12-9-2006

# Hyperspectral Dimensionality Reduction via Sequential Parametric Projection Pursuits for Automated Invasive Species Target Recognition

Terrance Roshad West

# HYPERSPECTRAL DIMENSIONALITY REDUCTION VIA SEQUENTIAL PARAMETRIC PROJECTION PURSUITS FOR AUTOMATED INVASIVE SPECIES TARGET RECOGNITION

By

Terrance R. West

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Electrical Engineering
in the Department of Electrical and Computer Engineering

Mississippi State University

December 2006

HYPERSPECTRAL DIMENSIONALITY REDUCTION VIA SEQUENTIAL PARAMETRIC PROJECTION PURSUITS FOR AUTOMATED INVASIVE SPECIES TARGET RECOGNITION

By

Terrance Roshad West

Approved:

| | |
|---|---|
| Lori Mann Bruce<br>Associate Professor of Electrical and Computer Engineering<br>(Director of Thesis) | Noel Schulz<br>Associate Professor of Electrical and Computer Engineering<br>(Committee Member) |
| Jenny Du<br>Assistant Professor of Electrical and Computer Engineering<br>(Committee Member) | Nicholas H. Younan<br>Professor of Electrical and Engineering<br>(Graduate Program Director) |

Roger King
Associate Dean of Bagley College of Engineering

Name: Terrance Roshad West

Date of Degree: December 8, 2006

Institution: Mississippi State University

Major Field: Electrical and Computer Engineering

Major Professor: Dr. Lori Mann Bruce

Title of Study: HYPERSPECTRAL DIMENSIONALITY REDUCTION VIA SEQUENTIAL PARAMETRIC PROJECTION PURSUITS FOR AUTOMATED INVASIVE SPECIES TARGET RECOGNITION

Pages in Study: 71

Candidate for Degree of Master of Science

This thesis investigates the use of sequential parametric projection pursuits (SPPP) for hyperspectral dimensionality reduction and invasive species target recognition. The SPPP method is implemented in a top-down fashion, where hyperspectral bands are used to form an increasing number of smaller groups, with each group being projected onto a subspace of dimensionality one. The Bhattacharyya distance is used as the SPPP performance index. The performance of the SPPP method is compared to two other currently used dimensionality reduction techniques, namely best spectral band selection (BSBS) and best wavelet coefficient selection (BWCS). The ATR system is tested on two invasive species hyperspectral datasets: a terrestrial case study of Cogongrass versus Johnsongrass and an aquatic case study of Waterhyacinth versus American Lotus. For both case studies, the SPPP approach either outperforms or

performs on par with the BSBS and BWCS methods in terms of classification accuracy; however, the SPPP approach requires significantly less computational time.

# DEDICATION

This thesis is dedicated to my mother, who never obtained a college degree and who has always had faith in me when no one else did or when others tried to bring be down. I also dedicate this thesis to minorities in mathematics, science, and engineering, who wish to obtain a Masters degree in their fields. This is a testament from God that it can be done. **"**Look Onward and Upward Toward the Light"

ACKNOWLEDGEMENTS

I would like to thank Dr. Lori Mann Bruce for being an excellent academic advisor through my graduate Masters program. I would like to thank Dr. Lori Mann Bruce and Dr. Noel Schulz for words of encouragement throughout this process. I would like to thank Dr. Jenny Du and Dr. Noel Schulz for serving on my committee. I would like to thank my family, friends, and fraternity brothers for their support throughout this time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

TABLE OF ACRONYMS

| | |
|---|---|
| ASDAnalytical | Spectral Device |
| AVIRIS | Airborne Visible-Infrared Imagery Spectrometer |
| ATR | Automated Target Recognition |
| BD | Bhattacharyya Distance |
| BSBS | Best Spectral Band Selection |
| BWCS | Best Wavelet Coefficient Selection |
| CASI | Compact Airborne Spectrographic Imager |
| DAFE | Discriminant Analysis Feature Extraction |
| DWT | Discrete Wavelet Transform |
| FWHM | Full Width Half Maximum |
| HYDICE | Hyperspectral Digital Imagery Collection Experiment |
| IR | Near-Infrared |
| LD | AFisher's Linear Discriminant Analysis |
| NM | Nearest Mean Classifier |
| PCA | Principal Component Analysis |
| ROC | Receiver Operating Characteristics |
| $S_B$ | Between Class Covariance |
| SPPP | Sequential Parametric Projection Pursuits |
| SW | Within Class Covariance |

# CHAPTER I

# INTRODUCTION

## 1.1 Hyperspectral Image Analysis

A variety of imagery can be used for remote sensing applications, including the following imaging modalities: panchromatic, multispectral, hyperspectral, radar, and lidar. From these modalities, hyperspectral imagery arguably has the most potential for subpixel target recognition. This potential stems from the nature of hyperspectral imagery having many spectral features per pixel. And as hyperspectral sensor technology advances, the number of spectral features per pixel increases. In theory, this increase in the number of features can improve subpixel target detection accuracy. In practice, however, this may not be true due to three main factors.

The first factor deals with the bandwidth and computational time requirements for processing the hyperspectral data. With the increase in spectral bands, these requirements for per-pixel analysis can become inordinate and preclude practical applications. For this reason, it is typically desirable for the hyperspectral data to be preprocessed such that the dimensionality of the data is significantly reduced.

The second factor is related to the training of automated target recognition (ATR) systems. When designing and utilizing supervised ATR systems, there exists a direct relationship between the number of potential features and the amount of

necessary training data. If the number of features is excessively increased without appropriately increasing the amount of training data, the ATR system can easily become overtrained. If the amount of training data remains constant, as the number of features increases, the target detection accuracy will increase to a critical point where, thereafter, the classification accuracy decreases. This trend is known by various names, including the "curse of dimensionality" and "Hughes phenomenon" [1,2].

The third and final factor deals with the limitations of not having the ability to employ necessary statistical methods on the data. This is evident in statistical methods which use the computation of the inverse of a covariance matrix, as is the case with Fisher's linear discriminant analysis and maximum likelihood classifiers. With these methods, when the numbers of features excessively increase with a limited number of training samples, the feature covariance matrices can become sparse, and their inverses may not be computable. As a result, these statistical methods, which are commonly found in supervised ATR systems, can fail.

These three factors necessitate the use of dimensionality reduction methodologies for hyperspectral data. These methodologies can be utilized as a preprocessing stage or as a feature extraction/optimization stage within the ATR system.

**1.2 Invasive Species**

An "invasive species" is defined as a non-native species whose introduction causes or is likely to cause harm to the environment, human health, or economy [3]. Over the years, remote sensing has become an important tool for the detection and mapping of invasive vegetation. Remote sensing technologies, such as aerial photography and multispectral digital imagery, have been used successfully to detect dense weed infestations when there is a unique spectral pattern expressed by the weed [4]. Due to the availability of hyperspectral data, new methods of detecting invasive species are being established everyday. The hyperspectral imagery may allow for the early detection of incipient infestations and increased detection accuracy via subpixel ATR systems [4]. In particular, hyperspectral imaging could prove beneficial to the early detection of two example invasive species, Cogongrass (*Imperata cylindrica*) and Waterhyacinth (*Eichhornia crassipes*). Both of these invasive species are becoming widespread in the southeastern United States. Cogongrass is an agressive noxious weed that forms thick clumps releasing toxins that smother out native plants. Cogongrass has invaded extensive acreage of roadways, pasturelands, and forests, and as a result, is causing significant ecologic and economic damage. For example, Cogongrass can quickly colonize in the open areas left after natural forests have been cleared, making it very difficult to be reforested or be converted to other agricultural uses. In addition, the Cogongrass infestations significantly damage the wildlife habitat and serve as wildfire fuel. Waterhyacinth is a floating aquatic invasive native to South America and has invaded

many lakes, ponds, canals, and rivers in the United States. Waterhyacinth can create dense mats of floating vegetation, causing considerable economic and ecologic damage. A few ways in which Waterhyacinth causes economic damage includes obstructing waterway navigation, blocking drainage which can cause flooding or prevent subsidence of floodwaters, clogging irrigation pumps, intensifying mosquito problems, and blocking access to recreational areas and decreasing waterfront property values. Clearly, it would be beneficial to use a cost effective method, such as remote sensing, to detect and monitor the spread of these types of invasive species.

### 1.3 Contributions of this Thesis

In this thesis, a method known as projection pursuits will be applied to hyperspectral data. Projection pursuits will be used as an automated means of determining optimum dimensionality reduction of the hyperspectral signatures, where the goal is detection of a target invasive species. The projection pursuits method will utilize a bank of four potential linear projections: averaging, Gaussian weighted averaging, principal component transform, and Fisher's discriminant transform. An automated band grouping system will be used to determine an optimum partitioning of the hyperspectral signature, where each partition or group of spectral bands will then be projected onto a lower dimensional subspace using one of the potential projections. Bhattacharyya distance will be used as a performance metric in the optimization process. The overall projection pursuits method will be incorporated into an automated target recognition system that uses a nearest mean

classifier, and the system will be applied to hyperspectral data of invasive species. Two case studies will be conducted: Cogongrass versus Johnsongrass and Waterhyacinth versus American Lotus, representing realistic scenarios where the target invasive could be misidentified as a similar non-invasive vegetation. The results of the projection pursuits approach will be compared to two currently popular dimensionality reduction methods, best spectral band selection and best wavelet coefficient selection.

The final outcome of the thesis will be a determination of how projection pursuits compares to the other dimensionality reduction methods in terms of accuracy and computation expense, as well as an in depth analysis of the results of the projection pursuits method, including an analysis of how the hyperspectral signal is partitioned, which potential projections are most often selected, and how much the supervised projections improves the performance.

# CHAPTER II

# BACKGROUND AND LITERATURE REVIEW

## 2.1 Invasive Species

### 2.1.1 Cogongrass

Cogongrass (*Imperata Cylindrica*) is an invasive species which originated from South Asia, the Philippines, and Japan [5]. Cogongrass is currently present in the south and southeastern sections of the United States. This plant has the potential to grow from 2 to 4 feet tall and have leaves which are 1 ½ inches wide. Cogongrass tends to grow in dense colonies in open fields, along roadsides, and in forests that are not fully canopied. They are considered an ecological threat because they have the capability of causing more frequent fires in a fire-driven ecosystem when the density of the colony is high [5]. Cogongrass has the ability to choke out the native plant life which in turn affects the habitat of insects, mammals, and birds. Furthermore, Cogongrass has the ability to decrease pine growth in a pine plantation if the infestation is very dense, thus reducing the amount of pine that can be harvested [6]. Weeds, such as Cogongrass, cause an overall reduction of 12 % in crop yield in US

agriculture [7]. This 12 % in reduction represents an annual crop production lost of $ 32 billion dollars. Cogongrass and Johnsongrass (*sorghum halepense*) were the two vegetations compared in this experiment. These two plant species were chosen for comparison for three reasons: (i)throughout the south and southeastern United States both species are commonly found adjacent to one another, (ii) they have spectral similarities, and (iii) they have physiologic similarities, which is shown in the Figure 2.1 and Figure 2.2.

(a) (b)

Figure 2.1 Cogongrass and Johnsongrass pictures. (a) Cogongrass image taken in southern Mississippi, U.S.A. (b) Johnsongrass image in northern Mississippi, U.S.A.

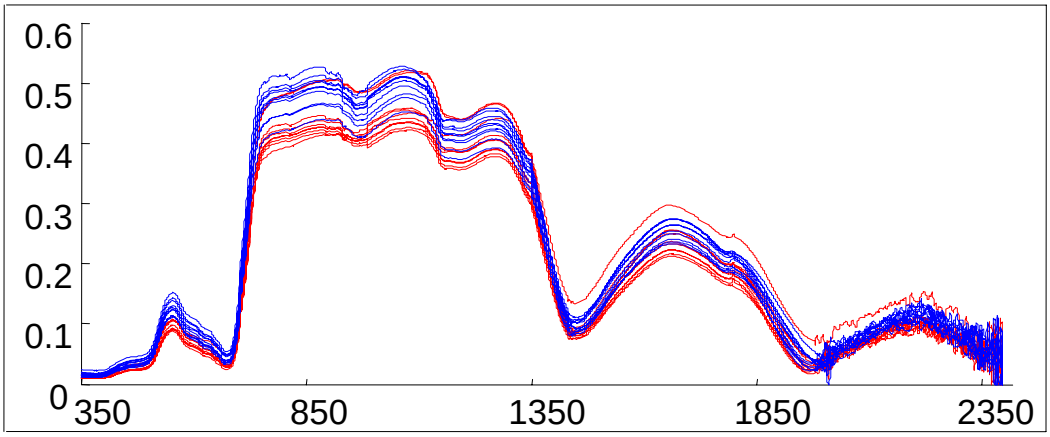


Figure 2.2 Ten randomly selected hyperspectral signatures of Cogongrass (red) and Johnsongrass (blue)

### 2.1.2 Waterhyacinth

Aquatic invasive species, such as Waterhyacinth (*Eichhornia Crassipes*), have the capability of reducing the growth of native plant life and altering the communities of fish and other wildlife [7]. Waterhyacinth is an invasive species which originated from South America. An annual cost of more than $100 million is invested in the United States for the control of nonindigenous aquatic weed species [7]. This invasive species is a floating tropical species which has the means of overtaking a body of water in 30 days. This characteristic prevents native aquatic plants from producing oxygen which makes the body of water uninhabitable [7]. This aquatic plant is known for hindering commercial and recreational traffic through waterways and blocking ports with its dense mats [8]. When large quantities of Waterhyacinth are present in flowing water, it has the ability to cause damage to transportation infrastructures. The weight combined with the current of the body of water has the capabilities of causing costly damage and repairs to structures in its

path [8]. Waterhyacinth and American lotus (*Nelumbo lutea*) were the two aquatic invasive species compared in this experiment. These two plant species were chosen for comparison for three reasons: they have spectral similarities, they have physiologic similarities and they are found throughout the south and southern regions of the United States, which are shown in the Figure 2.3.


(a) (b)

Figure 2.3 Waterhyacinth and American Lotus pictures taken in Starkville, Mississippi, U.S.A. in 2005. These vegetation were grown in tank systems in the plant and soil sciences research facility on the North Farm of Mississippi State University. (a) Waterhyacinth and (b) American Lotus.

## 2.2 Hyperspectral Sensors and Imaging

"Hyperspectral sensors (sometimes referred to as imaging spectrometers) are instruments that acquire images in many, very narrow, and contiguous spectral bands throughout the visible, near-infrared (IR), mid-IR, and thermal IR portions of the spectrum" [9]. Hyperspectral data can contain 10's, 100's or even 1000's of spectral bands of information. When the hyperspectral data is reported in terms of

reflectance, each band contains the percentage of reflected light for a narrow range of wavelengths across the electromagnetic spectrum. Hyperspectral data has the potential for being particularly useful in target detection applications. Figure 2.4 displays a hyperspectral signal for one sample of the invasive species Cogongrass (*Imperata cylindrica*).

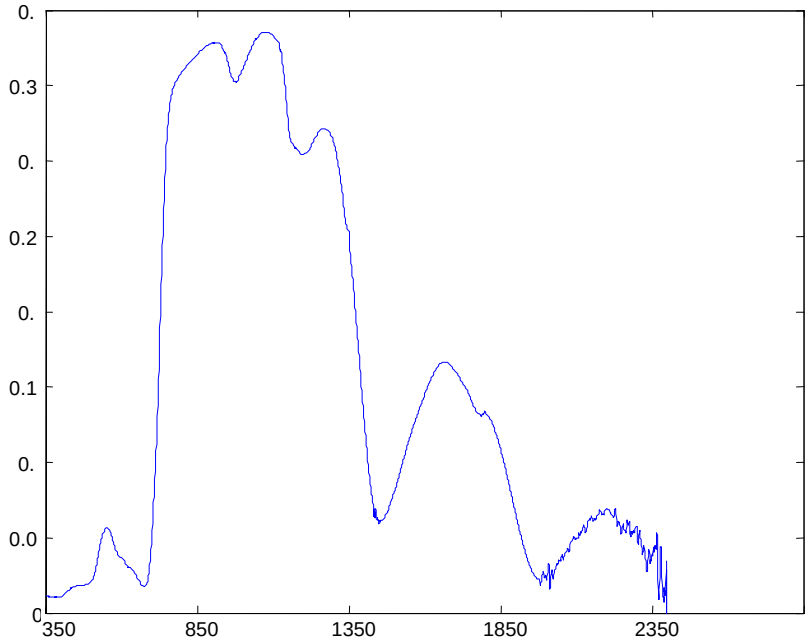


Figure 2.4 Example hyperspectral signature of Cogongrass.

There are three typical methods in which hyperspectral data can be obtained: airborne sensors, spaceborne sensors, or handheld sensors. Airborne sensors are mounted onboard planes, helicopters, unmanned aerial vehicles or other aircraft. Spaceborne sensors are mounted onboard satellites and are orbited around earth.

Handheld sensors are typically small and lightweight, where the instrument can be hand carried into the field for data collection.

Compact Airborne Spectrographic Imager (CASI) and Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) are examples of two types of airborne hyperspectral sensors. CASI has been in operation since 1989 and has 288 spectral bands. These bands are contained in the electromagnetic spectrum range between 0.40 and 1.0 μm in 1.8 nm spectral intervals. This system has a field of view of about 37.8 degrees [9]. AVIRIS is a hyperspectral sensor which has 224 bands with approximately 9.6 nm full width half maximum (FWHM). These 224 bands span the electromagnetic spectrum in the interval between .40 and 2.45 μm. This sensor is capable of having a ground pixel resolution of approximately 20 m and an across-track scanner swath width of about 10 km [9].

The Hyperion instrument is an example of a spaceborne sensor and was launched by NASA. Hyperion is a hyperspectral imager which has 220 spectral bands [10]. These bands are contained in the electromagnetic spectrum interval between 0.4 to 2.5 μm. Hyperion has a 30 meter spatial resolution and has the capability of obtaining a 7.5 km by 100 km land area per image [10].

The Analytical Spectral Devices (ASD$^{TM}$) Fieldspec Pro handheld spectroradiometer is an example of a handheld instrument which can obtain hyperspectral data [11]. The ASD has a spectral range of 350 – 2500 nm, spectral resolution of 3 nm @ 700 nm and 10 nm @ 1400/2100 nm, and uses a single 512 element silicon photodiode array for sampling 350 - 1000 nm and two separate,

graded index Indium-Gallium-Arsenide photodiodes for the 1000 - 2500 nm range [11]. As a result, an ASD hyperspectral signal contains an astounding 2151 spectral bands. The spatial resolution of the ASD is dependent on its height above the ground when it is in operation. A great advantage of the ASD is its capability of collecting hyperspectral signatures of individual targets (or endmembers) without mixing of neighboring materials. For example, the ASD can be held over a target vegetation, and a hyperspectral signature can be obtained for that exact target without mixing from other vegetation, soil, etc. This can be very useful when conducting feasibility studies such as those found in this thesis.

### 2.3 Hyperspectral Dimensionality Reduction

Hyperspectral data is by definition high dimensional data. However, much of this data is redundant, and the samples of hyperspectral data often reside in a space that is mostly empty. The redundancy is due to the spectral correlation between adjacent spectral bands. The spectral correlation present in hyperspectral data allows for the data to be projected on to a lower dimensional subspace without loss of significant information. The goal of this projection is to simultaneously remove redundancy while retaining pertinent information that can be used to discriminate between various classes of data, i.e. target hyperspectral signals vs. nontarget hyperspectral signals.

The ability to detect different objects and phenomena while reducing the dimensionality of a dataset is a task that has not been perfected. In current studies,

different methods have been employed in the reduction of dimensionality. Fisher's linear discriminant analysis (LDA), principal component analysis (PCA), and discrete wavelet transform (DWT) techniques are dimensionality reduction methods that are often utilized in remotely sensed hyperspectral imagery.

LDA is a dimensionality reduction technique which projects the input data onto a subspace in which the data has maximum class separation. This class separation is based on the minimization of the within class covariance ($\mathbf{S_W}$) and maximization of the between class covariance ($\mathbf{S_B}$). Since class-specific covariance are necessary, this method falls into the category of supervised dimensionality reduction techniques. The ratio of the $\mathbf{S_B}$ and $\mathbf{S_W}$ are used in calculating the transformation matrix for the LDA algorithm [12]. This transformation matrix is constructed by performing eigen-analysis which involves taking the inverse of the $\mathbf{S_W}$. A problem with this analysis arises when there are too many features with too few training vectors, and then the $\mathbf{S_W}$ may become sparse. The sparseness causes $\mathbf{S_W}$ to become ill-conditioned and can inhibit the ability to calculate its inverse. The inability to calculate the covariance matrix inverse makes this dimensionality reduction method problematic in many operational scenarios.

A common way of using LDA for dimensionality reduction is a method known by several names, including stepwise LDA, discriminant analysis feature extraction (DAFE), best spectral band selection (BSBS) and spectral greedy searches. All of these methods use the same basic approach. For example, the BSBS method has two primary phases: forward selection and backward rejection. During forward

selection, a subset of spectral bands is constructed in the following way. (i) Spectral bands are sorted according to their individual performance of class separation. (ii) The spectral band that provides maximum class separation is used as a seed to start the running subset. (iii) The spectral band with the next highest class separation is then added to the running subset to form a temporary subset. (iv) The temporary subset is then projected using LDA and tested to determine if it provides an increase in the class separation. If it does provide an increase, the temporary subset becomes the running subset. If it does not provide an increase, the running subset remains unchanged. (v) Steps iii-iv are repeated until all spectral bands are considered for addition to the running subset. Next, backward rejection is conducted in the following way. (vi) The first spectral band in the running subset is removed to form a temporary subset. (vii) The temporary subset is then projected using LDA and tested to determine if it provides an increase in the class separation. If it does provide an increase, the temporary subset becomes the running subset. If it does not provide an increase, the running subset remains unchanged. (viii) Steps vi-vii are repeated until all spectral bands in the running subset are considered for removal. (ix) The final running subset is then considered to be the "best" group of spectral bands and is projected using LDA.

Dimensionality reduction can also be performed by employing PCA. This method involves computing the covariance of the whole dataset by using eigen-analysis. The eigen-values and eigen-vectors are then used to construct the transformation matrix [13]. Dimensionality reduction is achieved by truncating the

transformation matrix, such that only the first few principal components are included. The number of principal components included in the matrix determines the dimensionality of the projected data. Although PCA is a leading method in dimensionality reduction, it has been shown that PCA is not an adequate method for feature extraction [14]. Cheriyadat and Bruce proved in their study that PCA is not a sufficient method for dimensionality reduction when feature extraction is used for classification or target detection [14]. The primary reason is that PCA is an unsupervised dimensionality reduction technique. In some operational scenarios, no class-specific training data is available, and unsupervised methods must be used. However, if class-specific data is available, the work by Cheriyadat and Bruce shows that PCA should be avoided and supervised methods, like LDA, should be utilized.

Another method that has been applied in hyperspectral dimensionality reduction is the DWT. The DWT decomposes the signal into detail and approximation coefficients using scaled and translated versions of the mother wavelet. The coefficients themselves or combinations of the coefficients are commonly used as features. Bruce *et al.* investigated the use of the receiver operating characteristics (ROC) to best select the coefficients for dimensionality reduction and feature extraction [15]. The method, called best wavelet coefficient selection (BWCS) is similar to BSBS described above. It contains the same forward selection and backward rejection phases. The main difference is that the selection process is conducted on the wavelet coefficients rather than the original spectral bands.

A more recent method of dimensionality reduction applied to hyperspectral signals is projection pursuits. The principal objective of projection pursuits is to overcome the "curse of dimensionality" while at the same time retaining information within the hyperspectral signal that is pertinent to target detection and classification. The importance of this objective is illustrated by the limitation in which humans can interpret data in higher dimensional space. A dataset which exceeds a dimension of three or more is difficult for humans to process and comprehend. A dataset which is composed of dimensions of three and less gives humans not just the analytical skills but also a visual means to interpret the data.

The idea of projection pursuits was coined in 1974 by Friedman and Tukey [16]. The technique used by Friedman and Tukey reduced the dimensionality of a large multivariate dataset by performing an orthogonal projection on the dataset. The driving force of their experiment was to formulate a low dimensional projection that reduced the dimensions of a point cloud by maximizing a function known as the projection index which was a combination of a trimmed standard deviation and a weighted count of the number of close pairs. Their implementation proved to be the birth of the projection pursuits approach. The projection index is a robust component in the projection pursuits scheme. This component provides a quantitative metric that can be used to determine the degree to which the dimension of the data will be reduced.

The idea of performing orthogonal projections on multispectral and hyperspectral data, such as projection pursuits, is not a new concept. However, it is

not nearly as commonly used as methods like PCA and LDA. The method of projection pursuits has been applied to a few types of multispectral and hyperspectral applications.

Jimenez and Landgrebe evaluated parallel and sequential projection pursuits methods in high dimensional feature reduction using AVIRIS hyperspectral data [17]. In their study, they seek to distinguish between two crops, corn and soybean, which was acquired from the Indiana's Pine test site. The parallel projection pursuits approach involves projecting each adjacent spectral band. The Bhattacharyya distance (BD), which is the projection index, is then applied to the features. The sequential approach takes into account all the adjacent bands by projecting the adjacent bands and only varying one band projection. In the sequential approach, there is a global maximization of the projection index and is not based a local maximization of the projection index. The two projection pursuits methods were compared against linear discriminant analysis. The two approaches proved to out perform the discriminant analysis based on the resulting classification accuracies and BDs.

Another study in which projection pursuits was used as a dimensionality reduction tool was performed by Ifarraguerri and Chang [18]. In their study, the hyperspectral imagery was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. From this data collection, a 256 x 256 section of the image which contained vehicles, roads, trees, and other features was analyzed. This study evaluated the use of projection pursuits in the analysis of hyperspectral images

in an unsupervised method. In this study, the authors sought to find a projection which locates low probability targets that have significant spatial resolution in a hyperspectral image. The projection pursuits method was performed by applying PCA to the area of interest. Next, the elements which had the largest eigen-values were obtained and then placed in the transformation matrix. Finally, the data was transformed and the information divergence index was applied, which is described in [18], to produce the pixel with the highest value of the projection index. The projection pursuits methods in this study proved that with the information divergence index the dimensionality of the hyperspectral image could be reduced while retaining the important characteristics of the image.

Lin and Bruce evaluated the use of projection pursuits for dimensionality reduction using hyperspectral data for applications involving agricultural target recognition [19]. The data was obtained by a handheld spectroradiometer which collected 2000 spectral bands in the range of 350 and 2350 nm of two vegetation species. The targets in their experiment were sicklepod and cocklebur which are species of weeds found amongst variety agricultural crops. In their study, parallel parametric projection pursuits, projection pursuits best band selection, and sequential parametric projection pursuits methods were used in reducing the dimensions of the hyperspectral data to allow for efficient target recognition. The two projection indexes used in this research were BD and the area under receiver operating characteristics curves [19]. The weights for the transformation matrix consisted of a vector that averaged the bands in a group, a vector that chose only one spectral band,

and a vector that maximized the performance metric. The projection pursuits preprocessing methods employed in their study proved to have higher classification accuracies than data that were not preprocessed with the projection pursuits.

From these previous studies on projection pursuits, we can see that the method certainly holds promise as a means of hyperspectral dimensionality reduction for invasive species detection applications.

# CHAPTER III

# METHODOLOGIES

## 3.1 Dimensionality Reduction

Three types of dimensionality reduction methods are investigated in this thesis. The primary method is projection pursuits, which is described in detail below. Two comparison methods are also investigated. These are the BSBS and BWCS methods, which are described in detail in Chapter 2.

### 3.1.1 Projection Pursuits

The mathematical form of projection pursuit is given by

$$Y = A^T X_j$$

where $\mathbf{X}_j$ is the original hyperspectral data ($d$ x $M$) of class $j$, **A** is the transformation matrix ($d$ x $n$), and **Y** is the projected data ($n$ x $M$). The $d$ in the original data is the number of spectral bands, $M$ is the number samples being projected, and $n$ is the dimension of the subspace in which the data will be projected. The columns of the transformation matrix **A** are orthogonal, which means that the dot product of any two columns would equal zero. The nonzero elements in each column of the transformation matrix contain the weights of the potential projections. The transformation matrix **A** is optimized by the projection index which is usually

expressed as $\mathbf{I}(\mathbf{A}^{T}\mathbf{X})$. In a two class problem, the optimization can be based on the classification separation of the two classes. The optimization is performed by transforming the data into a lower dimension subspace and applying the projection index . Projections are chosen or rejected based on the projection index measurement. Thus, the selection of the correct projection index will significantly affect how effective the projection pursuit method will perform.

#### 3.1.1.1 Transformation Matrix A

The dimension of the projection matrix is determined by the number of spectral bands or features within the training samples and the dimensionality of the projection subspace. The columns of the transformation matrix are orthogonal and contain zero and nonzero elements. Each column of the matrix **A** are zero for every entry except for the position of the adjacent bands. The structure of the transformation matrix **A** is shown in Figure 3.1.

$$A=\begin{bmatrix} A_{1,1} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & & & & & & & & \vdots & \vdots \\ A_{1,n_1} & 0 & & & & & & & & \vdots & \vdots \\ 0 & A_{2,1} & & & & & & & & \vdots & \vdots \\ \vdots & \vdots & & & & & & & & 0 & \vdots \\ \vdots & A_{2,n_2} & & & & & & & & A_{k-1,1} & \vdots \\ \vdots & 0 & & & & & & & & \vdots & \vdots \\ \vdots & \vdots & & & & & & & & A_{k-1,n_{k-1}} & 0 \\ \vdots & \vdots & & & & & & & & 0 & A_{k,1} \\ \vdots & \vdots & & & & & & & & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & A_{k,n_k} \end{bmatrix}$$

Figure 3.1. SPPP transformation matrix formulization

The $n_j$ denotes the number of bands in the $j^{th}$ partition or group and the $k$ denotes the total number of groups.

The construction of the sequential parametric projection pursuits (SPPP) transformation matrix is conducted as follows:

1. An initial matrix $A'$ is constructed, for which each adjacent band $\mathbf{A}_{k\text{-}1,1} - \mathbf{A}_{k\text{-}1,n}$ is chosen.
2. Using the initial matrix $A'$, the high dimensional data is transformed and an initial global BD is computed, $B'$.
3. While maintaining all $\mathbf{A}_{k,n}$'s constant, the $\mathbf{A}_{1,1} - \mathbf{A}_{1,n}$ group is split in half and the high dimensional data set is transformed and the global BD is computed, $\mathbf{B}_1$.

4. Step 3 is repeated for each value of $k$, until each group of adjacent bands has been spilt and a Bhattacharyya distance $B_i$ is calculated for $i$ =1:$k$ creating a BD vector $\vec{B}$.
5. The maximum value in vector $\vec{B}$ is obtained, $B_m$. Holding all groups constants, the $m^{th}$ group is split in the initial transform $A'$ and the matrix is stored.
6. The BD $B_m$ is compared to $B'$ and if $B_m > B'$ then $B' = B_m$ and steps 2 – 6 are repeated until the global Bhattacharyya index stops increasing or the number of groups in the transformation matrix exceeds the initial size of the groups (i.e. the minimum number of training samples per class).

Figure 3.2 shows the block diagram of the SPPP method which describes the implementation of the SPPP technique. Figure 3.3 shows the SPPP group splitting diagram which illustrates how the groups are split during optimization of the transformation.

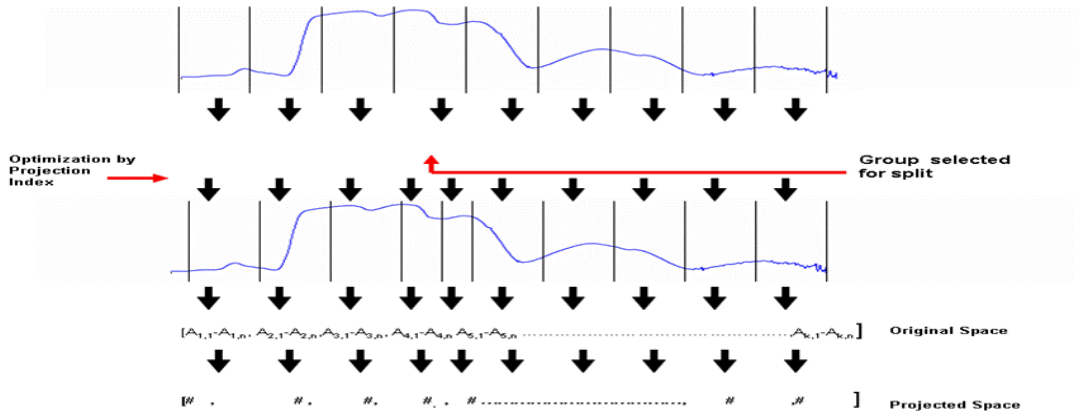


Figure 3.2 Sequential Parametric Projection Pursuits Group Splitting Diagram

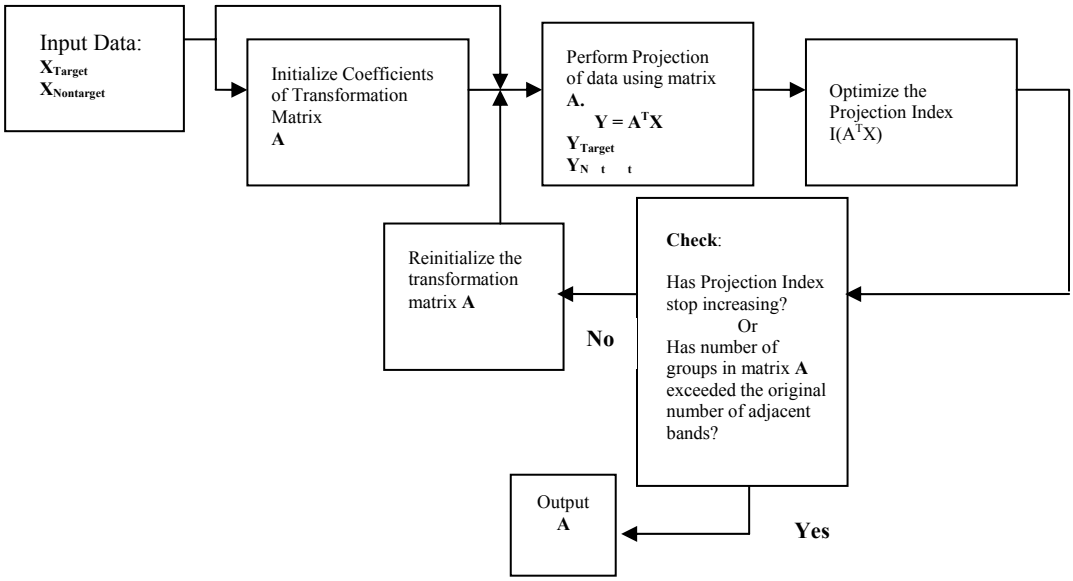


Figure 3.3. Block Diagram for Sequential Parametric Projection Pursuits

### 3.1.1.2 Constructing Vectors for SPPP Matrix A

In the section above, the overall approach for construction for the SPPP transformation matrix is defined. Each column of the matrix represents a projection of a group of adjacent spectral bands. The values of the non-zero elements in each column are the weights of that particular projection. In this study, these projections are selected from a bank of four potential projections. Three of the projects are unsupervised. They include a simple average, a Gaussian weighted average, and PCA. One potential supervised projection is included in the bank, and it is Fisher's LDA.

The first unsupervised projection is the averaging vector. This vector is the average of all the elements in the group of adjacent spectral bands in the original data space. The vector is

$$A_{kn} = \frac{1}{n_k}\left[a_{k1}, a_{k2}, \ldots, a_{kn_k}\right]$$

where $k$ is the $k^{th}$ group of adjacent bands, and $n_k$ is the number of the adjacent bands. The second unsupervised projection is a Gaussian-weighted average. The vector is the element by element multiply of the group of adjacent bands with a Gaussian window vector having a length that is equal to the number of adjacent bands in each group. Suppose that the $i^{th}$ group of adjacent bands have $n_i$ samples and is represented by vector $\vec{x} = [x_{1,\ldots\ldots,} x_{n_i}]$. Also, suppose that the vector $\vec{g} = [w_{1,\ldots\ldots,} w_{n_i}]$ contains the coefficients of an $n$-point Gaussian window with standard deviation of σ. The Gaussian guess vector is obtain by an array multiplication of the vectors $\vec{g}$ and $\vec{x}$.

$$A_{kn} = \vec{g} .* \vec{x} = \left\lfloor w_{k1} \times a_{k1}, w_{k2} \times a_{k2}, \ldots, w_{kn_k} \times a_{kn_k} \right\rfloor$$

The third and final unsupervised initial vector guess is a principal component vector. Unlike the two previous unsupervised methods, this method employs applying eigen-analysis to the set of adjacent bands. In this method, the eigenvectors are used as weights to transform the data. The eigenvalues and eigenvectors are computed for the adjacent bands. Next, the eigenvectors are placed in decreasing order based on the value of the eigenvectors. The eigenvector with the highest eigenvalue is then chosen as the weights for the adjacent bands. For the PCA projection, only the first principal component is utilized, so the resulting dimensionality of the lower subspace is one.

The supervised method that governs the vectors weights for the matrix **A** is the Fisher's LDA. This method is classified as a supervised method because of the *a priori* knowledge that must be known about the data's classification before the weights are calculated. The weights were based on the class separation of the adjacent bands of the two classes. This class separation is based on the minimization of the with-in class variance ($\mathbf{S_W}$) and maximization of the between class variance ($\mathbf{S_B}$). The ratio of the $S_B$ and $S_W$ are used in calculating the transformation matrix for the LDA algorithm. This transformation matrix is constructed by performing eigen-analysis which involves taking the inverse of the with-in class variance matrix. Finally, the LDA weights are calculated and these weights are used as the vector weights. LDA reduces the dimensionality to $c$-1, where $c$ is the number of classes. Since both case studies in this thesis are two class problems (Cogongrass vs. Johnsongrass and Waterhyacinth vs. American Lotus), the resulting dimensionality of the lower subspace is one.

#### 3.1.1.3 Bhattacharyya Distance

The BD is used as the performance index in the SPPP approach. It is a special form of the Chernoff distance. The Chernoff distance seeks to find the upper bounds of the error of probability by finding the value of s which produces the maximum value for $\mu(s)$ [20]. The Chernoff distance for a 2 class problem is define as the following:

$$\mu(s) = \frac{s(1-s)}{2}(M_2 - M_1)^T[s\Sigma_1 + (1-s)\Sigma_2]^{-1}(M_2 - M_1) + \frac{1}{2}\ln\frac{\left|s\Sigma_1 + (1-s)\Sigma_2\right|}{\left|\Sigma_1\right|^s\left|\Sigma_2\right|^{1-s}}$$

where $M_i$ is the mean of class *i* and the $\mathbf{\Sigma}_i$ is the covariance for class *i*.

The BD is formulated by selecting a specific *s* for the Chernoff distance. This selection of *s* is not the optimum value for *s*, however this selection of *s* allows for a less problematical upper bound. The value for which the BD is converted to from the Chernoff distance is ½. The BD for a 2 class problem is then defined as the following:

$$\mu(1/2) = \frac{1}{8}(M_2 - M_1)^T\left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1}(M_2 - M_1) + \frac{1}{2}\ln\frac{\left|\frac{\Sigma_1 + \Sigma_2}{2}\right|}{\sqrt{\left|\Sigma_1\right|\left|\Sigma_2\right|}}$$

The BD consists of two terms which governs its operation. The first term represents the difference of mean of the class separation and the second term represents the difference of covariance of the class separation [20]. This relationship of the mean and covariance of the two terms can be revealed by setting the mean of both classes equal to each other for the first term and the covariance of the both classes equal to each other for the second term.

## 3.2 Testing and Evaluation

### 3.2.1 Classification

The feature vectors are the driving force of any classifier. The principal objective of any classifier is to use the feature vectors to assign the data in question

to the correct class. The performance of the classifier will be optimum if the feature extraction method is chosen in such a manner in which the data which has very similar characteristics are classified together and the data which has dissimilar characteristics are not classified together.

Similar to nearest neighbor classifier, the Euclidean distance defines how the test samples are classified in the nearest mean classification scheme. The nearest mean algorithm uses the Euclidean distance between the test samples and the mean of each class to determine how the test samples are classified. The mathematical expression below expresses how this distance is calculated [1].

$$D_j = \sqrt{\sum_{n=1}^{d} \left(\vec{x}(n) - \vec{\mu}_j(n)\right)^2}$$

The $d$ is the dimension of the training samples. The $x(n)$ represents the $n^{th}$ component test sample and the $\mu_j(n)$ represents the $n^{th}$ component of the mean vector of the $j^{th}$ class.

### 3.2.2 N-Fold Cross Validation

The types of testing methods that can be used for validation are dependent on the amount of data that is available for a target recognition system. In the case in which the there is a small amount of data, the *N*-fold cross-validation approach is one of the common methods used [21]. The different testing methods are obtained by varying *N*. Leave-one-out cross-validation is obtained when k is equal to one. In this method, a set of *N* training instances is repeatedly divided into a training set of size

*N*-1 and a test set of size 1[21]. The partitioning of the data is repeated until each sample in the dataset has been designated as the test sample. Jackknifing cross-validation is obtained using the *N*-fold cross-validation approach when *N* is equal to two. In this method, the data is divided in half, where half is used for training and the other half is used for validation. The method used in this research is a combination of both jackknifing and leave-one-out cross-validation. The data is jackknifed during training of the dimensionality reduction methods. Then, a leave-one-out cross-validation is applied to the overall ATR system that includes both the dimensionality reduction stage and the classification stage. The leave-one-out cross-validation of the overall ATR system is applied to the testing half of the data (i.e. not the training half of the data used to train the dimensionality reduction methods).

### 3.3 Case Study Data

Pure endmember hyperspectral signatures were collected for this study using a handheld ASD spectroradiometer. Signatures were collected for Cogongrass and Johnsongrass (*sorghum halepense*) for the first case study and for Waterhyacinth and American Lotus (*Nelumbo lutea*) for the second case study.

Johnsongrass and American Lotus were chosen as the non-target vegetation for three primary reasons. In each of the case studies, (i) the target and non-target vegetation are often found adjacent to one another in the same habitats throughout the south and southeastern U.S.; (ii) the target and non-target vegetation have physiologic similarities which could cause them to be easily confused with one

another; and (iii) the target and non-target vegetation have spectral similarities which make them particularly challenging cases. Thus, both case studies are practical and challenging problems for validating the proposed SPPP method.

The hyperspectral signatures collected in this experiment were collected in good weather conditions in Mississippi, U.S.A., in 2000-2004 with the fiber optic sensor held NADIR at approximately shoulder height 4 feet above ground. A $25^{o}$ instantaneous field of view (IFOV) foreoptic was used, and the ASD unit was set to average ten signatures to produce each sample signature.

For this study, 260 samples were used for evaluation, 125 samples of Cogongrass and 125 samples of Johnsongrass. For each class of vegetation, 100 signatures were used to train the SPPP and classifiers, while the 25 remaining signatures were used to test the system. The Waterhyacinth and American Lotus dataset is multitemporal. ASD hyperspectral signatures were collected each week for a total of 16 weeks. The data was collected over the summer of 2005, starting in June and ending in October. For this study, 600 samples were used for evaluation, 300 samples of Waterhyacinth and 300 samples of American Lotus. For each class of vegetation, 200 signatures were used to train the SPPP and classifers, while the 100 remaining signatures were used to test the system, for the combined dataset. The division of the sample dataset for data organized my months is shown in Table 3.1.

Table 3.1 Division of sample dataset for Waterhyacinth and American Lotus data organized by months

| Months | Training | Testing | Total |
|---|---|---|---|
| June/July | 35 | 35 | 70 |
| August | 40 | 40 | 80 |
| September | 40 | 40 | 70 |
| October | 35 | 35 | 80 |

# CHAPTER IV

# RESULTS AND DISCUSSION

## 4.1 Case Study I – Johnsongrass versus Cogongrass

### 4.1.1 Limited Number of Groups Based on Amount of Training Data

For this thesis, the proposed projection pursuits method of dimensionality reduction, namely SPPP, is compared with two methods found in the current remote sensing literature, namely BWCS and BSBS methods. The first case study is the Cogongrass vs. Johnsongrass dataset, where a limitation is set on how large the number of groups can grow based on the amount of training data available for each class. Also, the number of groups is only allowed to increase if the BD is also increasing. The classification accuracies for all three methods were above 90 %. As one can observe from Figure 4.1, the BWCS method outperformed both the BSBS and SPPP dimensionality reduction methods in terms of overall classification accuracy. Although both the BWCS and BSBS methods had higher classification accuracies, the computation times were higher than that of the SPPP technique, which is shown in Tables 4.1 and 4.2.

Table 4.3 and Figure 4.2 show the classification accuracies for the SPPP approach when varying the size of the starting groups. That is, when initializing the SPPP, the hyperspectral signature is first partitioned into $N$ adjacent, equally sized

groups. Then each of the groups has the potential to be split into multiple smaller groups as the SPPP method proceeds. The goal in varying the starting group size is to determine if this parameter significantly affected the overall performance of the SPPP method. The overall accuracy was not significantly affected by the starting group size. The overall classification accuracies for all group sizes were above 94 %. The producers and users classification accuracies for all group sizes were above 92 %. The smaller group size resulted in a slight increase in accuracy. As shown in Figure 4.2, the confidence intervals for each group's accuracy place each group's accuracy in the same range. That is, the confidence intervals overlap, so we cannot say that any particular starting group size outperforms any other starting group size. Also, it should also be noted that the overall computation time is less when the group size is smaller, since less splitting is required in the construction of the transformation matrix **A**. So, it is recommended for this case study that the SPPP be initialized with smaller group sizes.

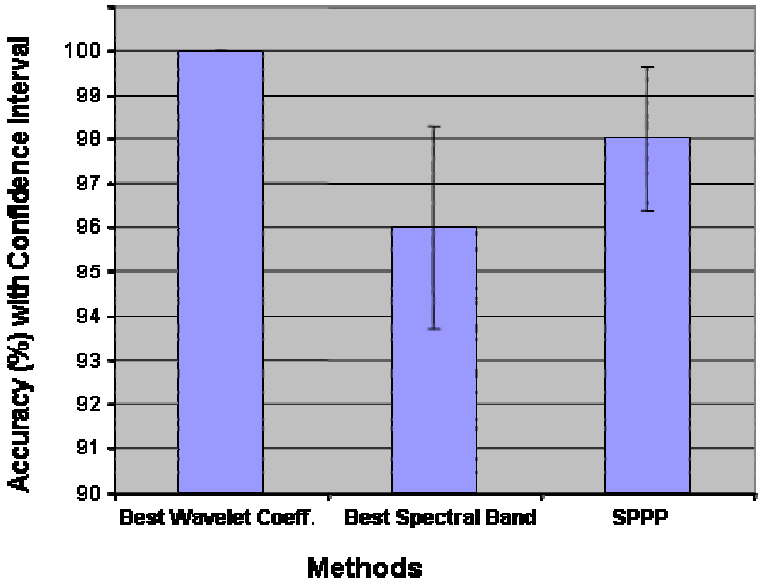


Figure 4.1 Classification accuracies for Cogongrass and Johnsongrass dataset for the dimensionality reduction methods of Best Wavelet Coefficients, Best Spectral Bands, and SPPP with the NM classifier with limitation on the number of groups and the BD

The groups selected for investigation included groups 50, 60, 65, and 75. The selection of these group sizes were governed by the stability of the SPPP method. This stability was based on the supervised potential projection. At certain group sizes, the supervised potential projection could not be obtained, because the inverse of the covariance matrix could not be computed.

Table 4.1 Computational time for the different group sizes for the SPPP technique

| Group Size | Computational Time (Minutes) |
|---|---|
| **50** | 1.4 |
| **60** | 3.4 |
| **65** | 1.3 |
| **75** | 7.4 |

Table 4.2 Computational time for the different Comparison methods

| Methods | Computational Time (Minutes) |
|---|---|
| **Best Wavelet Coefficients** | 45.7 |
| **Best Band Selection** | 66.3 |

Table 4.3 Producers, Users, and Overall Accuracies different starting group sizes for the Cogongrass and Johnsongrass dataset when the SPPP dimensionality reduction method being applied using all projection with the NM classifier

| | **Producers Accuracy** | | **Users Accuracy** | | |
|---|---|---|---|---|---|
| **Group Size** | **Cogongrass** | **Johnsongrass** | **Cogongrass** | **Johnsongrass** | **Overall Accuracy** |
| **50** | 96 | 100 | 100 | 96 | 98 |
| **60** | 96 | 96 | 96 | 96 | 96 |
| **65** | 92 | 96 | 96 | 92 | 94 |
| **75** | 96 | 96 | 96 | 96 | 96 |

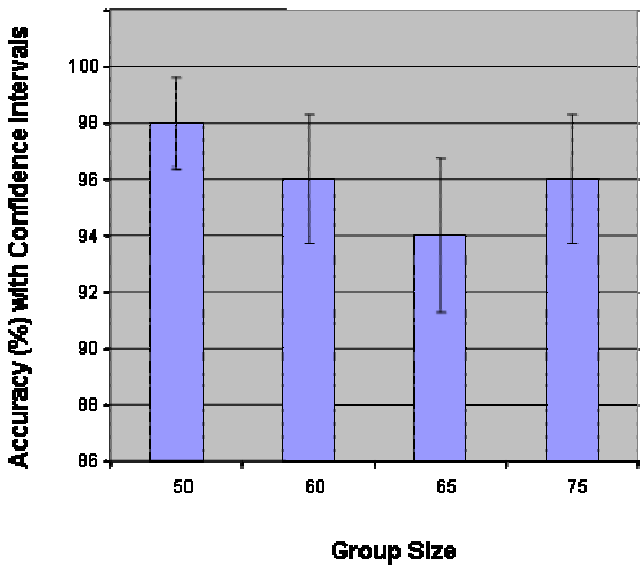


Figure 4.2 Classification accuracies for Cogongrass and Johnsongrass dataset for the dimensionality reduction method SPPP for group sizes 50, 60, 65, 75 with the NM classifier. Error bars indicate a 95% confidence interval.

Figure 4.3 shows a plot of BD versus number of group splits, i.e. a progress of the BD as the SPPP method progresses. The BD for all group sizes starts below 120. For the larger starting group sizes, the number of splits increases as well as the final BD. On one hand, this result is not surprising since SPPP method is designed to stop the splitting of groups when the number of groups reaches its limiting factor, generally specified to be the minimum number of training samples per class. So allowing the SPPP method to initialize with large groups and intelligently split the groups and select projections would be expected to provide improved results. On the other hand, this result is surprising since this plot conflicts with the Figure 4.2. Based on the plot in Figure 4.3, one would expect the scenario where we start with fewer, larger groups to result in significantly higher classification accuracies. But

that is not the case in this study. This indicates that maybe the NM classifier is not capitalizing on the lower subspace data generated by the SPPP method.

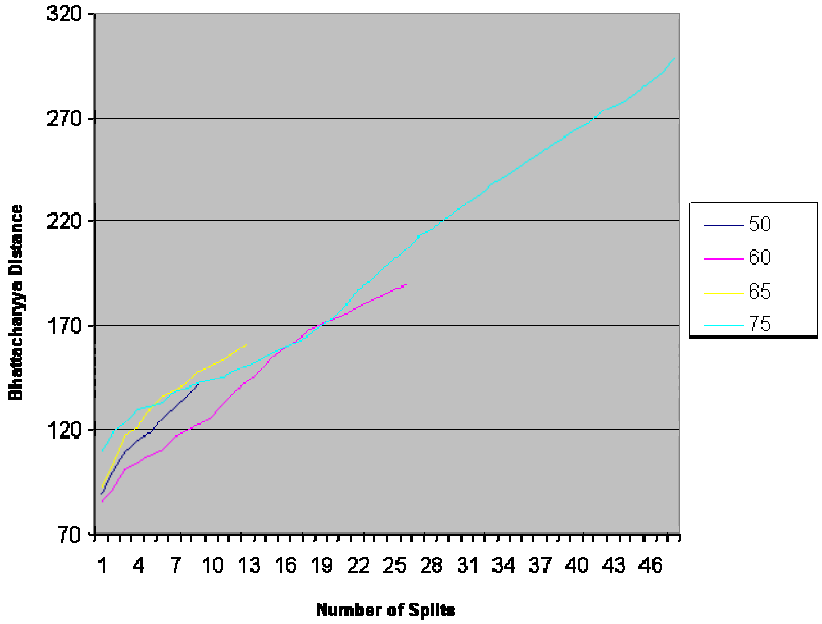


Figure 4.3 Plot of BD for each group size versus the number of times a group was split

Figure 4.4 shows a barchart of the number of times each potential projection is selected for the four cases of varying the starting group sizes. All four potential projections were selected for each case. This is an interesting result because one would expect that the supervised potential projection, LDA, would be selected every time. The second most selected potential projection was the unsupervised Gaussian-weighted average. This result is unexpected because one would not expect a simple Gaussian-weighted average potential projection to be selected as many or more times than the PCA potential projection. Another interesting result is that the starting group size of 65 results in fewer final groups than a starting group size of 60. One

would expect that the larger group sizes would have more potential projections because of its ability to have more splits.

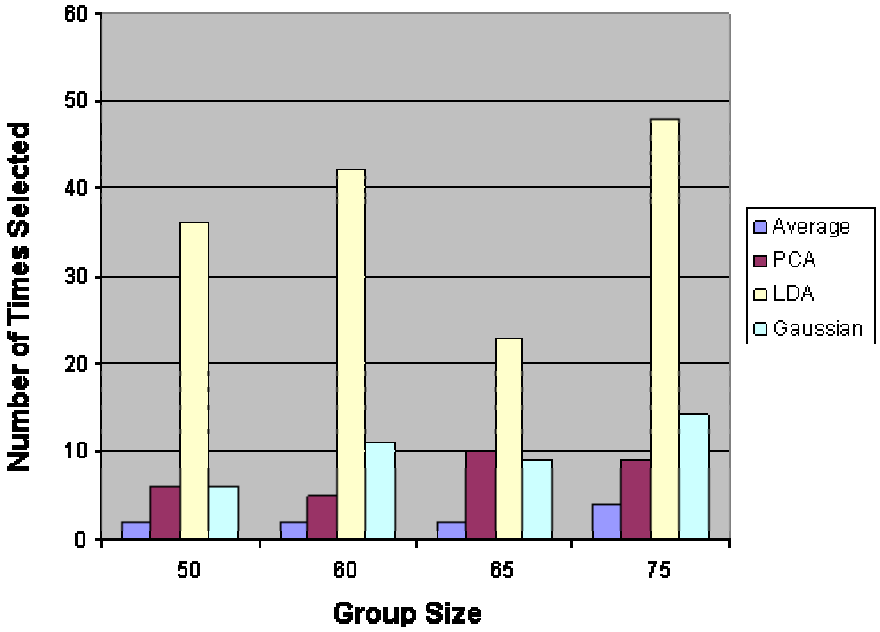


Figure 4.4 Bar graph of the number of times a potential projection was chosen for the different group sizes when applying SPPP to Cogongrass-Johnsongrass dataset

Figure 4.5 illustrates the spectral location of the final set of groups of projections when applying the SPPP method. The distributions of the groups are found throughout the spectrum range with different group sizes. One can observe that there exist small concentrations of groups in the spectral regions of 1100 to 1150, 1900 to1925, and 2150 to 2250. This result implies that these spectral regions have significant information which is essential in discriminating between the two targets of Cogongrass and Johnsongrass.

Figures 4.6 through 4.9 illustrate the progression of the SPPP groups and their optimum projections, starting with the initial group size and proceeding through the group splitting until the SPPP method terminates. The color coded projection map describes which potential projections were selected during the splitting of groups. The different potential projections are represented by four different colors. The region that is dark blue represents the area beyond the number of groups for that iteration of the SPPP method. The expansion of the number of groups is governed by the initial group size. As one can observe from Figures 4.6 to 4.9, the LDA potential projection is selected more often in the initial guess than that of the other projections. This result is as expected because the LDA potential projection is a supervised weighted vector which is constructed by having some *a priori* knowledge of the dataset. The color coded maps indicate that more unsupervised potential projections are selected during the splitting of the groups. The results from Figures 4.6 to 4.9 show that the unsupervised potential projection serves an important role in some spectral bands when using hyperspectral data to distinguish between Cogongrass and Johnsongrass. The last row in each color code projection map shows which final projections selected for each group.

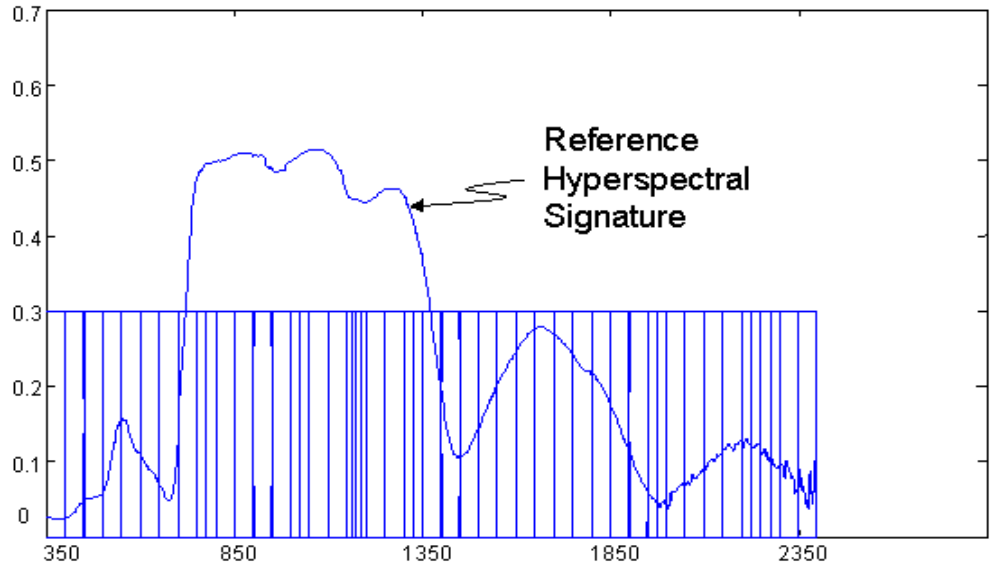


Figure 4.5 Spectral location of groups (final number of groups = 50) for final set of projections when applying SPPP to Cogongrass-Johnsongrass dataset; locations are plotted against a reference signature

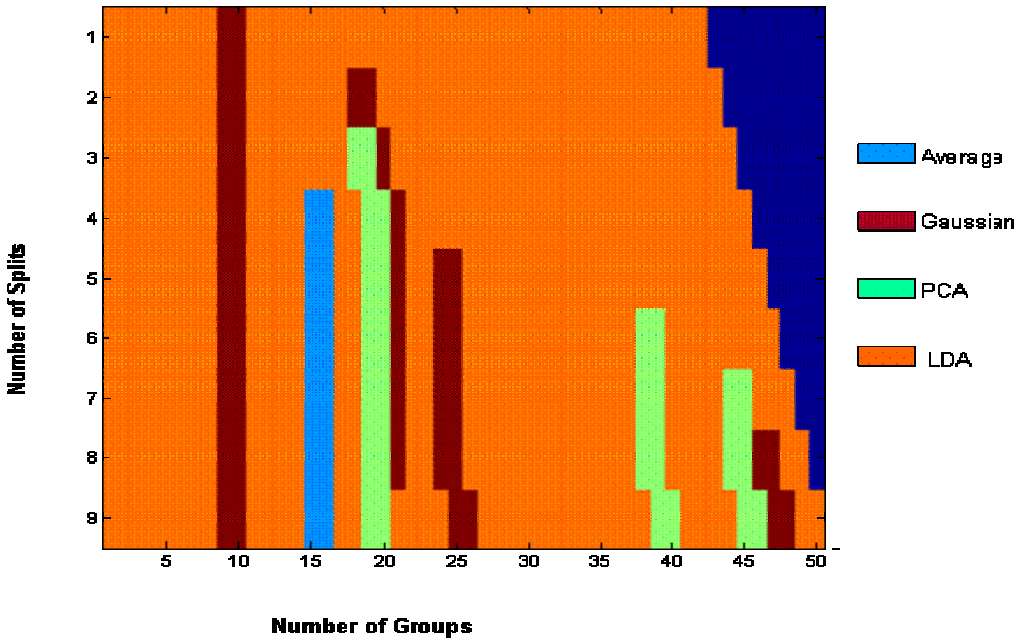


Figure 4.6 Color coded projection map for initial group size of 50 when applying SPPP to Cogongrass-Johnsongrass dataset

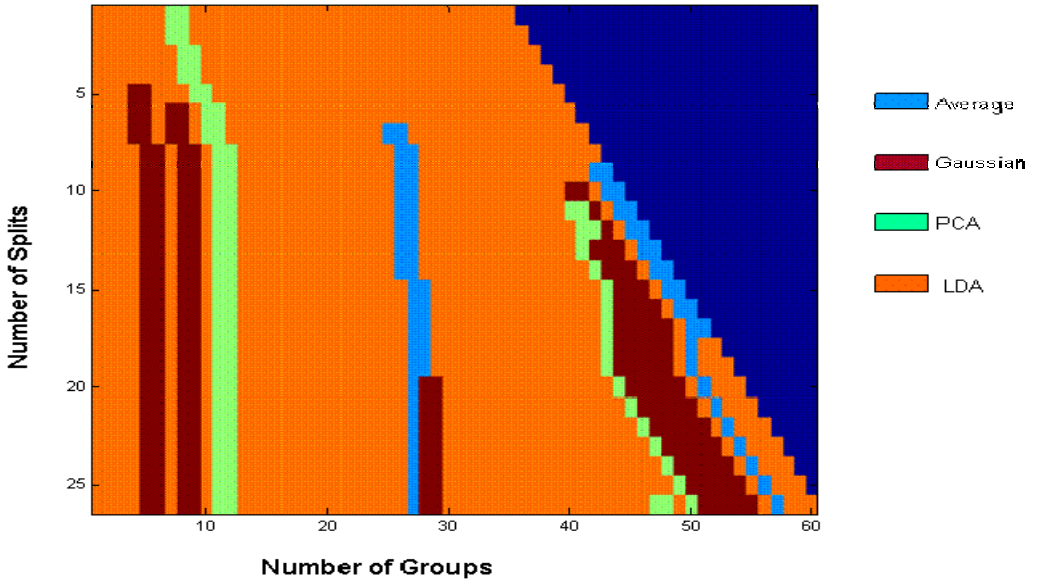


Figure 4.7 Color coded projection map for initial group size of 60 when applying SPPP to Cogongrass-Johnsongrass dataset

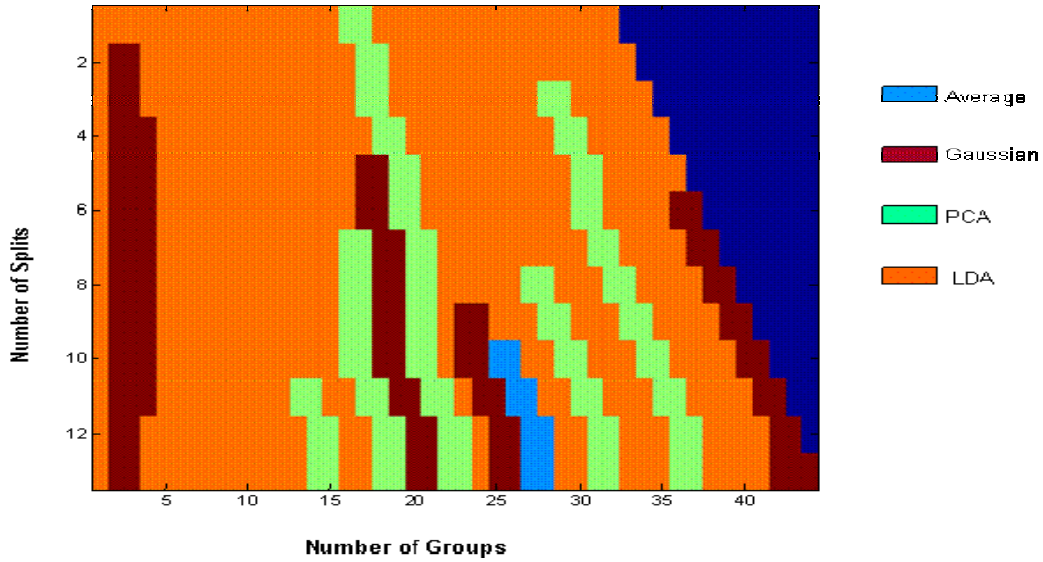


Figure 4.8 Color coded projection map for initial group size of 65 when applying SPPP to Cogongrass-Johnsongrass dataset

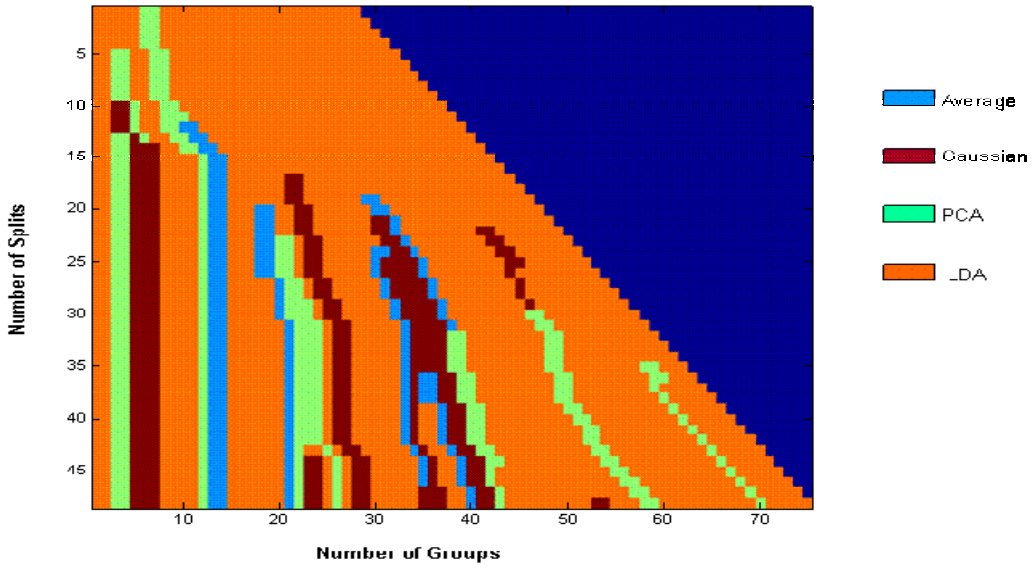


Figure 4.9 Color coded projection map for initial group size of 75 when applying SPPP to Cogongrass-Johnsongrass dataset

### 4.1.2 Unlimited Number of Groups

The second approach was to apply the SPPP technique to the Cogongrass and Johnsongrass datasets with no limitation on the number of groups, or projections, and no limiting factor on the BD. The system was allowed to run until the supervised potential projection was unable to calculate the coefficients for the weighted LDA vector. This phenomenon occurred when the inverse of the data could not be calculated for the LDA weighted vector. The classification accuracies for the two comparison methods was in the upper 90's and the classification accuracy for SPPP method was in the lower 80's, which is shown in Figure 4.10. These results are not surprising because there was no limitation on the BD, which is the optimization factor. We did not expected the classification accuracies to increases if there was no measure on how the SPPP technique was performing in the task of reducing the

dimensionalities of the datasets or increasing the ability to be distinguish between classes. Furthermore, these results were as expected because this approach does not take in account the amount of data that is available.

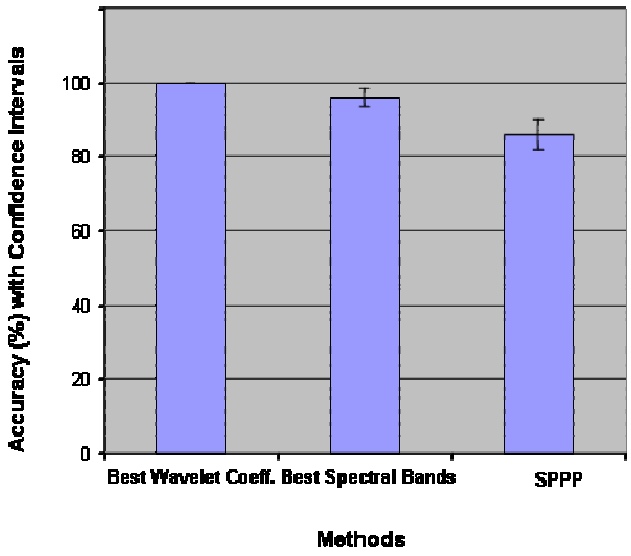


Figure 4.10 Classification accuracies for Cogongrass and Johnsongrass dataset for the dimensionality reduction methods of Best Wavelet Coefficients, Best Spectral Bands, and SPPP with the NM classifier with no limitation on the number of groups and the BD

Table 4.4 and Figure 4.11 show the classification accuracies for the Cogongrass and Johnsongrass dataset for varying initial group sizes. The classification accuracies were the same for all initial group sizes used in this study. The overall classification accuracies were all 86 %. The producers and users classification accuracies for all group sizes were above 84 %. The producers, users, and overall classification accuracies for the second approach are about 10% lower than that of the first method, i.e. when group splitting is bounded by either the amount of initial training data or a non-increasing BD. From the results, the

author suspects that with no monitoring of the optimization of the system the transformation matrix becomes suboptimum which decreases the discrimination between the Cogongrass and Johnsongrass.

Table 4.4 Producers, Users, and Overall Accuracies for different starting group sizes for the Cogongrass and Johnsongrass dataset when the SPPP dimensionality reduction method is applied using all four potential projections with the NM classifier with no limitation on the number of groups

| | **Producers Accuracy** | | **Users Accuracy** | | |
|---|---|---|---|---|---|
| **Group Size** | **Cogongrass** | **Johnsongrass** | **Cogongrass** | **Johnsongrass** | **Overall Accuracy** |
| **50** | 85 | 88 | 88 | 84 | 86 |
| **60** | 85 | 88 | 88 | 84 | 86 |
| **65** | 85 | 88 | 88 | 84 | 86 |
| **75** | 85 | 88 | 88 | 84 | 86 |

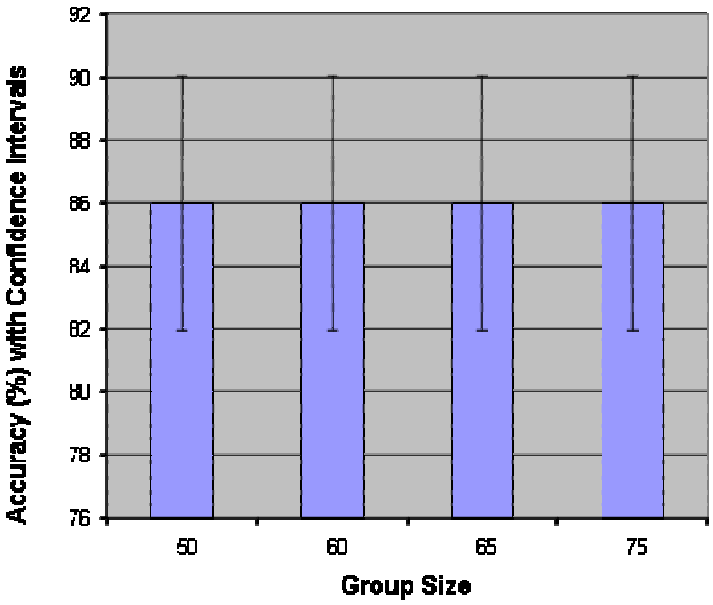


Figure 4.11 Classification accuracies for Cogongrass and Johnsongrass dataset for the dimensionality reduction method SPPP for group sizes 50, 60, 65, 75 with the NM classifier with no limitation on the number of groups

The BD for the second approach had higher values of separation than that of the first method. A very interesting result is that the BD for the group size of 60 decreases about the 13$^{th}$ split and began increasing at the 17$^{th}$ split and continued to increase, as shown in Figure 4.12. The result suggests that there exist local maxima and local minima which could affect the performances of the SPPP system. These local maxima and local minima have the potential to prevent the system from reducing the dimensionality of the dataset for the optimum separation. These phenomena exist throughout several of the cases of varying initial group sizes.

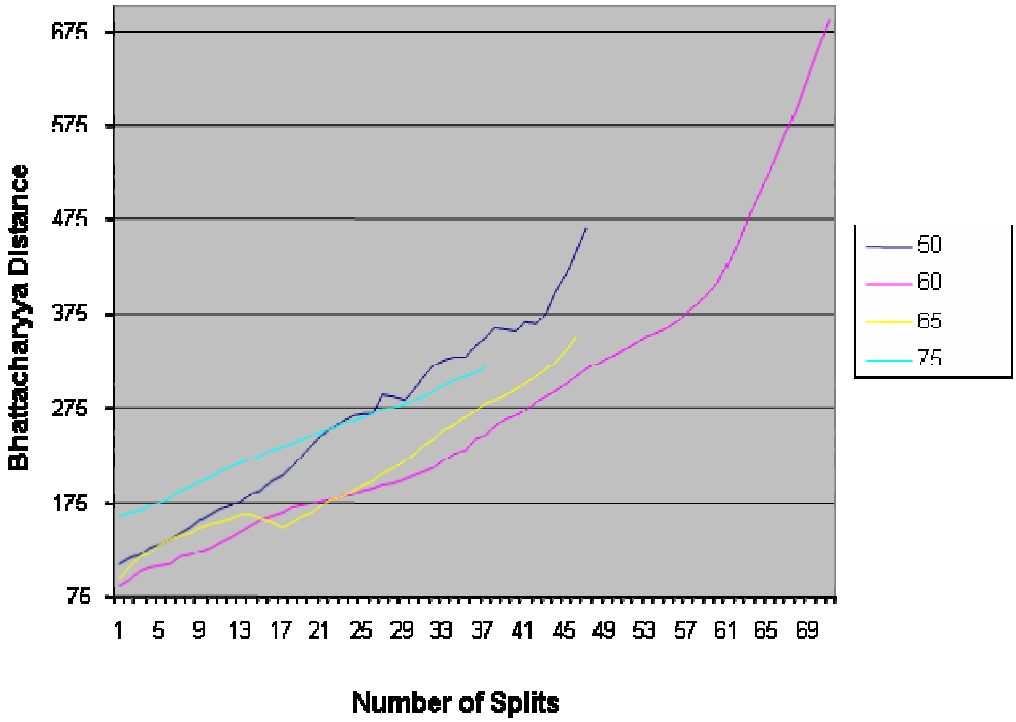


Figure 4.12 Plot of BD for each group size versus the number of times a group was split with no limitation on the number of groups

All four potential projections were selected for each case of initial group size, as shown in Figure 4.13. More total projections were selected for this approach than that of the approach described in section 4.1. This result is as expected, because the splitting of the groups had no limitation factors which implied that more projection would be selected. The supervised potential projection of LDA and the unsupervised Gaussian-weighted average were both again the top selected projections as in the approach described in section 4.1. The results from the color coded projection maps are shown in Figures 4.14 – 4.17. As in the previous results, the LDA potential projection was selected the most for the initial groups. An

interesting result is how the selection of the projection is distributed throughout the four potential projections as the splitting of the groups increase. The color coded map for the group size of 70 had less selected potential projections than that of the others. The number of groups created for each group size where all above 70.

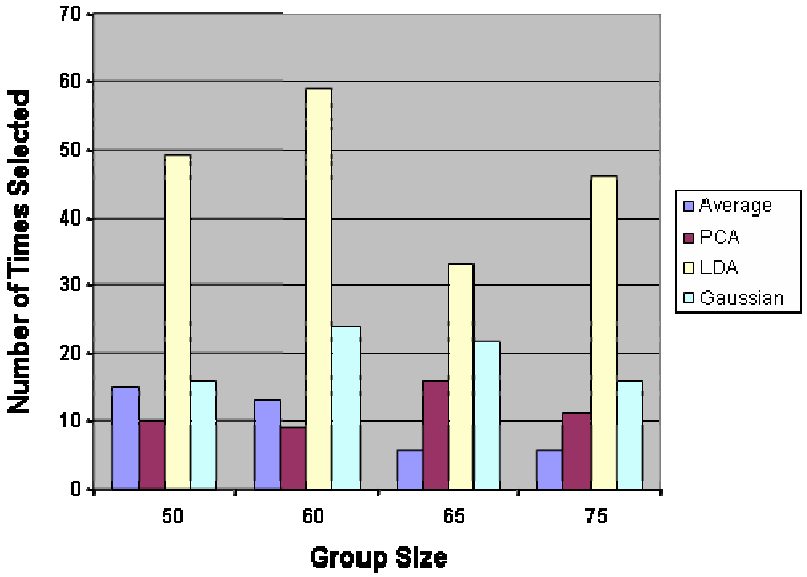


Figure 4.13 Bar graph of the number of times a potential projection was chosen for the different group sizes when applying SPPP to Cogongrass-Johnsongrass dataset, with no limitation on the number of groups.

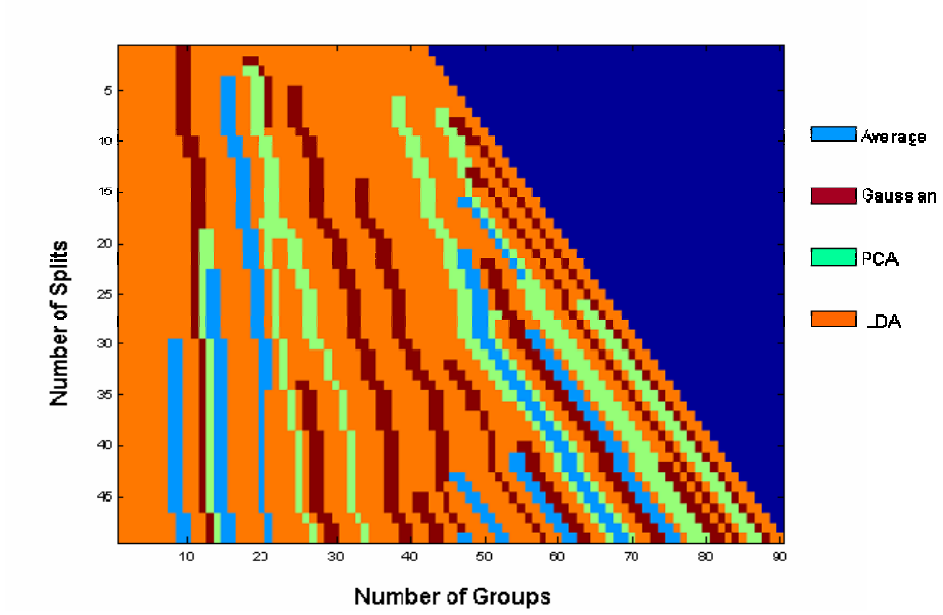


Figure 4.14 Color coded projection map for initial group size of 50 when applying SPPP to Cogongrass-Johnsongrass dataset with no limitation on the number of groups.

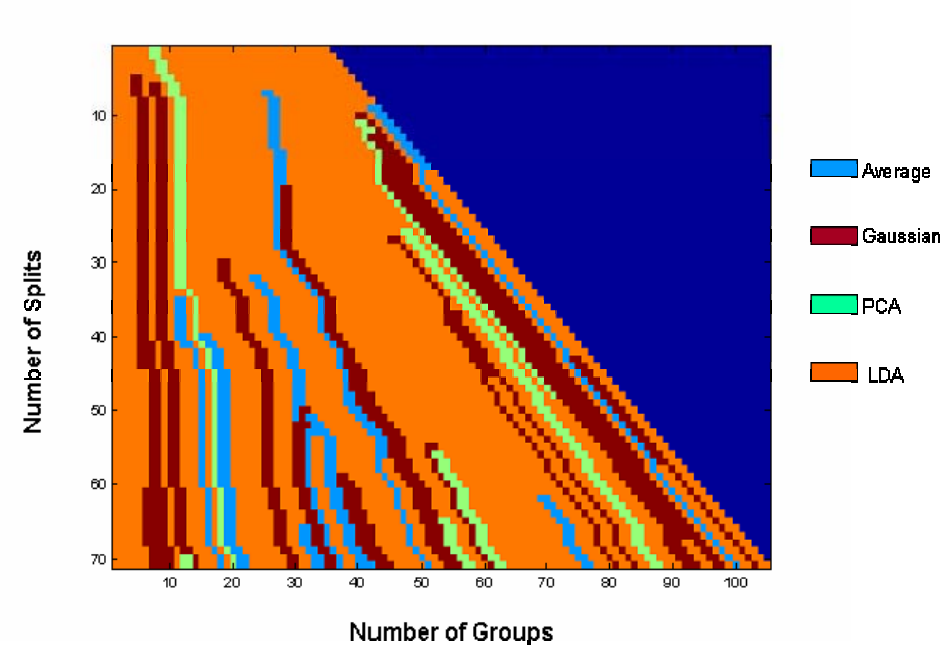


Figure 4.15 Color coded projection map for initial group size of 60 when applying SPPP to Cogongrass-Johnsongrass dataset with no limitation on the number of groups

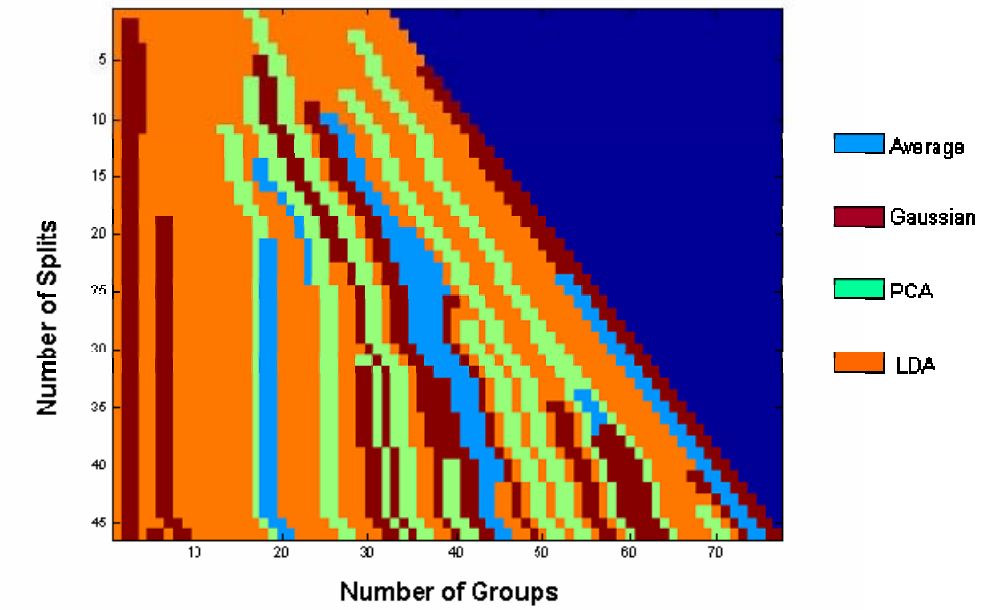


Figure 4.16 Color coded projection map for initial group size of 65 when applying SPPP to Cogongrass-Johnsongrass dataset with no limitation on the number of groups

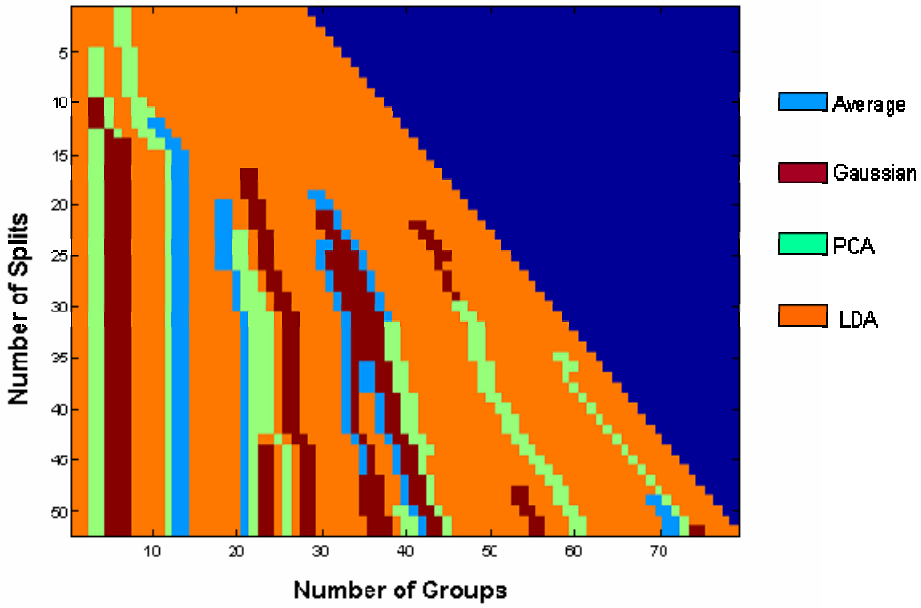


Figure 4.17 Color coded projection map for initial group size of 75 when applying SPPP to Cogongrass-Johnsongrass dataset with no limitation on the number of groups

### 4.1.3 Unsupervised SPPP

The third approach was to apply the SPPP technique to the Cogongrass and Johnsongrass datasets when including only the unsupervised potential projections, i.e. removing the LDA method from the bank of potential projections. In this case, as in section 4.1, the number of groups is limited by the amount of training data available, i.e. the groups cannot split unbounded. This third approach was performed for three reasons. The first was to evaluate what affect the supervised projection had on the system since it had been the projection that was selected the majority of the time. The second was to investigate which unsupervised projections would be selected the most if the supervised projection was not available. The third was to investigate how applying LDA in the classifier instead of within the SPPP technique would improve the classification accuracies of the Cogongrass and Johnsongrass datasets.

The comparison methods along with the SPPP method had classification accuracies in the upper 80's to 90's range, as is shown in Figure 4.18 and Tables 4.5 and 4.6. The SPPP method without LDA applied within the classifier typically had classification accuracies in the 80's. The results in classification for the classifier without LDA is as expected. One expects the classification accuracy of a dataset to be lower when there is no optimization within the classifier. However, when LDA was included with the NM classifier, i.e. applying LDA after all projections were conducted and before NM classification, resulted in surprisingly high accuracies. In fact, the overall classification accuracies are as high as those for the case described in

section 4.1, where LDA is included as on of the potential projections. This result implies that the benefits of the supervised projection, LDA, can be equally obtained either during the dimensionality reduction phase or the classification phase.

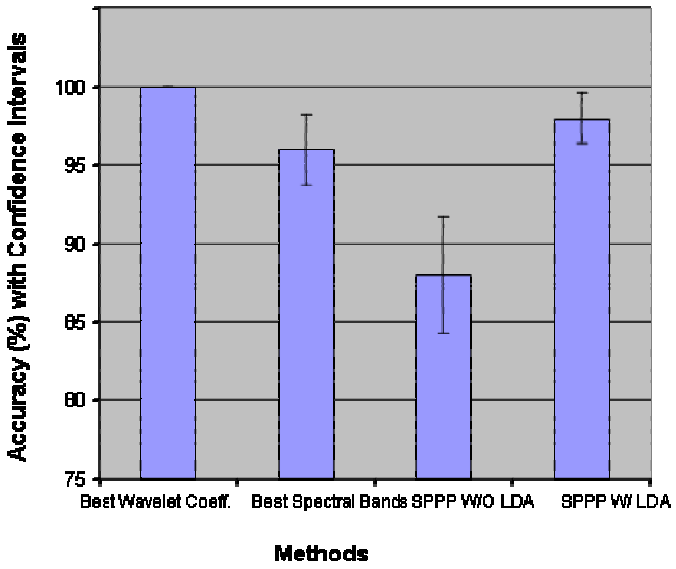


Figure 4.18 Classification accuracies for Cogongrass and Johnsongrass dataset for the dimensionality reduction methods of Best Wavelet Coefficients, Best Spectral Bands, and SPPP and without the potential projection LDA with the NM classifier without LDA and with LDA.

The producers, users, and overall classification accuracies for the unsupervised approach employing LDA within the classifier and exclude LDA from the classifier for the Cogongrass and Johnsongrass datasets are shown in Tables 4.5 and 4.6. The users and producers classification accuracies for the case in which LDA is excluded in the classifier are in the 91 % and above range. The users and producers classification accuracies for the case in which LDA is included in the classifier are in the 96 % and above range.

Table 4.5 Producers, Users, and Overall Accuracies different starting group sizes for the Cogongrass and Johnsongrass dataset when the SPPP dimensionality reduction method is applied using only unsupervised projections with the NM classifier without LDA.

| | **Producers Accuracy** | | **Users Accuracy** | | |
|---|---|---|---|---|---|
| **Group Size** | **Cogongrass** | **Johnsongrass** | **Cogongrass** | **Johnsongrass** | **Overall Accuracy** |
| **50** | 85 | 91 | 92 | 84 | 88 |
| **60** | 85 | 91 | 92 | 84 | 88 |
| **65** | 85 | 91 | 92 | 84 | 88 |
| **75** | 85 | 88 | 88 | 84 | 86 |

Table 4.6 Producers, Users, and Overall Accuracies different starting group sizes for the Cogongrass and Johnsongrass dataset when the SPPP dimensionality reduction method being applied using only unsupervised projections with the NM classifier with LDA within the classifier.

| | **Producers Accuracy** | | **Users Accuracy** | | |
|---|---|---|---|---|---|
| **Group Size** | **Cogongrass** | **Johnsongrass** | **Cogongrass** | **Johnsongrass** | **Overall Accuracy** |
| **50** | 96 | 96 | 96 | 96 | 96 |
| **60** | 96 | 100 | 100 | 96 | 98 |
| **65** | 96 | 100 | 100 | 96 | 98 |
| **75** | 96 | 96 | 96 | 96 | 96 |

The overall classification accuracies when the initial group sizes are 60 and 65 are the highest for each case, using a NM classifier without LDA and with LDA. The classification accuracies are shown to be in the 86 % - 88 % range in the case in which the LDA is excluded from the classifier which is shown Figure 19. The classification accuracies for the case in which the LDA is included with the classifier

vary between the 96 % - 98 % range which is shown Figure 20. In general, the classification accuracies are high for the case when LDA is included in the classifier. Again, however, the results are not significantly affected by the initial group size.

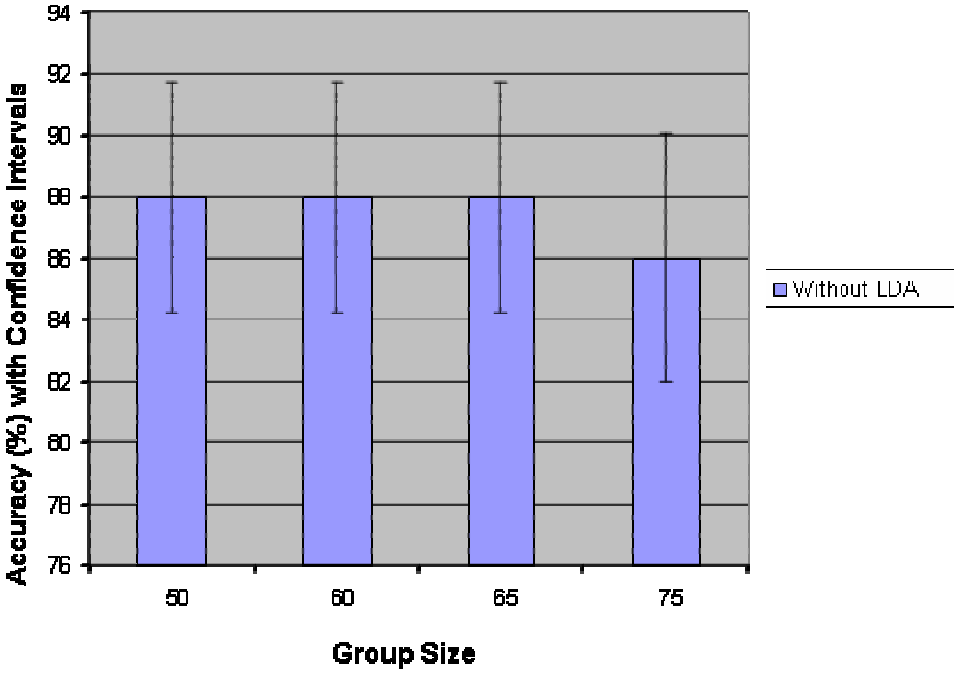


Figure 4.19 Classification accuracies for Cogongrass and Johnsongrass dataset for the dimensionality reduction method SPPP for group sizes 50, 60, 65, 75 with the NM classifier

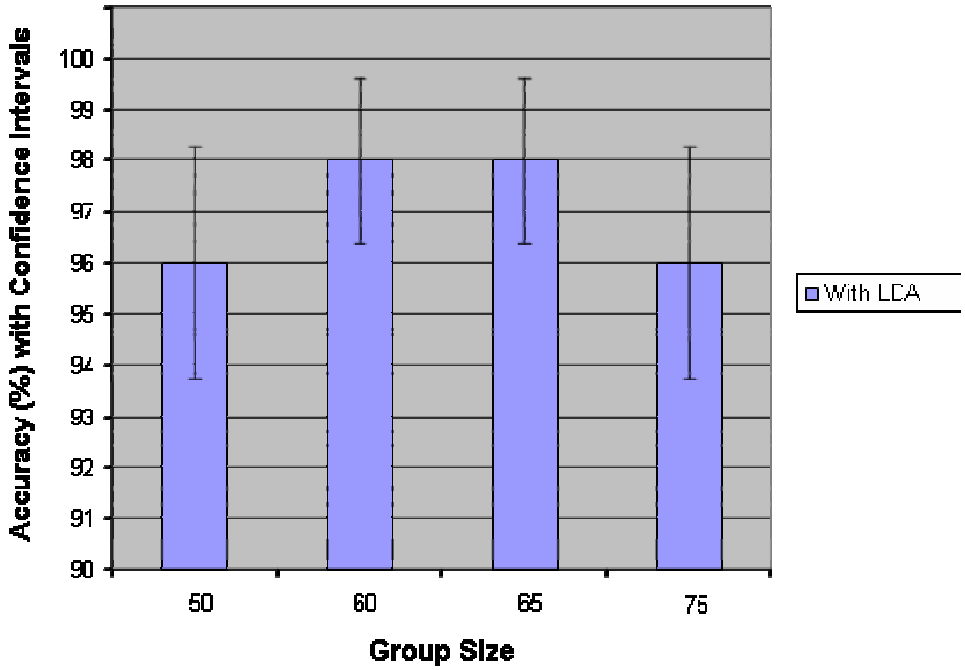


Figure 4.20 Classification accuracies for Cogongrass and Johnsongrass dataset for the dimensionality reduction method SPPP for group sizes 50, 60, 65, 75 with the NM classifier combined with LDA

Figure 4.21 shows the progression of the BD performance index as the SPPP iterates through splitting the groups and forming projections. The figure shows the BD curve for several initial group sizes. The final BD is significantly lower that that in the other approaches described in sections 4.1 and 4.2.

The results for the number of times a potential projection was selected for each group size for the unsupervised projection is shown in Figure 4.22. An interesting result is that the Gaussian-weighted average projection is selected the most for the group sizes 60, 65, and 75 and the PCA projection is selected the most for the lowest group size.

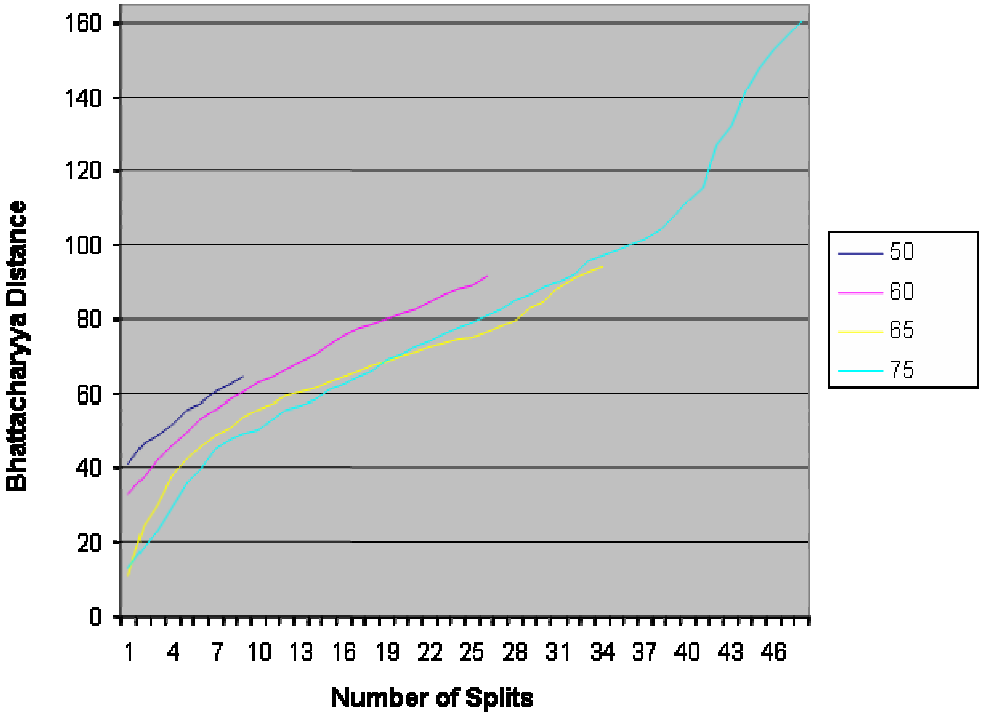


Figure 4.21 Plot of BD for each group size versus the number of times a group was split

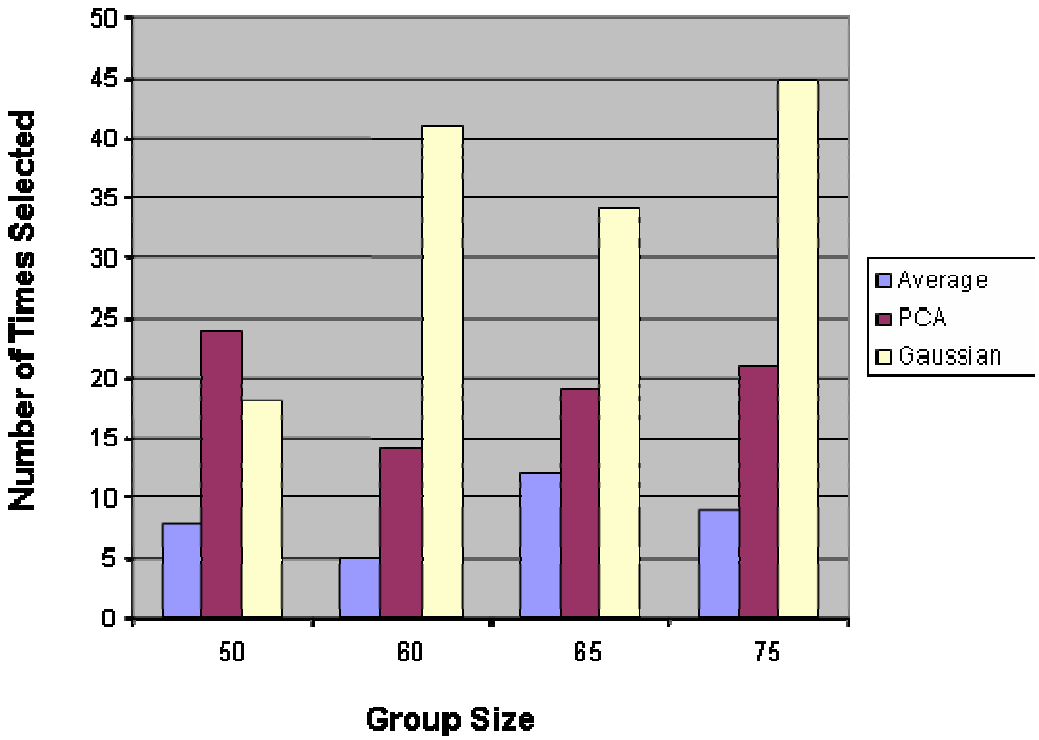


Figure 4.22 Bar graph of the number of times a potential projection was chosen for the different group sizes when applying SPPP to Cogongrass-Johnsongrass dataset

The results for the color coded projection maps for the unsupervised approach are shown in Figures 4.23 to 4.26. A very interesting result is that the unsupervised approach's initial groups of projections are not dominated by just a single projection, but the initial group of projections is a combination of all the potential projections. Also from these colors coded projection maps, the results show that different potential projection tend to migrate to or near to a certain group. For example, it can be seen in all group sizes the projection PCA migrates near the group number thirty.

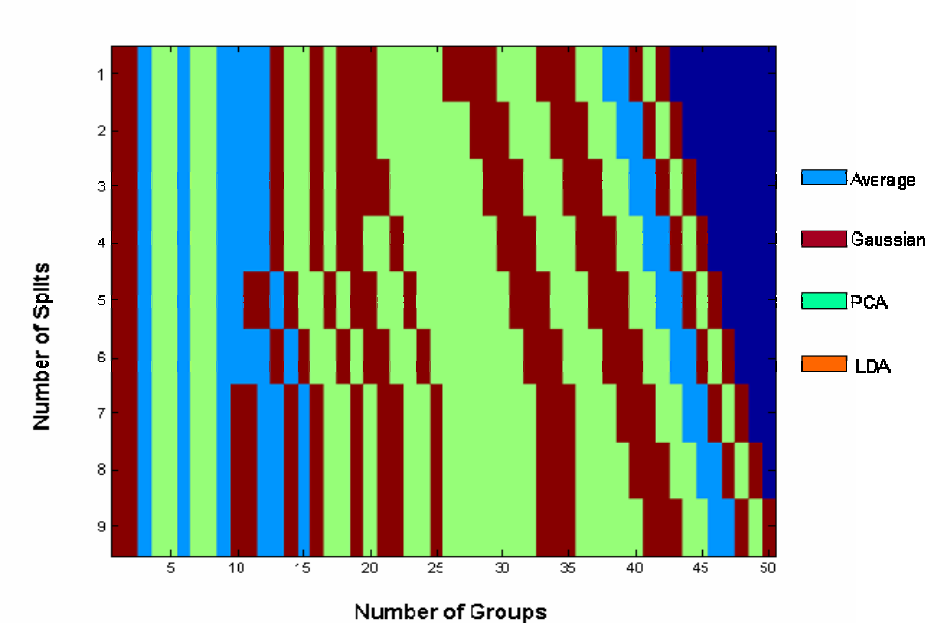


Figure 4.23 Color coded projection map for initial group size of 50 when applying SPPP to Cogongrass-Johnsongrass dataset for unsupervised approach

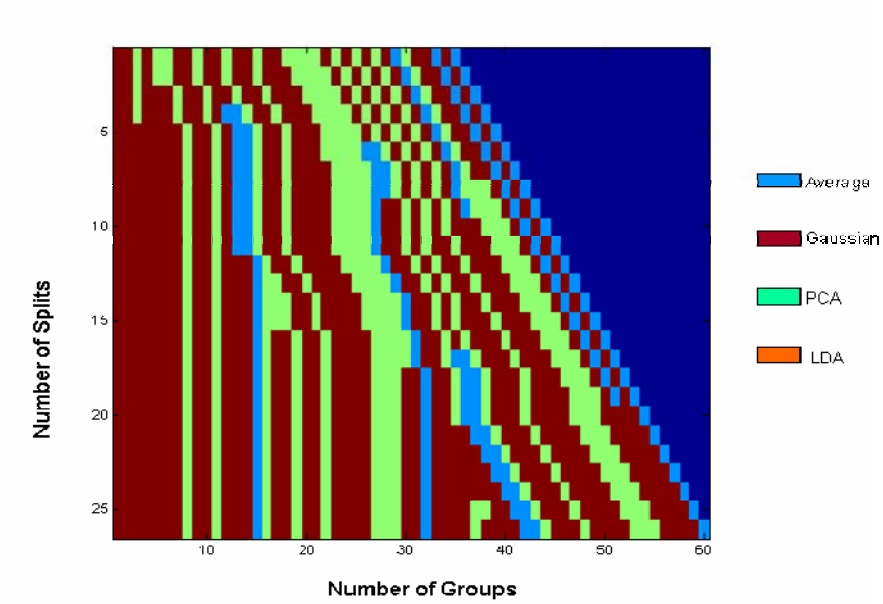


Figure 4.24 Color coded projection map for initial group size of 60 when applying SPPP to Cogongrass-Johnsongrass dataset for unsupervised approach

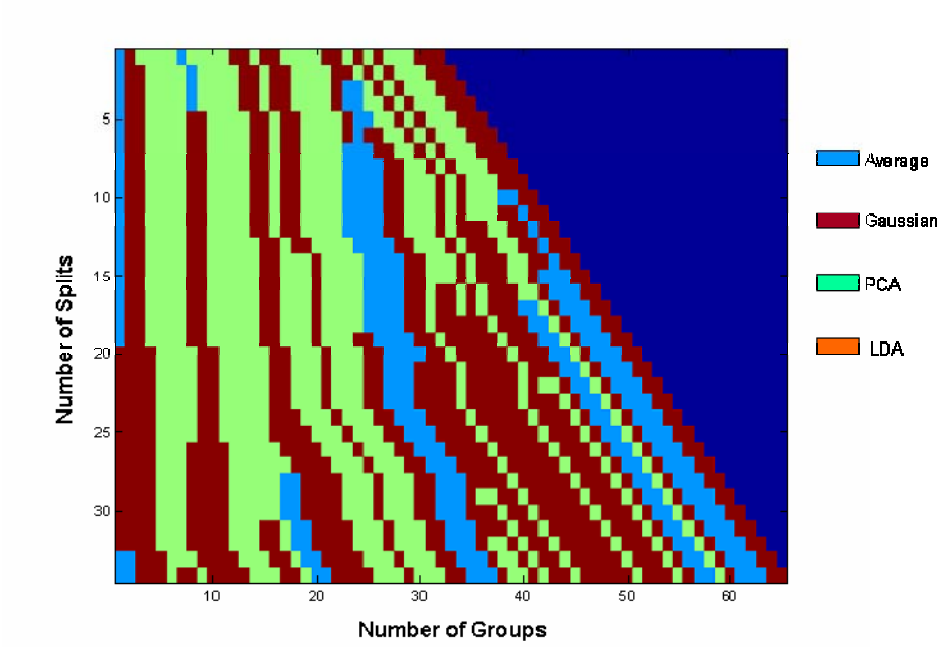


Figure 4.25 Color coded projection map for initial group size of 65 when applying SPPP to Cogongrass-Johnsongrass dataset for unsupervised approach

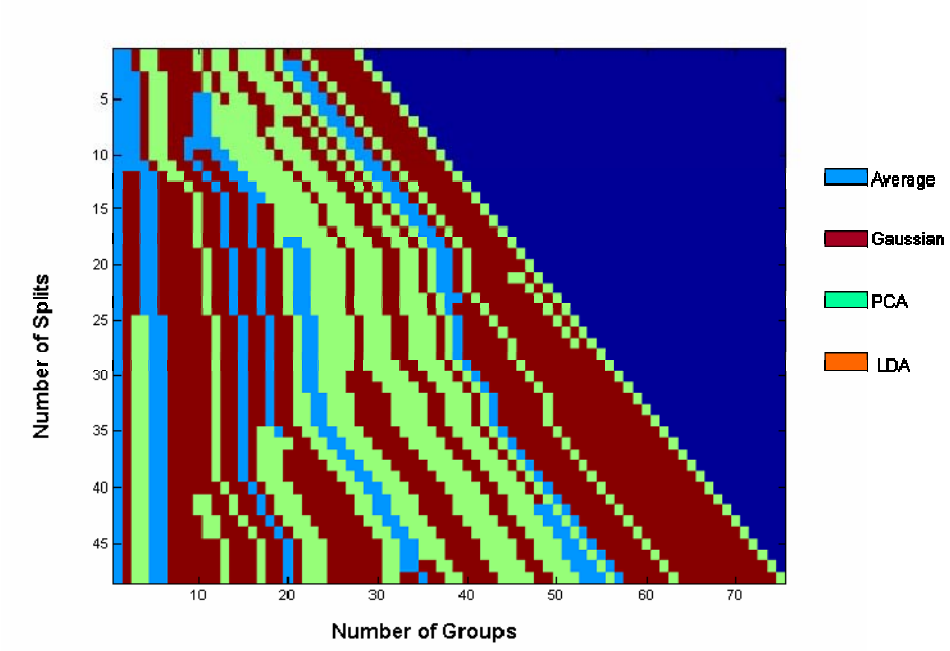


Figure 4.26 Color coded projection map for initial group size of 75 when applying SPPP to Cogongrass-Johnsongrass dataset for unsupervised approach

**4.2 Case Study II – Waterhyacinth versus American Lotus**

The fourth approach was to apply the SPPP technique to the Waterhyacinth and American Lotus dataset. The Waterhyacinth and American Lotus dataset is multitemporal. The data was collected each week for a total of 16 weeks. This dataset is analyzed in two approaches. The first approach was to lump all the data together and not account for the multitemporal aspect of the data. This approach is designed to mimic a scenario where an ATR system is designed to run on remotely sensed data regardless of when that data was collected. In the second approach, the multitemporal data is organized by months and analyzed. This approach is designed to mimic a scenario where an ATR system is designed to run on remotely sensed data for a particular month, i.e. the system is retrained for each month of the growing season. In all of the SPPP analysis of the Waterhyacinth and American Lotus data, the initial group size was set to be 60. The results were found to be very insensitive to the initial group size, and thus, only one initial group size is reported for this dataset.

The classification accuracies for the different approaches are shown in Figure 4.27. The combined dataset results in classification accuracies in the range of 60 % to 75 % . The classification accuracies for the monthly datasets are fairly high. The SPPP method performed well, high 90's, on each of the monthly datasets, whereas the BSBS and BWCS methods appeared to be more sensitive to the time in which the data was collected. However, all three methods performed very well for the August dataset.

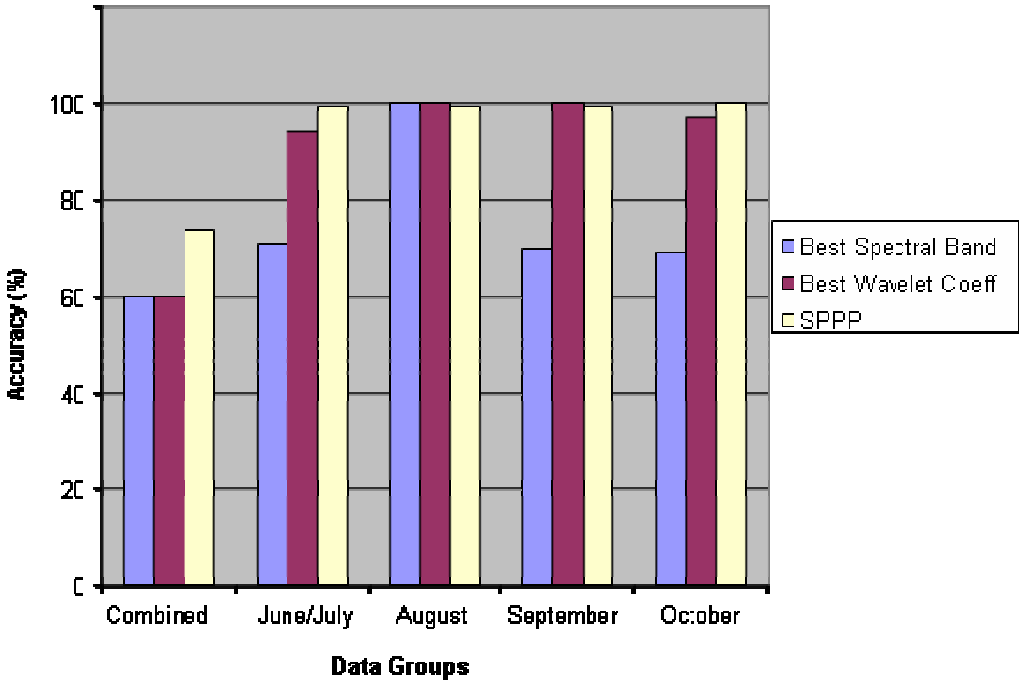


Figure 4.27 Classification accuracies for American Lotus and Waterhyacinth dataset for the dimensionality reduction methods of Best Wavelet Coefficients, Best Spectral Bands, and SPPP

The producer, user, and overall accuracies for the SPPP method are shown in Table 4.7. The producer and user accuracies for the monthly odataset are shown to range in the high 90's. The producers and users accuracies for the combined organized data were below 85 %. This demonstrates the benefit of having an ATR system that is designed for use of data that is collected within a particular timespan, rather than having a system that is trained on data from throughout the growing season.

Table 4.7 Producers, Users, and Overall Accuracies different starting group sizes for the American Lotus and Waterhyacinth dataset when the SPPP dimensionality reduction method being applied using all four projections with the NM classifier

| | Producer Accuracy | | User Accuracy | | |
|---|---|---|---|---|---|
| **Months (Group Size of 60)** | **American Lotus** | **Water Hyacinth** | **American Lotus** | **Water Hyacinth** | **Overall Accuracy** |
| **Combined Months** | 72 | 80 | 83 | 64 | 74 |
| **June and July** | 97 | 100 | 100 | 97 | 99 |
| **August** | 98 | 100 | 100 | 98 | 99 |
| **September** | 100 | 98 | 98 | 100 | 99 |
| **October** | 100 | 100 | 100 | 100 | 100 |

The Bhattacharyya curves for the American Lotus and Waterhyacinth dataset are shown in Figure 4.28. The BD for the combined dataset is significantly lower than the monthly datasets. The BD values for the monthly data have separation values which range from 100 to 5000. The combined dataset has separation values which range from 20 to100. These results explain the high overall classification accuracies for the monthly datasets and also explain the low overall classification accuracy for the combined dataset.

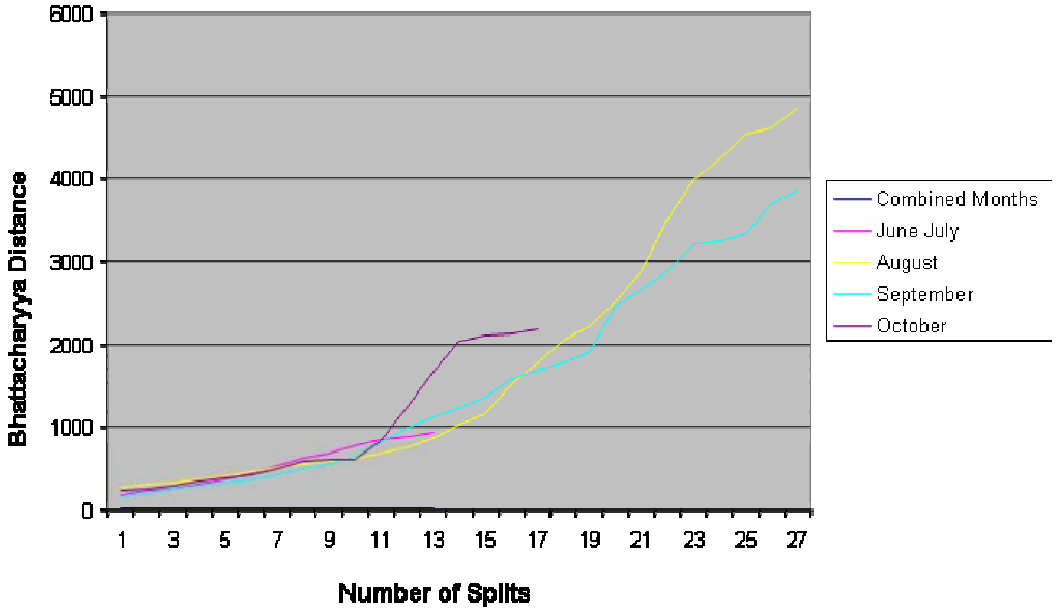


Figure 4.28 Plot of BD for each temporal dataset versus the number of times groups were split

As with the Cogongrass and Johnsongrass dataset, the LDA potential projection was the most commonly selected projection. In fact, for this dataset, the supervised LDA projection dominates, with a simple average and the Gaussian-weighted average being approximately equally selected for a few groups, which is shown in Figure 4.29. A very interesting result is that for the months of June, July, August, and October the potential projection of PCA was not selected at all.

The results for the color coded projection maps are shown in Figures 4.30 to 4.34 for the American Lotus and Waterhyacinth dataset. As in the previous approaches, the LDA potential projection is the dominant projection in the SPPP approach.

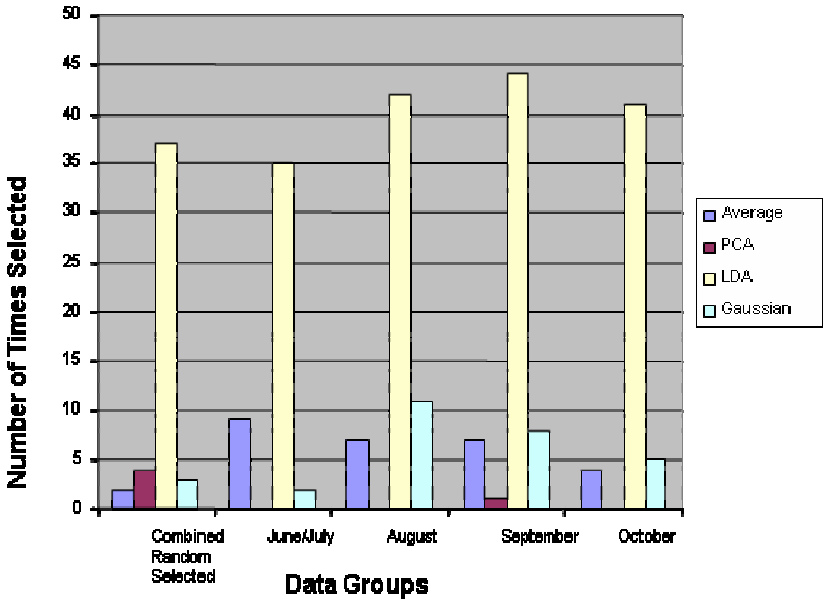


Figure 4.29 Bar graph of the number of times a potential projection was chosen for the different group sizes when applying SPPP to American Lotus-Waterhyacinth dataset

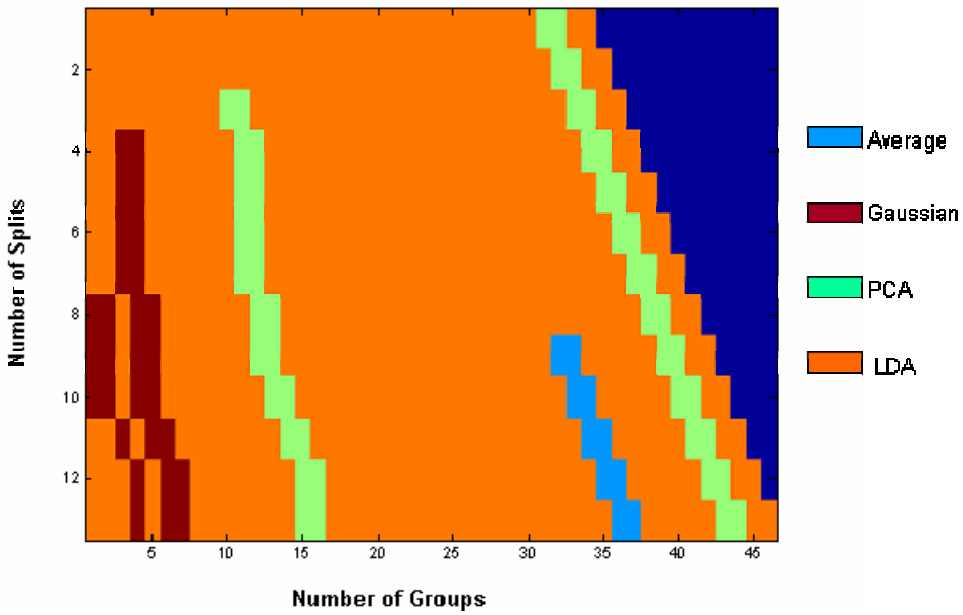


Figure 4.30 Color coded projection map for initial group size of 60 when applying SPPP to American Lotus-Waterhyacinth dataset for the Combined random selected organized data.

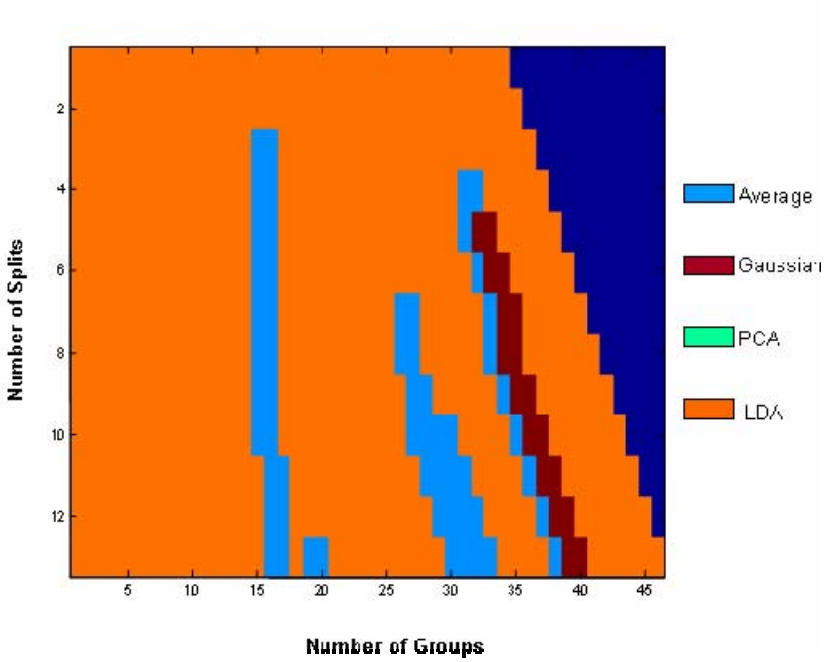


Figure 4.31 Color coded projection map for initial group size of 60 when applying SPPP to American Lotus-Waterhyacinth dataset for the months of June and July

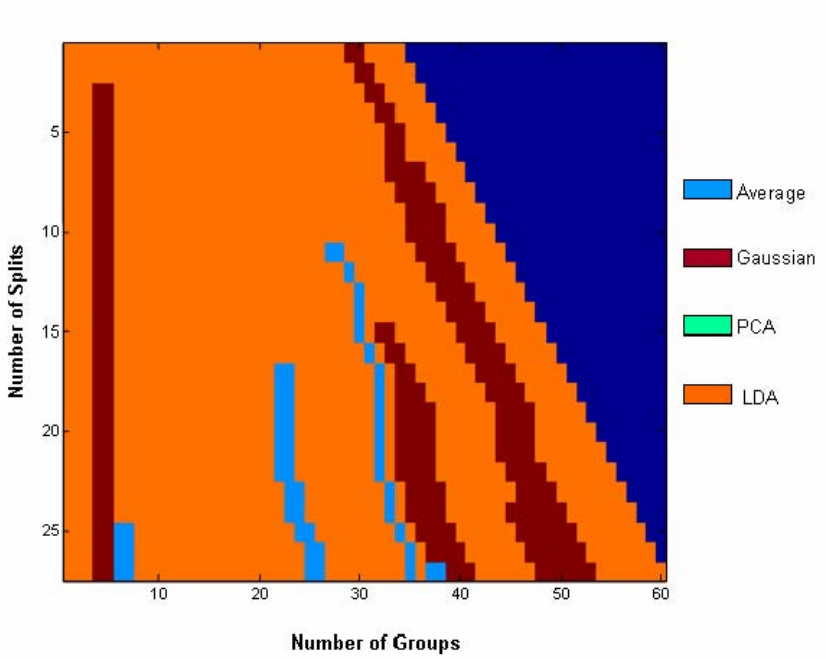


Figure 4.32 Color coded projection map for initial group size of 60 when applying SPPP to American Lotus-Waterhyacinth dataset for the month of August.

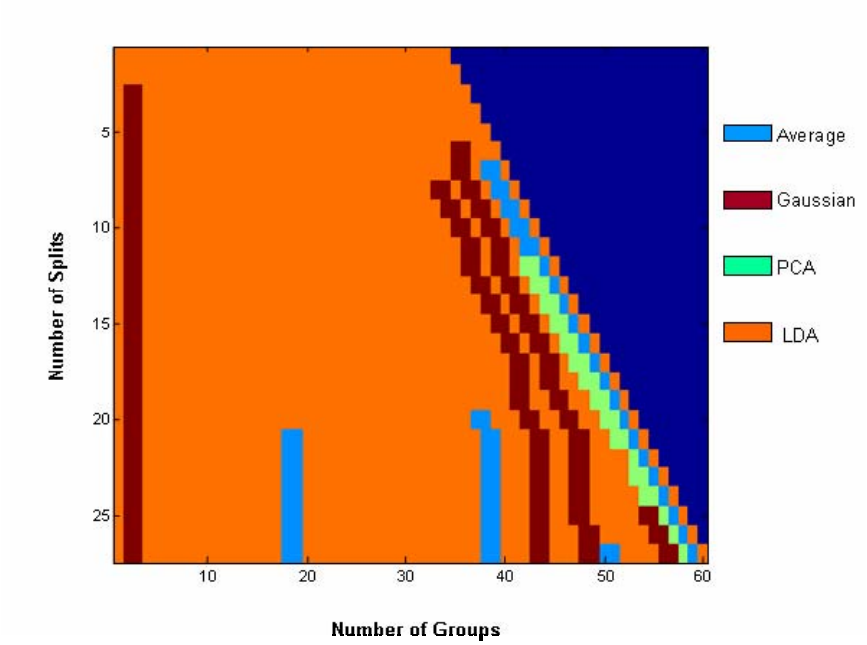


Figure 4.33 Color coded projection map for initial group size of 60 when applying SPPP to American Lotus-Waterhyacinth dataset for the month of September

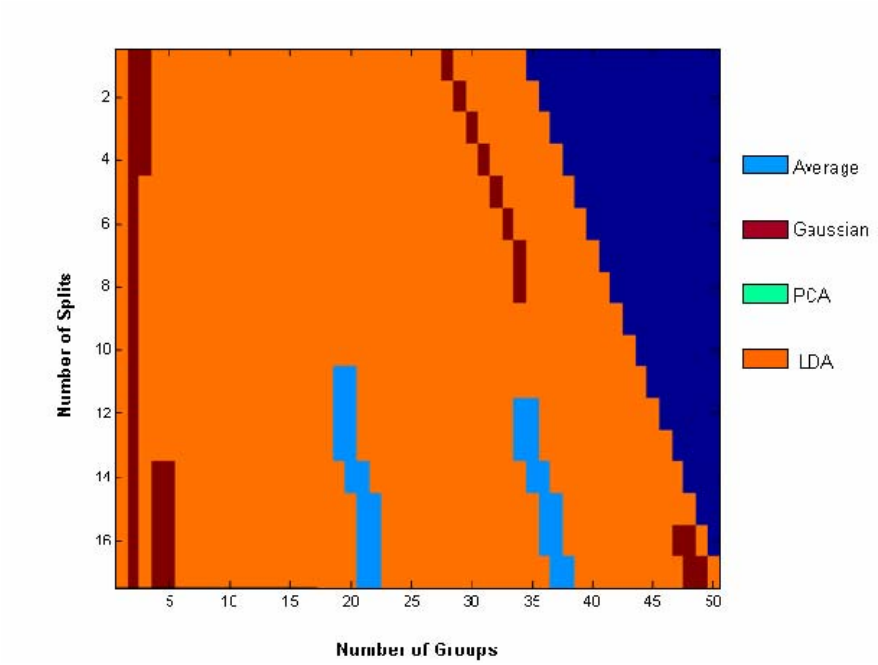


Figure 4.34 Color coded projection map for initial group size of 60 when applying SPPP to American Lotus-Waterhyacinth dataset for the month of October

# CHAPTER V

# CONCLUSION AND FUTURE WORK

## 5.1 Conclusion

In general, the proposed SPPP method performed as well as the BSBS and BWCS methods of dimensionality reduction. Their overall accuracies on the Cogongrass and Johnsongrass dataset were all in the high 90's. However, the computational time of the SPPP was surprisingly less than that of the BSBS and BWCS methods. The SPPP method required approximately 1 to 7 minutes to complete, depending on the initial group size, whereas the BSBS and BWCS methods required approximately 45 to 65 minutes to complete.

The SPPP performance, in terms of overall classification accuracy, was not significantly affected by group size. However, the computation time was substantially affected by the initial group size. Larger group sizes required additional computation time since the SPPP method was allowed to perform more group splitting. It should be noted however, that if the SPPP method were applied to a more challenging class discrimination problem, we might find that the larger initial group sizes would provide improved classification accuracies. Also, we found that the stability of the SPPP approach was quite dependent on the initial group size. Oftentimes, when varying group sizes were selected, the SPPP code would fail.

When the SPPP method was allowed to iterate without a limitation on the number of group splits, the overall accuracy was reduced. When a upper bound was imposed on the group splitting, where the two stopping criteria were based on the amount of training data and on the increase in BD, the overall accuracy was around 96%. However, when no upper bound was imposed on the group splitting, the overall accuracy was around 86%.

Supervised projections, LDA, are predominantly selected when they are included in the bank of potential projections. However, Gaussian-weighted average was also selected in many instances. When LDA was not included in the bank of potential projections, a simple average and Gaussian-weighted average were predominantly selected. And when only unsupervised projections were included, both in the SPPP dimensionality reduction and the classification, the overall accuracies decreased significantly. However, when the supervised projection, namely LDA, was reintroduced in the classification stage, the classification accuracies were on par with those of the system that included LDA in the SPPP stage.

For the multitemporal Waterhyacinth and American Lotus dataset, the SPPP approach performed as well as the BWCS method and outperformed the BSBS method. When the data was combined across dates, none of the three dimensionality reduction methods performed very well. The SPPP method performed the best and only achieved an overall classification accuracy of around 74%. When the data was combined on a monthly basis, the SPPP approach resulted in a near perfect target

detection. If one month were to be recommended for this study, it would be August, where all three dimensionality reduction methods result in approximately 100% classification accuracies.

### 5.2 Future Work

One recommendation for future work would be to include more types of projections, particularly supervised projections, into the bank of potential projections. Since LDA was predominantly selected, especially with the Waterhyacinth dataset, the SPPP approach would probably benefit from having more supervised projections to pursue.

A second recommendation would be to apply the SPPP method to a dataset where more training data is available. Since the group splitting is limited by the minimum amount of training data per class, the SPPP approach could significantly improve when more training data is available, particularly for the Waterhyacinth case study. Also, with the availability of more training data, alternative classifiers could be investigated. In this thesis, the NM classifier was used because more sophisticated classifiers like a maximum-likelihood classifier would fail due to the small number of training samples. It would be very interesting to know if how the SPPP approach performs in combination with other types of classifiers.

A third recommendation would be to combine the SPPP approach with multiclassifiers and decision fusion. Each group, or partition, resulting from the SPPP could be input to its own classifier, and all classifications could then be

combined via some sort of decision fusion, such as qualified majority voting. This type of ATR system could be very powerful.

# REFERENCES


[1] R. O. Duda, P. E. Hart, D. G. Strork, *Pattern Classification*, John Wiley & Sons, Inc., 2001.

[2] K. Fugunaga, R. R. Hayes, "Effects of Sample Size in Classifier Design," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 873-885, 1989.

[3] J. L. Schnase, J. A. Smith, T. J. Stohlgen, S. Graves, C. Trees, "Biological Invasions: a Challenge in Ecological Forecasting," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 1, pp. 122-124, 2002.

[4] L. W. Lass, D. C. Thill, B. Shafii, T. S. Prather, "Detection Spotted Knapweed (Centaurea maculosa) with Hyperspectral Remote Sensing Technology," *Weed Technology*, vol 16. pp. 426-432, 2002.

[5] E. R. R. L. Johnson and D. G. Shilling, "PCA Alien Plant Working Group-Cogon Grass (Imperata Cylindrical): Cogon Grass," [Online]. Available: http://www.nps.gov/plants/alien/fact/imcy1.thm.

[6] C. H. Bronson, M. C. Long, "Invasive Non-native Plants: Beware of Cogon Grass," [Online]. Available: http://fl-dolf.com/forest_management/fh_invasives_cogon.html.

[7] D. Pimentel, L. Lach, R. Zuniga, D. Morrison, "Environmental and Economic Costs of Nonindigenous Species in the United States," *Bioscience*, vol. 50, no. 1, 2000.

[8] M. E. Jakubauskas, D. L. Peterson, S. W. Campbell, F. deNoyelles Jr., S. D. Campbell, D. Penny, "Mapping and Monitoring Invasive Aquatic Plant Obstructions in Navigate Waterways Using Satellite Multispectral Imagery," *Conference Proceedings of Pecora 15/Land Satellite Information IV*, 2002.

[9] T. M. Lillesand, R. W. Kiefer, J. W. Chipman, *Remote Sensing Image Interpretation*, John Wiley & Sons Inc., 2004.

[10] N. Speciale, T. Johnson, L. Allen, "Hyperion Instrument," [Online] Available: http://e01.gsfc.nasa.gov/Technology/Hyperion.html.

[11] Analytical Spectral Devices Fieldspec Pro Spectroradiometer specifications, [Online] Available: http://asdi.com/products_specifications-FSP.asp

[12] S. Balakrishnama, A. Ganapathiraju, J. Picone, "Linear Discriminant Analysis for Signal Processing Problems," *IEEE Proceedings of the Southeastcon*, pp. 78 – 81, 1999.

[13] R. Gonzalez and R. Woods, *Digital Image Processing*, Addition-Wesley Inc., 2001.

[14] A. Cheriyadat, L. M. Bruce, "Why Principal Component Analysis is not an Appropriate Feature Extraction Method for Hyperspectral Data," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 6, pp. 3420 – 3422, 2003.

[15] L. M. Bruce, C. H. Koger, J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 40, pp. 2331-2338, 2002.

[16] J. H. Friedman, "Exploratory Projection Pursuit," *Journal of the American Statistical Association*, vol. 32, 1987.

[17] L. O. Jimenez, D. A. Landgrebe, "Projection Pursuits for High Dimensional Feature Reduction: Parallel and Sequential Approaches," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 1, pp. 148 – 150, 1995.

[18] A. Ifarraguerri, C. Chang, "Unsupervised Hyperspectral Image Analysis with Projection Pursuit," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 38, pp. 2529 – 2538, 2000.

[19] L. M. Bruce, H. Lin, "Projection Pursuits for Dimensionality Reduction of Hyperspectral Signals in Target Recognition Applications," *IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, pp. 960 – 963, 2004.

[20] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, California Academic Press Inc., 1990.

[21] T. M. Mitchell, *Machine Learning*, The McGraw-Hill Companies Inc., 1997.