Investigation of Nonlinearity in Hyperspectral Remotely Sensed Imagery

by

Tian Han

B.Sc., Ocean University of Qingdao, 1984

M.Sc., University of Victoria, 2003

A Dissertation Submitted in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Computer Science

© Tian Han, 2009
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

Investigation of Nonlinearity in Hyperspectral Remotely Sensed Imagery

By

Tian Han

B.Sc., Ocean University of Qingdao, 1984
M.Sc., University of Victoria, 2003

Supervisory Committee

Dr. David J. Goodenough, Supervisor
(Department of Computer Science)

Dr. Jens H. Weber, Co-Supervisor
(Department of Computer Science)

Dr. Dale Olesky, Departmental Member
(Department of Computer Science)

Dr. K. Olaf Niemann, Outside Member
(Department of Geography)

Supervisory Committee

Dr. David J. Goodenough, Supervisor
(Department of Computer Science)

Dr. Jens H. Weber, Co-Supervisor
(Department of Computer Science)

Dr. Dale Olesky, Departmental Member
(Department of Computer Science)

Dr. K. Olaf Niemann, Outside Member
(Department of Geography)

Abstract

Hyperspectral remote sensing excels in its high spectral resolution, which enables the generation of contiguous spectral profiles covering the visible to shortwave infrared region (400 – 2500 nm) of the solar electromagnetic spectrum. The high spectral resolution has greatly stimulated the applications of hyperspectral remote sensing in different disciplines. The initial applications have been found in mineral exploration, followed by applications in environmental research, forest health evaluation, vegetation species mapping, precision farming, water pollution monitoring, and military target identification.

It has been noticed, however, that there is an inconsistency between the statistical characteristics of hyperspectral remotely sensed data and the methods employed to model and process the data for information extraction. On the one hand, hyperspectral data are considered inherently nonlinear, due to the multiple nonlinear sources involved in the data formation. On the other hand, hyperspectral data has long been modeled and processed as realisations of some linear stochastic processes. What is the impact of this inconsistency on hyperspectral data analysis? This dissertation is prepared to address this question by firstly evaluating the significance of nonlinearity in hyperspectral data and

secondly examining the influence of nonlinearity on dimensionality estimation, and noise reduction.

This dissertation proved that nonlinearity existed in hyperspectral data and it was statistically significant. It was found that the dimension of hyperspectral data was substantially smaller when the nonlinearity was considered compared to estimations based on linear algorithms. It was demonstrated that improved noise reduction was achieved without compromising spectral absorption features if the nonlinearity was taken into consideration. The algorithms discussed in this dissertation were implemented, which provided a useful tool set for those who are interested in studying the nonlinear behaviours in hyperspectral data, which are not available in commercial remote sensing software packages.

Table of Contents

List of tables

List of Figures

Acknowledgements

Life is full of uncertainties. But there is one thing sure to me: without Dr. David G. Goodenough opening the door for me to the fascinating world of remote sensing and without his guidance and support in the past ten years, I would not have gone this far down the road: from a Mandarin-speaking oceanographer to an English-speaking remote sensing scientist. Thank you Dave! My sincere gratitude also goes to other supervisory committee members of my graduate study for their time and effort spent on reviewing, commenting, and examining this dissertation. They are Drs. Jens H. Weber, Dale Olesky, K. Olaf Niemann, and Jay Pearlman.

I would like to take this opportunity to acknowledge the following agencies for providing resources to my graduate study, including

- GeoBC, Integrated Land Management Bureau, Ministry of Agriculture and Lands, British Columbia, Canada;
- Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada;
- Natural Centre of Ocean Technology, State Oceanic Administration, China.

Dedications

To my family

Chapter 1

Introduction

Hyperspectral remote sensing, also called imaging spectroscopy, is a data acquiring technology that evolved from spectroscopy, a method for material identification that has been used in the laboratory by physicists and chemists for over 100 years. By measuring absorption features due to specific chemical bonds, spectroscopy is capable of detecting and identifying minerals, vegetation, and man-made materials. With the advance of electro-optical technology in the past 30 years, spectroscopy has found applications outside of the laboratory environment, leading to the development of field and imaging spectroscopy. Field spectroscopy has been established to characterize the reflectance of natural surfaces in situ [1], where the sensing instrument is installed on platforms that are close to the targets being observed. In contrast to field spectroscopy, imaging spectroscopy or hyperspectral remote sensing has the sensing instrument mounted on platforms that are far from the targets. A typical configuration of imaging spectroscopy is to install the sensing instruments on airplanes or spacecraft to measure targets on the surface of the Earth or other astronomical objects.

By combining imaging and spectroscopy, hyperspectral remote sensing systems acquire data that are generally composed of 100 to 500 spectral images with narrow bandwidths (3 - 10 nm). The acquired hyperspectral data sets are saved as digital images that contain both spatial and spectral information. In addition to being presented as digital images, hyperspectral data can also be viewed and studied in feature and spectral presentations. Each of these forms of presentation depicts certain aspects of hyperspectral data suitable for different applications. Compared to other optical remote sensing technologies, hyperspectral remote sensing excels in its high spectral resolution, which enables the generation of contiguous spectral profiles covering the visible to shortwave infrared region (400 – 2500 nm) of the solar electromagnetic spectrum. The high spectral resolution has greatly stimulated applications of hyperspectral remote sensing in different

disciplines. The initial applications have been found in mineral exploration, followed by applications in environmental research, forest health evaluation, vegetation species mapping, precision farming, water pollution monitoring, and military target identification.

Hyperspectral data are often modeled and processed by algorithms assuming that the data are realizations of some linear stochastic processes, though the physical processes involved in hyperspectral data collection have suggested that nonlinear dynamics may exist. “The nonlinear processes are originated from the interaction of light with atmosphere and surface mater, which is then observed in the data through the instrument collection process,” (Pearlman, private communication, 2009). The dilemma is likely due to the reason that either the nonlinearity in hyperspectral data is not strong enough and that algorithms based on the linear data assumption may still do the job, or that the effective algorithms that are capable of dealing with nonlinearity are limited. Regardless of the reasons, the assumption of linearity on data characteristics compromises the effectiveness and accuracy of information extraction from hyperspectral data.

This dissertation is dedicated to investigating the nonlinear issues in hyperspectral remotely sensed data from a nonlinear time series analysis perspective with the following three objectives: *1) mathematically prove the existence of nonlinearity in hyperspectral data; 2) determine the intrinsic dimensionality of hyperspectral data and extract the corresponding features by taking nonlinearity into consideration; and 3) improve the effectiveness of noise reduction by explicitly addressing the nonlinearity in hyperspectral data.* Nonlinearity and its impact on dimension estimation, noise reduction, and feature extraction have not been sufficiently studied but have implications for hyperspectral remote sensing applications. Two sets of hyperspectral remotely sensed data are used throughout this dissertation. They are a space-borne image acquired by Hyperion and an airborne image acquired by AVIRIS. Both were collected over an area of coastal forest during the growing season under ideal weather conditions. Though they have similar spectral resolution and coverage, these data sets differ markedly in spatial resolution and

signal-to-noise ratio. They provide a test bed to study the relationships between nonlinearity, spatial resolution and signal strength.

This dissertation is composed of the following six chapters, including *Introduction, Investigation of nonlinearity, Nonlinearity-counted dimension estimation and feature extraction, Nonlinearity-counted noise reduction, Algorithm implementation, and Conclusions and contributions.* The chapter of *Introduction*, i.e. the current chapter, is provided to set off the research objectives of this dissertation. It also describes the layout and structure of this dissertation and provides the necessary background information to connect subsequent chapters. The background information includes hyperspectral imaging systems, hyperspectral data characteristics, applications of hyperspectral data, and the hyperspectral data sets used in this dissertation. The second chapter, *Investigation of nonlinearity*, reviews the process of hyperspectral data formation and the physical processes involved that may introduce nonlinearity to hyperspectral data. The highlight of this chapter is to mathematically prove the existence of nonlinearity in the spectral domain of hyperspectral data. Pixels of different land-covers, including forest canopy and water, which are believed highly nonlinear, are selected for the investigation. In the third chapter, *Nonlinearity-counted dimension estimation and feature extraction*, a new approach to determining the intrinsic dimensionality of hyperspectral data is proposed based on phase space reconstruction using “spectral” delay embedding. The estimated dimensions are compared with those derived by using conventional eigenvector-based methods that assume linear data distribution. This chapter is finished up by applying a nonlinear feature extraction method to extract the corresponding number of features based on the determined dimensions. The fourth chapter, *Nonlinearity-counted noise reduction*, addresses the issues related to the linear noise reduction approaches, which either omit the existence of nonlinearity or mistake the nonlinearity for noise. These approaches often end up removing the information-bearing nonlinearity in hyperspectral data. A nonlinear noise reduction method is therefore proposed as the remedy. Implementing details of the proposed algorithms in this dissertation are given in Chapter Five, *Algorithm implementation*, whose emphasis is on the selection of application programming interface and the approaches to deal with memory and computing limitations. Key contributions of this dissertation are summarized in the sixth

chapter, *Conclusions and contributions*, where recommendations for future work are also included.

1.0 Hyperspectral remote sensing technology

Hyperspectral remotely sensed data are acquired by imaging systems, installed in either airborne or space-borne platforms. Those systems are capable of capturing the reflected solar radiation from the Earth's surface simultaneously in 100 to 500 contiguous spectral bands [2] [3]. A typical hyperspectral imaging system is conceptually composed of an optical lens, a wavelength dispersion device (either a prism or diffraction grating), and a two-dimensional detector array, as shown in Figure 1. The cross-track spatial dimension of the scene is projected onto one dimension of the detector array, while the spectral dimension is spread across the second dimension of the detector array. The along-track spatial dimension of the scene is covered as the airborne or space-borne platform glides forward. Since the 1980s, several hyperspectral imaging systems have been built and put into operation. Here is a non-exhaustive list of these systems:

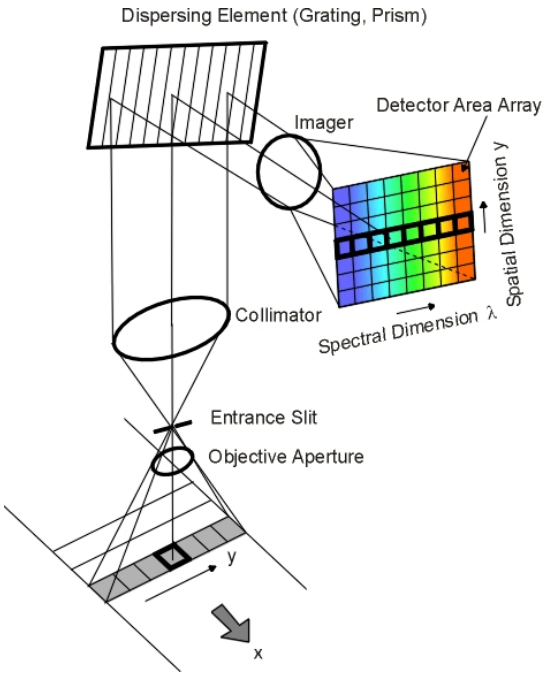


Figure 1. Hyperspectral imaging system.

- Airborne Imaging Spectrometer (AIS), JPL, NASA, United States [4].
- Compact Airborne Spectrographic Imager (CASI), ITRES Research Inc., Canada [5].
- Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), JPL, NASA, United States [6].
- Hyperspectral Mapper (HYMAP), Integrated Spectronics Inc., Australia [7].
- Hyperspectral Data and Information Collection Experiment (HYDICE), Naval Research Laboratory, United States [8].

- Modular Airborne Imaging Spectrometer (MAIS), Shanghai Institute of Technical Physics, China [9].
- Hyperion, TRW Inc. United States [10].
- Compact High Resolution Imaging Spectrometer (CHRIS), Sira Electro-Optics Ltd, United Kingdom [11].
- Airborne Imaging Spectrometer for Applications (AISA), Specim Spectral Imaging Ltd., Finland [3].

As a representative of the airborne systems, AVIRIS is a nadir-viewing whisk-broom instrument that was installed aboard aircraft flying at different altitudes. When operating on a high altitude aircraft (20 km above mean sea level), AVIRIS measures upwelling ground radiance with the spatial response of 1.0 mrad, forming pixels of 20 m × 20 m on the ground. The image width (swath) is 11 km wide, and the image length is typically 10 to 100 km. When operating on a low altitude aircraft (4 km above mean sea level), AVIRIS produces 4 m × 4 m pixels on the ground with a swath of 2 km. For both flying scenarios, the spectral response ranges from blue to shortwave-infrared (380 to 2450 nm) in the electromagnetic spectrum. The radiance entering the instrument is divided into four grating spectrometers and is further broken down into a total of 224 contiguous spectral bands approximately 10 nm wide each. Linear arrays are used, which are composed of silicon (for visible bands) and indium-antimonide detectors (for infrared bands). Data quantization was 10 bits before 1995 and was upgraded to 12 bits thereafter. The signal-to-noise ratio in 2000 was 1000:1 and 400:1 for visible and near infrared (VNIR) and shortwave infrared (SWIR) bands respectively. AVIRIS has been flown in North America, Europe, and portions of South America. An image cube acquired by the high altitude AVIRIS is shown in Figure 2 (left).

As the first space-borne hyperspectral imaging system, Hyperion is a nadir-viewing push-broom instrument aboard Earth Observing-1 (EO-1), a sun-synchronous satellite (flying at an altitude of 705 km) that was launched by NASA on November 21, 2000. Spatially, Hyperion produces pixels of 30 m × 30 m on the ground with a 7.5 km swath. Spectrally, Hyperion covers a similar range of the electromagnetic spectrum (400 to 2500 nm) as AVIRIS. The incoming radiance is divided into two grating spectrometers

corresponding to VNIR and SWIR spectral regions, respectively. The VNIR spectrometer splits the 400 – 1000 nm radiance into 70 bands, approximately 10 nm each, which are collected by Charge Coupled Device (CCD) detectors. Similarly the SWIR spectrometer splits the 900 – 2500 nm radiance into 172 bands, which are captured by HgCdTe detectors. The combination of the VNIR and SWIR results in 242 bands in total with a 22-band overlap between 900 – 1000 nm, which allows cross calibration between the two spectrometers. The signal-to-noise ratio (SNR) of Hyperion varies with the spectral region: ~150 between 400 and 700 nm, ~110 between 700 and 1125nm, and ~40 between 1125 and 2400 nm. A Hyperion image cube is shown in Figure 2 (right).

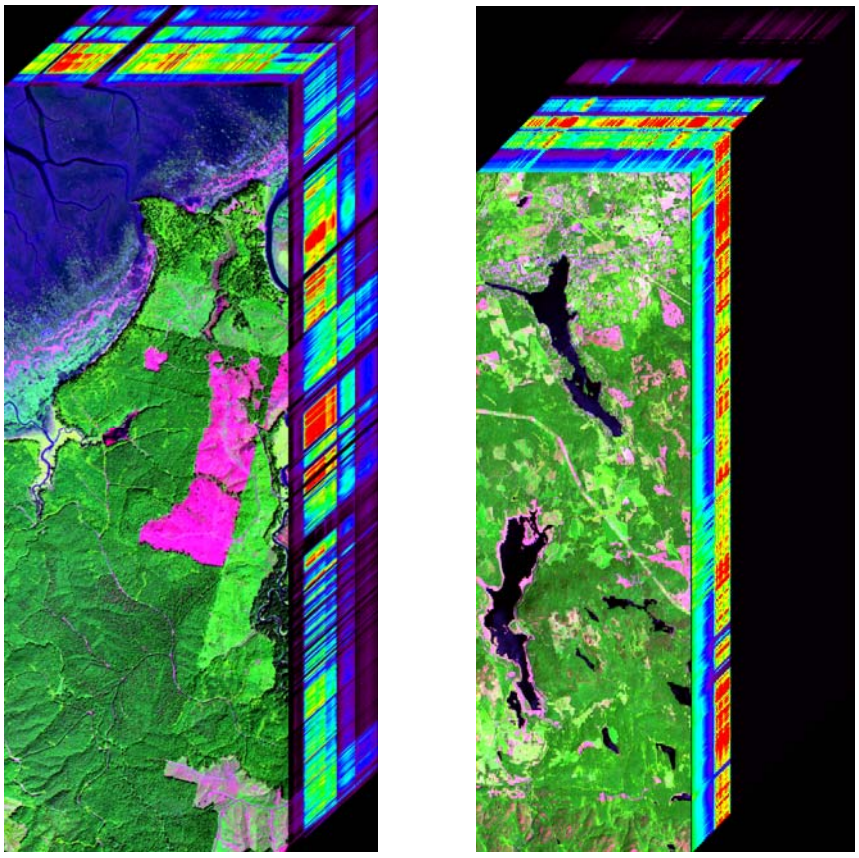


Figure 2. Examples of hyperspectral remotely sensed images: AVIRIS (left) and Hyperion (right)

The original EO-1 Mission was successfully completed in November 2001. Due to the high interest in continued acquisition of Hyperion data expressed by the remote sensing research and scientific communities, NASA and the United States Geological Survey (USGS) have made an agreement to allow continuation of the EO-1 Program as

an Extended Mission. The EO-1 Extended Mission is chartered to collect and distribute Advanced Land Imager (ALI) multi-spectral and Hyperion hyperspectral products in response to data acquisition requests. Under the Extended Mission provisions, image data acquired by EO-1 are archived and distributed by the USGS Center for Earth Resources Observation and Science (EROS) and placed in the public domain [12].

Though it began almost 30 years ago, hyperspectral remote sensing technology is in greater demand due to its unique ability to characterize the complex interaction between solar radiation and the Earth surface structures. Compared to the first generation imagers, the current hyperspectral imagers provide hyperspectral data with improved qualities, including wider spectral coverage, higher signal-to-noise ratio, finer resolution (both spectral and spatial), and wider swath. As an airborne example of such imagers, the Airborne Reflective Emissive Spectrometer (ARES) is built by the Integrated Spectronics, Australia, and co-financed by DLR German Aerospace Center and the GFZ GeoResearch Center Potsdam, Germany [13]. ARES is a whisk-broom sensor that provides ~160 channels in the visible and infrared region (0.45-2.45 μm) and the thermal region (8-13 μm) at 2 – 10 meter spatial resolution. It consists of four co-registered individual spectrometers, three of them for the reflective and one for the thermal part of the spectrum. The spectral resolution is between 12 and 15 nm in the solar wavelength range and about 150 nm in the thermal. ARES is used mainly for environmental applications in terrestrial ecosystems. The thematic focus is on soil sciences, geology, agriculture and forestry. The instrument is offered to the scientific community on a commercial basis as well as through planned national and international programs starting from 2007.

One of the major goals of ARES is to prepare the ground for a future space-borne hyperspectral mission – the Environmental Mapping and Analysis Program (EnMAP) [14]. The hyperspectral imager being developed in this program will work in push-broom configuration with two spectrometers (VNIR and SWIR) covering the spectral region of 420 – 2450 nm. Compared to the first generation space-borne hyperspectral imager, Hyperion, this instrument is similar in spatial resolution and bandwidth, but has a much wider swath: 30 km and hence a much larger daily spatial coverage (~150,000 km^2 in

nadir looking operation mode), which allows operations on a global scale. In addition, the EnMAP hyperspectral imager will produce hyperspectral data with an improved SNR (~500 in VNIR and ~150 in SWIR). This will help to reduce the uncertainty associated with the low SNR hyperspectral data. EnMAP is scheduled to be launched in 2010 [14].

1.1 Applications of hyperspectral remote sensing

The ability to acquire laboratory-like spectra remotely is a major advance in remote sensing capability. As many surface materials, although not all, have diagnostic absorption features that are 20 to 40nm wide at half the band depth, hyperspectral imaging systems, which acquire data in contiguous 10nm-wide bands, can produce data with sufficient resolution for direct identification of those materials [2]. This capability is not possessed by other remote sensing instruments. Hyperspectral remote sensing has been applied to a wide variety of areas, which can be categorized functionally as land-cover mapping, target detection/recognition, and indicator/index derivation.

Land-cover mapping based on remotely sensed data dates back to the early 1970s when multi-spectral data collected by the Multiple Spectral Scanner (MSS) aboard Landsat-1 were available. Due to the broad bandwidth of MSS, the data collected can only be used to map land-cover types with marked spectral differences, such as broadleaf and coniferous forests for forestry applications. Hyperspectral remote sensing, owing to the narrow bandwidth of the hyperspectral imager, has extended the conventional meaning of land-cover mapping based on multi-spectral remote sensing to include 1) mapping of land-cover types with similar spectral signatures, 2) mapping of properties other than land-cover, and 3) sub-pixel mapping (spectral unmixing). As an example of mapping land-cover types with similar spectral signatures, Goodenough et al. (2003) successfully executed mapping major forest species with Hyperion data [15]. The forest species mapped included Douglas-fir (*Pseudostuga menziesii* var. *Menziesii*), western red-cedar (*Thuja plicata)*, lodgepole pine (*Pinus contorta*), and red-alder (*Alnus rubra*). The classification reaches 90% accuracy, though the spectral signatures of these species are very similar. The work done by Underwood et al. [16] provided another example of this kind, where AVIRIS data were used to map invasive species, including ice-plant (*Carpobrotus edulis*), in California's Mediterranean-type ecosystems. Validation with

field sampled data showed high mapping accuracies for identifying the presence or absence of ice-plant (97%).

Hyperspectral remote sensing is capable of characterizing the Earth surface materials by measuring the position, depth, and width of the related absorption features, which make it possible to map other land-cover properties besides land-cover type. For example, Goodenough et al. (2003) mapped forest foliar nitrogen content based on the statistical relationship established between lab-measured nitrogen and AVIRIS data [17]. Schlerf et al. (2005) explored in [18] whether hyperspectral data improved estimation and mapping of biophysical forest variables compared to multi-spectral data. They found that the hyperspectral data contains more information than the multi-spectral data relevant to the estimation of forest stand variables, which led to better estimations of leaf area index and crown volume.

Sub-pixel mapping is another function made possible by hyperspectral remote sensing. Due to spatial heterogeneity of the terrestrial Earth surface, an area corresponding to a typical hyperspectral image pixel, say 30 × 30 m, often covers several different materials visible from remote sensing sensors, such as tree foliage, under-storey vegetation, soil, rock, etc. Sub-pixel mapping is developed to determine the spatial coverage of these components within the area corresponding to each image pixel. The problem is usually expressed mathematically as a set of linear equations, where the right hand sides are the pixel digital numbers at different bands and the left hand sides are the area-weighted summations of the spectral values at different bands of the above materials (also called endmembers) [19]. The areas used as the weights aforementioned above are the unknowns in these equations. To solve the equations, the number of equations, which is equivalent to the number of bands, should not be less than the number of unknowns (endmembers). Because spectral adjacent bands are very correlated, this leads to dependent equations. It is required that the number of bands needed should be more than the number of endmembers so that the problem can be solved in the sense of least squares. For hyperspectral data, this condition is guaranteed satisfactorily, where the number of bands available is more than enough. Among the numerous applications of sub-pixel mapping using hyperspectral data, Farrand and Harsanyi [20] employed a sub-

pixel mapping algorithm, called Constrained Energy Minimization (CEM), with AVIRIS imagery to map mine tailing. They found that CEM performed very well. Among a total of 484 ferruginous rich pixels, 472 of them were correctly identified by CEM. Goodenough et al. (2008) proposed a method based on fully-constrained least squares to estimate the spatial coverage of forest canopy within each image pixel [21]. This sub-pixel mapping, conducted by using both 20 m and 4 m AVIRIS data, produced reasonable results (72% correspondence) with the following endmembers: clear-cuts, grassy land, surface water, bare soil, forest canopy, shrub land, and false spectra.

Target detection/recognition is another active area of hyperspectral remote sensing application, which is often implemented by extracting extreme pixels from hyperspectral imagery and comparing them with the entries in a spectral library for target recognition. For example, Ren et al. (2006) developed in [22] a technique for target recognition. Realizing that most of interesting targets usually occur with low probabilities and small population, which may not be able to constitute reliable second-order statistics, they proposed to use high-order statistics to perform target detection. They demonstrated that the following five minerals: alunite, buddingtonite, calcite, kaolinite, and muscovite, were correctly identified within an AVIRIS scene. In the same study, they also placed 15 panels of different size and painted with different materials in a HYDICE scene. According to their experiment, all these panels had been correctly detected, though some of them were not visible from the image as the size of the panels was below image spatial resolution. Goovaertsa et al. (2005) described a novel technique in [23] to detect the disturbed soil in mine tailings near Yellowstone National Park, based on the hyperspectral data acquired by Probe-1, a 128-band airborne hyperspectral imager [24]. Compared to the detection approaches based solely on spectral information, the proposed technique was developed by capitalizing on both spatial and spectral correlations, which made it effective to deal with the complex landscape containing multiple targets of various sizes and shapes. The successful application of this technique on analyzing mine tailings also indicated its potential for other similar applications, including identifying locations of buried landmines or toxic waste.

Similar to the land-cover mapping discussed above, indicator/index derivation, a technique originated from multi-spectral remote sensing, is also greatly expanded by hyperspectral remote sensing. The Normalized Difference Vegetation Index (NDVI), a measure of density and vigor of green vegetation, is perhaps the most well-known indicator born from multi-spectral remote sensing. It is calculated by ratioing the difference to summation of the near-infrared and red reflectance. Vegetation NDVI typically ranges from 0.1 up to 0.6, with higher values associated with greater density and greenness of plant canopies, while the NDVI values of non-vegetation, such as soil and rock, are close to zero. As hyperspectral remote sensing is capable of collecting reflectance at precise spectral locations corresponding to different absorption features, many new indicators/indices, not possible with multi-spectral remote sensing, have been derived from hyperspectral data. Cho & Skidmore [25] proposed a linear extrapolation method to extract the red edge position (REP) from hyperspectral data for explaining a wide range of nitrogen concentration, where the hyperspectral data were simulated using field measured spectra. They concluded that the destabilizing effect of the double-peak feature on the REP/nitrogen relationship can be mitigated and spectral changes near the low and high nitrogen sensitive peaks can be determined by identifying the REP as an intersection of two straight lines extrapolated on the far-red and NIR flanks of the first derivative reflectance spectrum. Using high spatial resolution hyperspectral data acquired by ROSIS and Digital Airborne Imaging Spectrometer (DAIS), Zarco-Tejada et al. [26] explored several total chlorophyll indices, including the modified chlorophyll absorption index (MCARI), the transformed chlorophyll absorption index (TCARI), and the optimized soil-adjusted vegetation index (OSAVI). They found that the image-derived indices and those based on ground measured data agreed reasonably well when targeting crowns. They claimed that a radiative transfer model that accounts for shadow, soil, and crown reflectance was required when hyperspectral optical indices were applied to open-canopy situations.

It should be noted that the applications mentioned above in this sub-section are just a few examples indicating the proliferation of hyperspectral remote sensing in natural resource and environment sectors. This is a simple survey conducted based on the three most influential remote sensing journals: IEEE Transactions on Geoscience and Remote

Sensing, Remote Sensing of Environment, and International Journal of Remote Sensing. In the most recent three years, the numbers of hyperspectral-related papers published in these journals were 216, 114, and 137, respectively.

1.2 Characteristics of hyperspectral data

Hyperspectral data are usually processed using multivariate statistical analysis approaches, where each pixel of the image is treated as a vector in a high dimensional space. The entire image is organized as a matrix in such a way that a column vector represents an image pixel and a row vector represents a band image acquired at a certain wavelength between 400 to 2500 nm. The dimension of each pixel vector is defined as the number of spectral bands of the image. Hyperspectral data can be presented in three domains to facilitate data processing and analysis for different applications [27]. In addition to the image domain presentation as shown in Figure 2, hyperspectral data can also be presented in the spectral domain and feature domain. Figure 3 shows four pixels

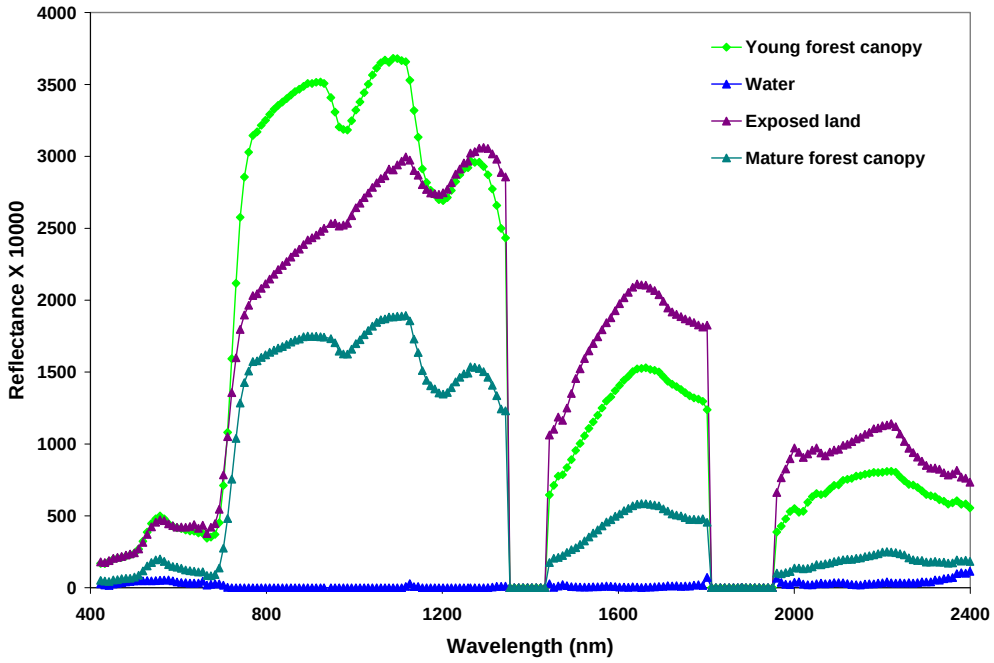


Figure 3. Examples of hyperspectral data presented in spectral domain (X axis represents wavelength and Y axis depicts the scaled reflectance)

of the AVIRIS image (Figure 2) presented in the spectral domain, where each of them is shown as a spectral profile across wavelength. These spectral profiles are also called spectra. Many materials on the Earth surface have unique spectral signatures. By

comparing and matching pixel spectra with the spectral signatures in a spectral library, one can identify targets of interest, as the police identify suspects by matching their finger prints.

The 3rd domain of hyperspectral data presentation is the feature domain that is spanned by the hyperspectral band images. For hyperspectral imagery with 200 bands, a coordinate system of 200 dimensions is required for a complete representation of the imagery, where each pixel in the image is shown as a point. The feature domain presentation facilitates information extraction from hyperspectral data, such as classification, spectral unmixing, and target recognition. Figure 4 shows an example of some AVIRIS pixels depicted in the feature domain with three bands, where pixels appear in four separated clusters. The pixels in the same cluster tend to represent the same materials or targets on the ground.

The high spectral resolution of hyperspectral remote sensing provides one with the capability to analyze and study subtle spectral difference and variation of the Earth surface targets. However, this capability comes with the price of high data dimensionality, where the dimension is defined as the number of spectral bands that an

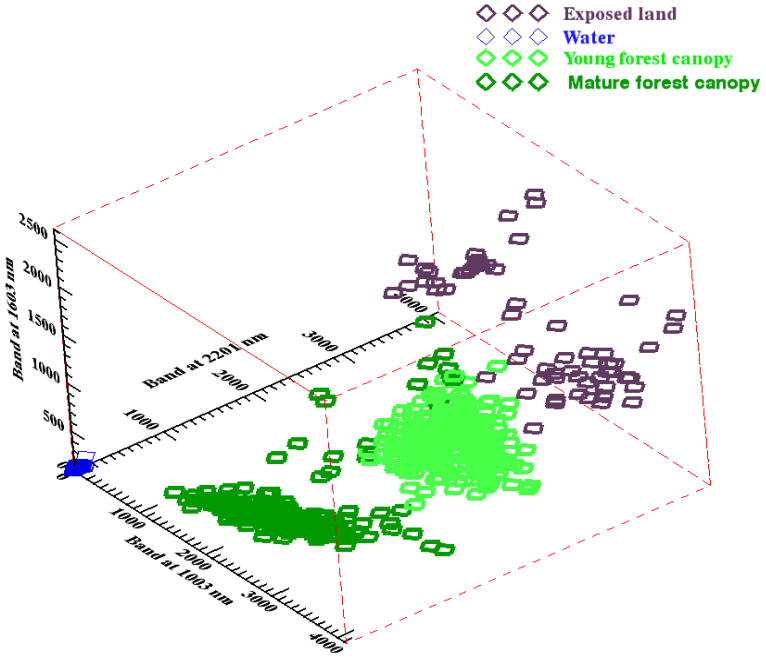


Figure 4. Examples of hyperspectral data presented in feature domain (X, Y, and Z axes represent three bands at 1003, 2201, and 1603 nm, respectively)

imaging system employs for data acquisition. Compared to the multispectral remotely sensed data, whose dimension is usually below 10, the dimension of hyperspectral data may be over 200. The direct impact of the high dimensionality is the dramatic increase of data volume and computing complexity. For example, with the same spatial coverage, the data volume of a Hyperion image is about 50 times larger than that of a Landsat TM image, because Hyperion has 242 bands with 12-bit quantization [10], while the Landsat TM has 7 bands with 8-bit quantization [28]. Instead of treating each pixel in the Landsat TM image as a 7-dimensional vector, one has to deal with a vector of 242 dimensions for each Hyperion pixel.

The size of hyperspectral imagery is gigantic. A normal-sized hyperspectral image, say 640 × 640 pixels of an AVIRIS image, consumes about 142-Mb disk space. In addition, the speed of imaging or data acquisition is incredibly fast for hyperspectral remote sensing. For example, Hyperion can acquire 782-Mb data in less than 5 minutes [10]. The gigantic image size and the fast-paced data acquisition have direct impact on data storage, distribution, access, and management. New systems and tools are demanded to efficiently handle hyperspectral data. SAFORAH [29] is such a system that is developed based on the advanced networking cyber-infrastructure with the ability to manage, catalogue, store, and disseminate large volumes of Earth observing (EO) data, share computing resources, and facilitate collaborative research and automation over broadband communications.

Besides the data volume, a significant impact of hyperspectral data is on information extraction methods, such as classification, spectral unmixing, and target detection. Hughes (1968) pointed out in [30] that, given a finite number of training samples, the mean recognition accuracy of a pattern classifier could always increase as one increased the measurement complexity (equivalent to the number of bands used in remote sensing). However, there was a saturation point. Once this point was reached, the accuracy would drop with the increasing of the measurement complexity (Figure 5). This is the so-called “Hughes Phenomenon”, the remote sensing version of “the curse of dimensionality”, which is often used to express the detrimental impact of general high-dimensional data for information extraction [31]. Landgrebe et al. verified that in the

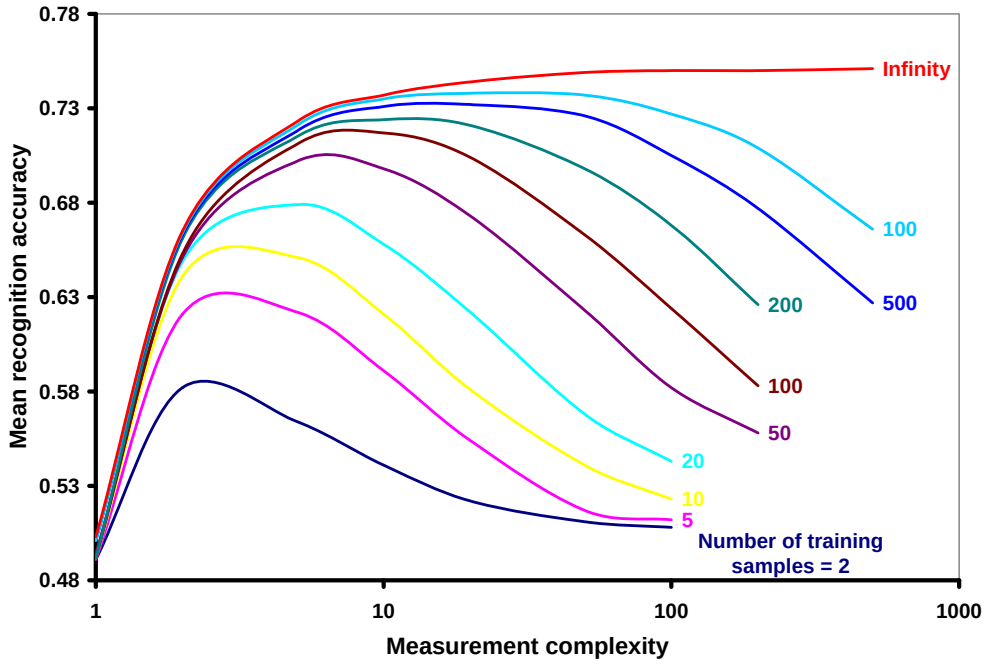


Figure 5. Hughes Phenomenon (X-axis represents measurement complexity in terms of number of bands and Y-axis depicts mean accuracy of land-cover recognition by classification)

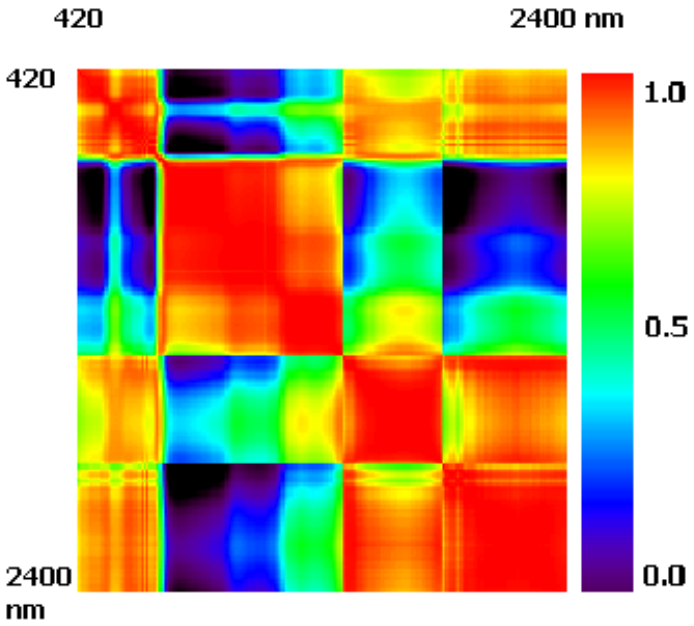


Figure 6. An example of AVIRIS data correlation image (the red and blue areas indicate bands of high and low correlations, respectively)

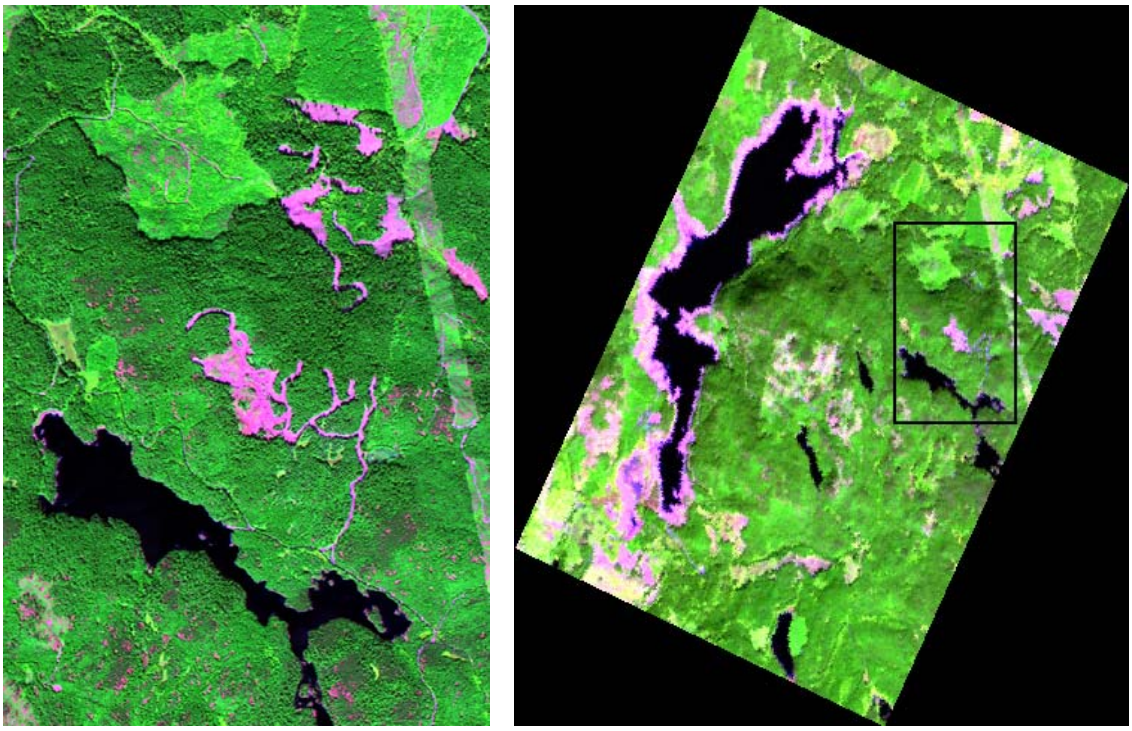
context of hyperspectral remote sensing and gave detailed exposition about the relationship between classification accuracy and the number of employed bands in

[27][32][33]. To mitigate the Hughes Phenomenon, one could either keep adding training information to pattern classifiers as the measurement complexity increased, or reduce the dimensionality of the measurement complexity. For hyperspectral remote sensing, the first option is not practical, because acquiring enough training information is expensive and often impossible. Therefore, the preferred choice is the reduction of data dimensionality.

Previous studies, including [32 – 37], showed that the high-dimensional space spanned by hyperspectral bands was mostly empty, where the data volume was concentrated at corners of the outside shell in the data hypercube. This indicates that the total dimensionality of a hyperspectral image defined by the number of spectral bands is much larger than the intrinsic dimensionality, and the space spanned by the hyperspectral bands is reducible without losing much information and discriminating capability. One can easily find that there exists significant redundancy in hyperspectral imagery, which is demonstrated in a correlation matrix image as shown in Figure 6. The blocky structures shown in Figure 6 originated from the between-band correlations due to the contiguous and narrow band configuration of the hyperspectral imagers. The high band correlation and the near emptiness of the hyperspace spanned by hyperspectral bands are dealt with by dimensionality reduction, which is often considered as one of the pre-processing steps prior to any information extraction from hyperspectral data. Principal Component Analysis (PCA) and Maximum Noise Fraction (MNF) [38] are two well-known linear algorithms for this purpose.

1.3 Hyperspectral data of interest

Two hyperspectral images are chosen in this dissertation: an airborne image acquired by AVIRIS and a space-borne image acquired by Hyperion, which are shown in Figure 7 (RGB composites of the 3 channels at ~1500 nm, ~750 nm, and ~640 nm, respectively). The AVIRIS image in 4 m spatial resolution was collected on August 10, 2002, while the Hyperion image in 30 m spatial resolution was collected on September 10, 2001. These images are selected out of consideration for the representation and coverage of current hyperspectral remote sensing in platform, spatial resolution, and signal-to-noise ratio so that the findings in this dissertation are representative and applicable. Both of the selected images were acquired under clear weather conditions



(a) **(b)**

Figure 7. Hyperspectral images of interest: AVIRIS (left) and Hyperion (right), where the enclosed area indicates the spatial coverage of the AVIRIS image)

over the Greater Victoria Watershed District (GVWD), Vancouver Island, British Columbia, Canada [15]. The GVWD test site is a coastal forest area, where the predominant forest species is Douglas-fir (*Pseudostuga menziesii* var. *Menziesii*) and the secondary species includes western hemlock (*Tsuga heterophylla*), western white pine (*Pinus monticola*), lodge-pole pine (*Pinus contorta*), red alder (*Alnus rubra*), western red cedar (*Thuja plicata*), and arbutus (*Arbutus menziesii*). The under-storey ground cover of GVWD is a heterogeneous mixture of salal (*Gaultheria shallon*) and Oregon grape (*Mahonia nervosa*). The GVWD area contains some of the oldest unmanaged stands of Douglas-fir in the southern area of Vancouver Island. The relief of this area is approximately 600 meters with an average elevation at 400 m above mean sea level. The slopes vary but attain gradients as great as 45 degrees. The above images were pre-processed by following similar steps as those employed in [15], including radiometric correction, orthorectification, atmospheric removal, and spectral correction. The water absorption bands around 1400 nm and 1900 nm were removed from both images, leaving 179 bands for further analysis.

The following notation is adopted throughout this dissertation. Matrices will be denoted by upper-case characters, vectors by array-capped lower-case characters, and scalars by lower-case characters. Based on this notation, a hyperspectral image with r rows, c columns, and n bands is mathematically represented by a matrix, X, which is composed of p column vectors, i.e. $X = [\vec{x}_1, \vec{x}_2, ..., \vec{x}_p]$, where $p = c \times r$, the number of pixels included in the image. The dimension of each column vector $\vec{x}_i$ is n, the number of spectral bands. For the AVIRIS image of consideration, $r = 1000$; $c = 550$; and $n = 179$. They are 387, 273, and 179, respectively, for the Hyperion image.

Chapter 2

Investigation of nonlinearity

Although hyperspectral remotely sensed data are believed to be nonlinear, they are often modeled and processed by algorithms assuming that the data are realizations of some linear stochastic processes. This is likely due to the reason that either the nonlinearity of the data may not be strong enough and the algorithms based on linear data assumption may still do the job, or the effective algorithms that are capable of dealing with nonlinear data are not widely available. The simplification on data characteristics, however, may compromise the effectiveness and accuracy of information extraction from hyperspectral imagery. This chapter is dedicated to investigating the existence of nonlinearity in hyperspectral data represented by AVIRIS and Hyperion imagery acquired over the GVWD test site. The method employed for the investigation is based on the statistical test using surrogate data, an approach often used in nonlinear time series analysis [39]. In addition to the high-order autocorrelation, spectral angle is utilized as the discriminating statistic to evaluate the differences between the hyperspectral data and their surrogates. To facilitate the statistical test, simulated data sets are created under linear stochastic constraints. Both the simulated and real hyperspectral data are rearranged into a set of spectral series where the spectral and spatial adjacency of the original data is maintained as much as possible. This study reveals that the differences are statistically significant between the values of discriminating statistics derived from the hyperspectral data and their surrogates. This indicates that the selected hyperspectral data are nonlinear in the spectral domain. Algorithms that are capable of explicitly addressing the nonlinearity are needed for processing hyperspectral remotely sensed data.

2.0 Motivation

There is an inconsistency between the statistical characteristics of hyperspectral remotely sensed data and the algorithms employed to model and process hyperspectral data for information extraction. On the one hand, hyperspectral data are considered

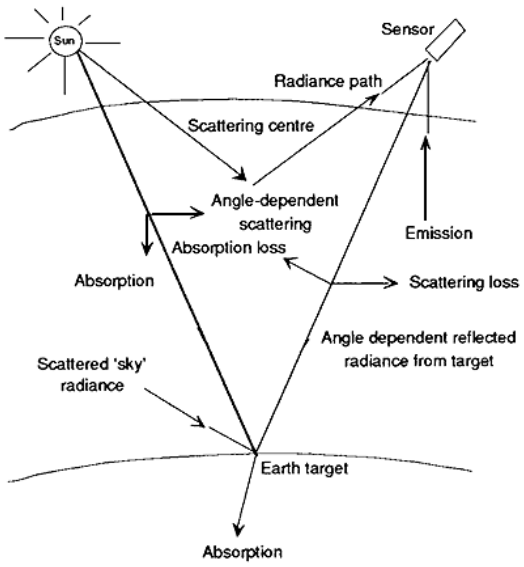


Figure 8. Physical processes involved in generating hyperspectral data (including interaction between solar radiation and atmosphere, interaction between solar radiation and targets on the ground, and sensor responsiveness)

inherently nonlinear [40] [41]. The nonlinearity in hyperspectral data is attributable to multiple sources involved in the process of data formulation, as depicted conceptually in Figure 8. Multiple scattering between solar radiation and targets coupled with the variations in sun-target-sensor geometry is often considered as the primary source of nonlinearity [42] [43]. The presence of a nonlinear attenuating medium, such as water, is another source of nonlinearity [44]. The heterogeneity of pixel composition contributes to the nonlinearity as well [45]. On the other hand, hyperspectral data have long been modeled and processed using algorithms based on the assumption that hyperspectral data originated from some linear stochastic processes, which are defined as summations of collections of random variables [46]. For example, developers of the Maximum Noise Fraction (MNF) [38], a popular feature extraction algorithm, treated hyperspectral data as realizations of a random process whose underlying random variables are linearly correlated. Orthogonal features are then derived by linearly transforming the original spectral features (channels). Linear spectral unmixing algorithms have been increasingly used to process hyperspectral data for retrieving sub-pixel information, such as endmember spectral signatures and endmember spatial compositions within each pixel. These algorithms are developed based on the hypothesis that a pixel spectrum is a linear combination of a set of endmember spectra [47]. Spatial and textural information derived from remotely sensed imagery has been utilized to improve classification accuracy obtained by spectral based classifiers [48]. The methods for extracting the spatial and textural information, such as spatial autocorrelation, however, only work under the assumption that the data under investigation are linear.

The inconsistency between data statistical characteristics and the conceived algorithms for information extraction has been realized in the remote sensing community, which has stimulated the development of new algorithms without assuming linear data distribution or explicitly addressing the nonlinearity. As one of such efforts, Zhong et al. (2006) proposed in [49] a novel algorithm, called unsupervised artificial immune classifier (UAIC), to label multi/hyperspectral remotely sensed images. In contrast to the conventional statistical classifiers, including K-means and ISODATA, which assume data to be linear stochastic, the UAIC is a nonlinear model that is data-driven and self-adaptive without requiring any form of data distribution. They reported that UAIC outperformed the conventional methods in classification accuracy. Based on support vector data description, Banerjee et al. [50] developed a kernel method to improve the conventional anomaly-detection algorithms that require the local background to be Gaussian and homogeneous, which is often violated in reality. The proposed algorithm was nonparametric and capable of modeling non-Gaussian background clutter. It was found that compared to the conventional methods, the proposed method markedly reduced the number of false alarms in anomaly detection. Bachmann et al. [40] [41] proposed an algorithm of feature extraction for hyperspectral remotely sensed data. The proposed algorithm, based on isometric mapping (ISOMAP), addressed the nonlinearity in hyperspectral data using geodesic distance, upon which a manifold coordinate system was derived. They demonstrated that, compared to the features derived through linear feature extraction algorithms, the features derived using the proposed algorithm had a better separability between class spectra and contributed to a more accurate classification. Han and Goodenough [51] proposed another nonlinear feature extraction algorithm based on Locally Linear Embedding (LLE). This algorithm addressed the nonlinearity by preserving the local neighborhood relationship, through which a new manifold was derived. Compared to the linearly extracted features, the features extracted this way produced improved separability among endmember spectra and better spatial information preservation. This work will be discussed in more detail in Chapter 5 of this dissertation. Mohan et al. [52] extended the neighborhood calculation in LLE to incorporate information of spatial coherence. They reported that the classification accuracy was about 10% higher with the 25 extracted features than that with the same number of features

derived from Principal Component Analysis (PCA).

The studies of nonlinearity in hyperspectral data usually commence by demonstrating the nonlinearity in terms of scatter plots built-up from multiple channels of the hyperspectral imagery. The existence of nonlinear features in the scatter plots are then considered as the indication of nonlinearity. This expression of nonlinearity, however, may not be optimal or complete. Firstly, to show the nonlinear features in scatter plots, the channels used to generate the plots have to be selected far apart spectrally from each other. Otherwise nonlinear features might not be discernable due to the high correlation among neighboring channels. This may cause the illusion that the nonlinearity only exists between well-separated channels in hyperspectral imagery. Secondly, the nonlinear features exhibited in scatter plots only represent the nonlinearity in the spectral domain of hyperspectral imagery, i.e. the between-channel nonlinearity. The nonlinearity may also exist in the spatial domain, i.e. the within-channel nonlinearity. Thirdly, scatter plots only provide a visual expression of nonlinearity. Other expressions, especially those related to data statistical properties, are more desirable as they can be used analytically to investigate the nonlinearity in hyperspectral data.

The primary objective of this chapter is to develop a quantitative method so that it can be used in a complementary way with the scatter plot to demonstrate the existence of nonlinearity in hyperspectral data. The initial effort was reported in [53], where a nonlinear time series analysis method was proposed to investigate the existence of nonlinearity in the hyperspectral data represented by an AVIRIS image. The study concluded that using high order autocorrelations as the discriminating statistic, nonlinearity was found markedly in the spectral domain of the hyperspectral data, but not in the spatial domain.

To improve and expand the initial effort reported in [53], this chapter introduces a new discriminating statistic and employs a more formal approach to evaluate the differences between the real hyperspectral data and the surrogate-derived discriminating statistical values. Specifically, compared to the previous work, this study strengthens the following aspects. (1) Linear stochastic data (Gaussian) are created and used as the

simulated data series to ensure the process of surrogate generation does not introduce any nonlinearity. (2) Both AVIRIS and Hyperion data are considered so that the conclusions drawn in this study are acquisition-platform independent. (3) More hyperspectral pixels are selected to generate testing data series so that the chance of a faulty statistical test is reduced. (4) Spectral angles between spatially adjacent pixels are calculated as the new discriminating statistic. (5) The statistical test based on the simulated data is conducted prior to the test using the real hyperspectral data. This reduces the possibility of false rejection of null hypothesis that may happen in the statistical test if surrogate data are not generated properly or the discriminating statistics are wrongly selected. (6) In addition to the rank ordering, an approach based on confidence interval is employed to examine the differences between the real data and the surrogate-derived discriminating statistical values.

2.1 Data sets

Hyperspectral data sets chosen in this study are an AVIRIS and a Hyperion image, as shown in Figure 9 (a RGB composite of 3 channels around 1500 nm, 750 nm, and 650 nm, respectively, the same data sets shown in Figure 7). The atmospheric correction was conducted using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) software [54]. Pixels of the following two land-cover types are considered: forest and water, as specified in Figure 9. Both of them are believed to be highly nonlinear. The forest is the primary land-cover of interest in this dissertation, whose nonlinearity is due mainly to the strong multiple scattering occurring within the forest canopy. Since water attenuates the intensity of light traveling through, it causes a strong nonlinear effect as well among water pixels.

To test the performance of the discriminating statistics and the correctness of the surrogate generation method (so that the false rejection of null hypothesis can be avoided), simulated data series were generated using the random number generator provided in IDL [55]. The simulated data series were created such that each of them was normally distributed. They could be considered as the realization of some linear stochastic processes, which is in close compliance with the null hypothesis of the statistical test (more on this in next section).

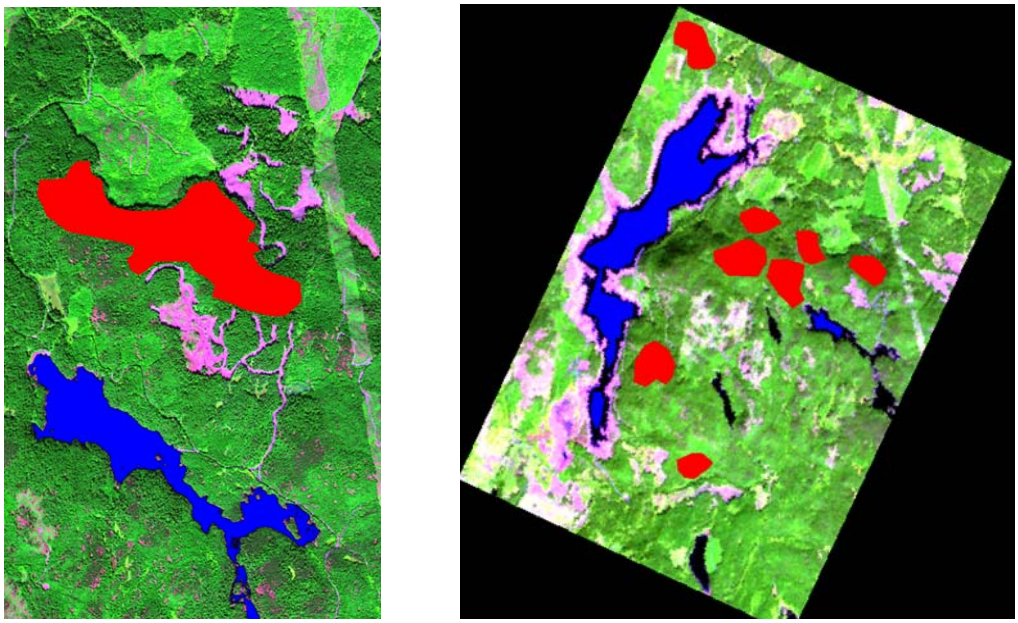


Figure 9. Areas of interest in AVIRIS (left) and Hyperion (right), where forest pixels selected are marked in red and water pixels are marked in blue.

2.2 Algorithm description

Before delving directly into the investigation of nonlinearity using a statistical test based on surrogate data [39], it may be necessary to have a brief introduction about this test, as the surrogate based statistical test is not commonly used in the remote sensing community. Originally this method was "an indirect approach to investigate the existence of nonlinear dynamics in time series by means of hypothesis testing using surrogate data" [56]. There were four components required to complete the test, including:

- a null hypothesis H_0
- a collection of data series D, upon which H_0 is to be tested
- a collection of surrogate series S_i for $i = 1,...,m$, which were generated corresponding to D and were in compliance with H_0
- a discriminating statistic q

where H_0 was defined in terms of the statistical process through which the data series D was produced; D was organized as an $l \times m$ matrix, representing m data series, each of

which had the length of l; S_i, corresponding to the data series i in D, was given as a matrix as well with the size of $l \times h$, where $h > 1$, indicating that each data series in D had more than one surrogate series corresponding to it; finally q was a numeric estimator that was used to quantify the characteristics of the data series and the corresponding surrogate series.

Surrogate series, as the name suggests, are representatives of the data series in some aspects specified by the null hypothesis H_0 to be tested. For example, if H_0 is stated as: the data series are linear stochastic, then the corresponding surrogate series are created so that they are realizations of some linear stochastic processes. The surrogate series have the same linear statistical properties, including histogram, autocorrelation, and power spectrum, as those possessed by the data series. This can also be phrased as: the surrogate series are created in compliance with the null hypothesis and they are identical to the data series from the perspective of linear statistics.

Once the above four components are all in place, the statistical test based on surrogates can be carried out by calculating the discriminating statistic q using the data series D and the companion surrogate series s_i. For each data series, one q-value is calculated. As there are h surrogate series associated with each data series, h q-values are derived from the surrogate series. Then the q-value calculated based on the data series and those from the surrogate series are compared. If the single q-value from the data series is not different enough from the others derived from the surrogate series, the null hypothesis is accepted, or equivalently, the data under investigation are linear stochastic. Otherwise it indicates that a discrepancy exists between the data series and the corresponding surrogates. As the data and the surrogates are similar in linear properties, the discrepancy is only possible due to the causes other than linearity, which leads to rejecting the null hypothesis.

At this point, however, the evidence is not sufficient enough to claim the existence of nonlinearity, even though the null hypothesis of linearity has been rejected. In the circumstance of hyperspectral remote sensing, additional evidence exists to support

the claim. For example, the physical processes involved in hyperspectral data formation suggest that hyperspectral data contain nonlinearity, which can be partially shown using scatter plots. Hopefully incorporating the additional evidence and the rejection of the null hypothesis will provide us with sufficient evidence to conclude that nonlinearity exists in the data series being tested. The full description regarding each of these components tailored to hyperspectral data is given in the following sub-sections.

A. Determination of the null hypothesis H_0

The null hypothesis should be specified according to the characteristics of the data series under investigation and the objective of the statistical test. An improperly specified null hypothesis may result in a meaningless test, regardless of whether the null hypothesis is rejected or sustained. As the existence of nonlinearity in hyperspectral data is the only interest of this study, the null hypothesis is stated as follows: "***The hyperspectral data represented by the 4m AVIRIS and Hyperion imagery are realizations of some linear stochastic processes***."

B. Creation of data series D

Five data sets were created as the data series to be tested, each of which had the size of $l \times m$, representing m series with the length of l each. The number and the length of the data series were chosen arbitrarily, just to ensure that the series were populous and long enough to conduct a meaningful statistical test (more on this in Section 2.3). The 1st set of the data series had the size of 17900 × 300, which is the simulated data as described in the previous section. The data series in this set were true realizations of a linear Gaussian process - a simplified version of a linear stochastic process, produced by a Gaussian random number generator [57].

The 2nd set (17900 × 300) was composed of forest data series, which was prepared using the AVIRIS data (left image in Figure 9). Each of the data series was created by concatenating the selected forest pixels in the spectral direction; i.e. the rear element of pixel (vector) j is connected to the head element of pixel (vector) $j+1$ and so on, where pixel j and $j+1$ are spatially immediate neighbors. The same method was used in [51] to create the spectral series. This arrangement maintained the pixel spatial connectivity as

much as possible when one transformed the 3D hyperspectral data (image cube) to 1D (series). The number of pixels to be concatenated relies on the desired length of the data series. To match the length of the data series (17900), 100 pixels were connected this way to make one data series. As 300 data series were needed, it consumed 30000 forest pixels in total. The 3rd set (17900 × 300) was composed of the water data series, which was also prepared with the AVIRIS data. Each of the data series was produced in exactly the same way as was the 2nd set (forest data series) by connecting water pixels (left image in Figure 9). The 4th and 5th data sets were created in the same way as the 2nd and 3rd sets were made, respectively, but using Hyperion pixels (right image in Figure 9). Due to the limitation in the number of pixels in the Hyperion image, the 4th and 5th sets were made smaller, each of which had a size of 8950 × 100. Each series was made by connecting 50 Hyperion pixels.

C. Creation of surrogate series s_i

As discussed previously, surrogate series need to satisfy the following two conditions. Firstly they have to be consistent with the null hypothesis to be tested. Secondly, they have to be similar enough to the companion data series in linear stochastic properties, including histogram, autocorrelation, and power spectrum. Meeting these conditions ensures that the surrogate series maintains all the linearity of the data series but are free of any nonlinearity that may or may not exist in the real data series.

A few algorithms have been developed to produce such surrogates based on the given data series. The earliest and best-known is called “Amplitude Adjusted Fourier Transform (AAFT)”, developed by Theiler et al. [57]. This algorithm includes the following five steps of data manipulation: 1) static transformation that “Gaussianizes” (makes Gaussian) the data series to be tested; 2) discrete Fourier Transform on the “Gaussianized” data series; 3) randomization of the phase coefficients of the above transform; 4) inverse Fourier Transform using the randomized phase coefficients and the unchanged amplitude coefficients; 5) inverse of “Gaussianization” transformation of step 1 applied to the results of step 4. Finally the outputs of step 5 are the desired data series that satisfy the above two conditions of being the surrogate series corresponding to the given data series. Among these steps, 1 and 5 are designed to match the data distribution

between the surrogate and the data series, while step 3 is to ensure that the generated surrogates are free of nonlinearity.

Kugiumtzis and Palus, however, found in [56] and [58], respectively, that AAFT does not match well the linear correlations between the data series to be tested and the corresponding surrogates, which may lead to a false rejection of the null hypothesis . Schreiber and Schmitz developed an alternative algorithm, called iterative or improved AAFT (IAAFT), in [59]. Instead of matching at the same time the power spectrum and distribution between the data series and the surrogates as was done in AAFT, they tried to achieve the matching separately and iteratively. It was shown in [59] that surrogates produced by IAAFT caused less false rejection than those produced by AAFT during a comparative statistical test using the same data series. IAAFT has been implemented as a tool for nonlinear time series analysis [60].

Using IAAFT, 100 surrogate series were generated for each data series included in the five data sets: simulated, AVIRIS-forest, AVIRIS-water, Hyperion-forest, and Hyperion-water, i.e. $h = 100$ (h is defined in Section 2.2). Though it is suggested in [61] that any $h > 30$ is considered sufficient for a statistical test, more surrogates for each data series were made here to improve the statistical significance of the test. Therefore the size of s_i is 17900×100 for each of the first three data sets where $i = 1,...,300$, and 8950×100 for each of the last two sets where $i = 1,...,100$.

D. Determination of discriminating statistic q

As the measure of nonlinearity, the discriminating statistic q is used to distinguish the data series and the corresponding surrogates. Ideally the q-values derived from the surrogates (free of nonlinearity) should be markedly different from those derived from the data series to be tested if nonlinearity exists. Based on the discussion in [62], the following two discriminating statistics were initially considered:

$$q^1(\tau) = < x_i x_{i-\tau} x_{i-2\tau} > \qquad (2.1)$$

$$q^2(\tau) = < (x_i - x_{i-\tau})^3 > \qquad (2.2)$$

where x_i is the data point at location i of a data series; τ is the time lag, representing the

wavelength difference in the spectral data series, and the angle brackets represent average operation. As the relationship between adjacent channels is of interest, τ is set to 1. Both of these discriminating statistics are high-order autocorrelations, which are inexpensive to calculate.

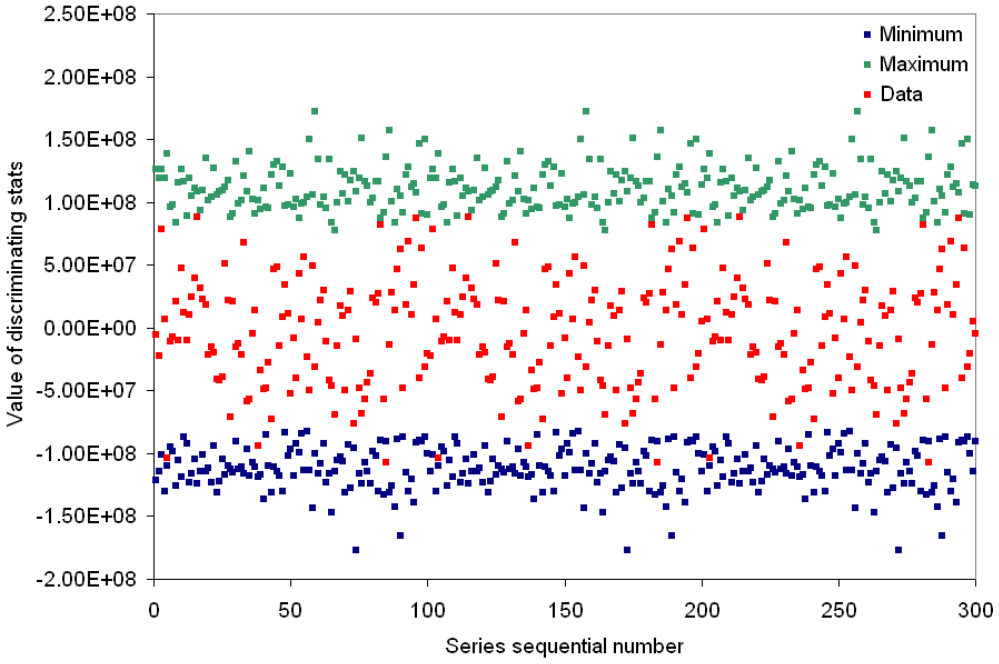
It was found in [53], however, that the 1st discriminating statistic above had less discriminating power than the 2nd one to distinguish between data series and the corresponding surrogates. Using an identical data series (spectral series), the 2nd discriminating statistic clearly depicted the difference between the data series and the surrogates, while the 1st one barely showed the difference. For this reason, the 2nd one is chosen in this study.

Although the autocorrelation-based discriminating statistics are able to show the nonlinearity in the spectral domain of the hyperspectral data in [53], it is of concern that they do not have any physical meaning in the sense of remote sensing. For example, the 2nd discriminating statistic is just an averaged cubic difference of a pixel between two adjacent hyperspectral channels. It is desirable to have an entity that is both physically meaningful and able to show nonlinearity. Spectral angle is a good candidate for this task. Firstly, it is a meaningful metric which is capable of minimizing the spectral shape variation associated with the different illumination and viewing geometry. It is related to many hyperspectral applications, including classification, target identification, and spectral unmixing. Secondly, the spectral angle is nonlinear itself and is sensitive to nonlinear changes. Lastly, spectral angle is calculated as a vector dot product that is less computationally expensive. The spectral angle between two pixel vectors, $\vec{x}_i$ and $\vec{x}_j$, is determined by the following formula: $\theta = \arccos\left(\frac{\vec{x}_i{}^T \vec{x}_j}{\| \vec{x}_i \| . \| \vec{x}_j \|}\right)$. Considering these advantages, spectral angle was selected as a new discriminating statistic.

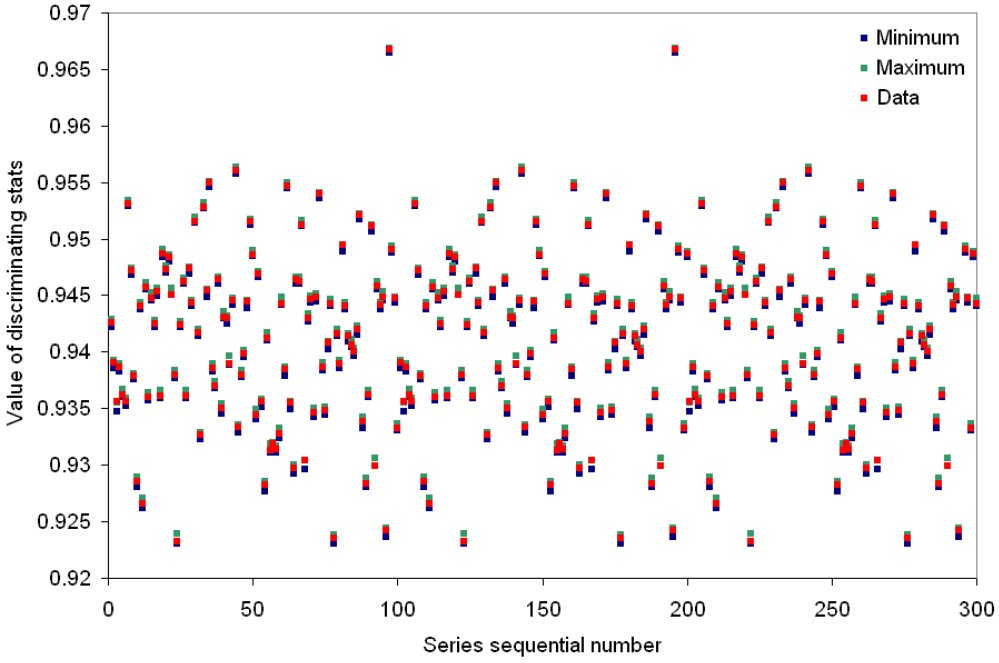
2.3 Results and discussion

One of the issues associated with the statistical test using surrogates is false positives. That is, the null hypothesis H_0 is rejected but there is no nonlinearity in the data series being tested. This is often due to the poor choice of the discriminating statistic q or

the improperly generated surrogate series s_i for $i = 1,...,m$. To mitigate false positives, prior to using real hyperspectral data, a test was conducted using the simulated data, which were created using the method described in Section 2.1. As the simulated data



(a)



(b)

Figure 10. Discriminating statistical values calculated in terms of (a) high order autocorrelation and (b) cosine of spectral angle, based on simulated data series and the corresponding

series are from a pure linear stochastic process, the discriminating statistic values calculated based on the data (simulated) series should not be different from those based on the corresponding surrogates, if both the discriminating statistic and the surrogate data are correct. The null hypothesis is therefore sustained. If this is not the case, there is something wrong with either the discriminating statistic or the surrogates, or both. As shown in Figure 10a, where the discriminating statistic values were calculated as the high order autocorrelation (Equation 2), all of the discriminating statistic values of the 300 simulated series fall in the range of variation (indicated as the minimum and maximum) defined by the discriminating statistic values of the corresponding surrogates. This means that these values are not separable statistically and the null hypothesis prevails; i.e. the simulated data series are from linear stochastic processes. The discriminating statistic values calculated using the cosine of the spectral angle between adjacent pixels are shown in Figure 10b. Similar to the previous result, it was found that the simulated series and the corresponding surrogates are also indistinguishable without exception. This extra step prior to the statistical test using real hyperspectral data demonstrated that the discriminating statistics and the surrogate series created by IAAFT are appropriate and they do not cause any false positives.

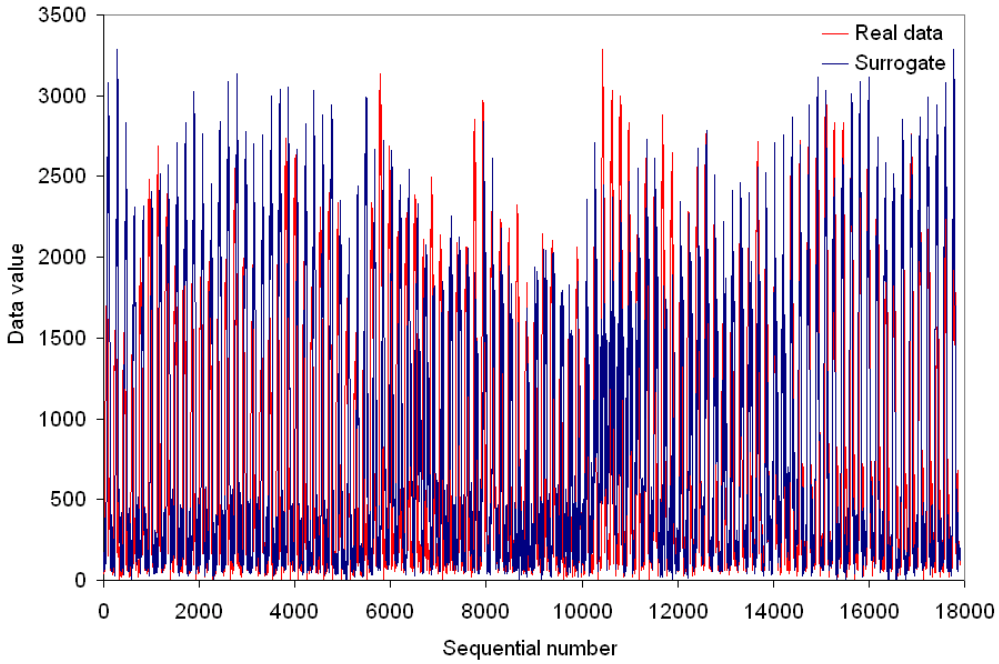


Figure 11. A generated AVIRIS-forest data series and one of its surrogate series.

As discussed in Section 2.2B above, spatial neighboring forest and water pixels were concatenated, respectively, in the spectral direction to generate forest and water data series. For each data series generated, 100 surrogate series were created. Figures 11 and 12 are examples showing these data series created using AVIRIS pixels and their corresponding surrogates. It is demonstrated that the surrogates look very similar to the corresponding data series in amplitude, profile, and distribution. Actually this similarity is guaranteed because of the constraints imposed, including the same autocorrelation, power spectrum, and histogram, between each data series and its surrogates. It should be noted that the data series and the corresponding surrogates are similar but not identical. The difference between them is likely due to the existence of nonlinearity, which is to be proved in this chapter.

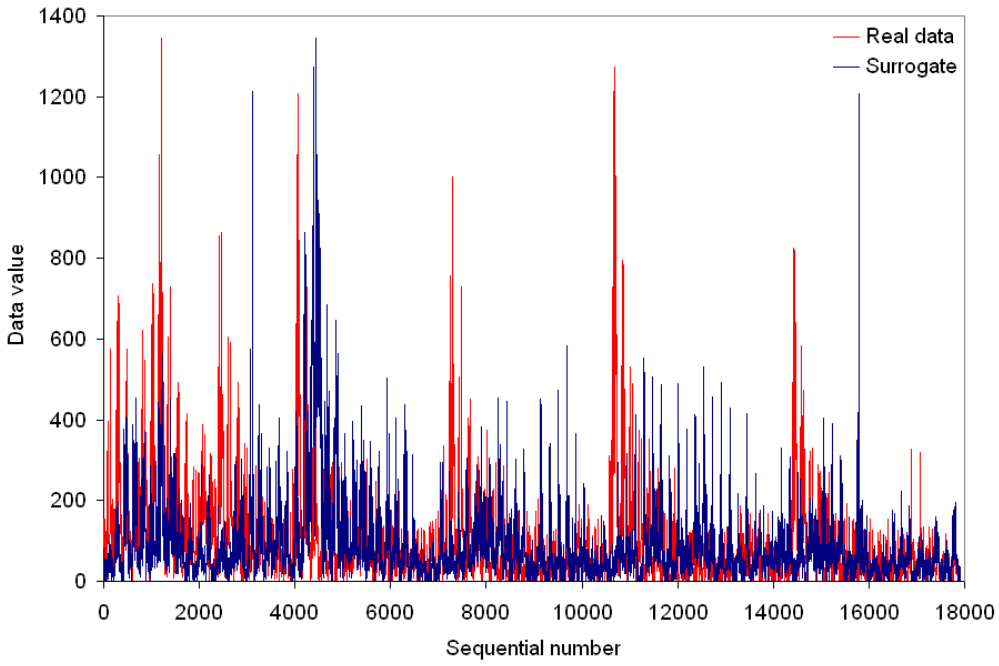
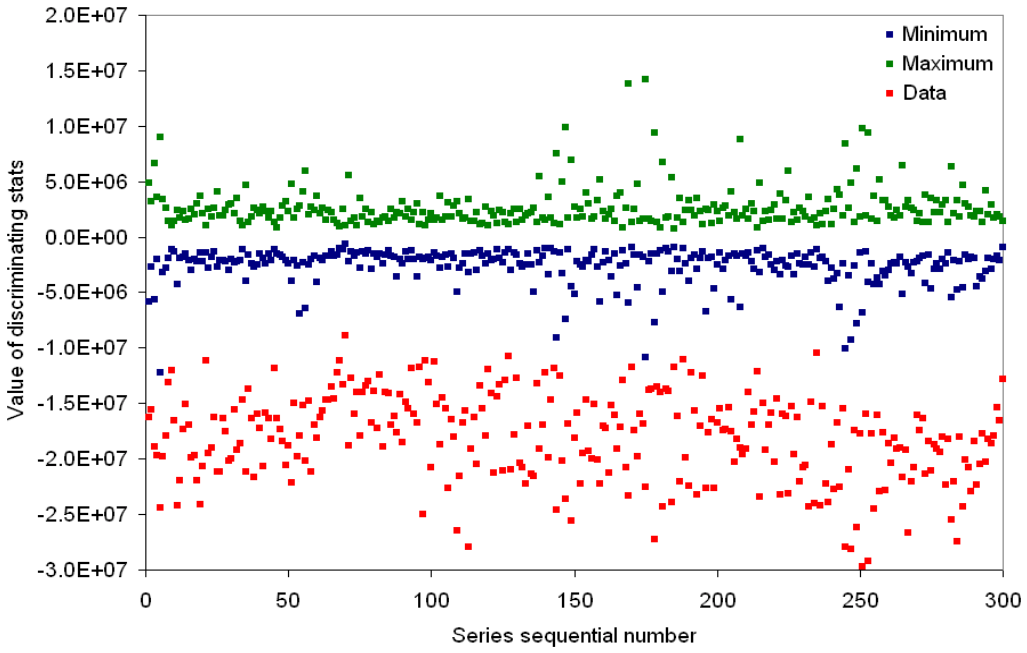


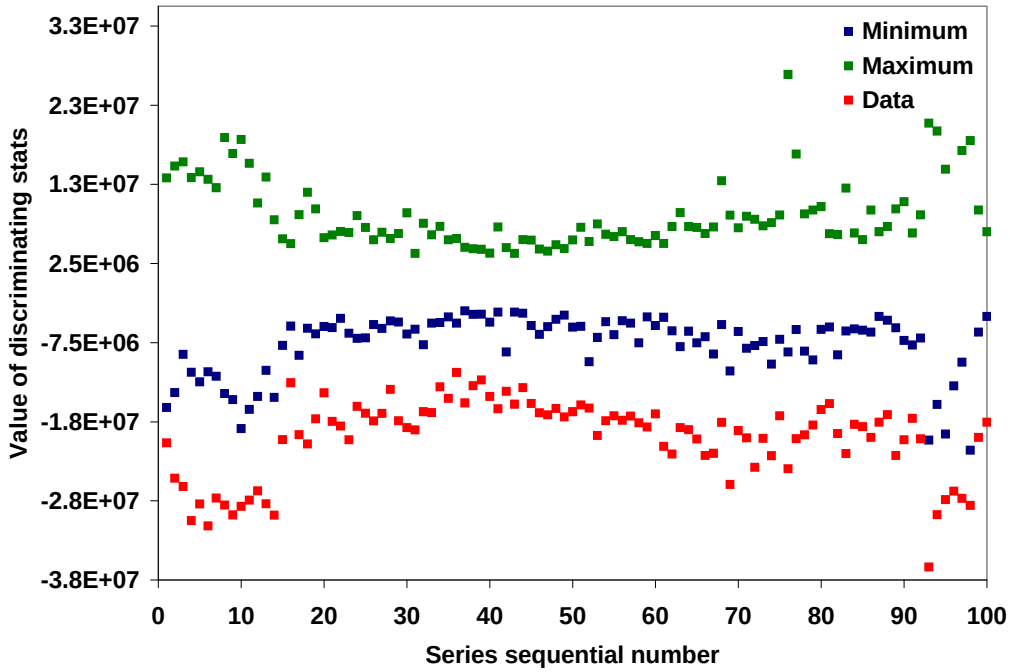
Figure 12. A generated AVIRIS-water data series and one of its surrogate series.

The discriminating statistic, either based on Equation 2.1, 2.2 or the spectral angle, is a single numeric estimator for each data series or surrogate. As there were 100 surrogate series corresponding to each data series, 100 discriminating statistic values were calculated with these surrogates, which defined a range of variation, upon which the discriminating statistic value from the data series was to be examined. This could be done with the following two methods: rank ordering and statistical significance [63]. To

facilitate describing the rank ordering method, let q^0 be the discriminating statistic value from a data series; q^i be the discriminating statistic value from one of the corresponding



(a)



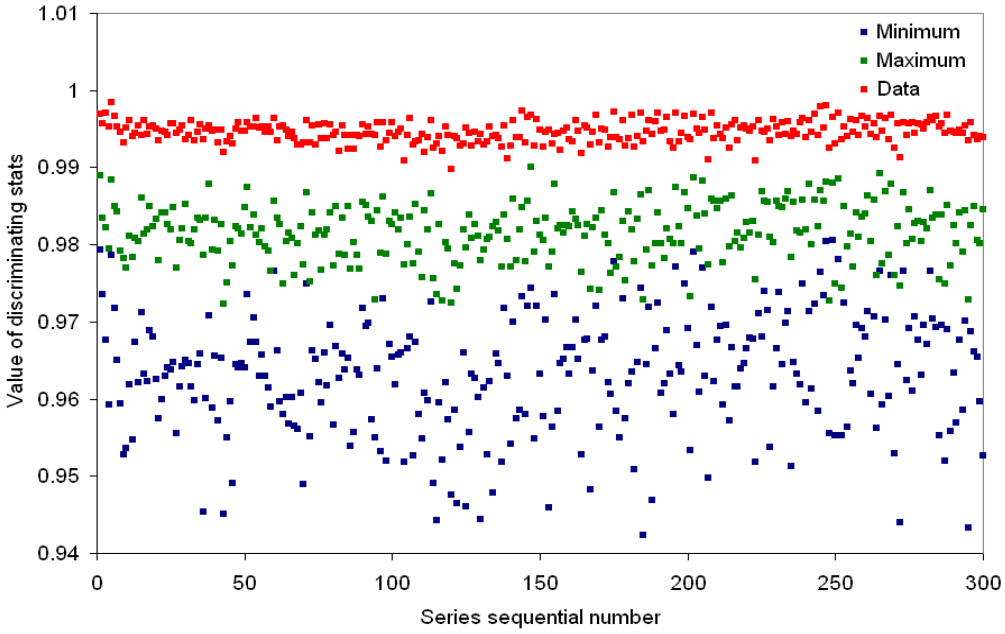
(b)

Figure 13. Discriminating statistical values calculated in terms of high order autocorrelation based on forest data series and the corresponding surrogates: (a) AVIRIS (b) Hyperion

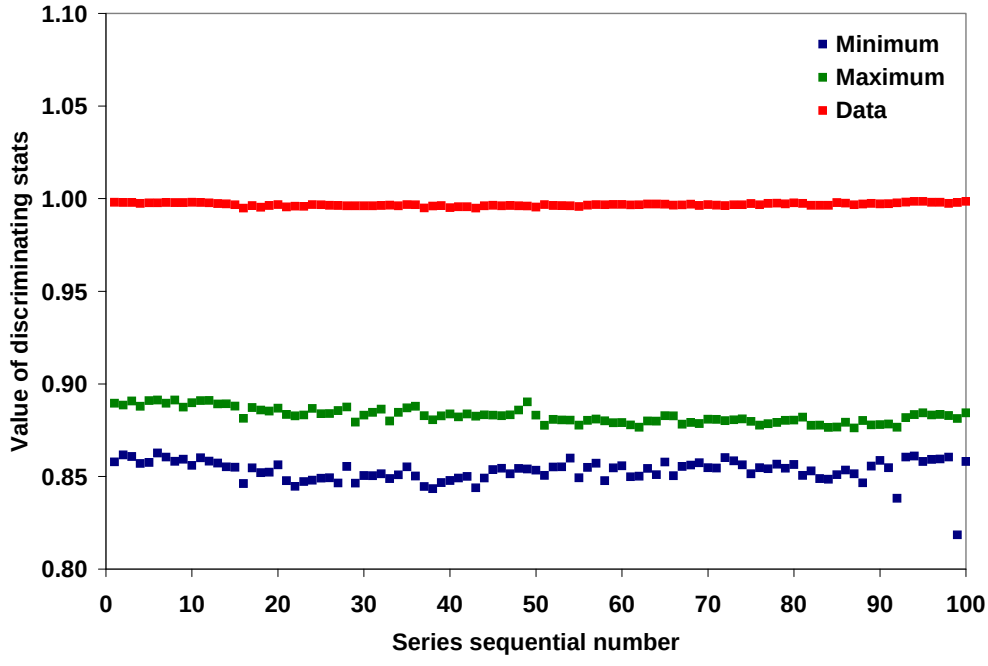
surrogate series, where $i = 1,...,100$. Let $q^{\min} = MIN\,(q^1, q^2, ..., q^{100})$ and $q^{\max} = MAX\,(q^1, q^2, ..., q^{100})$. If $q^{\min} \le q^0 \le q^{\max}$, then it indicates that the discriminating statistic value based on the data series is indistinguishable from those derived from its surrogates. This is the case for the test conducted using the simulated data series as shown in Figure 10.

The forest data series were tested first using the rank ordering method. The results are shown in Figures 13 and 14, representing the discriminating statistic values calculated using the high order correlation (Equation 2.2) and the spectral angle, respectively. It was found that for all the forest data series (both AVIRIS and Hyperion), using the high order autocorrelation as the discriminating statistic, $q^0 < q^{\min}$ held without exception (Figure 13). Using the spectral angle as the discriminating statistic, it was found that $q^0 > q^{\max}$ was always true as shown in Figure 14. For all the cases above, q^0 always stayed outside the range defined by ($q^{\min}, q^{\max}$). The discriminating statistic values obtained based on the forest data series were easily separated from those calculated using the corresponding surrogates. Thus, the null hypothesis H_0 was rejected. Linear stochastic processes cannot fully explain the characteristics of the forest data series. Therefore, the forest hyperspectral data cannot be produced from linear stochastic processes.

For the water data series, the situation became a little intricate. Figures 15 and 16 show the discriminating statistic values calculated in terms of the high order correlation (Equation 2.2) and the spectral angle, respectively. For both the AVIRIS- and Hyperion-based water series q^0 did not always stay within or outside ($q^{\min}, q^{\max}$), though it was outside for most cases. It is due to the fact that water is a strong attenuator across the entire wavelength, which greatly reduced the reflective energy. This caused water spectra to be extremely low in magnitude and the nonlinearity was weakened by noise. Using the rank ordering method described above, it was not possible to decide whether q^0 and q^i were separable confidently and whether the null hypothesis was accepted. The other method: statistical significance, therefore, was employed to evaluate the difference between the discriminating statistic values from the water series and their surrogates. The



(a)

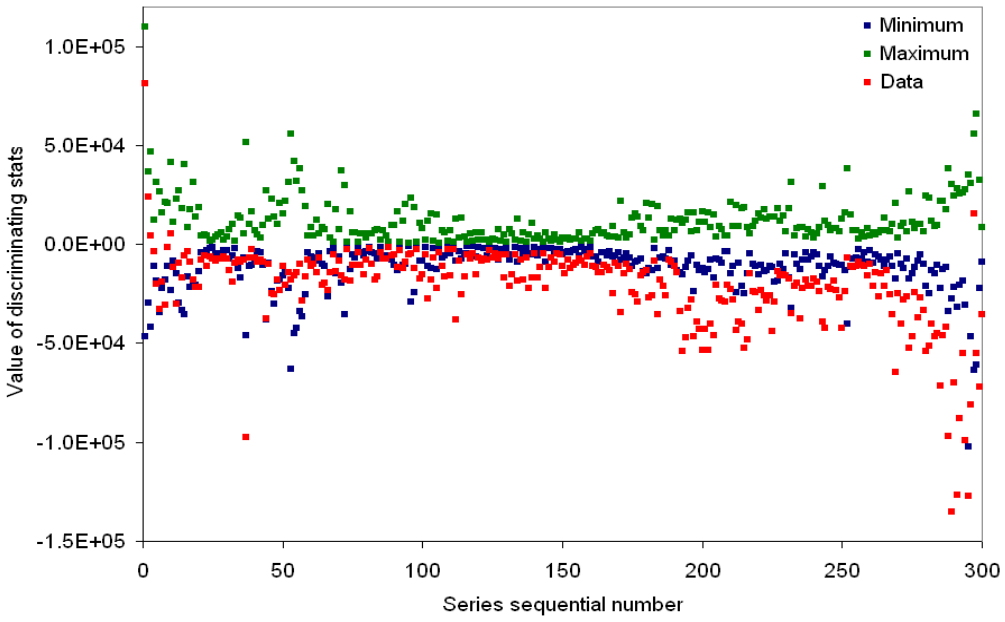


(b)

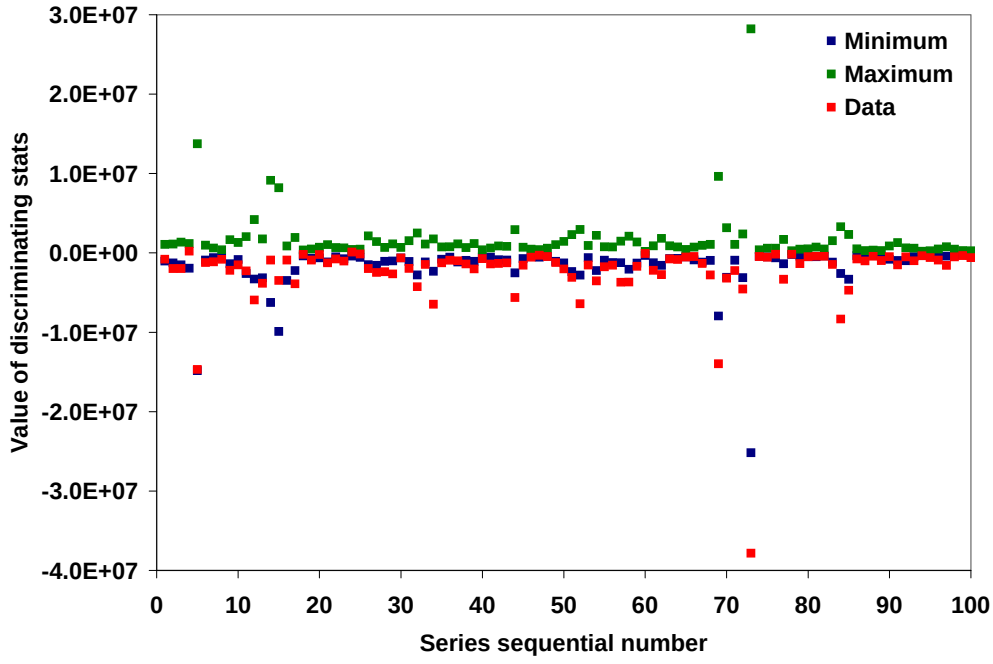
Figure 14. Discriminating statistical values calculated in terms of cosine of spectral angle based on forest data series and the corresponding surrogates: (a) AVIRIS (b) Hyperion

statistical significance s is calculated as follows:

$$s = \frac{| q^{0} - q^{mean} |}{q^{std}} \qquad (3)$$



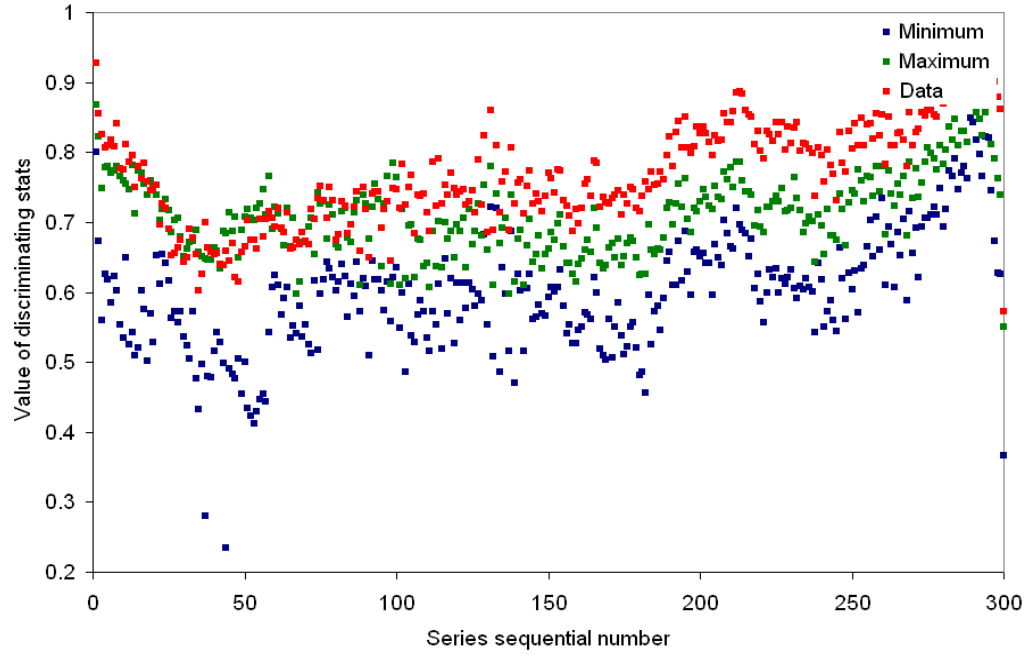
(a)



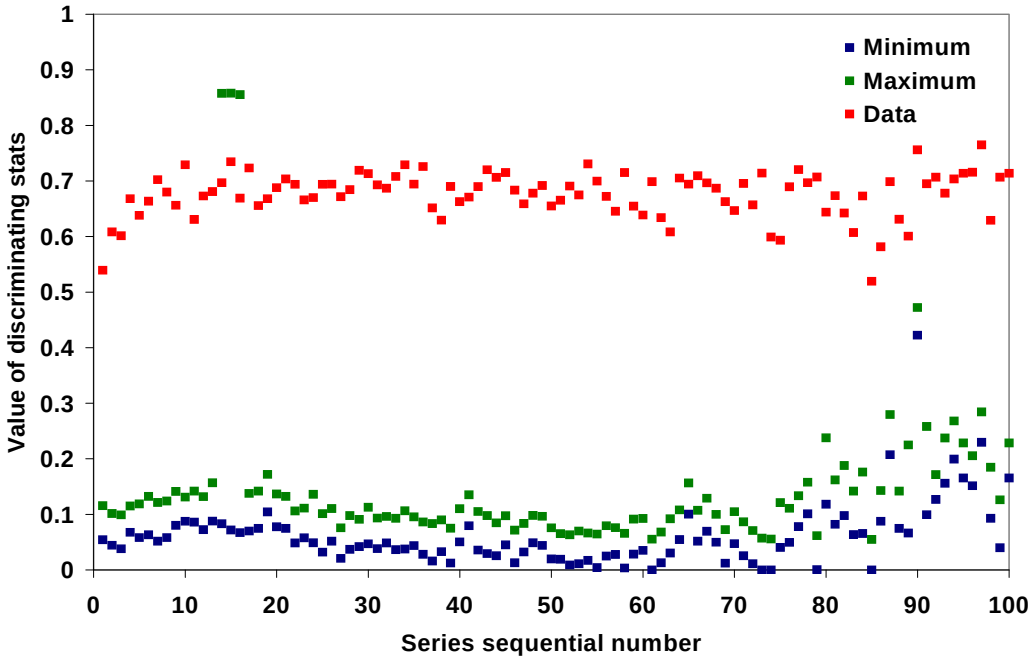
(b)

Figure 15. Discriminating statistical values calculated in terms of high order autocorrelation based on water data series and the corresponding surrogates: (a) AVIRIS (b) Hyperion

where q^{mean} is the average of q^i for $i = 1,...,100$ and q^{std} is the standard deviation. If $s \geq 2$, then q^0 is separable from the q^i s and the null hypothesis is rejected at the 95% level of confidence, as q^0 is located outside two standard deviations of q^i s. If $s < 2$, the



(a)



(b)

Figure 16. Discriminating statistical values calculated in terms of cosine of spectral angle based on water data series and the corresponding surrogates: (a) AVIRIS (b) Hyperion

separation of q^0 and q^i s is insignificant statistically. To apply this method, however, the q^i s need to be normally distributed. To check the distribution, skewness and kurtosis were employed, where the skewness is a measure of symmetry of data distribution and

the kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. The results of the calculation are given in Table 1. Both the skewness and

Table 1 Distribution and separation of the discriminating statistic values based on water series and surrogates

Water series	Discriminating stats	Skewness	Kurtosis	Significance
AVIRIS	High order autocorrelation	0.02	0.46	2.6
AVIRIS	Cosine of spectral angle	0.05	0.37	2.8
Hyperion	High order autocorrelation	0.82	2.92	2.1
Hyperion	Cosine of spectral angle	0.75	3.68	2.3

kurtosis were fairly small based on the water series and their surrogates, indicating that the distribution was not a perfect Gaussian, but close enough (both of them are smaller than 5). The significance calculated is greater than 2 for all cases, indicating that the hypothesis can be rejected at 95% level of confidence.

The significance was also calculated, for comparison purposes, using the forest data series and the corresponding surrogates. Compared to the significance calculated using water data, they were much larger, 3.8 and 3.9 for AVIRIS and 3.8 and 4.0 for Hyperion, respectively. This means that for the forest data series, q^0 is located outside 3 standard deviations of the q^i s. Therefore, the null hypothesis could be rejected at the 99% level of confidence. It was interesting that the forest data series exhibited stronger nonlinearity than the water data series, though they were both considered highly nonlinear. A possible explanation is that the signal-to-noise ratio (SNR) of water pixels is low, due to water strongly absorbing the incoming radiation, that the nonlinearity may be more affected by noise.

Compared to the forest data series, the water data series were more difficult to distinguish from their surrogates. This was likely due to the following two reasons: 1) Both AVIRIS and Hyperion are general-purpose instruments, but were developed having land-applications in mind [64]. They were challenged by the low signal intensity and

other water-related constraints in the aquatic environment, which may have prevented them from capturing the complete information to describe the spectral properties of inland and coastal water, including nonlinearity. Hyperspectral data collected by the dedicated sensors for aquatic studies should be tested. It would be of interest to repeat this test when such hyperspectral data are available. 2) Both data series of water and forest were generated using $\tau = 1$ (Equation 2.2), which is good to address the nonlinearity between adjacent channels. But the correlation between nearby channels of hyperspectral data is strong. It may obscure the nonlinearity from being detected. To determine the dependence between τ and the separation of the data-based and surrogate-based decimating statistical values, the entire process was repeated using $\tau = 2$ (Equation 2.2) for both AVIRIS-forest and water series. The series generated this way represented relationships between non-adjacent channels where the correlation was less significant. It was found, however, that the separation was barely changed. For the AVIRIS-forest series, the significance of separation was 4.0 and 3.8, calculated using the high order correlation and spectral angle, respectively. For the AVIRIS-water series, the corresponding numbers were 2.6 and 2.8, respectively. It seems that the separation is independent of τ, indicating that the between-band correlation is not a factor influencing the statistical test.

It was also noticed that the discriminating statistic values calculated using Hyperion series were relatively smoother compared to those based on the AVIRIS data (Figures 13 – 16). This was likely attributable to the differences in pixel size. A Hyperion pixel is 900 m^2, while a 4m-AVIRIS pixel is only 16 m^2. The larger pixel size may intensify the averaging effect on surface reflectance and produce uniform discriminating statistic values. This study demonstrated that, given this spatial averaging effect, hyperspectral data still exhibit significant differences in statistical properties from the linear stochastic data.

It should be noted that the rejection of the null hypothesis here is not equivalent logically to the existence of nonlinearity in hyperspectral data. What one can infer from the above statistical test is that hyperspectral data are not out of some linear stochastic processes. The statement that nonlinearity exists in hyperspectral data can only be

justified by incorporating the results of this test with other supporting evidence, including the nonlinearity suggested by the physical processes involved in the formation of hyperspectral data, existence of nonlinear media, and the nonlinearity demonstrated partially using scatter plots.

2.4 Chapter summary

Nonlinearity in hyperspectral data was investigated using a 4m AVIRIS and a Hyperion image. The investigation was focused on the spectral domain nonlinearity associated with specific land-cover types of interest: forest and water. Based on the previous studies, this study was conducted by incorporating new data sets, introducing a meaningful discriminating statistic metric, and using a more formal approach to evaluate the differences between the real data and surrogate-derived discriminating statistic values. Using the simulated data series and corresponding surrogates, it was demonstrated that both the discriminating statistics and the method of surrogate generation are working correctly. The null hypothesis that hyperspectral data were linear was rejected at the 99% level of confidence for the forest data series and at the 95% level of confidence for the water data series. The forest data series were found to exhibit stronger nonlinearity than the water data series did.

The nonlinearity identified in this study suggested that caution should be exercised when one applies linearity-assumed methods directly to hyperspectral data, as these methods do not consider explicitly or may not be capable of dealing with the nonlinearity. It should be noted also that the conclusions reached in this chapter are based on a 4m AVIRIS and a Hyperion image only. Further studies are encouraged to investigate the nonlinear issues in hyperspectral data collected by more hyperspectral sensors over diverse land-cover types.

Chapter 3

Nonlinearity-counted dimensionality estimation

Accurate estimation of dimensionality is a prerequisite step prior to many applications of using hyperspectral and other optical remotely sensed images. For example, the classification-based applications need to determine the dimensionality so that the corresponding number of features can be extracted and fed into the classifiers for the classification. The spectral unmixing-based applications need to estimate the dimensionality to determine how many endmembers exist in the image. Dimensionality estimation of hyperspectral imagery is usually conducted through linear transformations, including, among others, Principal Component Analysis (PCA), Maximum Noise Fraction (MNF), Projection Pursuit (PP), and Wavelet Transform. These approaches, though manifested in different mathematical forms, are based on treating hyperspectral images as the data sets produced by linear stochastic processes. In other words, hyperspectral data are assumed to be realizations of some linear stochastic processes that are free of nonlinear and dynamical variations. This assumption, however, may contradict the following fact: the formation of hyperspectral data involves several physical processes that result in nonlinearity. The existence of nonlinearity in hyperspectral imagery has been noticed by some researchers and was statistically demonstrated in Chapter 2 of this dissertation. Following on the preceding effort, this chapter is dedicated to investigating the intrinsic dimensionality of hyperspectral data while taking account of the associated nonlinearity. The investigation is conducted using the false near neighbour method, which was originally developed in nonlinear time series analysis [39].

3.0 Motivation

Hyperspectral data are often processed using multivariate analysis methods, which require determining the number of independent variables that are used to describe

the data. This is also called dimensionality estimation. Estimating the dimensionality spanned by hyperspectral data is usually conducted through feature extraction methods, including, among others, Principal Component Analysis (PCA) [65][66], Maximum Noise Fraction (MNF) [38], Projection Pursuit (PP) [67], and Wavelet Transform [68], wherein the dimensionality is determined either as the number of the eigenvectors with the significant eigenvalues derived from the corresponding covariance matrix or the components with significant information content. Based on these methods, the dimensionality of hyperspectral imagery is often estimated between 10 and 25 across diverse scenes. For example, using the Discriminant Analysis Feature Extraction (DAFE) and Decision Boundary Feature Extraction (DBFE) (both of them are variants of PCA), Jimenez and Landgrebe reported in [69] that an AVIRIS image was well represented by 22 extracted features, through which a land-cover classification achieved the highest accuracy. Ifarraguerri and Chang found in [67] that 11 and 12 features, based on the PP and MNF, respectively, were information bearing in a HYDICE image. After examining a suite of linear feature extraction algorithms, Jimenez-Rodriguez et al. (2007) pointed out in [70] that about 15 features were needed to achieve the best classification result given a certain amount of training information.

These approaches, though manifested in different mathematical forms, were all derived by treating hyperspectral images as data sets produced by linear stochastic processes. In other words, hyperspectral data were assumed to be realizations of some linear stochastic processes that are free of nonlinear and dynamical variations. This assumption, however, contradicted the following fact: the formation of hyperspectral imagery involved several physical processes, including solar radiation, atmospheric scattering, interactions between solar radiation and the Earth's surface, and responsiveness of the sensing instruments [71]. These processes could introduce nonlinearity into hyperspectral data. The existence of nonlinearity in hyperspectral imagery has been noticed by some researchers, among others, including Bachmann et al [49] and Mohan et al. [52], who explicitly developed algorithms to deal with the nonlinearity in hyperspectral data for feature extraction and mapping. Further more, it was demonstrated in Chapter 2 that hyperspectral data behave significantly different in

the spectral domain compared to linear stochastic data, which provided a statistical proof showing the existence of nonlinearity in hyperspectral data.

Omission of the nonlinearity in hyperspectral data may hinder the correct estimation of the intrinsic dimensionality. Bachmann et al. [40] found that by taking nonlinearity into consideration with the Isometric Mapping (ISOMAP) embedding, hyperspectral data can be compressed tighter than using the MNF, indicating the intrinsic dimensionality of hyperspectral data may be smaller than those estimated based on linear feature extraction algorithms. They believed that the improved feature extraction was attributable to the embedded coordinate system derived from the ISOMAP, which was more suitable to represent the underlying nonlinear characteristics of the data manifold.

This chapter is prepared based on the preceding effort reported in [72] to investigate the dimensionality of the hyperspectral data represented by the same hyperspectral images described in Chapter 2 using the false near neighbour method, where the distance changes between each data point and its nearest neighbour are recorded when the dimensionality is gradually increased. This method is capable of accommodating nonlinearity and is often employed in nonlinear time series analysis to estimate the dimensionality of phase spaces. It is expected that this method will facilitate identifying the true dimensionality of hyperspectral data. The investigation is conducted using a number of data series generated by “spectrally” concatenating the hyperspectral pixels of different land-cover types, including forest, clear-cut, and water. It is the primary interest of this chapter to look at the dimensionality separately corresponding to individual cover types, as they represent different ways of interaction between solar radiation and ground surface, which may introduce nonlinearity and noise of different intensity into hyperspectral data. After the true dimensions are determined, a corresponding number of independent features are extracted nonlinearly and compared with those extracted using a common linear approach.

3.1 Data sets

The same hyperspectral data sets were chosen for this chapter as shown in Figure 7. They were preprocessed identically as described in Section 2.1 of Chapter 2. For the AVIRIS image, the following three land-cover types were considered: forest, clear-cut, and water. Due to the limitation of available pixels, only two land-cover types: forest and water, were considered for the Hyperion image. For each land-cover type of each image, 100 data series were created in the same way as demonstrated in Section 2.2B of Chapter 2. The length of each series was 8950, consisting of 50 spectrally connected pixels. In total, 15000 AVIRIS and 10000 Hyperion pixels were used. These pixels are indicated in Figure 17.

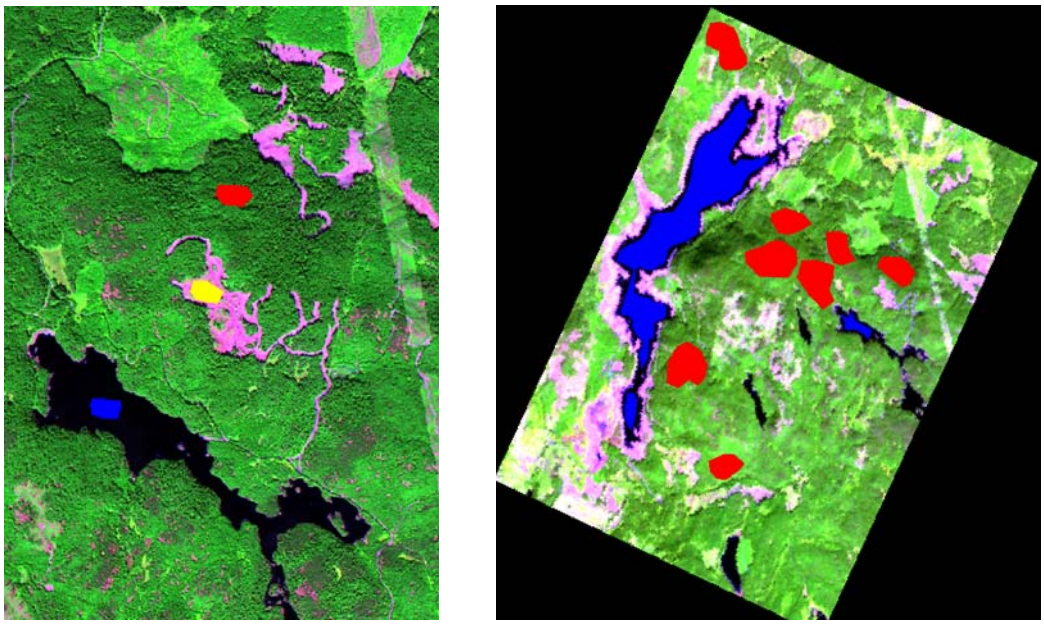


Figure 17. Areas of interest in AVIRIS (left) and Hyperion (right), where forest pixels selected are marked in red, water in blue, and clear-cut in yellow.

3.2 Algorithm description

False near neighbour, the method employed in this study to tackle the dimensionality of hyperspectral data, was originally developed in nonlinear time series analysis to determine the minimum embedding dimension for the phase space reconstructed using time-delay vectors derived from an observed time series [73 – 76]. The minimum embedding dimension is examined by looking at the behaviour of near neighbours when the embedding dimension is expanded from $d \to d+1$, where d

represents the current dimensionality. The idea is intuitive: if the reconstructed phase-space is not large enough; i.e. d is small, the number of near neighbours for each point in the space will change when d is raised to $d+1$. This is because a near neighbour in d-dimensional space may be someone faraway in $d+1$. As the space keeps unfolding, a saturation point will finally arrive, upon which the number of near neighbours becomes constant. This indicates that the space is fully unfolded and this dimension is therefore defined as the minimum embedding dimension, which represents the intrinsic dimensionality of the data under investigation.

To avoid specifying the range or scope required to search for the near neighbours (which is an open parameter and data dependent), the distance between each data point and its nearest neighbour is used instead as the alternative criterion to determine the minimum embedding dimension [74]. Here are the key steps to implement this method, given an observed time series of length l, $\vec{x} = (x_1, x_2, ..., x_l)$:

1. Construct delay vectors in d dimensional space based on the given time series $\vec{x}$ and time-delay τ, i.e., $\vec{p}_i(d) = (x_i, x_{i+\tau}, ..., x_{i+(d-1)\tau})$, where $i = 1,2,..l-(d-1)\tau$.

2. Identify the nearest neighbour of $\vec{p}_i(d)$ and record the corresponding distance, $s_i(d) = \| \vec{p}_i(d) - \vec{p}_{n(i,d)}(d) \|$, where $\vec{p}_{n(i,d)}(d)$ is the nearest neighbour of $\vec{p}_i(d)$.

3. Construct delay vectors in $d+1$ dimensional space based on the given time series $\vec{x}$ and time-delay τ, i.e., $\vec{p}_i(d+1) = (x_i, x_{i+\tau}, ..., x_{i+d\tau})$, where $i = 1,2,...l-d\tau$.

4. Retrieve the distance between $\vec{p}_i(d+1)$ and $\vec{p}_{n(i,d)}(d+1)$, i.e., $s_i(d+1) = \| \vec{p}_i(d+1) - \vec{p}_{n(i,d)}(d+1) \|$, where $\vec{p}_{n(i,d)}(d+1)$ is the nearest neighbour of $\vec{p}_i(d)$ in $d+1$ dimensional space.

5. Calculate the distance ratio of $s_i(d+1)$ to $s_i(d)$ across all delay vectors and compute the average, i.e., $e(d) = (\sum_{i=1}^{l-d\tau} s_i(d+1)/s_i(d))/(l-d\tau)$

6. Increase d by 1 and loop back to step 1 until $e(d)$ converges to 1 or a specified d is reached.

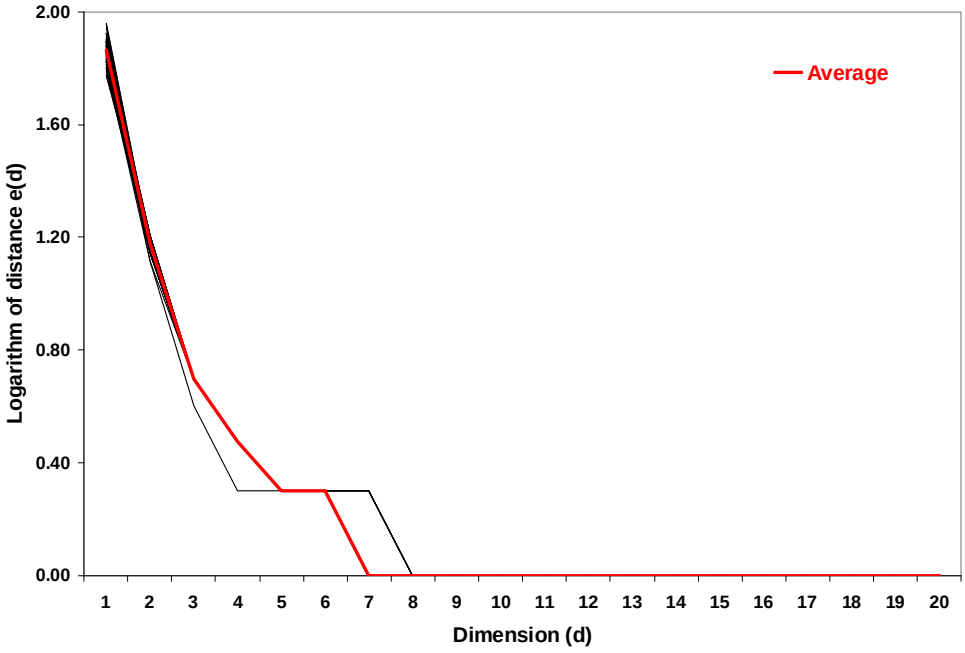
The rule of termination (Step 6 above) is based on the fact that once the embedding dimension is thoroughly expanded, the nearest neighbour of each delay-vector will be stabilized and the distance between them will not change as the dimensionality increases. Therefore the distance ratio, $e(d)$, will eventually reach 1.

3.3 Results and discussion

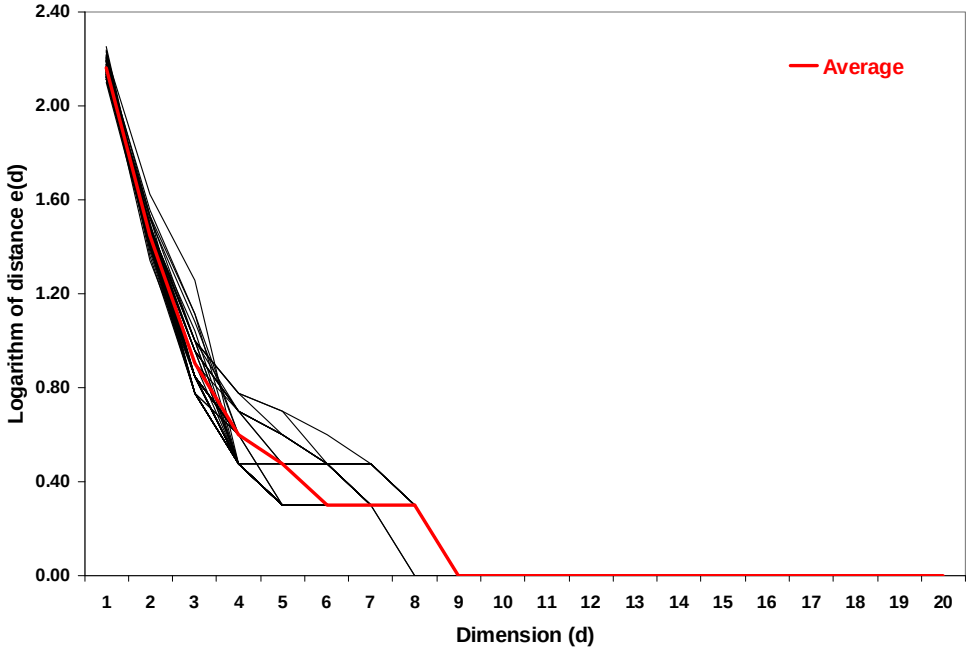
In the initial effort [72] to estimate the dimensionality of hyperspectral data using the false near neighbour method, 50 data series of 3 land-cover types were generated, each of which was made by spectrally connecting 30 AVIRIS pixels of 179 channels. The length of each series was therefore 5170. In this chapter, data series were constructed similarly as before by concatenating pixels extracted separately from areas of forest, clear-cut, and water (Figure. 17). It is intended to look at the impact of nonlinear behaviour of different land-cover types on dimensionality estimation. To further verify the findings reported in [72], both AVIRIS and Hyperion image were included. Meanwhile more pixels were extracted so that data series of increased length were created. As a result of this extension, for each land-cover type, 100 data series were generated. Each of them was at length of 8950 by concatenating 50 pixel spectra.

There are two open parameters in the algorithm described in Section 3.2: time-delay τ and dimensionality d. To maintain the spectral relationship between adjacent channels of the hyperspectral data, τ is always set to 1 so that most of the delay vectors are made using signals from adjacent channels. It is intuitive to start with $d = 1$ and stop with $d = 179$ (number of channels of the hyperspectral image), indicating that the minimum embedding dimension can be 1 and the maximum can go as large as 179. In reality, however, the maximum dimensionality of hyperspectral is usually much less than the maximum possible dimensionality.

Based on the forest data series (100 for AVIRIS and 100 for Hyperion) created using forest pixels, it was found that for the AVIRIS data series when dimensionality was



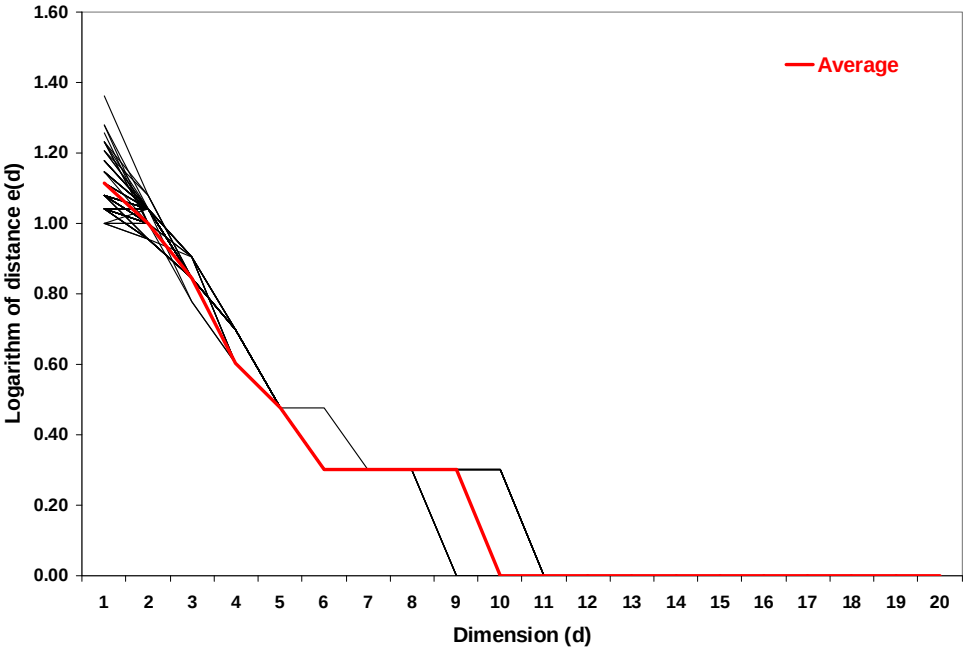
(a)



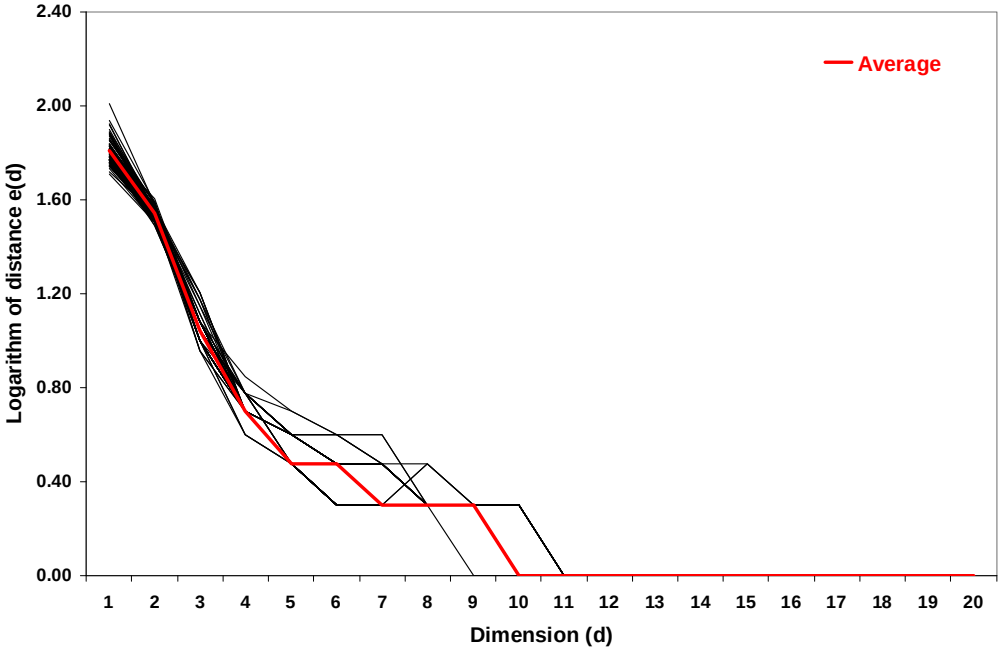
(b)

Figure 18. Estimated dimensionality of forest data series: (a) AVIRIS and (b) Hyperion (X-axis indicates embedding dimensions and Y-axis denotes the logarithm of Euclidean distance between nearest neighbours)

increased to 7 or 8 (average = 7), $e(d)$ converged to 1, or equivalently, $\log_{10}(e(d)) = 0$ (Figure 18a). This meant that the distance between each point and its nearest neighbour



(a)



(b)

Figure 19. Estimated dimensionality of water data series: (a) AVIRIS and (b) Hyperion (X-axis indicates embedding dimensions and Y-axis denotes the logarithm of Euclidean distance between nearest neighbours)

became constant upon these dimensions. Therefore, the dimensionality of these forest series was identified as between 7 and 8 with an average of 7. Similar results were found with the Hyperion forest series (Figure 18b), where the distance saturation dimensionality was between 8 and 9 and the average was 9. The dimensionality was estimated similarly using the water series (both AVIRIS and Hyperion) and the clear-cut series (AVIRIS only). The results are shown in Figure 19 and 20, respectively. The dimensionality of the water was found between 9 and 11 with an average of 10 (both AVIRIS and Hyperion), while it was estimated to be between 7 and 10 with an average of 9 (AVIRIS only) for the clear-cut series.

Four things are found interesting in this study. Firstly, the dimensionality estimated here was markedly smaller than those derived using linear methods, through which the dimensionality was usually between 10 and 25. This finding agreed with what

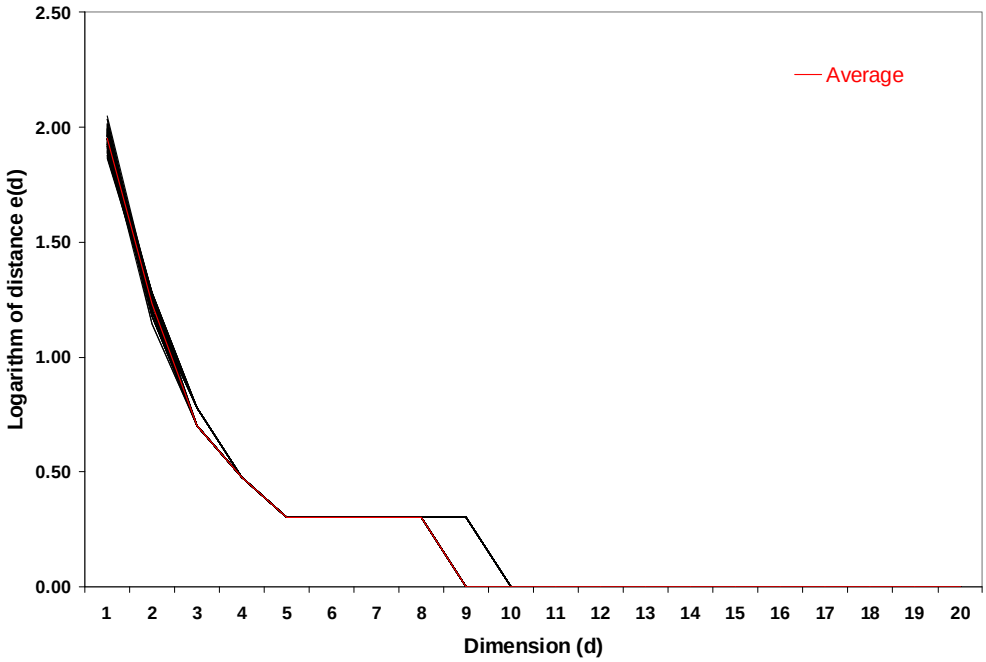


Figure 20. Estimated dimensionality of AVIRIS clear-cut data series (X-axis indicates embedding dimensions and Y-axis denotes the logarithm of Euclidean distance between nearest neighbours)

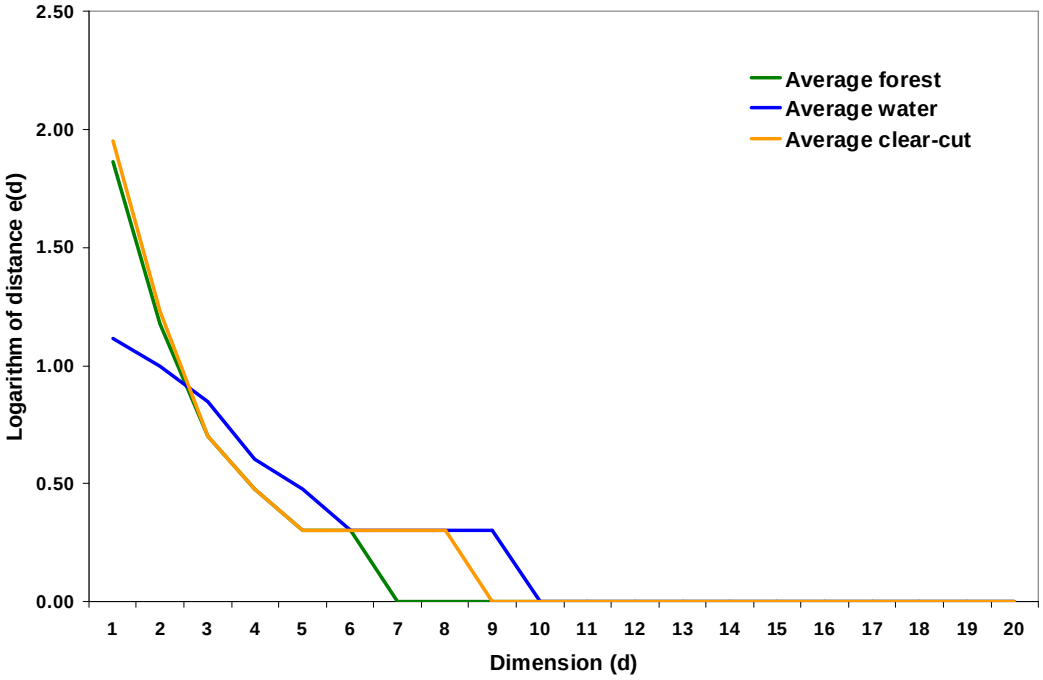
had been found in [40] [41], suggesting that hyperspectral data could be embedded in a tighter space if the nonlinear characteristics were well taken care of. Actually, there is a point of similarity between the method employed here and that in [40]: both of them

assume that hyperspectral data are linear only locally, in contrast to the assumption of global linearity considered by the linear methods. The locally linear assumption is more realistic, which may lead to more accurate dimensionality estimation for hyperspectral data.

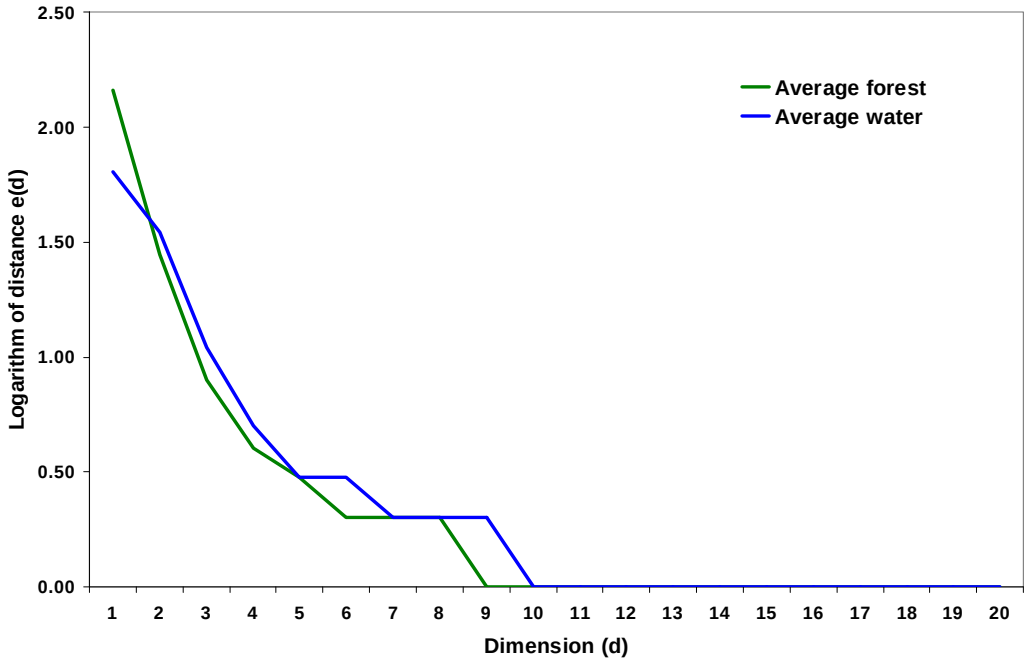
Secondly, the dimensionality of the AVIRIS data series was found to be the same as that identified in [72], though the data series investigated here were much longer. The result indicates that the dimensionality estimation using the false near neighbour method is not sensitive to the length of the data series under investigation. This is particularly pertinent to the remotely sensed data, where the connected spectral data series are used but they are often not as long as the normal time series data.

Thirdly, the average dimensionality of the forest data series was found to be smaller than that of clear-cut and water (AVIRIS) and water (Hyperion) (Figure 21). This is believed to be related to the intensity of nonlinearity and noise in the data stemming from the interaction between solar radiation and ground surface. The selected land-cover types represented 3 different ways of interaction. For forest surfaces, multiple scattering occurred within the forest canopy between solar radiation and leaves, twigs, and branches. The clear-cut areas, often scattered with bushes and grass, were relatively homogeneous and were close to Lambertian surfaces. As a notorious attenuator, water strongly absorbs the incoming radiation across multiple wavelengths, letting very limited radiation (in regions of short wavelength) be reflected back. Because of this, the noise is stronger in water than other land cover types, which caused the dimension of water to be over-estimated as noise is high dimensional in nature. These different ways of interaction gave rise to the data series with some dissimilar characteristics: strong nonlinearity and relatively low levels of noise for the forest series and weakened nonlinearity and significant noise for the water series, leaving the characteristics in between for the clear-cut series. As pointed out in [73-76], data series of strong nonlinearity can be embedded tightly by some lower dimensional space, while the existence of random noise (linear stochastic signals) dilutes the deterministic nonlinear property, which results in a higher estimate in dimensionality estimation. This may explain what has been found in this study.

Lastly it was noticed that the average dimensionality of the AVIRIS forest series was estimated as 7, while it was 9 for the Hyperion forest series (Figure 21). This



(a)



(b)

Figure 21. Comparison of the estimated dimensionality across different land-cover types: (a) AVIRIS and (b) Hyperion (X-axis indicates embedding dimensions and Y-axis denotes the logarithm of Euclidean distance between nearest neighbours)

discrepancy is likely due to the difference between AVIRIS and Hyperion on signal-to-noise ratio (SNR). The SNRs of the former are ~1000 and ~500 corresponding to the visible and near-infrared (VNIR) and shortwave infrared (SWIR) spectral regions [77], respectively. By comparison, for Hyperion, the SNRs are ~150 in the VNIR and ~50 in the SWIR [78]. The existence of the significant noise in the Hyperion data series may blur the nonlinearity, make them less distinguishable from random noise, and result in an increase of the estimated dimensionality (as discussed in the previous paragraph).

Dimensionality estimation and feature extraction are closely related in hyperspectral image analysis, the following questions are often asked once the dimensionality is determined: how can one extract the corresponding number of features to the estimated dimensions from hyperspectral data and what do the extracted features look like? These questions were addressed here using a nonlinear feature extraction algorithm, called Locally Linear Embedding (LLE) [79]. LLE maps high-dimensional data into a low-dimensional Euclidean space while preserving local topological structures. This algorithm assumes that a data point in a high-dimensional input dataset is located in a local patch that can be linearly reconstructed by its neighboring points, one patch for each data point. Eventually the global data topological structures in the high-dimensional space can be preserved by stitching up all the patches (linear hyperplanes). The details regarding LLE's derivation and implementation in the context of hyperspectral data were described in [51].

Considering that the maximum average dimensions were identified as 10 for both AVIRIS and Hyperion images across all the three land-cover types aforementioned, 10 features were extracted using LLE from each of the images. The false color composite images, which were generated using LLE-feature 1, 2, and 3, are shown in Figure 22. To further investigate the quality of the extracted features, an entropy-based comparison was arranged between features extracted by LLE and those by the Maximum Noise Fraction (MNF) [38]. MNF is a well-known linear algorithm that is often used to process hyperspectral data for dimensionality estimation, feature extraction, and noise reduction (the details about MNF are discussed in the next section). The entropy (E) of each extracted feature image was evaluated using the following formula: $E = -\sum_{i=1}^{g} d_i \log_2 d_i$,

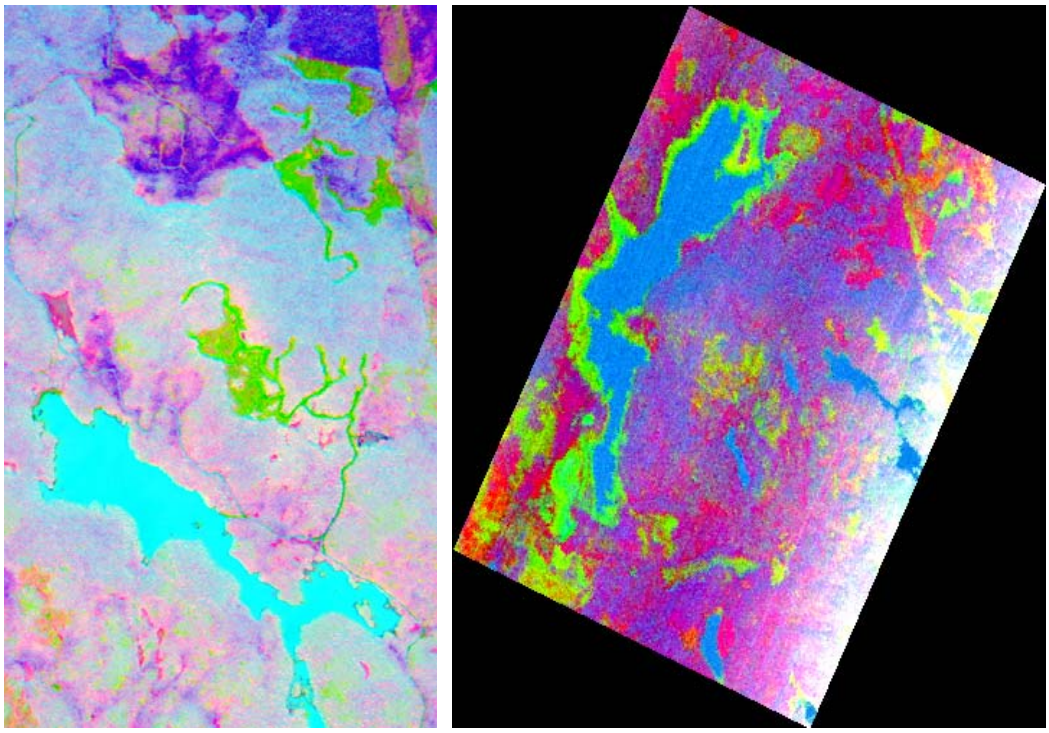


Figure 22. False color images generated by the first three LLE features: AVIRIS (left) and Hyperion (right), where LLE features 1, 2, and 3 were mapped to red, green, and blue, respectively.

where g is the number of grey levels assumed in the image; d_i is the probability that the grey level i occurred. The results of the comparison are summarized in Table 2, where the larger values of E indicate more information content. This comparison revealed that, in terms of image entropy, the 10 LLE-feature images contained more spatial information than the MNF-feature images ($p < 0.05$). This conclusion was supported by both the AVIRIS and Hyperion images. It was also in agreement with what was found in [51].

The following two phenomena were also noticed: 1) the entropy values descended against the feature bands extracted by both LLE and MNF; 2) the entropy values of the feature bands extracted from the AVIRIS data were greater than the corresponding values of the feature bands extracted from the Hyperion data. The first phenomenon indicated that the information content by entropy as calculated using the above equation was in full agreement with the variance, as the extracted feature bands by both LLE and MNF were arranged in descending order by variance. The second phenomenon is well understood as the 4m-AVIRIS image contains more information than the 30m-Hyperion image does. It

Table 2 Estimated dimensionalities of different land-cover types

Feature band	1	2	3	4	5	6	7	8	9	10
AVIRIS-LLE	7.28	6.97	6.94	6.80	6.74	6.43	6.35	6.14	6.13	5.49
AVIRIS-MNF	7.26	6.92	6.92	6.80	6.67	6.35	6.28	6.11	6.11	5.45
Hyperion-LLE	6.83	6.43	5.55	5.55	5.37	5.20	5.18	5.17	5.15	4.61
Hyperion-MNF	6.81	6.35	5.48	5.45	5.33	5.16	5.14	5.14	5.09	4.54

also showed that the above conclusion also held when the images were projected into LLE and MNF spaces.

3.4 Computational complexity

This section is prepared to analyze and compare the computational complexity in computing time between the Maximum Noise Fraction (MNF) and the false near neighbour method discussed previously in this chapter on estimating dimensionality of hyperspectral data. As the MNF originated from and is closely related to the Principal Component Analysis (PCA), the description and mathematical derivation of PCA are also included in this section. The analysis and comparison are carried out in Big O notation using the original version of these algorithms. Some improved versions of these algorithms may exist, but they are out of the scope of this dissertation.

A. Principal Component Analysis (PCA)

PCA has been around for a long time in Statistics. The general idea of PCA is to determine a set of new basis vectors, called principal components, and project the original data onto the space spanned by these principal components. In the new space, features are arranged in the order of their variance [80], represented by the corresponding eigenvalues of the covariance matrix. The number of significant eigenvalues corresponding to these principal components is a good hint of the data dimensionality. Normally, eigenvectors whose eigenvalues are greater than 1 are considered as information-bearing. Mathematically PCA is about the factorization of a covariance matrix. Here is the derivation of the PCA in the context of hyperspectral data.

Let n be the number of spectral bands and $\vec{s}_i$ be a pixel (column vector), for $i = 1,2,...p$, where p is the number of pixels in the image. Let $V \in R^{n \times n}$ be the covariance matrix of the image, which is constructed as

$$V = \frac{1}{p-1}\sum_{i=1}^{p}(\vec{s}_i - \vec{m})(\vec{s}_i - \vec{m})^T \quad (3.1)$$

where $\vec{m} = \frac{1}{p}\sum_{i=1}^{p}\vec{s}_i$, the mean spectral vector of the image. Because V is positive semi-definite for the hyperspectral data, the factorization of V can be achieved, among others, by using Singular Value Decomposition (SVD) as follows,

$$V = U\Sigma U^T \quad (3.2)$$

where $\Sigma \in R^{n\times n}$ is a diagonal matrix, composed of the eigenvalues of V as its diagonal entries, while $U \in R^{n\times n}$ is an orthogonal matrix whose columns are the eigenvectors of V. The eigenvalues represent how much global variance of the hyperspectral image is explained by the corresponding eigenvectors [80]. Considering that the first few eigenvalues, say d, account for a major part of the total eigenvalue (trace), their corresponding eigenvectors are selected as the principal components for data representation. Hence the number of these principal components, which is also d, is considered as the dimensionality of the hyperspectral data. Other eigenvectors, related to the trivial eigenvalues, are treated as redundant or noise and discarded. For example, the dimensionality of the Hyperion data in [15] was estimated as $d = 12$, which represent 95% of the total variance. The final step of the PCA is to project all the pixel vectors onto the space spanned by the identified principal components using the following linear transformation,

$$D_{PCA} = U_s D \quad (3.3)$$

where $U_s \in R^{d\times n}$ is a sub-matrix of U, which only contains d significant eigenvectors of U; $D \in R^{n\times p}$ is filled with column pixel vectors of the original image being considered; $D_{PCA} \in R^{d\times p}$ is the result of the PCA.

For the purpose of dimensionality estimation, the last step above (Equation 3.3) is not needed. What is required includes: calculation of the mean vector $\vec{m}$; construction of the covariance matrix V (Equation 3.1); and the factorization of V (Equation 3.2). The computational complexities of these steps are $O(pn)$, $O(pn^2)$, and $O(n^3)$ (based on QR method [81]), respectively, where p is the number of pixels in the image and n is the

number of spectral bands. Among the above three steps, the construction of the covariance matrix is the most expensive computationally, because $p >> n$ for a normal-sized hyperspectral imagery.

B. Maximum Noise Fraction (MNF)

There is an implicit assumption in the PCA when it is applied to optical remotely sensed data (including hyperspectral): the noise in the data is a standard Gaussian, i.e. zero mean and unit variance, and can be fully characterized by the covariance matrix. This assumption, however, is often violated in reality. Green et al. pointed out in [38] that ranking new features in an eigen space in terms of variance as done in PCA is not appropriate, because the noise in the data often contributes greatly to the total variance. They concluded that the total variance of an image is not a legitimate criterion for dimensionality estimation and feature extraction for the optical remotely sensed imagery. Instead they proposed to use signal-to-noise ratio (SNR) as the criterion. Lee et al. [82] further formulated the MNF into two PCAs and renamed it as the noise-adjusted PCA, where the first PCA converts the noise into a standard Gaussian (also called noise whitening) and the second one is applied to the noise whitened data (result of the first PCA). It has been shown in many publications, including [66][82][83], that MNF is more effective than PCA, especially when hyperspectral data are noisy (low SNR).

As the MNF is equivalent to two PCAs, its computational complexity is approximately doubled as well. The first PCA, designed to deal with the noise, includes the following four steps: calculate mean noise vector, construct noise covariance matrix, standardize it, and whiten the noise in the data. The first two steps are computationally identical to those in a normal PCA, which are in the magnitudes of $O(pn)$ and $O(pn^2)$, respectively. The third step is to determine the noise whitening matrix that converts the noise covariance matrix into an identity, i.e.,

$$NW = I \text{ or } W = N^{-1}I \tag{3.4}$$

where N is the noise covariance matrix, I the identity matrix, W the noise whitening matrix. This involves a matrix inversion and multiplication. Both of them are $O(n^3)$,

though they can be optimized to $O(n^{2.376})$ [84]. The last step is to apply W to D for the noise whitening,

$$D_w = WD \tag{3.5}$$

where $D_w \in R^{n \times p}$ is the noise whitened data. The cost of this step is $O(pn^2)$. The second PCA included in MNF is conducted using D_w as the input, which is identical to the process described in 3.4A. In summary, the most expensive computing involved in the MNF is $O(pn^2)$.

C. False near neighbour method

As discussed in Section 3.2, there are six steps in the false near neighbour method, including: 1) construction of delay vectors in d dimensional space; 2) searching for the nearest neighbour and recording the corresponding distance; 3) construction of delay vectors in $d+1$ dimensional space; 4) calculation of the distance between each delay vector in $d+1$ dimensional space and the nearest neighbour identified previously in d dimensional space; 5) ratioing the distances obtained in 2) and 4); and 6) increasing dimensionality by 1 until the distance ratio calculated in 5) converges to 1. Generally speaking, this is a problem of pair-wise nearest neighbour search. Contrary to the conventional nearest neighbour problems, where the dimensionality stays the same during the search process, this search needs to be conducted with increasing dimensionality.

The general nearest neighbour search requires $O(p^3)$, though it may be optimized to $O(p^2 \log(p))$ [84]. If the increment of dimensionality is considered, the requirement becomes $O(dp^3)$, where d and p denote the dimensionality and the number of pixels, respectively. Considering that a normal-sized hyperspectral image includes a few million pixels, the pair-wise nearest neighbour search in an incremental high dimensional space poses a real computing challenge, even for the most update-to-date computers. Compared to $O(pn^2)$ that is the upper bound for the MNF, $O(dp^3)$ required by the false neighbour method is a dramatic increase in computational complexity.

3.5 Chapter summary

The main focus of this chapter is about estimating the intrinsic dimensionality of hyperspectral data and the relationship between the dimensionality and certain land-cover types that interact differently with the incoming solar radiation. Using the false near neighbour method that is capable of accommodating the nonlinearity in hyperspectral data, the dimensionalities of hyperspectral data were estimated between 7 and 10 as listed in Table 3. With the linear transformations, those dimensionalities, however, were found to be between 10 and 25. This indicated that hyperspectral data could be embedded more tightly in a lower dimensional space if nonlinearity was addressed. Accurate estimation of dimensionality is important for hyperspectral data analysis based on multivariate analysis. This may lead to using a fewer number of independent variables to describe hyperspectral data and reducing the subsequent computational complexity.

Table 3 Average dimensionalities of different land-cover types

	Forest	**Water**	**Clear-cut**
AVIRIS	7	10	9
Hyperion	9	10	N/A

It was also found that the dimensionality changes among different land-cover types**.** This is believed to be related to the interaction between solar radiation and ground targets, which may introduce nonlinearity and noise of different intensities into hyperspectral data. The stronger the nonlinearity and the weaker the noise, the fewer the dimensions are required to describe the data. The differences in signal-to-noise ratio between AVIRIS and Hyperion were believed to be the cause of dimensionality discrepancy on forest data series, where the low SNR Hyperion series needed more dimensions than the high SNR AVIRIS series. These findings were in agreement with those reported for the nonlinear time series analysis. Based on the determined dimensions, 10 features were extracted using the LLE algorithm from each of the hyperspectral data sets. The extracted features were compared favorably in terms of information preservation to those derived using the MNF-based method.

Though the false near neighbour method can estimate the dimensionality of

hyperspectral data more accurately than linear transformations, it is very expensive in computational complexity. For example, the MNF based methods require $O(pn^2)$ in computing time, but the false near neighbour method requires $O(dp^3)$. To make it applicable to process normal-sized hyperspectral data within a reasonable time frame, this method needs to be optimized either using some alternative algorithms for the nearest neighbour search or by implementing it in a parallel computing environment (cluster computing for example). Considering the nearest neighbour search has to be conducted by incrementing the dimensionality, some of the alternative search algorithms may not be feasible. Implementing this method in parallel seems to be more appropriate.

Chapter 4

Nonlinearity-counted noise reduction

Like other types of digital signals, hyperspectral data are subject to a variety of effects associated with the physical processes involved during data acquisition, including interaction between solar radiation and atmosphere, interaction between solar radiation and targets on the Earth surface, and sensor responsiveness to the reflected radiation from the Earth surface. The noise from these efforts distorts signal statistical properties and limits applications of hyperspectral data for information extraction. Noise reduction or removal is, therefore, a prerequisite for many hyperspectral data applications based on classification, target identification, and spectral unmixing.

Principal Component Analysis (PCA) and Wavelet Transform (WT), among others, were two commonly adopted approaches to deal with the noise in hyperspectral data [85-88]. They both worked well under the assumption that hyperspectral data were composed of the following two separable and additive components: noise in white Gaussian and signal in linear stochastic distributions. However, studies [88] had found that hyperspectral data were far more complicated than indicated by the above assumptions. For example, the noise might be non-Gaussian and signal dependent. Moreover, as demonstrated in Chapter 2, hyperspectral data exhibited nonlinear characteristics, which suggests that the noise might exist in a broad band in frequency domain, which is often the case for nonlinear data sets. This indicates that the noise might not be separated from the signal by frequency. To reflect the real situation of hyperspectral data, an algorithm is introduced in this chapter with the intention to improve the noise reduction for hyperspectral data. The effectiveness of the introduced algorithm is evaluated using multiple metrics focusing on both noise reduction and spectral shape preservation. Improved results of noise reduction are expected with the introduced algorithm, as it takes into consideration the real statistical characteristics of hyperspectral data.

4.0 Motivation

Noise undermines every effort of information extraction from hyperspectral data. For the classification-based applications, noise blurs the statistical boundaries between classes upon which classifiers rely for class separation, and it ultimately results in less accurate classification results. For applications of target identification based on spectral analysis using a spectral library, excessive noise displaces the location or decreases the depth of certain absorption features. This could make targets of interest unrecognizable when comparing this spectral signature with those in a spectral library. Spectral unmixing using hyperspectral data also suffers from the impact of noise, wherein noise hinders determining the endmembers if they only exist with small spatial abundances.

To mitigate or reduce noise in hyperspectral data, a number of algorithms have been conceived in the past thirty years. Though they varied in mathematical details and were derived for different applications, most of them fell into the following two mainstreams: Principal Component Analysis (PCA) and Wavelet Transform (WT) based noise reduction. As suggested by their names, the former applied PCA, directly or indirectly, to the covariance matrix of the hyperspectral data of interest. Noise reduction was achieved by reconstructing the hyperspectral data with those expressed by a few high ranking eigenvectors of the covariance matrix. The low ranking eigenvectors were considered as noise and discarded. The second method employed WT to decompose hyperspectral data into fine- and large-scale components, where the fine-scale ones bore the high frequency information (such as subtle absorption features and noise) and the large-scale ones bore low frequency information (such as trending and continuum). Discarding the noise bearing fine scale components therefore reduced the noise in the hyperspectral data.

A popularly-cited PCA-based noise reduction algorithm, called Maximum Noise Fraction (MNF), was proposed in [38] for multi-spectral and hyperspectral remotely sensed data. MNF projected data into a space where the data were rearranged based on signal-to-noise ratio (SNR). Noise could be simply reduced by getting rid of the components with low SNRs. This algorithm was later rephrased in [82] as the noise-adjusted principal components transform. It was found that MNF was mathematically

equivalent to two concatenated PCAs, where the first one was about noise whitening and the second one was a variance-based projection for the noised-whitened data. Using HYDICE data, Staab reported in [85] the impact of noise reduction on target detection. What he did included: 1) projecting the HYDICE image and some target spectra in reflectance space into PCA and MNF spaces separately; 2) discarding the low ranking PCA and MNF components; 3) reducing noise by iteratively projecting the high ranking PCA and MNF components back into the reflectance space; 4) evaluating the impact of the noise removal using Spectral Angle Mapper (SAM) and Orthogonal Subspace Projection (OSP). It was found that both PCA and MNF based noise reductions had minimal effect on target detection as long as 20+ high ranking PCA or MNF components were used for data reconstruction.

Wavelet Transform (WT) based-algorithms for noise reduction became increasingly popular recently. These algorithms took advantage of the following two unique features of the WT: multi-resolution and local adaptability, which helped to achieve an optimal trade-off between noise reduction and the preservation of spectral absorption features. Among many examples of this type, Schmidt and Skidmore in [86] de-noised some field-measured spectra of salt-marsh vegetation using wavelet shrinkage, where noise reduction was achieved by putting to zero the coefficients representing the finest detail levels in the transform. They reported that the proposed WT-based approach outperformed other spectral smoothing algorithms (including Fourier transform-based smoothing, spline curve fitting, and least squares curve fitting) in maintaining wavelength positions of the local minima or maxima as well as the inflection, which were usually the locations of some absorption features. Instead of directly thresholding wavelet coefficients for noise reduction, Scheunders proposed in [87] to pre-process them by squaring and multiplying to improve the separation of probability density functions between signal and noise. It was demonstrated based on a Landsat image that the proposed denoising technique achieved a factor of 2 on SNR improvement. To address the signal-dependent noise (such as photon noise) in hyperspectral data, Othman and Qian developed a wavelet-based algorithm in [88] by marrying up spatial and spectral wavelet shrinkage which took advantage of the dissimilarity of the signal nature in the spatial and spectral dimensions. They reported that the proposed algorithm provided

significant and consistent improvement for reducing noise in both real and simulated hyperspectral data sets (AVIRIS). They believed that the proposed algorithm was superior in performance to the published algorithms assuming that noise existed in a fixed level.

Though hyperspectral denoising algorithms, including those discussed above, might differ dramatically in mathematical details from one to another, they often had three things in common, explicitly or implicitly: 1) assumption about statistical properties of hyperspectral data; 2) assumption about the nature of noise; 3) linear transformation based operations. The first thing in common was that hyperspectral data were assumed to be realizations of some linear stochastic processes. This has been proved wrong in Chapter 2 of this dissertation. This assumption omitted or ignored the nonlinearity in hyperspectral data, which might have negative impact on information extraction from hyperspectral data. For example, as demonstrated in Chapter 3 of this dissertation, omission of nonlinearity in hyperspectral data caused overestimation of data dimensionality. The second thing in common was the nature of noise, which assumed that the noise manifested itself in a narrow spatial frequency band and could be well modelled by some additive Gaussian processes. The existence of nonlinearity identified in Chapter 2, however, indicated that noise in hyperspectral data was likely non-Gaussian and existed in broad-band as shown in many nonlinear data sets [89 – 93]. Noise might spread out across the entire frequency domain, which made it impossible be fully separated from the true signals. The nature of the noise in hyperspectral data will be discussed in detail in Section 4.2. The third thing in common was that most of the proposed algorithms, including PCA, MNF, and WT, linearly transformed hyperspectral data from one space to another for noise and signal separation. These linear operations lacked the capability of dealing with nonlinearity, and either ignored it or mistook it for noise.

In this chapter, a denoising algorithm from nonlinear time series analysis is introduced and adapted to hyperspectral applications. The algorithm was developed using less stringent assumptions about data and noise distributions than those assumed by many denoising algorithms as discussed above. It takes explicit measures to deal with the nonlinearity in hyperspectral data. The effectiveness of the algorithm will be evaluated in

the following sections of this chapter using both AVIRIS and Hyperion images. The evaluation is based on three metrics representing: noise intensity, signal-to-noise ratio, and spectral shape preservation. It is the ultimate interest of this chapter to report a noise reduction algorithm that is capable of suppressing noise and boosting signal-to-noise ratio under the constraint of spectral shape preservation (spectral shape is the overall representation of all absorption features contained in a spectral signature).

4.1 Data sets

The same hyperspectral images were chosen for this chapter as shown in Figure 9. They were preprocessed identically as described in Section 2.1 of Chapter 2. Each hyperspectral image, as usual, was reformatted into multiple spectral data series prior to noise reduction. For convenience, each data series was generated by spectrally connecting an entire row of pixels. As the AVIRIS image had 1000 rows, 550 columns, and 179 bands, 1000 AVIRIS data series were generated, each of which was 98450 in length. For the Hyperion image of consideration, 387 data series were generated, each of which was 48867 in length since the image was comprised of 387 rows, 273 columns, and 179 bands. Both AVIRIS and Hyperion images will be denoised in the following sections of this chapter. In addition, some forest and inland water pixels, as indicated in red and blue respectively in Figure 9, will be selected from the denoised images. The reason of this selection is that the forest and inland water are the most prominent land cover types within the image scene. They are used to evaluate the impact of noise reduction on spectral shape preservation.

4.2 Algorithm description

The development of any denoising algorithms should be based on the nature of the data to be denoised. One has to consider the following realities regarding the nature of hyperspectral data. Firstly the noise in hyperspectral data manifested itself not only as signal-independent (e. g. dark noise) but also signal-dependent (e. g. photon noise) [88]. This suggested that the noise might not be fully represented by a white Gaussian process and the noise might not be simply additive to the signal. Chang et al. reported in [94] [95] that in addition to the additive noise, there existed another type of noise: interference in hyperspectral data, which gave rise to more serious signal deterioration than the former.

Secondly due to the narrow and contiguous (sometimes even slightly overlapping) band configuration of hyperspectral sensors, the between-band noise was unlikely independent with each other but interrelated, similar to the between-band data dependency as depicted by the correlation image (Figure 6). Lastly, as demonstrated in Chapter 2, the nonlinear processes associated with hyperspectral data generation caused irregular behaviours (nonlinearity-related) that could not be fully described by linear stochastic processes. Previous studies had found that the noise in nonlinear data series exhibited genuine broad band spectra in frequency domain [89–93], indicating that the noise did not exist isolated but spread out across all frequencies.

The above reality of hyperspectral data challenges the appropriateness of assuming that hyperspectral data are realizations of some linear stochastic processes, where the noise is between-band independent and additive to the signal. This is the assumption that many legacy denoising algorithms, including PCA, WT, and their variants, were based upon. The less realistic assumption about hyperspectral data might limit the effectiveness of noise reduction. Denoising algorithms that directly face the reality of hyperspectral data are therefore needed. In nonlinear time series analysis, a suite of denoising algorithms were developed for this purpose [89-93]. These algorithms took realistic or less stringent assumptions about the data to be denoised. Under these assumptions, the noise was classified into the following categories: measurement and dynamic. The former referred to the corruption of observations by errors that were independent of the phenomenon being measured and assumed to be white Gaussian. The latter, in contrast, referred to the distortions caused by the perturbed dynamic systems underlying the phenomena being measured in a feedback fashion. In the context of hyperspectral data, the measurement noise was the equivalent of the random noise as white Gaussian, which had been well taken care of by the legacy denoising algorithms. The dynamic noise, on the other hand, corresponded to other types of noise in hyperspectral data that had not been satisfactorily dealt with, which was signal dependent, between-band correlated, and broad band spreading. Based on these assumptions, the signal and the noise of both categories were separable only by the difference of dimensionality, wherein the noise manifested itself in a higher dimensional space than the signal.

Among the suite of noise reduction algorithms of interest, the Local Geometric Projection (LGP) method [96] was adapted in this chapter to denoise hyperspectral data. This decision was made due primarily to LGP's geometric feasibility, simplicity, robustness, and effectiveness. The LGP was proposed based on the observation that nonlinear signals were intrinsically low dimensional and their superficial high dimensionality was often due to the existence of noise that was high dimensional in nature [89–93]. Noise smeared out the low dimensional manifold where the nonlinear signals resided. This observation was consistent with what was believed about hyperspectral data, whose intrinsic dimensions were much smaller than they appeared to be. Denoising could be achieved by adjusting the trajectories of the noisy signals towards their true locations in the underlying manifold. As it was impossible to reproduce the physical processes through which the data series under consideration were generated, the underlying manifold could only be approximated using the data series themselves. The embedding theorems proposed in [97-98] provided the scheme for approximating the underlying manifold in terms of phase space reconstruction using time-delay embedding. LGP is similar to LLE, as both of them try to separate a data set into small patches in topology and treat the data points linearly within each patch.

The LGP was derived with the following geometric picture in mind: a nonlinear data series expressed in a smooth manifold (constructed by using time-delay embedding) could be linearly approximated in the vicinity of each point on it. The manifold was tessellated into small neighbourhoods, each of which was approximated by a locally linear representation provided that it was small enough. The linear representation of each neighbourhood was best represented, in terms of variance, by the eigenvectors derived from the covariance matrix of the data points within each neighbourhood. Noise reduction was achieved by projecting the data points using high ranking eigenvectors because the low ranking eigenvectors were contributions of noise. Specifically the LGP worked in four steps as follows: A) construction of state vectors in phase space; B) establishment of neighbourhood; C) determination of projection directions; D) noise reduction by projection. These steps are detailed below in the hyperspectral data setting using the following notations. Let $\vec{v} = (v_1, v_2, ..., v_l)$ be a data series with a length of l

elements, which is made by concatenating p hyperspectral pixels in the spectral direction as shown in Chapters 2 and 3. Therefore, $l = p \times n$, where n is the number of bands.

A. Construction of state vectors in phase space

The idea of phase space was briefly mentioned in Chapter 3. For any observed data, they could be expressed either as a series (the format in which the data were collected) or some multi-dimensional vectors in a Euclidean space, which was called phase space. The phase space consisted of all possible values of position, momentum, or other prosperities of interest that were governed by the underlying dynamic system. These representations were different in format, but equivalent to each other in data presentation, just like the different ways to represent hyperspectral data as discussed in Chapter 1. One representation is preferred to the other depending on applications. For noise reduction purposes, representing data in phase space was preferred. According to Takens [97] and Sauer et al. [98], given a high enough embedding dimension, a phase space could be reconstructed using a single data series by time delay embedding, wherein the geometric properties of the manifold and the dynamics borne by the data series were preserved. Specifically, given a data series $\vec{v} = (v_1, v_2, ..., v_l)$ and the embedding dimension of d, the data series could be expressed as the following d dimensional vectors, called state vectors, in the reconstructed phase space:

$$\vec{p}_i(d) = (v_i, v_{i+\tau}, ..., v_{i+(d-1)\tau})$$

where $\vec{p}_i(d)$ is the *ith* state vector for $i = 1,2,...l-(d-1)\tau$, and τ is the time-delay coefficient. The number of state vectors to be created depended on l (the length of $\vec{v}$), d, and τ.

B. Establishment of neighbourhood

The neighbourhood of each state vector in a phase space was established by nearest neighbour search in Euclidean distance. Because linear geometric structures derived from the identified nearest neighbours were used to approximate the nonlinear structures represented by the data series within the corresponding neighbourhood, the size of the neighbourhood could not be too small, otherwise unacceptable geometric distortions would occur. The neighbourhood could not be over specified either, as an oversized neighbourhood was ineffective for noise reduction. The size of a

neighbourhood can be defined either by the radius of a bounding hyper-ball or the required number of points (state vectors) that had to be included in a neighbourhood. The latter method was used here so that only enough points were maintained to approximate the local geometric structures in each neighbourhood. The optimal selection of the size of a neighbourhood depended on the density of the data cloud in the reconstructed phase space and properties of the data series under consideration. It was often data dependent and determined by experiments.

C. Determination of orthogonal vectors

The orthogonal vectors, onto which state vectors were projected for locally linear approximation, were derived based on maximizing the variance of state vectors within each neighbourhood. This was an eigen problem that could be solved by decomposing the covariance matrix of state vectors. Prior to the decomposition, a diagonal weight matrix was applied to the covariance matrix to penalize the outer coordinates of the state vectors to avoid over-correction due to a possible unstable manifold [90]. The required steps included: 1) computing the mean state vector within a neighbourhood $\vec{m}(d) = \frac{1}{k}\sum_{i=1}^{k} \vec{p}_i(d)$ where k is the number of state vectors; 2) constructing the covariance matrix of the state vectors $V = \frac{1}{k-1}\sum_{i=1}^{k}[\vec{p}_i(d) - \vec{m}][\vec{p}_i(d) - \vec{m}]^T$; 3) penalizing on outer coordinates of state vectors to avoid over-correction $Z = WVW^T$, where W is a diagonal weight matrix whose elements on the main diagonal are all 1s except for the first and last ones that are set larger than 1; 4) factorizing the penalized covariance matrix Z using Singular Value Decomposition (SVD) $Z = U\Sigma U^T$, where the columns of U are the eigenvectors of Z. These eigenvectors were arranged in a descending order in U in terms of variance representation, where the first one (first column of U) indicated the direction of the maximum variance, while the last one (last column of U) indicated the direction of the least variance, which was believed to be noise related.

D. Noise reduction by projection

Noise reduction was achieved by projecting state vectors onto the directions of the eigenvectors representing the majority of data variance. These eigenvectors were

considered as high ranking ones. The exact number of the high ranking eigenvectors used for the projection depended on the intrinsic dimension of the data series under consideration. The projection was done by $D_{nr} = U_s D$, where U_s is the simplified version of U, only consisting of partial columns of U (high ranking eigenvectors); D is the matrix consisting of state vectors as its columns; and D_{nr} contains the adjusted state vectors, which constitute the denoised data series. It should be noted that each element in the data series appeared in several state vectors and consequently was corrected several times. The final results of noise reduction would be the average of these corrections.

To denoise a given data series, the preceding steps, except for Step A, might need to be executed iteratively until no further denoising was obtainable. The number of iterations required was related to the size of neighbourhood and the data series under consideration, which was often determined by checking the convergence of the denoised results. Because the final denoised data were in the form of state vectors, they would be firstly transformed back to the form of data series, secondly reformatted into hyperspectral pixel vectors, and finally saved as a hyperspectral image.

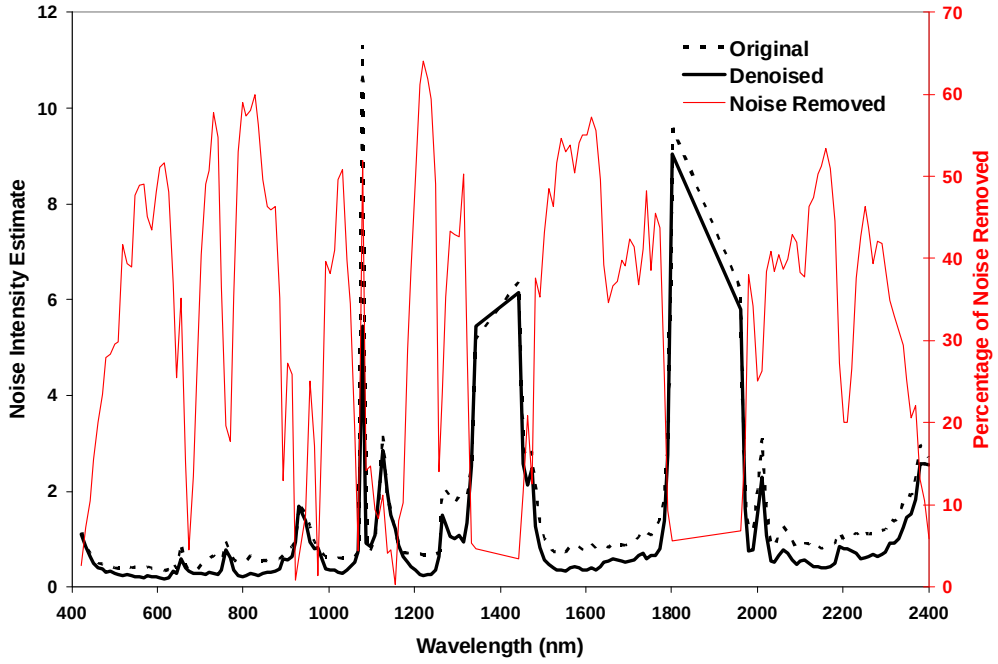
4.3 Results and discussion

To adapt the Local Geometric Projection (LGP) algorithm to hyperspectral applications, one had to determine the following input parameters in the above steps, including the embedding dimension (d) and delay coefficient (τ) required in Step A, size of the neighbourhood (k) required in Step B, number of eigenvectors to be maintained (s) required in Step D, and number of iterations that the above steps need to be executed. According to Takens' theory [97], to reveal the underlying nonlinear dynamics of a data series, the ideal embedding dimension should be at least twice the intrinsic dimension of the data series. As identified in Chapter 3, the average intrinsic dimensions of forest, clear-cut, and water data series were 7, 9, and 10 by AVIRIS and 9, 10, and N/A by Hyperion. Without losing generality, the number 10 was taken as the intrinsic dimension for both AVIRIS and Hyperion data. This gave an embedding dimension $d = 21$. As to the delay coefficient τ, it was set to unity, i.e. $\tau = 1$. This was consistent with what was used in Chapters 2 and 3. Using unit delay coefficient depicts the relationship between

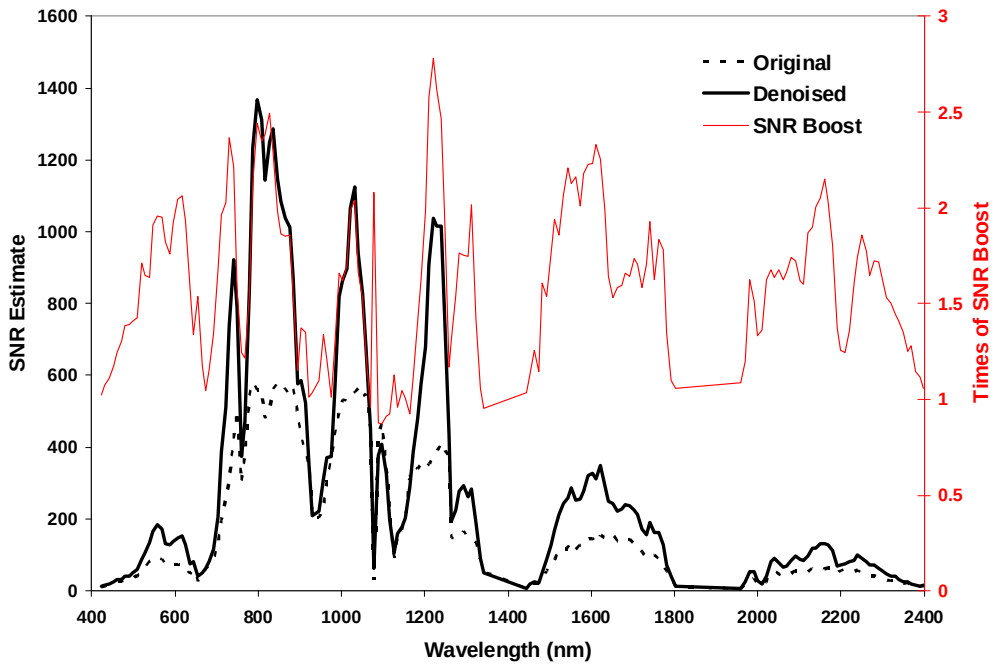
adjacent bands. The optimal size of a neighbourhood k was determined heuristically by testing multiple neighbourhoods of different sizes, starting with a few points up to a thousand points. It was straightforward to decide the number of eigenvectors s for the local projection, because it was equal to the intrinsic dimension of data series, i.e. $s = 10$. The last parameter to be determined was the number of iterations for the LGP steps to be executed. Previous studies in nonlinear time series analysis [89-93] had found that in most cases convergence could be achieved between one and three iterations.

In the following, the effectiveness of noise reduction based on the LGP was evaluated by comparing the noise intensity and signal-to-noise ratio (SNR) between the original and the denoised images. However, it was not an easy task to get an accurate noise estimate for a hyperspectral image, because it required additional information. The additional information included the sensor dark current image and large homogeneous areas within the image scene, which were very difficult to acquire for average hyperspectral data users. To alleviate the above limitations, a legacy method proposed by Roger and Arnold [99] [100] was employed here for noise estimation. Without relying on any additional information, this method was capable of producing reasonable noise estimates based on the covariance matrix of a hyperspectral image alone. In this method, the noise was estimated as the residual standard deviations, which were derived as follows: $V^{-1} = U_d C U_d$, where V is the covariance of the hyperspectral image under consideration; U_d is a diagonal matrix; C is a correlation-like matrix, i.e. 1s down its principal diagonal and all other elements having magnitudes less than 1. The reciprocals of the elements on the principal diagonal of U_d were interpreted as the noise estimate of the hyperspectral image. They represented the unexplained or residual standard deviation of the covariance matrix.

To demonstrate the effectiveness of LGP for noise reduction, pre- and post-denoised images were compared in terms of noise intensity and signal-to-noise ratio. Noise intensity, as discussed above, was estimated as the reciprocals of the elements on the principal diagonal of U_d, while signal intensity was estimated by taking the square root of the elements on the principal diagonal of V (covariance matrix). The signal-to-



(a)

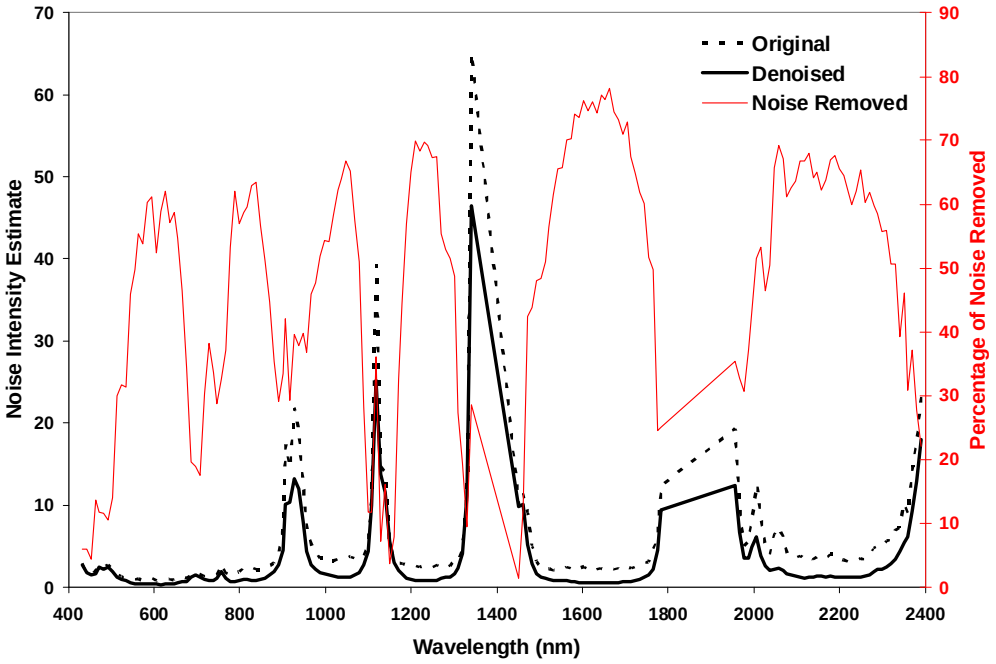


(b)

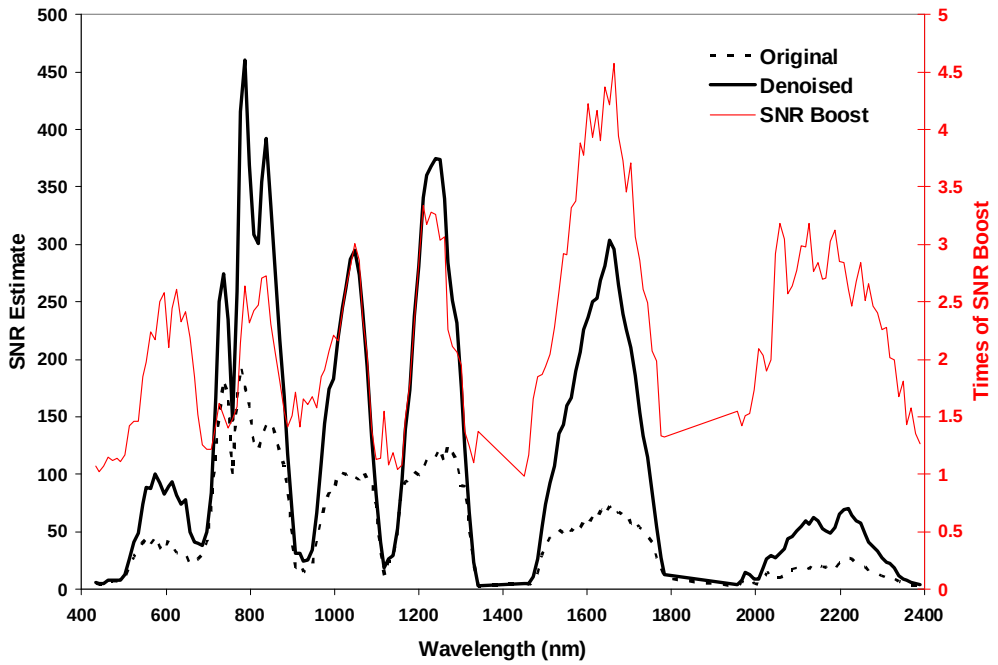
Figure 23. AVIRIS results of noise reduction by LGP (a) Noise intensity estimate and (b) Relative signal-to-noise ratio

noise ratio was then obtained by dividing the latter by the former. In the context of this chapter, both noise intensity and signal-to-noise are unit-less noise estimates. As shown in Figure 23a, the average noise intensity of the original AVIRIS image was 1.2, which

varied between 0.3 (minimum) and 11.3 (maximum) across all bands. The least noisy band occurred at 683.2 nm and the noisiest band was at 1078.8 nm. The LGP removed an



(a)



(b)

Figure 24. Hyperion results of noise reduction by LGP (a) Noise intensity estimate and (b) Relative signal-to-noise ratio

average of 34.4% of the noise in the AVIRIS image, which left the noise intensity in the

post-denoised image varying between 0.2 and 9.0 with the average of 0.9. The improvement on relative SNR by LGP was summarized in Figure 23b. The average, minimum, and maximum SNR post-denoising were 300.6, 5.8, and 1366.5 across all bands, which corresponded to 177.6, 5.4, and 581.4 respectively pre-denoising. The SNR was increased 1.6 times on average.

Similarly, LGP's ability on noise reduction for the Hyperion image was demonstrated in Figure 24, where LGP achieved an average of 48.6% noise reduction and 2.2-time SNR increase. Prior to noise reduction, the average noise intensity of the original Hyperion image was 4.9, which varied between 0.8 and 65.0 across all bands. The least noisy band occurred at 574.7 nm and the noisiest band was at 1340.9 nm. The average SNR was 52.5, which fluctuated between 2.5 and 193. After noise reduction, the noise values became 2.9, 0.4, and 46.5 for the average, minimum, and maximum noise intensity and 123.4, 3.4, and 460 for the average, minimum, and maximum SNR, respectively.

The results shown in Figures 23 and 24 were derived using uniform neighbourhoods, of which each contained 150 state vectors, i.e. $k = 150$. It was found that only a single iteration was required before the denoising converged. Their parameters were determined based on several tests by using different sizes of neighbourhood and numbers of iterations. They seemed to be the optimal choice for the hyperspectral images under consideration. It was also found from these tests that the denoising results were not sensitive to the choice of the input parameters. Usually good results could be derived by using a set of parameters, of which each might vary in a certain range. This finding was consistent with the findings of others [89 – 93] with the nonlinear time series data.

Two things were found interesting from the above results, which were worth further exploring. Firstly, though a significant amount of noise had been removed by the LGP for both AVIRIS and Hyperion images, there was still some left, about half on average, to be specific. One might expect a higher percentage of noise reduction. This lower-than-expectation noise reduction was likely due to the method of noise estimation. It should be noted that the noise estimated was actually "the unexplained standard

deviation" [99], which might include more than just noise. For instance, the inter-pixel variance was included. The existence of additional variance might cause overestimating the noise and underestimating the result of noise reduction. Secondly, denoising was considered as a two-sided sword. It might remove not only noise but useful information as well if not conducted properly. The ultimate goal of a noise reduction algorithm for hyperspectral data was to reduce the noise as much as possible without reducing the information content. Spectral absorption features, among others, are considered as the primary constituent of information content borne by a hyperspectral image. They provided diagnostic information for material identification. For forest applications of hyperspectral remote sensing, spectral absorption features have been utilized to study forest health and chemistry. It is the special interest of this chapter to investigate if the LGP denoising is capable of maintaining spectral absorption features while conducting noise reduction.

There are many absorption features across each hyperspectral spectrum, which can be characterized in terms of spectral shape. Assume that a hyperspectral image correctly captured all absorption features. It is desired that the denoising maintains these features so that the spectral shape stays the same as much as possible between pre- and post-denoising. The preservation of spectral shape was evaluated using the spectral angle between a paired pixel spectra, i.e. ($pixel_{pre\text{-}denoising}$, $pixel_{post\text{-}denoising}$). Selecting the spectral angle as the metric to examine spectral shape preservation is due to its insensitivity to variation of illumination intensity caused by changes of viewing geometry. Considering that forest and inland water were prominent land-cover types within the image scene, the forest and water pixels marked in Figure 17 were extracted from both the original and the denoised images for spectral angle calculation. They included 10,000 AVIRIS and 10,000 Hyperion pixels, respectively, which were equally separated into forest and water. The statistics of the calculated spectral angles were shown in Figure 25. It was revealed that the LGP denoising caused limited spectral shape distortion. For the AVIRIS forest pixels, the average distortion was 0.3 degrees, which varied between 0.1 and 2.2. The distortion became more apparent for AVIRIS water pixels, which was 4.9 degrees on average varying between 1.2 and 47.3. As to the Hyperion pixels, the average, minimum, and maximum distortions were 1.3, 0.4, and 8.4 for the forest and 6.0, 1.9, and 19.6 for the

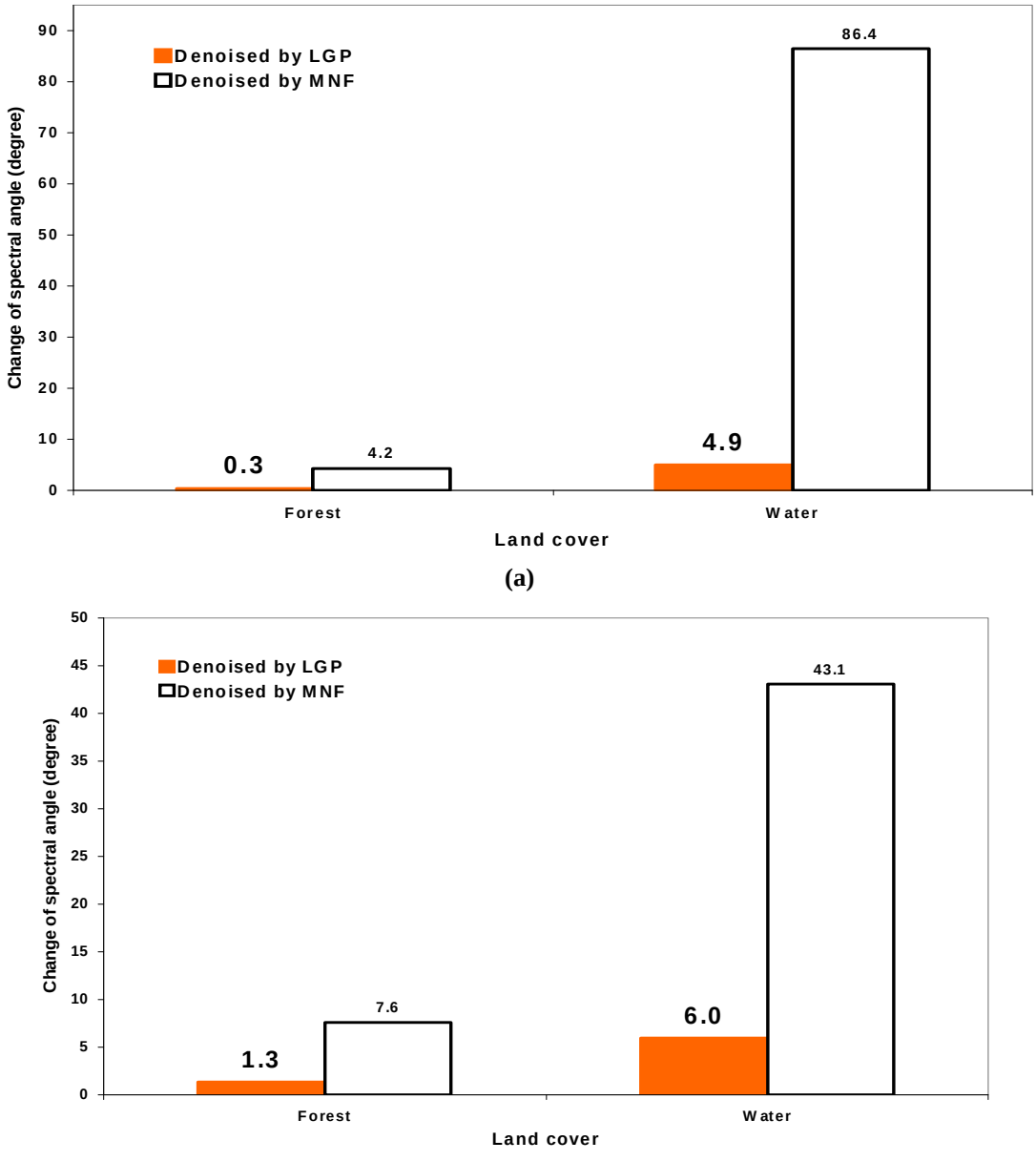


Figure 25. Impact of noise reduction on spectral shape evaluated as spectral angle changes (a) AVIRIS (b) Hyperion

water, respectively. It seemed that the LGP caused more distortions for the spectral shape of water pixels than those of forest pixels. This was due to the fact that water was a target of very low reflectance. Because of this, even a small amount of noise in low intensity was capable of contaminating the signal and destabilizing the airborne and space-borne water spectra. This caused less consistent spectral shape of water pixels between pre- and post-denoised images.

For comparison purposes, the legacy denoising algorithm, MNF, was employed to reduce the noise in the same hyperspectral images. The denoised image was constructed using the 10 high ranking MNF components. This number corresponded to the intrinsic dimensionality identified in Chapter 3. As depicted in Figure 25, for the AVIRIS image, the average, minimum, and maximum distortions were 4.2, 0.6, 27.1 based on the forest pixels, while the values based on the water pixels were 86.4, 10.7, and 143.7 correspondingly. For the Hyperion image, these values were accordingly 7.6, 2.4 and 21.8 for the forest and 43.1, 13.4, and 92.6. Apparently MNF caused more spectral shape distortions than the LGP for both forest and water pixels. This amount of distortion was enough to alter spectral absorption features, which might jeopardize applications relying on spectral analysis of hyperspectral data.

4.4 Computational complexity

As discussed in Section 4.2, the Local Geometric Projection algorithm could be implemented in the following steps: 1) construction of state vectors in phase space; 2) establishment of neighbourhood; 3) determination of orthogonal vectors; 4) projection for noise reduction; and 5) data restoration. Among these steps, Step 2 – 4 might need to be executed iteratively until no further noise reduction is discernable. The first and last steps were least expensive in computing, which were both linear, i.e. $O(dl)$ where d and l are the embedding dimension and the length of the data series, respectively. The second step involved a k-nearest neighbour search, which required $O(kdl^2)$, where k is the neighbourhood size. Step 3 included construction and decomposition of a covariance matrix, which was upper-bounded by $O(kd^2)$ for each state vector. As the total number of state vectors could be as many as l, this added up to $O(lkd^2)$. Step 4 was a linear transformation done by a matrix multiplication for each state vector. The cost of this step was $O(kd^2)$ per state vector, which was added up to $O(lkd^2)$ for the entire data series.

Considering that $l >> d$ and k for the data series generated using hyperspectral pixels, the most costly step above all was the second one: establishment of neighbourhood by k-nearest neighbours, which was $O(kdl^2)$. It should be noted that the above estimation was made based on a single iteration of the LGP algorithm. It would be

even higher if more-than-one iterations were required. Compared to the MNF that was upper-bounded by $O(ld^2)$ as discussed in Section 3.4, the LGP increased the computational complexity dramatically. In order for the LGP to be able to deal with normal-sized hyperspectral images, one has to either modify the nearest neighbour search algorithm, implement it in a parallel environment, or does both. For instance, the original nearest search algorithm could be improved for complexity reduction by some pre-processing techniques, including linear search, space partitioning, or locality sensitive hashing [101]. These approaches are out of consideration in this dissertation,

4.5 Chapter summary

A nonlinear noise reduction algorithm, called Local Geometric Projection (LGP), which was originally proposed in nonlinear time series data analysis, was adapted in this chapter for hyperspectral data denoising. This adaption was made by taking into consideration the true nature and characteristics of hyperspectral data, specifically including the nonlinear data distribution and the existence of signal-dependent and correlated noise, which were either ignored or overly simplified by many legacy noise reduction algorithms in hyperspectral remote sensing. The LGP was developed with the following geometric picture in mind that noise, existing as higher dimensional signals, displaced the signals from their true locations. The displacement could be locally restored by linearly projecting data onto the directions represented by the majority of data variance. This was quite different from many legacy algorithms, including PCA, MNF, and WT, which linearly transformed data on a global scale. As listed in Tables 3 and 4, the LGP-based denoising, on average, achieved ~50% noise reduction and ~2 times SNR increment.

Table 4. Noise reduction

	Mean	Minimum	Maximum	Standard Deviation
Original AVIRIS	1.2	0.3	11.3	1.3
Denoised AVIRIS	0.9	0.2	9.0	1.1
Original Hyperion	4.9	0.8	65.0	6.8
Denoised Hyperion	2.9	0.4	46.4	5.0

Table 5. SNR increment

	Mean	Minimum	Maximum	Standard Deviation
Original AVIRIS	177.6	5.4	581.4	178.9
Denoised AVIRIS	300.6	5.8	1366.5	338.7
AVIRIS SNR Boost	1.6	1.0	2.8	0.4
Original Hyperion	52.5	2.5	193.8	45.3
Denoised Hyperion	123.4	3.4	460.6	113.6
Hyperion SNR Boost	2.2	1.0	4.6	0.8

It is worth mentioning that the above noise reduction and SNR boost were obtained without compromising the fidelity of spectral absorption features. Comparing to the distortions introduced by the MNF-based denoising, the distortions caused by the LGP denoising was minimal (Table 5). It should be noted that selecting a proper noise reduction algorithm for hyperspectral data depends on the applications on hand. For classification-based applications, PCA- or MNF-based denoising may still be preferable, as it is capable of producing maximum noise reduction. On the other hand, for spectral analysis based applications, such as matching spectral absorption features and developing narrow band spectral indices, the LGP-based denoising is a better choice, which is able to provide decent noise reduction while preserve pixel spectral shapes.

Table 6. Spectral deformation

	Mean	Minimum	Maximum	Standard Deviation
AVIRIS Forest by LGP	0.3	0.1	2.2	0.2
AVIRIS Forest by MNF	4.6	0.6	27.1	2.9
AVIRIS Water by LGP	4.9	1.2	47.3	2.9
AVIRIS Water by MNF	86.4	10.7	143.7	25.3
Hyperion Forest by LGP	1.3	0.4	8.4	0.7
Hyperion Forest by MNF	7.6	2.4	21.8	2.7
Hyperion Water by LGP	6.0	1.9	19.6	1.8
Hyperion Water by MNF	43.1	13.4	92.6	12.5

The LGP denoising, similar to the false near neighbour algorithm discussed in Chapter 3, had a very expensive step of computation: nearest neighbour search. Compared to the MNF-based denoising that required $O(ld^2)$, the LGP-based algorithm needed $O(kdl^2)$ for a single iteration. To make it accommodate normal-sized

hyperspectral data within a reasonable time frame, this algorithm needs to be optimized either by using alternative methods for the nearest neighbour search or implementing it in a parallel computing environment (cluster computing for example).

Chapter 5
Algorithm implementation

Three algorithms were introduced in this dissertation to address the nonlinearity in hyperspectral remotely sensed imagery, including testing for nonlinearity, determination of intrinsic dimensionality, and noise reduction. This chapter is dedicated to the implementation of these algorithms so that they are accessible for those who are interested in studying nonlinear behaviours in hyperspectral data. As these algorithms are intrinsically intensive in computational complexity, approaches are discussed in details regarding how to relieve the computational restriction by replacing the exact solutions with some approximated ones. This provides a reasonable trade-off between the computing time required and the quality of the derived results. The selection of programming language and the associated application programming interface (API) are also discussed in this chapter.

5.0 Motivation

Though the existence of nonlinearity in hyperspectral data was known in the remote sensing community, the nonlinearity has not been well addressed in information extraction from hyperspectral data. This is due primarily to lacking of the dedicated application software off-the-shelf. For example, among the well recognized software packages for hyperspectral data processing and analysis, including ENVI [102], Geomatica [103], and ERDAS IMAGINE[104], there was no tool or function available to deal specifically with the nonlinearity. The algorithms implemented in these packages, such as segmentation, classification, and spectral mixture analysis, were all based on the assumption that hyperspectral data were realizations of some linear stochastic processes. This is contradicting to the physical processes of hyperspectral data formation and what had been found in Chapter 2 of this dissertation.

This chapter was prepared trying to fill the gap between the interest of investigating the nonlinearity in hyperspectral data and availability of proper tools by implementing the three algorithms introduced in Chapters 2, 3, and 4, respectively, including testing for nonlinearity, estimation of dimensionality, and noise reduction. It is expected that the availability of these tools would foster more interest and advocate more research efforts towards studying the nonlinear behaviour of hyperspectral data. As these algorithms were originally developed in the field of nonlinear time series analysis, implementing them and making them available in the context of hyperspectral remote sensing would also promote adapting other algorithms of nonlinear time series analysis for hyperspectral data processing and analysis.

5.1 Application programming interface (API)

The first thing encountered in order to implement these algorithms was to select a proper programming language that provided suitable data structures, powerful libraries of routines, and user-friendly developing environment, which are, in general, referred to as application programming interface (API). The API selection had to be conducted in the context of hyperspectral remote sensing, which is capable of acquiring a large volume of data in a short period of time and the acquired data are saved and delivered in digital images composed of hundreds of channels. Each channel corresponds to a spectral band operated by the spectrometer inside a hyperspectral imager.

A hyperspectral image is usually stored and manipulated in a 3-dimensional array internally, where the first 2-dimensions provide spatial information while the third dimension provides spectral information. The following three factors are often considered in the API selection. Firstly, hyperspectral data analysis relies heavily on array-based operations, which often require dynamic expansion or contraction on array size. Secondly, the introduced algorithms in this dissertation involve extensive multivariate analysis, including matrix arithmetic and factorization and eigen analysis, which needs the API of interest to have a powerful mathematic library to support this kind of operations. Lastly, the selected API has to provide programmable mechanisms for flexible and efficient memory management, because a normal-sized hyperspectral image consumes a significant amount of memory. Based on the above considerations, the

Interactive Data Language (IDL) [105] is chosen as the API for the implementation of the algorithms introduced in this dissertation.

IDL is a complete computing environment for both interactive and programmable applications, which provides the vectorized computing capability to process large amounts of data efficiently. The prominent advantage of IDL over other programming languages, such as C^{++} and JAVA, is its offering of array programming, wherein operations can be directly conducted to whole aggregates of data without having to resort to explicit loops of individual scalar operations. Though much of IDL's syntax is inherited from FORTRAN and C, IDL is designed from the beginning for high level image processing applications. The key features of IDL are summarized as follows.

- IDL is a complete, structured language that can be used in both interactive and programmable fashion to create sophisticated functions and procedures;
- IDL is byte-code compiled to a runtime engine, similar to Java, and implements an efficient array processing engine combined with a large library of analysis routines;
- Rapid multi-dimensional plotting, volume visualization, image display and animation make it possible to observe the computational results immediately;
- IDL widgets can be used to quickly create multi-platform graphical user interfaces (GUI) for applications;
- IDL supports dynamic binding that enables easy variable declaration and definition without type checking. Variable types can be modified during runtime;
- IDL applications run the same across all supported platforms (UNIX, LINUX, Windows, and Macintosh systems) with little or no modification.

What makes IDL an excellent choice for the task is, concisely, the combination of array processing, image visualization, and powerful library of data analysis routines. Programming in IDL is a timesaving and effective alternative to programming using other languages. A bonus for choosing IDL is that any applications developed using IDL

can be integrated seamlessly into ENVI, a well-known commercial hyperspectral data analysis package, which is coded in IDL itself.

5.2 Algorithm implementation

To facilitate describing algorithm implementation which is intended for hyperspectral data only, the following notation adopted throughout this dissertation is restated as follows: 1) matrices are denoted by upper-case characters; 2) vectors by array-capped lower-case characters; and 3) scalars by lower-case characters. Based on the above notations, a hyperspectral image with r rows, c columns, and n bands is represented by a matrix, X, composed of p column vectors, i.e. $X = [\vec{x}_1, \vec{x}_2, ..., \vec{x}_p]$, where $p = c \times r$, denoting the number of pixels included in the image. The dimension of each column vector $\vec{x}_i$ included in X is n, equivalent to the number of spectral bands. For the AVIRIS image of interest in this dissertation, $r = 1000$; $c = 550$; and $n = 179$. For the Hyperion image of interest, these numbers are 387, 273, and 179, respectively. The algorithms introduced in this dissertation are detailed, using the aforementioned notations, in Figures 26 – 28, where Figure 26 depicts the dataflow of testing for nonlinearity, Figure 27 the dataflow of determination of dimensionality, and Figure 28 the data flow of noise reduction. Thought the implementation of these algorithms is straightforward, two things are worth mentioning here: how to deal with large image files given the limited memory space; and how to speed up the process so that the waiting time is tolerable.

A. Dealing with the limited memory space

As shown in Figures 26 – 28, hyperspectral image input and transformation are common steps included in all three algorithms to be implemented. Considering that hyperspectral images are often given in raw binary format for efficient storage and the associated metadata are provided in separate ASCII header files, hyperspectral images can be conveniently loaded into memory using corresponding IDL I/O routines if enough memory space is available. The loaded images are stored in 2-dimensional arrays (matrices), whose sizes can be predetermined by parsing the associated ASCII header files. In a Windows XP environment where these algorithms were implemented,

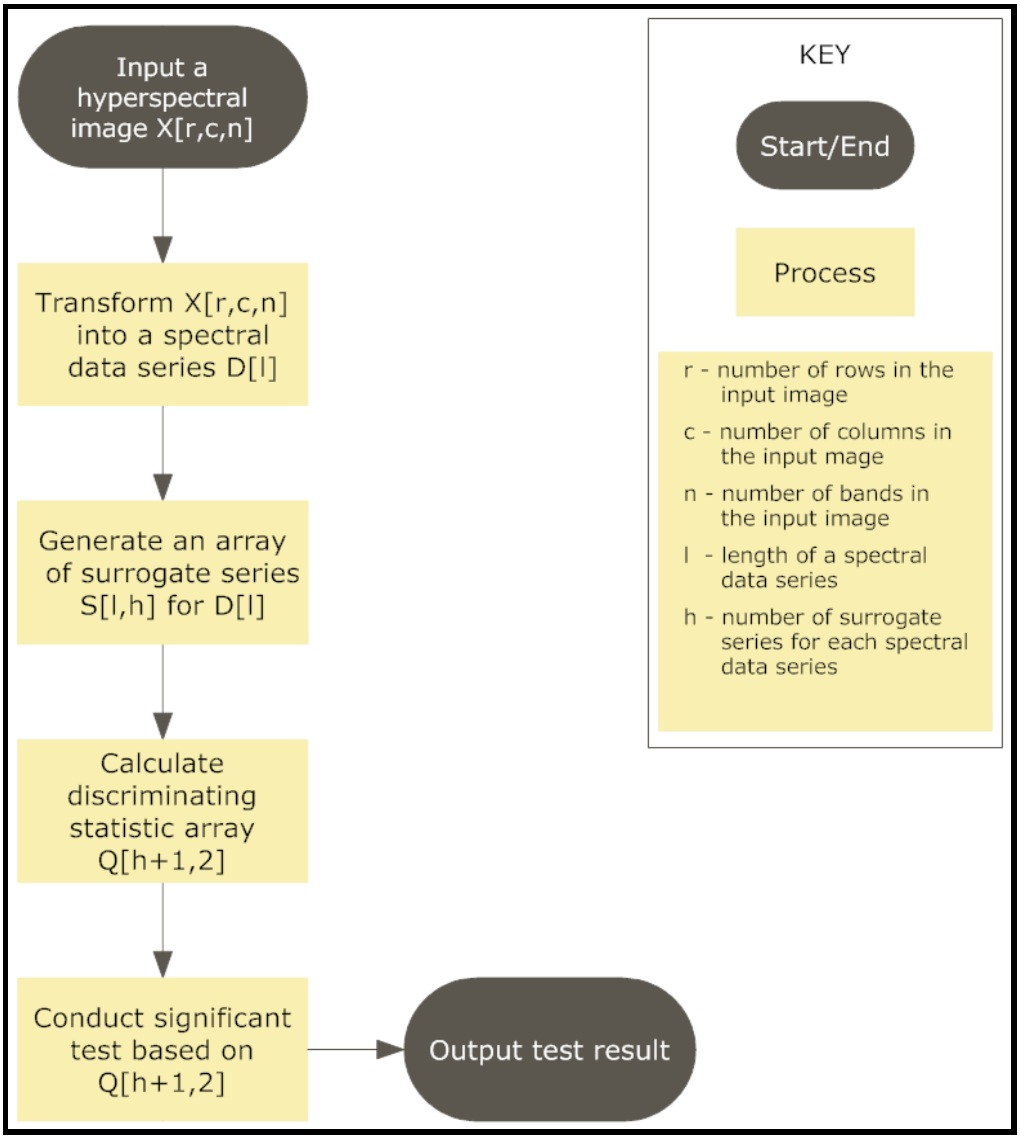


Figure 26. Dataflow of testing for nonlinearity

however, the maximum memory space allocated for applications is limited to 2 GB. This is not a whole lot of memory space considering the sheer size of hyperspectral image files, which are often composed of hundreds of bands in 16-bit depth. Though the 2-GB limitation is fine for the two hyperspectral data sets considered in this dissertation, consideration should be taken to accommodate hyperspectral images in a larger volume. To this end, both complete and incremental data input mechanisms were implemented to deal with normal- and over-sized hyperspectral data. For a normal-sized hyperspectral image that is under the 2-GB limit, the entire image is loaded into memory in one shot.

This ensures that the loading time is minimized, as it does not require frequently hard drive access.

For an over-sized hyperspectral image, the incremental data input kicks in. It

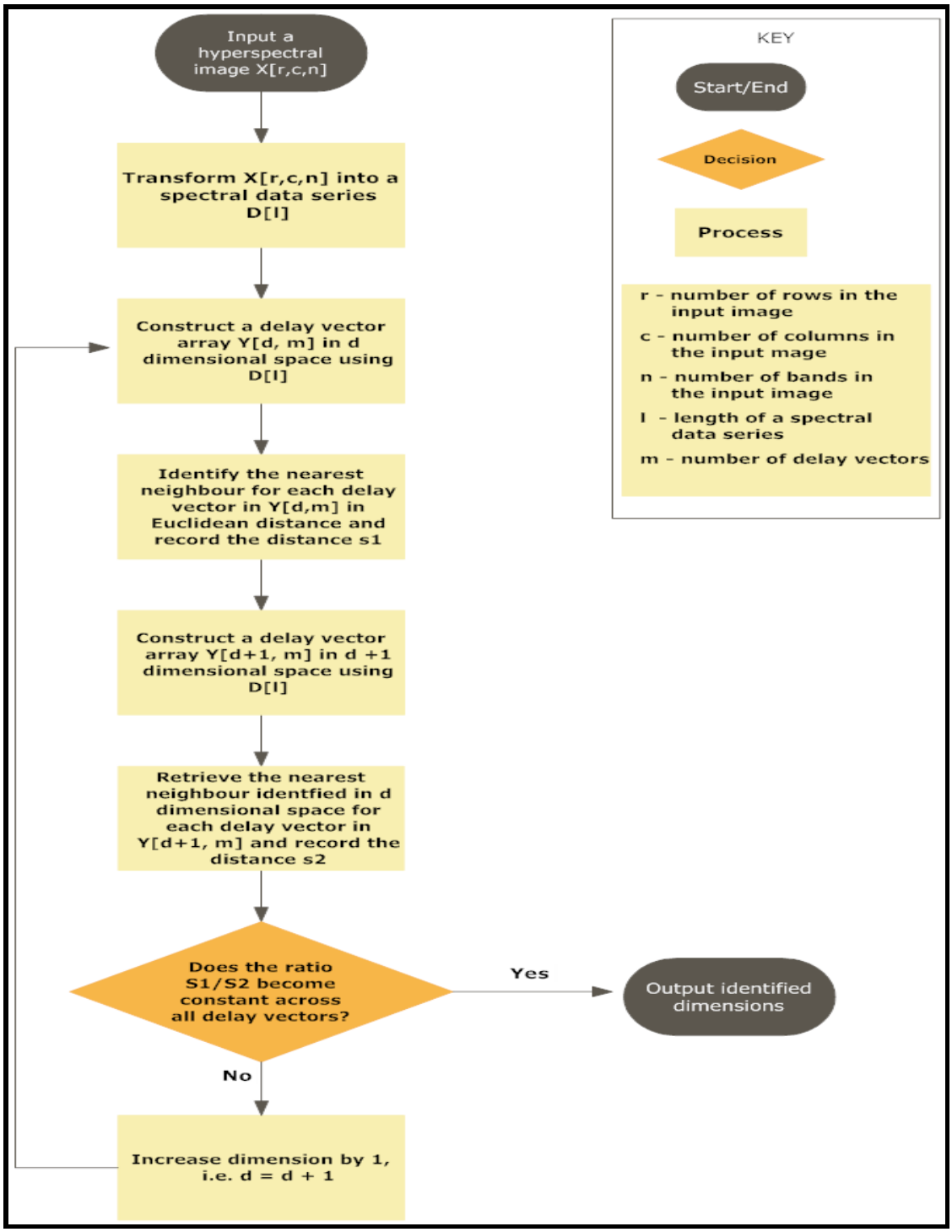


Figure 27. Dataflow for determination of dimensionality

loads only enough pixels to generate a partial set of data series. To ensure the performance stability, each data series should be generated with a nominal length, which was experimentally determined at 5,000. This is a soft threshold that can be adjusted based on the hyperspectral data of consideration. For example, given a hyperspectral

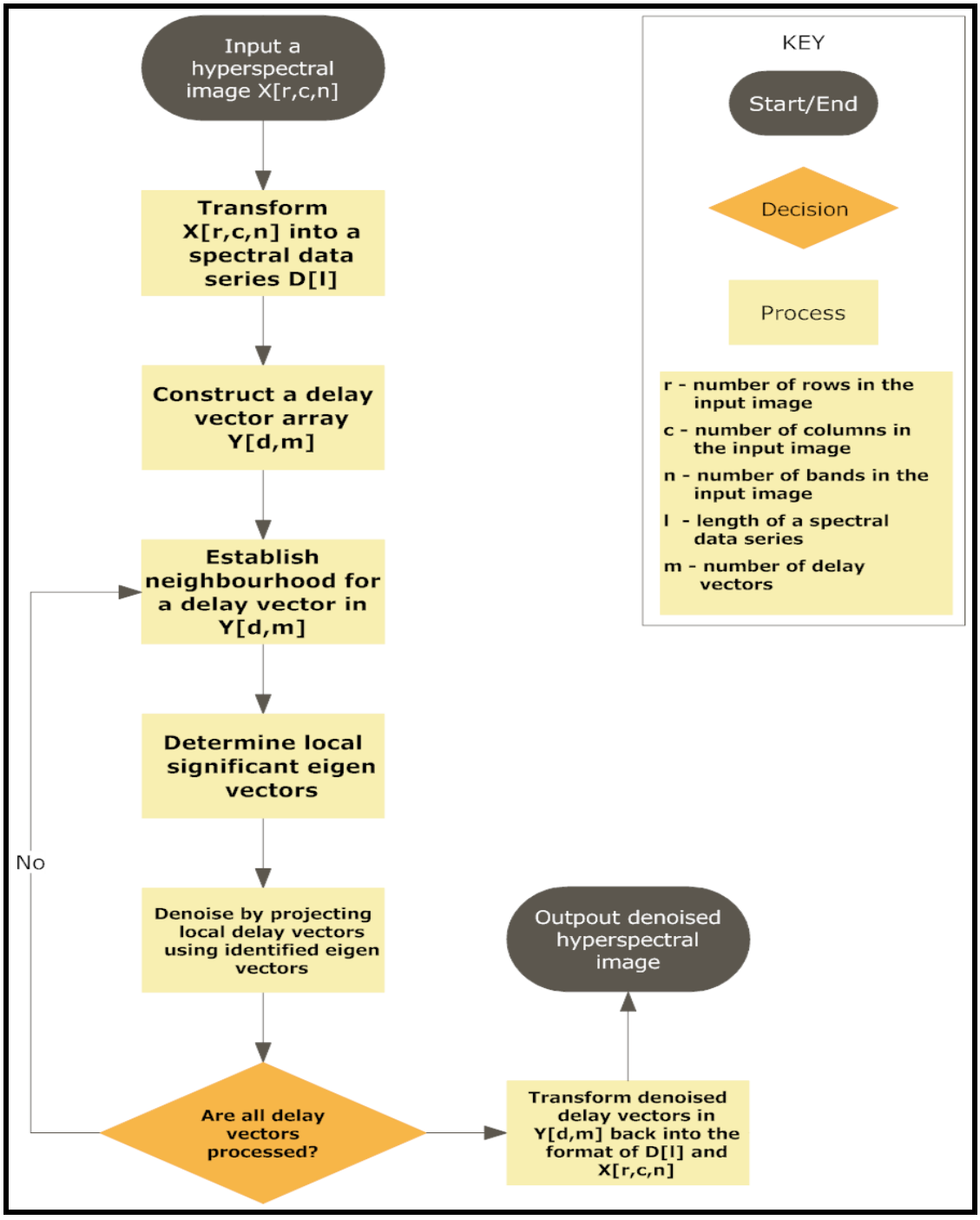


Figure 28. Dataflow of noise reduction

image composed of 200 bands and 10,000,000 pixels in 16-bit depth, which amounts to ~3.7 GB in total, it can be loaded incrementally in two steps. In the first step, about 1.8 GB of image or equivalently 5,000,000 pixels are loaded, from which 200,000 data series in the above nominal length are generated. The second step of loading is followed after the data loaded in the first step finishes processing and the occupied memory is released.

B. Improving processing speed

The computational complexity of the three algorithms described in the previous chapters is tremendous. For example, compared to the MNF-based algorithms that only needed $O(ld^2)$ for noise reduction, the proposed LGP algorithm required $O(kdl^2)$, where d, k, and l denoted embedding dimension, neighbourhood size, and length of data series, respectively. The surge in computational complexity causes significant increase on processing time, which makes it less practical to apply these algorithms to process hyperspectral images in any realistic sizes. To improve the processing speed, the following two approaches were taken during the algorithm implementation.

Firstly take full advantage of IDL's strength on array programming. Unlike object oriented programming that implicitly breaks down data to its constituent parts, array programming looks to group data and apply a uniform handling, which, in a sense, resembles parallel computing. IDL can be very fast doing array operations but quite slow if elements need to be processed individually. For numerically heavy computations using IDL, it is critical to treat each array as an undividable entity and avoid as much as possible accessing individual array elements. In addition, it is always time-saving by making use of the built-in array operations, instead of resorting to use any explicit loops.

Secondly if possible use approximate solutions instead of exact ones when implementing the computationally expensive steps. The noise reduction algorithm discussed in Chapter 4 included a very expensive step: neighbourhood establishment for each delay vector, which was $O(kdl^2)$. This step could be approximated by building up a common neighbourhood for all delay vectors within the neighbourhood. This approximation cut back on computational cost to $O(dl^2)$. As the number of neighbourhood to be established was greatly reduced because of this approximation, the

computational complexity of the subsequent steps to this one in the noise reduction algorithm was also reduced. For example, the cost of covariance matrix construction and decomposition dropped accordingly, from $O(lkd^2)$ then to $O(ld^2)$ now. Experiments revealed that the above approximation achieved ~20% computing time reduction with a very limited influence on the denoised results..

The algorithms introduced in this dissertation were coded using IDL by the full implementation of the approaches for memory saving and computing time reduction discussed above. The platform chosen for the implementation was a 32-bit Windows XP PC configured with a 2.33-Ghz Intel Core 2 Duo processor and 2-GB memory. Though more powerful platforms were available for this task, the one chosen here was more representative of common users and widely available. The coded algorithms were assembled into a software package, called “a system for nonlinear spectral series analysis of hyperspectral data”, which can be accessed through the graphic user interface as shown in Figure 29. This interface was coded using IDL’s widget functions, which provided a single entry point to access the implemented algorithms, including testing for nonlinearity, determination of dimensionality, noise reduction, and feature extraction. Options were offered in the interface to generate data series of different type and length. For example, in addition to the spectral data series that were used throughout this dissertation, spatial data series could be generated as well. Data series could be made with the minimum length of 5,000 up to 100,000. Execution feedback message was displayed in the “Status Update” panel in the interface. The feedback message was shown in multi-line text, which, depending on the input, included the progress of execution, error description if applicable, and summary of execution. This interface had the same feel and look as ENVI, as they were coded with the same language and in a similar fashion. This resemblance ensured that the interface and the implemented algorithms could be incorporated seamlessly into ENVI.

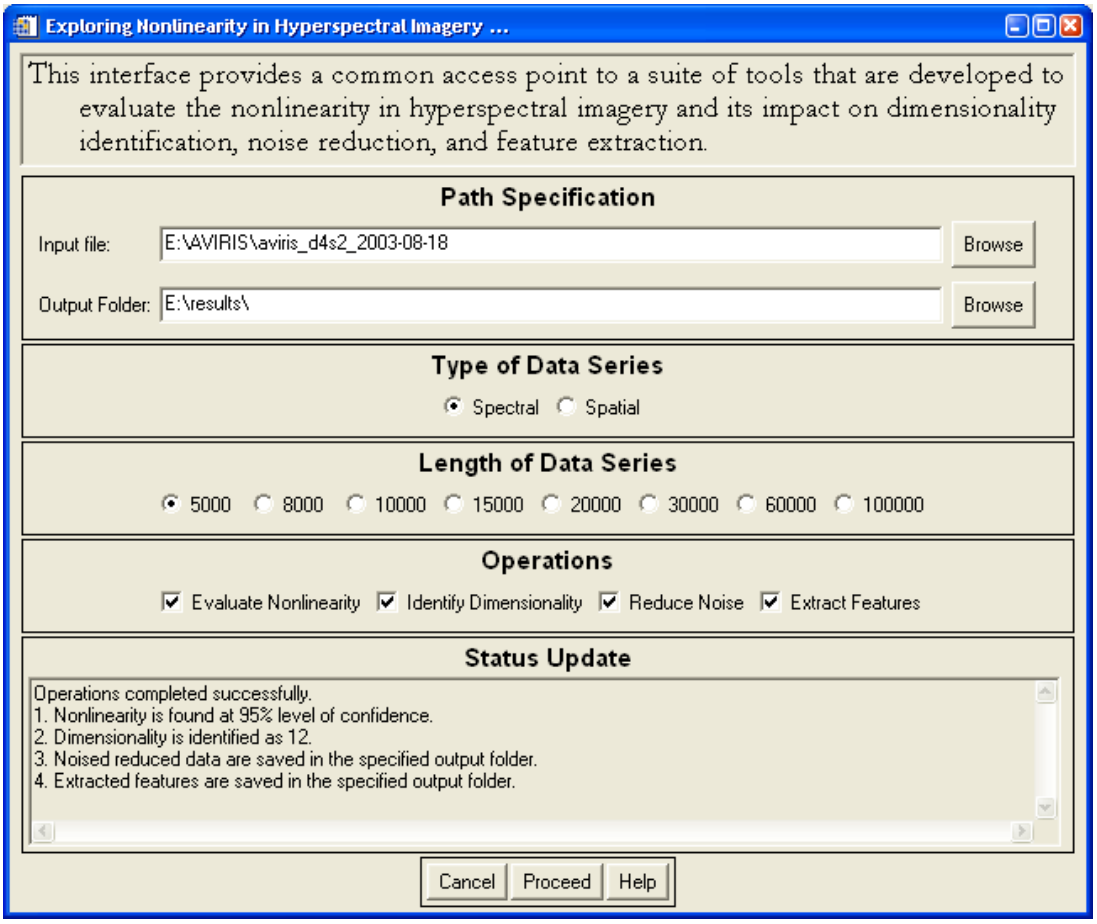


Figure 29. Common access interface

5.3 Chapter summary

This chapter provided details regarding implementing the algorithms introduced in this dissertation with the emphasis on the selection of application programming interface and the approaches to deal with memory and computing limitations. IDL, due primarily to its ability on effective array computing, was selected for the algorithm implementation. For large hyperspectral data sets, incremental data loading was offered to avoid running out of memory. Approximate algorithms were proposed and implemented, which achieved a ~20% computing time reduction. The approximation had a very limited impact on the derived result compared to those based on the exact algorithms, which was described in [39]. Despite adoption of these approaches, the required computing time is longer than that required by the conventional linear algorithms. For example, given the computing platform as described in the previous

paragraph and the AVIRIS image used in this dissertation, the MNF-based algorithm for noise reduction only needed an hour to finish while the proposed LGP algorithm required about 5 hours. Further reduction on computing time is still needed.

Chapter 6

Conclusions and contributions

It has been noticed that there was an inconsistency between the statistical characteristics of hyperspectral remotely sensed data and the methods employed to model and process the data for information extraction. On the one hand, hyperspectral data are considered inherently nonlinear, due to the multiple nonlinear sources involved in the data formation, including the interaction between solar radiation and atmosphere. On the other hand, hyperspectral data has long been modeled and processed as realisations of some linear stochastic processes. What is the impact of this inconsistency on hyperspectral data analysis? This dissertation was prepared dedicatedly to address this question by firstly evaluating the significance of nonlinearity in hyperspectral data and secondly examining the influence of nonlinearity on dimensionality estimation, feature extraction, and noise reduction. This chapter summarizes what had been found throughout this dissertation, contributions of this dissertation, and considerations for further work. Though the conclusions drawn in this dissertation are based on AVIRIS and Hyperion imagery only, these conclusions are expected to be applicable for other airborne or space-borne hyperspectral imagery due to the popularity and representativeness of these two hyperspectral sensors.

6.0 Conclusions

A. Hyperspectral data are nonlinear in spectral domain

The investigation of nonlinearity was conducted using the surrogate method and focused on the spectral domain nonlinearity associated with specific land-cover types of interest: forest and water. Using the simulated data series and the corresponding surrogates, it was demonstrated that both the discriminating statistics and the method of surrogate generation were working correctly. The difference was found statistically significant between the discriminating statistic values derived from hyperspectral data and those from the corresponding surrogates. The null hypothesis that hyperspectral data

were linear was rejected at the 99% level of confidence for the forest data series and at the 95% level of confidence for the water data series. It was also found that the nonlinear behaviours of hyperspectral data became less apparent for low signal targets. This indicated that the statistical characteristics of nonlinear data might be indistinguishable from those of linear ones when the data signal-to-noise ratio was very low.

B. The intrinsic dimensionality of hyperspectral data is lower than what was believed

Using the false near neighbour method that is capable of accommodating the nonlinearity in hyperspectral data, the dimensionalities of hyperspectral data were estimated between 7 and 10 across different land-cover types. With the legacy algorithms based on linear transformations, those dimensionalities, however, were found to be between 10 and 25. This indicated that hyperspectral data could be embedded tightly in a lower dimensional space if nonlinearity was addressed. Accurate estimation of dimensionality is important for hyperspectral data analysis using multivariate analysis. This may lead to using a fewer number of independent variables to describe hyperspectral data and reducing the subsequent computational complexity.

It was also found that the dimensionality might change among different land-cover types**.** This was believed to be related to the way of interaction between solar radiation and ground targets that introduced nonlinearity and noise of different intensity into hyperspectral data. The stronger the nonlinearity and the weaker the noise, the fewer dimensions were required to describe the data. The difference on signal-to-noise ratio was believed to be the cause of dimension discrepancy on forest data series between AVIRIS and Hyperion, where the low SNR Hyperion data series exhibited higher dimensions than the high SNR AVIRIS data series. These findings were agreeable to those reported in the nonlinear time series analysis.

C. The LPG-based nonlinear noise reduction algorithm is able to reduce noise while maintain spectral absorption features

The Local Geometric Projection (LGP), originally proposed in nonlinear time series analysis, was adapted for denoising hyperspectral data. This adaption was made by taking into consideration the nonlinear distribution and existence of signal-dependent and

correlated noise in hyperspectral data, which were either ignored or overly simplified by many legacy noise reduction algorithms. The LGP was developed with the following geometric picture in mind that noise displaced signals from their true locations. The displacement could be locally restored by linearly projecting data onto the directions represented by the majority of data variance. The LGP-based denoising, on average, achieved ~50% noise reduction and ~2 times SNR increment. The real attraction of this algorithm was that the above noise reduction and the SNR boost were obtained without compromising the fidelity of spectral absorption features. Compared to the spectral distortions introduced by the MNF-based denoising, the distortion caused by the LGP denoising was minimal.

It should be noted, however, that selecting a proper noise reduction algorithm for hyperspectral data analysis is application dependent. For classification-based applications, PCA- and MNF-based denoising may still be preferable, as they are capable of producing maximum noise reduction. On the other hand, for the spectral analysis based applications, such as matching spectral absorption features and developing narrow band spectral indices, the LGP-based denoising is a better choice, for it provides decent noise reduction while preserves spectral shapes.

D. Nonlinear algorithms for hyperspectral data analysis are more computationally expensive than their linear counterparts

Though the false near neighbour method could provide more accurate dimension estimation for hyperspectral data than the linear transformation-based methods, it was more expensive in computational time. Compared to MNF that required $O(pn^2)$ in computing time, the false near neighbour method needed $O(dp^3)$, where d, n, and p represented embedding dimension, number of bands, and number of pixels, respectively. Similarly, the LGP denoising had an expensive step of computation: the nearest neighbour search, whose upper bound was $O(kdl^2)$, where k and l denote neighbourhood size and the length of data series, respectively. Given the typical computing platform and the AVIRIS image used in this dissertation, the MNF-based noise reduction was finished within an hour, while the LGP-based one required about 5

hours. Efforts on reducing computing time are still needed.

6.1 Contributions

A. Provided a new perspective to analyze hyperspectral data

Hyperspectral data were usually analyzed from the following three perspectives: image, spectral, and feature, wherein a hyperspectral image was represented correspondingly by a set of image arrays, spectral profiles, and high-dimensional vectors. Many hyperspectral data analysis algorithms were developed from the feature perspective. Mathematically each hyperspectral pixel was treated as a random vector following the law of linear stochastic process, which led to adapting many of the methods developed in multivariate analysis to process hyperspectral data. From the feature perspective, one considered a hyperspectral image as some repetitive samples collected by a number of independent variables. However, as demonstrated by the hyperspectral covariance images, these variables were far from independent, but strongly correlated. Motivated by this observation, this dissertation provided a new perspective to look at hyperspectral data: spectral series, which was constructed by spectrally concatenating adjacent pixels. In this new perspective, hyperspectral data were thought as samples collected by a single variable and represented by one or a few spectral data series. The new perspective was preferable to other perspectives for studying inter-band relationship in hyperspectral data.

B. Statistically proved the existence of nonlinearity in hyperspectral data

Though hyperspectral data were believed to be nonlinear, which was inferred from the nonlinear processes and media involved in the formation of hyperspectral data, the existence of nonlinearity had never been formally proved. The only way to show the nonlinearity was through multi-band scatter-plots, wherein any observed nonlinear features were considered as the sign of nonlinearity. In this dissertation, the existence of nonlinearity in hyperspectral data was formally proved using a hypothesis test. The method consisted of generating surrogates, designing and calculating discriminating statistic metrics, and statistically testing the difference between the discriminating statistic values derived from real hyperspectral data and those from the surrogates. It was

revealed that hyperspectral data are nonlinear in spectral domain and the intensity of nonlinearity was dependent on land-cover types and signal-to-noise ratio.

C. Introduced and implemented a new algorithm to estimate dimensionality of hyperspectral data

The false neighbour method, originated in nonlinear time series analysis, was first introduced and implemented in this dissertation for the purpose of dimensionality estimation of hyperspectral data. The algorithm was invented based on an intuitive observation that the distance between a pair of neighbouring points in a high dimensional space became consistent if the geometric structure containing these points was fully unfolded. Compared to the legacy methods based on linear transformations, this algorithm was able to accommodate nonlinearity and produce more accurate and clear dimension estimates. As identified by this algorithm, data dimensions varied from one land-cover type to another. For a hyperspectral image composed of a multiple land-cover types, it was suggested to use the largest dimension estimate as the dimensionality of the entire image.

D. Introduced and implemented a new algorithm to reduce noise in hyperspectral data

Local Geometric Projection, the noise reduction algorithm implemented in this dissertation, also originated in nonlinear time series analysis, which could be considered as an extension to the algorithms that globally projected data into the space depicted by major eigenvectors. The algorithm was designed to restore data to their true geometric locations that were displaced by noise. It was demonstrated that this algorithm was superior to the global projection-based counterparts in maintaining spectral absorption features. The LGP-based denoising would find more applications in areas like spectral analysis, feature matching, and spectral index development.

6.2 Future work

The common theme of all the algorithms discussed in this dissertation was to look at hyperspectral data from a new perspective by converting the data into spectral data series, wherein they were processed separately in geometric locality, instead of global as performed by many linear algorithms. This approach is advantageous for maintaining the

geometric shape and topological relationship between the features in hyperspectral data, and hence improves the quality of information extraction. This conversion (from multivariate to univariate) and the geometric localization are great facilitators for examining the nonlinear behavior of hyperspectral data and assessing its impact on information extraction. But inevitably they caused an increase in computational complexity. Though approaches had been taken in Chapter 5 to reduce the complexity, further reduction is still needed before the introduced algorithms are applicable for operational uses requiring large hyperspectral images to be processed within a short time frame. This is an achievable goal if these algorithms can be implemented in a more powerful computing environment, such as parallel or cluster computing.

In this dissertation, the nonlinearity was addressed across the entire wavelength range of the spectra. Considering that the nonlinearity is originated from the interactions between solar radiation and atmosphere and between solar radiation and targets of interest, the nonlinearity exhibits in shorter wavelengths may be stronger than that in the longer wavelengths. Further research is required to address this topic.

Appendix A

List of publications

I. Peer-reviewed publications

[1] T. Han and D. G. Goodenough, “Investigation of nonlinearity in hyperspectral imagery using surrogate data methods,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2840 – 2847, 2008.

[2] D. G. Goodenough, T. Han, and A. Dyk, “Comparing and validating spectral unmixing algorithms with AVIRIS imagery,” *Canadian Journal of Remote Sensing*, vol. 34, pp. 82-91, 2008.

[3] M. A. Wulder, J.C. White, T. Han, N.C. Coops, J.A. Cardille, T. Holland, and D. Grills, “Monitoring Canada's forests. Part 2: National forest fragmentation and pattern,” *Canadian Journal of Remote Sensing*, vol. 34, no. 6, pp. 563-584, 2008

[4] N. C. Coops, M. A. Wulder, D. C. Duro, T. Han, and S. Berry, “Large area characterization of habitat using satellite data across Canada,” *Ecological Indicators*, vol. 8, pp. 754 – 766, 2008.

[5] T. Han, M. A. Wulder, J. C. White, N. C. Coops; M. F. Alvarez, and C. Butson, “An efficient protocol to process Landsat images for change detection with Tasselled Cap Transformation, *IEEE Geoscience and Remote Sensing Letters*, vol.4, pp. 147–151, 2007.

[6] M. A. Wulder, T. Han, J. C. White, C. R. Butson, and R. J. Hall, “An approach for edge matching large area satellite image classifications,” *Canadian Journal of Remote Sensing*, vol. 33, pp. 266-277, 2007.

[7] M. A. Wulder, T. Han, J. C. White, T. Sweda, and H. Tsuzuki, “Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization,” *Remote Sensing of Environment*, vol. 110, pp. 123-137, 2007.

[8] N. C. Coops, D. C. Duro, M. A. Wulder, and T. Han, “Estimating afternoon MODIS land surface temperatures (LST) based on morning MODIS overpass, location, and elevation information,” *International Journal of Remote Sensing*, vol. 28, pp. 2391-2396, 2007.

[9] D. C. Duro, N. C. Coops, M. A. Wulder, and T. Han, “Development of a large area biodiversity monitoring system driven by remote sensing,” *Progress in Physical Geography*, vol. 31, no. 3, pp. 1-26, 2007.

[10] D. G. Goodenough, A. Dyk, K. O. Niemann, J. Pearlman, H. Chen, T. Han, M. Murdoch, and C. West, "*Processing Hyperion and ALI for Forest Classification*," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1321-1331, 2003.

[11] T. Han, "An automatic system for hyperspectral remote sensing endmember unmixing," *Master thesis*, University of Victoria, Canada, 2003.

II. Conference proceedings

[1] T. Han and D. G. Goodenough, "Estimating dimensionality of hyperspectral data using false neighbour method," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 83-86, July 6–11, 2008, Boston, USA.

[2] T. Han and D. G. Goodenough, "Investigation of nonlinearity in hyperspectral remotely sensed imagery – a nonlinear time series analysis approach," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1556 – 1560, July 23–27, 2007, Barcelona, Spain.

[3] T. Han and D. G. Goodenough, "Nonlinear dimensionality reduction of hyperspectral data based on Locally Linear Embedding," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1237 – 1240, July 25–29, 2005, Seoul, Korea.

[4] D. G. Goodenough, T. Han, A. Dyk, J. Gour, and J. Li, "Forest biomass estimation with AVIRIS," *Presented in AVIRIS Earth Science and Applications Workshop*, May 24–27, 2005, Pasadena, USA.

[5] D. G. Goodenough, A. Dyk, T. Han, A. Hollinger, A. Jazayeri, and H. Zwick, "Hyperspectral sensor trades for forest applications," *Presented in International Astronautical Federation Congress*, Oct. 4–8, 2004, Vancouver, Canada.

[6] T. Han, D. G. Goodenough, A. Dyk, and H. Chen, "Feature selection for forest classification," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1471 – 1474, Sept. 20–24, 2004, Anchorage, USA.

[7] D. G. Goodenough, A. Dyk, T. Han, A. Jazayeri, and J. Li, "Impacts of lossy compression on hyperspectral products for forestry," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, pp. 465-468*, Sept. 20–24, 2004, Anchorage, USA.

[8] D. G. Goodenough, T. Han, J. Pearlman, A. Dyk and S. McDonald, "Forest chemistry mapping with hyperspectral data," *Proceedings of IEEE workshop on advances in techniques for analysis of remotely sensed data*, pp. 395 – 398, Oct. 27 - 28, 2003, Washington DC., USA.

[9] T. Han and D. G. Goodenough, "Hyperspectral endmember detection and unmixing based on linear programming," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1973-1766, July 21–25, 2003, Toulouse, France.

[10] T. Han, D.G. Goodenough, A. Dyk, and J. Love, “Detection and correction of abnormal pixels in Hyperion images,” *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pp. 1327-1330, June 24–28, 2002, Toronto, Canada.

Appendix B

List of Acronyms

AIS – Airborne Imaging Spectrometer

AISA – Airborne Imaging Spectrometer for Applications

ALI – Advanced land Imager

API – Application programming interface

AVIRIS – Airborne Visible/Infrared Imaging Spectrometer

BRDF – Bi-directional Reflectance Distribution Function

CASI - Compact Airborne Spectrographic Imager

CCD - Charge Coupled Device

CEM – Constrained Energy Maximization

CHRIS - Compact High Resolution Imaging Spectrometer

CSA – Canadian Space Agency

DAFE – Discriminant Analysis Feature Extraction

DBFE – Decision Boundary Feature Extraction

EnMAP – Environmental Mapping and Analysis Program

ENVI – Environment for Visualizing Images

EO-1 – Earth Observation 1

EROS - Earth Resources Observation and Science

fPAR - fractional photo-synthetically active radiation

HERO - Hyperspectral Environment and Resource Observer

HUDFA – HERO Utility Development for Forestry Applications

HYDICE - Hyperspectral Data and Information Collection Experiment

HYMAP - Hyperspectral Mapper

GIS – Geographic Information System

GVWD – Greater Victoria Watershed District

IDL – Interactive Data Language

IGARSS – International Geoscience and Remote Sensing Symposium

ISOMAP – Isometric Mapping

JPL – Jet Propulsion Laboratory

LAI – Leaf Area Index

Landsat ETM – Landsat Enhanced Thematic Mapper

Landsat MSS – Multiple Spectral Scanner

LIDAR – Light Detection and Ranging

LLE – Locally Linear Embedding

LPNNR - Locally Projective Nonlinear Noise Reduction

MAIS - Modular Airborne Imaging Spectrometer

MNF – Maximum Noise Fraction

NASA - National Aeronautics and Space Administration

OSP – Orthogonal Subspace Projection

PCA – Principal Component Analysis

PDF – Probability density function

PFC – Pacific Forestry Centre

PLS – Partial Least Squares

PP – Projection Pursuit

PROBA - Project for On Board Autonomy

RADAR – Radio Detection and Ranging

SAM – Spectral Angle Mapper

SNR – Signal to Noise Ratio

SVD – Singular Value Decomposition

USGS - United States Geological Survey

Bibliography

[1] E. J. Milton, N. P. Fox, and M. E. Schaepman, “Progress in field spectroscopy,” *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1966 – 1968, Denver, U.S.A., July 31 – Aug. 4, 2006.

[2] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, “Imaging spectrometry for Earth remote sensing,” *Science*, 228, pp. 1147–1153, 1985.

[3] Specim Spectral Imaging LTD, http://www.specim.fi/

[4] J. E. Conel, R. O. Green, G. Vane, C. J. Bruegge, R. E. Alley, and B. J. Curtiss, “Airborne imaging spectrometer-2: radiometric spectral characteristics and comparison of ways to compensate for the atmosphere,” *SPIE*, vol. 834, pp. 140–157, 1987.

[5] ITRES, ITRES Research Ltd., “CASI-2 Sensor,” 2001 http://arsf.nerc.ac.uk/documents/casi2.pdf

[6] W. M. Porter and H. T. Enmark, “A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS),” *SPIE*, vol. 834, pp. 22-31, 1987.

[7] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, “The HYPMAP airborne hyperspectral sensor: the system, calibration and performance,” *Proceedings of 1st EARSEL Workshop on Imaging Spectroscopy*. Zurich. Oct. 1998.

[8] R. W. Basedow, D. C. Armer, and M. E. Anderson, “HYDICE system: implementation and performance,” *SPIE*, vol.2048, pp. 258–267, 1995.

[9] Q. Tong, L. Zheng, and Y. Xue, “Development and application of hyperspectral remote sensing in China,” *SPIE*, vol. 3502, pp. 2-9, 1998.

[10] J. Pearlman, S. Carman, C. Segal, P. Jarecke, and P. Barry, “Overview of the Hyperion imaging spectrometer for the NASA EO-1 mission,” *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3036-3038, Sydney, Australia. July 9-13, 2001.

[11] M. A. Cutter, L. S. Johns, D. R. Lobb, and T. L. Williams, “Flight experience of the compact high resolution imaging spectrometer (CHRIS),” *SPIE*, vol.5159, pp. 392-404, 2004.

[12] USGS Earth Observing 1 (EO-1), http://eo1.usgs.gov/

[13] A. Mueller, R. Richter, M. Habermeyer, H. Mehl, S. Dech, H. J. Kaufmann, K. Segl, P. Strobl, P. Haschberger, and R. Bamler, "ARES: a new reflective/emissive imaging spectrometer for terrestrial applications," *SPIE*, vol. 5574, pp. 120–127, 2004.

[14] T. Stuffler, C. Kaufmann, S. Hofer, K. P. Forster, G. Schreier, A. Mueller, A. Eckardt, H. Bach, B. Penne, U. Benz, and R. Haydn, "The EnMAP hyperspectral imager – An advanced optical payload for future applications in Earth observation programmes," *Acta Astronautica 61*, pp. 115–120, 2007.

[15] D. G. Goodenough, A. Dyk, K. O. Niemann, J. S. Pearlman, H. Chen, T. Han, M. Murdoch, and C. West, "Processing Hyperion and ALI for forest classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1321–1331, 2003.

[16] E. Underwood, S. Ustin, and D. DiPietro, "Mapping nonnative plants using hyperspectral imagery," *Remote Sensing of Environment*, vol. 86, pp. 150–161, 2003.

[17] D. G. Goodenough, T. Han, J. Pearlman, A. Dyk, and S. McDonald, "Forest chemistry mapping with hyperspectral data," *Proceedings of IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp. 395-398, Washington DC., U.S.A., Oct. 27 – 28, 2003.

[18] M. Schlerf, C. Atzberger, and T. Hill, "Remote sensing of forest biophysical variables using HyMap imaging spectrometer data," *Remote Sensing of Environment***,** vol. 95, pp. 177-194, 2005.

[19] H. M. Horwitz, R. F. Nalepka, and P. D. Hyde, " Estimating the proportions of objects within a single resolution element of a multispectral scanner," *Proceedings of 7th International Symposium on Remote Sensing of the Environment*, University of Michigan, pp. 1307-1320, 1971.

[20] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Couer d'Alene river valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sensing of Environment*, vol. 59, pp. 64–76, 1997.

[21] D. G. Goodenough and T. Han, and A. Dyk, "Comparing and validating spectral unmixing algorithms with AVIRIS imagery," *Canadian Journal of Remote Sensing*, vol. 34, pp. S82-s91, 2008.

[22] H. Ren, Q. Du, J. Wang, C.-I. Chang, J. O. Jensen, and J. L. Jensen, "Automatic target recognition for hyperspectral imagery using high-order statistics," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 4, pp. 1372-1385, 2006.

[23] P. Goovaertsa, T. Geoffrey, M. Jacqueza, and A. Marcusb, "Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies," *Remote Sensing of Environment*, vol. 95, no 3, pp. 351–367, 2005.

[24] ESSI, "Earth Search's Probe-1 Sensor," 2004 http://www.earthsearch.com/Earth_Search's_Probe_1_Sensor.htm

[25] M. A. Cho and A. K. Skidmore, "A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method," *Remote Sensing of Environment,* vol. 101, pp. 181-193, 2006.

[26] P. J. Zarco-Tejada, J. R. Miller, A. Morales, A. Berjón, and J. Agüera, "Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops," *Remote Sensing of Environment*, vol. 90, pp. 463-476, 2004.

[27] D. A. Landgrebe, "Signal Theory Methods in Multispectral Remote Sensing," John Wiley & Sons, 2003.

[28] NASA LANDSAT 7, 2005, http://landsat.gsfc.nasa.gov/

[29] D. G. Goodenough, H. Chen, D. Horspool, J. Dechka, G. Hicks, and J. Whybra, "Transforming national forest resource information management with SAFORAH," Pacific Forestry Centre, Natural Resources Canada, and BC CITS, 2003.

[30] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, IT-14(1), 1968.

[31] D. L. Donoho, "High-dimensional data analysis: the curse and blessing of dimensionality," Lecture to the American Mathematical Society - Math challenges of the 21st century, 2000.

[32] Q. Jackson and D. A. Landgrebe, "An adaptive classifier design for high dimensional data analysis with a limited training data set," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, no. 4, pp. 518-528, 1991.

[33] B. Kim and D. A. Landgrebe, "Hierarchical classifier design in high dimensional, numerous class cases," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 12, pp. 2664-2679, 2001.

[34] D. A. Landgrebe, "Information extraction principal and methods for multispectral and hyperspectral image data," *Information Processing for Remote Sensing*, Scientific Publishing, 2000.

[35] C Lee,and D. A. Landgrebe, "Analyzing high-dimensional data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 4, pp. 792-800, 1993.

[36] D. W. Scott," Multivariate Density Estimation," Wiley, 1992.

[37] E. J. Wegman, "Hyperdimensional data analysis using parallel coordinates," *Journal of the American Statistical Association*, vol. 185, no. 411, pp. 664-675, 1990.

[38] A. A. Green, M. Berman, P. Switzer, M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, pp. 65-74, 1988.

[39] H. Kantz and T. Schreiber, "*Nonlinear Time Series Analysis*," Pub. Cambridge University Press, pp. 93, 1997.

[40] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2786 – 2803, 2006.

[41] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441 – 454, 2005.

[42] D. G. Goodin, J. Gao, and G. M. Henebry, "The effect of solar illumination angle and sensor view angle on observed patterns of spatial structure in Tallgrass Prairie," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 42, no. 1, pp. 154 – 165, 2004.

[43] S. R. Sandmeier, E. M. Middleton, D. W. Deering, and W. Qin, "The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization," *Journal of Geophysical Research*, vol. 104, no. D8, pp. 9547 – 9560, 1999.

[44] C. D. Mobley, *Light and Water: Radiative transfer in Natural Water,* San Diego, CA: Pub. Academic, 1994.

[45] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no.1, pp.44-57. 2002.

[46] I. F. Blake and J. B. Thomas, "The linear random process," *Proceedings of the IEEE*, vol.56, no. 10, pp. 1969 – 1703, 1968.

[47] J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," *Proceedings of 4th Annual JPL Airborne Geoscience Workshop*, vol. 1, Pasadena, CA, 1993.

[48] J. Campbell, "Spatial correlation effects upon accuracy of supervised classification of land cover," *Photogrammetric Engineering & Remote Sensing*, vol. 47, pp. 355-363, 1981.

[49] Y. Zhong, L. Zhang, B. Huang, and P. Li, "An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 420 – 431, 2006.

[50] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2282 – 2291, 2006.

[51] T. Han and D. G. Goodenough, "Nonlinear feature extraction of hyperspectral data based on locally linear embedding (LLE)," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1237 - 1240, Seoul, Korea. July 25 – 29, 2005.

[52] A. Mohan, G. Sapiro, and E. Bosch, "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 206–210, 2007.

[53] T. Han and D. G. Goodenough, "Investigation of nonlinearity in hyperspectral remotely sensed Imagery – a nonlinear time series analysis approach," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1556 - 1560, Barcelona, Spain, July 23 – 27, 2007.

[54] G. W. Felde, G. P. Anderson, T. W. Cooley, M. W. Matthew, S. M. adler-Golden, A. Berk, and J. Lee, "Analysis of hyperion data with the FLAASH atmospheric correction algorithm," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 90 - 92, Toulouse, France, July 21 - 25 2003.

[55] ITT Visual Information Solutions**,** IDL User's Guide 6.3.

[56] D. Kugiumtzis, "Test your surrogate data before you test for nonlinearity," *Physical Review E*, vol. 60, no.3, pp. 2808 – 2816, 1999.

[57] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D*, vol. 58, pp.77-94, 1992.

[58] M. Palus, "Testing for nonlinearity using redundancies: quantitative and qualitative aspects," *Physica D*, 80:186-205, 1995.

[59] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Physical Review Letters*, vol. 77, no. 4, pp. 635 – 638, 1996.

[60] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos*, vol. 9, no. 2, pp. 413- 435, 1999.

[61] D. Kugiumtzis, “Surrogate data test on time series", in "Modeling and Forecasting Financial Data, Techniques of Nonlinear Dynamics", edited by A. Soofi and L. Cao, Kluwer Academic Publishers, 2002, Chapter 12, pp 267 – 282.

[62] T. Schreiber and A. Schmitz, “Discrimination power of measures for nonlinearity in a time series,” *Physical Review E*, vol. 55, no. 5, pp. 5443 – 5447, 1997.

[63] J. B. Cromwell, W. C. Labys, and M. Terazza, “Univariate Tests for Time Series Models,” Number 07-099 in Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage, Thousand Oaks, CA, 1994.

[64] P. Mouroulis, R. O. Green, and D. W. Wilson, "Optical design of a coastal ocean imaging spectrometer," *Optics Express*, vol. 16, no. 12, pp. 9087-9096, 2008.

[65] L. Eklundh and A. Singh, “A comparative analysis of standardised and unstandardised principal components analysis in remote Sensing,” *International Journal of Remote Sensing*, vol. 14, no. 7, pp. 1359–1370, 1993.

[66] R. E. Roger, “Principal component transform with simple, automatic noise adjustment,” *International Journal of Remote Sensing*, vol. 17, no. 14, pp. 2719-2727, 1996.

[67] A. Ifarraguerri and C. Chang, “Unsupervised hyperspectral image analysis with projection pursuit,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 6, pp. 2529–2538, 2000.

[68] L. M. Bruce, C. H. Koger, and J. Li, “Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331–2338, 2002.

[69] L. O. Jimenez and D. A. Landgrebe, “Hyperspectral data analysis and supervised feature reduction via projection pursuit,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2653–2677, 1999.

[70] L. O. Jimenez-Rodriguez, E. Arzuaga-Cruz, and M. Velez-Reyes, “Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 2, pp. 469–483, 2007.

[71] J. P. Kerekes and D. A. Landgrebe, “An analytical model of Earth-observational remote sensing systems,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 1, pp. 125–133, 1991.

[72] T. Han and D. G. Goodenough, “Estimating dimensionality of hyperspectral data using false neighbour method,” *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, pp. 83-86, Boston, USA, July 6-11, 2008.

[73] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Physical Review A*, vol. 45, no. 6, pp. 3403-3411, 1992.

[74] L. Cao, “Practical method for determining the minimum embedding dimension of a scalar time series,” *Physica D*, vol. 110, pp. 43-50, 1997.

[75] S. P. Garcia and J. S. Almeida, “Nearest neighbour embedding with different time delays,” *Physical Review E*, vol. 71, pp. (037204)1-4, 2005.

[76] M. B. Kennel and H. D. I. Abarbanel, “False neighbors and false strands: A reliable minimum embedding dimension algorithm,” *Physical Review E*, vol. 66, pp. (026209)1-18, 2002.

[77] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, “Imaging spectroscopy and the airborne Visible/Infrared Imaging Spectrometer (AVIRIS),” *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227-248, 1998.

[78] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, “Hyperion, a space-based imaging spectrometer,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1160-1173, 2003.

[79] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000.

[80] I. T. Jolliffe, *Principle Component Analysis*, Springer-Verlag, 1986.

[81] R. A. Horn, and C. R. Johnson, *Matrix Analysis*, Cambridge: Cambridge University Press, 1990.

[82] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal components transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, pp. 295–304, 1990.

[83] R. E. Roger, “A faster way to compute the noise-adjusted principle components transform matrix,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 6, pp. 1194-1196, 1994.

[84] D. Coppersmith and S. Winograd, “Matrix multiplication via arithmetic progressions” *Journal of Symbolic Computation*, vol. 9, pp.251–280, 1990.

[85] B. Staab, “Investigation of noise and dimensionality reduction transforms on hyperspectral data as applied to target detection,” *Internal Report*, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 2005.

[86] K. S. Schmidt and A. K. Skidmore, “Smoothing vegetation spectra with wavelets,” *International Journal of Remote Sensing*, vol. 25, no. 6, pp. 1167 – 1184, 2004.

[87] P. Scheunders, “Wavelet thresholding of multivalued images,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 475 – 483, 2004.

[88] H. Othman and S. Qian, “Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 397 – 408, 2006.

[89] H. Kantz, T. Schreiber, and I. Hoffmann, “Nonlinear noise reduction: a case study on experimental data,” *Physical Review E*, vol. 48, no. 2, pp. 1529 – 1538, 1993.

[90] P. Grassberger, R. Hegger, H. Kantz, and C. Schaffrath, “On noise reduction methods for chaotic data,” *Chaos*, vol. 3, no. 2, pp. 127 – 141, 1993.

[91] T. Schreiber, “Extremely simple nonlinear noise-reduction method,” *Physical Review E*, vol. 47, no. 4, pp. 2401 – 2404, 1993.

[92] E. J. Kostelich and T. Schreiber, “Noise reduction in chaotic time-series data: a survey of common methods,” *Physical Review E*, vol. 48, no. 3, pp. 1752 – 1762, 1993.

[93] M. Davis, “Noise reduction schemes for chaotic time series,” *Physica D*, vol. 79, no. 2-4, pp. 174- 192, 1994.

[94] C-I. Chang, T-L. Sun, and M. Althouse, “Unsupervised interference rejection approach to target detection and classification for hyperspectral imagery,” *Optical Engineering*, vol. 37, no. 3, pp. 735 – 743, 1998.

[95] C-I. Chang and Q. Du, “Interference and noise-adjusted principal components analysis,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2387-2396, 1999.

[96] R. Cawley and G-H. Hsu, “Local-geometric method for noise reduction in chaotic maps and flows,” *Physical Review A*, vol. 46, no. 6, pp. 3057 – 3082, 1992.

[97] F. Takens, ``Detecting Strange Attractors in Turbulence,” *Lecture Notes in Math*. vol. 898, Springer, New York, 1981.

[98] T. Sauer, J. Yorke, and M. Casdagli, “Embedology,” *Journal of Statistical Physics*, vol. 65, vol. 3-4, pp. 579-616, 1991.

[99] R. E. Roger, “Principal Components transform with simple, automatic noise adjustment,” *International Journal of Remote Sensing*, vol. 17, no. 14, pp. 2719 – 2727, 1996.

[100] R. E. Roger and J. F. Arnold, “Reliably estimating the noise in AVIRIS hyperspectral images,” *International Journal of Remote Sensing*, vol. 17, no. 10, pp. 1951- 1962, 1996.

[101] P. Indyk, “Nearest neighbours in high-dimensional spaces,” *chapter of Handbook of Discrete and Computational Geometry*, CRC Press, 2004.

[102] ITT Visual Information Solutions, 2008, ENVI 4.5 User Manual.

[103] PCI Geomatics, 2008, Geomatica 10.1.3 User Manual.

[104] ERDAS, 2009, ERDAS IMAGINE 9.3.2 User Manual.

[105] ITT Visual Information Solutions, 2008, Interactive Data Language (IDL) 7.0 User Manual and Programmer’s Guide.