[ |t(x)| = 1.753 \times 10^{-1} \quad \text{(4-14)} \]

Hence, the selected model easily passes this test.

2) Serial Independence Test

The selected model gives the test statistic

\[ n(x) = 2.00 \times 10^{1} \quad \text{(4-15)} \]

for \( n_1 = 12 \). 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

\[ |t(x)| = 1.598 \times 10^{-1} \quad \text{(4-16)} \]

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

\[ n(x) = 2.1250 \times 10^{1} \]  

(4-17)

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

\[ |t(x)| = 2.495 \times 10^{-1} \]  

(4-18)

Hence, the selected model easily passes this test.

2) Serial Correlation Test

The selected model yields the test statistic

\[ n(x) = 1.787 \times 10^{1} \]  

(4-19)

for \( n1 = 11 \). Therefore, the selected model passes 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

$$|t(x)| = 3.622 \times 10^{-2}.$$  \hspace{1cm} (4-20)

Thus, the selected model easily passes this test.

5) Serial Independence Test

The selected model gives the test statistic

$$n(x) = 1.628 \times 10^1$$ \hspace{1cm} (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, Fig. IV-28. Correlogram, Band 5, 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

\[ |t(x)| = 1.299 \times 10^{-1} \quad (4-22) \]

Thus, the selected model easily passes this test.

2) Serial Independence Test

The selected model gives the test statistic

\[ n(x) = 2.345 \times 10^1 \quad (4-23) \]

for \( n_1 = 11 \). 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

\[ |t(x)| = 2.861 \times 10^{-1} \quad (4-24) \]

Hence, the selected model easily passes this test.

5) Serial Independence Test

The selected model yields the test statistic

\[ n(x) = 8.322 \quad (4-25) \]

for \( n_1 = 11 \). Hence, the selected model also easily passes this test.
Fig. IV-31. Correlogram, Band 5, Combined Scene

Fig. 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.

6) 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

$$|t(x)| = 1.853 \quad (4-26)$$

Hence, the selected model easily passes this test.

2) Serial Independence Test

The test statistic for the selected model is

$$\eta(x) = 1.807 \times 10^1 \quad (4-27)$$

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

Fig. 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

Fig. IV-42. Cumulative Periodogram, Band 7, Combined Scene
4) Zero Mean Test

The selected model gives the test statistic

\[ |t(x)| = 1.208 \times 10^{-1} \]  \hspace{1cm} (4-28)

Thus, the selected model easily passes this test.

5) Serial Independence Test

The test statistic for the selected model is

\[ \eta(x) = 5.568 \]  \hspace{1cm} (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

\[ |t(x)| = 7.217 \times 10^{-1} \]  \hspace{1cm} (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, Fig. IV-46. Correlogram, Band 8, Band 8, Wheat Scene  Wheat Scene
2) Serial Independence Test

The selected model gives the test statistics

\[ n(x) = 2.516 \times 10^1 \]  

for \( n_1 = 13 \). 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

\[ |t(x)| = 1.033 \]  

Thus, the selected model easily passes this test.

5) Serial Independence Test

The test statistic for the selected model is

\[ n(x) = 3.141 \]  

for \( n_1 = 13 \). 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.

1) 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

\[ |t(x)| = 8.142 \times 10^{-1} \]  \hspace{1cm} (4-34)

Thus, the selected model easily passes this test.

2) Serial Independence Test

The test statistic for the selected model is

\[ \eta(x) = 9.915 \]  \hspace{1cm} (4-35)

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
Thus, the selected model easily passes this test.

5) Serial Independence Test

The selected model gives the test statistic

\[ \eta(x) = 1.362 \times 10^1 \]  \hspace{1cm} (4-37)

for \( n_1 = 15 \). 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

<table>
<thead>
<tr>
<th>Band</th>
<th>Wheat Scene</th>
<th>Combined Scene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AR(6)</td>
<td>IAR(11)</td>
</tr>
<tr>
<td>2</td>
<td>AR(2)</td>
<td>AR(2)</td>
</tr>
<tr>
<td>3</td>
<td>IAR(11)</td>
<td>IAR(11)</td>
</tr>
<tr>
<td>4</td>
<td>ARC(1)</td>
<td>ARC(1)</td>
</tr>
<tr>
<td>5</td>
<td>AR(1)</td>
<td>AR(3)</td>
</tr>
<tr>
<td>6</td>
<td>ARC(2)</td>
<td>AR(1)</td>
</tr>
<tr>
<td>7</td>
<td>IAR(5)</td>
<td>ARC(9)</td>
</tr>
<tr>
<td>8</td>
<td>IAR(8)</td>
<td>IAR(8)</td>
</tr>
<tr>
<td>9</td>
<td>IAR(6)</td>
<td>AR(1)</td>
</tr>
</tbody>
</table>
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
\[ y(k) = S(k) + n(k) \quad k \in [\lambda_1, \lambda_2] \]  

(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, \( \sigma_n^2 \). Since the noise disturbance is assumed to be of constant power spectral density level for each spectral band, considering several values of \( \sigma_n^2 \) 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 \( \sigma_n^2 \) 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 $\sigma_n^2$. 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

<table>
<thead>
<tr>
<th>Band</th>
<th>Noise Variance, $a_n^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>1</td>
<td>57.07</td>
</tr>
<tr>
<td>2</td>
<td>50.05</td>
</tr>
<tr>
<td>3</td>
<td>51.02</td>
</tr>
<tr>
<td>4</td>
<td>61.64</td>
</tr>
<tr>
<td>5</td>
<td>57.30</td>
</tr>
<tr>
<td>6</td>
<td>57.11</td>
</tr>
<tr>
<td>7</td>
<td>77.19</td>
</tr>
<tr>
<td>8</td>
<td>67.56</td>
</tr>
<tr>
<td>9</td>
<td>80.04</td>
</tr>
</tbody>
</table>

Note: The information values are given in nats here.

Table V-2. Average Information for Combined Scene Bands

<table>
<thead>
<tr>
<th>Band</th>
<th>Noise Variance, $a_n^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>1</td>
<td>56.19</td>
</tr>
<tr>
<td>2</td>
<td>53.51</td>
</tr>
<tr>
<td>3</td>
<td>52.73</td>
</tr>
<tr>
<td>4</td>
<td>61.49</td>
</tr>
<tr>
<td>5</td>
<td>56.36</td>
</tr>
<tr>
<td>6</td>
<td>56.21</td>
</tr>
<tr>
<td>7</td>
<td>80.48</td>
</tr>
<tr>
<td>8</td>
<td>69.93</td>
</tr>
<tr>
<td>9</td>
<td>79.93</td>
</tr>
</tbody>
</table>
Table V-3. Order of Preference of Spectral Bands for the Wheat and Combined Scenes

<table>
<thead>
<tr>
<th>Rank</th>
<th>Wheat Scene Band</th>
<th>Combined Scene Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

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 average 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 research 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 selection 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 techniques 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
SIBLIOGRAPHY


Kumar, R. and Silva, L. F., "Statistical 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.


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.


ORIGINAL PAGE IS OF POOR QUALITY


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 [53] and Meditch [41]. The algorithm is stated here in the manner of Sage and Melsa [53, Chapter 7].

The system model is given as

\[ x(k+1) = \phi(k+1, k)x(k) + r(k)w(k) \]  \hspace{1cm} (1)

and

\[ z(k) = H(k)x(k) + v(k) \]  \hspace{1cm} (2)

where

- \( x(k) \) is the state vector
- \( \phi(k+1, k) \) is the state transition matrix
- \( r(k) \) is a matrix
- \( w(k) \) is the driving noise vector
- \( z(k) \) is the observation vector
- \( H(k) \) is a matrix
- \( v(k) \) is the observation noise vector.

The assumed prior statistics are given as
\[ E[x(k)] = 0 = E[y(k)] \quad : \forall k \quad (\ast) \]

\[ E[x(0)] = \bar{x}(0) \quad (1) \]

\[ \text{cov}[u(k), u(j)] = \gamma(k) \delta(j-k) \quad (5) \]

\[ \text{cov}[v(k), v(j)] = \gamma(k) \delta(j-k) \quad (6) \]

\[ \text{cov}[u(k), v(j)] = \text{cov}[x(0), u(k)] \]

\[ = \text{cov}[x(0), v(k)] = 0 \quad (7) \]

where

\[ \delta(j-k) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \quad (9) \]

These assumptions give the Kalman filter algorithm for the estimate, \( \hat{x}(k) \), of \( x(k) \). The estimate is

\[ \hat{x}(k) = \phi(k, k-1) \hat{x}(k-1) + K(k) \left[ z(k) - H(k) \hat{x}(k; k-1) \hat{x}(k-1) \right] \quad (9) \]

where

\[ K(k) = \frac{\gamma_x(k/k-1)}{\hat{x}(k/k-1)} H^T(k) \left[ H(k) \gamma_x(k/k-1) H^T(k) + \gamma_v(k) \right]^{-1} \]

\[ = \frac{\gamma_x(k) H^T(k)}{\hat{x}(k)} \gamma_v^{-1}(k) \quad (10) \]
\[ V(k) = \left[ I - \frac{K(k)H(k)}{\lambda} \right] V(\lambda/k - 1) \]  

\[ V(k+1/k) = \frac{\phi(k+1,k) V(k) \phi^T(k+1,k)}{\lambda} + \frac{\Gamma(k) W(k) \Gamma^T(k)}{\lambda} \]  

\[ \hat{x}(k) = x(k) - \hat{x}(k) \]  

The ease with which this algorithm can be implemented on a digital computer is evident from the above equations. Sage and Molsa [SS, Chapter 7] give a good discussion of all the terms used above.
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 Malta [SJ, p. 499-500].

Matrix Inversion Lemma

If for any N x N nonsingular matrix \( A \) and any two N x N matrices \( B \) and \( C \), the two matrices \((A + BC^T)\) and \((I + C^TA^{-1}B)\) are nonsingular, then the matrix identity

\[
(A + BC^T)^{-1} = A^{-1} - A^{-1}B(I + C^TA^{-1}B)^{-1}C^TA^{-1} \tag{1}
\]

is valid.

Proof

Define

\[
D = A + BC^T \tag{2}
\]

Then since by assumption \( D \) is nonsingular, we can write

\[
D^{-1}D = I = D^{-1}A + D^{-1}BC^T \tag{3}
\]

Now postmultiply (3) by \( A^{-1} \) to obtain
\[ A^{-1} = D^{-1} + D^{-1}BCA^{-1} \] \hspace{1cm} (4)

Next postmultiply each side of (4) by \( D \) to obtain

\[ A^{-1}D = D^{-1}D + D^{-1}BCA^{-1}D \]

\[ = D^{-1}D(I + CTA^{-1}B) \] \hspace{1cm} (5)

Now by assumption \( I + CTA^{-1}B \) is nonsingular. Hence, we can write from (5)

\[ D^{-1}B = A^{-1}B(I + CTA^{-1}B)^{-1} \] \hspace{1cm} (6)

Postmultiply by \( CTA^{-1} \) to obtain

\[ D^{-1}BC = A^{-1}B(CTA^{-1}B)^{-1}CTA^{-1} \] \hspace{1cm} (7)

But it is seen from (4) that we can write

\[ A^{-1} - D^{-1} = D^{-1}BCA^{-1} \] \hspace{1cm} (8)

Hence using (3) in (7) we can write

\[ A^{-1} - D^{-1} = A^{-1}B(I + CTA^{-1}B)^{-1}CTA^{-1} \] \hspace{1cm} (9)

and using (2) we can finally write

\[ (A + BC)^{-1} = A^{-1} - A^{-1}B(I + CTA^{-1}B)^{-1}CTA^{-1} \] \hspace{1cm} (10)

This is the desired result.
C**..****. TO IDENTIFY "POOLS OF SPECTRAL HANDS BY MAXIMUM LIFELIKE**
C**..****. TECHNIQUE.
C
DIMENSION Y(1500), Z(1500), ZP(31+30), ZPT(31+30), ZPT(31+30),
L(31+30), TAP(30), Z(36+30), S(36+30), Z2(36+30), C2(36+30),
AX(1500), Y(1500).

C**..****. DATA INPUT
C
READ(*,1) N
1 FORMAT(I5)
READ(*,2) (I(I), I=1, N)
2 FORMAT(6F10.5)

C**..****. L1 MUST BE CHANGED TO CHANGE ORDER OF AR SYSTEM
C
L1=1
WRITE(16,71) L1
71 FORMAT(I5)
LI=0
C
C**..****. L1 IS THE NUMBER OF TRAVN TERMS
C
C..**..INITIALIZATION OF THETA AND S(I,J)
C
I=2
L=0
DO 70 I=1, 4
THETA(I)=.10
70 CONTINUE

C..**.. INITIALLIZATION COMPLETE
C
C.**.. DEINATION OF THETA ESTIMATE FOLLOWS
C
N=M+1
DO 13 I=1, M
THETA(I)=THETA(I)
13 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 40 I=1, M
SZ(I)=0.0
40 CONTINUE

C.**.. IF (Z(I,J)=.0) GO TO 41
C
IF (5*I,J)=.0 GO TO 41

C.**.. COMPUTATION FOLLOWS
C
DO 30 I=1, M
THETA(I)=.10
DO 30 J=1, M
Z(I,J)=Z(I,J)+ZP(I,J)*ZP(I,J)
30 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 40 I=1, M
SZ(I)=0.0
40 CONTINUE

C.**.. IF (Z(I,J)=.0) GO TO 41
C
IF (5*I,J)=.0 GO TO 41

C.**.. COMPUTATION FOLLOWS
C
DO 50 I=1, M
SZ(I)=0.0
50 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 40 I=1, M
SZ(I)=0.0
40 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 50 I=1, M
SZ(I)=0.0
50 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 40 I=1, M
SZ(I)=0.0
40 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 50 I=1, M
SZ(I)=0.0
50 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 40 I=1, M
SZ(I)=0.0
40 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 50 I=1, M
SZ(I)=0.0
50 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 40 I=1, M
SZ(I)=0.0
40 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 50 I=1, M
SZ(I)=0.0
50 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 40 I=1, M
SZ(I)=0.0
40 CONTINUE

C.**.. COMPUTATION FOLLOWS
C
DO 50 I=1, M
SZ(I)=0.0
50 CONTINUE

C.**.. COMPUTATION FOLLOWS
CONTINUE
THETA=0.0
DO 20 I=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
20 CONTINUE

DEMP=1.0+7TSZ
DO 43 J=1,4
ZTS(I,J)=ZTS(I,J)+1.0
GO TO 45
43 CONTINUE
21 CONTINUE

DO 60 I=1,4
TOP(I,J)=TOP(I,J)+1.0
GO TO 65
60 CONTINUE
22 CONTINUE

CONTINUE
THETA=0.0
DO 29 I=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
29 CONTINUE

CONTINUE
DO 58 J=1,4
IF (S(I,J) = 0.0) GO TO 67
58 CONTINUE

CONTINUE
DO 64 J=1,4
IF (S(I,J) = 0.0) GO TO 67
64 CONTINUE

CONTINUE
THETA=THE14ZTHETA {I) Z(I• 1)
THETA=THETA+2(Z(I,1)+Z(I,1)
50 CONTINUE

CONTINUE
DO 59 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
59 CONTINUE

CONTINUE
DO 65 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
65 CONTINUE

CONTINUE
DO 66 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
66 CONTINUE

CONTINUE
DO 67 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
67 CONTINUE

CONTINUE
DO 68 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
68 CONTINUE

CONTINUE
DO 69 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
69 CONTINUE

CONTINUE
DO 70 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
70 CONTINUE

CONTINUE
DO 71 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
71 CONTINUE

CONTINUE
DO 72 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
72 CONTINUE

CONTINUE
DO 73 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
73 CONTINUE

CONTINUE
DO 74 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
74 CONTINUE

CONTINUE
DO 75 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
75 CONTINUE

CONTINUE
DO 76 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
76 CONTINUE

CONTINUE
DO 77 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
77 CONTINUE

CONTINUE
DO 78 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
78 CONTINUE

CONTINUE
DO 79 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
79 CONTINUE

CONTINUE
DO 80 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
80 CONTINUE

CONTINUE
DO 81 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
81 CONTINUE

CONTINUE
DO 82 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
82 CONTINUE

CONTINUE
DO 83 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
83 CONTINUE

CONTINUE
DO 84 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
84 CONTINUE

CONTINUE
DO 85 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
85 CONTINUE

CONTINUE
DO 86 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
86 CONTINUE

CONTINUE
DO 87 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
87 CONTINUE

CONTINUE
DO 88 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
88 CONTINUE

CONTINUE
DO 89 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
89 CONTINUE

CONTINUE
DO 90 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
90 CONTINUE

CONTINUE
DO 91 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
91 CONTINUE

CONTINUE
DO 92 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
92 CONTINUE

CONTINUE
DO 93 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
93 CONTINUE

CONTINUE
DO 94 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
94 CONTINUE

CONTINUE
DO 95 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
95 CONTINUE

CONTINUE
DO 96 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
96 CONTINUE

CONTINUE
DO 97 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
97 CONTINUE

CONTINUE
DO 98 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
98 CONTINUE

CONTINUE
DO 99 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
99 CONTINUE

CONTINUE
DO 100 J=1,4
THETA=THETA+2(Z(I,1)+Z(I,1)
100 CONTINUE
PROGRAM TO PERFORM ZENO MEAN VALIDATION TEST.

DIMENSION X(1/300), Y(J/170), Z(J/170), U(J/170)
READ(X,1) Y
1 FOR I=1(J) 
  P=1(1) 
  DO I=1(J) 
  N=I 
  T=0.0 
  DO 2 N=1(J) 
  X=1(1) 
  T=0.0 
  2 CONTINUE 
  X(I)=X(I)/FLOAT(J) 
  T=0.0 
  DO 3 N=1(J) 
  X=1(1) 
  T=0.0 
  3 CONTINUE 
  R=.5'FLOAT(J) 
  T=0.0 
  DO 4 N=1(J) 
  X=1(1) 
  T=0.0 
  4 CONTINUE 
  X(I)=X(I)/FLOAT(J) 
  T=0.0 
  DO 5 N=1(J) 
  X=1(1) 
  T=0.0 
  5 CONTINUE 
  X(I)=X(I)/FLOAT(J) 
  T=0.0 
  DO 6 N=1(J) 
  X=1(1) 
  T=0.0 
  6 CONTINUE 
  X(I)=X(I)/FLOAT(J) 
  T=0.0 
  DO 7 N=1(J) 
  X=1(1) 
  T=0.0 
  7 CONTINUE 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 FORMAT(I6.0) 
  WRITE(1,2) J, T 
  2 Format...
**...PROGRAM TO PERFORM 1-HF SERIAL INDEPENDENCE VALIDATION TEST...**

```c
DIMENSION X(1400), Z(120), R(120)

1 FOR I=1(1) N
2 FOR J=1(1) I
3 F(I,J) = X(I,J) + X(J,I)
4 CONTINUE
5 CONTINUE
6 DO 4 I=1(N)
7 WRITE(14,7) I, F(I)
8 CONTINUE
9 FORMAT(5X,15,14,14,5X)
10 END
```

---

**...PROGRAM TO COMPUTE 1-HF SERIAL INDEPENDENCE...**

```c
DIMENSION X(1-20), Y(1-20), Z(1-20)
1 DO 10 I=1(N)
2 X(I) = 1.0 + I
3 Y(I) = 2.0 + I
4 Z(I) = 3.0 + I
5 DO 20 J=1(N)
6 F(I,J) = X(I) + Y(J) + Z(K)
7 CONTINUE
8 PRINT 10, I, F(I)
9 10 FORMAT(5X,15,14,14,5X)
11 END
```
****PROGRAM TO COMPUTE AVERAGE (MUTUAL) INFORMATION****

DIMENSION A(I,J),X(I,J)
REAL X00,X10,X01,X11,XX(I,J),YY(I,J)

OPEN (5,UNIT=10,FILE='IPL1974.COM')

DO 10 J=1,20
10 CONTINUE

OPEN (5,UNIT=11,FILE='IPL1974.COM')

READ (5,30) X00,X10,X01,X11

CONTINUE

WRITE (6,40) X00,X10,X01,X11

STOP

30 FORMAT (4F20.8)
40 FORMAT (4F20.8)
50 FORMAT (1X,15X,5X,15X,15X,2X,15X)
60 FORMAT (1X,15X,5X,15X,15X,4X,15X,15X)
70 FORMAT (1X,15X,5X,15X,15X,6X,15X,15X)
80 FORMAT (1X,15X,5X,15X,15X,7X,15X,15X)
90 FORMAT (1X,15X,5X,15X,15X,8X,15X,15X)
100 FORMAT (1X,15X,5X,15X,15X,9X,15X,15X)
110 FORMAT (1X,15X,5X,15X,15X,10X,15X,15X)
120 FORMAT (1X,15X,5X,15X,15X,11X,15X,15X)
130 FORMAT (1X,15X,5X,15X,15X,12X,15X,15X)
140 FORMAT (1X,15X,5X,15X,15X,13X,15X,15X)
150 FORMAT (1X,15X,5X,15X,15X,14X,15X,15X)
160 FORMAT (1X,15X,5X,15X,15X,15X,15X,15X,15X)

CALL X00

STOP

RETURN

END
DO 0, J=1,10
*END OF SYSTEM PARAMETERS

C.....INITIALIZATION OF FILTER PARAMETERS FOR COMPUTATION OF ENTROPY

DO 13 I=1,10

13 CONTINUE
CALL MITTPPCHT(I,J)
CALL TTMTPCHT(I,J)
DO 12 J=1,11

12 CONTINUE

IF (J .GT. 11) GO TO 62
00 00
DO 20 J=1,10

20 CONTINUE
*END OF SYSTEM PARAMETERS
ORIGINAL PAGE IS OF POOR QUALITY
CALL A(I,J) = A(I,J) + A(J,I)

DO 29 I = 1, N
   DO J = 1, J-1
      PA(I,J) = PA(I,J) + A(I,J) - A(J,I)
   END
29 CONTINUE

CALL A(I,J) = A(I,J) - PA(I,J)

DO 30 I = 1, N
   DO J = 1, J-1
      PA(I,J) = PA(I,J) - A(I,J) + A(J,I)
   END
30 CONTINUE

DO 30 I = 1, N
   DO J = 1, J-1
      A(I,J) = PA(I,J)
   END
30 CONTINUE

C...END OF MUTUAL INFORM. COMPUTATION

C...MUTUAL INFORMATION COMPUTATION FOLLOWS

C...MUTUAL INFORMATION COMPLETE

DO 201 K = 1, M
   WRITE(6,52) K, AVGINF(1,K), AVGINF(2,K), AVGINF(3,K), AVGINF(4,K), AVGINF(5,K)
52 FORMAT(1,5SF12.5)
END
SUBROUTINE ATANH(Z, A, C)
A = ATANH(Z, A, C)
END

SUBROUTINE ARCSIN(Z, A, C)
C = ARCSIN(Z, A, C)
END

SUBROUTINE ARCCOS(Z, A, C)
C = ARCCOS(Z, A, C)
END