Using Random Matrix Theory to Determine the Intrinsic Dimension of a Hyperspectral Image

Kerry Cawse-Nicholson

Primary supervisor:
Michael Sears

Wits co-supervisor:
Amandine Robin

CSIR co-supervisors:
Konrad Wessels, Frans van den Bergh, Renaud Mathieu

May 18, 2012

Declaration

I declare that this thesis is my own, unaided work. It is being submitted for the Degree Doctor of Philosophy at the University of the Witwatersrand, Johannesburg, South Africa. It has not been submitted before for any degree or examination at any other University.

Kerry Cawse-Nicholson
Johannesburg, 2012

Abstract

Determining the intrinsic dimension of a hyperspectral image is an important step in the spectral unmixing process, since under- or over- estimation of this number may lead to incorrect unmixing for unsupervised methods. In this thesis we introduce a new method for determining the intrinsic dimension, using recent advances in Random Matrix Theory (RMT). This method is not sensitive to non-i.i.d. and correlated noise, and it is entirely unsupervised and free from any user-determined parameters. The new RMT method is mathematically derived, and robustness tests are run on synthetic data to determine how the results are affected by: image size; noise levels; noise variability; noise approximation; spectral characteristics of the endmembers, etc. Success rates are determined for many different synthetic images, and the method is compared to two principal state of the art methods, Noise Subspace Projection (NSP) and HySime. All three methods are then tested on twelve real hyperspectral images, including images acquired by satellite, airborne and land-based sensors. When images that were acquired by different sensors over the same spatial area are evaluated, RMT gives consistent results, showing the robustness of this method to sensor characterisics.

Acknowledgements

This research was funded by Meraka Institute, Council for Scientific and Industrial Research (CSIR) as part of their studentship program. Additional funding was received from the National Research Foundation (NRF), through the Professional Development Program (PDP).

This research formed part of the Centre for High Performance Computing Flagship Project on Computational Imaging and Remote sensing at the University of the Witwatersrand.

One of the datasets examined in this thesis was acquired by Carnegie Airborne Observatory (CAO) over the Kruger National Park and surrounding areas, South Africa. The airborne campaign and analysis was funded by the Andrew Mellon Foundation. The CAO is made possible by the W.M. Keck Foundation and William Hearst III.

I am grateful to Professor Steven Damelin of Georgia Southern University for introducing me to Random Matrix Theory, and for providing many thought provoking ideas throughout the writing of this thesis. Professor Damelin was also a co-supervisor of this thesis until the end of 2010, when he resigned from the University of the Witwatersrand.

I am indebted to my co-supervisors Drs Amandine Robin, Konrad Wessels, Frans van den Bergh and Renaud Mathieu. You have all guided me through this thesis in different ways, suggesting great ideas, and continuously pushing me to be better. I would like to express my deepest gratitude to all of you. I would especially like to express my heartfelt thanks to my primary supervisor Prof Michael Sears, who has been inspirational to me during my PhD. You have put so much effort into supervising this thesis and your input has been invaluable.

Finally I would like to thank my family. This thesis would not be possible without the unequivocal support of my husband, my parents and my brother. This thesis is dedicated to you.

Contents

List of Figures

List of Tables

Notation

Define the following standard variables:

p - an integer representing the number of spectral bands

N - an integer representing the number of pixels or samples per band

$\underline{\tilde{x}}$ - a random vector of $\mathbb{R}^p$, where each element $\tilde{x}^j$ corresponds to the spectral reflectance at the j^{th} wavelength

$\underline{x}_i$ - $\forall i\ 1 \leq i \leq N$, $\underline{x}_i \in \mathbb{R}^p$ is a realisation of $\underline{\tilde{x}}$, where each element x_i^j corresponds to the spectral reflectance at the j^{th} wavelength for the i^{th} pixel

σ^2 - a real number representing the noise variance

K - an integer representing the Intrinsic Dimension of the dataset

When assuming measurement $\underline{x}_i$ may be decomposed as $\underline{x}_i = \underline{s}_i + \underline{n}_i$:

$\underline{s}_i$ - a vector in $\mathbb{R}^p$ representing the pure signal component of $\underline{x}_i$

$\underline{n}_i$ - a vector in $\mathbb{R}^p$ representing the pure noise component of $\underline{x}_i$

$\underline{v}_j$ - a vector in $\mathbb{R}^p$ representing j^{th} endmember or pure substance, where the q^{th} element v_j^q corresponds to the spectral reflectance of the endmember at the q^{th} wavelength

$\tilde{u}_j$ - a random scalar representing the proportion of endmember $\underline{v}_j$ in $\underline{\tilde{x}}$.
This variable is subject to the following conditions: $\tilde{u}_j \geq 0$ for each j, and $\sum_{j=1}^{K} \tilde{u}_j = 1$

u_{ij} - a real valued scalar realisation of $\tilde{u}_j$ for the i^{th} measurement or pixel.
This variable is subject to the following conditions: $u_{ij} \geq 0$ for each i and j, and $\sum_{j=1}^{K} u_{ij} = 1$ for each i

1. Introduction

1.1. Background

The aim of this thesis is to introduce a new method for determining the intrinsic dimension (ID) of multi-dimensional data, specifically hyperspectral imagery, in order to determine the mixture of pure substances within the pixels of the image. Due to the pigeon-hole principal, the maximum number of targets that may be identified by their spectral signatures is limited by the number of bands [1]. The high dimension of hyperspectral images (typically more than 100 bands) therefore enables the identification and separation of many targets. Such targets are also considered on a sub-pixel level, where each pixel is a mixture of pure substances. These concepts are defined in more detail in the rest of the Chapter. It is important to understand the nature of a hyperspectral image and the definition of the intrinsic dimension, which are both discussed below.

1.1.1. Hyperspectral Imagery and Basic Definitions

Imaging spectroscopy is defined as the acquisition of an image from a remote sensing platform, where the spectral responses from each pixel are transformed into appropriate units. This image is made up of spatially co-registered images simultaneously acquired over spectrally contiguous bands [2]. Hyperspectral images are acquired by imaging spectroscopy sensors [2].

Due to the nature of acquisition, a hyperspectral image is a series of images acquired over hundreds of wavelengths. The resulting image is known as the three dimensional (3-D) data cube (Figure 1.1), where an x-y slice shows a grayscale image taken at a specific wavelength, and each pixel has an associated z-axis graph that shows the spectral properties of that pixel over the entire wavelength range [3]. The variables defined in the afore-mentioned notation may now be introduced: $\underline{x}_i$ is the vector in the direction of the z-axis, containing p spectral values for the i^{th} pixel; p is the number of spectral bands along the z-axis; and N is the number of pixels in the x and y directions.

A single pixel of a hyperspectral image is assumed to be made up of a number of base constituents, "pure" substances or targets, each having a unique spectral signature

(Figure 1.2) which may be used to identify these substances remotely [2]. While pure targets may be well understood in mineral applications, this idea becomes more subjective in data that contain vegetation spectra, which may vary according to species, phenology, etc. In this thesis it is assumed that each target may be represented by a single spectral signature, where variation around this target is understood to be smaller in magnitude than the noise present in the image. If the variation is larger than this, then certain vegetation classes may be represented by more than one spectral signature. These targets are known as endmembers, and all pixels in the image are made up of some combination of these endmembers [1].

The spectrum of each endmember is dependent on the chemical and physical or structural (e.g. surface roughness, leaf structure) properties of the targets. It is the incident energy (such as sunlight) causing both reactions in the energy levels of the electrons and various (multiple) scattering processes at the interface between ground and atmosphere that determines the amount of reflected light at each spectral frequency [4].

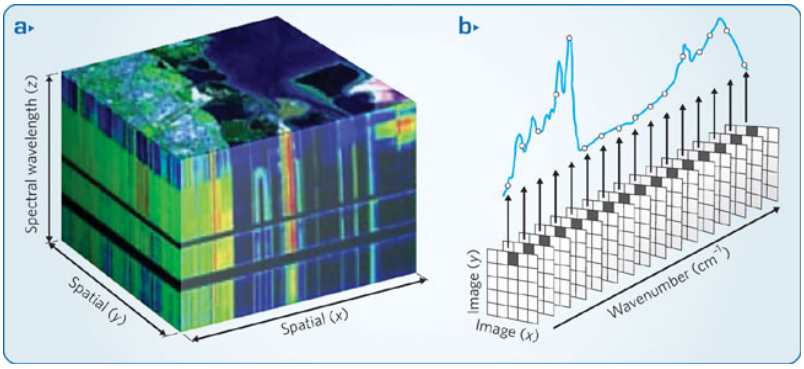


Figure 1.1.: A hyperspectral image is made up of many images acquired over hundreds of spectrally contiguous bands. Together, these images make up a hyperspectral image cube. (Taken from [3])

Depending on the spatial resolution of the sensor, each pixel may be a mixture of components or endmembers. In addition, geometry and sensor errors, the noise of the sensor, and the effects of the atmosphere all interfere with the surface reflectance signal, making it more difficult to identify surface components [1]. With all of these considerations in mind, there has been a large amount of research into separating an image into its key components [1, 5, 6]. Images may be unmixed by comparison with libraries of endmember spectra, or unsupervised algorithms can estimate these endmembers and partition the image into classes representing the abundance of each endmember [1].

The process of unmixing a hyperspectral image is illustrated in Figure 1.3. A good

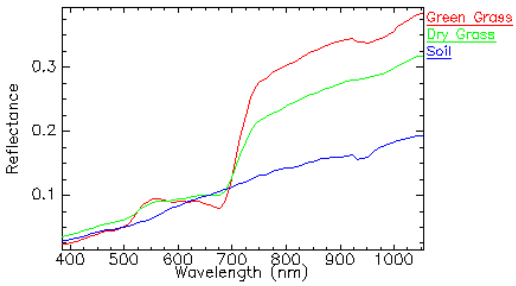


Figure 1.2.: Different substances have unique spectral signatures. These are endmembers collected from a hand-held spectrometer in the vicinity of the Kruger National Park.

summary of this process may also be found in [7]. The first step in unmixing the image, once atmospheric effects have been corrected (see Section 1.1.2), is to determine the number of base components or endmembers in the image. This number may be used to determine the endmember spectra, and these are in turn used to determine the endmember abundances in each pixel.

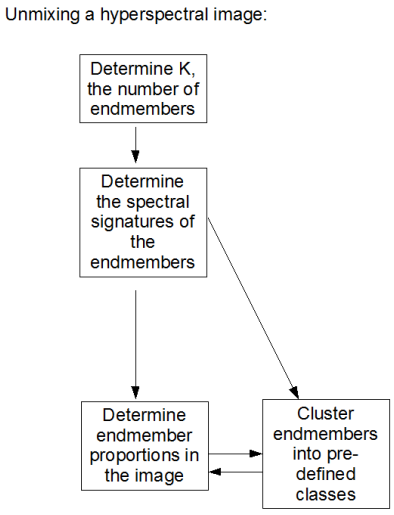


Figure 1.3.: The process of unmixing a hyperspectral image. Determining the number of endmembers in the image is the initial step in this process, and it determines the accuracy of the final unmixing result.

It is important to note the difference between the number of endmembers as opposed to the number of classes. Both entities may be the same if classes are defined in terms of the signal, but often classes are user-driven or application-specific. Endmembers are defined to be data-driven and independent of the application. For example, in a particular application, there may be four classes of interest — bare ground, green or photosynthetically active grass, dry or non-photosynthetically active grass, and tree cover. But, if the image is unmixed using a clustering algorithm with number of classes $K = 4$, it is likely that the image would not unmix into the above-mentioned classes. The classes could just as easily have been rearranged to group the different grass types together, separated the tree cover, or added shadow or water classes to still have $K = 4$. This is why it is important to first unmix the image correctly into the data-driven endmembers and then group these endmembers into the user-driven classes.

Further, the definition of the number of endmembers must be specified. The dataset involves significant redundancy in the spectral dimension, but the data will be of full dimension p (where p is the spectral dimension) because of the presence of random noise. If the data were presumed noise free, then they would be contained in a proper vector (signal) subspace of $\mathbb{R}^p$.

Bioucas-Dias and Nascimento [8] define the Intrinsic Dimension of the image (ID) as the dimension of the signal subspace. Chang and Du [9] define ID as the "minimum number of parameters required to account for the observed properties of the data". They also define a separate estimate, called *Virtual Dimension*, which is the number of endmembers necessary to give accurate unmixing, which may be larger than the number of so-called "idealized substances". Here substances may be understood as pure targets or endmembers, and could represent different objects depending on the application, e.g. chemical powders, cover types such as tree species, etc. By definition, VD may be dependent on the unmixing method that is used, unlike ID. Wu *et al.* have compared estimates of VD and ID in their survey paper [10] and found them to be comparable for the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) scene of Cuprite.

Bajorski [11], however, found that the concept of VD may be misleading in certain circumstances, since the value may change when the image is shifted and rotated. Bajorski [11] instead defines the Effective Dimensionality (ED) as "the dimensionality of the affine subspace giving an acceptable approximation to all pixels". Schlamm *et al.* [12] define an Inherent Dimension K, where "the entire spectral image can lie in the same K-dimensional hyperplane". These authors also claim that the Inherent Dimension is not equivalent to VD, or the number of spectral signatures present in the image. In contrast, the Spanning Dimension is defined as the "minimum number of basis vectors

required to span the space" of pixel observations [12]. Schlamm *et al.* define many other types of dimensions which may apply to hyperspectral imagery, to display the confusion around the term. These authors also define their own intrinsic dimension as "the smallest number of parameters needed to contain all of the variability in the data though a mapping function". This differs from Bioucas-Dias and Nascimento's definition.

The many different definitions of Intrinsic Dimension may be confusing, and although some of the above definitions are closely connected, it is core for this thesis to state a formal definition for ID that is independent of the method used to calculate it.

Let $\underline{x}_i \in \mathbb{R}^p$ be the observed measurement in each pixel $i, 1 \leq i \leq N$. Assuming the measurement may be linearly decomposed into signal and noise, write $\underline{x}_i = \underline{s}_i + \underline{n}_i$, where $\underline{s}_i$ represents the information in pixel i and $\underline{n}_i$ represents the noise. Then ID is defined as follows.

Definition 1 (Intrinsic dimension)**.** *The Intrinsic Dimension (ID) of the dataset is the dimension, K, of the vector subspace spanned by $\{\underline{s}_i, 1 \leq i \leq N\}$.*

Definition 1 is considered to be equivalent to the Bioucas-Dias and Nascimento definition [8], since $\{\underline{s}_1, \ldots, \underline{s}_N\}$ spans the signal subspace, and so K is therefore the dimension of the signal subspace. This is also equivalent to the spanning dimension defined in [12]. This definition is chosen so that the results in this thesis may be compared with HySime results, and also because this value may be used in the linear mixture model and in simplex methods for determining the endmember spectra (discussed in more detail below).

Determining the ID is a difficult problem in practice, and is still an ongoing area of research [1]. Many signal processing methods are applied to hyperspectral imagery, but these tend to make assumptions about the image that are not applicable in practice, in particular that the noise in the image is independent and identically distributed (i.i.d.) and uncorrelated between bands (these concepts will be discussed in the sections below) [10]. Other methods are sensitive to user-determined thresholds [9]. Supervised unmixing methods require prior knowledge about the image and these methods depend on the substances in the image being accurately represented in a spectral library. These methods do not always produce good results when spectra present in the library are not contained in the image [13], and may be computationally complex. In this thesis an unsupervised model will be developed, which does not assume i.i.d. and uncorrelated noise in the image.

Determining the intrinsic dimension is important for the processing of many different types of data, including chemical unmixing [14], extracting speech signals in a noisy line [15], unmixing minerals [16] and unmixing imagery of natural landscape [6], among

many others. Determination of this number is necessary for classification methods, unmixing methods and target detection [1]. It is common practice for the user to select the number of endmembers to suit the application, but this does not necessarily agree with the intrinsic dimension of the image, and an incorrect estimation of this number may have detrimental effects on unmixing results [10].

Although the intrinsic dimension is a concept that may be applied to many areas of research, this thesis will focus on applications to hyperspectral imagery. A hyperspectral image contains large amounts of information that can be of great benefit if used correctly. The techniques for analysing these images are largely target-driven and are developed according to specific applications. Hyperspectral imagery users are interested in which objects are contained within the image. For example, mining companies would want to know which minerals are present in a piece of rock [17], environmental scientists would want to know the vegetation cover versus bare soil to estimate how land has been degraded [18], the military would want to identify targets such as enemy vehicles [19], and so on.

The method developed in this thesis will be based exclusively on the intrinsic information in the image and will be independent of the particular application. It will not depend on any prior knowledge of the scene or user-determined thresholds, and the derivation does not restrict the method to hyperspectral imagery, although that will be the application investigated in this study. The method will be tested on twelve real images and performance will be evaluated on simulated data.

The ID of a hyperspectral image is calculated by considering the linear mixture model, which is described below. Non-linear mixing is a more accurate description of the physical interaction between scattered light and multiple substances or targets (some of which are 3-dimensional, e.g. trees and buildings), and the radiative transfer model describes this behaviour well [7]. The non-linear model does however require many parameters for the image, which are difficult to obtain, and therefore the linear mixture model is considered an acceptable approximation [7].

The ID, defined above, is calculated by assuming that the mixtures of pure substances — defined "endmembers" — are linear, and using this to reveal the base constituents and their abundances. This may be done by using a common model, namely the linear mixing model [20].

Linear Mixture Model In satellite remote sensing in particular, physical constraints of the sensor imply that spatial resolution is often low to enable high spectral resolution. The high number of spectral bands is necessary to identify targets, since a maximum of

p endmembers may be identified by an image with p spectral bands [1]. The resulting low spatial resolution means that each pixel is a mixture of endmembers. Even future satellite designs plan for 30 m spatial resolution (see Section 1.1.2), which means that many targets (e.g. trees, houses etc.) will only make up a fraction of a pixel. As a result, unmixing techniques that allow for sub-pixel endmembers play a very important part in earth observation remote sensing.

Mathematically, measurement vector $\underline{x}_i \in \mathbb{R}^p$ for each pixel i, where p is the number of spectral bands, may be written as:

$$\underline{x}_i = \sum_{j=1}^{K} u_{ij}\underline{v}_j + \underline{n}_i \tag{1.1.1}$$

where u_{ij} is a scalar value representing the proportion of endmember vector measurement $\underline{v}_j$ in the mixed pixel measurement $\underline{x}_i$, K is the number of endmembers that occur in the image, and $\underline{n}_i$ is a $(p \times 1)$ vector representing some noise function. The abundances u_{ij} must fulfill the positivity and sum-to-one conditions, so that $u_{ij} \geq 0$ for each $i, 1 \leq i \leq N$ and $j, 1 \leq j \leq K$, and $\sum_{j=1}^{K} u_{ij} = 1$ for each $i, 1 \leq i \leq N$.

It is important that the number of endmembers, K, in the linear mixing model is equivalent to the Intrinsic Dimension, $\tilde{K}$, as defined in Definition 1. If we assume linearly independent endmembers and noise-free pixels, then this is shown in the proposition below.

Proposition 1. *Suppose vector $\underline{x}_i \in \mathbb{R}^p$ may be written as $\underline{x}_i = V\underline{u}_i$, where V is a $(p \times K)$ matrix, containing linearly independent column vectors, $\underline{u}_i$ is a $(K \times 1)$ vector, $\underline{u}_i \geq \underline{0}$ and $\underline{1}^T\underline{u}_i = 1$, for $i, 1 \leq i \leq N$. Let $U_N = \frac{1}{N}\sum_{i=1}^{N} \underline{u}_i\underline{u}_i^T$ have full rank K. Suppose $\tilde{K}$ is the dimension of the vector subspace spanned by $\{\underline{x}_i, 1 \leq i \leq N\}$, where N is the number of pixels in the image.*

Then $\tilde{K} = K$.

Proof. Each $\underline{x}_i$ is a convex linear combination of the K columns of V, and the columns of V are linearly independent. The set $\{\underline{x}_1, \ldots, \underline{x}_N\}$ spans a vector space of dimension $\tilde{K}$. (The basis vectors of the subspace are not found in V.) Therefore $\tilde{K} \leq K$.

Since $U_N = \frac{1}{N}\sum_{i=1}^{N} \underline{u}_i \underline{u}_i^T$ has full rank K, it is invertible.

$$\begin{aligned}
\underline{x}_i &= V\underline{u}_i \\
\frac{1}{N}\sum_{i=1}^{N} \underline{x}_i \underline{u}_i^T &= \frac{1}{N}\sum_{i=1}^{N} V\underline{u}_i\underline{u}_i^T \\
&= VU_N \\
\frac{1}{N}\sum_{i=1}^{N} \underline{x}_i \underline{u}_i^T U_N^{-1} &= V \\
\text{Let } \underline{w}_i^T \text{ be the } (1 \times K) \text{ vector} \quad & \quad \underline{u}_i^T U_N^{-1} \\
\text{Thus } V &= \frac{1}{N}\sum_{i=1}^{N} \underline{x}_i \underline{w}_i^T \\
&= [\frac{1}{N}\sum_{i=1}^{N} w_{i1}\underline{x}_i, \ldots, \frac{1}{N}\sum_{i=1}^{N} w_{iK}\underline{x}_i]
\end{aligned}$$

Thus each column of V is a linear combination of the pixels $\underline{x}_i$. Therefore $\tilde{K} \geq K$.

Thus $\tilde{K} = K$. □

This proposition shows that $K = \tilde{K}$, and so the number of endmembers in the linear mixing model is equivalent to the Intrinsic Dimension as defined in Definition 1. However, even though $K = \tilde{K}$, the basis vectors of the spanning space should be distinguished from the endmember vectors, which must represent physical substances.

Note that it is necessary to assume that the endmembers are linearly independent and that $\frac{1}{N}\sum_{i=1}^{N} \underline{u}_i \underline{u}_i^T$ has full rank K. (These assumptions are used in practice by Kritchman and Nadler [14] in determining the intrinsic dimension of chemical mixtures.)

In order to better understand the assumption that $\frac{1}{N}\sum_{i=1}^{N} \underline{u}_i \underline{u}_i^T$ has full rank, consider an example where two vectors in V always occur in the same proportions, resulting in reduced rank of $\frac{1}{N}\sum_{i=1}^{N} \underline{u}_i \underline{u}_i^T$. In this case, however, we would expect that the value of K for the scene would be lower than the number of pure substances. Thus this assumption that $\frac{1}{N}\sum_{i=1}^{N} \underline{u}_i \underline{u}_i^T$ has rank K is reasonable.

The linear mixture model will be used to determine K, but it is important to understand the effects of noise and image acquisition on the data. These effects are discussed below.

1.1.2. Image Acquisition

Hyperspectral images are acquired using remote sensing techniques. The science of remote sensing is theoretically the act of observing an object from a distance, without making contact with it, to discover some properties of the object. In the case of interest, remote sensing is defined as the acquisition, processing and analysis of electromagnetic radiation [4] that is reflected off or emitted from the earth's surface [21] in order to determine some properties of the surface of the earth.

According to the Blackbody principle, all matter above a temperature of absolute zero continuously emits electromagnetic radiation [4]. In the case of thermal imaging, the sensor measures the radiation emitted by the surface of the earth, and in other passive remote sensing, sensors measure the target reflection of the highest electromagnetic radiation producer in the earth's vicinity — the sun [21]. In active remote sensing, such as LiDAR and radar, an artificial energy produced by the sensor itself is reflected off the earth's surface [21]. All these areas of remote sensing have wide areas of application, and this study will focus on the reflection of radiation from the sun off the surface of the earth. This domain is also called optical remote sensing.

Different portions of the electromagnetic spectrum may be used, including ultraviolet light, visible light and infrared light, seen in Figure 1.4. The range that humans can see makes up only a small portion of the measurable spectrum. Solar radiation may be measured in the ultraviolet to infrared ranges, but the ultraviolet region is not used because the atmosphere absorbs most of the incoming and reflected signal in that range [21]. The remaining bands are referred to as VNIR (visible, 0.4–0.7 μm, to near-infrared, 0.7–1.1 μm) and SWIR (short wave infrared, 1.1–2.5 μm). The images in this thesis cover VNIR, SWIR, and both ranges combined. Thermal emittance is also commonly used, covering the electromagnetic range 8 – 15 μm.

Image spectroscopy began with the sensor positioned near to the object being analysed. Similar technology is currently used in hand held devices used to establish ground reflectance which is used as ground truth, the true reflectance as measured on the ground. These sensors were adapted for airborne spectroscopy and finally for space-borne imagers. The first satellite remote sensing sensor, Landsat, was launched in 1972, with 4 VNIR bands [22] and research into satellite-based remote sensing has grown significantly since then. The purpose of such studies has been to evaluate objects on the surface of the earth, and if data are acquired at enough spectral wavelengths, identification of these substances may be possible [21]. This identification is possible because of the unique intensity of light that each object reflects at certain wavelengths, which is dependent on the composition and structure of the object [4]. This unique reflectance is known as the

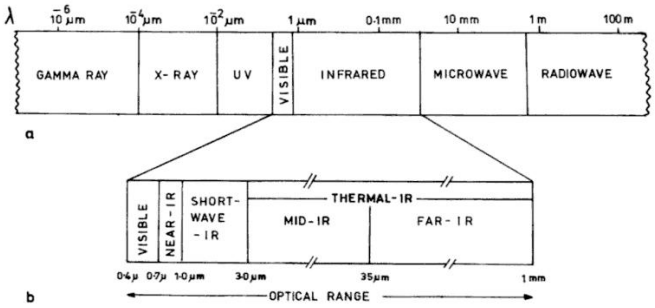


Figure 1.4.: The electromagnetic spectrum, taken from [4]. The optical range is covered by hyperspectral scanners, most commonly visible to SWIR, and the visible range makes up a very small portion of the spectrum.

spectral response curve, or spectral signature. Figure 1.2 shows the spectral signatures of different land cover classes, where each spectrum is shown to differ from the others.

According to the Conservation of Energy law, all the energy that comes into contact with an object must be scattered, absorbed and/or transmitted. Remote sensing focuses on the energy that is transmitted from the sun to the ground, and reflected back to the sensor. The reflection may follow Snell's law (angle of incidence equals angle of reflectance) or the reflection may be off a Lambertian surface (energy is reflected equally in all directions). In practice, the reflection is somewhere between the two extremes, with most of the energy reflecting back in one direction [4]. The proportion of light that is reflected in all directions is known as *albedo*. Before the reflected light reaches the sensor, it must pass through the atmosphere. The effects of this path are examined in more detail below.

The atmosphere All but one of the datasets considered in this thesis are obtained by air- or space-borne sensors. These collect the reflectance of the sun's radiance off the surface of the earth. An important consideration for this type of remote sensing is that the sun's energy is passing through the atmosphere once as it is emitted towards the earth's surface, and again as it is reflected back to the sensor. The atmosphere absorbs or scatters a certain amount of energy, and these effects must be considered before an image is processed. Some applications, such as blind source separation, are able to estimate atmosphere and endmembers, but other methods require that the atmospheric effects are corrected for before the algorithm is implemented. In particular, oxygen, carbon dioxide, water vapour and ozone absorb energy in certain wavelengths, converting the

energy into heat [21] (see Figure 1.5), so that it does not reach the ground to reflect back to the sensor. The regions of the spectrum that are not affected by these absorptions are known as "atmospheric windows" [4]. The light that is not absorbed may be scattered or transmitted. There are two dominant scattering mechanisms, namely Rayleigh scatter and Mie scatter, that cause large radiometric distortions. Rayleigh scatter is due to the actual molecules of air and occurs when the particles causing the scatter are smaller than the wavelength of the radiation. Since shorter wavelengths are scattered more than longer wavelengths, this explains why the sky appears blue (blue is a shorter wavelength than red) [21]. Mie scatter can result in an apparent "haze" over images [4]. Mie scatter is also known as aerosol scatter and is due to larger particles such as dust and smoke. These particles are larger than the wavelength of the radiation, and so this effect is more prominent in the longer wavelengths.

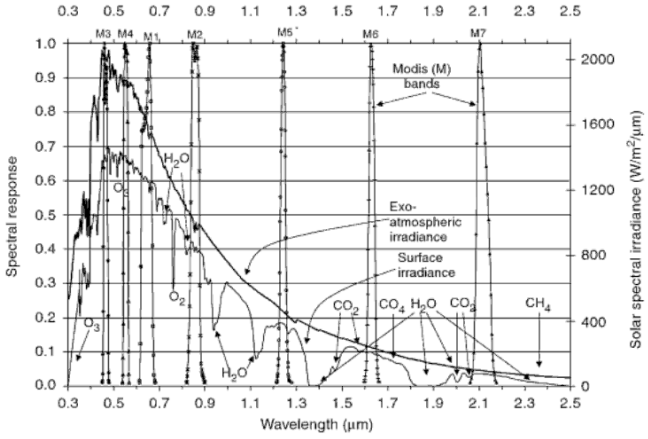


Figure 1.5.: The solar irradiance received by the top of the atmosphere is dependent on the wavelenth. Note the effect of the absorptions, and the effect of scatterings on the surface irradiance. (Image from [23].)

The energy transmitted by the sun reaches the atmosphere with a power density called *irradiance* (which describes the strength of any body emitting electromagnetic energy) and may be measured in Watts per square metre. *Radiance* is the irradiance at a given angle [24]. The absorptions and scatter described above mean that not all energy emitted by the sun actually penetrates the atmosphere to reach the surface of the earth. The percentage that does is known as *transmittance*, which is dependent on the angle of the sun, the atmospheric conditions and the wavelength of the light or energy. According to Sahu in [24], as little as 31% of the sun's energy reaches the earth as a direct beam. *Path radiance* is the amount of energy which reflects off the earth

to reach the sensor, which is the energy measured by the sensor. This measurement is also affected by energy scattered within the instantaneous field of view (IFOV) of the sensor by pixels surrounding the pixel under measurement (environmental effect) or by the atmosphere. (See a simplified diagram of these effects in Figure 1.6.)

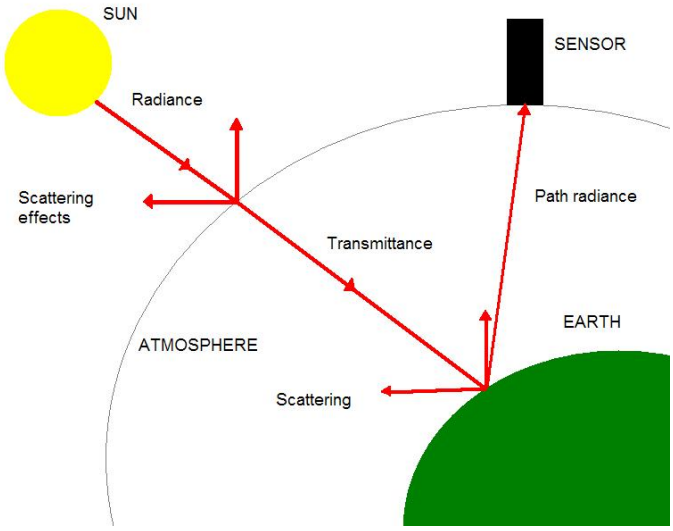


Figure 1.6.: Energy transmitted by the sun reaches the atmosphere, where some energy is scattered and some energy is absorbed. The absorbed energy is known as transmittance, and this reaches the earth, where (again) the energy is scattered, reflected or absorbed. The reflected energy that reaches the sensor is known as path radiance.

Radiance and Reflection Optical remote sensing sensors capture the amount of light reflected off particles on the earth's surface, at different wavelengths. However, this reflection is impacted by the atmosphere in both directions, from the sun to the ground and from the ground to the sensor, where the measurement received is known as path radiance, which must be corrected in order to be meaningful. Ideally, information about the atmosphere, such as the content of water vapour and the density of certain aerosols is necessary to model the atmosphere accurately before reflectance values may be derived using radiative transfer models [24] in software such as ATCOR (Atmospheric/Topographic Correction) [25, 26]. There are versions of ATCOR for satellite and airborne sensors, and the method uses the Modtran® 5 radiative transfer code. However, the detailed atmospheric inputs required by radiative transfer models are often unknown, and then the atmospheric effects may be approximately removed using algorithms based on simpli-

fied atmospheric parameters (e.g. visibility) or generic models (e.g. continental versus maritime atmospheric models). See for instance ATCOR or FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) [27] which is built into the ENVI[1] framework.

If the atmospheric effects are not known, some errors such as the "haze" effect caused by Rayleigh scattering may be corrected in other ways. This is done by assuming that at least one pixel in each band should be zero or close to zero. The minimum pixel value from each band is subtracted from all values in the band, correcting the haze and increasing the dynamic range of the image [21].

The atmospheric correction is an important consideration when evaluating the image, as an accurate correction will improve the quality of the signal in the end product.

Sensor types Solar reflectance from the surface of the earth may be collected in different ways. For all sensor types, *across track* is the direction perpendicular to the flight path, *along track* is the direction parallel to the flight path, *field of view (FOV)* is the angular coverage in the across track direction, and the *swath width* is the corresponding coverage on the ground [22] (see Figure 1.7). In a line scanner, the data are collected one pixel at a time, resulting in a vector output. Whiskbroom scanners contain an array of detectors (approximately 30) with a mirror that scans perpendicularly to the path of the sensor, allowing rows of pixels to be captured with each scan of the mirror. Pushbroom scanners contain a panel of detectors to simultaneously collect a block of pixel values (typically $> 1{,}000$) at each position along the sensor's flight line. A prism is used to disperse the light into different wavelengths, and whiskbroom and pushbroom sensors may consist of several detectors for different wavelength ranges.

Sensor characteristics include *spectral resolution* (the width of the spectral bands), *spatial resolution* (the size of the pixel on the ground, linked to IFOV, see Figure 1.7) and *temporal resolution* (for satellites, this is the time between scans of the same geographical area). These characteristics may not be as exact as they seem, for instance, while each pixel is considered to represent a square on the ground, the *point spread function (PSF)* describes the area that reflects back to a single point on the sensor, as shown in Figure 1.8. This often means that there is overlap between adjacent pixels, resulting in image pixels that contain components from neighbouring pixels [22], which may cause spatial correlations. The same is true for spectral correlations, which are significant in hyperspectral data.

[1] http://www.RSInc.com/envi

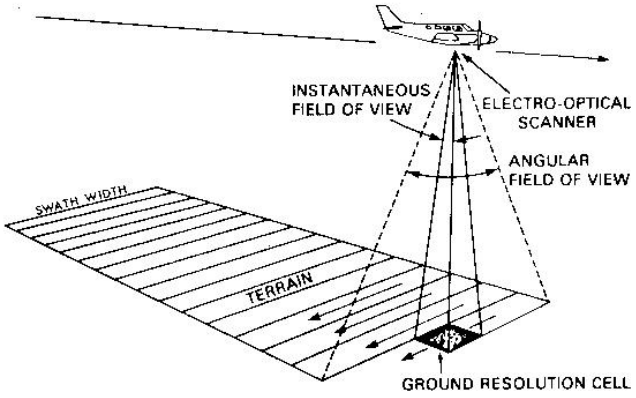


Figure 1.7.: The relationship is displayed between swath width, the across track distance covered by the sensor, and angular field of view (FOV), the angular range of the sensor. Taken from [28].

Geometry and Sensor Errors and Influences The reflectance received by the remote sensing platform may not represent the exact position and brightness of the surface reflections at each pixel, due to atmospheric influences or some instability of the remote sensing platform, etc. Radiometric errors can occur in a single pixel or an image band. Geometric errors affect the entire image and are usually due to the motion or position of the instrument, such as altitude (height above the ground), velocity, pitch (upward/downward angle), roll (horizontal angle) and yaw (direction of flight) [21]. These motions are illustrated in Figure 1.9 (Taken from NASA[2]). Errors are also caused by the curve of the earth or by panoramic distortion (pixels further from the sensor appear larger than those directly beneath it), especially if the sensor field of view is wide.

Pushbroom sensors may have within band striping effects if the individual sensors are not properly calibrated. This may be improved by assuming that the mean and standard deviations of neighbouring bands should be similar, and correcting accordingly (gain and offset method) [21]. Alternatively, Acito *et al.* [29] have recently proposed an effective destriping technique especially derived for hyperspectral images.

Other errors include electronic "cross talk" that occurs when the signal received by one array element is transmitted to another element, mixing signals for certain spectral bands, and "spectral smile" where there is spectral variation across the sensor's field of view [30].

[2]www.src.nasa.gov/WWW/K-12/airplane/rotations

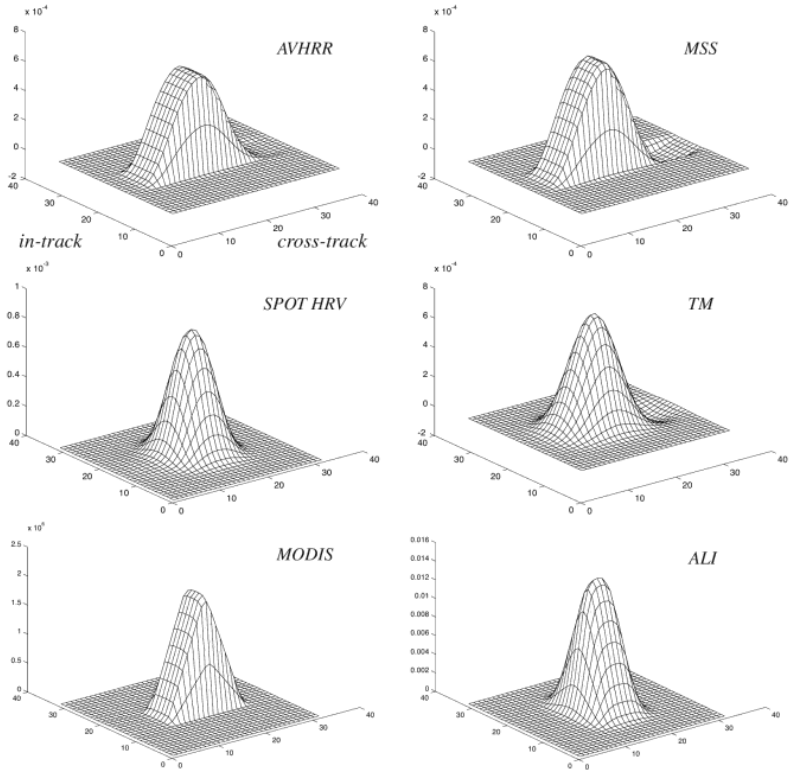


Figure 1.8.: The point spread function is shown for several different sensors. Note that none of the sensors image a square, like the pixels seen in the image, and some point spread functions are in fact asymmetrical, especially the whiskbroom sensors (AVHRR, MSS and TM), caused by the low pass electronic filter. (Image from [22].)

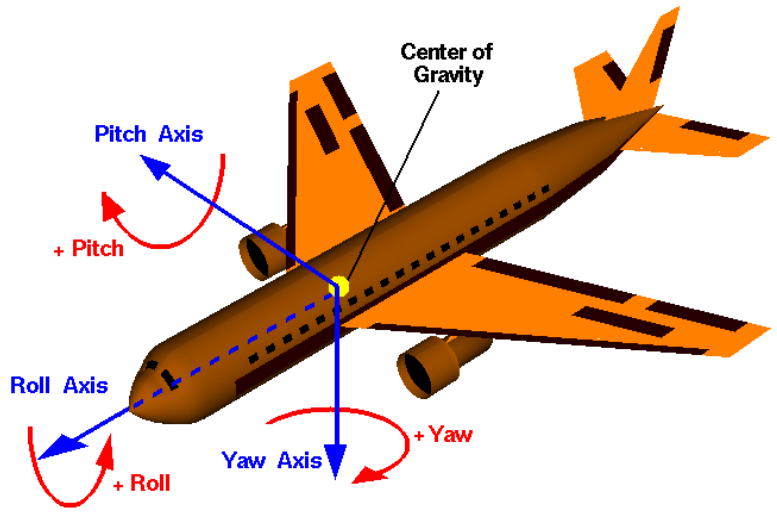


Figure 1.9.: This image illustrates the effect of roll (horizontal angle), pitch (upward/downward angle) and yaw (direction of flight) on the sensor. (Image from NASA.)

The future of satellite remote sensing Hyperion was NASA's first satellite based hyperspectral sensor, and is mounted on the Earth Observing Mission 1 (EO1) satellite that was launched in 2000 as part of a one-year project [31]. The sensor is still providing data, long past its planned mission dates. The spectral range covers 0.4 – 2.5 μm with 220 spectral bands, 30 m spatial resolution and 7.5 km swath width.

Europe and USA are both currently designing hyperspectral satellites that will be launched in the near future. EnMAP (Environmental Mapping and Analysis Program) is being managed by the German Aerospace Center (DLR) and is due to be launched in 2013-2014. The spectral range will cover 0.42 – 2.45 μm with 232 spectral bands, the swath width will be 30 km, with spatial resolution of 30 m, and a possible revisit time of 4 days [32, 33]. EnMAP has a budget of nearly a hundred million Euros [34].

HyspIRI (Hyperspectral Infrared Imager) is a pushbroom scanner, managed by NASA and due to be launched in 2015. The hyperspectral sensor will cover the optical range 0.4 – 2.5 μm and there is a separate multispectral thermal infrared scanner that will cover the range 8 – 12 μm. The hyperspectral scanner will have 210 bands, spatial resolution of 45 m and a swath width of 90 km; the thermal scanner will have 5 bands and a spatial resolution of 90 m. HyspIRI will cost about $300 million [35].

The South African sensor MSMI (Multi-Sensor Microsatellite Imager) has been designed at a much lower budget, and is currently awaiting funding for integration into

a satellite bus. MSMI will have a 14.9 km swath width at 660 km altitude. It will cover the spectral range 0.4 – 2.35 μm over 200 bands, with 14.5 m spatial resolution [36]. Such large investments in hyperspectral instruments show the great demand for the data produced by such sensors. As a result, there will be a demand for methods to process the resulting images.

Although satellite technology is growing fast, hyperspectral sensors are still mostly mounted on airborne platforms, such as AVIRIS, HyMap, SpecTIR, and many other privately owned sensors. The advantages of airborne platforms include easy maintenance of the sensor, high spatial resolution (depending on the height of the aircraft), and less atmosphere to influence the surface reflectance.

AVIRIS was designed to identify constituents of the earth's surface and atmosphere, with a focus on the global environment and climate change [37]. SpecTIR has application in geology, vegetation, water and emergency response [38]. HyMap has marine and geological applications [39]. Other sensors are available for specialised applications according to the clients' needs.

In order for these applications to be achieved, the object of interest must be identified. For example AVIRIS scenes could be unmixed into base minerals [40], SpecTIR images are used to detect the conditions of asphalt roads [41], and HyMap data may be used to classify plant species [42]. All of these applications may benefit from accurate unsupervised unmixing methods, which require the intrinsic dimension of the dataset.

Most methods for determining the intrinsic dimension require a noise estimate of the image [10][14], and the characteristics of this noise are discussed below.

1.1.3. Noise

It is assumed that the atmospheric and geometric errors discussed above have already been sufficiently corrected in the datasets used in this thesis. The image may also be affected by dark current noise, which results from small electric currents that flow through the sensor even when the sensor is not receiving any light. This noise may be modelled by analysing a so-called black image or target (i.e. non-reflecting or non-emitting target). In the datasets used it is also assumed that dark current noise has been corrected.

The noise considered in this study is due to low absolute reflectance values after atmospheric absorptions and small errors in the preprocessing of the image. This noise follows random behaviour and is usually considered Gaussian, for example see [10] [14].

Signal to noise ratio (SNR) SNR is defined as the ratio $s : n$, where s represents some measurement of the signal and n represents some measurement of the noise. This is an important tool for evaluating the quality of an image, especially where certain algorithms are only applicable if the SNR is "high enough". This evaluation may be used to identify certain bands that have high noise due to atmospheric or sensor effects, so that they may be removed. The presence of a single bad band can cause problems for many algorithms, and it is standard practice to remove these as a preprocessing step [10].

The quantities s and n can be written in several different ways. SNR may be measured in decibels (dB) by using the following definition in [8]

$$\mathrm{SNR}_{\mathrm{dB}} = 10 \log_{10} \frac{E[\underline{\tilde{s}}^T \underline{\tilde{s}}]}{E[\underline{\tilde{n}}^T \underline{\tilde{n}}]}, \tag{1.1.2}$$

where $\underline{\tilde{s}}$ is the $(p \times 1)$ signal present in each pixel and $\underline{\tilde{n}}$ is the $(p \times 1)$ noise vector. This approximation requires knowledge or approximation of the noise.

Another SNR definition in [1] is given as the ratio of the mean signal to the standard deviation (std) of the noise (here the noise should be assumed i.i.d. in order for the mean and standard deviation to give meaningful results).

$$\mathrm{SNR}_{\mathrm{var}} = \frac{\mathrm{mean}(\underline{s(i)})}{\mathrm{std}(\underline{n(i)})}. \tag{1.1.3}$$

Often the standard deviation of the noise is approximated by taking the standard deviation of a small neighbourhood of pixels that are assumed to be homogeneous.

The issue is further confused by the definition in [10], where

$$\mathrm{SNR}_{\mathrm{var2}} = \frac{50\% \text{ of the mean observed reflectance value}}{\text{standard deviation of the noise}}. \tag{1.1.4}$$

All of these definitions describe different measurements, and so for the sake of consistency, the last definition will be used when referring to SNR, unless specified otherwise. This is chosen since it is not dependent on a noise estimate and several state of the art methods, including HySime and HFC are compared in [10] using this measurement. However, using the standard deviation of the noise implies that the noise should be i.i.d., and using reflectance means that the results may be affected by atmospheric corrections, which the user should bear in mind.

1.2. Objectives

In remote sensing, classification is important for image interpretation, and even when sensors have high spatial resolution, pixels may contain a mixture of endmembers [1]. Thus unmixing these pixels into their base constituents is necessary for almost all images. This is an active area of research, and often supervised unmixing is used, where endmembers are selected from large libraries or a-priori knowledge. It has been shown that incorrect choice of endmembers will adversely affect the unmixing results [13], and so unsupervised unmixing techniques are considered in methods such as N-FINDR, SGA and VCA, amongst others. The first step in unsupervised unmixing is the choice of the intrinsic dimension, and this has a large impact on the overall unmixing [10]. Section 2 will show that this is an active area of research.

This thesis contains the development of a new unsupervised method, using Random Matrix Theory (RMT), to determine the intrinsic dimension of a hyperspectral image. Random Matrix Theory is chosen as it was shown to be robust in chemical applications in [14], and does not require user-determined thresholds. In this thesis the RMT method in [14] is adapted to be applicable to hyperspectral imagery, and this will be tested on several real and synthetic datasets and compared to other accepted methods. The performance of the method will also be analysed with respect to several variables such as image size, number of spectral bands, noise levels, etc. This is done by setting a range of values for these variables in synthetic data, and evaluating the accuracy of the method over the range.

The method will be further tested by comparing the resulting signal subspace with that of other well-known methods, and the image will be unmixed to assess the homogeneity of the endmember abundances. The real images considered will include the well-studied AVIRIS scene of Cuprite, along with the same scene imaged by another airborne sensor (SpecTIR) and a satellite sensor (Hyperion). Other images from these three sensors are studied, as well as an image from the Hyperspectral Core Imager (HCI) and images acquired from the Carnegie Airborne Observatory (CAO) instrument. Images are considered with and without atmospheric effects, and mineral and vegetation scenes are compared. This wide collection of data will allow reliable analysis of the method in many different scenarios.

2. Literature review

This chapter describes some of the principal methods found in the literature to address the problem of unmixing a hyperspectral image. This process may be separated into three steps: ID estimation (Section 2.1); endmember approximation (2.3); and abundance unmixing (Section 2.4). Some methods are derived for other fields (such as signal processing or chemical applications) and have been applied to hyperspectral imagery as an extension of the method. Other methods have been designed specifically for hyperspectral imagery. Many of these methods require an estimation of the noise, and noise approximation methods will be discussed in more detail in Chapter 4.

2.1. Determining the Intrinsic Dimension (ID) of a hyperspectral image

An important part of unmixing the image is determining K, the number of endmembers, or intrinsic dimension of the image. Some existing methods for determining the ID of a hyperspectral image will be examined. Not all of these methods calculate a value that determines the ID as defined in Definition 1, but all estimates are considered to be comparable [10] [43].

Following the procedure in [14], define the model for the i^{th} pixel, $1 \leq i \leq N$, so that the observed hyperspectral measurement vector $\underline{x}_i \in \mathbb{R}^p$ described in (1.1.1) is a realisation of a random vector $\underline{\tilde{x}}$, where

$$\underline{\tilde{x}} = V\underline{\tilde{u}} + \sigma\underline{\tilde{\xi}}, \tag{2.1.1}$$

where $\underline{\tilde{u}}$ is a $(K \times 1)$ random vector representing the proportions of pure components, subject to the constraints $\underline{\tilde{u}} \geq \underline{0}$ and $\underline{1}^T\underline{\tilde{u}} = 1$; V is a $(p \times K)$ deterministic matrix, with columns corresponding to the spectral measurements characterizing the pure components; $\underline{\tilde{\xi}}$ is a random vector in $\mathbb{R}^p$ representing the noise; σ is a scalar weighting factor; and K is the total number of pure components.

It is assumed that $\underline{\tilde{\xi}}$ follows a Gaussian distribution $\mathcal{N}(0, I_p)$. This assumption has

been successfully used in chemical unmixing [14] and in many methods described in the survey paper by Wu *et al.* [10]. The scalar σ must be estimated from the data, and in the models it is assumed that the spectra of the pure components must remain constant over the entire spatial scene in order to estimate the global ID.

Most of the methods to determine the ID use the $(p \times p)$ observation covariance matrix, $S(N)$, defined as

$$S(N) = \frac{1}{N} \sum_{i=1}^{N} (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^T \tag{2.1.2}$$

where N is the number of pixels in the image, p is the number of spectral bands and $\bar{\underline{x}}$ is the mean pixel value. In particular, these methods use the eigenvalues, λ_j, of $S(N)$, sorted in descending order.

2.1.1. Information criteria methods

Two early methods (1974 and 1978 respectively) are Akaike Information Criterion [44] (AIC) and Minimum Description Length [45] (MDL); both are used in array signal processing as follows:

$$\begin{aligned}
\mathrm{AIC}(i) &= -2\log \left[\frac{\left(\prod_{j=i+1}^{p} \lambda_j^{1/(p-i)} \right) \cdot (p-i)}{\sum_{j=i+1}^{p} \lambda_j} \right]^{(p-i)N} \\
&\quad +2i(2p-i) \\
\mathrm{MDL}(i) &= -\log \left[\frac{\left(\prod_{j=i+1}^{p} \lambda_j^{1/(p-i)} \right) \cdot (p-i)}{\sum_{j=i+1}^{p} \lambda_j} \right]^{(p-i)N} \\
&\quad +0.5i(2p-i)\log N \\
\mathrm{VD}_{\mathrm{AIC}} &= \underset{i}{\operatorname{argmin}}(\mathrm{AIC}(i)) \qquad (2.1.3) \\
\mathrm{VD}_{\mathrm{MDL}} &= \underset{i}{\operatorname{argmin}}(\mathrm{MDL}(i)) \qquad (2.1.4)
\end{aligned}$$

where p is the number of bands, N is the number of pixels, and $\{\lambda_j\}$ are the ordered eigenvalues of the observation covariance matrix (2.1.2). In each of these methods, the noise is assumed to be independent and identically distributed (i.i.d.), uncorrelated and Gaussian.

Akaike [44] uses the definition: AIC = -2log(maximum likelihood) + 2(number of independently adjusted parameters within the model); to create an alternative to hy-

pothesis testing — model selection. Schwartz [46] adjusted this to choose the appropriate dimension for a model, given certain observations and using Bayes estimators. This definition is: dimension = -log(maximum likelihood) + 1/2 log(number of observations)×(dimension of model). Rissanen [45] determined the number of parameters in a time series, by using linear models and Gaussian distributions, with reference to the above methods. Wax and Kailith [47] determined that Schwartz and Rissanen's theories reduced to the same limit in large samples, and this is defined as MDL. Wax and Kailith then used AIC and MDL to successfully count the number of signals in a multichannel time series, using the linear mixture model described above. It is therefore natural to test these methods, which are simple and easy to implement, on hyperspectral data.

To summarise, the information criteria methods assume i.i.d., uncorrelated and Gaussian noise. These assumptions may not be applicable to hyperspectral images, but are considered popular methods for determining the intrinsic dimension [9]. These methods require only the eigenvalues of the observation covariance matrix, are entirely unsupervised, and are easy and fast to implement.

2.1.2. Malinowski's method

Malinowski [48] (1977) derived functions to determine the IE (Imbedded Error) and IND (Indicator function) in order to evaluate if a matrix is factor analyzable. His application is in the field of Chemistry, but Chang and Du [9] apply this to hyperspectral images with some success. Chang and Du [9] define the Imbedded Error and Indicator function as follows:

$$\begin{aligned} \mathrm{RE}(q) &= \frac{(\sum_{j=q+1}^{p} \lambda_j)^{\frac{1}{2}}}{N^{\frac{1}{2}}(p-q)^{\frac{1}{2}}} \\ \mathrm{IND}(q) &= \frac{\mathrm{RE}(q)}{(p-q)^2}. \end{aligned}$$

Chang and Du [9] further use these to define Malinowski's Empirical Indicator Function as:

$$\mathrm{EIF}(q) = \frac{(\sum_{j=q+1}^{p} \lambda_j)^{\frac{1}{2}}}{N^{\frac{1}{2}}(p-q)^{\frac{3}{2}}} \tag{2.1.5}$$

$$\mathrm{VD}_{\mathrm{EIF}} = \underset{q}{\operatorname{argmin}}\{\mathrm{EIF}(q)\} \tag{2.1.6}$$

where N is the number of pixels, p the number of bands, and $\{\lambda_q\}$ are the ordered (descending) eigenvalues of the observation covariance matrix. This method assumes i.i.d., uncorrelated Gaussian noise.

Malinowski [48] states that the origin of the above formula is unknown.

Although the noise assumptions are not necessarily applicable to hyperspectral imagery, this method has been used to determine the ID of a hyperspectral image in [9]. This method requires only the eigenvalues of the observation covariance matrix, is entirely unsupervised, and is easy and fast to implement.

2.1.3. Gerschgorin radius based method

This method assumes that the noise is i.i.d. and uncorrelated but not necessarily Gaussian. Given a $p \times p$ matrix $A = [a_{ij}]$, for each row, define the Gerschgorin radius as follows:

$$\rho_i = \sum_{j=1, j\neq i}^{p} |a_{ij}|. \tag{2.1.7}$$

Then the Gerschgorin disk is defined by a set $D_i(A)$ of complex numbers, where

$$D_i(A) = \{z \in \mathbb{C} \mid \rho_i \geq |z - a_{ii}|\}. \tag{2.1.8}$$

The Gerschgorin circle theorem states that all the eigenvalues of A lie in the union of all the Gerschgorin disks of A. Further, if a disk is disjoint from all other disks, then it contains only one eigenvalue. This method manipulates the given observation covariance matrix so that all disks are distinct. The manipulation is as follows: If S is the observation covariance matrix, then

$$S_{p\times p} = \begin{pmatrix} s_{1,1} & \cdots & s_{1,p} \\ \vdots & & \vdots \\ s_{p,1} & \cdots & s_{p,p} \end{pmatrix}$$

$$\text{Define } S_{(p-1)\times(p-1)} = \begin{pmatrix} s_{1,1} & \cdots & s_{1,(p-1)} \\ \vdots & & \vdots \\ s_{(p-1),1} & \cdots & s_{(p-1),(p-1)} \end{pmatrix}$$

Let $V = \{v_1, \cdots, v_{(p-1)}\}$ be the matrix of eigenvectors of $S_{(p-1)\times(p-1)}$, so that $\Lambda = V^T S_{(p-1)\times(p-1)} V$, where Λ contains the eigenvalues $\{\lambda_j\}$ of $S_{(p-1)\times(p-1)}$.

Define $\tilde{V} = \begin{pmatrix} V & 0 \\ 0 & 1 \end{pmatrix}$ and $\tilde{S}$ by

$$\begin{aligned} \tilde{S} &= \tilde{V}^T S \tilde{V} \\ &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & \rho'_1 \\ 0 & \lambda_2 & & 0 & \rho'_2 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & \lambda_{p-1} & \rho'_{p-1} \\ \rho'_1 & \rho'_2 & \cdots & \rho'_{p-1} & \rho'_{pp}. \end{pmatrix} \end{aligned}$$

Since those rows associated with noise should have a Gerschgorin radius of zero, this implies that $\rho'_j = 0 \; \forall j > K$, where K is the number of endmembers. If $\rho'_j > 0$, then that row is associated with signal. $\rho'_j = |v_j^H C|$, where C is $[s_{1p}, \cdots, s_{(p-1)p}]^T$.

Computationally, the noise rows do not have a Gerschgorin radius of exactly zero. There are two interpretations on how to proceed. Wu *et al.* repeats the procedure above for the centered observation covariance matrix as well, and where the transformed Gerschgorin radii ρ' are the same (within some user-defined ϵ) for both centered and non-centered observation covariance matrices, then those rows are said to represent noise. The method described in [49] relies on MDL and AIC, described above.

To summarise, the Gerschgorin radius based method assumes i.i.d. and uncorrelated noise that is not necessarily Gaussian. These assumptions may not be applicable to hyperspectral imagery, but this method has been used to determine the ID of a hyperspectral image in [10]. This method is unsupervised and requires the calculation of eigenvectors and eigenvalues and matrix multiplication, making it much more computationally complex than the information criteria methods and Malinowski's method.

2.1.4. Bayesian methods

When using the linear mixture model, most methods assume that the endmembers are deterministic, even though they may be unknown. A Normal Compositional Model (NCM) has been introduced in [50], which assumes that each endmember is an independent Gaussian vector with known mean. These means may be calculated using an endmember extraction algorithm. Eches *et al.* [51] use this model to determine the number of endmembers in a hyperspectral image. The noise in the dataset may be explained by the variable endmembers, and so the noise term in the LMM vanishes. The variance,

ϵ_i^2, of the endmembers is fixed for each band and endmember per pixel, although more complex models may be developed [51]. Then the linear mixture model becomes

$$\underline{x}_i = \sum_{j=1}^{K} \tilde{u}_{ij} \underline{\tilde{e}}_{ij}, \tag{2.1.9}$$

where $\underline{\tilde{e}}_{ij} \in \mathcal{N}(\underline{v}_j, \epsilon_i^2 I_p)$ for each pixel $i, 1 \leq i \leq N$. This algorithm is semi-supervised, since an endmember library is assumed known. The image means belong to this library, but the number of endmembers and exact spectra are considered unknown. The abundances $\tilde{u}_{ij}$ and the variances of the endmembers must also be estimated.

Algorithm 1 Semi-supervised unmixing using NCM [51]

$K(0) \leftarrow$ ID chosen from a uniform distribution [1,K_{max}]
$M(0) \leftarrow K(0)$ spectra chosen from the endmember library
$\underline{u}(0) \leftarrow$ initial abundances, with positivity and sum-to-one constraints
$\epsilon^2(0) \leftarrow$ initial value, calculated using reversible jump MCMC (Markov Chain Monte Carlo)
for $t = 1 \rightarrow T$ **do**
  μ is chosen from a uniform distribution [0,1]
  if $\mu \leq b(t-1)$ **then**
    Propose a birth move: add an endmember to M and normalise the new $\underline{u}$
  else
    if $b(t-1) < \mu \leq (b(t-1) + d(t-1))$ **then**
      Propose a death move: remove an endmember from M and normalise $\underline{u}$
    end if
  else
    if $\mu > b(t-1) + d(t-1)$ **then**
      Propose a switch move: a single endmember in M is replaced by one from the library
    end if
  end if
  μ_2 is chosen from a uniform distribution [0,1]
  if $\mu_2 < \rho$ **then**
    (where ρ is the acceptance probability)
    set$(\underline{u}(t), M(t), K(t)) \leftarrow (\underline{u}^*, M^*, K^*)$
  else
    set$(\underline{u}(t), M(t), K(t)) \leftarrow (\underline{u}(t-1), M(t-1), K(t-1))$
  end if
  sample $\underline{u}(t), M(t), K(t)$ from probability density distributions
end for

As seen in Algorithm 1, the algorithm works by constructing a matrix of K spectra,

M, which is chosen to represent the endmembers in the image. The endmember matrix M initially contains only one spectrum, and the matrix is grown or changed according to random moves. A birth move is when a new spectrum is chosen from the endmember library and added to M; a death move is when a spectrum is removed from M; and a switch move is when a spectrum from M is replaced by a library spectrum. The spectra are selected randomly, but the probabilities of a birth move (b), death move (d) and switch move (s) must sum to one. In most cases the three are considered equally probable, except when $K = 1$ (no death move allowed) and when $K = K_{max}$ (no birth move allowed). The maximum K_{max} is the size of the input library. This algorithm is run for each pixel in the image, and the global K is determined by evaluating the size of the union of M over all pixels.

After every update of the endmember matrix M, the abundances for each endmember are updated. The positive abundances are stochastically chosen and must sum to one. This in turn influences the variance ϵ_i^2 of the endmembers in the i^{th} pixel.

Eches *et al.* announce that their method performs better than the well-known HySime method, especially at high noise levels. This method may be useful especially in vegetation scenes, where there is inherent variation in a single endmember.

To summarise, this method assumes that all noise in the data may be explained by variation in the endmember spectra. These spectra have i.i.d. variance and are chosen from a spectral library, making the method semi-supervised.

2.1.5. MOCA

Maximum Orthogonal-Complements Algorithm (MOCA) was introduced in 2007 by Kuybeda, Malah and Barzohar [52]. This method aims to preserve rare vectors or anomalies in image reduction. To do so, the signal subspace is determined, and one of the key steps in the method is to estimate the size of this subspace, which is equivalent to the defined Intrinsic Dimension.

The signal subspace E may be separated into the $(K - h)$ dimensional subspace representing abundant endmembers, E_A, and the h dimensional subspace E_R representing rare endmembers. The abundant subspace E_A is made up of the $(K - h)$ largest eigenvectors of $S(N)$, and the rare subspace is made up of certain pixels in the image. This results in a noisy estimate of the rare subspace. Kuybeda *et al.* found that the traditional l_2 norm does not accurately retain rare vectors, and so the l_2^∞ norm is used, where

$$|||[\underline{z}_1, \ldots, \underline{z}_N]|||_{2,\infty} = \max_{1 \leq i \leq N} ||\underline{z}_i||_2.$$

The signal subspace is then determined by

$$E = \underset{\mathcal{L}}{\operatorname{argmin}} ||P_{\mathcal{L}}^{\perp} X||_{2,\infty}^2 \tag{2.1.10}$$

where $X = [\underline{x}_1, \ldots, \underline{x}_N]$ is the $(p \times N)$ matrix of observed pixel values, $P_{\mathcal{L}}^{\perp} X = X - P_{\mathcal{L}} X$ is the matrix of residuals, and $P_{\mathcal{L}}$ is the projection matrix onto an estimate of the signal subspace, $\mathcal{L} \subset \mathbb{R}^p$, where p is the number of spectral bands.

Every value for the size of the abundant signal subspace is tested, beginning with $K - h = 1$. The rare signal subspace is made up of pixels that are selected from the image, which maximise the l_2 norm of the residuals. In [52] all possible values for K are tested, and the computational complexity of this is improved in [53], where the Iterative Rare Vector Estimation (IRVE) algorithm is used. In the latter there is a stopping criterion on the l_2^{∞} norm of residuals, which is dependent on the size of the image and the number of bands.

Acito *et al.* [43] further develop this method by combining it with NWHFC (discussed below), which resulted in higher estimates of signal subspace size than NWHFC, when rare substances were present. MOCA was shown in [52] to out-perform AIC and MDL. This method assumes i.i.d., Gaussian noise, but may be adapted for non-i.i.d. and correlated noise [52]. Note that the use of the infinity norm may not be appropriate in images with dropped pixels.

2.1.6. Harsanyi-Farrand-Chang (HFC)

This method, developed by Chang and Du [9] in 2004, is based on the premise that the eigenvalues of the centered and non-centered covariance matrices will become the same when the eigenvalues represent only noise (which has a mean of zero). This means that the two sets of eigenvalues will converge at the point where signal ends and noise begins. However, computationally, the difference between the two sets is never exactly zero, and

so a false-alarm probability P_F is derived in order to find the threshold.

$$P_F = \int_{\tau}^{\infty} p_0(z)\mathrm{d}z \tag{2.1.11}$$

$$P_D = \int_{\tau}^{\infty} p_1(z)\mathrm{d}z \tag{2.1.12}$$

where τ is the detection threshold, and the probabilities arise from a set of hypothesis tests

$$H_0 : z_l = \hat{\lambda}_l - \lambda_l = 0 \text{ in the case of noise}$$
$$H_1 : z_l = \hat{\lambda}_l - \lambda_l > 0 \text{ in the case of signal,}$$

where $\hat{\lambda}_l$ is the l^{th} largest eigenvalue of the observation correlation matrix, and λ_l is the l^{th} largest eigenvalue of the observation covariance matrix. The probability $p_0(z) = P\{z_l|H_0\}$, where $z_l|H_0$ follows a Gaussian distribution $\mathcal{N}(0, \sigma^2_{z_l})$, is known as the false alarm probability and is fixed (user-determined) to determine τ. This results in the maximization of the detection probability P_D, where $p_1(z) = P\{z_l|H_1\}$. When the number of samples, N, is sufficiently large, then $\sigma^2_{z_l} \approx \frac{2}{N}\left(\hat{\lambda}^2_l + \lambda^2_l\right)$. Note that for each band, the two eigenvalues are compared with a different τ, due to the possibility of the noise characteristics differing in each band.

$$\begin{aligned}\int_{\tau_l}^{\infty} p_0(z)\mathrm{d}z &= \int_{\tau_l}^{\infty} \mathcal{N}(0, \sigma^2_{z_l})\mathrm{d}z_l \\ &= \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{\tau_l}{\sqrt{2\sigma^2_{z_l}}}\right)\right)\end{aligned}$$

Once τ_l has been determined for each band, the two sets of eigenvalues are said to converge if they differ by less than this threshold, and their convergence determines K.

This method performed the best of all algorithms analysed in Wu *et al.*, with its only downfall being its dependence on a user-determined threshold, P_F. Wu [10] has also used noise whitening to improve the accuracy of this method, using Roger's [54] residual based estimation in order to approximate the noise covariance matrix, described in Section 4.3.1. This is known as Noise Whitened HFC (NWHFC).

This method has also evolved into a Noise Subspace Projection (NSP) algorithm [9]. The HFC and NWHFC rely on enough samples for their approximations of the normal

distributions. If this is not the case, the whitened observation covariance may be used without comparison, since the noise eigenvalues will become identity when whitened. So the comparison becomes

$$H_0 : y_l = \lambda_l = 1 \text{ in the case of noise}$$
$$H_1 : y_l = \lambda_l > 1 \text{ in the case of signal.}$$

The failure probability P_F is defined by $\int_\tau^\infty p_0(z)\mathrm{d}z$ where $p_0(z) = p\{y_l|H_0\}$, $y_l|H_0$ follows a Gaussian distribution $\mathcal{N}(1, \sigma_{y_l}^2)$, and $\sigma_{y_l}^2 \approx \frac{2}{N}\lambda_l^2$. Now eigenvalues are considered noise if they are close to unity, within the threshold τ_l for each band l.

To summarise, HFC assumes Gaussian, zero-centered noise, and requires only the eigenvalues of two matrices. NWHFC and HSP also require the calculation of a whitening matrix. All three methods are dependent on a user-determined threshold.

One of the drawbacks of the above methods, is that the correct VD may depend on the endmember extraction method used. This is where the definition of VD differs from the definition of ID. In OSP (Orthogonal Subspace Projection) [55], a particular endmember extraction technique is used to estimate the endmembers, which are used as the projection matrices. This method was improved in [56], where MOSP (Maximum OSP) was introduced, by combining elements from MOCA with NWHFC. The VD derived from these were comparable with NWHFC and HySime.

2.1.7. HySime

Hyperspectral signal identification by minimum error (HySime) was developed by Bioucas-Dias and Nascimento [8] in 2005. As a preprocessing step, multiple regression theory (Section 4.3.2) is used to estimate the noise per pixel. This noise is then subtracted from every pixel, and the remaining values are used to create the estimated signal correlation matrix. The intrinsic dimension of the image is then calculated as follows,

$$K_{\text{HySime}} = \underset{1\leq i\leq p}{\operatorname{argmin}}\{\underline{r}^T P'(i)\underline{r} + 2\text{trace}[P(i)C_n/N]\} \tag{2.1.13}$$

where $\underline{r}$ is the mean $(p \times 1)$ pixel vector over the image, $P(i)$ is the projection of the pixel values onto the subspace spanned by $\{e_1, \ldots, e_i\}$, where $\{e_j\}$ are the eigenvectors ordered (descending) by the singular values generated by the SVD of the estimated signal correlation matrix, $P'(i)$ is the projection of the pixel values onto the subspace spanned by $\{e_{i+1}, \ldots, e_p\}$, and C_n is the noise covariance matrix, determined by the multiple

regression theory method in Section 4.3.2. An alternative HySime implementation is given in [57] (Section 2.2.3), where the signal subspace may be calculated as well.

This method has the advantage of not requiring any user-determined thresholds, and it produces reliable results on synthetic and real data [8]. However, HySime requires the calculation of eigenvalues and eigenvectors, matrix multiplication, and an estimate of the noise per pixel, making this one of the most computationally complex methods discussed. In addition, it assumes Gaussian noise which can be restrictive for some types of images, although this is a common assumption [10] [14].

2.1.8. Random Matrix Theory (RMT)

In the area of chemical unmixing, Kritchman and Nadler [14] (2008) worked with new results in Random Matrix Theory (RMT) to determine which eigenvalues are due to noise and which are due to signal. The advantage to this method is that there are no parameters that need to be set by the user, and it achieved good results in chemical testing. The false alarm probability may be varied by the method described in [58], but it may also be fixed as in [14]. In this study the false alarm probability is fixed for all images and is not considered a user-determined variable.

Johnstone [59] provides an excellent summary of the background to RMT. Some basics in the field of Random Matrix theory are covered before describing the method.

Wishart distributions Wishart matrices form the core of RMT. Johnstone [60] describes a random cross-product matrix $\tilde{A} = \tilde{X}\tilde{X}^T$, where $\tilde{X}$ contains N independent column vectors, each following a p-variate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. Then $\tilde{A}$ has a p-variate Wishart distribution with N degrees of freedom, $W_p(N, \Sigma)$. Johnstone [60] derives results for the case where the mean of each column is zero and the standard deviation of the p-variate Gaussian distribution is the identity matrix, *i.e.* each column of X follows the normal distribution $\mathcal{N}(0, I_p)$, and then $\tilde{A} \sim W_p(N, I_p)$.

Now suppose a hyperspectral image of pure noise is considered, *i.e.* the measurement in each pixel is a realisation of the random vector $\underline{\tilde{x}}_i$ following a Gaussian distribution $\mathcal{N}(0, \sigma^2 I_p)$. Define $\tilde{X} = [\underline{\tilde{x}}_1, \ldots, \underline{\tilde{x}}_N]$, then as stated above, $\tilde{A} = \tilde{X}\tilde{X}^T$ follows a Wishart distribution $W_p(N, \sigma^2 I_p)$. Denoting $\tilde{S}(N) = \frac{1}{N}\sum_{i=1}^{N} \underline{\tilde{x}}_i \underline{\tilde{x}}_i^T$, which may also be written as $\tilde{S}(N) = \frac{1}{N}\tilde{X}\tilde{X}^T$, then $\frac{N}{\sigma^2}\tilde{S}(N)$ follows a Wishart distribution $W_p(N, I_p)$. Hence, $S(N)$ defined in (2.1.2) may be seen as a realisation of $\tilde{S}(N)$ in the case of a pure noise image.

Random matrices Random matrices were first used in physics in the 1950s, to determine quantum energy levels [60]. In this setting, the analysis was only possible when the

energy levels were low. The existing statistical methods were only applicable for $N \to \infty$ and p fixed, and so new models were developed to analyse high-energy systems [59]. The case where N and p are both large has many applications in modern statistical theory [59], therefore in RMT it is assumed that $N \to \infty$, $p \to \infty$, with $\frac{p}{N} \to c$, where $c > 0$ is constant. A large body of research in RMT has been dedicated to the distribution of the largest eigenvalue of a matrix following a Wishart distribution [14, 60].

SemiCircle Law Wigner proposed that the local behaviour of high energy levels may be approximated by the distribution of eigenvalues of a random matrix [59]. The statistical analysis led to the development of the famous SemiCircle Law, where the density of the eigenvalues of a matrix containing i.i.d. entries with zero mean and variance σ^2, converges to a limit given by a semicircle [59]:

$$dF_N(x\sigma\sqrt{N}) \to \frac{1}{4\pi}\sqrt{4 - x^2}dx.$$

Tracy-Widom Limits If $A \sim W_p(N, I_p)$, and λ_1 is the largest eigenvalue of A, then the distribution of this eigenvalue tends to a distribution described by Tracy and Widom.

$$Pr\{\lambda \leq \sigma^2(\mu_{N,p} + s(\alpha)\sigma_{N,p})\} \to F_\beta(s), \tag{2.1.14}$$

where $\beta = 1$ corresponds to real valued data [14], α is a significance level and $s(\alpha)$ may be found by inverting the Tracy-Widom distribution.

$$F_1(s)^2 = \exp\left(-\int_s^\infty [(x-s)^2 - 1]q(x)dx\right)$$

and q is the solution to the second order Painlevé differential equation

$$q'' = sq + 2q^3.$$

The value for q may be approximated by the Airy function as $s \to \infty$ [59].

The centering and scaling factors were initially set to

$$\mu_{N,p} = \frac{1}{N}\left(\sqrt{N}+\sqrt{p}\right)^2 \tag{2.1.15}$$

$$\sigma_{N,p} = \frac{1}{N}\left(\sqrt{N}+\sqrt{p}\right)\times \left(\frac{1}{\sqrt{N}}+\frac{1}{\sqrt{p}}\right)^{1/3} \tag{2.1.16}$$

in a random growth model by Johansson [59] for a convergence $O(p^{-1/3})$, and this may be improved to convergence $O(p^{-2/3})$ for

$$\mu_{N,p} = \frac{1}{N}\left(\sqrt{N-\frac{1}{2}}+\sqrt{p-\frac{1}{2}}\right)^2 \tag{2.1.17}$$

$$\sigma_{N,p} = \frac{1}{N}\left(\sqrt{N-\frac{1}{2}}+\sqrt{p-\frac{1}{2}}\right)\times \left(\frac{1}{\sqrt{N-\frac{1}{2}}}+\frac{1}{\sqrt{p-\frac{1}{2}}}\right)^{1/3} \tag{2.1.18}$$

Phase transition There is a limit to the size of the eigenvalue that may be successfully detected by RMT. The phase transition phenomenon described in [61] results in the limit λ_{crit}, below which the noise eigenvalue will not be successfully identified. The limit is defined as

$$\lambda_{\text{crit}} = \sigma^2\sqrt{\frac{p}{N}}. \tag{2.1.19}$$

In hyperspectral images, N is typically much larger than p, and in this study it is assumed that the eigenvalues tested are above this limit.

RMT method Kritchman and Nadler [14] assume that in an image with both signal and noise, the largest noise eigenvalue will behave like the largest eigenvalue of a pure noise image.

They used some of Johnstone's results [60] to determine the ID in a chemical application. To determine the ID, the eigenvalues of the observation covariance matrix $S(N)$ are evaluated and one of these eigenvalues is expected to behave like the largest eigenvalue from a Wishart distribution. This eigenvalue is determined to be the largest eigenvalue resulting from noise. By the Tracy-Widom laws above, the largest eigenvalue

of a real-valued Wishart matrix, λ, fulfills the following condition with probability one as $N, p \to \infty$, $p/N \to c > 0$:

$$\lambda \leq \sigma^2(\mu_{N,p} + s(\alpha)\sigma_{N,p}), \tag{2.1.20}$$

where σ^2 is the variance of the Gaussian noise, α is a significance level and $s(\alpha)$ may be found by inverting the Tracy-Widom distribution (in [14], $\alpha = 0.5\%$, and this value is fixed for all images investigated). The scaling and centering values, $\mu_{N,p}$ and $\sigma_{N,p}$ are given by 2.1.17 and 2.1.18.

Note that these functions do not depend on the intrinsic dimension, K. It is also important that $\sigma_{N,p} \to 0$ as $p, N \to \infty$, $p/N = c$ fixed. This means that, especially for large images, the formula is not sensitive to the choice of α.

Traditionally, statistical techniques consider the scenario p fixed, with $p \ll N$, and in this case, the observed noise eigenvalues should be very close to σ^2 [62]. Although the number of pixels in a hyperspectral image are several orders of magnitude larger than the number of spectral bands, the dividing line between signal and noise eigenvalues is in practice still unclear. El Karoui [62] found that high dimensional problems (a hyperspectral image may contain hundreds of spectral bands) might be better solved by assuming that p and N are both large, with their ratio fixed. Similarly, Kritchman and Nadler [14] state that (2.1.20) still holds in the case of finite but large N and p, and in fact their algorithm deals specifically with small N, with good results. Another paper, also by Kritchman and Nadler [58] showed good results for $N \gg p$ when compared with AIC and MDL.

Then, in a hyperspectral image, if the eigenvalues $\{\lambda_j\}_{j=1}^p$ of the observation covariance matrix are arranged in descending order, so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, K is defined as the largest value so that $\lambda_j > \sigma^2(\mu_{N,p} + s(\alpha)\sigma_{N,p})$ for all $j \in \mathbb{Z}, 1 \leq j \leq K$.

When using (1.1.1), the separation between noise and signal eigenvalues is possible because the eigenvalues of S, where $S(N) \to S$ in probability for $N \to \infty$, are given by

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \cdots & \\ & \lambda_K & \\ 0 & & 0 \end{pmatrix} + \sigma^2 I_p$$

This definition uses the assumptions in [14] that the columns $\underline{v}_j$, $1 \leq j \leq K$ of the endmember matrix V are linearly independent, that the abundance covariance matrix $\sum_{i=1}^N \underline{u}_i \underline{u}_i^T$ has full rank, that the noise is uncorrelated with the signal, and that the

samples $\underline{x}_i$ are i.i.d.. Here it is assumed that $\underline{x}_i = V\underline{u}_i + \underline{n}_i$, where $\underline{x}_i$ is the observed i^{th} pixel vector, V is the matrix of endmembers, $\underline{u}_i$ is the vector of endmember proportions, bound by sum-to-one and positivity constraints, and $\underline{n}_i$ is the noise present in the pixel.

To summarise, this method assumes Gaussian, i.i.d. and uncorrelated noise. It requires only the eigenvalues of the observation covariance and a reliable estimation of the noise. This method is not dependent on any user determined values.

A new method based on these results will be the main contribution of this thesis. To validate this method, it will be compared with several methods described above, and the number, K, obtained will be used to unmix the image. In this method noise is assumed to be Gaussian, which is a common assumption in hyperspectral imagery [10] [14].

The first step in unmixing the image is to reduce the image dimension. Some popular methods will be used for this reduction, and these are described below.

2.2. Dimension reduction methods

Once the intrinsic dimension of the data has been calculated, the data may be transformed in such a way that information is not lost by neglecting certain bands. The image may be stored at this lower dimension (the ID) to reduce redundancy, reduce storage space, and enable the detection of the endmember spectra, discussed in Section 2.3. A necessary pre-processing step to all the methods discussed is that the dimension be reduced to the intrinsic dimension, K. Several methods for dimension reduction are discussed below, including simplex methods and a projection method.

2.2.1. Principal Components Analysis (PCA)

PCA as a statistical technique dates back to 1901 [59], and is a linear transform that orders the data in the directions of decreasing variance. If the image data are arranged so that the pixels $\underline{x}_i \in \mathbb{R}^p$ are represented in the $(p \times N)$ matrix $X = [\underline{x}_1, \ldots, \underline{x}_N]$, and $\bar{\underline{x}}$ is the mean pixel value, then $S_c(N) = \frac{1}{N}\sum_{i=1}^{N}(\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})^T$ is the $(p \times p)$ centered observation covariance matrix. Consider the principal component eigenvalues of $S_c(N)$, λ_i (in descending order) and their respective eigenvectors $\underline{z}_i$. The principal components of X are these eigenvectors, ordered by the eigenvalues, and they may be used to transform each pixel $\underline{x}_i$ as follows

$$\hat{X} = Z^T X \tag{2.2.1}$$

where $Z = [\underline{z}_1, \ldots, \underline{z}_p]$.

In hyperspectral data there is a certain amount of redundancy, and data may often be represented in a lower dimensional space without much loss of data [1]. This may be done by reducing the size of Z to remove those eigenvectors associated with eigenvalues below a certain threshold. ID will determine the size, K, of the signal subspace, so that $\hat{X}$ will become $(K \times N)$.

2.2.2. Minimum Noise Fraction (MNF)

This method is similar to PCA, but a noise covariance matrix is estimated and used to decorrelate and rescale the noise in a preprocessing step. Green *et al.* [63] observed that PCA does not always produce images ordered by quality, and so to improve this, MNF was introduced in [63]. This method does not maximise variance (as does PCA) but rather maximises the signal to noise ratio (SNR), essentially separating the observed vector space into a signal subspace and a noise subspace.

The image is again represented as a $(p \times N)$ matrix X, and the image is transformed as follows:

$$\hat{X} = H^T X \tag{2.2.2}$$

where $H = AD$; $A = B\Lambda^{-1/2}$ where Λ contains the (descending) eigenvalues of the noise covariance matrix, and B contains the respective eigenvectors. If the observation covariance matrix $S(N)$ is whitened by A to become S_W, then D contains the eigenvectors of S_W (ordered with respect to decreasing eigenvalues). Now the new dataset contains bands that are ordered according to their noise fractions. This method is considered comparable to noise-adjusted PCA, and if the noise is i.i.d. across bands, the two methods are equivalent [57].

As in PCA, the data dimension may be reduced by removing vectors in H that correspond to eigenvalues smaller than a certain threshold.

2.2.3. HySime

Hyperspectral signal identification by minimum error (HySime) was adapted for subspace identification by Bioucas-Dias and Nascimento [57] in 2008. As in (Section 2.1.7), the noise is calculated using a multiple regression technique, and removed from the observed data on a per pixel basis. This leads to the approximation of the signal covariance matrix, S_s, and the noise covariance matrix S_n. As in HySime $\underline{e}_i$ are the eigenvectors of S_s. This method is derived in [57], where Bioucas-Dias and Nascimento define

$$\delta_i = -\underline{e}_i^T S(N)\underline{e}_i + 2\underline{e}_i^T S_n \underline{e}_i \tag{2.2.3}$$

Arrange δ_i in ascending order, saving the permutation. Then K is the number of terms for which $\delta_i < 0$, and the signal subspace is the collection of eigenvectors $\underline{e}_i$ for which $\delta_i < 0$.

This method is derived by minimising the mean square error between the approximate signal value and the observed data value that has been projected onto a subspace created from some perturbation of K eigenvectors of S_s.

Once the image dimension has been reduced, several methods may be considered to determine the spectra of the endmembers. The most commonly used methods are described below in more detail.

2.3. Endmember extraction methods

Theoretically, a user could use all known endmember spectra in the linear mixing model described in (2.1.1), and the abundances of those endmembers not present would be zero. However, this does not work well in practice [13]. Certain sparse unmixing techniques may be used in an optimization problem to overcome this difficultly [64], but these methods are dependent on the endmembers being accurately represented in a known library.

It is therefore necessary to find the endmember spectra using other methods, and these may then be identified by comparing to a library of known spectra. A commonly used algorithm is Pixel Purity Index (PPI), introduced by Boardman *et al.* [65], which is built into the ENVI software. Because of the method's propriety, the algorithm is not publicly available, but several similar algorithms are discussed in [66]. This method uses randomly generated "skewers" on which to project the data. The extreme points of these projections are given a score, and the pixels with the highest scores are determined to be endmembers. Other methods have been shown to perform better than PPI, and three such methods are discussed. Two methods use simplex theory to determine the endmembers, or vertices of the K-dimensional simplex, and one method uses projection. All the methods assume the presence of pure pixels in the data.

2.3.1. N-FINDR

This algorithm was introduced by Winter in 1999 [67]. It assumes that each endmember is represented by at least one pure pixel, and attempts to find these pure pixels by maximizing the volume of the simplex formed with the pixels as the vertices. The volume is defined as $V(E)$, where

$$E = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \underline{e}_1 & \underline{e}_2 & \cdots & \underline{e}_K \end{pmatrix} \tag{2.3.1}$$

$$V(E) = \frac{1}{(K-1)!}|\det(E)|, \tag{2.3.2}$$

where $\{\underline{e}_i\}$ are the endmembers found in the data, and $(K-1)$ is the dimension of the data.

First, K must be calculated. Several methods for doing this have been discussed. Second, the data are reduced to dimension $(K-1)$.

The algorithm is initialised by choosing random pixels as the endmembers, and then calculating the volume. Then, each endmember is consecutively replaced by every pixel

in the image, with the combination that maximises the volume being stored for the next iteration. This process is repeated until a maximum number of iterations is reached, or until the volume is no longer increased. The details of the algorithm are shown in Algorithm 2.

Algorithm 2 N-FINDR

$K \leftarrow$ Intrinsic Dimension
$I \leftarrow$ original image with dimension reduced to $(K-1)$
$E \leftarrow \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \underline{e}_1 & \underline{e}_2 & \cdots & \underline{e}_K \end{pmatrix}$ where $\underline{e}_i$ are randomly selected pixels from I
$volume \leftarrow$ volume of E
$v2 \leftarrow volume$
$v1 \leftarrow -1$
$maxiter \leftarrow$ maximum number of iterations
$iter \leftarrow 1$
while $iter < maxiter$ **and** $v2 > v1$ **do**
  for $k = 1 \rightarrow K$ **do**
    for $r = 1 \rightarrow I.rows$ **do**
      for $c = 1 \rightarrow I.columns$ **do**
        $cE \leftarrow E$
        $cE(2:p, k) \leftarrow I(r, c, 1:(K-1))$
        $cvolume \leftarrow$ volume of cE
        if $cvolume > volume$ **then**
          $E \leftarrow cE$
          $volume \leftarrow cvolume$
          $P(1, k) \leftarrow r$
          $P(2, k) \leftarrow c$
        end if
      end for
    end for
  end for
  $iter \leftarrow iter + 1$
  $v1 \leftarrow v2$
  $v2 \leftarrow volume$
end while
Return P, the pixel co-ordinates of the endmembers
And return $I(P)$, the endmember vectors

This method was tested in [67] on synthetic data and Cuprite (AVIRIS). The application to real data was considered to yield good results in terms of comparison with laboratory spectra, comparison of derived endmember abundance maps, and visual examination of how the simplex encloses the scatter plot of the original image pixels.

2.3.2. Vertex Component Analysis (VCA)

VCA was introduced by Nascimento and Bioucas-Dias [68] in 2005. As in N-FINDR above, this method assumes the presence of pure pixels, and it must first reduce the dimension of the data. However, experimental results showed that different reduction methods are applicable for different noise levels. Specifically, the authors have found a threshold in the SNR (measured in dB), above which, SVD is more applicable (reduction to dimension K), and below which, PCA is more applicable (reduction to dimension $(K-1)$). This threshold was found experimentally to be $\text{SNR}_{th} = 15+10\log_{10}(K)$. VCA was shown in [68] to perform better than Pixel Purity Index (PPI) [65] and comparably to N-FINDR (but with much lower computational complexity) in Cuprite (AVIRIS). Once the image has been reduced, the algorithm iteratively creates new basis vectors, by taking the maximum vector in the projection orthogonal to all the other endmembers that have been determined. This uses the fact that all endmembers are orthogonal, and avoids the necessity for comparing each pixel combination. The algorithm is shown in Algorithm 3.

2.3.3. Simplex Growing Algorithm (SGA)

SGA was introduced by Chang and his colleagues [69] in 2006. The authors noted that N-FINDR and VCA found an initial set of endmembers using random selection, and this could affect the end result (i.e. the same algorithm run twice could yield different sets of endmembers). To avoid this, only one pixel is randomly selected, and the simplex is grown from there, with the new vertices chosen by maximizing the simplex volume, similar to N-FINDR. Chang *et al.* have done extensive experiments to show that the random initial choice of pixel does not affect the end results, so that the resulting endmembers are the same, for every run of the algorithm. As in N-FINDR and VCA, K must be known a-priori. Chang *et al.* use the HFC method described above to determine K. The details of the algorithm are shown in Algorithm 4.

Once the endmember spectra have been determined, the abundances of each endmember may be calculated for each pixel in the image. This is an optimization problem that is discussed in detail in the following section.

Algorithm 3 VCA (taken from [68])

$K \leftarrow$ Intrinsic Dimension
$R \leftarrow [\underline{x}_1, \dots \underline{x}_N]$, where $\underline{x}_i$ are pixels in the image
$\text{SNR}_{th} = 15 + 10\log_{10}(K)$
SNR $\leftarrow$ SNR in dB of the image
if SNR $>$ SNR$_{th}$ **then**
  $d \leftarrow K$
  $X \leftarrow U_d^T R$ (where U_d is calculated from SVD)
  (U_d is the matrix given by the first d eigenvectors, and is commonly used in dimension reduction)
  $\underline{u} \leftarrow mean(X)$
  $Y(:,j) \leftarrow X(:,j)/(X(:,j)^T \underline{u}) \forall j$
else
  $d \leftarrow K - 1$
  $X(:,j) \leftarrow U_d^T(R(:,j) - mean(\underline{x}_i))$ (where U_d is calculated from PCA)
  $c \leftarrow \arg\max_j\{||X(:,j)||\}$
  $\underline{c} \leftarrow [c, \dots, c]$ where C is a $(1 \times N)$ vector
  $Y \leftarrow \begin{pmatrix} X \\ \underline{c} \end{pmatrix}$
end if
initialize $A \leftarrow \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \end{pmatrix}$
for $i = 1 \rightarrow K$ **do**
  $\underline{w} = randn(0, I_K)$
  $\underline{f} = [(I - AA^{\#})\underline{w}]/||(I - AA^{\#})\underline{w}||$
  $\underline{v} = \underline{f}^T Y$
  $k = \arg\max_j\{||V(:,j)||\}$
  $A(:,i) = Y(:,k)$
  $index(i) = k$
end for
if SNR $>$ SNR$_{th}$ **then**
  $E = U_d X(:, index)$
else
  $E = U_d X(:, index) + mean(\underline{x}_i)$
end if

Algorithm 4 Simplex Growing Algorithm

$K \leftarrow$ Intrinsic Dimension
$I \leftarrow$ original image with dimension reduced to 2
Choose $\underline{y}$ to be a random pixel from I
$\underline{e}_1 = \arg\max_{\underline{x}} |\det \begin{pmatrix} 1 & 1 \\ \underline{y} & \underline{x} \end{pmatrix}|$, where $\underline{x} \in I$
($|\det(X)|$ refers to the absolute value of the determinant of X)
for $t = 2 \rightarrow K$ **do**
  $I \leftarrow$ original image with dimension reduced to t
  for $\underline{x} \in I$ **do**
    $E(\underline{e}_1, \ldots, \underline{x}) \leftarrow \begin{pmatrix} 1 & \cdots & 1 \\ \underline{e}_1 & \cdots & \underline{x} \end{pmatrix}$
    $V(\underline{e}_1, \ldots, \underline{x}) \leftarrow \frac{|\det(E(\underline{e}_1, \ldots, \underline{x}))|}{t!}$
  end for
  $\underline{e}_t \leftarrow \arg\max_{\underline{x}} V(\underline{e}_1, \ldots, \underline{x})$
end for
Return E, the endmember vectors

2.4. Unmixing methods

To follow the procedure of unmixing; first the number of endmembers are determined using Section 2.1; this number is then used to reduce the data dimension using Section 2.2, and the endmember spectra are approximated using Section 2.3; finally, these spectra are used to determine the abundances $\underline{u}_i$ for every pixel value $\underline{x}_i$. The abundance estimation is an optimization problem that determines which combination of endmembers matches the observed pixel value most accurately, while still obeying the sum-to-one and positivity constraints imposed on the abundances.

2.4.1. Non-negative Least Squares

The abundance fractions are bound by sum-to-one and positivity constraints, but the implementation of both constraints is difficult [13]. This method uses only the positivity constraints and simply minimises the sum-of-squares between the observed pixel value and the approximated pixel value using the known endmembers and unknown abundances. If each pixel is defined as in (2.1.1), then

$$\begin{aligned} \underline{u}_i &= \arg\min_{\underline{a}_i} ||\underline{x}_i - V\underline{a}_i|| \\ \text{such that } \underline{u}_{ij} &> 0 \ \forall j,\ 1 \leq j \leq K. \end{aligned} \tag{2.4.1}$$

where V is the $(p \times K)$ matrix of endmembers found using Section 2.3.

To implement the sum-to-one constraint, each pixel could be normalised so that the sum of abundances is one. However, Heinz and Chang [13] show that this does not lead to optimal solutions. Alternatively, the extra constraint may be added into the variables $\underline{x}_i$ and V so that the sum of abundances without weights sums to one. However, this leads to an overdetermined system which may not have a solution.

2.4.2. Fully constrained Least Squares (FCLS)

To avoid the limitations described above, a new method was developed in [13] to adapt an iterative non-negative least squares.

In unconstrained Least Squares, the abundances $\underline{u}_i$, for each pixel i, are given by:

$$\underline{u}_i^{LS} = (V^T V)^{-1} V^T \underline{x}_i \tag{2.4.2}$$

where V is the $(p \times K)$ matrix of endmembers found using Section 2.3, and $\underline{x}_i \in \mathbb{R}^p$ is the observed vector of the i^{th} pixel.

The implementation of FCLS is similar to the adding of rows to the input of the non-negative least squares, described above, but the algorithm works in an iterative way in order to avoid the pitfalls of the latter method. So again, an extra row of ones is attached to V and a corresponding extra one to $\underline{x}_i$ in the algorithm input, for each pixel i, $1 \leq i \leq N$. A Lagrangian, L, is formed

$$L = 0.5(V\underline{u}_i - \underline{x}_i)(V\underline{u}_i - \underline{x}_i)^T + \underline{\delta} \cdot (\underline{u}_i - \underline{c}) \tag{2.4.3}$$

where $\underline{c}$ is a positive constant vector, and $\underline{\delta}$ is a $(K \times 1)$ penalty vector for negative abundances. If it is assumed that the derivative of L with respect to $\underline{u}_i$ is zero, an iterative set of equations is used to determine the abundances:

$$\underline{u}_i^{FCLS} = (V^T V)^{-1} V^T \underline{x}_i - (V^T V)^{-1} \underline{\delta} \tag{2.4.4}$$

$$= \underline{u}_i^{LS} - (V^T V)^{-1} \underline{\delta} \tag{2.4.5}$$

$$\underline{\delta} = V^T (\underline{x}_i - V \underline{u}_i^{FCLS}) \tag{2.4.6}$$

The initial estimate for $\underline{u}_i$ is given by (2.4.2).

The algorithm (shown in Algorithm 5) defines two sets, an active set and a passive set. The active set contains the indices of all negative components of $\underline{u}_i$, while the passive set contains the remainder. The active set is used to "steer" the negative components

towards the non-negativity constraint. After every calculation of $\underline{u}_i^{FCLS}$, determine the active and passive sets and use only the indices corresponding to the active set to determine the new $\underline{\delta}$. Iteration is continued until all components of $\underline{u}_i$ are positive.

Algorithm 5 FCLS (Taken from [70])

Set the passive and active sets, so that active set is empty.
for i pixels in the image **do**
  Initialise $\underline{\hat{u}} = \underline{u}_i^{LS}$
  while All components of $\underline{\hat{u}}$ are negative **do**
    Move all indices corresponding to negative $\underline{\hat{u}}$ to the active set.
    Remove rows in $\underline{\hat{u}}$ and columns in V that correspond to the passive indices.
    $\underline{\delta} = (V^T V)\underline{\hat{u}}$.
    while Not all components of $\underline{\delta}$ are negative **do**
      Move the index corresponding to the maximum positive component of $\underline{\delta}$ to the passive set. (This may iterate; if the same index is selected twice, it must remain in the active set.)
      Remove columns in V corresponding to the new passive index.
      $\underline{\hat{u}} = \underline{u}_i^{LS} - (V^T V)^{-1}\underline{\delta}$
      If they exist, move all indices corresponding to negative $\underline{\hat{u}}$ to the active set
      Consider the rows in $\underline{\hat{u}}$ and columns in V that correspond to the active indices.
      $\underline{\delta} = (V^T V)\underline{\hat{u}}$.
    end while
    Remove columns in V corresponding to the passive index.
    $\underline{\hat{u}} = \underline{u}_i^{LS} - (V^T V)^{-1}\underline{\delta}$
  end while
end for

For a complete description of the algorithm, see [70].

2.4.3. Particle Swarm Optimization (PSO)

PSO was introduced by Kennedy and Eberhart in 1995 [71] in order to optimize continuous nonlinear functions. The method simulates the social behaviour of swarming insects, where the swarm is made up of many particles, and the movement of each particle is determined by the best overall position in the swarm [71]. Each particle $\underline{y}_i$ is initialised by a random vector, along with a record of its personal best value $\underline{p}_i$, and the global swarm best value $\underline{g}$, as determined by the optimization function $f(\underline{y})$. The particle is then allocated a vector of velocities $\underline{v}_i$, which is dependent on the previous velocity, the personal best value of the particle, and the global best value of the swarm. The user may set the weights of each of these dependencies. The particles are updated iteratively until they converge to the global solution.

If function $f(\underline{y})$ is to be minimized, then the PSO algorithm is shown in Algorithm 6. The constants c_1 and c_2 are set by the user.

In the case of hyperspectral unmixing, for each p-dimensional pixel vector, $\underline{x}_i$, and a global $(p \times K)$ matrix of endmembers, V, the K-dimensional abundances $\underline{u}_i$ are calculated by minimizing $||\underline{x}_i - V\underline{u}_i||$ for each $i = 1, \ldots, N$. To enforce the sum-to-one and positivity constraints, set $u_{iK} = 1 - \sum_{j=1}^{K-1} u_{ij}$ for each i, and introduce a penalty term for negative abundances, so that the objective function becomes

$$f(\underline{u}) = ||\underline{x}_i - V\underline{u}_i|| + q\delta(\underline{u})I_K \tag{2.4.7}$$

$$\text{where } \delta(\underline{u}) = \begin{cases} 0 & \text{if } u_{ij} \geq 0, \forall j \\ 1 & \text{if } u_{ij} < 0, \end{cases} \tag{2.4.8}$$

where q is a constant penalty term, and I_K is the $(K \times K)$ identity matrix.

Algorithm 6 PSO (Taken from [71])

Initialise all particles and velocities, $\underline{y}_i^1$ and $\underline{v}_i^1$ with random values for all i
for $j = 1 \rightarrow j_{max}$ **do**
  for all i **do**
    if $f(\underline{y}_i^j) < \underline{p}_i$ **then**
      $\underline{p}_i = f(\underline{y}_i^j)$
      $\underline{p}_i^{\text{index}} = \underline{y}_i^j$
    end if
    if $f(\underline{y}_i^j) < \underline{g}$ **then**
      $\underline{g} = f(\underline{y}_i^j)$
      $\underline{g}^{\text{index}} = \underline{y}_i^j$
    end if
    Choose random numbers r_1 and r_2 for each dimension d, so that
    $\underline{v}_{id}^{j+1} = \underline{v}_{id}^j + c_1 r_1(\underline{p}_{id} - \underline{y}_{id}^j) + c_2 r_2(\underline{g}_d - \underline{y}_{id}^j)$
    $\underline{y}_i^{j+1} = \underline{y}_i^j + \underline{v}_i^{j+1}$
  end for
end for

In conclusion, only FCLS and PSO allow for proper implementation of the sum-to-one constraint. PSO does not require the formation of a Lagrangian and so it is expected to perform substantially faster. PSO will be used to unmix some real images for validation in Chapter 8.

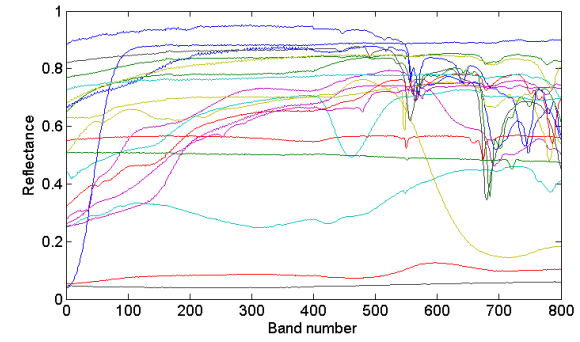
3. Data

Data are an important part of this project, since the method developed here must be tested on several datasets to prove consistent results. The hyperspectral images used in this study contain as wide a selection of images as possible, including simulated sets for robustness testing, different scenes acquired by the same sensor, the same scene acquired by different sensors, images acquired with satellite, aerial and ground based (hand held or mounted) platforms, and mineral and vegetation datasets.

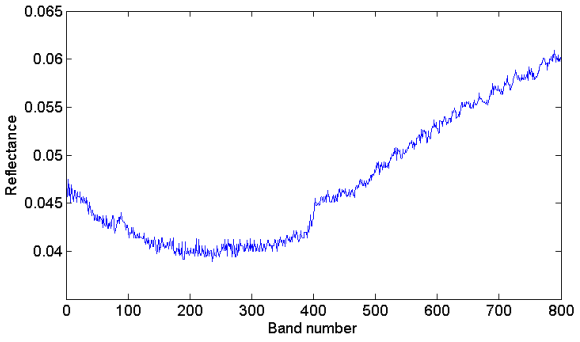
3.1. Synthetic data

A synthetic dataset is simulated using 20 minerals taken from the JPL spectral library. At each iteration of the algorithm, K unique minerals are chosen randomly from the library, and the proportions of each of the basis vectors in each pixel are randomly selected, with the only restrictions being the positive and sum-to-one conditions that are enforced i.e. the endmember fractions are uniformly distributed within a simplex. The values are also varied for p (number of bands), N (number of pixels), and σ (standard deviation of noise) in these images. As seen in Figure 3.1, the spectra in the synthetic dataset vary widely — some are similar and others easily separable. By mixing random spectra together, images are simulated that may be easy or difficult to process. Since the proportions are randomly selected as well, one or more basis vector may occur in very small proportions in the image. All of these properties were created in order to simulate as realistic a dataset as possible.

Figure 3.1(a) shows that there are two spectra in the dataset that appear to be flat and are low amplitude. A lack of features means that the signal may not be identifiable. Figure 3.1(b) shows that Graphite is not only flat, it is also noisy. These two spectra will be removed from the standard synthetic set and will be evaluated separately. This dataset will be used to execute performance analyses of all methods. Since all variables are known, success rates may be calculated based on certain criteria.



(a)



(b)

Figure 3.1.: 3.1(a) All the spectra used to make up the test dataset. Note that there is a mix between similar and easily separable spectra, flat spectra and those with sharp features, spectra with high and low amplitude, etc. This is done to mimic a real environment. 3.1(b) The spectra (Graphite 1A) with the lowest amplitude is shown at a larger scale. Take note of the noisiness of the spectrum.

3.2. Summary of sensor characteristics

Several different sensors are considered for real applications. The images themselves will be discussed in greater detail in the sections below, and the sensor characteristics are summarised in Table 3.1.

Table 3.1.: The characteristics of five sensors; Hyperion, AVIRIS, SpecTIR, HCI and CAO.

Sensor:	Hyperion	AVIRIS	SpecTIR	HCI	CAO
Sensor type	Pushbroom	Whiskbroom	Pushbroom	Whiskbroom	Pushbroom
Number of bands	220	224	360	640	72
Spectral range (nm)	400 – 2500	370 – 2500	400 – 2500	500 – 2500	380 – 1000
Spectral resolution	10 nm	10 nm	2.3 nm (VNIR) 5.8 nm (SWIR)	5 nm	10 nm
Spatial resolution	30 m	20 m	1 m	0.5 mm	1.12 m

3.3. AVIRIS

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) is flown by NASA's Jet Propulsion Lab on the NASA ER-2 aircraft at an altitude of 20 km. All the images considered are freely available online[1].

3.3.1. Cuprite

Figure 3.2 shows an AVIRIS image over Cuprite, Nevada, USA. The image is a subset from the reflectance image acquired in 1997, and has 350×350 pixels, 189 bands (bands 1–3, 105–115, and 150–170 were removed from the original data due to water absorption and low SNR, as in [10, 69]), and an average SNR of approximately 500:1 [40]. This was calculated by dividing the mean by the standard deviation of a homogeneous area (Stonewall Playa) and normalising to 50% reflectance [40]. While the exact intrinsic dimension of this dataset is unknown, it is a well-studied image, and Wu *et al.* [10] have tested 6 other methods to determine K in this image. In addition, ground truth

[1] aviris.jpl.nasa.gov/html/aviris.freedata.html

collected by Swayze *et al.* found at least 18 substances [72] (which may not include rarer minerals), and Chang *et al.* reported that $K = 22$ was the minimum number to guarantee that his endmember extraction algorithm could identify 5 ground truth points in the image [69].

Cuprite is located in the Nevada desert, USA, and is a useful test site as there is very little vegetation, so the scene does not vary in different seasons. Some common minerals found in Cuprite include kaolinite, alunite and hydrothermal silica [40].

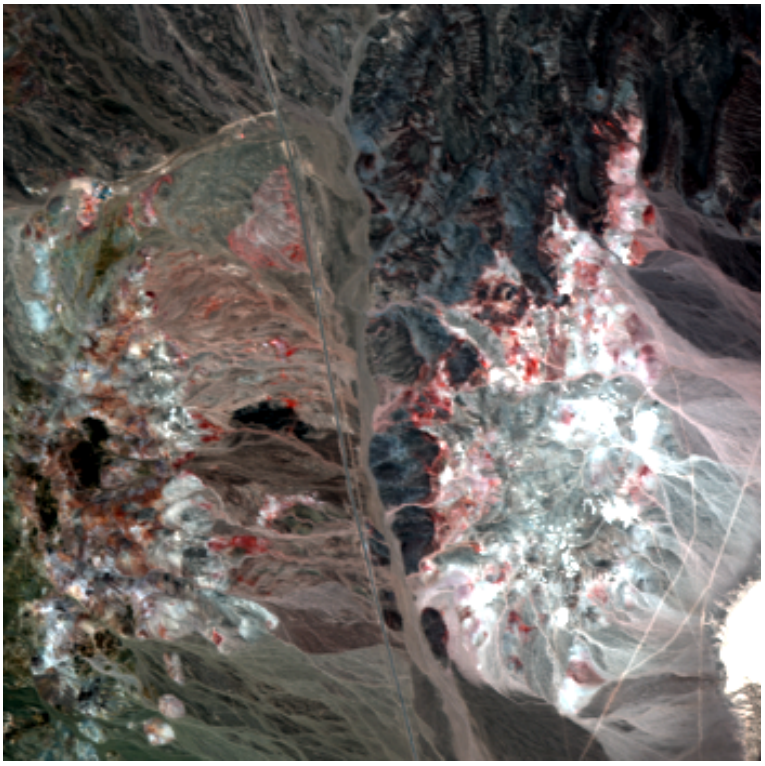


Figure 3.2.: A subset of the 1997 AVIRIS image over Cuprite, Nevada. The RGB image has been chosen for this illustration, with red approximated by band 35, green approximated by band 15, and blue approximated by band 9. This image has been chosen as it is a well studied test site.

3.3.2. Lunar Lakes

This image is a subset from the radiance image acquired in 2009, shown in Figure 3.3. It has 350×350 pixels and 200 bands. Lunar Lakes is contained in the Lunar Crater Volcanic Field located in the Nevada desert, USA, and so like Cuprite, it does not vary

over the seasons.

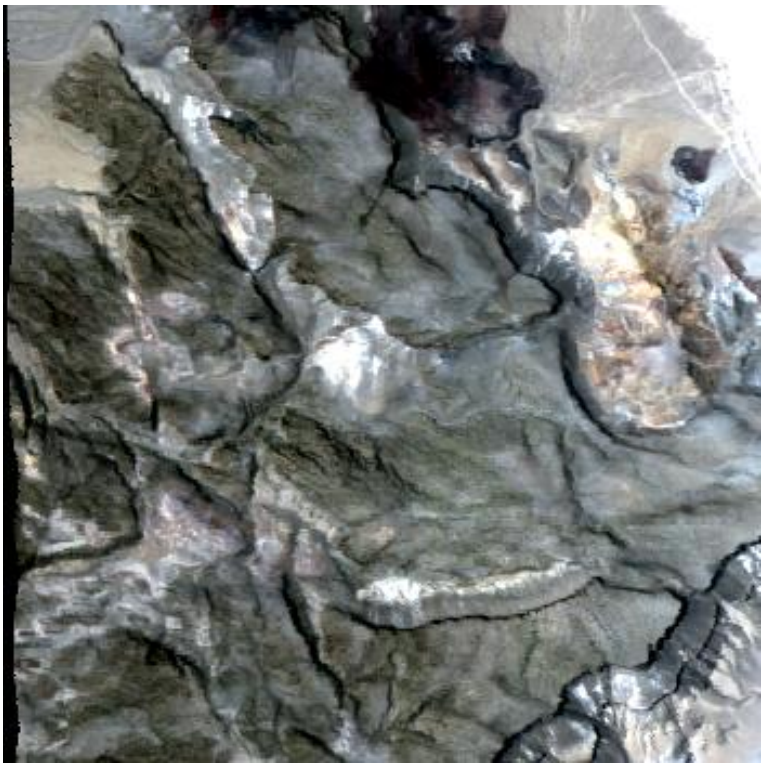


Figure 3.3.: A subset of the 2009 AVIRIS image over Lunar Lakes, Nevada. The RGB image is displayed, with red approximated by band 29, green approximated by band 20, and blue approximated by band 11.

3.3.3. Moffet Field

This AVIRIS image was aquired over Moffet field, California, USA. The image is a subset from the radiance image acquired in 2008, and has 350×350 pixels, 189 bands (bands 1–3, 105–115, and 150–170 were removed from the original data, as in Cuprite above). This image is of special interest as it includes water, urban settlements and vegetation (see Figure 3.4), where other AVIRIS scenes show mainly mineral data.

3.4. Hyperion

Hyperion is a hyperspectral sensor mounted on the EO-1 satellite. Hyperion is capable of producing 220 spectral bands over the spectral range, but the datasets acquired had

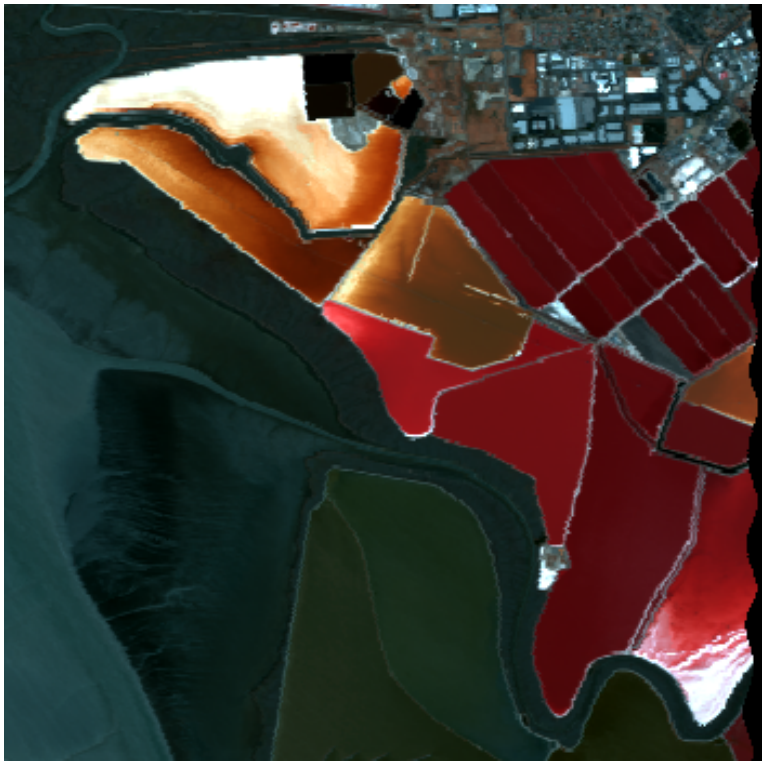


Figure 3.4.: A subset of the 2008 AVIRIS image over Moffet Field, California, USA. The RGB image has been chosen for this illustration, with red approximated by band 35, green approximated by band 15, and blue approximated by band 9.

several bands with zero values throughout and so these were removed to leave 184 bands (bands 10–56 and 82–218 were used). The original images are 7.5 km wide and 100 km long and two separate grating image spectrometers detect VNIR and SWIR wavelengths. Data are available for free online[2]. Subsets of the original data are considered.

3.4.1. Cuprite

The Hyperion subset shown in Figure 3.5 was acquired in 2011 over Cuprite, Nevada, USA. It has 200×200 pixels and is over an area that is comparable to the AVIRIS Cuprite scene, as seen in 3.6.

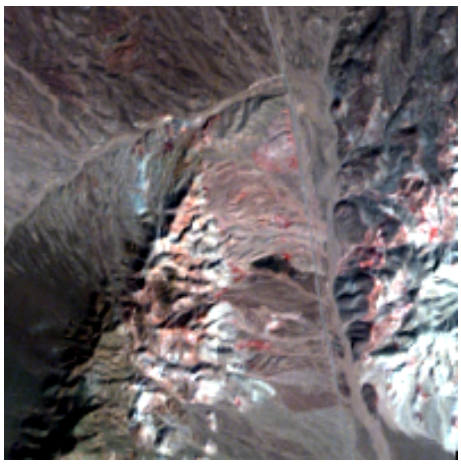


Figure 3.5.: A subset of the 2011 Hyperion image over Cuprite, Nevada. The RGB image is used in this illustration, with red approximated by band 26, green approximated by band 9, and blue approximated by band 3.

3.4.2. Lunar Lakes

The Hyperion subset shown in Figure 3.7 was aquired in 2010 over Lunar Lakes, Nevada, USA. It has 250×250 pixels.

3.4.3. Moffet Field

The Hyperion subset shown in Figure 3.8 was acquired in 2011 over Moffet Field, California, USA. It has 200×200 pixels and is over an area that is comparable to the AVIRIS Moffet Field scene.

[2]http://edcsns17.cr.usgs.gov/NewEarthExplorer/

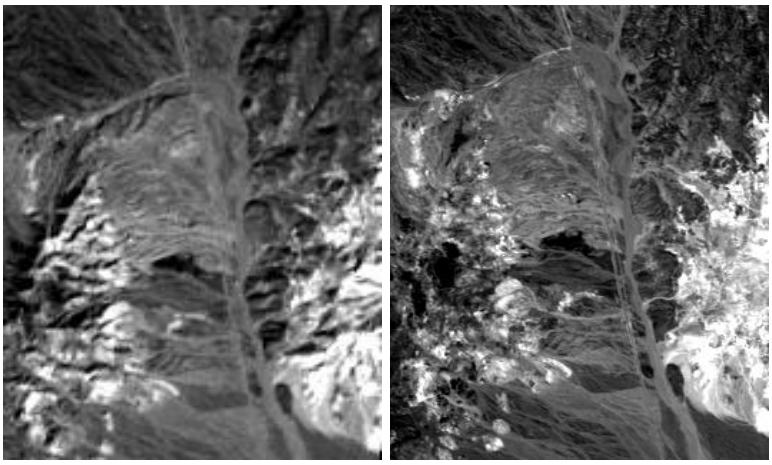


Figure 3.6.: Subsets of Cuprite from Hyperion (left) and AVIRIS (right) show the overlapping spatial region.

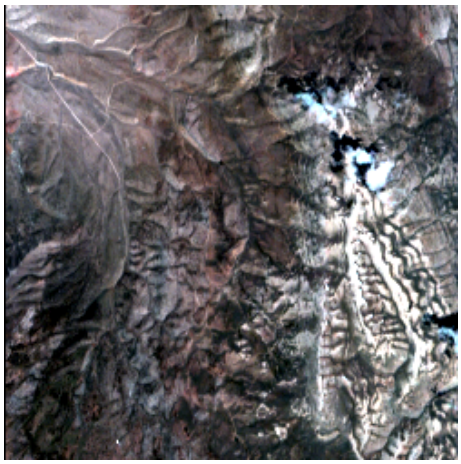


Figure 3.7.: A subset of the 2010 Hyperion scene over Lunar Lakes, Nevada, USA. The RGB image is displayed, with red approximated by band 26, green approximated by band 9, and blue approximated by band 3.

Figure 3.8.: A subset of the 2011 Hyperion scene over Moffet Field, California, USA. The RGB image has been used in this illustration, with red approximated by band 26, green approximated by band 9, and blue approximated by band 3.

3.5. SpecTIR

The ProSpecTIR VS sensor is a privately owned airborne sensor, and subsets are considered with approximate spectral range 400 – 1700 nm. The data was processed to reflectance using the same procedures as AVIRIS, and is available online[3].

3.5.1. Cuprite

The subset used in this study is a reflectance image containing 320×320 spatial pixels, and 230 spectral bands. The three Cuprite sets that are considered (acquired by AVIRIS, Hyperion and SpecTIR) do not cover the exact same spatial area, but the SpecTIR scene (see Figure 3.9) is a subset of the AVIRIS scene (see Figure 3.10.

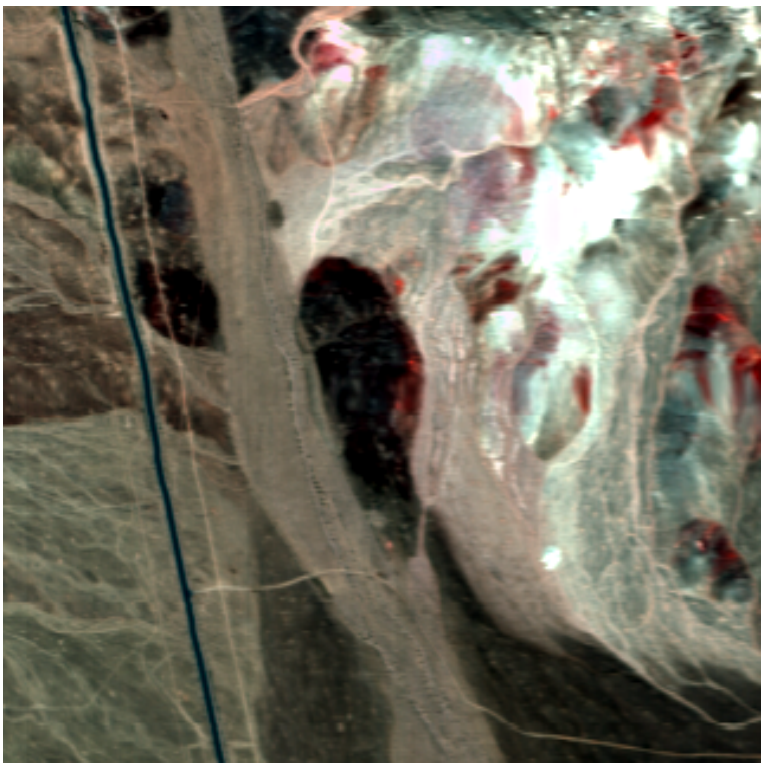


Figure 3.9.: A subset of a SpecTIR Cuprite scene. The RGB image of this subset is displayed, with red approximated by band 66, green approximated by band 30, and blue approximated by band 17.

[3]http://www.spectir.com/download.html

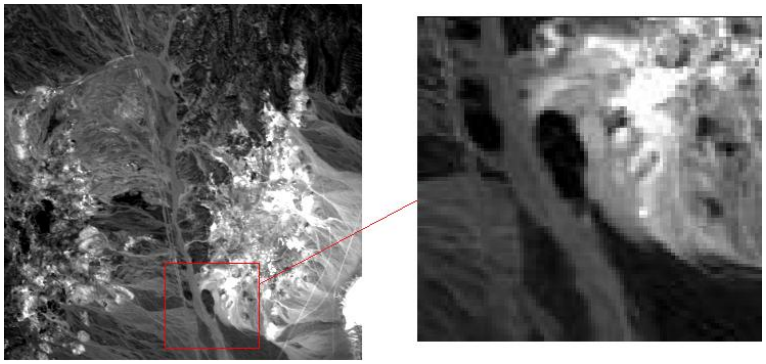


Figure 3.10.: A subset of the AVIRIS Cuprite scene, shown to match up spatially with the SpecTIR scene in Figure 3.9.

3.5.2. Suwannee

The SpecTIR subset shown in Figure 3.11 is a 2010 reflectance image taken in the Gulf of Mexico over a wetland area. The image contains 320×320 spatial pixels, (at 2 m spatial resolution) and 230 spectral bands.

3.6. Hyperspectral Core Imager (HCI)

HCI is an instrument owned by Anglo Gold Ashanti. The subset considered in this study has 264×75 pixels and 141 spectral bands from the short wave infrared (SWIR) spectrometer. This is a sensor that sits directly over the object it is scanning, and so it has negligible atmosphere. This means that signal is not affected by absorptions in the atmosphere at certain frequencies, and preprocessing algorithms are not needed to remove atmospheric effects. Experts have estimated the number of endmembers in the scene at $3 \leq K \leq 7$. A sample from HCI is displayed in Figure 3.12.

3.7. CAO

The Carnegie Airborne Observatory (CAO) flew their sensor over the Kruger National Park (South Africa) and surrounding areas in 2008. Images were taken over eight landuse areas and both granite and gabbro regions. Landuse 7 is a dataset over the Bushbuckridge area, containing degraded communal areas, trees, grass, soils, houses etc. Three spatial subsets are considered from landuse 7, including settlements, grasslands and tree cover, shown in Figures 3.13 – 3.15. Each image contains 350×350 pixels, and the 72

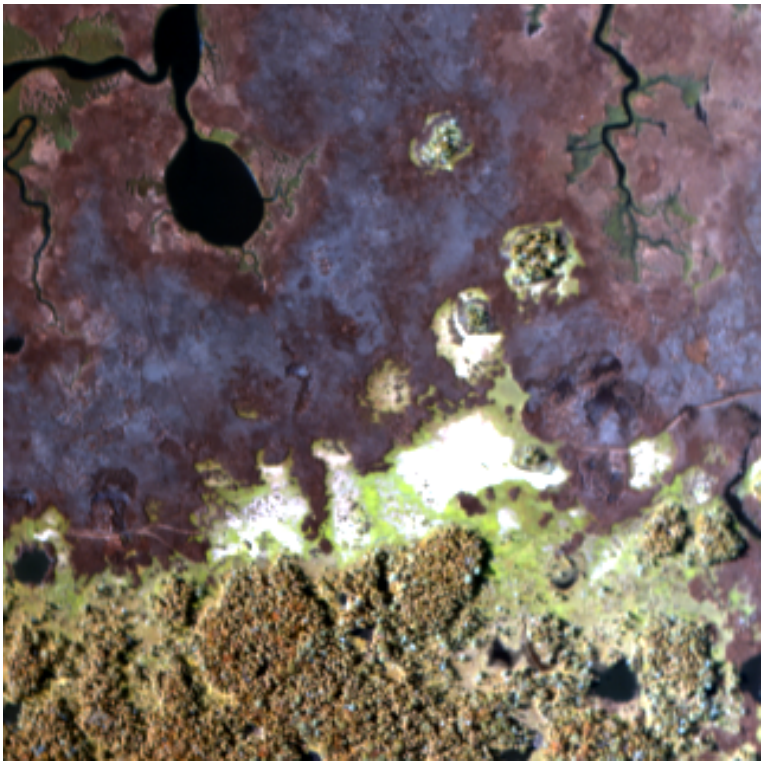


Figure 3.11.: A subset of the SpecTIR Suwannee scene over the Gulf of Mexico. The RGB image of this subset is displayed, with red approximated by band 66, green approximated by band 30, and blue approximated by band 17.

Figure 3.12.: A subset of a Hyperspectral Core Imager (HCI) image. The tenth band of this subset is displayed.

band product that is considered is a combination of 288 original bands. This improved the quality of the spectral responses, and resulted in high signal-to-noise ratios.

Figure 3.13.: A subset from the CAO scene over the Bushbuckridge area showing houses, roads and fields. The RGB image is used for this illustration, with red approximated by band 35, green approximated by band 16, and blue approximated by band 10.

Figure 3.14.: A subset from the CAO scene over the Bushbuckridge area showing trees, grass and water. The RGB image is used in this illustration, with red approximated by band 35, green approximated by band 16, and blue approximated by band 10.

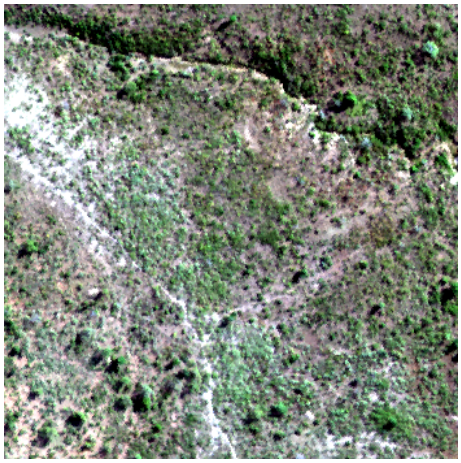


Figure 3.15.: A subset from the CAO scene over the Bushbuckridge area showing grass, trees and bare soil. The RGB image is used in this illustration, with red approximated by band 35, green approximated by band 16, and blue approximated by band 10.

4. Noise approximation

Consider the scenario where signal and noise are linearly separable, so that the p-dimensional measurement $\underline{x}_i$ of the ith pixel may be written $\underline{x}_i = \underline{s}_i + \underline{n}_i$ where $\underline{s}_i$ is a p-dimensional vector representing the information in the i^{th} pixel and $\underline{n}_i$ represents a p-dimensional random process present in all pixels. This process is called noise, and it is assumed that this noise follows a zero-centered Gaussian distribution. When independently identically distributed (i.i.d.) noise is considered, the noise has the same variance in every spectral band, so that $\underline{n}_i$ follows a Gaussian distribution $\mathcal{N}(0, \sigma^2 I_p) \forall i, 1 \leq i \leq N$. In non-i.i.d. noise, the noise is described by $\underline{n}_i \sim \mathcal{N}(0, \Sigma)$, where Σ is the noise covariance matrix that need not be proportional to the identity matrix. If Σ is diagonal, then the noise is uncorrelated among the spectral bands; if the noise is correlated among the spectral bands, the covariance matrix Σ is not diagonal.

It is important to estimate the noise for many of the methods that determine the intrinsic dimension of a hyperspectral image. Some methods require an estimate of the noise covariance to whiten the observation covariance matrices. These may use a residual method for the approximation. Other methods require an estimate of the noise on a per pixel basis and use multiple regression theory to approximate this.

Since the noise in a hyperspectral image is often difficult to characterise, it would be an advantage for all noise approximations to require only the image as an input, instead of requiring *a priori* knowledge of the noise. The resulting noise variance needs to be reliable in order for the methods described above to be used successfully.

Below, several methods are discussed for both i.i.d. and non-i.i.d. noise approximations, and each method is tested on real and synthetic data so that its performance may be evaluated. Even though noise in real hyperspectral images may not be i.i.d., we would like to test the sensitivity of the noise estimation methods to this assumption.

4.1. i.i.d. noise approximation methods

The noise in an image is said to be independent and identically distributed (i.i.d.) if the noise distribution has the same variance, σ^2, in each band. This assumption is common to several methods described in Section 2.1. In this section i.i.d. noise is assumed, *i.e.* the noise value $\underline{n}_i$ in the ith pixel follows a Gaussian distribution $\mathcal{N}(0, \sigma^2 I_p)$.

The noise approximation algorithms that are considered all use the eigenvalues of the observation covariance matrix $S(N)$. This matrix is defined in Section 2.1 for centered data as

$$S(N) = \frac{1}{N} \sum_{i=1}^{N} \underline{x}_i \underline{x}_i^T,$$

where $\underline{x}_i \in \mathbb{R}^p$ is a column vector of measurements of the i^{th} pixel, over p spectral bands, and N is the number of samples. For each i, $\underline{x}_i$ is a realisation of the random p-dimensional vector $\underline{\tilde{x}}$. Kritchman and Nadler [14] show that $S(N)$ tends to the model covariance $\Sigma = \mathbb{E}(\underline{\tilde{x}}\underline{\tilde{x}}^T)$ with measure one (as $N \to \infty$) and may be diagonalised as

$$\begin{pmatrix} \lambda_1 & & & 0 \\ & \cdots & & \\ & & \lambda_K & \\ 0 & & & 0 \end{pmatrix} + \sigma^2 I_p, \tag{4.1.1}$$

where K is the intrinsic dimension of the image, also thought of as the number of signals, or the number of endmembers (see Definition 1), and λ_i is the i^{th} eigenvalue of the signal covariance matrix. This matrix decomposition shows that if $N \to \infty$, constant noise variance σ^2 may be approximated by

$$\hat{\sigma}^2_{REF} = \frac{1}{p-K} \sum_{j=K+1}^{p} \lambda_j, \tag{4.1.2}$$

where $\hat{\sigma}^2_{REF}$ will be considered as the *reference noise approximation*. Kritchman and Nadler [14] report that (4.1.2) may be improved by using a scaling factor, and the new approximation is described as the *scaled noise approximation* and is given by

$$\hat{\sigma}^2_0 = \frac{\hat{\sigma}^2_{REF}}{1 - K/N} \tag{4.1.3}$$

where p is the spectral dimension, K is the number of endmembers in the mixture and N is the number of samples. Using $\hat{\sigma}_0^2$ as an initial estimate, Kritchman and Nadler [14] derive an iterative method for more reliable noise estimation, especially in their case where the number of samples is limited. The method diagonalises only the upper left submatrix of $S(N)$ to capture the signal subspace. The diagonal elements of this submatrix are approximated by $\hat{\rho}_j$ for $j = 1, \ldots, K$. The *Kritchman-Nadler* variance, $\hat{\sigma}_{KN}^2$, may then be calculated by solving the following equations iteratively

$$\hat{\sigma}_{KN}^2 - \frac{1}{p-K}\left(\sum_{j=K+1}^{p}\lambda_j + \sum_{j=1}^{K}(\lambda_j - \hat{\rho}_j)\right) = 0 \tag{4.1.4}$$

$$\hat{\rho}_j^2 - \hat{\rho}_j\left(\lambda_j + \hat{\sigma}_{KN}^2 - \hat{\sigma}_{KN}^2\frac{p-K}{N}\right) + \lambda_j\hat{\sigma}_{KN}^2 = 0, \tag{4.1.5}$$

where $\hat{\rho}$ is considered unknown, and K is assumed known in order to estimate the noise $\hat{\sigma}_{KN}^2$. It is shown in [14] that this noise approximation is very reliable even for low values of N.

These i.i.d. noise approximations are computationally cheap and simple to implement.

4.2. Spatially based non-i.i.d. noise approximation methods

The i.i.d. noise assumption may not be applicable to hyperspectral images, since hyperspectral sensors are known to have different noise characteristics across bands. All the methods considered in this study for finding the intrinsic dimension of an image are sensitive to the accuracy of the noise approximation. Spatially based methods such as Meer's method calculate the noise on a band-by-band basis. This means that no relationship is assumed across bands, so noise need not be i.i.d. or uncorrelated.

4.2.1. Meer's spatially based method

This method [73], developed in 1990, uses the spatial information in the image to approximate the noise. It requires no prior information about the image, and works by dividing the image into blocks and calculating the variance per block. A pyramid of block sizes is created so that the blocks become bigger in each iteration of the algorithm. If the image contains homogeneous areas then large block sizes will give a more reliable estimation of the noise variance, but large blocks may contain variance due to signal rather than noise. The smallest variance is therefore collected for each level of the pyramid, but outliers (variances not due to noise) may exist. So instead, the four smallest noise variances are collected, and various tests are performed to eliminate outliers. Slippage tests are then used to select or interpolate variances from the levels of the pyramid.

Without loss of generality, assume that the image has only one band (the noise variance in each band will be calculated separately). Assume that the image is of size $N_1 \times N_1$, where N_1 may be written as $N_1 = 2^m$. At each level in the pyramid, divide the image into blocks of size $c_l = 2^l, l = 1, 2, ..., m$. The blocks in the image at level l in the pyramid are indexed by $k_l = 1, 2, ..., K_l = 4^{m-1}$. The sample variance s_g is calculated for each block k_l and is given by:

$$s_g(k_l) = \frac{1}{4^l - 1} \sum_{i=1}^{2^l} \sum_{j=1}^{2^l} [x_{k_l}(i,j) - \bar{x}_{k_l}]^2,$$

where $x_{k_l}(i,j)$ represents the pixel measurement at row i and column j within the block k_l, and $\bar{x}_{k_l}$ is the mean pixel value in block k_l.

In order to compute the overall variance of the image, the four smallest sample variances are considered per pyramid level, in order to eliminate outliers. Let q_l be a vector containing the four smallest variances, where l refers to the level of the pyramid and q_l^i

represents the i^{th} smallest variance in q_l, then

$$\begin{aligned} q_l &= \min(s_g(k_l)) \text{ where } q_l^1 < q_l^2 < q_l^3 < q_l^4 \\ r_1 &= (q_l^2 - q_l^1)/(q_l^4 - q_l^1) \\ r_2 &= (q_l^3 - q_l^2)/(q_l^4 - q_l^2) \\ r_3 &= 1 - r_2. \end{aligned}$$

Slippage tests are used to determine if any of the q_l are outliers i.e. variance not due to noise. These outliers are removed. So, for each level of the pyramid, l, calculate the noise variance, V:

$$\begin{aligned} \text{if } r_1 &\leq 0.5 \text{ or } l = m - 1 \text{ (There are only 4 blocks at this level)} \\ \text{then } V_l &= \frac{1}{4}\sum_{i=1}^{4} q_l^i \\ \text{elseif } r_2 &\leq 0.7 \\ \text{then } V_l &= \frac{1}{3}\sum_{i=2}^{4} q_l^i \\ \text{elseif } r_3 &\leq 0.7 \\ \text{then } V_l &= \frac{1}{2}\sum_{i=3}^{4} q_l^i \\ \text{else } V_l &= q_l^4. \end{aligned}$$

There is a special case in the final level of the pyramid, where there is only one block. In that case, if $\bar{x}$ is the mean pixel value over the entire block (image), then

$$V_L = \frac{1}{(4^L - 1)} \sum_i \sum_j (x(i,j) - \bar{x})^2.$$

Variance estimates are obtained for each level of the pyramid. This sequence is the input

into a blind noise estimation algorithm defined by a set of rules listed below. First define

$$\begin{aligned}
\beta(l) &= 1 - 0.1 \times 2^{-l+6} \\
\alpha(l) &= \frac{V_{(l-1)}}{V_l} - \beta(l), l = 3, 4, ..., m \\
l_u &= \underset{l}{\text{argmin}}[\alpha(l) < 0], l = 3, 4, ..., m-2 \\
l_0 &= \underset{l}{\text{argmin}} \left[\sum_{i=l_u}^{l} \alpha(i) < T \right] \text{ so that } \sum_{i=l_u}^{l_0+1} \alpha(i) < T,\ l = 3, 4, ..., m-2
\end{aligned}$$

Rule 1

This detects an image uncorrupted by noise.

if $\exists l \geq 2$ such that $q_l^4 < \eta$ then $\sigma^2 = 0$.

Rule 2

if $l_0 = m$ then $\sigma^2 = V_m$

Rule 3 and 4

if $l_0 = 3$ or $l_0 = 4$ then 'Warning, cannot estimate'

Rule 5

Define $\rho = \alpha(5) + \alpha(6)$

Define: for $\rho_{i-1} \leq \rho \leq \rho_i$, $\delta = \frac{\rho - \rho_i}{\rho_{i-1} - \rho_i}$

if $-2 < \rho \leq -1.5$ then $\sigma^2 = V_3$

elseif $-1.5 < \rho \leq -1$ then $\sigma^2 = \delta \cdot V_3 + (1 - \delta) \cdot V_4$

if $-1 < \rho \leq -0.5$ then $\sigma^2 = \delta \cdot V_4 + (1 - \delta) \cdot V_5$

if $-0.5 < \rho \leq T$ then $\sigma^2 = \delta \cdot V_5 + (1 - \delta) \cdot V_6$, where $T = -0.1$

Rule 6 and 7

Define $\delta = |\alpha(l_0)|$

if $T < \alpha(l_0 - 1) \leq 0$ then $\sigma^2 = \delta \cdot V_{(l_0-2)} + (1 - \delta) \cdot V_{(l_0-1)}$,

where $T = -0.1$

Define: for $\alpha_{i-1} \leq \alpha(l_0) \leq \alpha_i$, $\delta = \frac{\alpha(l_0) - \alpha_i}{\alpha_{i-1} - \alpha_i}$

elseif $-1 < \alpha(l_0) \leq -0.5$ then $\sigma^2 = \delta \cdot V_{(l_0-2)} + (1 - \delta) \cdot V_{(l_0-1)}$

elseif $-0.5 < \alpha(l_0 - 1) \leq T$ then $\sigma^2 = 0.5[(1 + \delta) \cdot V_{(l_0+1)} + (1 - \delta) \cdot V_{l_0}]$,

where $T = -0.1$

The advantage of this method is that nothing is assumed about the noise in the image,

for instance the noise is not assumed to be uncorrelated or i.i.d.. The only requirement is the presence of homogeneous regions. The disadvantage is the computational complexity, although it may easily be implemented in parallel [73]. This is an old algorithm (20 years) but it provides good results for the images studied by Meer [73]. In a hyperspectral image, each band is analysed individually. This may lead to redundancy, especially with high numbers of bands, but to our knowledge this redundancy has not been exploited in Meer's method. Although Meer's method is a spatial method in a 2-dimensional image, it uses the noise estimates in each band separately to estimate the spectral noise. This is done by repeating the application of the method band by band.

4.3. Statistical non-i.i.d. noise approximation methods

Spatially based methods require large enough homogeneous regions in the image for reliable noise approximation. Alternative noise approximation methods include statistical methods, which also do not assume i.i.d. noise. Two such methods are discussed below, namely regression and residual methods.

4.3.1. Residual-based method

This method was developed by Roger [54] in 1996 and uses spatial and spectral correlation in the signal values to determine the variance in each band. The image is once again divided into spatial blocks, and multiple regression is used to approximate the values in each band using only the data in neighbouring bands. The residuals approximate the noise, and these residuals may be related to the observation covariance matrix. The method is described in [9], where Chang and Du show that the observation covariance matrix, $S(N)$, defined in (2.1.2), may be decomposed into

$$S(N) = D_S E_S D_S, \tag{4.3.1}$$

where D_S is a diagonal matrix containing the square roots of the diagonal elements of $S(N)$, denoted by $\{s_j\}_{j=1,\ldots,p}$. The matrix E_S contains ones down the diagonal, and correlation coefficients elsewhere. Similarly, the matrix $S(N)^{-1}$ may be decomposed into

$$S(N)^{-1} = D_{S^{-1}} E_{S^{-1}} D_{S^{-1}}, \tag{4.3.2}$$

where $D_{S^{-1}}$ is a diagonal matrix containing the square roots of the diagonal elements of $S(N)^{-1}$, denoted by $\{\zeta_j\}_{j=1,\ldots,p}$. The matrix $E_{S^{-1}}$ contains ones down the diagonal, and correlation coefficients elsewhere. Note that $D_{S^{-1}} \neq D_S^{-1}$ and $E_{S^{-1}} \neq E_S^{-1}$.

As described in [9], the diagonal elements of $S(N)$ and $S(N)^{-1}$ may be related by

$$\zeta_j^2 = \frac{1}{s_j^2(1 - r_j^2)}, \tag{4.3.3}$$

where r_j^2 is the multiple correlation coefficient of band j on the other $(p-1)$ bands. If a whitening matrix W is applied where $W = \text{diag}\{1/\zeta_1^2, \ldots, 1/\zeta_p^2\}$, then the noise correlation is removed as well.

The whitened sample covariance becomes

$$S_W = W^{-1/2} S(N) W^{-1/2}. \tag{4.3.4}$$

The resulting noise variance becomes 1 in each band. Use of the residual noise approximation requires the assumption that the noise is uncorrelated. In this case, the whitened noise becomes i.i.d.. This method is simple and fast to implement, and has been used successfully in [9] to calculate the number of endmembers in a hyperspectral image with noise variance that differs in each band.

4.3.2. Multiple regression based method

As described in [8], the noise may be approximated on a per pixel basis, using multiple regression theory as follows:

Let X be a $(N \times p)$ matrix so that X^i is a column vector containing all pixel values at band i. Let $X^{\partial i}$ be a $(N \times (p-1))$ matrix, where $X^{\partial i} = [X^1, \ldots, X^{i-1}, X^{i+1}, \ldots, X^p]$. Then the pixel values for each band i can be expressed in terms of the pixel values for all other bands, so that $X^i = X^{\partial i}\beta_i + \epsilon_i$, where β_i is the regression vector and ϵ_i is the modeling error. This error, ϵ_i, may be used to approximate the noise per pixel in the i^{th} band. (Bioucas-Dias and Nascimento [8] assume that $\beta_i = [X^{\partial i}]^{\#} X^i$, where $^{\#}$ indicates the pseudo-inverse.)

In the HySime method [8], the noise is calculated on a per pixel basis using the above method, and removed from each pixel as a preprocessing step to the algorithm that determines K, the intrinsic dimension of the image.

This method is computationally complex, and heavily reliant on linear input. However, a distinct advantage of this method is that it takes into account the spatial and spectral variability of the noise. Modern sensors contain noise that is highly dependent on the signal, and this photon noise has been studied in several recent papers [74] [75] [76]. HySime has been successfully used in [8] to calculate the intrinsic dimension of a hyperspectral image.

4.4. Hybrid noise approximation methods

Some of the algorithms for ID estimation allow only a single input for noise variance across the image. The non-i.i.d. approximations may be more accurate than the i.i.d. approximations, especially if the assumption of i.i.d. noise is unreasonable. A hybrid approach is introduced by taking the mean noise variance across all bands as the input.

Then the mean non-i.i.d. approximations may be used as an input to these single-input algorithms in an attempt to improve their accuracy. The following hybrid approximations are considered: Meer's method ($\bar{\sigma}^2_{Meer}$); the residual method ($\bar{\sigma}^2_R$); and the multiple regression method ($\bar{\sigma}^2_M$).

4.5. Performance of i.i.d. and hybrid noise approximations

Chapter 2 shows that some methods that determine the ID of a hyperspectral image, such as RMT, require an i.i.d. noise estimation as an input. The method is dependent on the accuracy of the noise approximation, and in this section we will test the reliability of three i.i.d. noise approximations as well as three hybrid noise approximations. First, the methods will be analysed on synthetic data, which will show that all methods are accurate when the assumptions of i.i.d. and uncorrelated noise are met. All methods are sensitive to these assumptions, and in Section 4.6 some non-i.i.d. noise approximation methods will be evaluated. In the current section, the three i.i.d. noise approximation methods and the two statistical hybrid methods will be tested.

The i.i.d. noise approximation methods require K as an input for the noise estimation and so it is important that these methods are reliable for all values of K. It is also important that all 5 methods can reliably estimate σ no matter what its value, and so the reliability of the methods should be independent of this variable too. It is important to know the minimum image size and minimum number of bands that an image should contain in order for these methods to be reliable. Meer's method is the last method to be examined and this will be discussed in Section 4.5.2. It is a spatially based method that calculates the noise variance for a single band, and so it is not suitable to test on the standard synthetic set.

4.5.1. Experiments for i.i.d. and statistical hybrid methods

Initially the simulated mineral dataset described in Section 3.1 will be used. Unless otherwise specified, the variables are fixed as follows: $p = 200$, N=10,000, $\sigma = 0.001$, $K = 5$.

For each point in the graphs displayed, the test was run 20 times, where the endmember spectra, abundances, and noise per pixel are randomly generated using the fixed N, p, K and σ for each iteration, creating a pool of different images. All tests were based on discrete values, but the graphs are shown as continuous for clarity between methods. The result, e, is the root mean squared error between the actual i.i.d. noise variance, σ^2, and approximated i.i.d. variance $\hat{\sigma}_i^2$ for $i = 1, \ldots, 20$ experiments. The root mean squared error is divided by the actual i.i.d. noise variance, to give a percentage error:

$$e = \frac{\sqrt{\frac{1}{20}\sum_{i=1}^{20}(\sigma^2 - \hat{\sigma}_i^2)^2}}{\sigma^2}. \tag{4.5.1}$$

This section will compare the accuracy of the i.i.d. noise approximations from equations (4.1.2) (σ^2_{REF}, the reference noise variance), (4.1.3) (σ^2_0, the scaled noise variance) and (4.1.5) (σ^2_{KN}, the Kritchman-Nadler variance). These methods all assume that K, the intrinsic dimension is known, and so this will be assumed true in the synthetic tests, to test only the noise approximation method.

Two hybrid methods discussed in Section 4.3 will also be compared; namely the residual based method ($\bar{\sigma}^2_R$) and the multiple regression method ($\bar{\sigma}^2_M$). Both of these methods allow for variable noise per band, so for these experiments the mean noise variance will be used. K is not assumed to be known in the hybrid methods.

For the multiple regression noise approximation, Matlab code will be used that has been provided online[1] by Bioucas-Dias and Nascimento (the function *estNoise*).

The synthetic dataset will be used to separately test the dependence of each of the 5 methods on the variables N, p, K and σ. All these interactions will be analysed below.

[1] http://www.deetc.isel.ipl.pt/jnascimento/public.html

Performance with respect to the number of pixels

In this test, the percentage root mean squared error over 20 experiments (Equation (4.5.1)) is shown for several image sizes. It is interesting to consider the dependence on N, since the Kritchman-Nadler variance σ^2_{KN} was derived in [14] especially for the case of small N, and the reference variance σ^2_{REF} is only a good approximation in the limit $N \to \infty, p \to \infty$. While all the methods produce acceptable error levels, especially for higher N, Figure 4.1 shows that the Kritchman-Nadler (σ^2_{KN}) and scaled (σ^2_0) variances significantly out-perform the reference variance (σ^2_{REF}). All three methods become more reliable as N increases, but it is interesting to note that the Kritchman-Nadler variance (σ^2_{KN}) is nearly the same as the scaled variance (σ^2_0) for the larger values for N considered in the synthetic images (σ^2_0 is the initial value for the iterative procedure used to calculate σ^2_{KN}). It is also interesting that the regression and multiple residual variances are very similar (for a large difference in computational time), and they are the worst performing methods for the tested values of N. Note that the residual method encountered singularities or badly scaled matrices when performing the inversion for $N < 1,000$. For shorter computational time, the image sizes were kept fairly small — the largest $N = 10^4$ which is only a 100×100 pixel image. However, even for these small sizes, all the approximations have an error of less than 1% for $N \geq 10,000$.

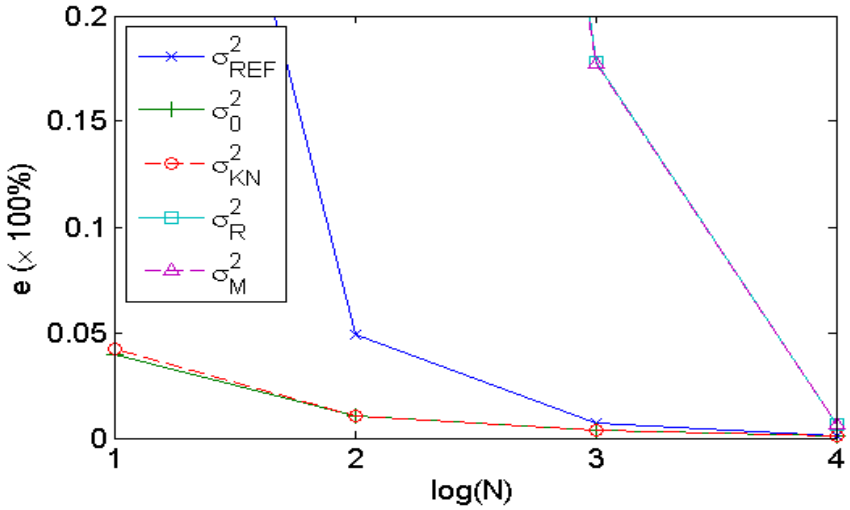


Figure 4.1.: The error of all noise variance approximations relative to the number of pixels in the image. All the approximations perform well with larger N. The Kritchman-Nadler (σ^2_{KN}) and scaled (σ^2_0) variances perform similarly and out-perform the reference variance (σ^2_{REF}), although all three approximations have only a 1% error for $N \geq 1,000$. The residual and multiple regression variances are also similar and they are the worst performing methods, although their error also drops below 1% for $N \geq 10,000$.

Performance with respect to the number of bands

The noise approximation is also dependent on p, but the improvement with large p is less obvious for the i.i.d. variances. Figure 4.2 shows that all three i.i.d. approximations are reliable, with a starting error near to 1%, and improvements as p becomes larger. So the noise approximation is less dependent on the number of bands, but the Kritchman-Nadler (σ^2_{KN}) and scaled (σ^2_0) variances remain superior. The residual and multiple regression variances follow the same trend as the other methods, although with much higher errors. In fact, the errors only drop below 1% when $p = 200$. One should bear in mind that the i.i.d. approximations use all $p \times N$ pixels in the image, whereas the non-i.i.d. approximations use only the N pixels in each band. This may account for the better performance of the i.i.d. approximations. Nevertheless, these latter methods still perform well, with errors of less than 1% for $N = 10,000$ and $p = 200$.

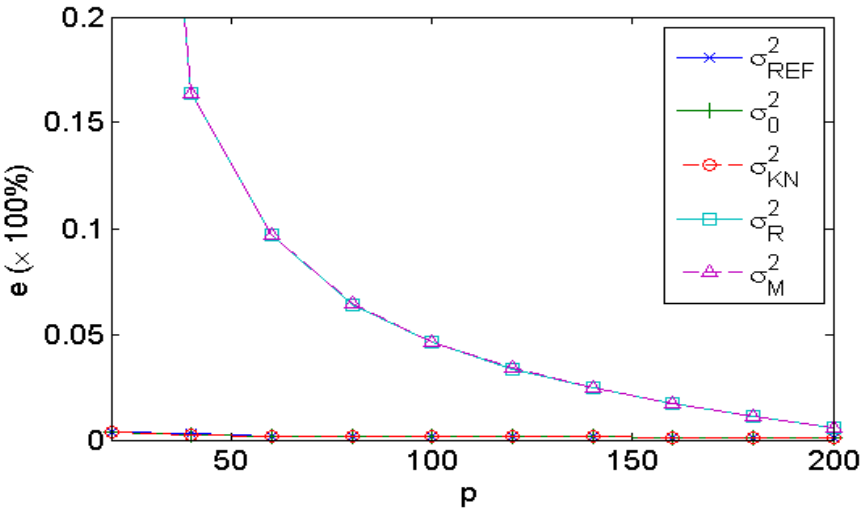


Figure 4.2.: The error of all noise variance approximations relative to the number of spectral bands. All three i.i.d. approximations perform well and improve with larger p. The Kritchman-Nadler (σ^2_{KN}) and scaled (σ^2_0) variances perform similarly and out-perform the reference variance (σ^2_{REF}). The residual and multiple regression variances perform similarly, but have much higher errors and are in general the worst approximations.

Performance with respect to the number of endmembers

The noise approximation method should not vary according to the number of endmembers, especially the three i.i.d. noise approximation methods, since all are dependent on this number. In Figure 4.3, the i.i.d. approximations behave as expected, with no dependence on K and a very small error. The residual and multiple regression methods are not dependent on K, but they become less reliable as K increases. This is possibly because complex mixtures in pixels, and especially pixels that vary from the image mean by a large amount, make it difficult to differentiate the variance due to noise from the variance due to signal mixtures.

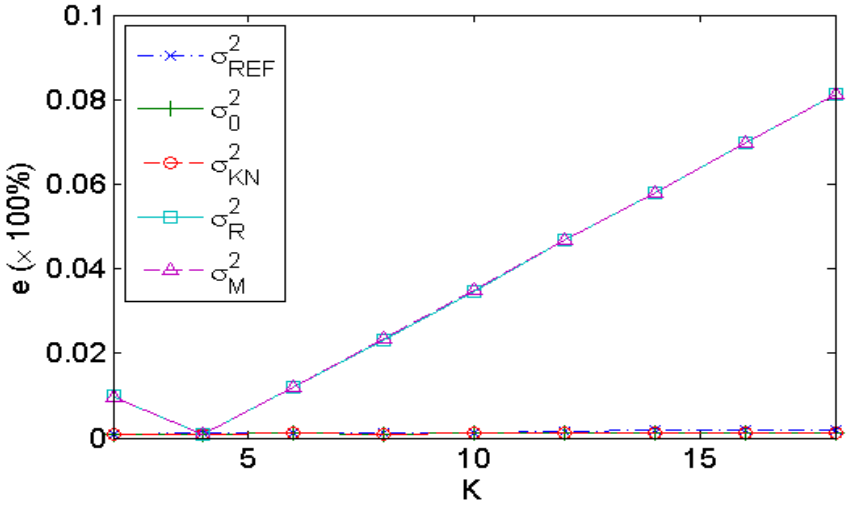


Figure 4.3.: The error of all noise variance approximations with respect to the number of endmembers. The Kritchman-Nadler (σ^2_{KN}), scaled (σ^2_0) and reference variances (σ^2_{REF}) perform similarly and are not dependent on K. However, the residual and multiple regression variances become less reliable as K increases.

Performance with respect to noise variance

The noise approximations should be able to reliably approximate the noise, no matter what their level (as long as the noise is not high enough to be confused with signal, but since K is assumed known this situation is not considered). In Figure 4.4, the i.i.d. approximations show a slight increase in errors as the noise levels become higher, especially the reference variance (σ^2_{REF}), but all three maintain very low error levels. The residual and multiple regression methods do not assume known K and so in this case large noise may be confused with small signal. However, these methods appear independent of the noise, even when it reaches high levels. As discussed previously, their lower performance may be due to fact that they require more pixels for success than the i.i.d. approximations.

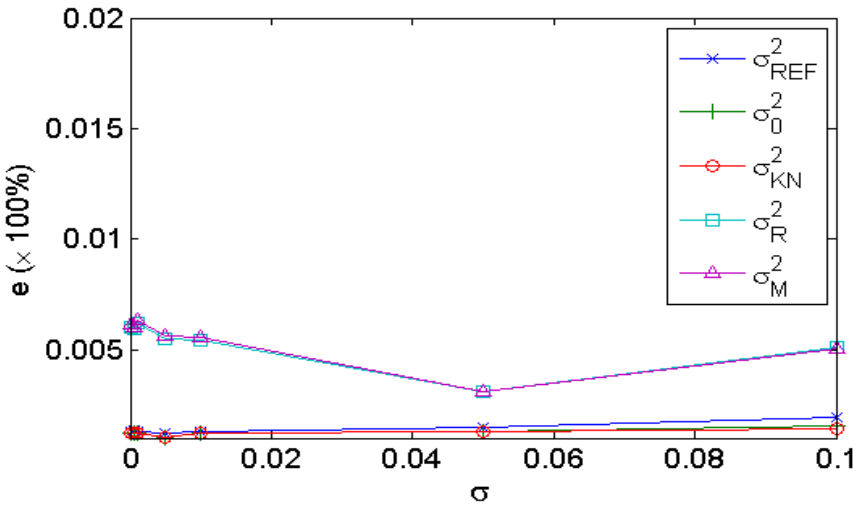


Figure 4.4.: The error of all noise variance approximations with respect to real variance. All five approximations appear largely independent of the standard deviation of the noise, although the i.i.d. approximations show a small increase in errors for high noise standard deviation, especially the reference variance (σ^2_{REF}). All five approximations maintain an error level of less than 1% for all noise values.

Performance with respect to variation in the noise across bands

It is known that hyperspectral images do not have i.i.d. noise in all bands. To test the dependence of the methods on the assumption of i.i.d. noise, noise is simulated that is close to the mean variance σ^2 in each band, where the range of the variation between bands may be adjusted. To simulate this, the standard deviation of the noise in each band is selected from a Gaussian distribution $\mathcal{N}(\sigma, \delta)$. The range considered is $0.1 \leq \gamma \leq 1$, with the upper limit chosen to reduce the likelihood of selecting negative variances, since $\gamma > 1$ results in $\delta > \sigma$. (In the event that a negative variance is selected, the variance is resampled from the distribution, forcing positivity.) In this simulation the approximation methods are expected to estimate the mean noise across all bands. Unfortunately this is not the case for any of the five approximations, as shown in Figure 4.5. Even for a small range of noise variances across bands, the accuracy of determining the mean noise is reduced. This means that it is unlikely that the mean noise variance will be applicable to real images. It is interesting that the residual and multiple regression variances, that have performed worst in all other tests, seem to perform slightly better than the i.i.d. approximations here. This is probably due to the fact that these approximations make no assumption of i.i.d. noise and mean noise variance is only used for comparison.

In all the other synthetic tests, the standard deviation around the mean squared error was small, $\sim 10^{-3}$. If the i.i.d. noise assumption is not met, not only does the mean accuracy decrease over the 20 experiments, but the standard deviation also becomes substantially larger. If in (4.5.1) the mean is replaced by the standard deviation, then these values are given in Table 4.1 for $\sigma/\delta = 0.1, 0.5, 1$.

Table 4.1.: The percentage standard deviation with respect to σ^2 for all noise approximation methods, in the presence of non-i.i.d. noise.

σ/δ	σ^2_{REF}	σ^2_0	σ^2_{KN}	σ^2_R	σ^2_M
0.1	0.0184	0.0186	0.0186	0.0214	0.0215
0.5	0.2025	0.2028	0.2028	0.1936	0.1937
1	0.5673	0.5677	0.5677	0.5493	0.5493

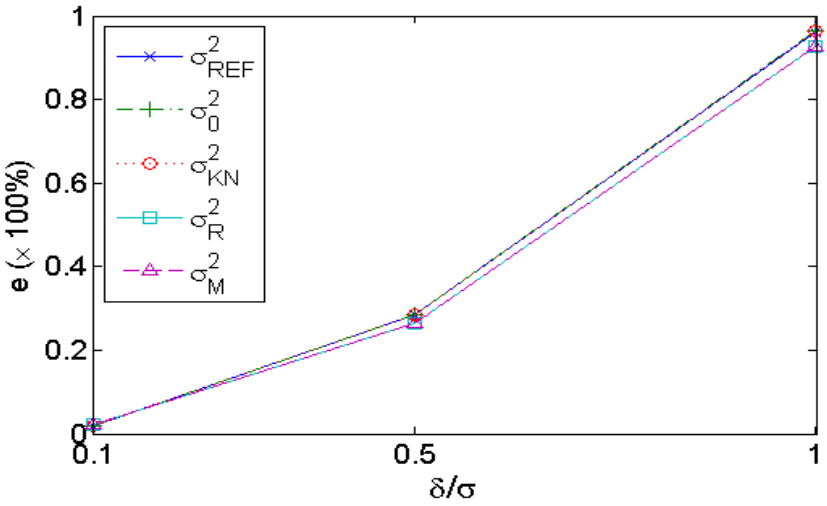


Figure 4.5.: The error of all noise variance approximations with respect to varying noise across bands. All five approximations tolerate only small variations between bands, and are unable to approximate the mean noise variance when the noise differs substantially across bands. If the standard deviation of the noise in each band is selected from $\mathcal{N}(\sigma, \delta)$, then consider the fraction δ/σ as a measure of the changing noise across bands. Even a 25% variation between bands results in errors of more than 10% for all approximations, and the errors worsen with higher variation.

Performance with respect to correlation between bands

In the derivation of the reference variance (σ^2_{REF}), the model covariance matrix, Σ, may be diagonalised as (Matrix of rank K) + $\sigma^2 I_p$ (see (4.1.1)). This is only true for random noise $\underline{\tilde{\xi}} \sim \mathcal{N}(0, I_p)$, so if correlations are introduced, the product $\sigma^2 \underline{\tilde{\xi}} \underline{\tilde{\xi}}^T$ will contain off-diagonal elements, instead of being diagonal and equal to $\sigma^2 I_p$ as in (4.1.1). This will increase the rank of the signal matrix that was previously rank K, resulting in inaccurate determination of the noise and of K. The scaled variance (σ^2_0) is derived from the reference variance (σ^2_{REF}) and so correlated noise bands will affect this approximation in the same way. Kritchman and Nadler [14] also use the assumption $\underline{\tilde{\xi}} \sim \mathcal{N}(0, I_p)$ in the derivation of σ^2_{KN}, so this approximation is also expected to be sensitive to correlated bands.

The residual and multiple regression variances both use the theory of multiple linear regression in their approximation. In this theory, signal correlation is assumed, but correlation between predictor variables (which is equivalent to noise correlation) is known as multicollinearity, and is considered part of the curse of dimensionality [77]. As the number of bands increases, it becomes more likely that some of these bands are correlated [77]. Clarke *et al.* [77] explains that when multicollinearity occurs, all observations are clustered, and the regression fit will be accurate near these observations, but the fit will become unreliable for points further away.

From the theory stated above, noise or pixel observations that are correlated between bands may adversely affect noise approximations. In real images, correlation often occurs and may be caused by sensor characteristics such as narrow spectral bands and spectral overlap, or by preprocessing procedures. A synthetic experiment is designed in order to test the sensitivity of the noise approximations to correlation. A certain number, p_{cor}, of correlated bands are chosen so that for $j = 2, 4, \ldots, 2 \times p_{\text{cor}}$, the noise is chosen from the multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & c\sigma_2^2 & 0 & \cdots & & & 0 \\ c\sigma_2^2 & \sigma_2^2 & 0 & & & & \\ 0 & & \ddots & & & & \vdots \\ & & 0 & \sigma^2_{(2\times p\text{cor}-1)} & c\sigma^2_{(2\times p\text{cor})} & 0 & \\ \vdots & & 0 & c\sigma^2_{(2\times p\text{cor})} & \sigma^2_{(2\times p\text{cor})} & 0 & \\ & & & & & \ddots & 0 \\ 0 & & \cdots & & & 0 & \sigma_p^2 \end{pmatrix}.$$

For the results illustrated in Figure 4.6, $c = 0.5$. In this particular example all the noise values in a correlated band become correlated with the values in the previous band. This is just one example; the same behaviour holds even when the bands considered are not neighbouring bands. Figure 4.6 shows that all noise variances become less reliable in the presence of correlated bands.

In particular, correlated bands result in a consistent underestimation of the noise when the noise variance is estimated using the i.i.d. or hybrid methods discussed above. If this noise approximation is used to separate signal from noise in order to determine the ID of a hyperspectral image, the ID may be overestimated.

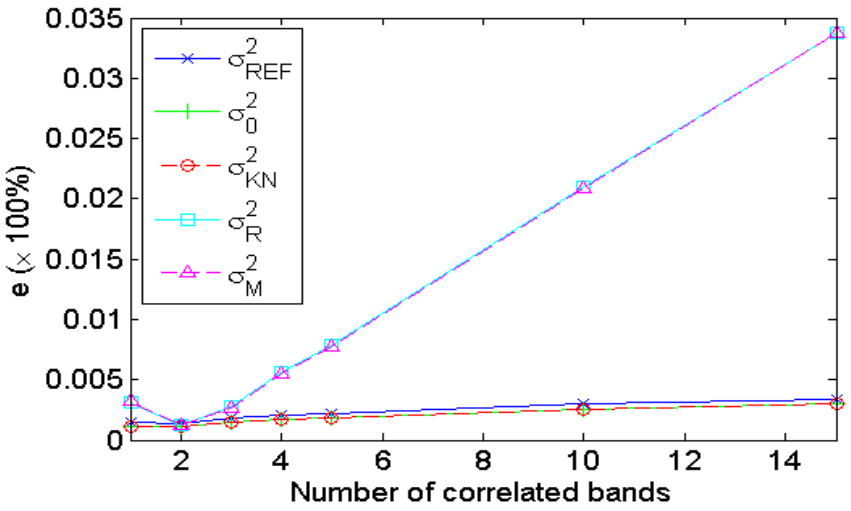


Figure 4.6.: The error of all noise variance approximations with respect to correlated bands within the noise. All five variances perform poorly when noise is correlated across bands. The worst approximations are given by the residual and multiple regression methods.

4.5.2. Experiments for the spatially based hybrid method

The last noise approximation method, namely Meer's method ($\bar{\sigma}^2_{Meer}$), is a spatially based method and so cannot be tested with the synthetic dataset. Recall that Meer's method divides the image into blocks and calculates the variance over each of these blocks. The four smallest variances are used to estimate the overall noise variance for the pyramid level that determines the block size. If all four variances are below a certain threshold, then the image is determined to contain no noise. Meer [73] sets this threshold at $\eta = 1$, since he is assuming that pixel values are between 0 and 255. The pixel values in the synthetic set are between 0 and 1, and so a lower threshold must be set. In the synthetic experiments, the case of no noise is not considered. For Meer's method, the effect of noise variance will be tested by adding Gaussian noise to a real image.

Since the synthetic set is not applicable, the tenth band of Cuprite (AVIRIS) is considered and Meer's method is used to approximate the noise. In the synthetic data the spectral values were bounded between zero and one, and so the real images are scaled so that they are comparable. The initial variance determined by Meer is σ^2. Next, add noise of known variance τ^2 to the image (I) so that each pixel x_i in I becomes $x_i + n_i$, where $n_i \in \mathcal{N}(0, \tau^2)$. The "known" total variance for the error calculation is taken to be $\sigma^2 + \tau^2$, and this is compared with the approximation that Meer gives for the new image. The experiment is repeated 50 times for each point on the graph, and Figure 4.7 shows errors relative to the ratio τ/σ. The errors are calculated using (4.5.1). The results show a slight upward trend, but errors remain below 10%. Since the original noise characteristics of the image are unknown, assuming that the initial noise approximation is good, then this test shows that Gaussian, spatially independent noise is reliably approximated. If the original noise estimate σ^2 is inaccurate, then the error in approximating τ is linearly related to the error in σ.

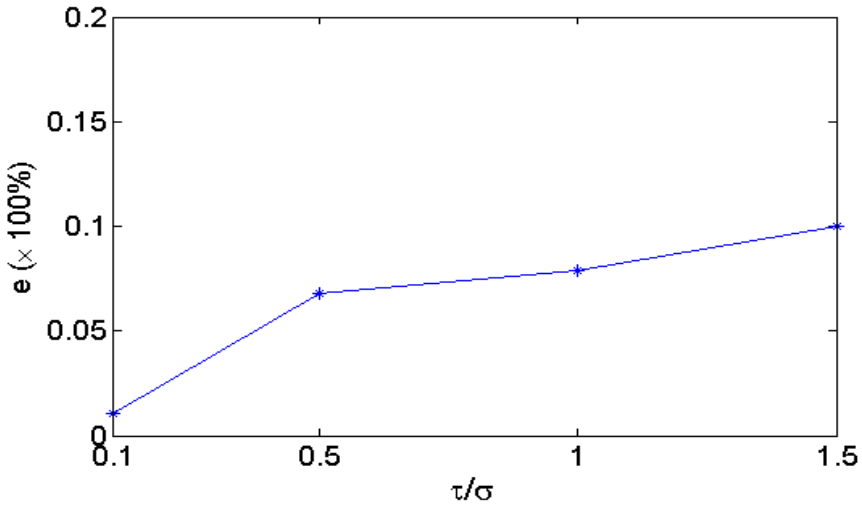


Figure 4.7.: The error of Meer's hybrid noise approximation. When known noise of variance τ^2 is added to the tenth band of normalised Cuprite (AVIRIS), Meer's method is used to determine the new variance, if the initial variance (no added noise) is assumed to be correct.

4.6. Performance of non-i.i.d. noise approximation

The assumption of i.i.d. noise may not be applicable to hyperspectral imagery [10], and Chapter 2 shows that some of the methods that determine the ID of a hyperspectral image require non-i.i.d. noise estimates. For instance, HySime depends on a noise approximation per pixel and HFC requires an estimate of the noise in each band. These methods are dependent on the accuracy of the noise approximation, and in this section we will test the reliability of three non-i.i.d. noise approximations, namely the residual method, multiple regression method, and Meer's method. The former two methods are statistical methods, whereas Meer's method is spatially based, and will be tested separately.

To test the residual and multiple regression methods, the synthetic dataset described in Section 3.1 will be used. An image is created with default parameters of N=10,000 pixels, p=200 bands, K=5 endmembers, and a noise standard deviation of $\sigma = 10^{-3}$.

The residual method estimates the variance per band, whereas the multiple regression method estimates the actual noise values per pixel. To illustrate their accuracy, the real and approximate noise (multiple regression) for a single pixel is shown in Figure 4.8 and the real and approximate noise variance (both methods) for all bands is shown in Figure

4.9.

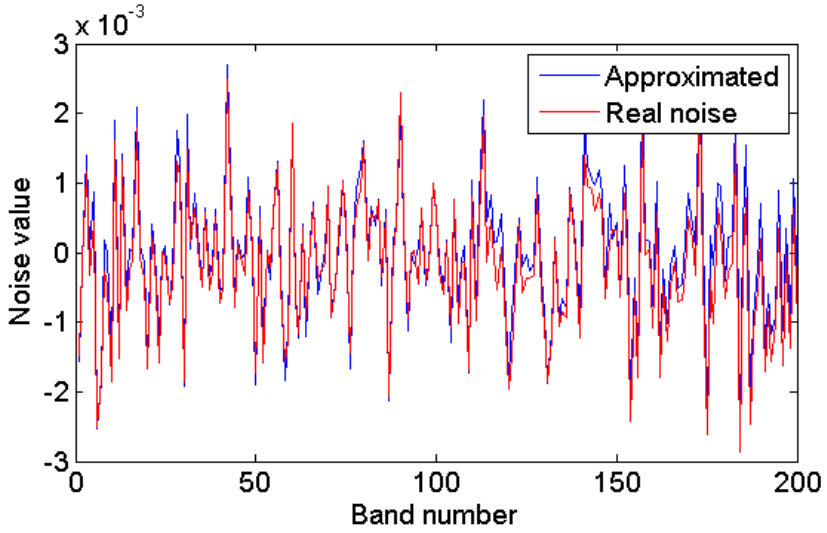
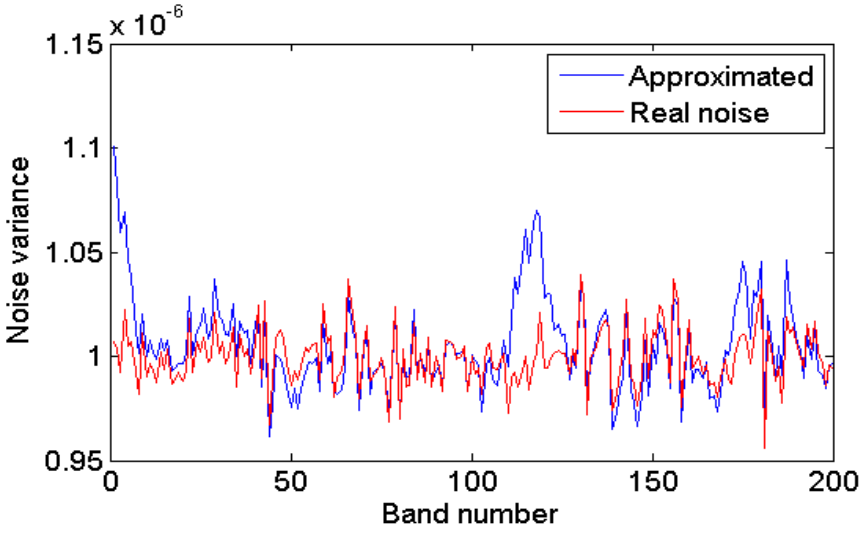


Figure 4.8.: A comparison between the simulated noise in a pixel (red), to the approximated noise in a pixel (blue) using the regression theory method for approximation. Note the good agreement produced by the method.

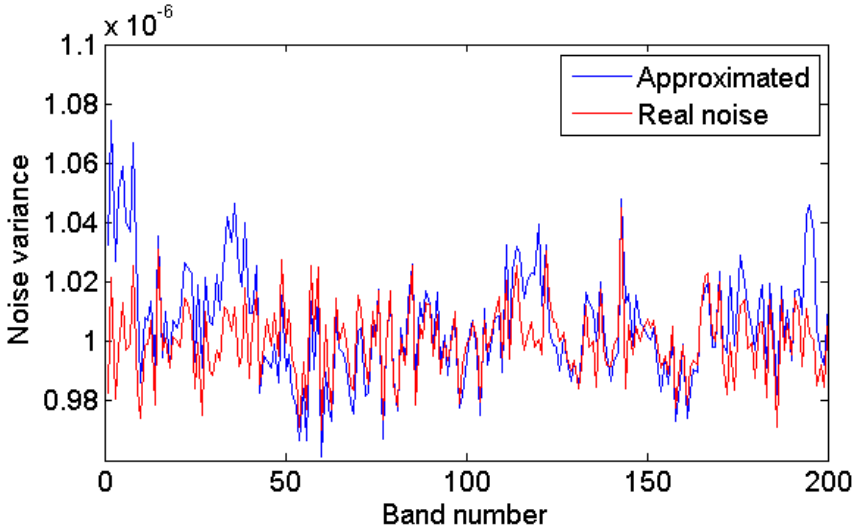
Figure 4.8 shows excellent agreement with the real noise for a single pixel across all bands. The variance for each band over the entire image shown in Figure 4.9 is less reliable for both methods, but there is still good agreement. In both Figures, the input noise variance for all bands is $\sigma^2 = 10^{-6}$ and the approximated variances are distributed around this value. This allows one to visually appreciate the accuracy of the approximation, and in the following section the effects of varying certain variables will be investigated.

4.6.1. Experiments for statistical methods

For each point on Figures 4.10 to 4.14, 20 synthetic images were created, with randomly chosen endmembers (only the number is fixed), abundances and noise per pixel (only the variance is fixed). In the above section, i.i.d. noise was examined and so the error was calculated with respect to noise over the image. Here non-i.i.d. noise is considered, and so the error is calculated for each band, over all experiments $i = 1, \ldots, 20$. The displayed error is the mean error per band.



(a)



(b)

Figure 4.9.: A comparison between the noise variance per band simulated in an image (red), to the approximated noise in the image (blue) using (a) the multiple regression method and (b) the residual-based method for approximation.

$$e = \text{mean}_{1 \leq j \leq p} \left(\frac{\sqrt{\frac{1}{20} \sum_{i=1}^{20} (\sigma_j^2 - \sigma_{ij(\text{approx})}^2)^2}}{\sigma_j^2} \right), \tag{4.6.1}$$

where σ_j^2 is the noise variance in the j^{th} band, and $\sigma_{ij(\text{approx})}^2$ is the noise in the j^{th} band approximated by the i^{th} experiment.

This error value will be used to measure the dependence of the two methods on the variables N, p, K, and σ^2. It is important to know the minimum image size, minimum number of bands and maximum noise level for these methods to be reliable. It is also important to approximate the noise independently of the signal contents of the scene.

The following sections evaluate the dependence of the methods on all the parameters described above.

Performance with respect to the number of pixels

In this test, the percentage root mean squared error over 20 experiments (Equation (4.6.1)) is shown for several image sizes. In all the synthetic comparisons, the multiple regression and residual methods perform very similarly. The i.i.d. tests done in Section 4.5 show that both methods performed well for more than 10,000 pixels. This corresponds to a very small 100×100 image. Figure 4.10 shows that the accuracy of both approximations is increased with higher number of pixels, but the error shown does appear to reach a plateau after about 6,000 pixels. This is smaller than the default value for N in the synthetic experiments, and so the tests in the following sections are valid in terms of N.

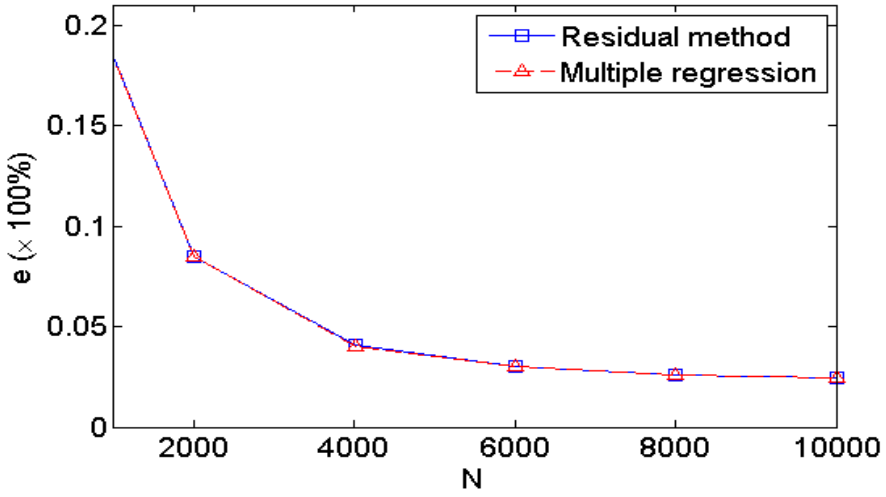


Figure 4.10.: The error of residual and multiple regression noise variance approximations relative to the number of pixels. The approximation error decreases for increased N, but reaches a plateau after about $N = 6,000$ pixels. After this, both methods produce low errors and appear almost independent of N.

Performance with respect to the number of bands

Figure 4.11 shows the error of the residual and multiple regression noise approximation methods when applied to synthetic images with differing numbers of bands, p. The accuracy is strongly dependent on p, with errors only dropping below 5% for $p > 100$. The residual and multiple regression variances both depend on optimising the fit of the approximation across the bands. More bands allow a closer fit. Hyperspectral images tend to contain more than 100 bands and so this limit is not considered a problem when applying these methods to real images.

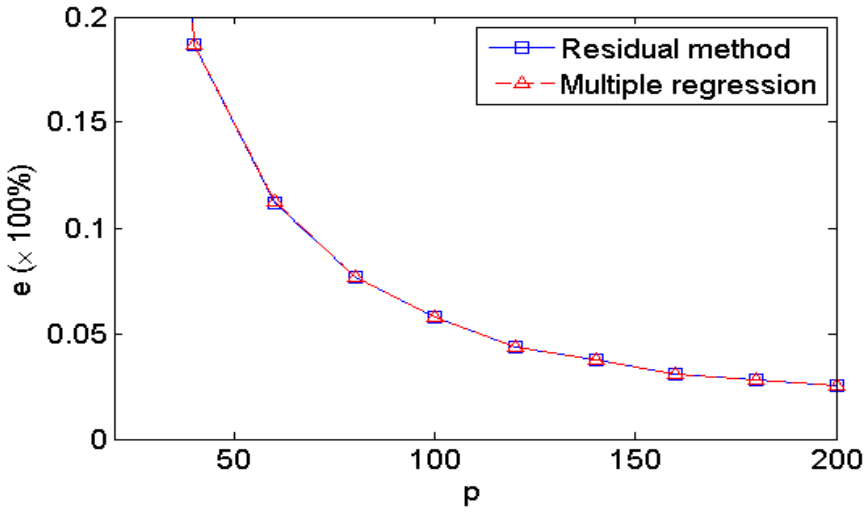


Figure 4.11.: The error of residual and multiple regression noise variance approximations relative to the number of spectral bands. Both methods perform similarly, and the errors decrease in relation to p. The errors only drop below 5% for $p > 100$.

Performance with respect to the number of endmembers

Figure 4.12 shows the errors in the residual and multiple regression noise approximation methods relative to the number of endmembers in a synthetic image. The errors in both variances are dependent on K, the number of endmembers. Where the image is more complex, it may be more difficult to determine the variance due to noise compared with the variance due to complex signal mixtures, which may result in the increase in errors for large K.

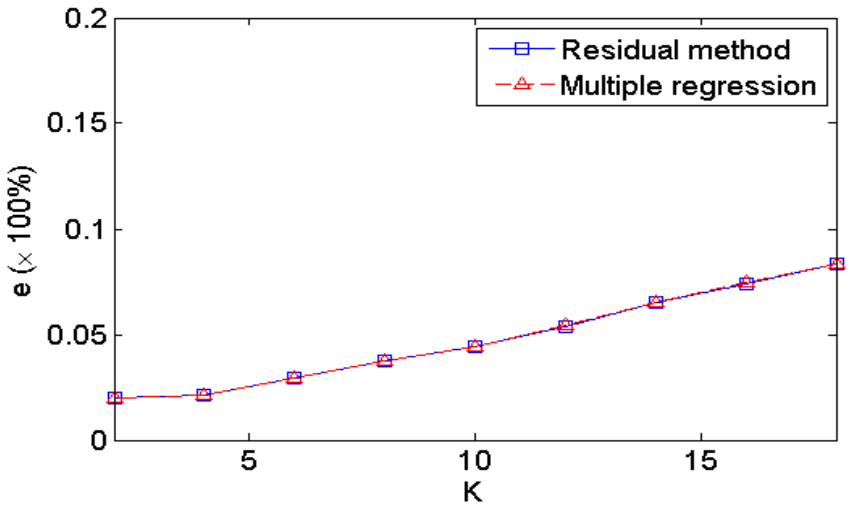


Figure 4.12.: The error of residual and multiple regression noise variance approximations relative to the number of endmembers. The accuracy decreases with increased K, for both methods. The noise estimates are nonetheless good approximations, with errors of less than 10% for all points displayed.

Performance with respect to noise variance

It is important that neither variance approximation is dependent on the noise level. The noise should be reliably estimated, no matter the true variance. Figure 4.13 shows the errors in the residual and multiple regression noise approximation methods, and the noise is reliably estimated by both methods, with no dependence on the noise variance, σ^2.

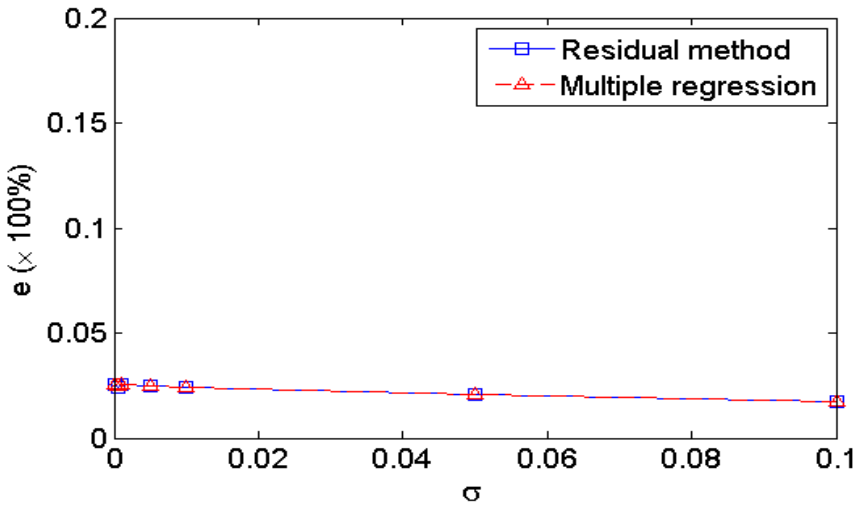


Figure 4.13.: The error of residual and multiple regression noise variance approximations relative to real noise variance. Both variances perform well, with errors of approximately 2%. Neither variance approximation is dependent on the level of the real noise variance.

Performance with respect to variation in noise across bands

To simulate noise variances that are close to the mean noise variance σ^2, but may vary from the mean in each band, noise is simulated where the standard deviation of the noise in each band is chosen from $\mathcal{N}(\sigma, \delta)$. The ratio $\gamma = \delta/\sigma$ is considered a measure of the changing noise across bands. Figure 4.14 shows the errors in the residual and multiple regression noise approximations relative to the percentage variation in noise across bands. Such variation is very well tolerated. Even when the variance of the noise distribution is larger than the mean noise variance, the errors are a fraction of a percent. Both the residual and multiple regression variances may be used for real images.

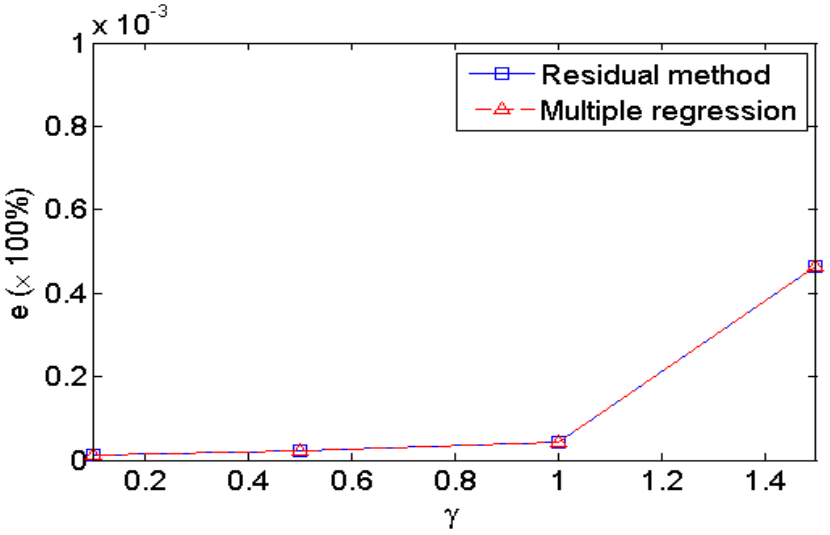


Figure 4.14.: The error of residual and multiple regression noise variance approximations relative to varying noise across bands. If the variance spread is defined by $\mathcal{N}(\sigma, \delta)$, then consider the fraction $\gamma = \delta/\sigma$ as a measure of the changing noise across bands. Small values for γ are well tolerated.

Performance with respect to correlation between bands

The residual and multiple regression variances both use the theory of multiple linear regression in their approximation. In this theory, correlation between signal is necessary, but correlation in the predictor variables (which is equivalent to correlation in the noise) is known as multicollinearity, and is considered part of the curse of dimensionality [77]. As the number of bands increases, it becomes more likely that some of these bands are correlated [77]. Clarke *et al.* [77] explains that when multicollinearity occurs, all observations are clustered, and the regression fit will be accurate near these observations, but the fit will become unreliable for points further away.

From the theory stated above, noise that is correlated between bands may adversely affect noise approximations. As in Section 4.5.1, a synthetic experiment is designed in order to test the sensitivity of the noise approximations to correlation. Figure 4.15 shows the error in the residual and multiple regression noise approximations, and all noise variances become less reliable in the presence of correlated bands.

It is interesting to note that the error is not limited to the noise variance of the correlated bands; the entire dataset becomes compromised. For the residual and multiple regression variance approximations it is necessary to remove the correlated bands before the noise computation in order to reliably approximate the noise variances in the uncorrelated bands. This is not true for all noise approximation methods. For example, Meer's spatially based method calculates the noise separately for each band, and so this method will not be compromised by the presence of correlated noise bands.

In particular, correlated bands result in a constant underestimation of the noise when the noise variance is estimated using the residual or multiple regression methods discussed above. If this noise approximation is used to separate signal from noise in order to determine the ID of a hyperspectral image, the ID may be overestimated.

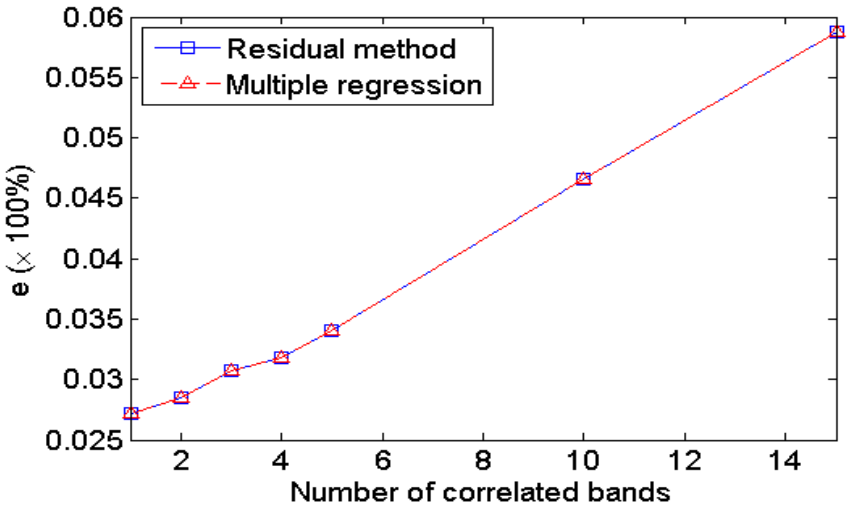


Figure 4.15.: The error of residual and multiple regression noise variance approximations with respect to correlated bands within the noise. Both methods deteriorate when noise is correlated across bands.

4.6.2. Experiments for the spatially based method

Meer's method is more difficult to test as it is, conversely to the statistical methods, a spatially based method. This method is not dependent on the number of pixels in the image, but rather on the size of homogeneous regions in the image. Since the method calculates the noise variance in each band separately, it is not dependent on the number of bands. Similarly, an endmember is defined by the unique spectral response over all bands, so that if there is only one band (as Meer's method considers), then the image is made up of real-valued pixels, rather than mixtures of endmembers. Similarly, i.i.d. or non-i.i.d. noise is defined by comparing the noise variance for all the bands. So if only one band is considered, the effects of i.i.d. or non-i.i.d. noise need not be tested. Therefore the only remaining parameters to test are the noise variance and the size of the homogeneous area.

To test the performance of Meer's method relative to the size of homogeneous regions and the noise variance in a single band, a 2-D greyscale image is simulated. The image is of size $(N \times N)$ with discrete points between 1 and 10 randomly distributed throughout the image, and a black block of size $(m \times m)$ occurring somewhere in the image (assume a square homogeneous region for simplicity). Gaussian noise of variance σ^2 is added to this image, so that the dependence of the method on m and σ^2 may be tested. The

default parameters are $N = 200$, $m = 20$, $\sigma = 1$ (remember that pixel values are discrete and between 0 and 10).

Since there is only one band, follow the procedure of Section 5.3 and define the error function as follows

$$e = \frac{\sqrt{\frac{1}{10}\sum_{i=1}^{10}(\sigma^2 - \sigma^2_{i\text{approx}})^2}}{\sigma^2}, \tag{4.6.2}$$

where σ^2 is the known noise variance added to the image, and $\sigma^2_{i\text{approx}}$ is the approximated noise variance from the i^{th} experiment. For each point displayed, the experiment is run 10 times.

The following sections evaluate the dependence of the methods on all the parameters described above.

Performance with respect to the size of homogeneous regions

Meer's method is not dependent on the number of pixels but rather on the size of the homogeneous regions within the image. It was previously stated that the synthetic dataset designed above was not appropriate for the testing of Meer's method. A new synthetic image is designed specifically for the testing of homogeneous regions.

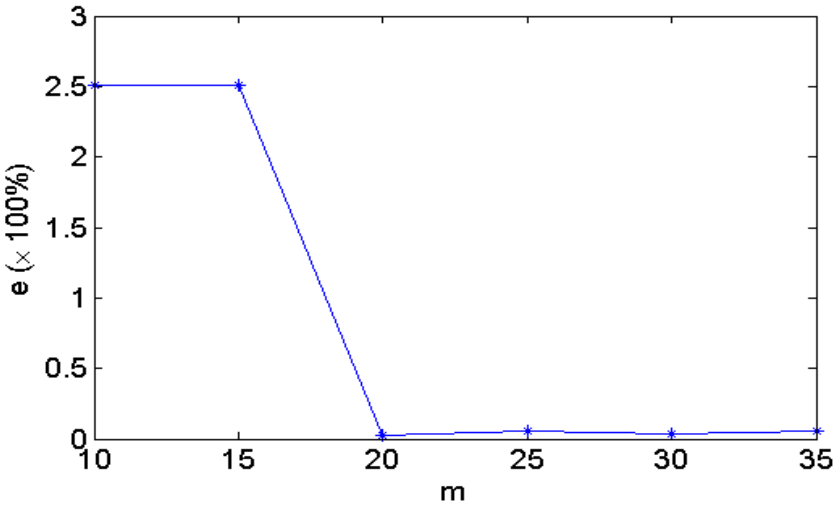


Figure 4.16.: The error of Meer's noise variance approximation with respect to homogeneous regions. The success of the method depends on the size of the homogeneous region present in the image. The graph reaches a plateau of approximately 2% error for a homogeneous region of size $(m \times m)$ where $m \geq 20$.

Figure 4.16 shows that the errors taper off at $m = 20$. This (20×20) block is extremely small when compared with the sizes of standard hyperspectral images.

The minimum error is approximately 2%, making this method the least reliable of the three examined, but Meer's method does have an advantage in that it does not assume linear or uncorrelated data. This method may be least reliable overall, but it may be more robust under certain circumstances, and in this case a 2% error is considered acceptable.

Performance with respect to noise variance

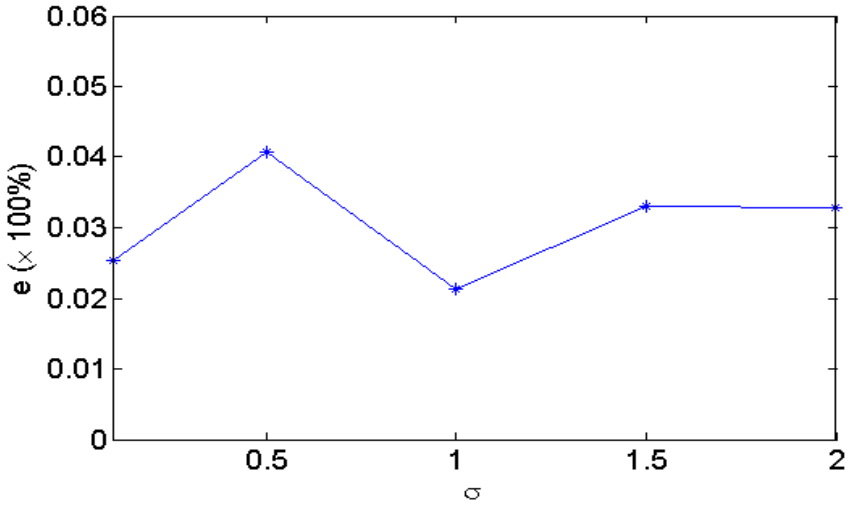


Figure 4.17.: The error of Meer's noise variance approximation with respect to real noise variance. Meer's method does not show any trend in accuracy with regard to noise variance. The variance should be understood in terms of the pixel values, which are discrete and range from 0 to 10.

Meer's method does not display a noticeable trend in errors with regard to noise level (Figure 4.17). This is good and follows the trends of the other two methods analysed. Meer's method once again shows lower accuracy than the other methods, with errors of approximately 3% compared to the errors of approximately 2% of the residual and multiple regression noise approximations. All three methods do however produce good results.

4.7. Results on Cuprite and HCI

Cuprite (AVIRIS) and HCI have been chosen since approximate values for the number of endmembers are known. The images are scaled so that the values fall between zero and one, in order that the variances may be compared to the synthetic experiments above. Wu *et al.* [10] have determined that the intrinsic dimension of Cuprite (AVIRIS) is between 22 and 28, and experts consulted about HCI estimate the intrinsic dimension to be between 3 and 7.

4.7.1. Results with i.i.d. noise approximations

Since all noise approximation methods have performed well in the synthetic experiments above, noise will be estimated in two real images. For all the i.i.d. noise approximations K needs to be known, so graphs are shown with respect to a range of values for K. Figure 4.18 shows the i.i.d. noise approximations for Cuprite for $22 \leq K \leq 28$. As expected, the noise approximation increases as K decreases. The i.i.d. noise approximation for HCI displays the same behaviour, as seen in Figure 4.19. For this image the range $3 \leq K \leq 7$ is considered.

In both these images, all three i.i.d. noise approximation methods (the reference, scaled and Kritchman-Nadler variances) produced the same result up to the nearest integer. This is the value displayed. The method developed in [14] uses the relationship between noise and ID estimations, and this will be discussed further in Section 5.3.

Table 4.2.: The noise variances of normalised Cuprite and HCI. For Cuprite (AVIRIS), the methods that don't assume i.i.d. noise (Meer, Residual and Regression) approximate a much higher mean noise variance than the i.i.d. methods. In HCI, Meer's method is once again much higher than the i.i.d. methods, but the Residual and Regression methods approximate an extremely low noise variance.

Noise variance	Meer	Residual	Regression	i.i.d. Methods
Cuprite ($* \times 10^{-5}$)	2.89	3.44	3.45	0.29 – 0.40
HCI ($* \times 10^{-4}$)	2.73	0.01	0.01	0.06 – 0.30

Table 4.2 shows the large variation between methods when estimating the noise in a real image. While the i.i.d. noise approximation methods performed best in the simulated tests, they had an advantage in that i.i.d. noise is an assumption of these methods that

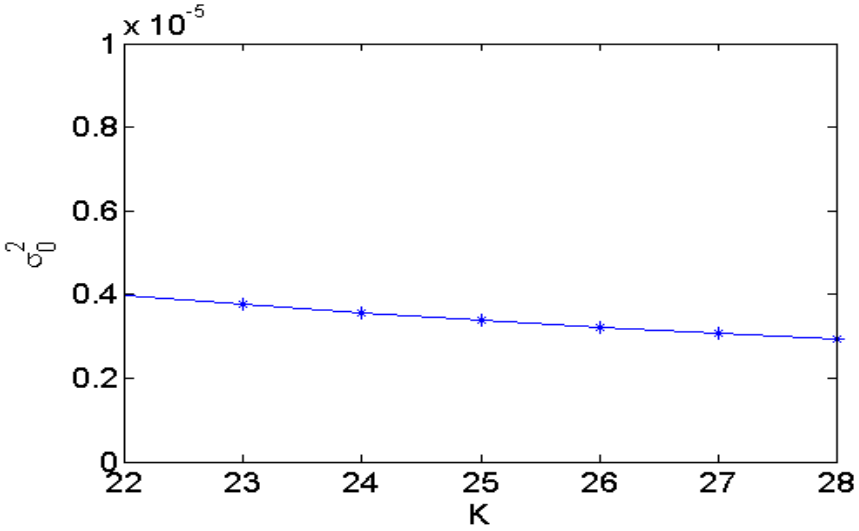


Figure 4.18.: The noise variance is approximated by σ^2_{REF}, σ^2_0 and σ^2_{KN} for normalised Cuprite (AVIRIS). All three methods produced the same result as σ^2_0 when rounded to the nearest integer, which is displayed above. The noise decreases with K since more variance is accounted for by signal with more endmembers in the image.

was true in these tests. The other methods did not assume anything about the noise across bands. In the synthetic tests, the latter methods performed better in the presence of non-i.i.d. noise and so it is likely that the higher noise estimations shown in the hybrid variances in Table 4.2 are better approximations.

The experiments above show that correlated bands may result in an underestimation of noise. Figure 4.20 shows that HCI contains higher partial correlation coefficients than Cuprite, and this would explain why Meer's approximation differs from the other hybrid approximations more noticeably in HCI than in Cuprite. Meer's method, unlike the others, is not affected by correlated bands as it considers each band separately. Table 4.3 shows that the noise variance approximations increase when some of the correlation is removed, however in the case of HCI the correlation between bands was so strong that when correlated bands are removed, very few bands are retained. The simulated experiments show that the residual and multiple regression approximations are unreliable for small p, and so for HCI, Meer's variance is the only value that may be considered reasonable.

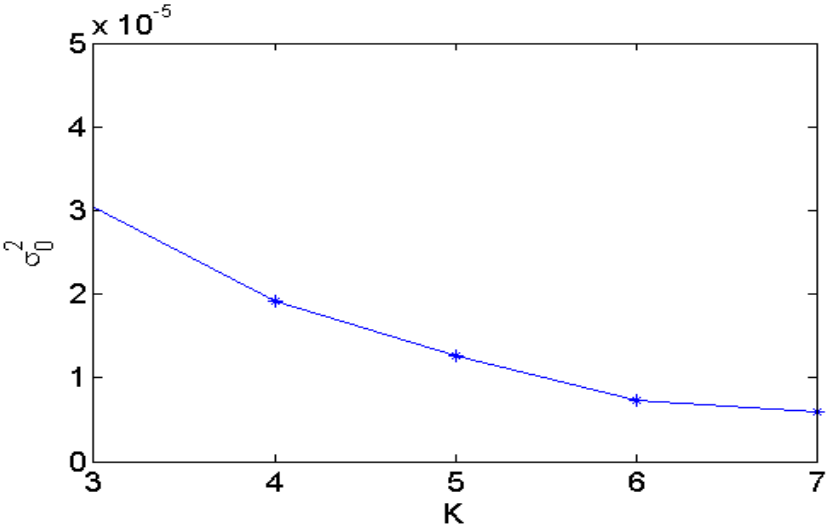


Figure 4.19.: The noise variance is approximated by σ^2_{REF}, σ^2_0 and σ^2_{KN} for normalised HCI. All three methods produced the same result as σ^2_0 when rounded to the nearest integer, which is displayed above. The noise decreases with K since more variance is accounted for by signal with more endmembers in the image.

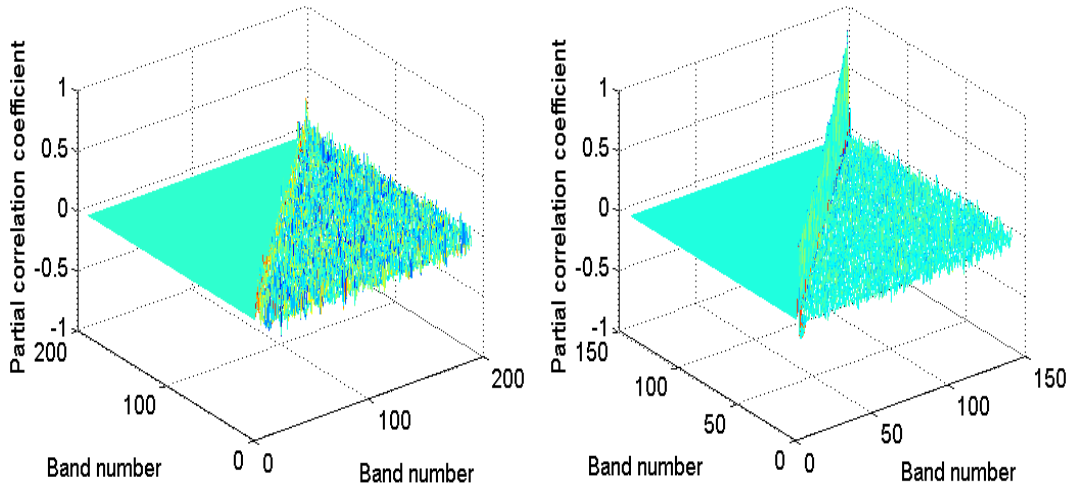


Figure 4.20.: The upper triangular matrices of partial correlation coefficients (original matrices are symmetric). The correlation of Cuprite (left) is much lower than that of HCI (right). The correlation in HCI occurs mainly between neighbouring bands.

Table 4.3.: The noise variances of normalised HCI, with correlated bands removed. Consider only every fourth band of HCI (35 bands), then the correlation decreases and the noise estimations all increase, except for Meer's method, which is not dependent on correlation.

Noise variance	Meer	Residual	Regression	i.i.d. Methods
HCI (subset) ($* \times 10^{-4}$)	2.73	0.07	0.07	0.07 – 0.32

4.7.2. Results with non-i.i.d. noise approximations

When noise is non-i.i.d. and correlated, i.i.d. methods have been shown to perform poorly, and these properties often occur in real images. In this section the non-i.i.d. noise approximation methods will be applied to real images, namely Cuprite (AVIRIS), shown in Figure 4.21 and HCI, shown in Figure 4.22. In order to compare the variances to the values tested in the synthetic section, the images are normalised so that all pixel values are between zero and one.

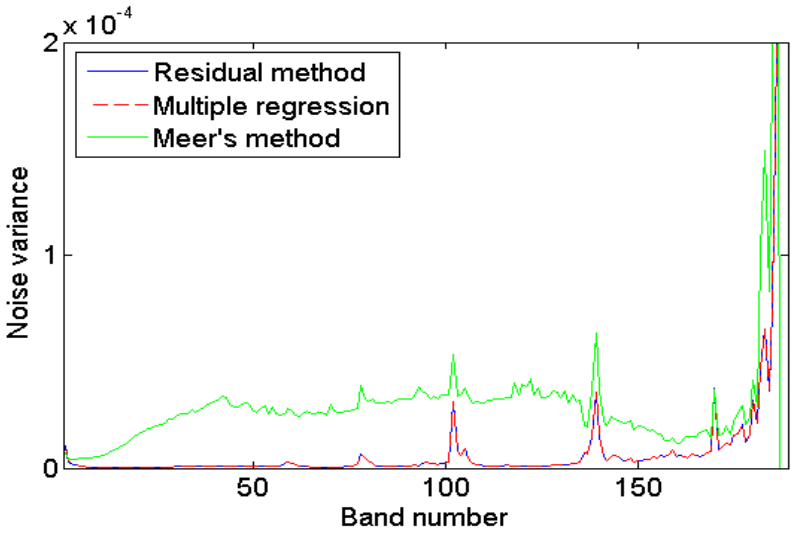


Figure 4.21.: The noise variances of normalised Cuprite. The regression and residual based methods produce similar results, which show lower noise variance than Meer's method. All three methods produce a similar shape of the noise variance per band.

Figure 4.21 shows that the noise variance in Cuprite appears to be consistently ap-

Table 4.4.: The amount of time in seconds that it took each method to calculate the noise in Cuprite. The residual method took only a fraction of a second, and Meer was the slowest method at approximately 6 minutes.

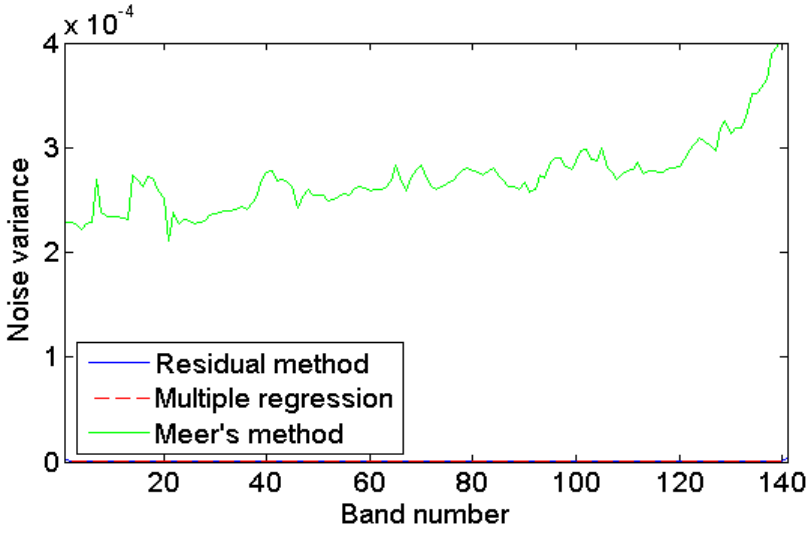
Meer	Multiple Regression	Residual method
362.94	26.10	0.01

proximated by all methods in terms of general shape. Meer's method shows the highest noise variance, and this is seen to a greater extent in HCI (Figure 4.22). Figure 4.20 shows that HCI contains higher levels of correlation, which may explain the low noise estimates determined by the multiple regression and residual methods, since these methods are sensitive to correlation. Recall that Meer's method is not sensitive to correlated noise since it estimates the noise on a band-by-band basis. The noise estimates of all three methods correspond to the behaviour of the i.i.d. estimates seen in Table 4.2 for HCI. It is interesting that the presence of correlated noise seems to have more impact than the presence of non-i.i.d. noise.

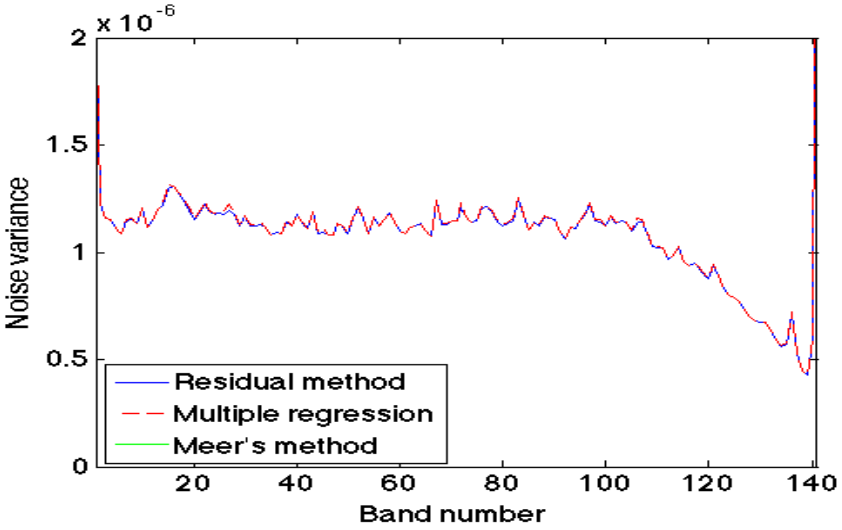
There are only minimal differences between the multiple regression and residual based methods, despite the difference in computational complexity.

The time taken to compute the noise is an important consideration when choosing an approximation method. To evaluate this, all three noise approximation methods have been tested on Cuprite, on an HP notebook with Intel Core2Duo 1.83 GHz CPU and 2GB RAM. The results are displayed in seconds in Table 4.4. Meer's method is by far the most time-consuming method, followed by the multiple regression method, and the fastest method is the residual method, at a fraction of a second.

In all the tests that have been done, the residual and multiple regression methods have produced almost indistinguishable results. The difference between them is that the multiple regression method calculates the noise present in each pixel, whereas the residual method calculates only the noise variance in each band. The effect of spatially dependent noise is tested by creating a synthetic image with noise variance $\sigma_1^2 = 10^{-6}$ in the first half of the image and noise variance $\sigma_2^2 = 10^{-8}$ in the second half of the image. Both methods determined the same variance per band, with mean variance $\sigma^2 = 5 \times 10^{-7}$. This scenario is unlikely to occur in real hyperspectral images, but the experiment is designed to compare the residual and regression noise estimation methods under the condition of spatially non-constant noise. To investigate this idea further, recent research [74, 75, 76] has been done into noise estimation where the noise contains a



(a)



(b)

Figure 4.22.: The noise variance in HCI is shown. The regression and residual based methods are barely visible in graph (a) due to their low amplitude. Graph (b) shows the variance determined by the regression and residual methods in more detail. Meer's method shows a more realistic noise variance.

Poisson component. This scenario will require different noise estimation techniques, and is discussed in the Open Questions section of this thesis. The comparison between the residual and multiple regression estimates shows that if image statistics are considered sufficient, then the two methods are no different in terms of results. The residual method has a much lower computational complexity and so is preferable in this case. If noise statistics are required for each pixel however, then the multiple regression method must be used.

As discussed above, the residual and multiple regression approximations are underestimated in HCI due to the high levels of correlation found in this image.

Next, the effect of non-linearities is tested. A synthetic dataset is simulated with three endmembers. Two of these are real spectra, and the third is the dot product of the first two. In this simple test, the noise variance displays low errors, comparable with those of linear data. This is surprising, as the methods are both based on linear mixing assumptions, but by simply identifying the non-linear mixture as a third endmember, the residual and multiple regression methods return good results.

4.8. Discussion

In the presence of real i.i.d. noise, the methods σ^2_{REF}, σ^2_0 and σ^2_{KN} performed better than taking the mean noise variance from the methods that do not assume i.i.d. noise. The approximation σ^2_{REF} performed the worst of the three methods, but all three had acceptable error levels for almost all the tests. The exception to this is when the noise varied in different bands. Even a small variation reduced accuracy, making these methods unsuitable for application to real images. The non-i.i.d. methods handled varying noise across bands with high accuracy. Of these methods, the multiple regression and residual methods performed similarly, while Meer's method performed the worst. However, Meer's method has the advantage in that it estimates noise one band at a time so that the results are independent of any correlation between spectral bands. This explains why in HCI, which contains correlation, Meer's method is the only method that produced satisfactory results. Because of this discrepancy, all three methods should be considered as noise approximations to determine the ID.

5. Applying Kritchman and Nadler RMT method to Hyperspectral Images

The use of Random Matrix Theory (RMT) in chemical applications has been discussed in Section 2.1.8. In the application of Kritchman and Nadler [14], multiple samples are taken from the same linear mixture of chemicals, and so if enough samples are collected, the mean observed measurement will tend to the value in a model that describes the expected measurement. This chapter will discuss the application of this method to hyperspectral imagery, where the samples are pixels in an image that may vary significantly from the mean. The method is applied as described in [14] where the noise variance is assumed constant in every band (i.i.d.) and the noise is uncorrelated.

5.1. Method validation

Kritchman and Nadler [14] show that signal and noise eigenvalues are separable in the case of i.i.d. uncorrelated noise, and they use this to derive a technique to determine the ID of their chemical mixture. This method may be applied to hyperspectral imagery by defining a model for each pixel in a hyperspectral image as follows (based on equation (1.1.1)):

$$\underline{\tilde{x}} = \sum_{j=1}^{K} \tilde{u}_j \underline{v}_j + \sigma \underline{\tilde{\xi}}, \tag{5.1.1}$$

where $\underline{\tilde{x}}$ is the $(p \times 1)$ vector of values for any pixel (for all p bands), $\tilde{u}_j$ a random variable representing the proportion of endmember $\underline{v}_j \in \mathbb{R}^p$ in the mixed pixel, $\underline{\tilde{\xi}} \in \mathbb{R}^p$ is chosen from a multivariate normal distribution $\mathcal{N}(0, I_p)$ and σ is a constant weighting factor for the noise. Here the abundances in $\underline{\tilde{u}}$ are bound by positivity and sum-to-one constraints.

In matrix form, (5.1.1) becomes

$$\underline{\tilde{x}} = V \underline{\tilde{u}} + \sigma \underline{\tilde{\xi}}, \tag{5.1.2}$$

where $V = [\underline{v}_1, ..., \underline{v}_K]$ and $\underline{\tilde{u}} = [\tilde{u_1}, ..., \tilde{u_K}]^T$.

Kritchman and Nadler [14] make the following assumptions: the endmembers are linearly independent, the noise is zero-centered, Gaussian, i.i.d. and uncorrelated, and $\mathbb{E}(\underline{\tilde{u}}\underline{\tilde{u}}^T)$ has full rank. Then the eigenvalues of the model covariance matrix may be partitioned into signal and noise eigenvalues [14]. If K is the intrinsic dimension, the i^{th} model covariance eigenvalue will be $\lambda_i + \sigma^2$ for $i = 1, \ldots, K$, and the remaining eigenvalues will be σ^2, where λ_i are the signal eigenvalues. Given the model in (5.1.2), where the noise is uncorrelated with the signal, then signal and noise eigenvalues are separable, assuming mean-centered data throughout (i.e. the mean pixel value over the image is $\underline{0}$). This is shown by defining the model covariance matrix by Σ (for centered data), then

$$\begin{aligned}
\Sigma &= \mathbb{E}[\underline{\tilde{x}}\underline{\tilde{x}}^T] && (5.1.3)\\
&= \mathbb{E}[(V\underline{\tilde{u}})(V\underline{\tilde{u}})^T + \sigma^2\underline{\tilde{\xi}}\underline{\tilde{\xi}}^T + \sigma(V\underline{\tilde{u}})\underline{\tilde{\xi}}^T + \sigma\underline{\tilde{\xi}}(V\underline{\tilde{u}})^T] && (5.1.4)\\
&= \mathbb{E}[(V\underline{\tilde{u}})(V\underline{\tilde{u}})^T] + \sigma^2 I_p && (5.1.5)\\
&= V\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]V^T + \sigma^2 I_p, && (5.1.6)
\end{aligned}$$

since $\underline{\tilde{\xi}} \in N(0, I_p)$ and the other products are zero since the noise is uncorrelated to the signal.

Then, V^T is $(K \times p)$ with $K < p$, and, since the columns of V are linearly independent,

$$\begin{aligned}
& V^T : \mathbb{R}^p \text{ onto } \mathbb{R}^K\\
\Rightarrow\ & \mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]V^T : \mathbb{R}^p \text{ onto } \mathbb{R}^K\\
& \text{since } \mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T] \text{ has full rank}\\
\Rightarrow\ & V\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]V^T : \mathbb{R}^p \text{ onto } \mathbb{R}^K \subset \mathbb{R}^p\\
& \text{since columns of V are linearly independent.}
\end{aligned}$$

Therefore the rank of $V\mathbb{E}[\underline{\tilde{u}}\underline{\tilde{u}}^T]V^T$ is K and so there are exactly $(p - K)$ eigenvalues of value 0.

Therefore,

$$W^T \Sigma W = W^T V \mathbb{E}[\tilde{\underline{u}}\tilde{\underline{u}}^T] V^T W + \sigma^2 I_p \tag{5.1.7}$$

$$= \begin{pmatrix} \lambda_1 & 0 & \dots & & 0 \\ 0 & \ddots & & & \\ \vdots & & \lambda_K & & \vdots \\ & & & \ddots & \\ 0 & & \dots & & 0 \end{pmatrix} + \sigma^2 I_p, \tag{5.1.8}$$

where W diagonalises Σ, giving the eigenvalues in decreasing order. Therefore the signal and noise eigenvalues of the model are separable.

In summary, if the endmembers are linearly independent, the noise is zero-centered, Gaussian, i.i.d. and uncorrelated, and $\mathbb{E}(\tilde{\underline{u}}\tilde{\underline{u}}^T)$ has full rank, then the signal and noise eigenvalues of the model covariance matrix, Σ, are separable so that the largest K eigenvalues are equal to $\lambda_i + \sigma^2$, and the remaining eigenvalues are equal to σ^2. This separation will enable determination of the ID of the model.

This separation must also apply to the observation covariance matrix (derived from pixel values). Kritchman and Nadler [14] show that the observation covariance matrix tends to the model covariance in the case of repeated sampling. Note that in [14], repeated samples are taken of the same mixture, whereas in the hyperspectral imagery case pixels are sampled from the ground and may vary substantially. In (5.1.2), the abundances for each pixel i, $\underline{u}_i$ will form a tight distribution around the model abundance $\tilde{\underline{u}}$ in the repeated sampling experiments, whereas the abundances in a hyperspectral image will form a much larger distribution. To show that the convergence still holds in the hyperspectral case, let $\underline{x}_i$ be a sample chosen from $\tilde{\underline{x}}$ described in (5.1.2), then

$$\underline{x}_i = V\underline{u}_i + \sigma\underline{\xi}_i. \tag{5.1.9}$$

Define the observation covariance matrix $S(N)$ (for centered data) by

$$\begin{aligned}
S(N) &= \frac{1}{N}\sum_{i=1}^{N} \underline{x}_i \underline{x}_i{}^T \\
&= \frac{1}{N}\sum_{i=1}^{N} [(V\underline{u}_i + \sigma\underline{\xi}_i)(V\underline{u}_i + \sigma\underline{\xi}_i)^T] \\
&= \frac{1}{N}\sum_{i=1}^{N} [V\underline{u}_i\underline{u}_i{}^T V^T + \sigma V\underline{u}_i\underline{\xi}_i{}^T + \sigma\underline{\xi}_i\underline{u}_i{}^T V^T + \sigma^2\underline{\xi}_i\underline{\xi}_i{}^T] \\
&= V\left(\frac{1}{N}\sum_{i=1}^{N}\underline{u}_i\underline{u}_i{}^T\right)V^T + \sigma V\left(\frac{1}{N}\sum_{i=1}^{N}\underline{u}_i\underline{\xi}_i{}^T\right) \\
&\quad +\sigma\left(\frac{1}{N}\sum_{i=1}^{N}\underline{\xi}_i\underline{u}_i{}^T\right)V^T + \sigma^2\left(\frac{1}{N}\sum_{i=1}^{N}\underline{\xi}_i\underline{\xi}_i{}^T\right) \\
&\to V\mathbb{E}(\tilde{\underline{u}}\tilde{\underline{u}}^T)V^T + 0 + 0 + \sigma^2 I_p \text{ as } N \to \infty \\
&\quad \text{with convergence in probability measure} \\
&\quad \text{since } V\left(\frac{1}{N}\sum_{i=1}^{N}\underline{u}_i\underline{\xi}_i{}^T\right) \to \mathbb{E}(V\tilde{\underline{u}}\tilde{\underline{\xi}}^T) = 0 \\
&\quad \text{since the noise is uncorrelated with the signal.} \\
&= \Sigma
\end{aligned}$$

by 5.1.6. Therefore the observed sample covariance eigenvalues tend to the model eigenvalues. This means that the signal and noise eigenvalues are also separable when observed data are used.

In summary, if the endmembers are linearly independent, the noise is zero-centered, Gaussian, i.i.d. and uncorrelated, and $1/N\sum_{i=1}^{N}(\underline{u}_i\underline{u}_i^T)$ has full rank in the limit $N \to \infty$, then the signal and noise eigenvalues of the observation covariance matrix, $S(N)$, are separable so that the largest K eigenvalues tend to $\lambda_i + \sigma^2$, and the remaining eigenvalues tend to σ^2.

Once the ID of the image has been calculated, this number may be used to unmix the image. This could be used to determine that an image contains, for example, 40% grass, 25% soil and 35% rock. The difference between this case and the case described by Kritchman and Nadler [14], is that the latter considered repeated samples of the same mixture, whereas the hyperspectral case considers pixels that may differ substantially from the mean pixel value.

As shown in Section 2.1.8, the eigenvalues of the observation covariance matrix $S(N)$

are arranged in descending order. An eigenvalue λ_i of S(N) is considered as a noise eigenvalue if

$$\lambda_i \leq \sigma^2(\mu_{N,p} + s(\alpha)\sigma_{N,p}), \tag{5.1.10}$$

and then the definition of intrinsic dimension (Definition 1) may be translated in terms of K, where K is defined as the largest number for which $\lambda_i > \sigma^2(\mu_{N,p}+s(\alpha)\sigma_{N,p})\forall i, 1 \leq i \leq K$. The number of pixels is represented by N, σ^2 is the variance of the Gaussian noise, α is a significance level and $s(\alpha)$ may be found by inverting the Tracy-Widom distribution (in [14], $\alpha = 0.5\%$, and this value is fixed for all images investigated). Since the data are real valued, define

$$\mu_{N,p} = \frac{1}{N}\left(\sqrt{N-\frac{1}{2}}+\sqrt{p-\frac{1}{2}}\right)^2 \tag{5.1.11}$$

$$\sigma_{N,p} = \frac{1}{N}\left(\sqrt{N-\frac{1}{2}}+\sqrt{p-\frac{1}{2}}\right)\times \left(\frac{1}{\sqrt{N-\frac{1}{2}}}+\frac{1}{\sqrt{p-\frac{1}{2}}}\right)^{1/3} \tag{5.1.12}$$

In order to calculate the ID, the noise variance, σ^2 must be approximated. Hence the RMT method for determining the intrinsic dimension requires a method for noise approximation. Any method may be used, and some such methods are discussed in Chapter 4.

5.2. Analysing the RMT threshold

Before applying the RMT method to hyperspectral images, the relevance of the Random Matrix Theory threshold is numerically analysed by testing the accuracy of the assumption that the largest eigenvalue of a random matrix fulfills the following condition:

$$\lambda_1 < \sigma^2(\mu_{N,p-K} + s(\alpha)\sigma_{N,p-K}), \tag{5.2.1}$$

where λ_1 is the first (largest) eigenvalue of a random matrix as described in Johnstone [60], σ^2 is the noise variance, N is the number of pixels, p is the number of bands, K is the intrinsic dimension and α is the significance level.

A $(p \times N)$ random matrix $\tilde{X}$ is generated, where each column is drawn from $\mathcal{N}(0, \sigma^2 I_p)$. Then $A = \frac{1}{N}\tilde{X}\tilde{X}^T$ can be considered an observation covariance matrix for a pure noise image. This is a Wishart matrix by construction.

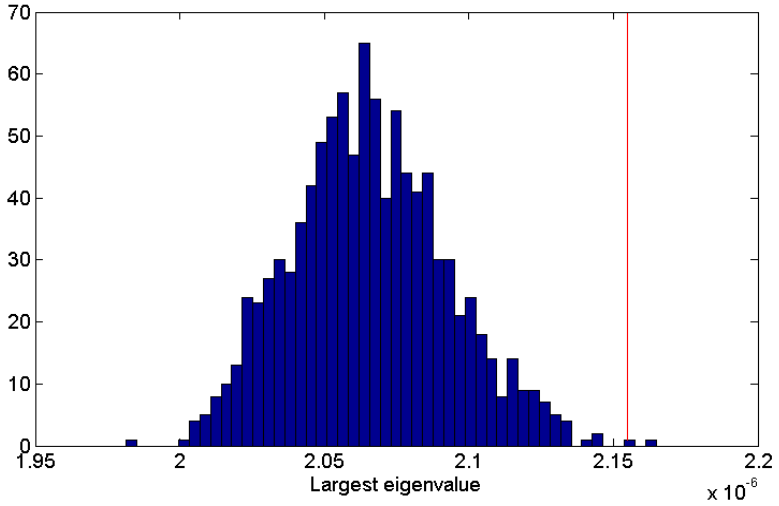


Figure 5.1.: The histogram of largest eigenvalues from Wishart matrices. One thousand Wishart matrices were created, which are equivalent to pure noise images with 1,000 pixels, 200 bands, and a standard deviation of 10^{-3}. The threshold is shown by a red line, and it may be seen that 99.8% of the simulated images had largest eigenvalues that fulfilled the RMT evaluation described in (5.2.1).

In order to better understand the behaviour of λ_1, the largest eigenvalue of a Wishart matrix, with respect to the threshold, a thousand Wishart matrices have been randomly simulated ($N = 1,000$, $p = 200$, $\sigma = 10^{-3}$, $\alpha = 0.5\%$) and their largest eigenvalues are

plotted with the threshold. In this simulated scenario, σ is known for approximation of the threshold, but in a real context σ will have to be estimated. As seen in Figure 5.1, there was a 99.8% success rate when considering (5.2.1). Note that the condition will still be fulfilled if the threshold is overestimated (but non-noise eigenvalues may be identified as noise), and an underestimation of the threshold may result in noise eigenvalues being identified as signal. Therefore it is important that σ is approximated reliably.

Table 5.1.: The success rate of eigenvalues correctly identified as noise by the RMT threshold, with respect to all parameters. The RMT threshold proves to be very robust with respect to variation in the number of pixels, number of bands, and noise variance. High values for α result in a lower threshold, and so fewer eigenvalues are correctly recognised.

N	p	σ	α	Success (%)
1	200	10^{-3}	0.5	100
10				99.8
100				99.8
1000				99.7
10000				99.7
1000	1	10^{-3}	0.5	100
	10			99.9
	100			99.9
	800			99.7
1000	200	10^{-4}	0.5	99.8
		10^{-2}		99.8
		1		100
		100		99.9
1000	200	10^{-3}	0	100
			5	97.4
			10	91
			100	56.9

In order to determine the accuracy of the RMT threshold, N, p, σ and α are adjusted in order to analyse the effect of these variables on the success of the threshold. Once again, assume that σ is known in order to determine the threshold. For each success rate, 1,000 Wishart matrices were created with the given variable values. Table 5.1 shows the effect of several values for N, p, σ and α on the success of the RMT threshold. This threshold proved to be robust, with no apparent dependencies on the number of pixels, number of bands and noise variance. The overall accuracy for the variation in N, p and

σ was again 99.8%, as in figure 5.1.

To further evaluate the effect of signal on the other variables, we added a constant vector $\underline{c} \in \mathbb{R}^p$ to every pixel in the image, where $c = [1, \ldots, 1]$. This also gave good results — the largest eigenvalue no longer satisfied the threshold (0% success for 1,000 Wishart matrices), but the second largest eigenvalue behaved as described in Table 5.1 (99.8% success for 1,000 Wishart Matrices). The same was true for non-constant signal, and multiple signals, showing that experimentally the largest noise eigenvalue behaves like the largest eigenvalue of a Random Matrix.

From these experiments, the RMT threshold appears reliable, and independent of N, p and σ. The significance level, α is the only variable to effect the accuracy of the RMT threshold. Higher values for α result in a lower RMT threshold, and so fewer eigenvalues meet the RMT condition. A lower value for α is therefore preferable, but if a signal is added to the image (in this case, add a constant signal of value 1 to the entire image), then $\alpha = 0$ results in all eigenvalues meeting the RMT condition, and therefore signal eigenvalues are incorrectly identified as noise. Kritchman and Nadler [14] use $\alpha = 0.5$, which yields a success rate of 99.7% when there is no signal, and no signal eigenvalues are recognised as noise when a single signal is included in the image. From these experiments this is indeed a reasonable value for α and everywhere in this thesis it will be assumed that $\alpha = 0.5\%$. It should be noted that α is not dependent on the image and should not be considered a tuneable parameter.

To conclude, the RMT threshold correctly identifies noise eigenvalues in a synthetic environment where σ is known. It is therefore reasonable to use this method to determine the intrinsic dimension of a hyperspectral image. In the next section, this method will be investigated in terms of sensitivities to certain variables, such as size of the image, number of bands, noise levels, etc.

5.3. Performance of determining the Intrinsic Dimension

In this section, the accuracy of RMT is evaluated by synthetic experiments and by comparing it to AIC and MDL, which are existing methods for determining the ID. For the real datasets, Cuprite (AVIRIS) and HCI are considered, since the ID of these scenes has been approximated in other studies and by experts' opinions. As discussed in Sections 3.3.1 and 3.6, for Cuprite (AVIRIS), $22 \leq K \leq 28$ and for HCI, $3 \leq K \leq 6$.

First the RMT method is applied to synthetic data, where 50 test images are created for each set of variables to determine a success rate. The default values are $N = 10,000$, $p = 200$, $\sigma = 10^{-3}$, $K = 5$. In this section we consider the noise approximations σ^2_{REF}, σ^2_0, σ^2_{KN}, σ^2_R and σ^2_M, which are discussed in Section 4.5.

The accuracy of RMT will be tested for all 5 noise approximations, with respect to the variables N, p, K and σ^2. By testing the relationship between each variable and the performance of the method, image specifications may be determined that are necessary for reliable ID calculation. RMT also assumes that the noise is i.i.d. and uncorrelated. The sensitivity of the method to such assumptions will also be tested, and these results are shown in Figures 5.3 and 5.4. Once again, Meer's method cannot be tested with the synthetic dataset, and this will be discussed later in the section.

The following section will evaluate how many pixels are necessary in an image for accurate determination of the intrinsic dimension, using RMT with several noise approximation methods.

5.3.1. Performance with respect to the number of pixels

Figure 5.2 shows that the approximations σ_0^2 and σ_{KN}^2 both allowed RMT to determine K with 100% accuracy for any N over 10 pixels. The estimation σ_{REF}^2 allowed RMT to reach full accuracy at 1,000 pixels, and σ_R^2 and σ_M^2 both required more than 10,000 pixels for RMT to reliably estimate K. Although the latter methods performed the worst, this restriction is not detrimental in practice, since 10,000 pixels corresponds to a very small 100×100 image.

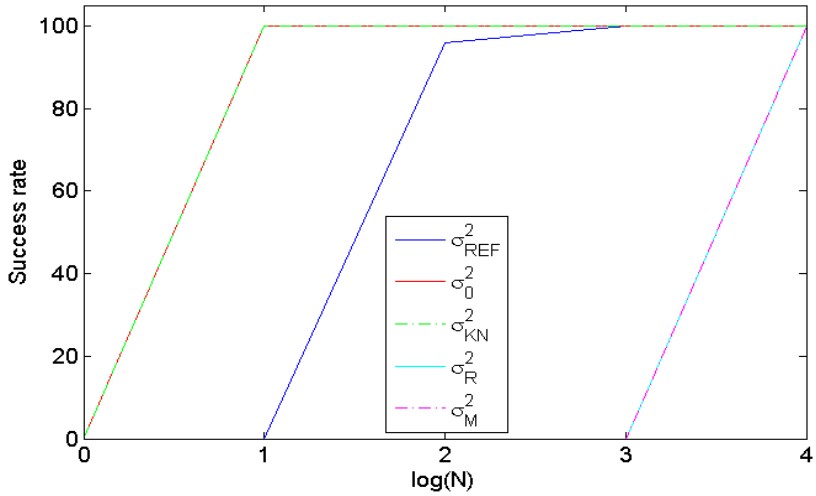


Figure 5.2.: The performance of RMT with 5 noise approximations, with respect to the number of pixels. The approximations σ_0^2 and σ_{KN}^2 required very few pixels for reliable determination of K. The approximation σ_{REF}^2 required an intermediate number of pixels and σ_R^2 and σ_M^2 required the most pixels. All of the displayed values for N represent very small images, and so all methods perform well.

5.3.2. Performance with respect to the number of bands, noise and endmembers

All five noise approximation methods allowed RMT to determine K with 100% accuracy irrespective of the number of bands.

For the given default values, RMT with σ^2_{REF}, σ^2_0 and σ^2_{KN} was able to determine K with 100% accuracy, independently of the value of K. Since this section investigates the accuracy of ID calculation, rather than that of the noise approximation, K is assumed known when determining the noise, but not when using that noise to approximate K. The other two approximation methods did not use prior knowledge, and for $N = 10,000$ these methods also allowed RMT to determine K with 100% accuracy independent of K.

As above, all methods performed with maximum accuracy for all tested values of $\sigma \leq 0.01$. This translates into an SNR of 25:1, which is lower than the SNR of Cuprite as seen in Table 4.2.

RMT assumes i.i.d. noise, and the sensitivity of the method to this assumption is tested. Consider noise in each band that is close to the mean noise variance, σ^2, and slowly increase the range around this mean. To simulate this, a synthetic dataset is created where the standard deviation of the noise in each band is chosen from the distribution $\mathcal{N}(\sigma, \delta)$. The ratio $\gamma = \delta/\sigma$ is used as a measure of variation in the noise across bands. For even a 10% variation in the noise all five methods overestimate K by a large margin. This is further tested by assuming that the mean noise is known, and Figure 5.3 shows that the approximations become inaccurate for even a 10% difference in band variances.

5.3.3. Performance with respect to correlated bands

RMT assumes uncorrelated noise, which is equivalent to assuming uncorrelated pixel values, since the signal is intrinsically correlated. To test the sensitivity of RMT to this assumption, a synthetic experiment is designed where a certain number, p_{cor}, of correlated bands are simulated so that for $j = 2, 4, \ldots, 2 \times p_{\text{cor}}$, the noise is chosen from

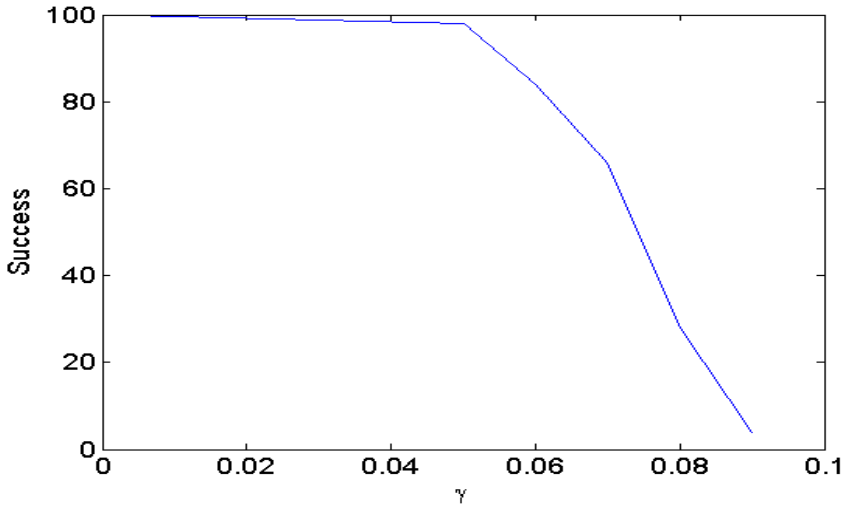


Figure 5.3.: The performance of RMT with 5 noise approximations, with respect to varying noise across bands. If the standard deviation of the noise in each band is chosen from the distribution $\mathcal{N}(\sigma, \delta)$, then the ratio $\gamma = \delta/\sigma$ is a measure of the variation in the noise across bands. Accuracy of the method decreased rapidly as variation between bands increased.

the multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & c\sigma_2^2 & 0 & \dots & & & 0 \\ c\sigma_2^2 & \sigma_2^2 & 0 & & & & \\ 0 & & \ddots & & & & \vdots \\ & & 0 & \sigma^2_{(2\times p\text{cor}-1)} & c\sigma^2_{(2\times p\text{cor})} & 0 & \\ \vdots & & 0 & c\sigma^2_{(2\times p\text{cor})} & \sigma^2_{(2\times p\text{cor})} & 0 & \\ & & & & & \ddots & 0 \\ 0 & & \dots & & & 0 & \sigma_p^2 \end{pmatrix}.$$

For the results illustrated in Figure 5.4, $c = 0.5$. In this particular example all the noise values in a correlated band become correlated with the values in the previous band. This is just one example of correlation, and the same behaviour holds even when the bands considered are not neighbouring bands.

Even a single pair of correlated noise bands results in zero accuracy for all methods. K is substantially overestimated as the number of correlated bands increases, as is shown by the median K values determined by all methods in Figure 5.4.

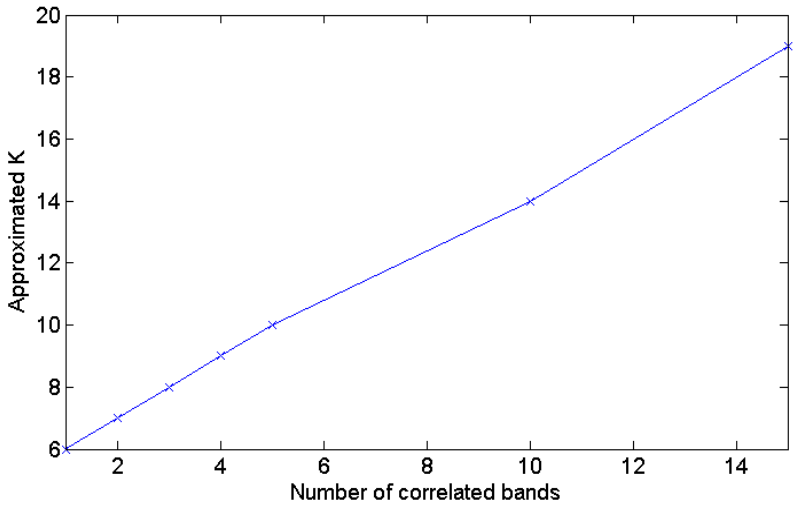


Figure 5.4.: The performance of RMT with 5 noise approximations, with respect to the number of correlated noise bands. This graph shows the median K value determined by all methods (all methods produced the same results) in the presence of correlated bands. Any correlation results in an overestimation of K. The correct value should be $K = 5$.

If the correlations are reduced so that, for instance, $c = 0.1$ for all correlations, then K is accurately determined by all methods. Figure 5.5 shows the median K determined by all methods for 10 correlated bands with respect to c. This shows that higher degrees of correlation result in higher inaccuracies when determining the ID of a hyperspectral image, although the values for K plateau after a certain amount of correlation. These experiments show that highly correlated bands should be removed when using the i.i.d., hybrid multiple regression and hybrid residual noise approximation methods, but some correlation is tolerated.

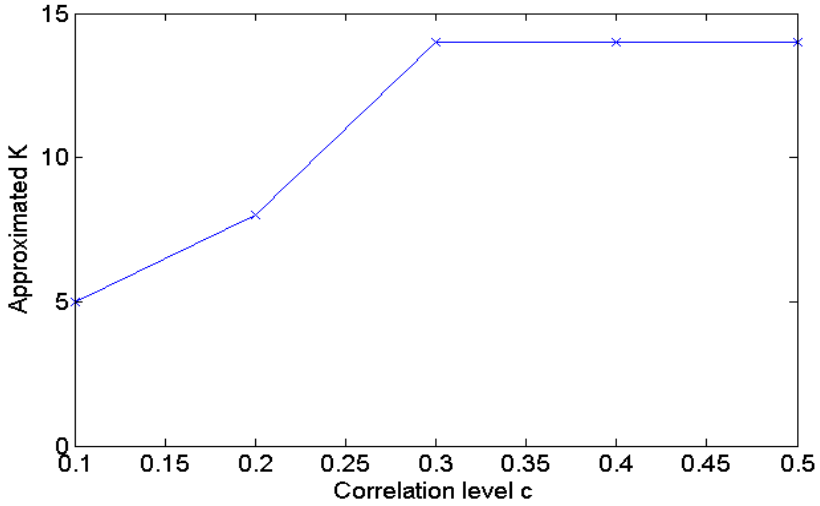


Figure 5.5.: The performance of RMT with 5 noise approximations, with respect to the level of correlated noise. This graph shows the median K value determined by all methods (all methods produced the same results) in the presence of ten correlated bands. Any correlation $c > 0.1$ results in an overestimation of K. The correct value should be $K = 5$.

5.4. Results

RMT has shown to be reliable in synthetic tests when the basic assumptions of the method are met, i.e. that the image contains Gaussian, i.i.d., uncorrelated noise. AIC and MDL are established methods that make the same assumptions about the noise, and so these will be compared with RMT, for real and synthetic images.

5.4.1. Comparison with AIC and MDL

According to Chang and Du [9], AIC and MDL are some of the most popular methods to determine the ID. These methods assume i.i.d. noise and so will be tested in this section. RMT also assumes i.i.d. noise and has been discussed in detail above. AIC and MDL were tested on the synthetic dataset described in Section 3.1, with various values for the number of pixels, number of bands, noise, and number of endmembers. Even though the noise in hyperspectral images is known to be non-i.i.d., and these methods assume i.i.d. noise, we would like to test the sensitivity of all methods to the i.i.d. assumption.

In all cases, the ID was largely overestimated, as seen in some examples in Table 5.2. In contrast, RMT detected the correct value for K for all variable combinations

displayed, and for all i.i.d. and hybrid noise approximations. These findings agree with

Table 5.2.: A comparison between AIC, MDL and RMT for determining the ID of several synthetic datasets. When estimating the number of endmembers in a synthetic dataset with Gaussian, i.i.d. noise, both AIC and MDL overestimated K for each combination of number of pixels, number of bands, noise variance and number of endmembers. RMT was tested with five different noise approximation methods, and all methods result in perfect determination of K for these experiments.

N	p	σ	K	K_{AIC}	K_{MDL}	RMT
1000	200	0.0001	10	154	154	10
10000	400	0.0001	10	339	339	10
10000	200	0.0001	20	147	147	20
10000	200	0.001	10	144	144	10

studies done by Wu *et al.* [10] and Chang and Du [9]. While these methods may be applicable to the signal processing problems for which they were designed, they are not applicable to hyperspectral imagery, even in simulated tests.

5.4.2. Results on Cuprite and HCI

Now RMT, AIC and MDL will be applied to real images, namely Cuprite (AVIRIS) and HCI. These real images were used in the testing of noise approximations (Section 4.3) because the ID of these datasets has been estimated in previous studies. The ID of Cuprite (AVIRIS) is approximately $22 \leq K \leq 28$ and the ID of HCI is approximately $3 \leq K \leq 7$.

First AIC and MDL are applied to Cuprite, and both methods find $K = 188$. When applied to HCI, both methods find $K = 140$. Once again, these values are far higher than the approximations from more reliable methods.

In the synthetic tests for RMT, K was assumed known in order to calculate the i.i.d. noise approximations. For application to real images, both K and σ^2 must be simultaneously estimated. In [14], the dependence of the i.i.d. noise approximations on K is used to determine K, using a hypothesis test. First, K is assumed to be $K_{est} = 1$, which means that the $(K_{est} + 1)^{th}$ largest eigenvalue must be noise, and fulfill the RMT evaluation. The value K_{est} is increased until this is true.

In order to visually compare the threshold when applied to real and synthetic data,

re-write the random matrix evaluation as follows,

$$\begin{aligned} f(N,p,K) &< 1 \text{ for signal eigenvalues, where} \\ f(N,p,K) &= \frac{\sigma^2_{\text{approx}}}{\lambda_K}(\mu_{N,p} + s(\alpha)\sigma_{N,p}). \end{aligned} \tag{5.4.1}$$

For the three i.i.d. noise approximations, the hypothesis testing method is used, and for the other three methods, the noise results from Table 4.2 are used, and the largest λ_K that satisfies (5.4.1) will determine K. The validity of this method for synthetic data is

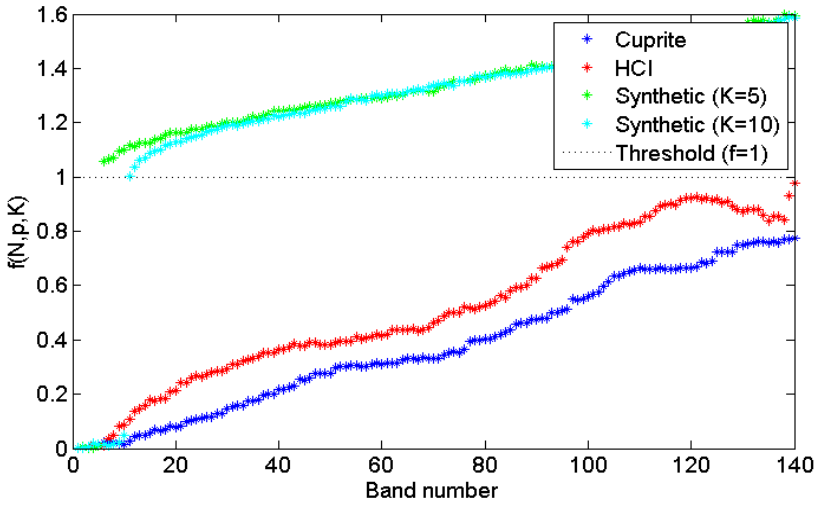


Figure 5.6.: The ratio of RMT threshold to the i^{th} eigenvalue of the observation covariance matrix, described in (5.4.1), is shown for synthetic images, Cuprite (AVIRIS) and HCI. For clarity, the synthetic images are shown with low noise (standard deviation = 0.001). The ID is correctly identified for synthetic images, but neither of the real images follow the same pattern of clear separation between noise and signal eigenvalues.

displayed in Figure 5.6. Only two synthetic images were displayed for clarity, but the same behavior was observed for all the synthetic images generated. For the synthetic images, the line $f(N,p,K) = 1$ is a good threshold between signal and noise eigenvalues, as expected from the Random Matrix behavior. But the Cuprite and HCI scenes behaved differently. These real scenes were further tested using all the noise methods discussed.

Table 5.3 shows the intrinsic dimensions of Cuprite and HCI, as determined by RMT using all i.i.d. and hybrid noise approximation methods. Recall that previous studies [10] and experts' opinions have determined the ID of Cuprite to be between 22 and 28

Table 5.3.: All six noise approximations have been used by RMT to evaluate the ID of Cuprite (AVIRIS) and HCI. All methods significantly overestimated K, with the notable exceptions of the residual and multiple regression on Cuprite, and Meer's method on both datasets.

	σ^2_{REF}	σ^2_0	σ^2_{KN}	σ^2_R	σ^2_M	σ^2_{Meer}
K_{Cuprite}	181	180	180	23	23	25
K_{HCI}	140	140	140	82	82	6

and the ID of HCI to be between 3 and 7. Meer's algorithm provided a mean noise value that allowed reliable ID approximations for both images. This is encouraging! When Table 4.2 is compared with Table 5.3 the noise is significantly underestimated in those methods that failed to properly estimate K, as is intuitive.

Meer's method is able to approximate the mean noise variance in Cuprite (AVIRIS) and HCI to an extent where the ID is in line with previous works (Wu *et al.* [10] described K between 22 and 28 for Cuprite) and experts' opinions (HCI should contain approximately 3 to 7 endmembers). The residual and multiple regression methods were able to determine reasonable ID values for Cuprite, but not for HCI. The consistency of Meer's algorithm must be tested, and so it will be applied to Cuprite (SpecTIR), which is a subset of Cuprite (AVIRIS) but at a higher resolution. This gives $K = 26$, which seems realistic when compared to the $K = 25$ estimate for the larger Cuprite area (at reduced resolution). One of the major drawbacks of Meer's method is that it requires homogeneous regions in an image for reliable approximation, although it does not require the noise to be uncorrelated.

It is interesting to evaluate the range of variances that produce the same value for K in normalised Cuprite and HCI. RMT determines the same K for a 2% underestimation and a 4.3% overestimation of Meer's noise variance in Cuprite. The tolerance seen in HCI is even wider, with RMT returning the same value for a 38% underestimation and a 125% overestimation of Meer's noise variance. RMT appears to be remarkably insensitive to the exact noise approximation, especially for HCI.

5.5. Discussion

In the synthetic tests, several noise approximation methods were used by RMT to determine the ID of a hyperspectral image. The i.i.d. noise approximations (σ^2_{REF}, σ^2_0 and σ^2_{KN}) yielded the best RMT results when the noise was i.i.d. and uncorrelated. RMT was seen to be sensitive to these noise assumptions in the synthetic tests, for the i.i.d. and statistical hybrid noise approximation methods.

When applied to real images, the residual and multiple regression methods estimated a much higher mean noise variance than i.i.d. methods for Cuprite. When using the former noise approximation for RMT, the estimate for K was similar to the values reported in the literature. The same was not true for HCI, but RMT with Meer's noise approximation yielded realistic values for K for Cuprite (AVIRIS), Cuprite (SpecTIR) and HCI. From these tests this method appears to be a success.

Unfortunately, Figures 5.3 and 5.4 show that K is not reliably determined when the i.i.d. and uncorrelated noise assumptions are not met, so this method may not apply to all images. Meer's method is also not always applicable, as it requires homogeneous regions within the image.

There exist methods to transform image data so that the noise has zero mean and unit variance in each band, and each noise vector is chosen from a multivariate normal distribution (*i.e.* the noise is i.i.d. and uncorrelated). These methods are called noise whitening and will be discussed in the next section. If the noise is i.i.d. and uncorrelated, then synthetic results show that the method investigated above could be successfully applied to all images.

6. RMT with image whitening

Noise whitening methods are based on linear transforms that normalise the distribution of the Gaussian noise to zero mean and unit variance. The noise is also decorrelated, so that the noise covariance becomes the identity matrix. This is particularly useful for methods that assume i.i.d. and uncorrelated noise (such as RMT), since the noise variance is reduced to unity in every band, which would create a whitened dataset with i.i.d. noise that has been decorrelated.

In order to successfully whiten the noise, the noise must be reliably estimated. Also, the whitening is applied to the observation covariance matrix, and so the signal is also transformed. In this chapter several whitening methods will be evaluated in order to discover if RMT is applicable to whitened images.

6.1. Methods

There are several techniques for whitening a hyperspectral image. Although the transformation is applied to the model covariance matrix from the entire image, Σ, the goal is to whiten the noise covariance, Σ_n. To do this, the signal, Σ_s, and noise covariance matrices must be separable as follows:

$$\Sigma = \Sigma_s + \Sigma_n. \tag{6.1.1}$$

If noise covariance is approximated by $\hat{\Sigma}_n$, then the noise may be approximately whitened by symmetric methods, asymmetric methods, and eigenvector methods.

6.1.1. Symmetrical whitening

If $\hat{\Sigma}_n$ approximates the noise covariance matrix Σ_n, then the noise may be whitened symmetrically as follows:

$$\hat{\Sigma}_n^{-1/2}\Sigma\hat{\Sigma}_n^{-1/2} = \hat{\Sigma}_n^{-1/2}\Sigma_s\hat{\Sigma}_n^{-1/2} + \hat{\Sigma}_n^{-1/2}\Sigma_n\hat{\Sigma}_n^{-1/2} \tag{6.1.2}$$

$$\approx \hat{\Sigma}_n^{-1/2}\Sigma_s\hat{\Sigma}_n^{-1/2} + I \tag{6.1.3}$$

It is advantageous to whiten symmetrically as the signal covariance matrix remains symmetric.

6.1.2. Asymmetrical whitening

The noise may also be whitened asymmetrically, as in [78] by

$$\begin{aligned} \hat{\Sigma}_n^{-1}\Sigma &= \hat{\Sigma}_n^{-1}\Sigma_s + \hat{\Sigma}_n^{-1}\Sigma_n && (6.1.4)\\ &\approx \hat{\Sigma}_n^{-1}\Sigma_s + I && (6.1.5) \end{aligned}$$

Asymmetrical whitening still reduces the noise covariance to the identity matrix, but the signal covariance is no longer symmetrical.

6.1.3. Eigenvector whitening

Levy [78] also suggests using the estimated noise covariance matrix indirectly in order to whiten. To do this define F, such that

$$F = E\Delta^{-1/2}, \tag{6.1.6}$$

where E contains the eigenvectors of $\hat{\Sigma}_n$ and Δ is a diagonal matrix containing the eigenvalues of $\hat{\Sigma}_n$. Then the observed covariance may be whitened by

$$\begin{aligned} F^T\Sigma F &= F^T\Sigma_s F + F^T\Sigma_n F && (6.1.7)\\ &= F^T\Sigma_s F + \Delta^{-1/2}E^T\Sigma_n E\Delta^{-1/2} && (6.1.8)\\ &\approx F^T\Sigma_s F + \Delta^{-1/2}\Delta\Delta^{-1/2} && (6.1.9)\\ &= F^T\Sigma_s F + I. && (6.1.10) \end{aligned}$$

This method results in a symmetric signal covariance matrix.

6.1.4. Whitening using the residual based method

The residual based method described in Section 4.3.1 approximates the noise covariance matrix by $\hat{\Sigma}_n$, where $\hat{\Sigma}_n$ is a diagonal matrix containing the inverse diagonal elements of Σ^{-1}, i.e. if

$$\Sigma^{-1} = \begin{pmatrix} \zeta_{1,1} & \cdots & \zeta_{1,p} \\ \vdots & \zeta_{i,i} & \vdots \\ \zeta_{p,1} & \cdots & \zeta_{p,p} \end{pmatrix} \tag{6.1.11}$$

$$\text{then } \hat{\Sigma}_n = \begin{pmatrix} \zeta_{1,1}^{-1} & 0 & \cdots & 0 & 0 \\ 0 & \zeta_{2,2}^{-1} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \vdots & 0 & \zeta_{(p-1),(p-1)}^{-1} & 0 \\ 0 & 0 & \cdots & 0 & \zeta_{p,p}^{-1} \end{pmatrix}. \tag{6.1.12}$$

This approximation is used in [9], where the observation covariance is whitened symmetrically as in (6.1.5).

6.1.5. Discussion

In the previous sections, all the whitening methods were shown to produce the same analytical results. Each method involves different computational complexity, however. If asymmetrical whitening is performed, the noise covariance matrix must be estimated and inverted. If the estimation is not a true covariance matrix, then this matrix may not be invertible. If symmetric whitening is performed, the square root of the inverted matrix must be calculated, which adds to the computational cost.

The eigenvalue whitening method requires the eigenvectors of the noise approximation method, which is also computationally complex. The advantage of this method is that the inversion and square root are only performed on the diagonal eigenvalue matrix, so that these operations may be performed on each eigenvalue separately, rather than processing the entire matrix. The residual whitening method is even more advantageous since it does not require an estimate of the noise covariance matrix, which is often difficult to obtain. While the observation covariance matrix must be inverted, this is a true covariance matrix so the inversion is possible.

Due to these advantages, the residual method will be used for whitening in this chapter.

6.2. Do whitening methods whiten accurately?

The residual based noise approximation method produced diagonal matrices for the whitening, whereas the multiple regression method also displays correlations between bands. The known noise covariance is symmetrically whitened as in (6.1.3), and both methods correctly whiten the noise covariance for i.i.d. noise, as seen in Figure 6.1. This graph was created by simulating a synthetic dataset with 10,000 pixels, 200 bands, 5 endmembers and a noise standard deviation of 10^{-3} (in each band). The whitening matrices were calculated independently of the known noise, and then applied to the known noise covariance. The result is a close approximation to the identity matrix, since the mean of the diagonal elements is 0.9951 (standard deviation 0.0314) and the mean of the off-diagonal elements is 7.67×10^{-5} (standard deviation 0.0099).

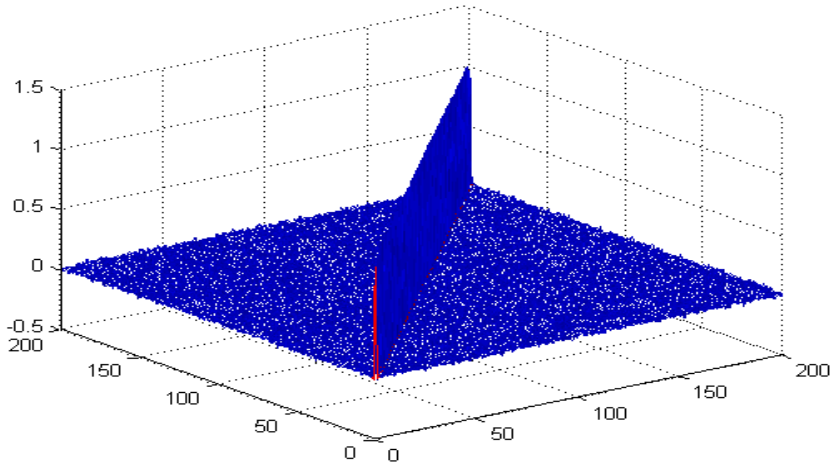


Figure 6.1.: Whitening the known i.i.d. noise covariance using estimated noise. Both the residual and multiple regression methods were successfully able to whiten the i.i.d. noise covariance matrix to a close approximation of the identity matrix.

The same experiment was applied to a synthetic scene with non-i.i.d. noise across bands. Consider an image where the noise is almost i.i.d. but with small difference in the noise variance across bands. To simulate this, choose the noise standard deviation for the i^{th} band, σ_i from the normal distribution $\mathcal{N}(\sigma, \delta)$, where δ controls the range of noise values across bands. For this experiment $\sigma = 10^{-3}$ and $\delta = 10^{-2}$. The effect of varying noise across bands will be examined further in Section 6.3. As in previous synthetic

tests there are 10,000 pixels, 200 bands and 5 endmembers. The noise covariance was successfully whitened, as seen in Figure 6.2, since the mean of the diagonal elements is 0.9945 (standard deviation 0.1265) and the mean of the off-diagonal elements is -7.81×10^{-5} (standard deviation 0.0099).

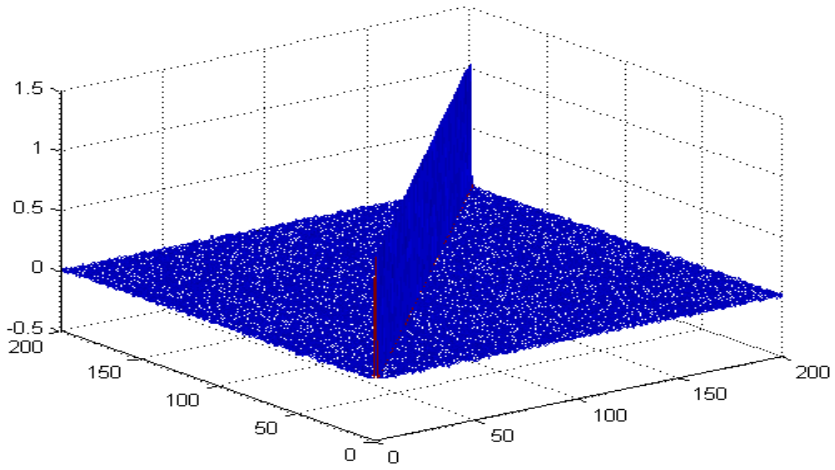


Figure 6.2.: Whitening the known non-i.i.d. noise covariance using estimated noise. Both the residual and multiple regression methods were successfully able to whiten the non-i.i.d. noise covariance matrix to a close approximation of the identity matrix.

Whitening methods also claim to decorrelate the noise, in particular the residual whitening method [1]. The same experiment was applied to a synthetic scene with correlated bands. In this test, suppose there are p_{cor} correlated bands. Then the noise is chosen from the multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & c\sigma_2^2 & 0 & \dots & & & 0 \\ c\sigma_2^2 & \sigma_2^2 & 0 & & & & \\ 0 & & \ddots & & & & \vdots \\ & & 0 & \sigma^2_{(2\times p\text{cor}-1)} & c\sigma^2_{(2\times p\text{cor})} & 0 & \\ \vdots & & 0 & c\sigma^2_{(2\times p\text{cor})} & \sigma^2_{(2\times p\text{cor})} & 0 & \\ & & & & & \ddots & 0 \\ 0 & & \dots & & & 0 & \sigma_p^2 \end{pmatrix}.$$

for $j = 2, 4, \dots, 2 \times p_{\text{cor}}$. For the results illustrated in Figure 4.6, all bands are correlated

and $c = 0.5$. The noise covariance was not successfully whitened in this case, as seen in Figure 6.3, since the diagonal values are consistently greater than one and the off-diagonal correlation terms remain (in this experiment, they are just off-diagonal since neighbouring bands are correlated). Specifically, the mean of the diagonal elements is 1.3434 (standard deviation 0.0146) and the mean of the off-diagonal elements is 0.0031 (standard deviation 0.0493). As seen in Chapter 4, this is due to the inaccuracy of the noise approximation rather than any inaccuracy in the method itself.

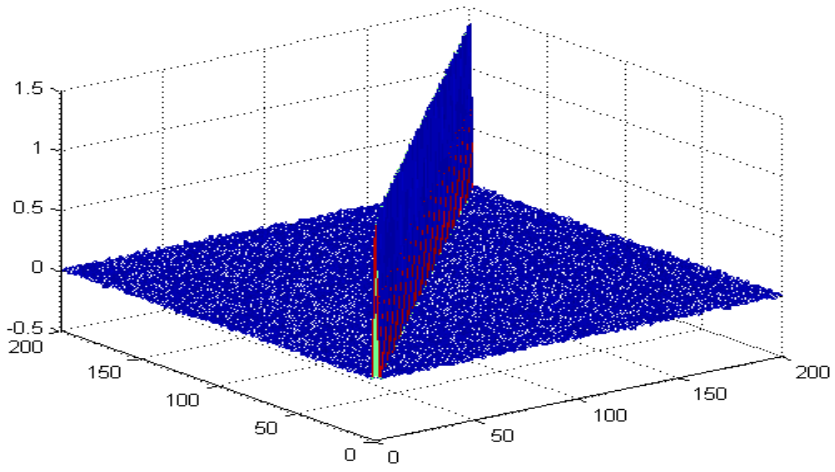


Figure 6.3.: Whitening the known correlated noise covariance using estimated noise. Neither the residual nor multiple regression methods were successfully able to whiten the correlated noise covariance matrix. Note that the diagonal values are greater than one, where they should be whitened to unity, and off-diagonal correlation terms remain.

6.3. Performance of determining the Intrinsic Dimension

Section 6.2 shows that the residual and multiple regression methods are able to reliably estimate the noise in data with uncorrelated noise and successfully whiten synthetic data. These methods will now be tested in 3 methods that can use whitened data to determine the ID, namely RMT, Malinowski's EIF, and Chang's HFC method.

In the case of i.i.d. noise, noise whitening will have no effect on RMT, since the residual whitening matrix W is equal to $\sigma^{-1}I_p$, where σ^2 is the constant noise variance. This is shown by evaluating the effect of whitening on the RMT threshold as follows:

$$\begin{aligned}
\text{Whitening matrix } W &= \sigma^{-2}I_p \\
\text{Whitened } S(N) \text{ becomes} & \quad WS(N)W \\
&= \sigma^{-1}I_pS(N)\sigma^{-1}I_p \\
&= \sigma^{-2}S(N) \\
\text{If } S(N)\underline{x} &= \lambda\underline{x} \\
\text{then } (\sigma^{-2}S(N))\underline{x} &= (\sigma^{-2}\lambda)\underline{x} \\
& \quad \text{where } \lambda \text{ is an eigenvalue of } S(N) \\
& \quad \text{and } \underline{x} \text{ is the corresponding eigenvector.} \\
& \quad \text{Eigenvalues of the whitened observation} \\
& \quad \text{covariance matrix may be evaluated by:} \\
\sigma^{-2}\lambda_i &< (\mu_{N,p-K} + s(\alpha)\sigma_{N,p-K}) \\
& \quad \text{since the whitened noise variance has value} \\
& \quad \text{unity for calculation of the RMT threshold} \\
\rightarrow \lambda_i &< \sigma^2(\mu_{N,p-K} + s(\alpha)\sigma_{N,p-K}),
\end{aligned}$$

where λ_i are the ordered (descending) eigenvalues of $S(N)$ for $i, 1 \leq i \leq p$. So this reduces to the original RMT evaluation, showing that whitening will have no effect on the RMT evaluation of an image with i.i.d. noise.

The performance of RMT will therefore only be evaluated in the case of non-i.i.d. noise. Figure 6.3 shows that correlated noise is not effectively whitened by the residual method, and so the sensitivity of the methods to this correlation will also be tested.

6.3.1. Performance with respect to non-i.i.d. noise

A synthetic dataset is simulated with $N = 10,000$, $p = 200$, $K = 5$ and the noise in the i^{th} band, σ_i is chosen from $\mathcal{N}(\sigma, \delta)$. The results are shown for 20 tests in Figure 6.4, where whitened EIF performed with maximum success for all values of δ, whitened HFC appears independent of δ but had lower success rates, and whitened RMT was shown to depend on the amount of variation in noise between bands.

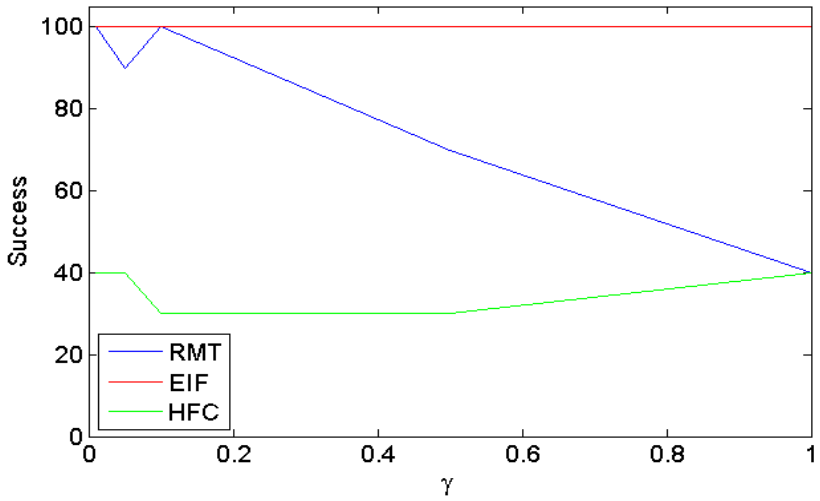


Figure 6.4.: The success of whitened RMT, EIF and HFC for non-i.i.d. noise. The noise in the i^{th} band is chosen from $\mathcal{N}(\sigma, \delta)$, and the figure above shows the ratio $\gamma = \delta/\sigma$. Whitened EIF gave excellent success rates, independent of the variation of the noise. HFC was also independent of the noise, but with the lower success rates consistent with this method. RMT performance decreased with larger noise variation across bands.

For the whitening, both the multiple regression and residual methods were tested, and they produced the same respective success rates when used to whiten the observation covariance matrix in RMT, EIF and HFC (with $F_D = 10^{-3}$) for determining ID. HFC was found to be very sensitive to the variation in abundances and exact noise values per pixel, so even in the best circumstances, this method has a success rate of only 30–40%. For fair comparison, Bioucas-Dias and Nascimento compared their HySime method against noise whitened HFC in [57] by displaying the results for K after 50 runs of their experiment. (If the median values for K were considered for ten runs, then HFC determined that $K = 5$ for all noise variation ratios displayed.) Malinowski's method was particularly reliable for variable noise per band, independent of the noise variation.

RMT results were significantly improved from the i.i.d. evaluation shown in Figure 5.3, but this method still showed a decrease in success when noise variation between bands increased.

6.3.2. Performance with respect to correlated bands

Once again a synthetic dataset is simulated with $N = 10,000$, $p = 200$ and $K = 5$. In this test, suppose p_{cor} correlated bands, so that bands $(j-1)$ and j are correlated, for $j = 2, \ldots, 2 \times p_{\text{cor}}$. Then the noise is chosen from the multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & c\sigma_2^2 & 0 & \cdots & & & 0 \\ c\sigma_2^2 & \sigma_2^2 & 0 & & & & \\ 0 & & \ddots & & & & \vdots \\ & & 0 & \sigma^2_{(2\times p\text{cor}-1)} & c\sigma^2_{(2\times p\text{cor})} & 0 & \\ \vdots & & 0 & c\sigma^2_{(2\times p\text{cor})} & \sigma^2_{(2\times p\text{cor})} & 0 & \\ & & & & & \ddots & 0 \\ 0 & & \cdots & & & 0 & \sigma_p^2 \end{pmatrix}.$$

For the results illustrated in Figure 6.5, $c = 0.5$. In this example, each band is correlated with its neighbour, but the same behaviour is observed if bands are correlated in a different order. Figure 6.5 shows that Malinowski's EIF is not affected by the correlation and HFC only overestimates K by one. The performance of RMT was linearly dependent on the number of correlated bands. This is interesting, since RMT and EIF use the same whitened covariance matrix, but HFC compares eigenvalues of the observation correlation and covariance matrices, which are both whitened by the same matrix.

The simulated tests show improved results for whitening in the presence of non-i.i.d. noise, although RMT was not accurate in the presence of correlated bands. In the next section, all three methods will be applied to real images.

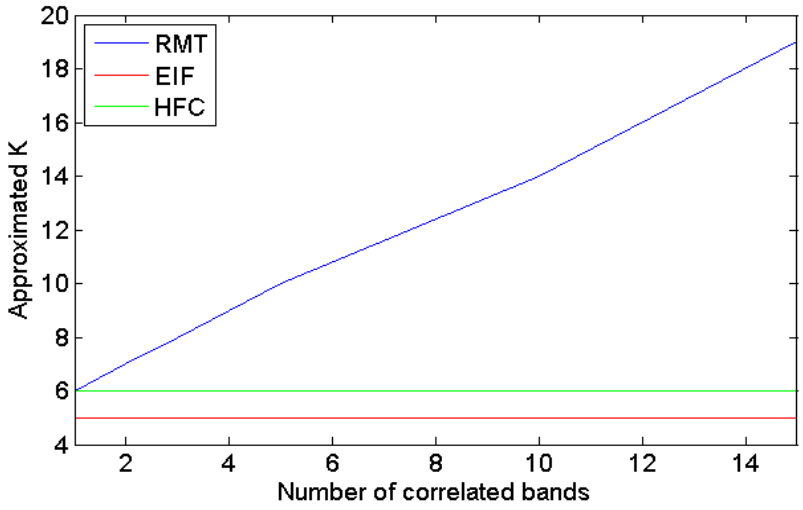


Figure 6.5.: The ID approximated by whitened RMT, EIF and HFC for correlated bands. Figure 6.5 shows that Malinowski's EIF is not affected by the correlation and HFC only overestimates K by one. The performance of RMT was linearly dependent on the number of correlated bands. (The correct value is $K = 5$).

6.4. Results on Cuprite and HCI

First the three methods are tested on Cuprite (AVIRIS), and the results (before and after whitening) for each of the methods are given in Table 6.1.

The accuracy of EIF is dramatically improved when using a noise whitening step. When applied to Cuprite, $K_{EIF} = 75$, but when applied with a noise whitening step (using the residual method), $K_{EIF} = 25$, which is similar to other methods studied. The advantage of this method is that it has very low computational complexity — it only needs the eigenvalues from a single matrix. HFC did not show large differences between the original and whitened methods, but this may be expected, since the original method does not assume i.i.d. noise. This method was largely dependent on the user-defined threshold. RMT was improved from $K = 181$ (when using the i.i.d. noise variance) to a somewhat overestimated K of 37. This is still higher than the K of 25 and 23 estimated by using the mean noise variance approximated by Meer and the residual/regression methods respectively.

HFC was the only method that showed different results when using the residual and multiple regression noise approximations. This may be due to the fact that the residual

Table 6.1.: RMT, EIF, and HFC (with $F_D = 10^{-4}, 10^{-3}, 10^{-2}$) applied to Cuprite (AVIRIS), before (K_O) and after (K_W) whitening with various noise approximation methods. Note that RMT without whitening still required a noise approximation, and the different methods tested account for the large range (see Table 5.3).

	K_O	K_W ($\sigma^2_{Residual}$)	K_W ($\sigma^2_{Regression}$)	K_W (σ^2_{Meer})
RMT	23 – 181	37	37	21
EIF	75	25	25	83
HFC	22, 24, 30	19, 24, 27	26, 26, 29	19, 20, 22

method calculates different whitening matrices for the centered and non-centered covariances, whereas the multiple regression method would use the same whitening matrix for both.

Meer's method was less reliable for noise approximation, but it does not assume uncorrelated bands, and so a whitening matrix was created by using the noise approximated by Meer. This method had approximated a noise of zero for the last three bands, which created problems when inverting the whitening matrix. These bands were set to the minimum non-zero noise variance over all bands. As expected, this method produced different results to the other whitening methods, including a notably more reasonable value of $K = 21$ for RMT.

Next, these methods are applied to HCI and the results are shown in Table 6.2. Previous tests show that the regression and residual methods underestimate the noise in HCI, since HCI contains higher levels of correlation than Cuprite, and so overestimation of the ID may be expected. HFC was the only method to give reasonable values for K, which is to be expected since this method performed well in synthetic tests even with correlated bands. This method is not sensitive to the noise underestimation since the method is based on a comparison of two matrices which are derived from the same image.

The multiple regression method produced the same results as the residual method. Using Meer's method to create the whitening matrix, the results are similar for HFC and RMT, where both results are considered reasonable. This result is an improvement over i.i.d. noise estimates, but it is the same as the K obtained by using the mean noise variance as approximated by Meer. It is important to note that Meer's method was the most reasonable noise approximation method tested with RMT since it calculates the

Table 6.2.: RMT, EIF, and HFC (with $F_D = 10^{-4}, 10^{-3}, 10^{-2}$) applied to HCI, before (K_O) and after (K_W) whitening with various noise approximation methods. Note that RMT without whitening still required a noise approximation, and the different methods tested account for the large range (see Table 5.3).

	K_O	K_W (Residual)	K_W (Regression)	K_W (Meer)
RMT	6 – 140	82	82	6
EIF	51	45	45	53
HFC	5, 6, 7	4, 7, 7	4, 7, 7	4, 5, 5

noise on a band-by-band basis and so is not affected by correlation.

6.5. Discussion

The residual and multiple regression methods are indistinguishable in many situations, including the two real images that were tested. The residual method has a much lower computational complexity, and appears to be sufficient in many scenarios. The only scenario that may show some difference is when the noise is spatially dependent. Both methods were unable to reliably estimate the noise in HCI, due to the high levels of correlation. Meer's method for noise approximation is a spatially based method that does not require uncorrelated bands, and this method was the only one surveyed that was able to reliably estimate the noise in HCI when using RMT.

HFC gave reasonable results (with and without whitening) for the two real datasets, however the method is dependent on a user-determined threshold. Malinowski's method performed well on Cuprite (AVIRIS) when whitened, but was unable to produce reasonable results for HCI, even when using Meer's method for the whitening. Whitened RMT gave reasonable results in both Cuprite and HCI (whitened with Meer). However, these results were no better than when using the mean noise variance as approximated by Meer's method.

The results of all the whitening methods are promising when the noise is uncorrelated, but the ID of real datasets was only reliably determined by RMT when Meer's noise approximation was used. Whitening the image does not improve on the results in the i.i.d. section (when using the mean of Meer's noise approximation), and so further options will be explored in Chapter 7.

7. A new RMT method which does not assume i.i.d. and uncorrelated noise

RMT has produced excellent results in synthetic data in Sections 5 and 6 where noise was i.i.d. and uncorrelated. However, when applied to real images, the i.i.d. method only produced good results when using the mean noise variance approximated by Meer's method, which was the least reliable noise approximation method used. The more reliable noise approximation methods did not always produce good results (especially in HCI). Whitening the noise produced an improvement in real data, but the ID was still overestimated.

To solve these issues, a model will be designed that makes no assumptions of i.i.d. or uncorrelated noise. This method will be independent of the noise approximation method used, relying only on a certain level of accuracy.

7.1. Model development

The i.i.d. model developed by Kritchman and Nadler in [14] relied on a separation between signal and noise eigenvalues. This separation was easy in the case of i.i.d. noise. If the noise covariance matrix is not a multiple of the identity matrix, the eigenvalues are still separable. To prove this the proof in Section 5.1 is revisited, but with noise covariance matrix $\Phi = \Delta\Delta^T$, where Φ is not necessarily diagonal.

Proposition 2. *Suppose $\underline{\tilde{x}} \in \mathbb{R}^p$ is a random column vector described by $\underline{\tilde{x}} = V\underline{\tilde{u}} + \Delta\underline{\tilde{\xi}}$, where V is a $(p \times K)$ matrix with linearly independent columns, $\underline{\tilde{u}} \in \mathbb{R}^K$, Δ is a $(p \times p)$ weighting matrix, and $\underline{\tilde{\xi}} \in \mathbb{R}^p$, where $\tilde{\xi}_i \sim N(0,1)\ \forall i$. Assume that V is chosen so that $\mathbb{E}\left[\underline{\tilde{u}}\underline{\tilde{u}}^T\right]$ has full rank K, and assume that $(V\underline{\tilde{u}})$ is independent to $\underline{\tilde{\xi}}$.*

If $S = \mathbb{E}\left[\underline{\tilde{x}}\underline{\tilde{x}}^T\right]$, then S may be written as $S = \Pi + \Phi$, where Π is rank K, and $\Phi = \Delta\Delta^T$.

Proof.

$$\begin{aligned}
\text{Define } S &= \mathbb{E}\left[\tilde{\underline{x}}\tilde{\underline{x}}^T\right] \\
&= \mathbb{E}[(V\tilde{\underline{u}})(V\tilde{\underline{u}})^T + (V\tilde{\underline{u}})(\Delta\tilde{\underline{\xi}})^T + \\
&\quad (\Delta\tilde{\underline{\xi}})(V\tilde{\underline{u}})^T + (\Delta\tilde{\underline{\xi}})(\Delta\tilde{\underline{\xi}})^T] \\
&= \mathbb{E}\left[(V\tilde{\underline{u}})(V\tilde{\underline{u}})^T\right] + \Delta I_p \Delta^T \\
&= V\mathbb{E}\left[\tilde{\underline{u}}\tilde{\underline{u}}^T\right] V^T + \Delta\Delta^T \\
\text{Let } \Pi &= V\mathbb{E}\left[\tilde{\underline{u}}\tilde{\underline{u}}^T\right] V^T \\
\text{Let } \Phi &= \Delta\Delta^T \\
\text{Then } S &= \Pi + \Phi
\end{aligned}$$

□

Proposition 1 shows that the signal and noise are separable in the correlation matrix. However, the Random Matrix Theory formulae rely on the fact that $\Delta\Delta^T$ is a diagonal matrix, where the i^{th} diagonal entry $\sigma_i^2 = \sigma^2\ \forall i$. Write $\Delta\Delta^T$ as $\sigma^2 I_p + \bar{\Lambda}$, where σ^2 is the mean of the diagonal entries of Δ, and $\bar{\Lambda}$ is the perturbation of the noise from $\sigma^2 I_p$. Then the separation of the signal and noise eigenvalues is shown in Proposition 3.

Proposition 3. *Suppose S is a $(p \times p)$ positive definite matrix described by $S = \Pi + \Phi$, where Π is symmetric and $\Phi = \sigma^2 I_p + \bar{\Lambda}$ (σ^2 is scalar). Then λ_S^i, the i^{th} eigenvalue of S (eigenvalues indexed in descending order), is given by*

$$\lambda_S^i = \lambda_\Pi^i + \frac{\underline{x}_\Pi^{iT}\Phi\underline{x}_S^i}{\underline{x}_\Pi^{iT}\underline{x}_S^i} \tag{7.1.1}$$

$$= \lambda_\Pi^i + \sigma^2 + \frac{\underline{x}_\Pi^{iT}\bar{\Lambda}\underline{x}_S^i}{\underline{x}_\Pi^{iT}\underline{x}_S^i}, \tag{7.1.2}$$

where $\underline{x}_S^i$ is the i^{th} eigenvector of S and $\underline{x}_\Pi^i$ is the i^{th} eigenvector of Π, ordered by the decreasing order of their respective eigenvalues, provided $\underline{x}_\Pi^{iT}\underline{x}_S^i \neq 0$.

Proof.

$$\begin{aligned}
\lambda_S^i \underline{x}_\Pi^{iT} \underline{x}_S^i &= \underline{x}_\Pi^{iT} S \underline{x}_S^i \\
&= \underline{x}_\Pi^{iT} (\Pi + \Phi) \underline{x}_S^i \\
&= \underline{x}_\Pi^{iT} \Pi \underline{x}_S^i + \underline{x}_\Pi^{iT} \Phi \underline{x}_S^i \\
&= \lambda_\Pi^i \underline{x}_\Pi^{iT} \underline{x}_S^i + \underline{x}_\Pi^{iT} \Phi \underline{x}_S^i \\
\lambda_S^i &= \lambda_\Pi^i + \frac{\underline{x}_\Pi^{iT} \Phi \underline{x}_S^i}{\underline{x}_\Pi^{iT} \underline{x}_S^i}
\end{aligned}$$

provided that $\underline{x}_\Pi^{iT} \underline{x}_S^i \neq 0$. $\Phi = \sigma^2 \ I_p + \bar{\Lambda}$, where σ^2 is constant, leading to (7.1.2). □

Note that this result is very similar to the case of i.i.d. noise shown in (5.1.8), but with extra terms involving the eigenvectors of Π and S.

$$\text{Define } \rho_i = \frac{\underline{x}_\Pi^{i\ T} \bar{\Lambda} \underline{x}_S^i}{\underline{x}_\Pi^{i\ T} \underline{x}_S^i} \ \forall 1 \leq i \leq p. \tag{7.1.3}$$

Then the threshold condition for noise eigenvalues (5.1.10) may be rewritten as:

$$\lambda_S^i < (\sigma^2 + \rho_i)(\mu_{N,p} + s(\alpha)\sigma_{N,p}) \ \forall 1 \leq i \leq p. \tag{7.1.4}$$

Now ρ_i depends on knowledge of $\bar{\Lambda}$ (the perturbation of the noise from $\sigma^2 I_p$), the eigenvectors of Π (remember that the observation covariance matrix $S = \Pi + \sigma^2 I_p + \bar{\Lambda}$), and the eigenvectors of S. As S is known, the noise statistics must be approximated in order to calculate Π and $\bar{\Lambda}$.

Equation (7.1.4) may be tested in a synthetic experiment. 1,000 Wishart matrices were created, with $N = 10,000$ pixels, $p = 200$ spectral bands, and the mean standard deviation of the noise is $\sigma = 10^{-3}$. In this case, 99.8% of the largest eigenvalues fulfilled (7.1.4). This experiment contained non-i.i.d. noise, showing that the Random Matrix Theory statistics are applicable in this case, if used as in (7.1.4).

7.2. Performance of determining the Intrinsic Dimension

A theoretical adaptation of RMT has been designed, and it will be tested on the synthetic dataset, with default variables $N = 10,000$, $p = 200$, $K = 5$ and $\sigma = 10^{-3}$. These variables will first be tested with noise chosen from $\mathcal{N}(0, \sigma^2 I_p)$ (i.i.d. noise) in order to show that in this case the results of the new method are no worse than the original RMT method shown in Chapter 5. The effects of non-i.i.d. and correlated noise will be evaluated separately, to show that the new method is an improvement over the original method, in the more general situation. This dataset will enable the evaluation of the performance of the method and detection of any sensitivities or limits with respect to specific variables.

Various noise environments will be considered. First, to test the accuracy of the RMT method, rather than that of the noise approximation, real noise will be considered, i.e. the generated noise in each pixel is assumed known, and this will be used to calculate the noise covariance matrix (or, in the case of HySime, the noise in each pixel is required). The results of RMT are more accurate when the distribution is known, but the noise approximation methods are based on the generated noise per pixel. So the generated noise is considered, and with enough samples, this will approximate the actual distribution of the noise, which is not assumed known. Next, three noise approximation methods will be considered: residual method, multiple regression method, and Meer's method, as discussed in Section 4.3. Meer's method is not applicable to the synthetic set and will only be used in the real images. The RMT adaptation will also be compared to two well known methods, namely NSP (Section 2.1.6) and HySime (Section 2.1.7). An interesting link is that NSP uses the residual method for noise approximation and HySime uses the multiple regression method. Since RMT uses these same approximations, a fair comparison may be made.

For the implementation of HySime, Matlab code has been provided by Bioucas-Dias and Nascimento (the functions *estNoise* and *hysime*), available online[1].

NSP displayed interesting behaviour by consistently underestimating K by one. In the paper where Chang and Du introduced the method [9], the ID of the synthetic set was also underestimated by one. The authors considered this to be due to the nature of similar spectra included in their dataset, however in this section this behaviour has been consistent independent of which spectra were chosen from the library. For fair comparison, one is added to all K estimates from the NSP method.

The following graphs show the dependence of all three methods on several variables,

[1] http://www.deetc.isel.ipl.pt/jnascimento/public.html

including the image size, number of bands, noise level, number of endmembers, etc. Each point on the graphs displayed represents the percentage correct estimation of K over 20 experiments.

7.2.1. Performance with respect to the number of pixels

In this experiment, RMT is tested with real noise (actual noise values) and two noise approximations, and is compared to HySime and NSP, for varying numbers of pixels. Figure 7.1 shows that RMT with real noise, NSP and HySime (the latter two with approximated noise) all perform with maximum success for any tested image size. RMT with both noise approximation methods performed well for an image larger than (80×80) pixels, which is much smaller than most hyperspectral images. The standard number of pixels, $N = 10,000$, in the synthetic dataset is therefore a valid test.

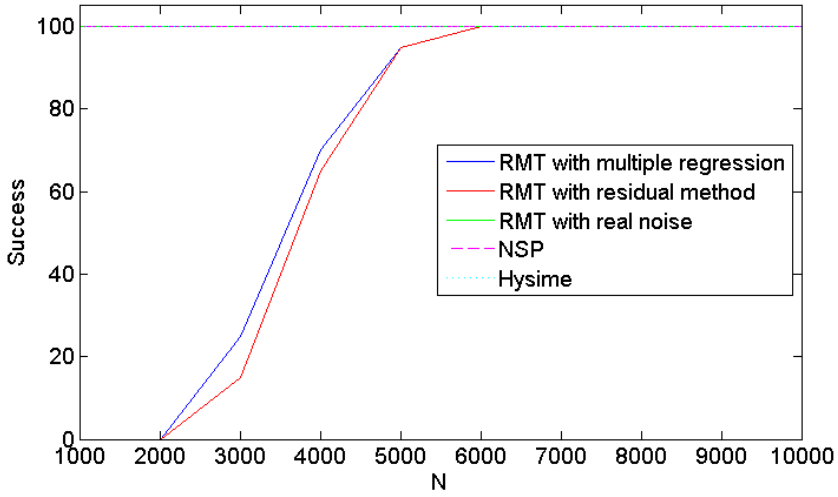


Figure 7.1.: The accuracy of RMT, NSP and HySime with respect to number of pixels. RMT with real noise, NSP and HySime have 100% success for all values of N. RMT with both noise approximations reached 100% for $N > 6,000$.

7.2.2. Performance with respect to the number of bands

RMT is tested with real noise values and two noise approximations, and is compared to HySime and NSP, for varying numbers of spectral bands. All five methods obtained maximum accuracy for the full range of p that was tested, $20 \leq p \leq 200$. It is encouraging that all methods are independent of this variable, as was the case in Chapter 5.

7.2.3. Performance with respect to the number of endmembers

Here RMT is tested with real noise values and two noise approximations, and is compared to HySime and NSP, for varying numbers of endmembers. All five methods obtained maximum accuracy for the full range of K that was tested, $2 \leq K \leq 18$. The only exception was HySime, which underestimated K for 80% of the tests when $K = 18$. This may be due to a lower accuracy in the multiple regression noise approximation for complex pixels. HySime had 100% accuracy for $K < 18$, and it is encouraging that all methods are independent of this variable for $K < 18$, as was the case in Chapter 5.

7.2.4. Performance with respect to variation of noise variance across bands

In this experiment, RMT is tested with real noise values and two noise approximations, and is compared to HySime and NSP, for non-i.i.d. noise. All five methods obtained maximum accuracy for the full range of γ that was tested, where $\gamma = \delta/\sigma$ and the noise standard deviation in each band is selected from $\mathcal{N}(\sigma, \delta)$. The range considered is $0.01 \leq \gamma \leq 1$, with the upper limit chosen to reduce the likelihood of selecting negative variances, since $\gamma > 1$ results in $\delta > \sigma$. (In the event that a negative variance is selected, the variance is resampled from the distribution, forcing positivity.) It is encouraging that all methods are independent of this variable, since the i.i.d. methods were not accurate for large γ (see Section 5.3.2).

If the experiments for N, p and K are repeated with $\gamma = 0.1$, we find that the behaviour is generally consistent with that of i.i.d. noise, which was tested above. All five methods are independent of the number of bands and number of endmembers (except for HySime at $K = 18$, once again), and it is interesting to note that RMT in fact performs better with respect to N for non-i.i.d. noise. RMT with both noise approximations reached 100% success at only 2,000 pixels (rather than 6,000 in the i.i.d. case). With non-i.i.d. noise, HySime and NSP also overestimate K for $N < 2,000$, but have 100% success rates for larger values of N. It is encouraging that the non-i.i.d. results are at least as good as the i.i.d. results.

7.2.5. Performance with respect to flat spectra

In the standard synthetic dataset, flat spectra have been eliminated. Figure 3.1(a) shows that there are two spectra in the dataset that appear to be flat and are low amplitude. A lack of features means that the signal may not be identifiable. Figure 3.1(b) shows that Graphite is not only flat, it is also noisy.

When both of these spectra were included in the synthetic dataset, only Graphite (the noisy, low amplitude flat spectrum) was not detected. In other words, K is underestimated by one when Graphite is included in the dataset. The other flat spectra was not noisy and so was correctly identified as signal. This means that some noisy, low amplitude and flat spectra may not be identified. It is interesting to note that when the standard deviation of the noise is reduced so that $\sigma = 10^{-4}$, then Graphite is correctly identified as signal, so it is the relative noisiness of the flat signal that should be taken into consideration.

Note that this behaviour occurred both with i.i.d. and non-i.i.d. ($\gamma = 0.1$) noise.

7.2.6. Performance with respect to noise variance

While SNR is constantly improving in new hyperspectral sensors, some of the older satellite sensors still contain high levels of noise, so it is important to test the noise limit in the image for reliable determination of K. This is the first case where NSP and HySime perform worse than the RMT methods, in particular the latter (see Figure 7.2). RMT requires a noise estimate, and is tested with the real noise values and two noise approximations, and it is interesting that the results with approximated noise just are as good as the results with real noise. It is also encouraging that these results are better than those seen in Chapter 5.

When non-i.i.d. noise is considered, with $\gamma = 0.1$, the behaviour of NSP and HySime remains unchanged. RMT however, is more accurate for non-i.i.d. noise, with successes of 100% up to $\sigma = 0.05$ (compared to $\sigma = 0.01$ for i.i.d. noise).

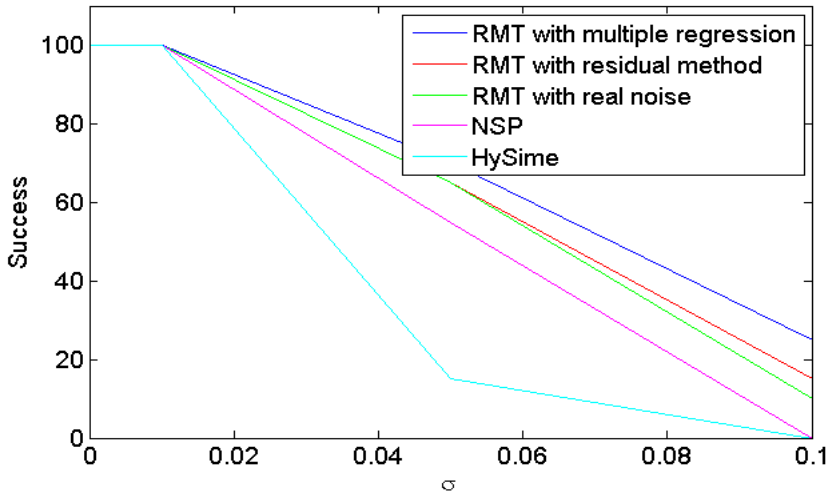


Figure 7.2.: The accuracy of RMT, NSP and HySime with respect to noise variance. RMT with approximated noise performs best for all noise approximations, followed by NSP and then HySime.

7.2.7. Performance based on accuracy of noise approximation

As previously stated, RMT does not depend on a specific noise approximation method. However, the method that is used must be reliable enough. Figure 7.3 shows that RMT only tolerates small underestimations, but overestimations of i.i.d. noise are well tolerated. NSP and HySime were also simulated with an error added to real noise values. These methods displayed the same asymmetry, although they were more tolerant of underestimation. NSP was the most robust, tolerating underestimations as high as 30%.

However, for slightly higher noise and K, say $\sigma = 2 \times 10^{-3}$, and $K = 12$, an overestimation of 10% results in underestimation of K for HySime (see Figure 7.4), and so for the best results, the noise approximation method should be chosen with care.

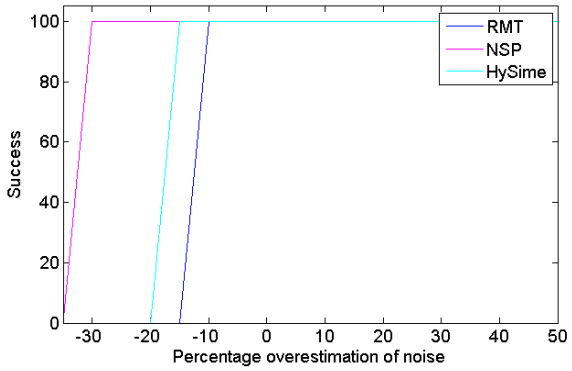


Figure 7.3.: The accuracy of RMT, NSP and HySime with respect to the accuracy of the noise estimation. If the correct noise standard deviation σ is in fact estimated at $\sigma + \delta$ (where δ is constant), then the results are given for the overestimation percentage. 20 different images per accuracy overestimation were simulated in order to calculate these success rates. RMT only tolerates small under-estimations of the noise, but is widely tolerant for over-estimation. The same pattern applies to NSP and HySime, although they are slightly more tolerant of underestimation.

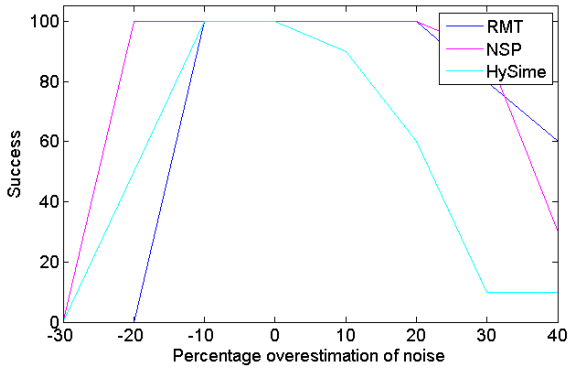


Figure 7.4.: The accuracy of RMT, NSP and HySime with respect to the accuracy of the noise estimation, for $K = 12$ and $\sigma = 2 \times 10^{-3}$. If the correct noise standard deviation σ is in fact estimated at $\sigma + \delta$ (where δ is constant), then the results are given for the overestimation percentage. 20 different images per accuracy overestimation were simulated in order to calculate these success rates. All methods are still more tolerant of overestimation, but there is a limit to this tolerance. RMT is least tolerant of underestimation, but most tolerant of overestimation.

7.2.8. Performance with respect to correlation

RMT and HySime calculated the ID with 100% accuracy when real noise values were used as an input, and when each band was correlated with its neighbour. In this test, suppose half the bands are correlated with their neighbours, so that bands $(j-1)$ and j are correlated, for $j = 2, 4, \ldots, p$ (for p assumed even). Then the noise is chosen from the multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & c\sigma_2^2 & 0 & \ldots & 0 \\ c\sigma_2^2 & \sigma_2^2 & 0 & & \\ 0 & & \ddots & & \vdots \\ \vdots & & 0 & \sigma_{(p-1)}^2 & c\sigma_p^2 \\ 0 & & 0 & c\sigma_p^2 & \sigma_p^2 \end{pmatrix}.$$

For the results above, $c = 0.5$. All methods resulted in overestimations of K when noise approximations were used (see Figure 7.5), due to the inaccuracy of these approximations in the presence of correlated bands (see Figure 4.15), and the sensitivity of all three methods to underestimations in noise (see Figure 7.3).

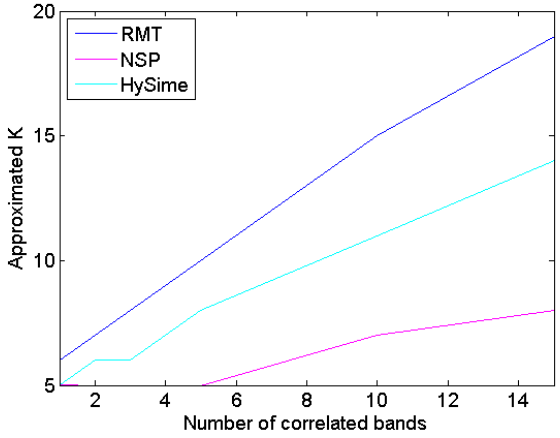


Figure 7.5.: The accuracy of RMT, NSP and HySime with respect to correlated noise. All three methods overestimate K for more than 10 correlated bands. NSP is the least affected, while RMT is almost linearly dependent on the number of correlated bands. The correct number for K is 5, so even a single correlated band results in overestimation of K for RMT.

If the correlations are reduced so that, for instance, $c = 0.1$ for all correlations, then K is accurately determined by all methods. Figure 7.6 shows the median K determined by all methods for 10 correlated bands with respect to c. This shows that higher degrees of correlation result in higher inaccuracies when determining the ID of a hyperspectral image, especially for RMT. These experiments show that highly correlated bands should be removed, but some correlation is tolerated.

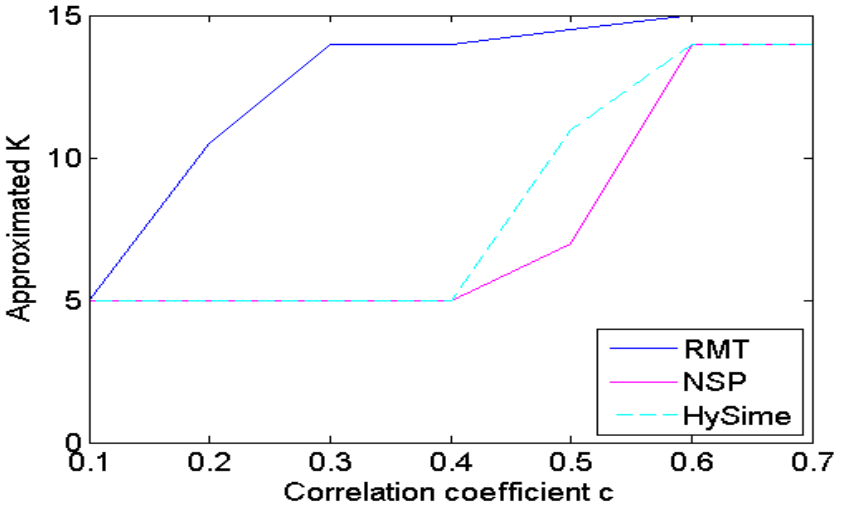


Figure 7.6.: The performance of RMT, NSP and HySime with respect to the level of correlated noise. This graph shows the median K value over 20 experiments for each method, in the presence of ten correlated bands. Any correlation $c > 0.1$ results in an overestimation of K for RMT, and correlation $c > 0.4$ results in an overestimation of K for NSP and HySime. The correct value should be $K = 5$.

7.2.9. Performance with respect to outliers and rare substances

Certain rare substances may be present in only a single pixel in the image. They may even be present only in a mixture in a single pixel. It is often important that such substances are detected, especially in images acquired by satellite sensors such as Hyperion that have a 30 m spatial resolution. In order to test the sensitivity of RMT, NSP and HySime to rare substances, the noise is assumed known in each method, and the fraction of rare substance present in a pixel is decreased. Figure 7.7 shows that RMT is most likely to detect substances that make up only a fraction of a pixel, by determining the correct K, and HySime is least likely. All methods detect substances that make up at least one pixel in the image at the noise level $\sigma = 10^{-3}$, but at higher noise levels the success rates drop for all methods.

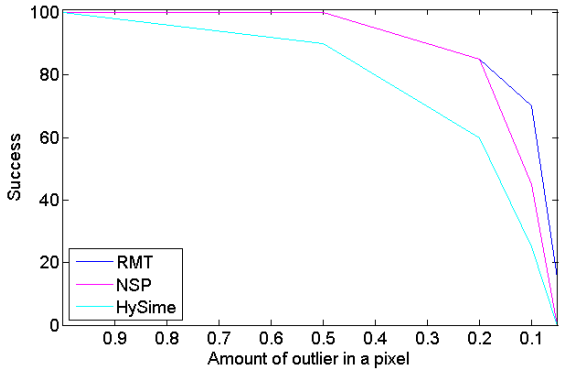


Figure 7.7.: The accuracy of RMT, NSP and HySime with respect to rare substances. RMT is the most robust in terms of rare substances, and HySime is the least likely to detect substances that make up a fraction of a pixel.

7.2.10. Analysis of subspaces

HySime, NSP and RMT all claim to produce a signal subspace. This section tests if these three methods produce the same or similar subspaces in the case where RMT and HySime produce the identical intrinsic dimension (recall that NSP consistently underestimates the ID by one). The basis vectors for the NSP subspace are the eigenvectors from the whitened observation covariance matrix whose corresponding eigenvalues satisfy the signal evaluation; the basis vectors for the HySime subspace are the eigenvectors of the signal covariance matrix (calculated by subtracting the noise from each pixel) whose corresponding eigenvalues satisfy the signal evaluation; and the basis vectors for RMT are the eigenvectors of the signal covariance matrix (calculated by subtracting the noise covariance from the observation covariance) whose corresponding eigenvalues satisfy the signal evaluation.

RMT and HySime both look at eigenvalues from the signal covariance matrices, although each method calculates this matrix in a slightly different way. Both methods are equivalent if the noise is independent from the signal. In this test a synthetic dataset was simulated with 10,000 pixels, 200 bands, $\sigma = 10^{-3}$ and $K = 5$. For all three methods real noise values were used instead of approximated noise. As discussed in previous simulations, NSP consistently underestimated K by one, and so it produced one less eigenvector than the other two methods. The basis vectors are expected to be different as the signal is said to be contained in a subspace of smaller dimension. To evaluate whether this lower dimensional subspace is contained within the RMT subspace, the Matlab function *subspace* was used, which calculates the angle between two subspaces. With real noise values, the angle was exactly zero, meaning that the NSP subspace is contained within the RMT and HySime subspaces. In this test, HySime and RMT produced exactly the same basis vectors, as seen in Figure 7.8.

When using approximated noise, the angle between the RMT and HySime subspaces is 2.4×10^{-6}, which means that the subspaces are almost exactly linearly dependent (essentially the same). The angle between the RMT and NSP subspaces is 0.02 which is also small.

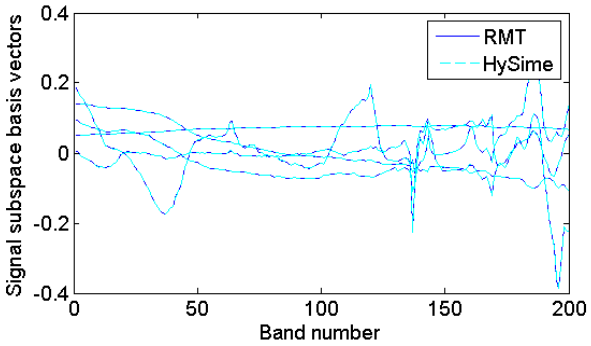


Figure 7.8.: The signal subspace basis vectors produced by HySime and RMT. Both methods produced exactly the same basis vectors.

7.2.11. Computation time

The time taken to compute the intrinsic dimension is an important consideration when choosing a method. To evaluate this, RMT, HySime and NSP, are tested assuming real noise values (i.e. leaving the noise computation out of the time) on Cuprite (AVIRIS). The test was done on an HP notebook with Intel Core2Duo 1.83 GHz CPU and 2GB RAM. The results are displayed in seconds in Table 7.1.

Table 7.1.: The time in seconds taken to calculate the ID of Cuprite (AVIRIS). NSP and RMT took almost exactly the same amount of time, and HySime was the fastest method. In this calculation, the time taken to calculate the noise was excluded.

HySime	NSP	RMT
16.18	29.68	29.32

RMT and NSP completed in almost exactly the same amount of time. HySime completed in approximately half the amount of time when compared to the other methods, but remember that HySime is dependent on a noise method that calculates the noise on a per pixel basis, which is more time consuming. If the noise approximation is included into the equation, HySime takes 42.28 seconds, whereas NSP takes only 29.69 seconds. RMT may be the fastest or slowest method, depending on the noise approximation used.

7.3. Discussion

The effect of non-linearities was also tested on RMT, NSP and HySime. A synthetic dataset was simulated with three endmembers. Two of these are real spectra, and the third is created by multiplying the components of the first two spectra in a pairwise fashion. In this simple test, all methods accurately determined that $K = 3$, for the same limits as linear data. This means that if a scene contains non-linearities, more endmembers may be detected than the number of "pure substances", but this higher number will be necessary to correctly unmix the image.

All methods performed well in the synthetic tests. In fact, for reasonable image sizes, all methods had maximum accuracy for all tested values of N and p. All methods were accurate to high noise levels, but HySime performed worst, followed by NSP. It is interesting to note that RMT performed better with approximated noise than real noise. However, these approximations (regression and residual methods) yielded the highest K values in real images when compared to other methods. As has been seen in Section 4.3, the regression and residual methods are unable to accurately estimate the noise in HCI and Cuprite (SpecTIR), due to the high levels of correlation in the data. RMT with Meer's method has performed consistently well, and is the only noise approximation method investigated that is able to process HCI and SpecTIR, as it does not assume uncorrelated noise.

RMT is most sensitive to noise underestimation and therefore to correlated bands, but it is most robust in terms of rare substances and has the fastest computation time when the residual noise approximation method is used. RMT also performed better in high noise environments, and was comparable to the well known methods NSP and HySime in most other tests. An advantage of RMT is that it is not dependent on a specific noise approximation method.

8. Results and Validation

RMT adapted for non-i.i.d. and correlated noise performed well on synthetic images, as did NSP and HySime. Now these methods will be tested on real images.

The results for all real images (using all spectral bands) are shown in Table 8.1. HySime almost always produced the lowest values for K, which is consistent with the synthetic experiments in Chapter 7, where HySime was least likely to identify rare substances, especially if such a substance makes up less than a pixel, or if the noise in the image is high. In SpecTIR however, HySime produced results much higher than expected, which may be seen especially in Cuprite ($K = 140$), where we expect values for K that are less than 30, according to previous studies on a larger subset of the area [10]. Although synthetic tests in Section 7.2.8 showed that HySime was able to tolerate some correlation between bands, high enough correlation would result in underestimation of the noise, which would explain the high values for K in these images.

RMT with Meer's method for noise approximation most often displayed values for K which are much smaller than the values determined when using the multiple regression and residual noise approximation methods. Synthetic experiments in Section 4.6.1 have shown that the latter two methods do not approximate the noise accurately where the noise is correlated between bands, and so RMT with Meer's noise approximation is likely to be more accurate in this case. There are however, certain images, such as Moffet Fields (acquired by both AVIRIS and Hyperion) where Meer's noise approximation yields values for K that are higher than all other approximations. In these images, there may not be homogeneous regions large enough for Meer's method to accurately estimate the noise.

NSP estimates $K = 28$ for Cuprite (AVIRIS), which is consistent with published results for the same image. The other Cuprite scenes are largely overestimated when using NSP, once again pointing to problems with correlation. None of the methods investigated have performed satisfactorily, and the effects of bad bands and correlation in each image is investigated in more detail in Sections 8.1 and 8.2.

Table 8.1.: The ID of all real datasets as determined by RMT, NSP and HySime.

	$RMT_{Residual}$	$RMT_{Regression}$	RMT_{Meer}	NSP	HySime
Hyperion:					
Cuprite	80	80	78	71	34
Lunar Lakes	82	82	38	57	18
Moffet Field	84	84	114	78	47
AVIRIS:					
Cuprite	31	30	21	28	15
Lunar Lakes	41	41	16	32	21
Moffet Field	62	62	173	55	32
SpecTIR:					
Cuprite	156	156	24	139	140
Suwannee	128	128	69	105	105
CAO:					
Scene 1	19	19	9	14	4
Scene 2	15	15	10	11	2
Scene 3	15	15	6	11	4
HCI	82	82	6	66	59

8.1. Removing bad bands

Wu *et al.* [10] and Chang et al. [9] remove certain bands from Cuprite (AVIRIS) (bands 1–3, 105–115 and 150–170) because of water absorptions and low SNR in these bands. The AVIRIS Moffet fields image also had these bands removed. Lunar lakes contains all original bands, and if we examine the bands that were removed from the other AVIRIS images, particularly the latter two band segments, the variance of the band was low, approximately 10^{-5}. Statistical noise techniques perform badly in the presence of these bands, and so these will be removed for all images. In Hyperion, we notice that certain bands corresponding to the frequencies of the AVIRIS bad bands were visibly noisy, and displayed artifacts, seen in Figure 8.1.

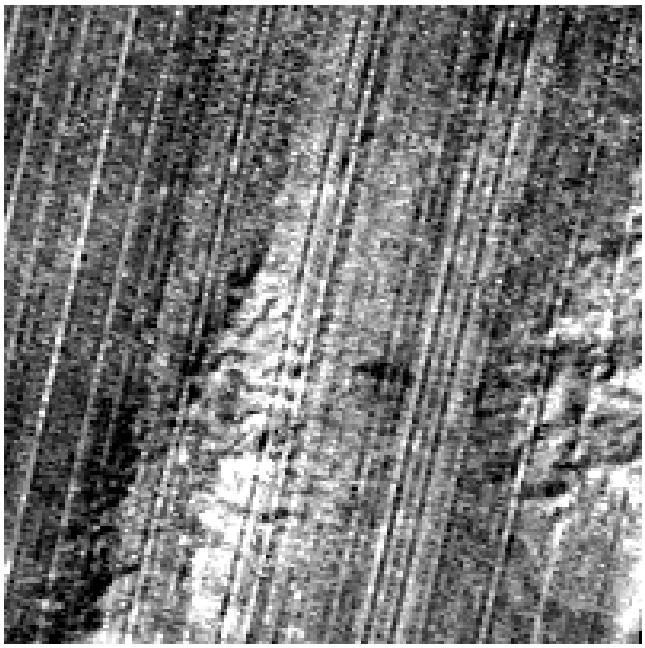


Figure 8.1.: Band 128 of Cuprite (Hyperion) appears to be noisy with spatial artifacts. The inclusion of bands such as these may adversely effect the estimation of noise in the image.

SpecTIR, CAO and HCI did not contain bad bands by our definition, and AVIRIS Cuprite and Moffet fields had already had their bad bands removed. So the only images affected by the removal of bad bands are the Hyperion images and AVIRIS Lunar Lakes. For Lunar Lakes we removed bands 1–3, 105–115 and 150–170 (as for Cuprite and Moffet Fields), and for the Hyperion images we removed bands 85–100 and 130–147. The ID

Table 8.2.: The ID of all real datasets as determined by RMT, NSP and HySime, where Hyperion and AVIRIS have had bad bands removed.

	$RMT_{Residual}$	$RMT_{Regression}$	RMT_{Meer}	NSP	HySime
Hyperion:					
Cuprite	62	62	62	54	23
Lunar Lakes	64	63	24	41	16
Moffet Field	66	66	99	60	35
AVIRIS:					
Cuprite	31	30	21	28	15
Lunar Lakes	37	37	13	29	20
Moffet Field	62	62	173	55	32
SpecTIR:					
Cuprite	156	156	24	139	140
Suwannee	128	128	69	105	105
CAO:					
Scene 1	19	19	9	14	4
Scene 2	15	15	10	11	2
Scene 3	15	15	6	11	4
HCI	82	82	6	66	59

of these images was reduced by this process, but the results are still inconsistent, and in particular the Hyperion Cuprite ID values are too large for all except HySime. Correlated noise may cause such overestimations, as seen in the synthetic tests in Section 7.2.8.

8.2. Correlations

Correlations between bands have been shown to adversely affect the calculation of the intrinsic dimension of the image. To evaluate the correlation in each image, the partial correlation coefficient of each band, relative to every other band, is considered. These coefficients are represented in a matrix, where the coeffient displayed at row i and column j represents the partial correlation between band i and band j, with the rest of the bands held fixed. This is calculated using Matlab's built-in function *partialcorr*. This method calculates the correlation within the entire image (signal and noise), and we assume that because scenes are considered that contain the same materials (e.g. Cuprite acquired with Hyperion, AVIRIS and SpecTIR) that the signal correlation is constant. This will be seen in Figure 8.7 later on in this Chapter. The *partialcorr* function does not remove correlation; it simply identifies high correlation bands. This is a time consuming calculation, and tests have shown that the correlation may be approximated by selecting a random sample of pixels in the image. (This sample must be at least larger than p, the number of spectral bands, and for the matrices shown, the sample used was larger than $2p$.)

Correlation may be caused by sensor characteristics such as narrow spectral bands and spectral overlap, or by preprocessing procedures. The high levels of correlation seen in Hyperion are not due to bandwidth or spectral overlap, since AVIRIS has the same characteristics as Hyperion, with a maximum spectral overlap of 13.6% and an approximate bandwidth of 10 nm [79]. AVIRIS does not show high levels of correlation, and so we conclude that the correlation is due to another sensor characteristic or preprocessing procedures.

8.2.1. Correlation in Hyperion

Hyperion is a pushbroom sensor, and such sensors have been known to suffer from spectral effects such as spectral smile (see Chapter 2). Hyperion is also the only satellite sensor considered in this study, and some of the preprocessing techniques (such as spectral smile correction, etc.) may lead to the correlation seen in Figure 8.2. The high levels of correlation between neighbouring bands are seen in all three images. This correlation may explain the high values for K seen in Table 8.1.

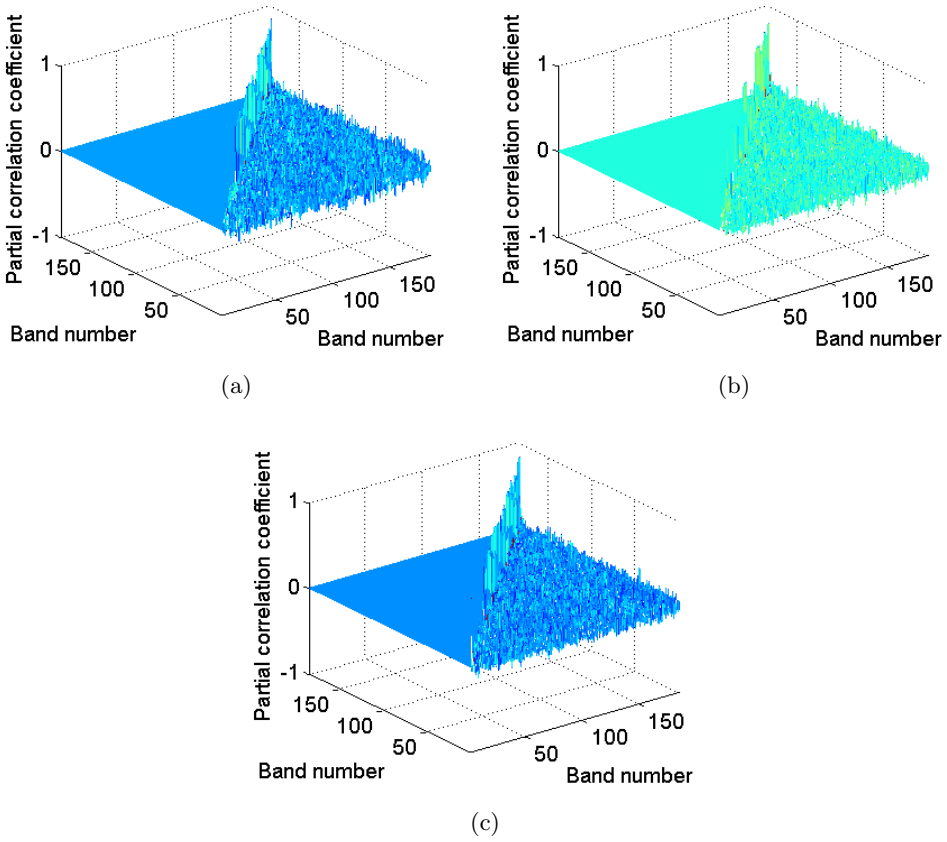


Figure 8.2.: The partial correlation coefficients for Hyperion 8.2(a) Cuprite, 8.2(b) Lunar Lakes, and 8.2(c) Moffet Fields. Cuprite and Lunar Lakes show similar high correlation graphs, and Moffet Fields shows even stronger positive correlations between neighbouring bands.

8.2.2. Correlation in AVIRIS

AVIRIS is a whiskbroom sensor and has been used to test many methods for determining the ID of a hyperspectral image, as the correlations between bands are low. As discussed above, AVIRIS and Hyperion share approximately the same spectral overlap and bandwidth, so the correlation seen in Hyperion is not due to either of these properties. The correlation graph for Cuprite, shown in Figure 8.3 may be slightly different from the other two since it was acquired in 1997, whereas Lunar Lakes and Moffet Field were acquired in 2009 and 2008 respectively.

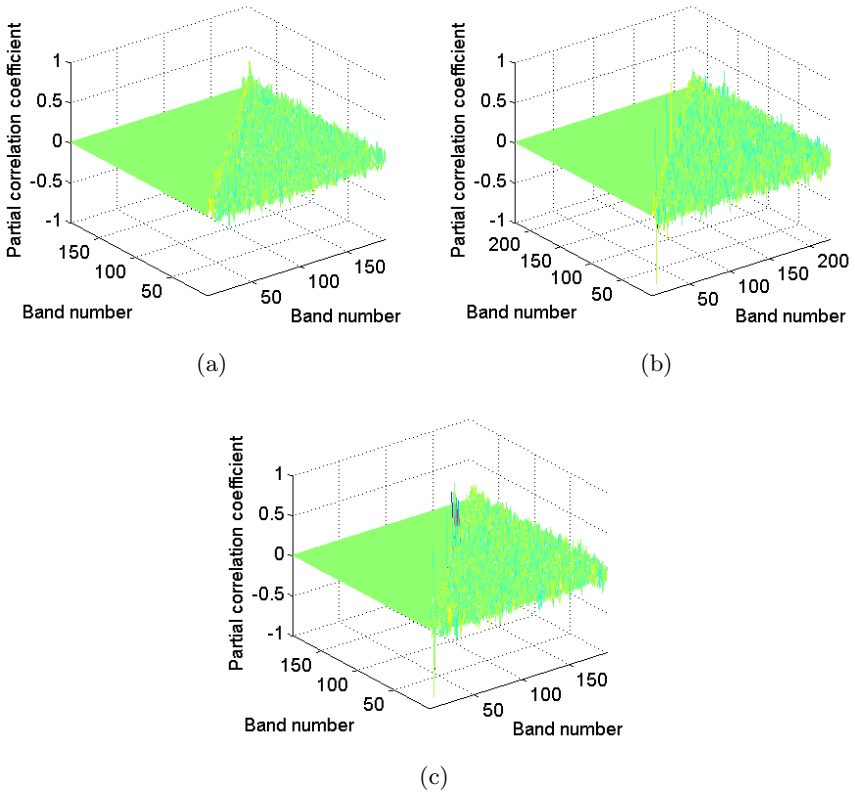
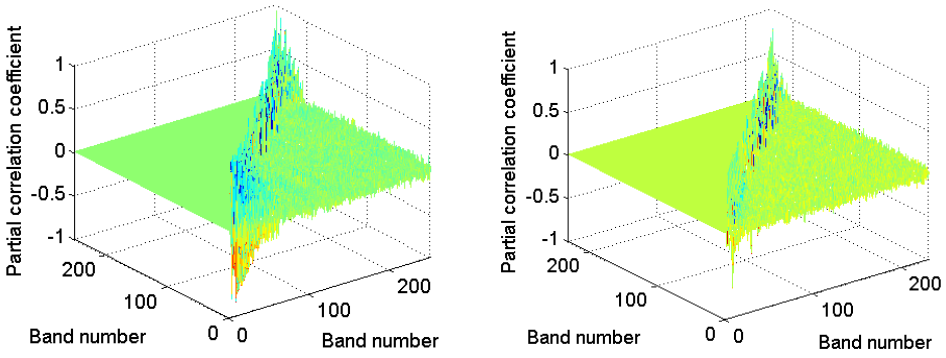


Figure 8.3.: The partial correlation coefficients for AVIRIS 8.3(a) Cuprite, 8.3(b) Lunar Lakes, and 8.3(c) Moffet Fields. All three images show low levels of correlation between bands, with Lunar Lakes and Moffet Field showing a small number of correlated bands.

8.2.3. Correlation in SpecTIR

SpecTIR is a pushbroom sensor and shows very high levels of correlation between neighbouring bands, illustrated in Figure 8.4. The Suwannee scene has lower levels of correlation, and has lower spatial resolution (2 m compared with Cuprite's 1 m), since the aircraft flew at higher altitude. It is interesting that SpecTIR displays high levels of positive and negative correlation, whereas Hyperion displays mostly positive correlation. This negative correlation may be due to pre-processing techniques, such as noise reduction methods.



(a) (b)

Figure 8.4.: The partial correlation coefficients for SpecTIR 8.4(a) Cuprite and 8.4(b) Suwannee. Both images show high levels of positive and negative correlation between neighbouring bands.

8.2.4. Correlation in CAO

CAO is also a pushbroom sensor, but the bands are resampled from 288 to 72 to achieve better SNR and higher spatial resolution. The resulting data shows very low levels of correlation, illustrated in Figure 8.5.

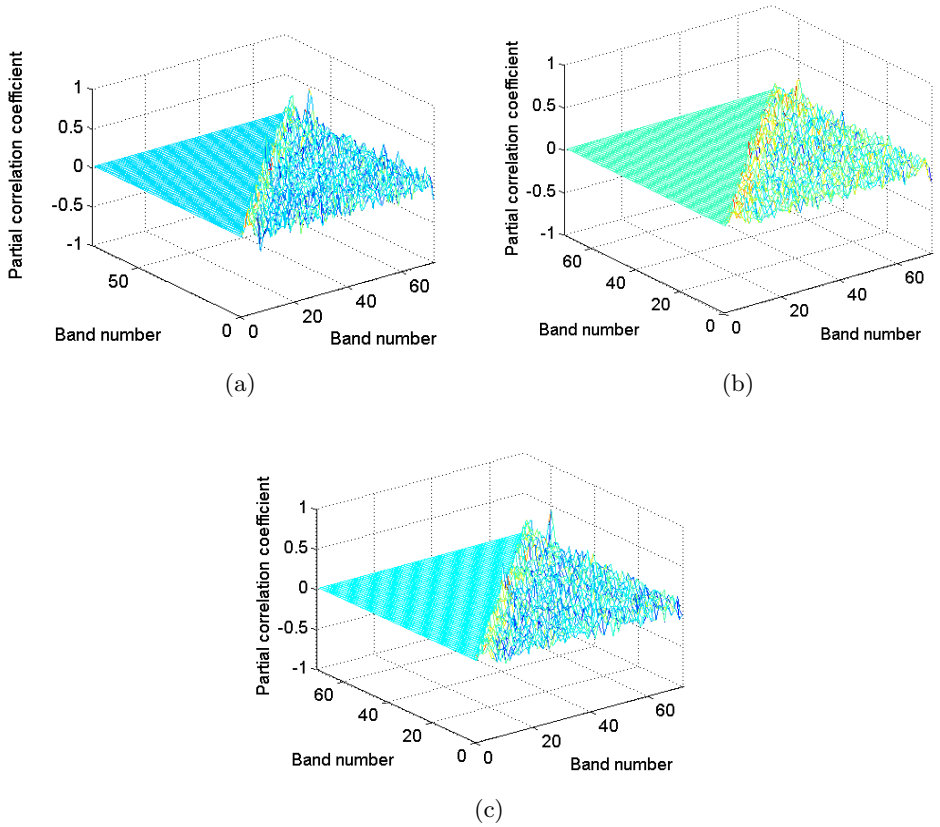


Figure 8.5.: The partial correlation coefficients for CAO 8.5(a) scene 1, 8.5(b) scene 2, and 8.5(c) scene 3. All three images show very low levels of correlation.

8.2.5. Correlation in HCI

HCI is a whiskbroom sensor and is the only such sensor to display high levels of correlation as seen in Figure 8.6. There is high positive correlation between neighbouring bands, which is the cause of the high estimates for K that may be seen in Table 8.1. RMT with Meer's method for noise approximation is the only method that estimates $K = 6$, which corresponds with experts' opinions of the intrinsic dimension of the dataset. Meer's method is the only noise approximation method that is not affected by correlation between bands.

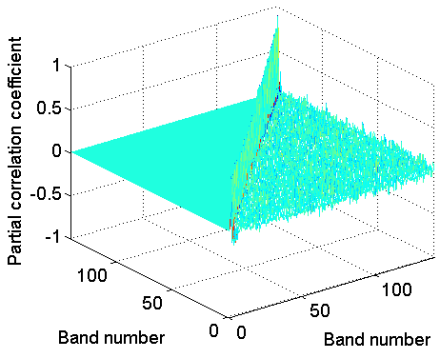


Figure 8.6.: The partial correlation coefficients for HCI. Neighbouring bands are very highly correlated in this image. In fact, there does not appear to be a single band that is not correlated with its neighbour.

8.3. Removing the effects of correlation

The optimal solution to the problem of correlation is to use a noise approximation technique that does not require uncorrelated bands. Meer's method is one such method, and the intrinsic dimension estimated by RMT with Meer's noise approximation for Cuprite (AVIRIS and SpecTIR) and HCI corresponds to previous studies and experts' opinions of these datasets. Unfortunately, Meer's method is not always able to reliably estimate the noise, as seen in Moffet Fields (AVIRIS and Hyperion).

When bands are strongly correlated, no extra information is gained, and so data will not be lost by removing correlated bands [1]. Correlated bands are removed by iteratively removing rows in the partial correlation matrix that contain the highest (or lowest negative) partial correlation coefficients, until the largest absolute correlation coefficent does not exceed a certain threshold. Since the noise approximations are more accurate with many bands, as few bands as possible must be removed. Conversely, the presence of correlation results in inaccurate noise estimation. Testing in synthetic sets showed that a reasonable threshold of the partial correlation coefficients is 0.3.

This threshold may also be visually appreciated in Figure 8.7, where the partial correlation coefficients of all three Cuprite scenes are displayed as a scatter plot.

Correlated bands are removed according to the threshold of 0.3, and the intrinsic dimension of these spectral subsets are shown in Table 8.3 for all real images. RMT with the multiple regression noise approximation is not displayed in this table since the ID calculated is mostly identical to RMT with the residual noise approximation. Note that nearly two thirds of the bands were removed for SpecTIR and HCI, and approximately 70 out of 184 bands were removed from Hyperion, whereas AVIRIS and CAO retained most of their bands. The results shown in Table 8.3 are much more consistent with known results than Table 8.1, especially for SpecTIR, where all methods (except with Meer's noise approximation) had previously over-estimated K. When correlated bands have been removed, RMT with residual noise approximation, NSP and HySime all provide comparable results for SpecTIR, and all are similar to published results in the larger Cuprite area ($22 \leq K \leq 28$). Overall, RMT (with residual noise approximation) estimates higher K than NSP and HySime, and HySime estimates are lower than all other methods.

Now the RMT estimates (using the residual noise approximation) for all Cuprite scenes produce values between 22 and 29, which is close to the range of 22 to 28 that was calculated in Wu *et al.* [10].

Another method for handling correlated noise is to only remove those bands in the

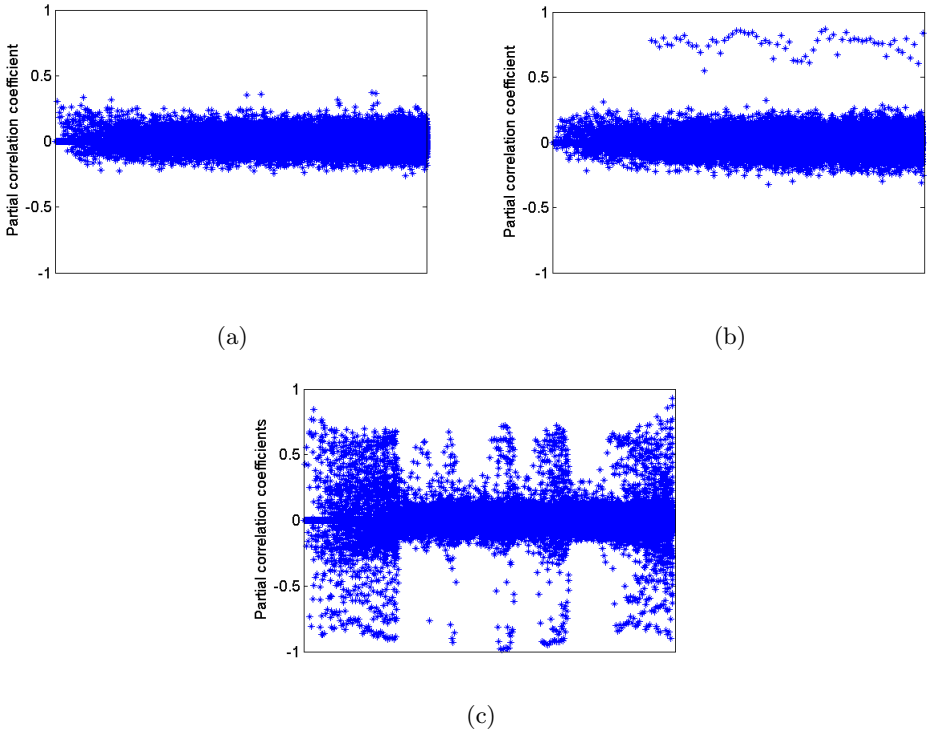


Figure 8.7.: The partial correlation coefficients are displayed as a scatter plot for Cuprite 8.7(a) AVIRIS, 8.7(b) Hyperion, and 8.7(c) SpecTIR. The AVIRIS image has low levels of correlation, Hyperion has few highly correlated bands, and SpecTIR displays many correlated bands.

Table 8.3.: RMT, SNR and HySime are applied to five real images, where correlated and bad bands have been removed. Correlated bands were identified as those bands that contained partial correlation values of more than 0.3 when compared to all remaining bands. Bad bands are those with a normalised variance less than 10^{-5}.

	Bands used (Total bands)	RMT_{Residual}	RMT_{Meer}	NSP	HySime
Hyperion:					
Cuprite	98(184)	23	47	13	10
Lunar Lakes	97(184)	29	15	14	9
Moffet Field	96(184)	21	92	16	12
AVIRIS:					
Cuprite	181(189)	29	20	26	17
Lunar Lakes	169(224)	30	13	23	19
Moffet Field	171(189)	55	157	45	22
SpecTIR:					
Cuprite	74(250)	22	11	21	17
Suwannee	85(230)	33	15	29	25
CAO:					
Scene 1	61(72)	16	8	12	9
Scene 2	69(72)	14	8	11	5
Scene 3	68(72)	13	5	10	7
HCI	54(141)	18	5	12	12

noise approximation, and apply this to the entire image. To do this, multiple regression theory is used on a band by band basis, where the noise in band i is calculated based on a regression with only those bands least correlated with band i. Specifically, in row i, we use only those bands corresponding to entries in row i with partial correlation values less than 0.1. This threshold was determined by synthetic tests.

Since the most correlated bands are often neighbouring bands, this method often has to estimate the noise in band i by using bands quite far away from band i, which may result in unreliable estimates. The correlation removal noise approximation method was variable in its success for synthetic data, sometimes overestimating and sometimes underestimating the noise in each band. The mean noise approximation however, was accurate even in the presence of correlated noise. This noise approximation method is then an i.i.d. method and such methods do not work well in the presence of non-i.i.d. noise. While the idea behind this method showed promise, the results were not consistent enough for this application.

The scenes from different sensors over Cuprite, Lunar Lakes and Moffet Fields that have been considered so far are over the same region but not necessarily the same spatial areas. More comparable results will be expected when the subsets are the same, and this will be considered in Section 8.6. First, some of the ID estimates will be used in the unmixing of an image.

8.4. Testing results by unmixing the image

Figure 8.8.: An image of CAO Scene 1. The RGB image is used for this illustration, with red approximated by band 35, green approximated by band 16, and blue approximated by band 10.

The results displayed in Tables 8.1-8.3 show that the ID estimations of different methods may vary. This section shows the unmixing results when using each ID estimate for the image CAO1. Note that it is not possible for this section to determine which ID estimate is correct, and is for illustrative purposes only. The purpose of a robust method for determining the ID is to avoid "trial and error" methods where the ID is estimated by testing the unmixing for different values of K until a satisfactory unmixing result is reached. Such an estimate is subjective and dependent on the requirements of the user, and does not necessarily produce the intrinsic dimension of the image.

The ID of CAO Scene 1 (Figure 8.8) has varying approximations, with values between 4 and 17 (Table 8.1). This image had low levels of correlation, and so the correlation thresholding method will not be used. This image will be unmixed and analysed.

The Matlab code for N-FINDR was written by the Neural Networks and Signal Processing group (GRNPS), Computer Science Department, University of Extremadura in Spain, and this group chose to use PCA for the dimension reduction. The Matlab code

for VCA was provided by Nascimento and Bioucas-Dias, and is available online[1]. The C++ code for PSO was written by Frans van den Bergh.

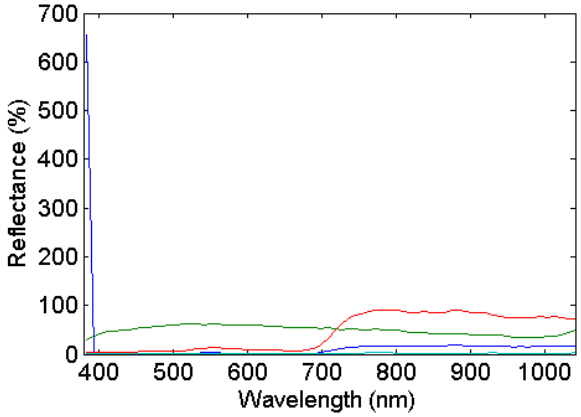


Figure 8.9.: The endmembers of CAO scene 1 with $K = 4$. The endmember depicted in blue appears to be an outlier due to its value in the first band, which is much higher than values seen in the rest of the image. This may be due to sensor error since the reflectance is well over 100%.

Using NFINDR to determine the endmembers may result in outliers becoming the simplex vertices, and this is the case in the CAO scene 1. Figure 8.9 shows when K=4 one endmember may be an outlier, and further examination of the image shows that in band 1 several pixels are widely separated from the rest of the pixel values in that band. This behaviour appears to be unique to band 1. While this may be corrected with preprocessing and the identification of outliers is important in target detection, HySime in particular has been shown to be sensitive to outliers, so this band is removed for this exercise. When the first band has been removed, NSP remains unchanged at $K = 14$, and surprisingly, HySime now gives $K = 10$, which is much closer to the results of the other methods. RMT with the residual noise approximation method behaves as expected, now yielding $K = 18$, as does RMT with Meer's noise approximation, which now gives $K = 8$.

To summarise, the following values will be tested on the original image without the first band: 8 (RMT - Meer), 10 (HySime), 14 (NSP), 18 (RMT - Residual).

The endmembers were determined by NFINDR, SGA and VCA. NDFINDR and SGA (both simplex methods) produced very similar endmembers, while VCA (a projection method) produced noticeably different endmembers. In general, NFINDR will be used

[1]http://www.deetc.isel.ipl.pt/jnascimento/public.html

to determine the endmember spectra, and will be compared to the VCA results for the HySime ID of K=10 (VCA is written by the same authors as HySime).

PSO is then used to estimate the endmember abundances, which are the fractions of a particular endmember present in each pixel. None of the methods that determine the ID use the spatial information of the pixels, and especially in images such as CAO, with 1 m spatial resolution, we expect objects on the ground to form congruent clumps, clusters or small homogeneous regions in the image. This will result in similar abundances for neighbouring pixels. If the ID is underestimated, then different objects in the image that may be separable by eye will be characterised by a single endmember, and if the ID is overestimated, then single classes may be separated. Both of these scenarios are difficult to see in the abundance images shown below, hence the need for an objective method for determining the ID.

8.4.1. Unmixing with $K = 8$

Figure 8.11 shows that all abundance maps show spatially congruent clumps corresponding to objects on the ground for $K = 8$. This is encouraging, since spatial information is not used at any stage of the calculation. Figure 8.10 shows distinct endmembers. Endmembers (d) and (e) have similar spectral signatures, but Figure 8.11 shows that these correspond to different targets in the image, for instance (e) represents mostly trees, while (d) represents mostly small trees (possibly a different species) and green grass. This means that $K = 8$ is visually a reasonable number for the ID in the image. Figure 8.12 removes the low abundances for clarity and shows only those abundances that are over 35% in each pixel. These show spatially congruent regions of higher abundance pixels, even though some are small. Figure 8.12(h) however, shows that the abundances associated with the endmember of lowest amplitute make up a large part of many pixels, and this suggests that it may be possible to divide the classes further.

Note that Figures 8.12 (c) and (f) show some overlapping areas over 80%. This means that the sum to one condition may not be true in some cases, but this problem exists only for a small percentage of pixels. Also, Figure 8.10 shows some vegetation spectra with very high reflectance values, over 70%. Recall that NFINDR is likely to choose extreme points in the image to represent endmembers, and it is encouraging to note that these vegetation spectra show clear unmixing graphs in Figures 8.11 and 8.12.

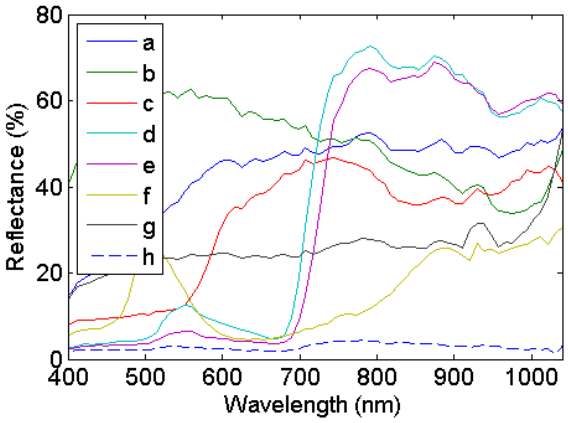


Figure 8.10.: The endmember spectra for CAO scene 1 with 8 endmembers. All spectra are distinct, with none appearing to behave like noise.

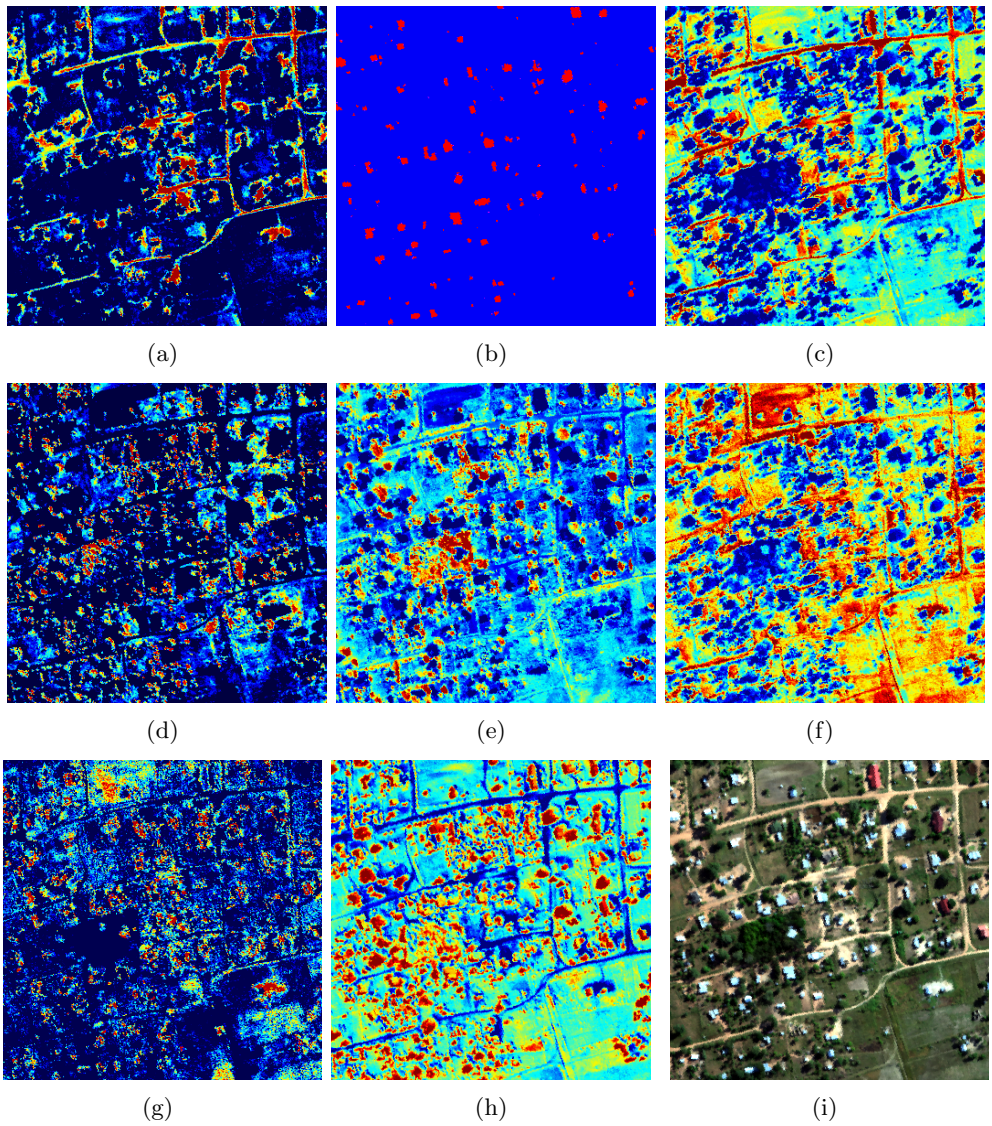


Figure 8.11.: The abundance maps for CAO scene 1 with 8 endmembers, and the colour image for comparison in (i). Blue areas represent zero abundance, whereas dark red represents almost pure pixels, and all maps show clusters of pixels with similar abundance values.

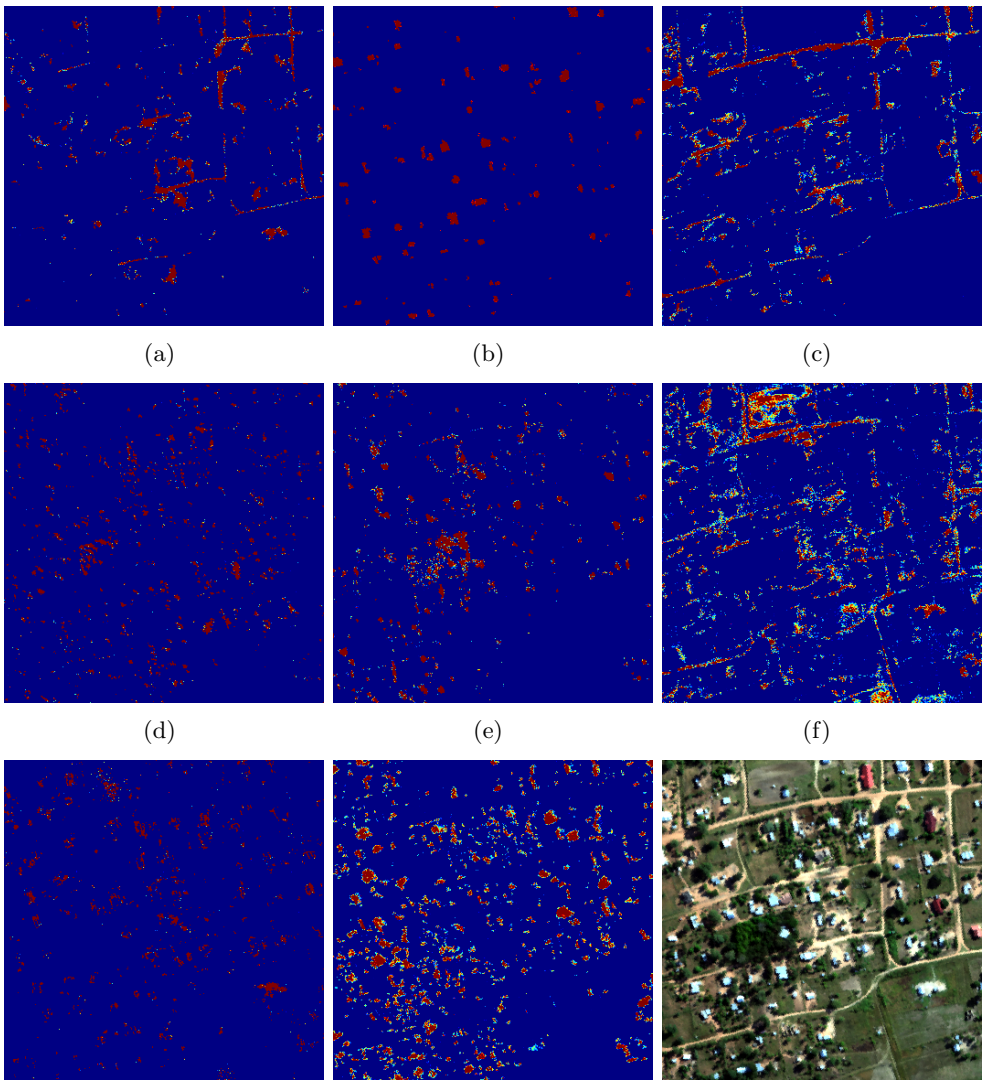


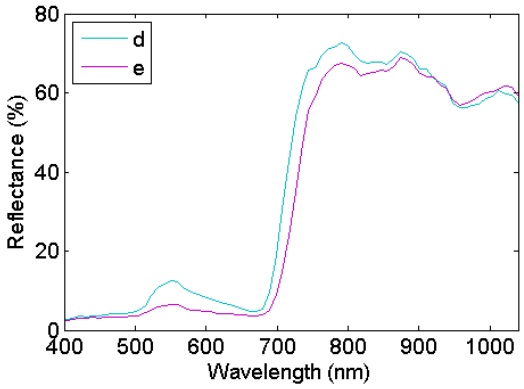
Figure 8.12.: The abundance maps for CAO scene 1 with 8 endmembers, and the colour image for comparison in (i). To remove pixels with low abundances for clarity, values are only shown when the abundance of the endmember is more than 80%.

8.4.2. Unmixing with $K = 10$

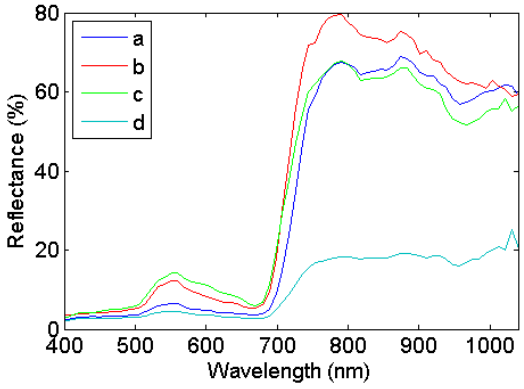
Unmixing with $K = 10$ gives similar endmembers to $K = 8$, when using NFINDR. The two new endmembers are formed by adding a third vector to the cluster of vegetation spectra, and a new endmember is introduced, as shown in Figure 8.13. All the endmembers are also distinct for $K = 10$. The abundance maps for the four new spectra shown in Figure 8.13 are displayed in Figure 8.14. All maps show congruent clumps of similar abundances and the three vegetation spectra map different objects in the image. This means that $K = 10$ is visually more likely the ID of the image than $K = 8$, and it may be possible to divide the endmembers yet further.

The thresholded abundance maps are shown in Figure 8.15 for the new endmembers. The areas of high abundances form congruent classes, even though some are small. For $K = 8$ the abundances associated with the low abundance endmember were mostly large, seen in Figure 8.12 (h). For $K = 10$, the abundances associated with the low abundance endmember now appear to form a well defined class. This is evidence that $K = 10$ provides a better unmixing than $K = 8$, although this judgment is subjective.

When VCA is used to determine the ten endmembers, all vegetation remain in the same class, and low amplitude spectra are separated, as seen in Figure 8.17. The abundances for the three low amplitude spectra are shown in Figure 8.18, and may correspond to shadowed areas. These three abundances do not show the same levels of homogeneity seen in the abundances where NFINDR was used, and in general, VCA produces abundances with lower accuracies for this scene.



(a)



(b)

Figure 8.13.: The difference in endmembers for CAO scene 1 with $K = 8$ (8.13(a)) and $K = 10$ (8.13(b)). Visually similar spectra have been removed. The effect of increasing the estimated ID by two is that the vegetation class previously represented by 2 spectra is now represented by 3 spectra, and a new endmember is introduced.

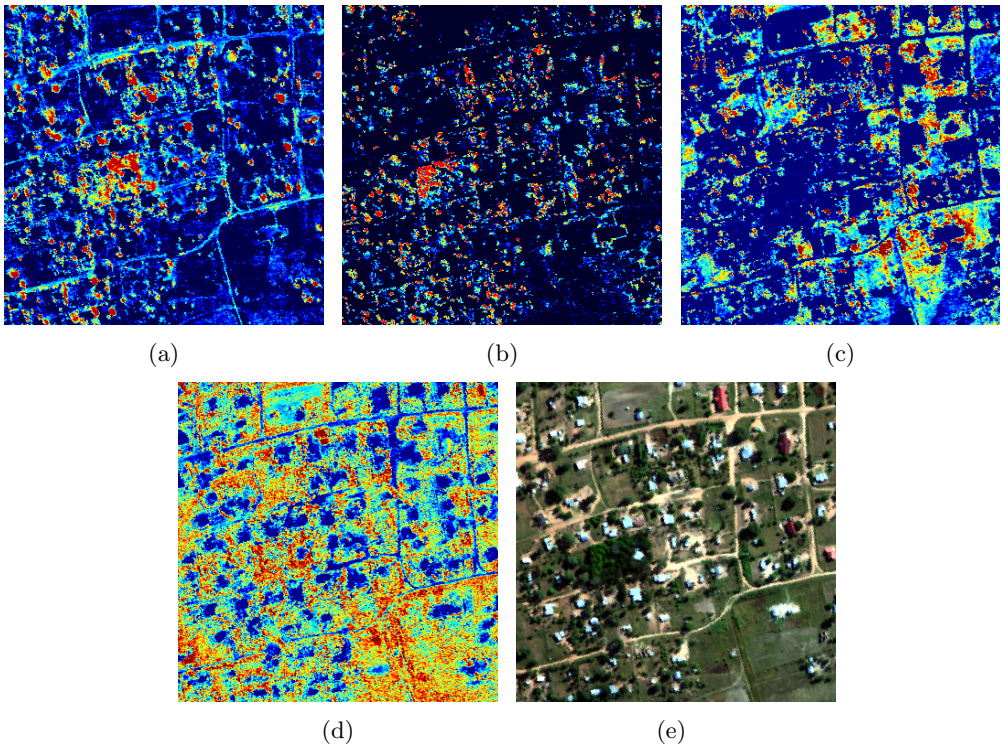


Figure 8.14.: The abundance maps of new endmembers for CAO scene 1 with $K = 10$, and the color image for comparison in (e). Blue areas represent zero abundance, whereas dark red represents almost pure pixels. Figures 8.14(a)–8.14(c) correspond to the clustered group of vegetation spectra and are seen to map different objects in the image. Figure 8.14(d) corresponds to the new endmember when compared with the $K = 8$ unmixing. These abundances are more convincing than those estimated for $K = 8$.

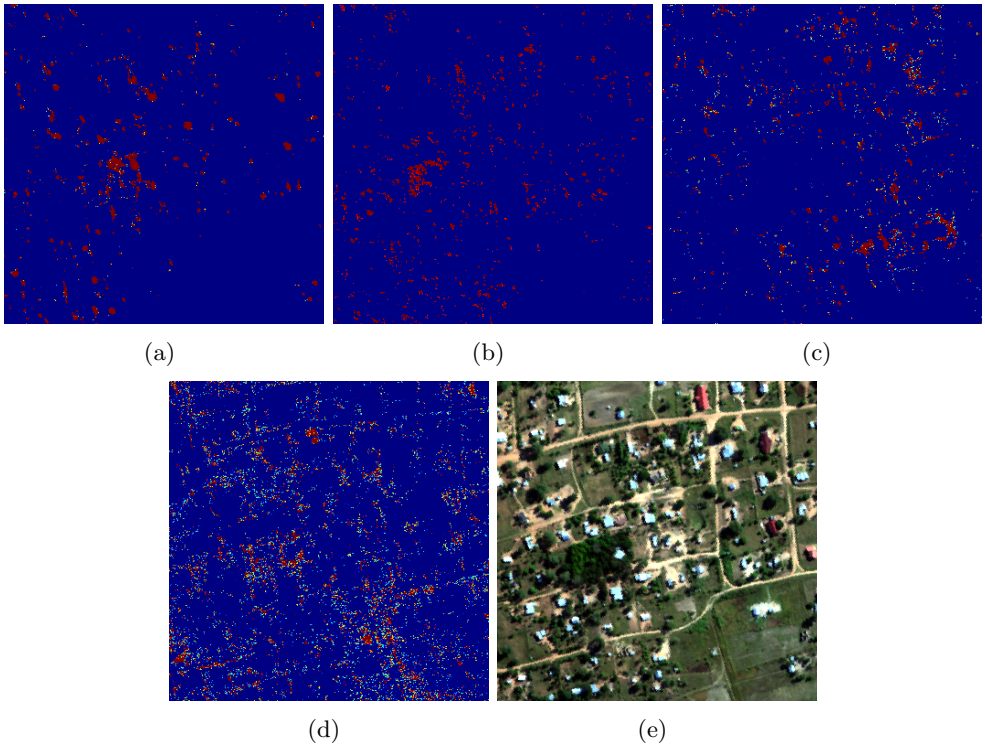


Figure 8.15.: The abundance maps for CAO scene 1 with 10 endmembers, and the colour image for comparison in (e). To remove small abundances, values are only shown when the abundance of the endmember is more than 80%.

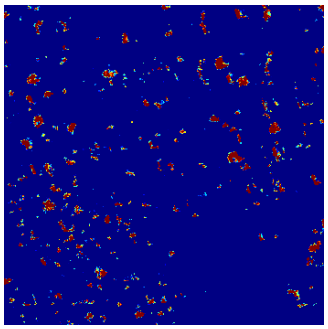
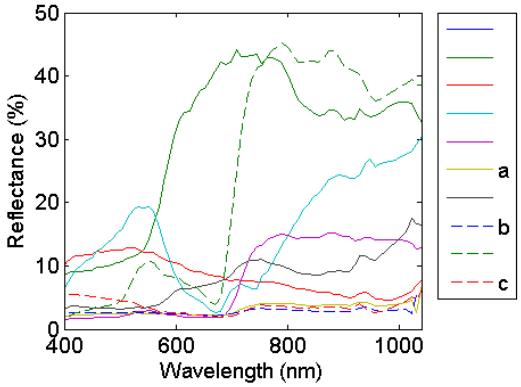
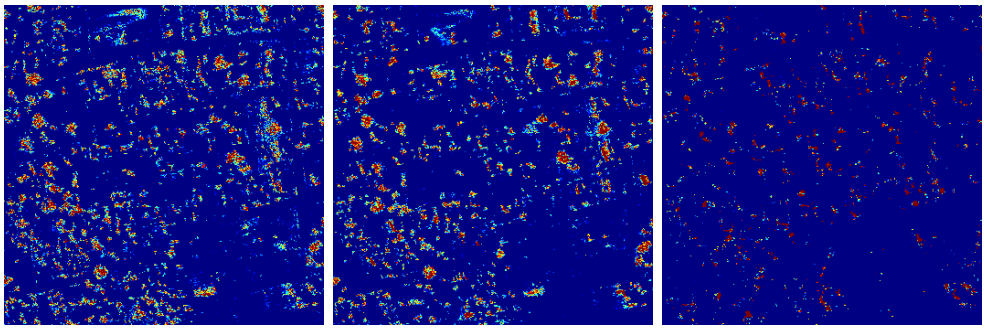


Figure 8.16.: The abundance map associated with the low amplitude endmember for CAO scene 1 with 10 endmembers. For $K = 8$ the abundances associated with this endmember were mostly large, seen in Figure 8.12 (h). For $K = 10$, the abundances associated with the low abundance endmember now appear to form a well defined class. To remove low abundances for clarity, values are only shown when the abundance of the endmember is more than 80%.



(a)

Figure 8.17.: The endmember spectra for CAO scene 1 with $K = 10$, using VCA. The vegetation class is still represented by only one spectral signature, whereas the low amplitude spectrum that may represent shadow, is now separated into three spectra.



(a) (b) (c)

(d)

Figure 8.18.: The abundances associated with the low amplitude endmembers determined by VCA for K=10.

8.4.3. Unmixing with $K = 14$

When CAO scene 1 is unmixed with $K = 14$, 10 of the endmembers and their abundances are visually similar to those for $K = 10$. The four new endmembers are shown in Figure 8.19, and their abundances are displayed in Figure 8.20. All of the new endmembers relate to very specific structures, such as certain buildings. Note that in Figure 8.11(h) the trees and roofs are clearly in the same class, corresponding to the spectrum seen in Figure 8.10 (h), which means that the number of classes is too low. Figure 8.20 (a) and (c) show the red and blue roofs respectively in separate classes, with no vegetation included in either class. All the abundance maps show spatial clusters of similar values, and so we conclude it is visually likely that $K \geq 14$. All the thresholded abundances, seen in Figure 8.21 show that the pixels that contain abundances over 80% also form congruent clumps, even though some may be small, which further supports this claim.

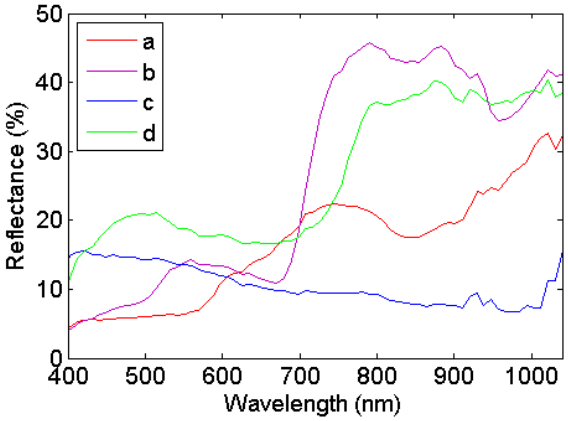
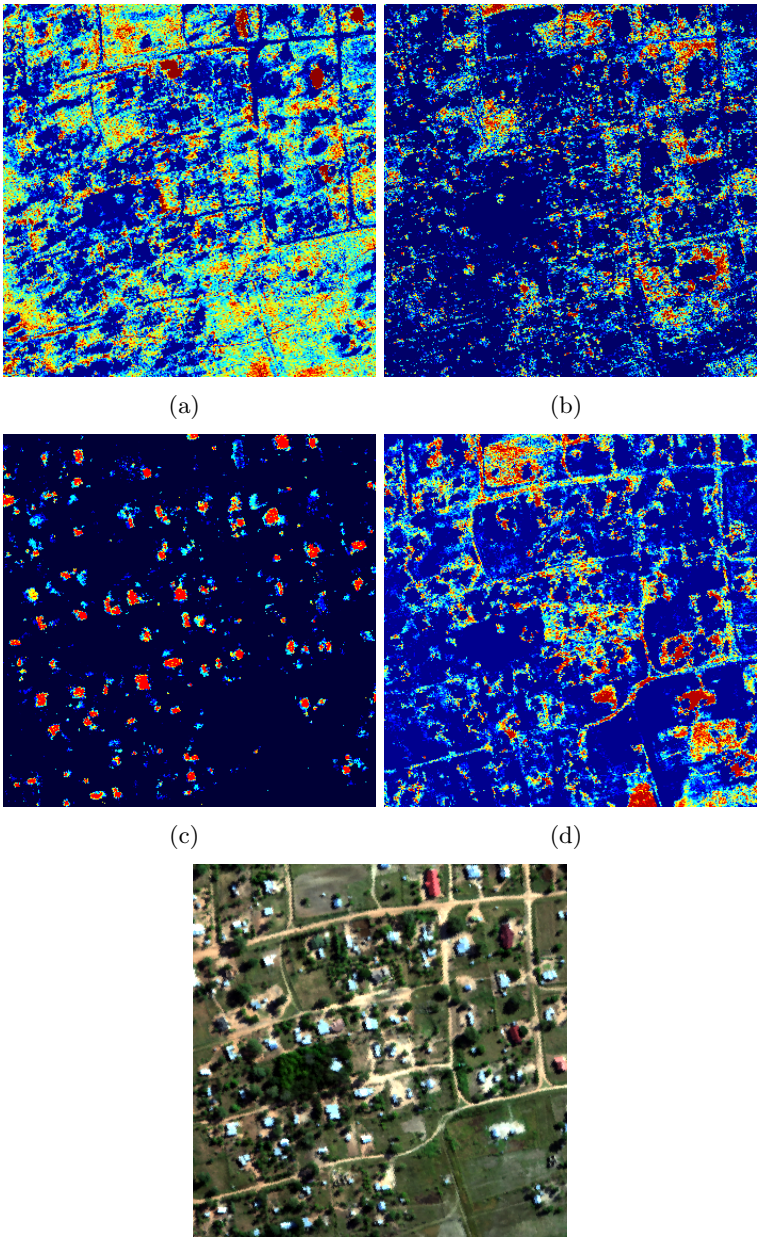
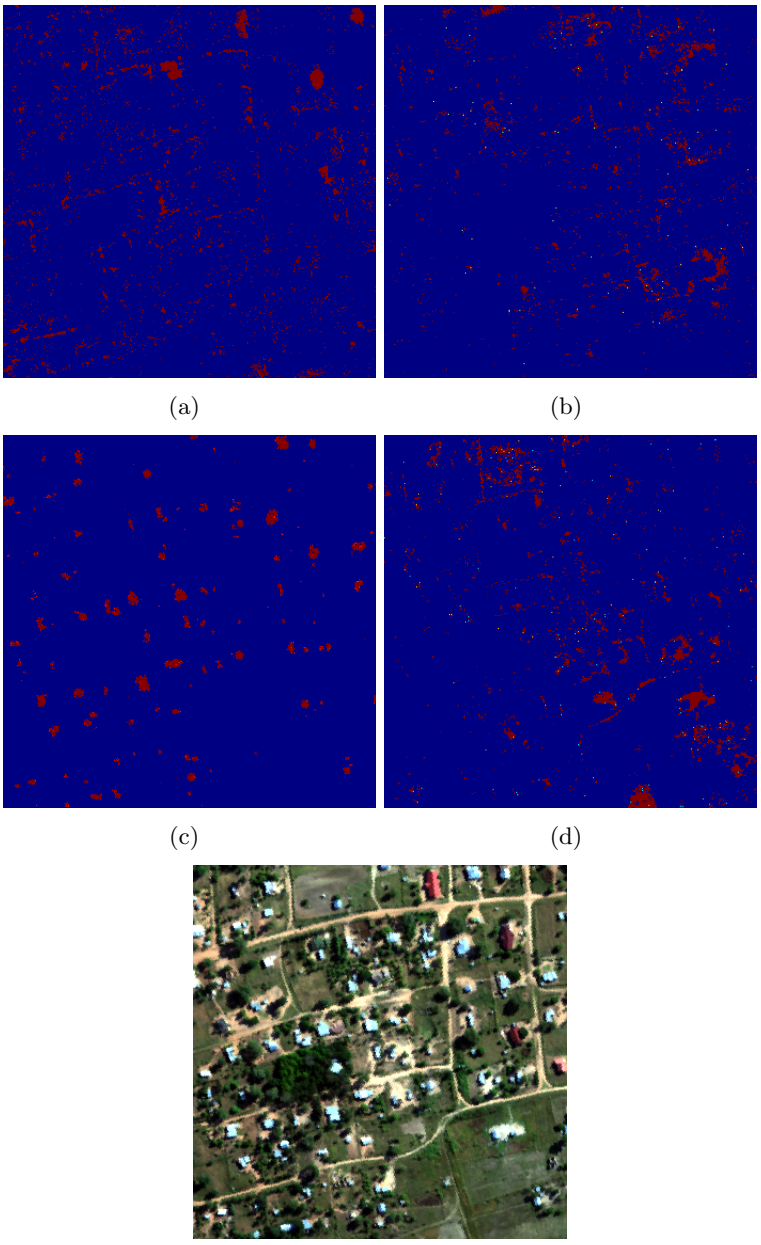


Figure 8.19.: The endmember spectra of CAO scene 1 for $K = 14$ that are not similar to those for $K = 10$. All spectra are clearly separable and do not behave like noise.



(a) (b) (c) (d) (e)

Figure 8.20.: The abundance maps of four endmembers for CAO scene 1 with $K = 14$. These four endmembers are chosen by removing those visually similar to the endmembers determined by $K = 10$. All endmembers correspond to abundance values that form congruent clumps, which are associated with structures in the image.



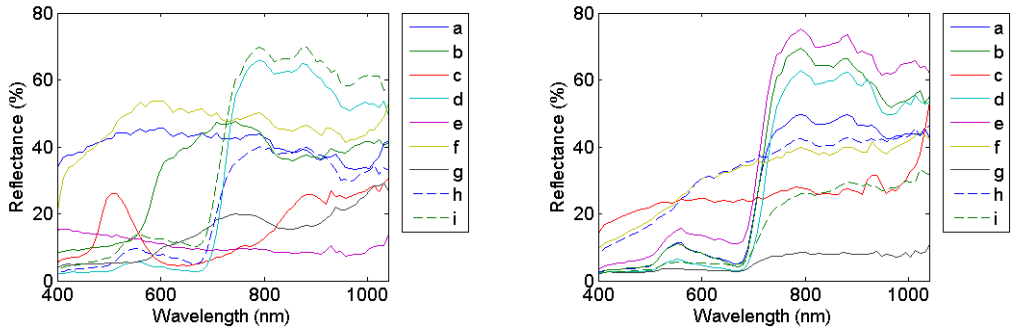
(a) (b) (c) (d) (e)

Figure 8.21.: The thresholded abundance maps of four endmembers for CAO scene 1 with $K = 14$. All endmembers correspond to abundance values that form congruent clumps, which are associated with structures in the image.

8.4.4. Unmixing with $K = 18$

RMT with the residual noise approximation gives the largest ID estimate, with $K = 18$. The endmembers are not the same as those determined by $K = 14$, although some are similar, seen in Figure 8.22. So not only have new classes been added, some classes have been refined.

This refinement may be seen in Figures 8.23 and 8.24, where all the abundance images are shown. The abundances form spatially congruent clusters, and separate classes do not appear to be combined as one endmember. Figures 8.23(g) and 8.24(e) do not display the same level of homogeneity as the other abundance images.



(a) (b)

Figure 8.22.: The endmember spectra of CAO scene 1 for $K = 18$, separated into the first 9 (a) and last 9 (b) spectra. All spectra are separable and not similar to noise.

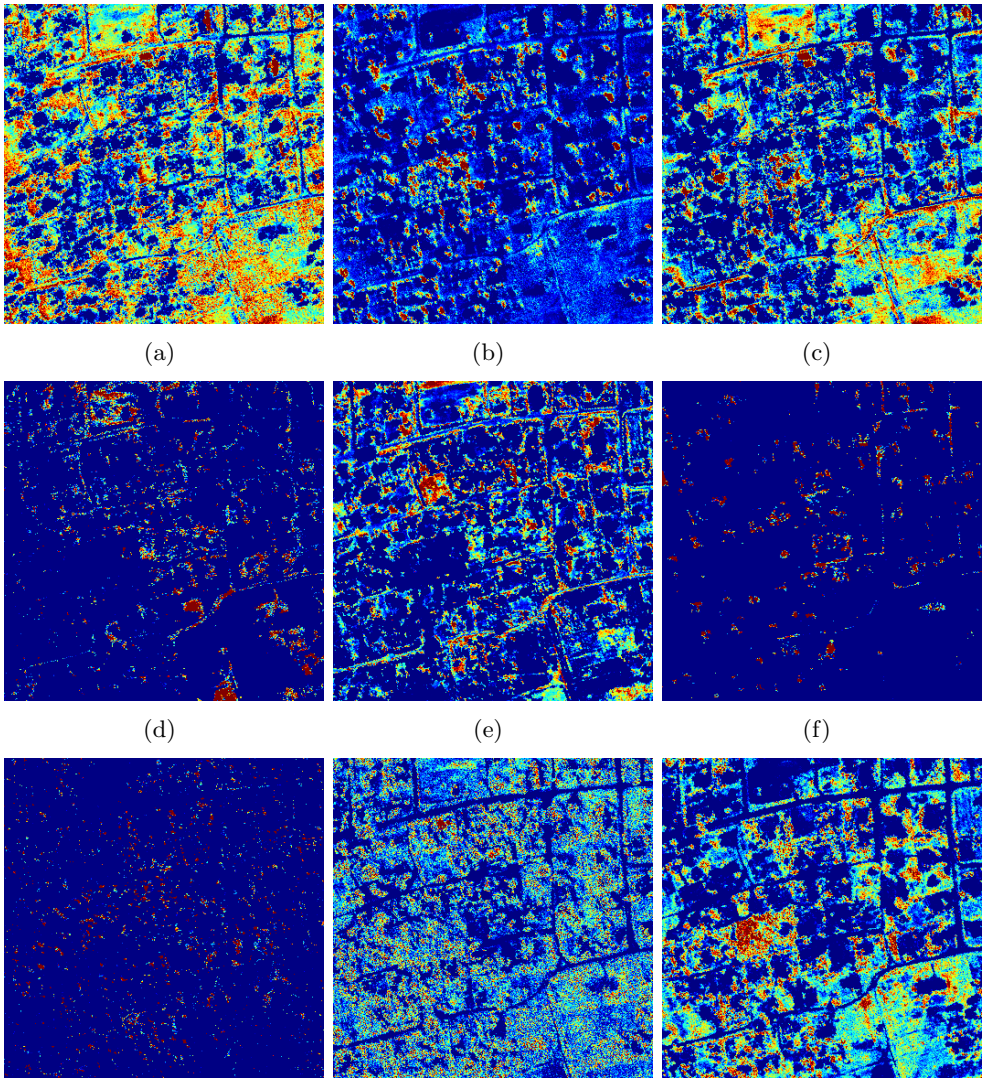


Figure 8.23.: The first nine abundance maps for CAO scene 1 with $K = 18$.

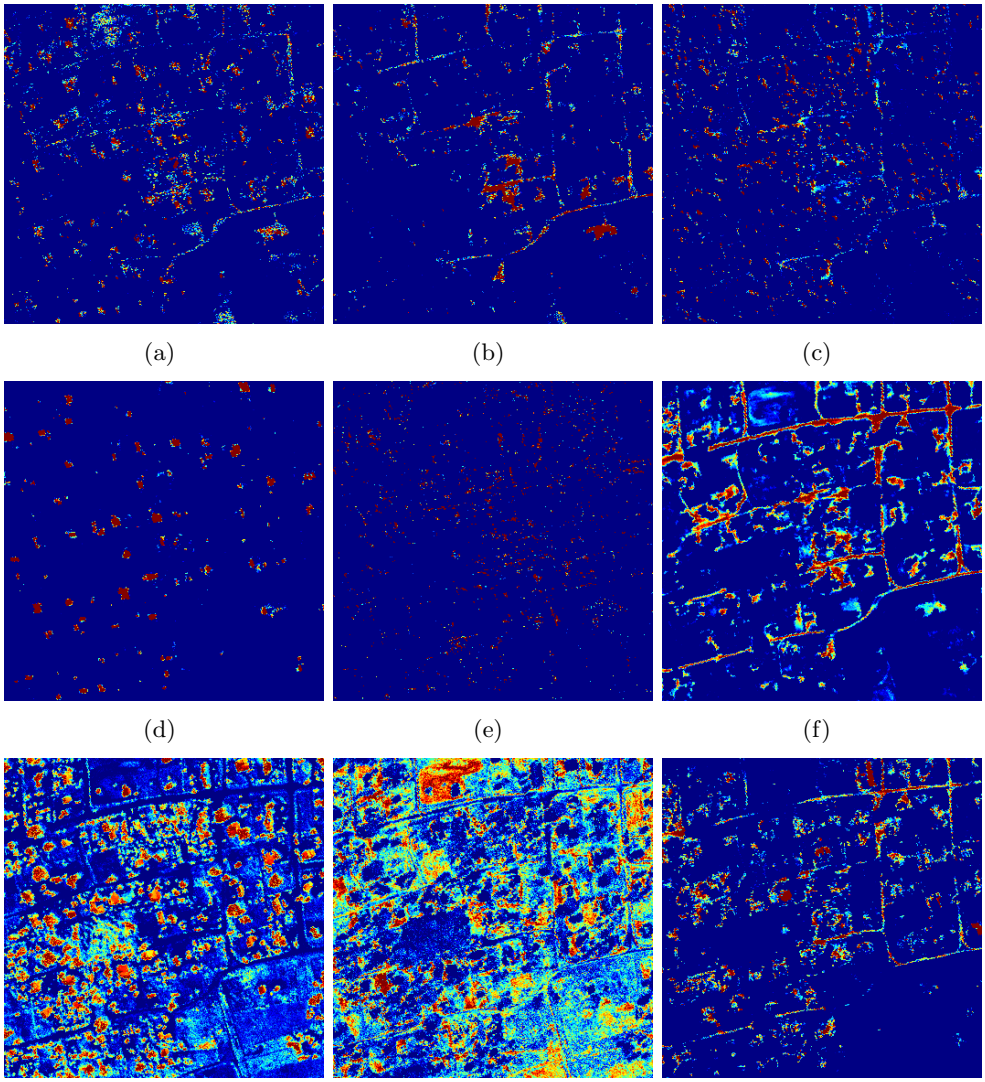


Figure 8.24.: The last nine abundance maps for CAO scene 1 with $K = 18$.

NFINDR uses PCA in an image reduction step, and we also tested the effect of using the RMT signal subspace for this reduction. The endmembers are shown in Figure 8.25 and the abundances are shown in Figure 8.26 and 8.27. The endmembers are distinct and the abundances appear to form spatially congruent regions, but more than this visual appreciation is required to determine which method is more accurate.

To remove the low abundances for clarity, values less than 80% are removed from the abundance images. These results, when using the original NFINDR extraction are shown in Figures 8.28 and 8.29. Some classes are well separated, while other contain very few pixels with more than 80% abundance. Extensive ground truth would be required to verify the accuracy of this. Also note that $K = 18$ may be too high a number for NFINDR since it requires pure pixels for each endmember, whereas RMT can detect endmembers that make up a fraction of a pixel.

This test illustrates the difficulty of determining the ID based only on unmixing results. Other methods are needed to determine the accuracy of the ID estimation for each method. All that can be concluded from these tests is that overall, all the tested values for K give visually reasonable abundance images, especially $K \geq 14$.

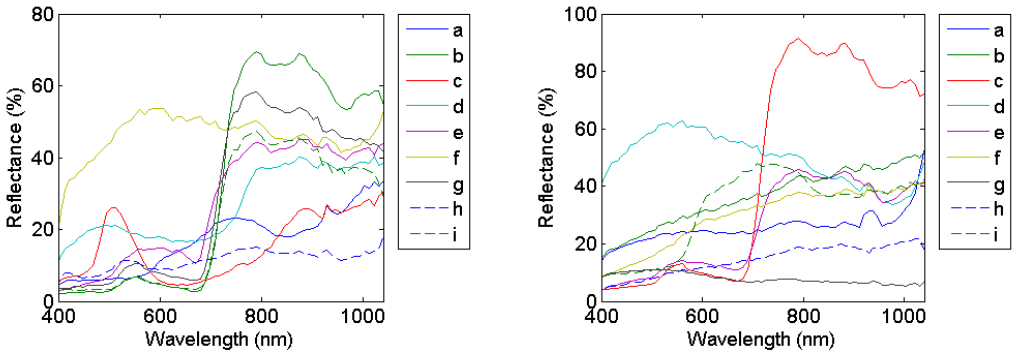


Figure 8.25.: The endmember spectra of CAO scene 1 for $K = 18$, separated into the first 9 (a) and last 9 (b) spectra. The image reduction is done in NFINDR by using the RMT signal subspace instead of PCA. All spectra are separable and not similar to noise.

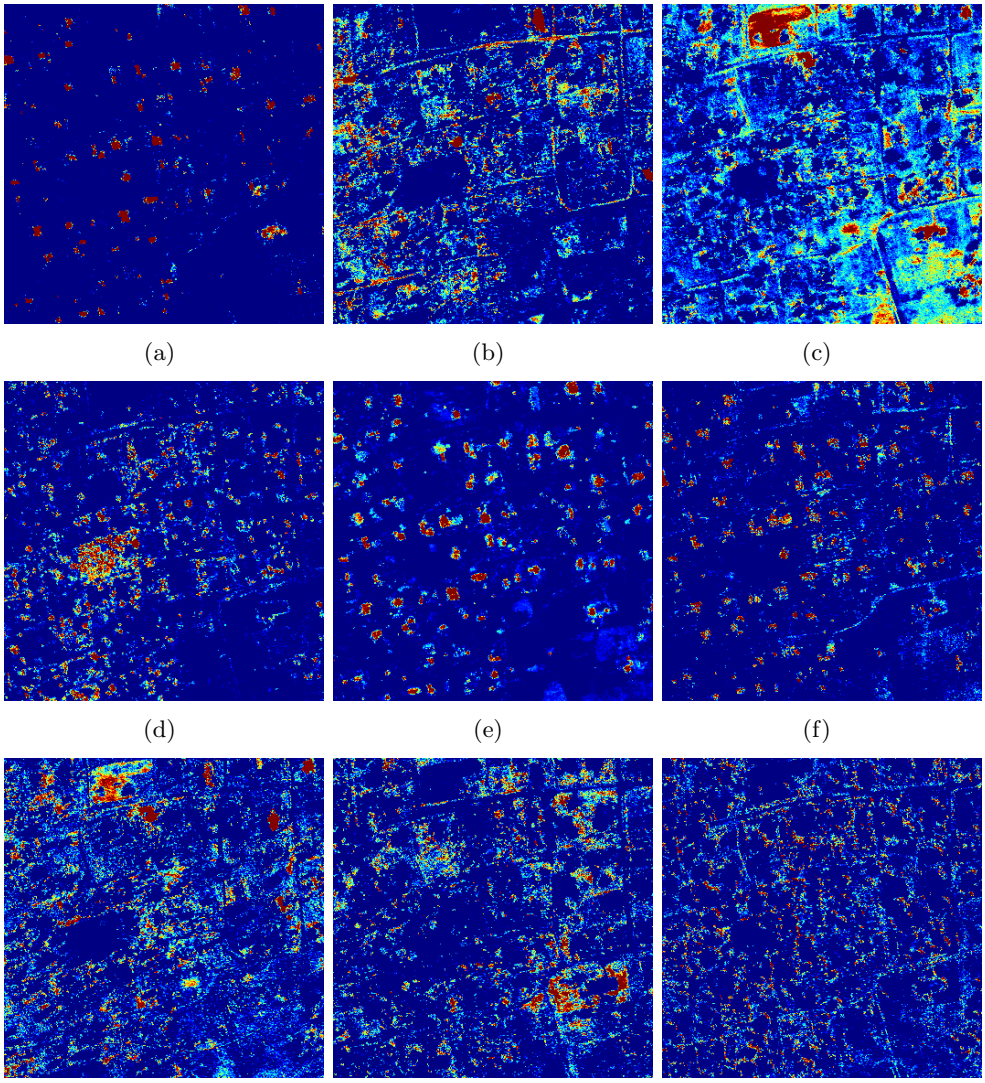
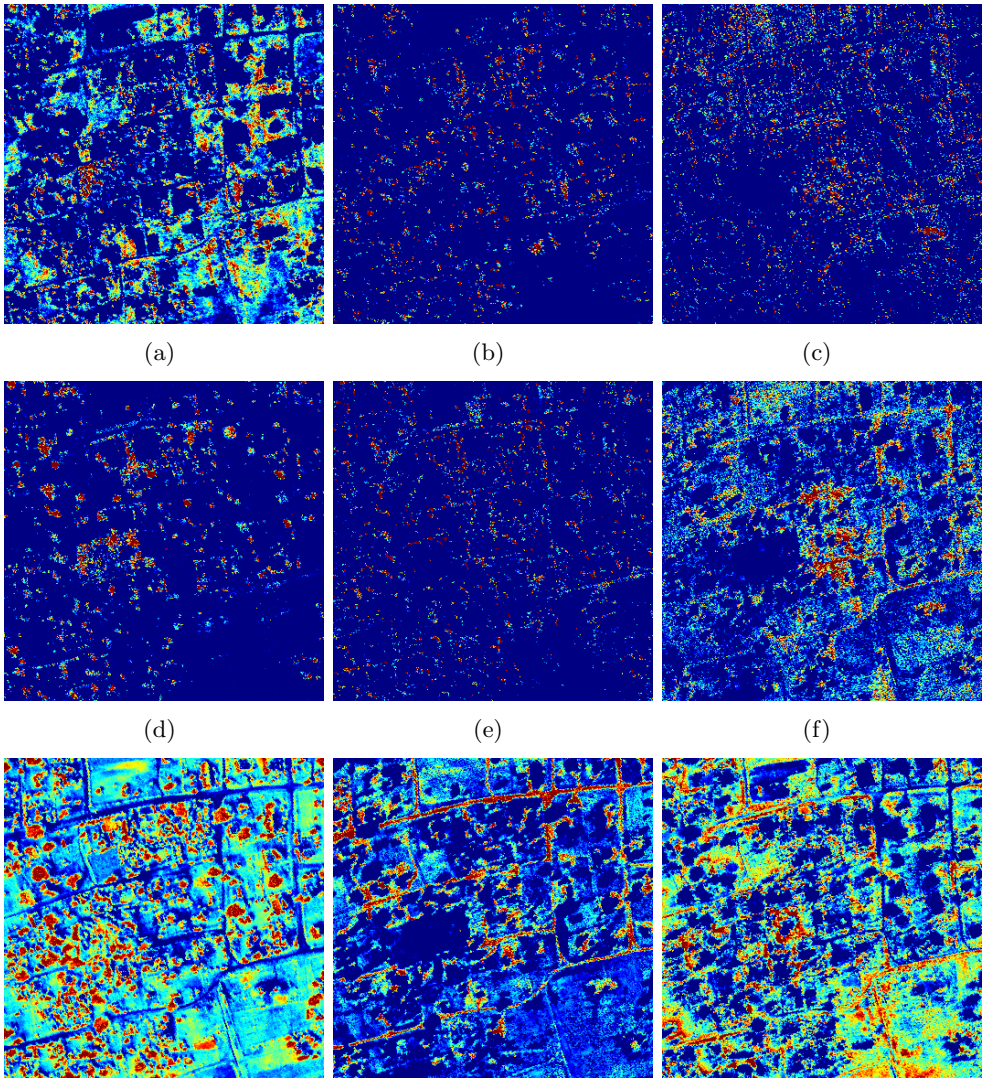


Figure 8.26.: The first nine abundance maps for CAO scene 1 with $K = 18$ and RMT subspace.

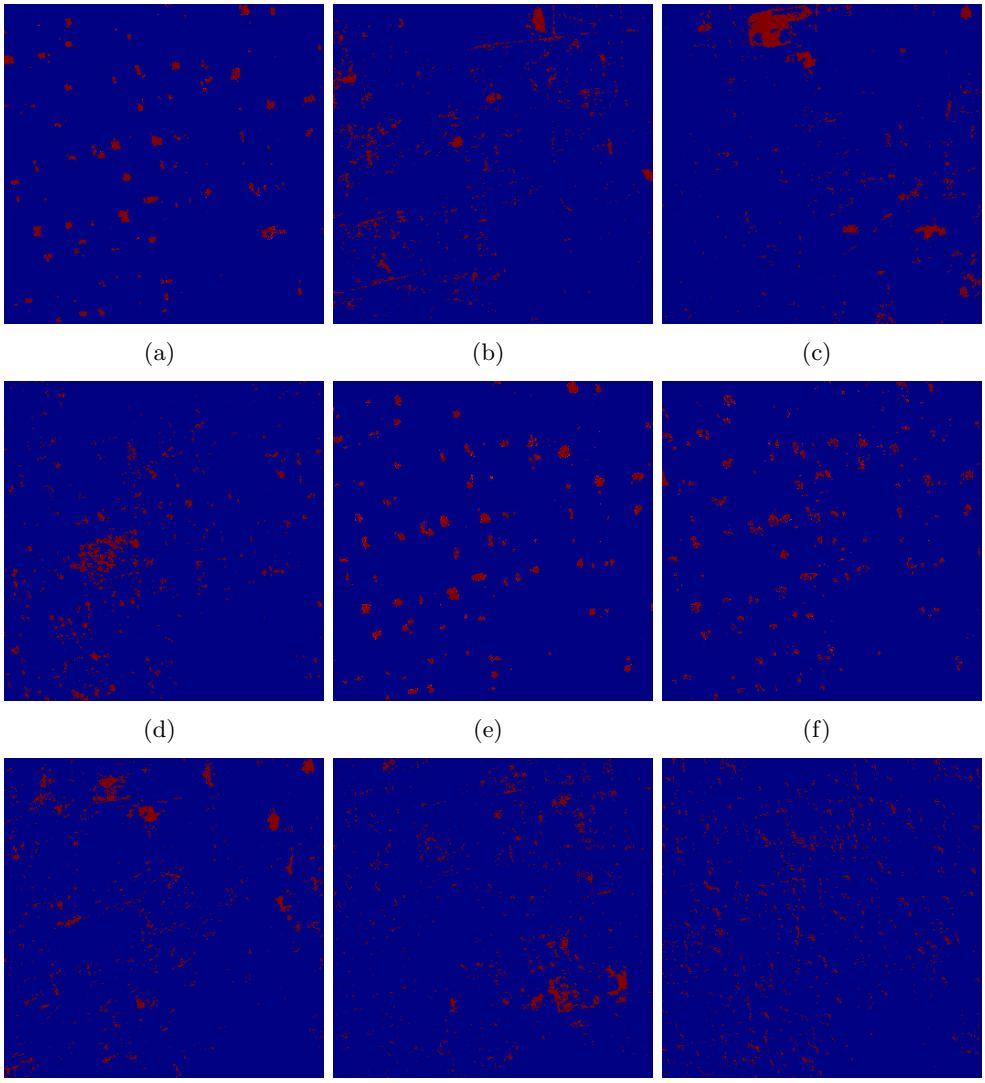


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 8.27.: The last nine abundance maps for CAO scene 1 with $K = 18$ and RMT subspace.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 8.28.: The first nine thresholded abundance maps for CAO scene 1 with $K = 18$ and RMT subspace.

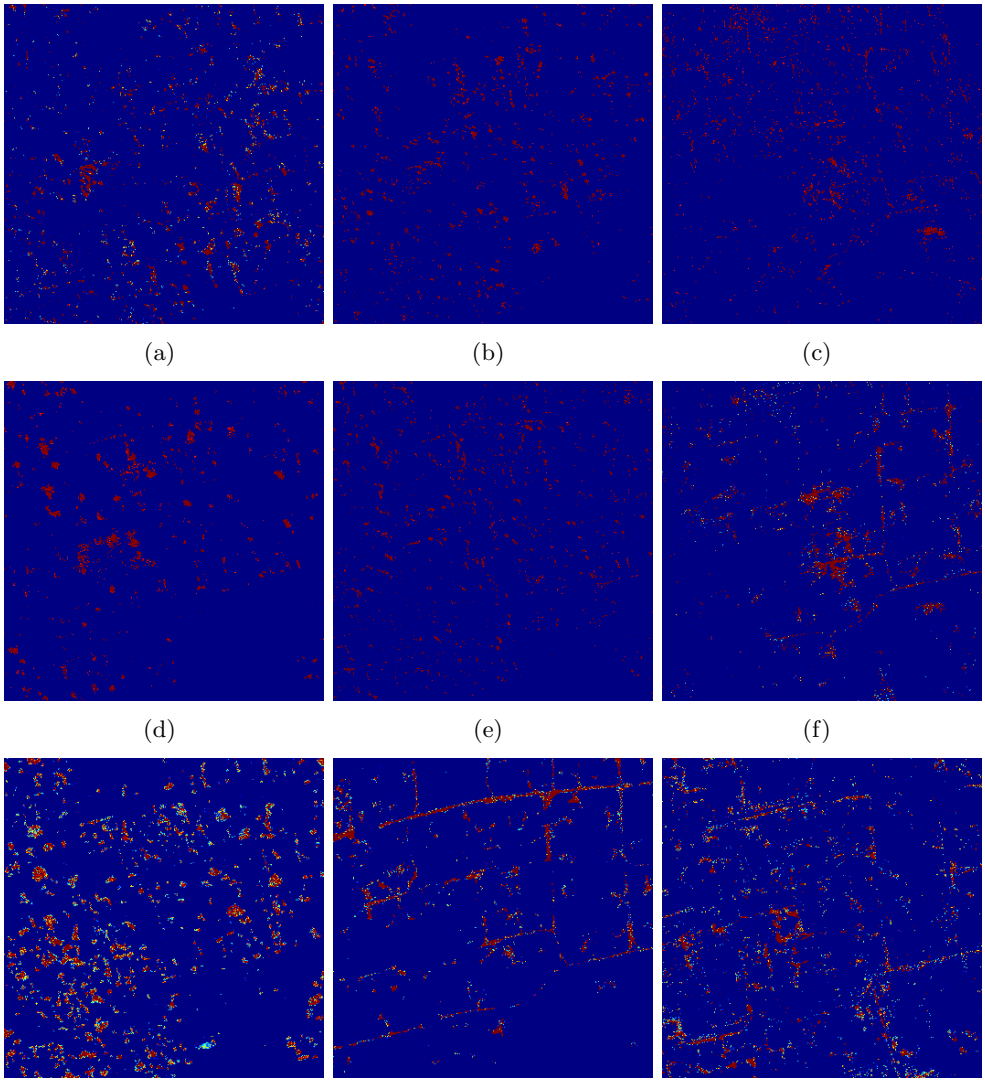


Figure 8.29.: The last nine thresholded abundance maps for CAO scene 1 with $K = 18$ and RMT subspace.

8.5. Dependence on image size

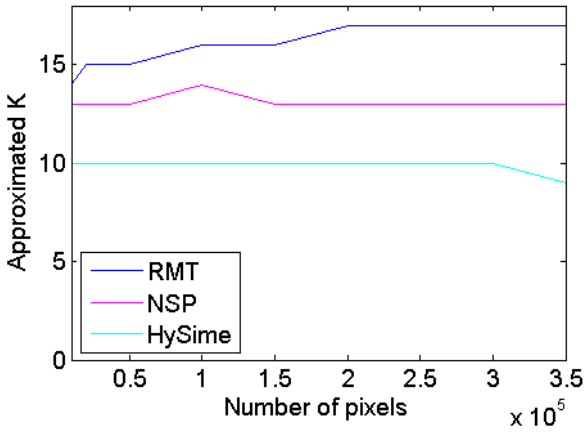


Figure 8.30.: The dependence of RMT, NSP and HySime on image size for CAO. RMT showed a decrease in its estimates of K for small image sizes, possibly due to the exclusion of rare substances. NSP and HySime determined a constant K, indepedent of image size, except for a single subset for NSP.

In the synthetic tests, RMT with the residual and multiple regression noise approximations required a minimum image size of approximately 6,000 pixels. This is smaller than the test images considered in this thesis, which mostly contain more than 100,000 pixels, but in this section we will test if this threshold is still applicable to real images.

To do so, a larger CAO subset (without the first band) is chosen with 367,500 pixels. Random spatial subsets were chosen, of varying sizes, to test if RMT, NSP or HySime were dependent on the size of the image. Recall that RMT was most sensitive to rare substances, and so by choosing a subset of the original image, some substances may be excluded. Figure 8.30 shows that the RMT value of K is stable for a CAO subset larger than 200,000 pixels. In real images, this limit is considered small. NSP produced constant values for K, except for a single value of $K = 14$, and HySime showed constant values of $K = 10$, except for a single value of $K = 9$.

The same behaviour is seen in the other images. Figure 8.31 shows the sensitivity of all three methods to image size for SpecTIR Cuprite. The ID estimated by RMT drops for SpecTIR subsets smaller than 20,000 pixels, and the ID estimated by HySime drops by a single value at that threshold as well. The NSP estimates remain constant for every subset size.

It is interesting to note that the plateau of RMT values begins at a different number

of pixels for CAO and SpecTIR Cuprite. This may be attributed to the more complex vegetation scene in CAO requiring more pixels to cover all endmembers, whereas Cuprite is a mineral scene.

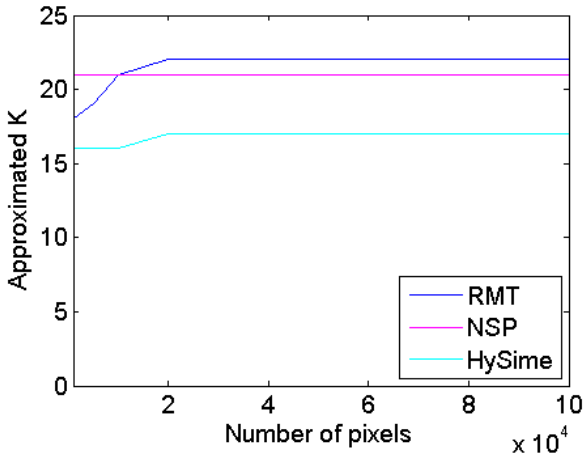


Figure 8.31.: The dependence of RMT, NSP and HySime on image size for SpecTIR Cuprite. RMT showed a decrease in its estimates of K for small image sizes, possibly due to the exclusion of rare substances. NSP determined a constant K, indepedent of image size, and the ID estimated by HySime dropped only by a single value for smaller images.

8.6. Comparing two datasets over the same spatial area

Figure 3.6 shows that the Hyperion and AVIRIS Cuprite datasets overlap on a spatial subset. The overlapping subset is compared for both sensors, and the intrinsic dimension is expected to be the same or similar. The Hyperion image is 14 years more recent, so there is the possibility that the contents of the scene have changed, but due to the nature of Cuprite, which is in the Nevada desert, large changes are unlikely. Hyperion pixels cover a larger spatial area (30 m) than AVIRIS (20 m), and Hyperion contains much higher levels of correlation than AVIRIS. Hyperion also requires different atmospheric corrections as it is a satellite sensor whereas AVIRIS is an airborne sensor. All of these qualities will affect the algorithms used. The results for both images are given in Table 8.4 for RMT, NSP and HySime.

Table 8.4.: RMT, NSP and HySime are used to calculate the intrinsic dimension of two datasets (AVIRIS and Hyperion) over the same spatial area in Cuprite. Correlated and bad bands are removed for RMT. RMT with the residual noise approximation method was the only methods to achieve any kind of similarity in the intrinsic dimensions between both datasets.

	RMT_{Meer}	RMT_{Residual}	HySime	NSP
K_{AVIRIS}	21	28	16	28
K_{Hyperion}	46	23	31	71

It has been noted that, while NSP and HySime are able to tolerate some level of correlation, they overestimate K when the noise is underestimated too severely. This may be seen in the high K values for Hyperion, especially for NSP ($K = 71$). The only method that resulted in similar values for K between the two datasets was to remove the most highly correlated bands. RMT with band removal gave $K = 28$ and $K = 23$ for AVIRIS and Hyperion respectively.

In RMT (with the residual noise approximation method), the difference in ID approximations may be attributed to new substances in the scene (there is a 14 year difference in acquisition times), small errors in the spatial matching, the difference in spatial resolution, correlation errors that have still not been corrected, etc. The most likely cause is that Hyperion has only 98 bands remaining after correlated and bad bands have been removed, while AVIRIS has 181 bands (nearly double), so more substances may be identified. When the ID of the AVIRIS scene is calculated with the same bands as the Hyperion scene, the ID is $K = 25$, which is very similar to the $K = 23$ determined for

Table 8.5.: RMT, NSP and HySime are used to calculate the intrinsic dimension of AVIRIS and SpecTIR datasets over the same spatial area in Cuprite. Correlated and bad bands are removed for RMT.

	RMT_{Meer}	RMT_{Residual}	HySime	NSP
K_{AVIRIS}	13	16	17	23
K_{SpecTIR}	11	22	140	139

Hyperion. NSP and HySime still did not produce similar results for the two sets, even when correlated and bad bands were removed and the two images were evaluated for the same bands.

The AVIRIS Cuprite scene has been successfully compared to the Hyperion Cuprite scene, and now it may also be compared to the SpecTIR Cuprite scene, although the difference in spatial resolution is high (20 m for AVIRIS and 1 m for SpecTIR). As a result, the AVIRIS subset has only 87×87 pixels in order to match up to the SpecTIR scene. The closest results are given by RMT with Meer's noise approximation method (Table 8.5), which determined $K = 13$ for AVIRIS and $K = 11$ for SpecTIR.

SpecTIR and AVIRIS do not cover the same spectral ranges, and so the same bands may not be compared in this case. If correlated and bad bands are removed for NSP and HySime however, there is a marked improvement in the correlation between the SpecTIR and AVIRIS images over Cuprite. In particular, NSP shows good results with $K = 22$ for AVIRIS and $K = 21$ for SpecTIR. HySime produced $K = 13$ for AVIRIS and $K = 17$ for SpecTIR. RMT is more dependent on image size than NSP (seen in Figure 8.31), and the lower estimate for AVIRIS (when using the residual noise approximation method) may be due to the small image size (87×87 pixels). The RMT estimate of 22 for SpecTIR is in line with the NSP estimates for both scenes once correlated and bad bands have been removed.

Unfortunately, the final set of images that were acquired over the same spatial area are the AVIRIS and Hyperion scenes over Moffet Fields. Moffet Fields contains substantial agricultural areas, and images acquired in different seasons will be expected to show different results, as seen in Table 8.6.

Table 8.6.: RMT, NSP and HySime are used to calculate the intrinsic dimension of AVIRIS and Hyperion datasets over the same spatial area in Moffet Fields. Correlated and bad bands are removed for RMT.

	RMT_{Meer}	RMT_{Residual}	HySime	NSP
K_{AVIRIS}	153	46	26	50
K_{Hyperion}	90	18	42	78

8.7. Discussion

All three methods were affected by correlated bands, and these effects were especially apparent in SpecTIR and HCI. Removing the correlated bands using a threshold improved the results of all methods. HySime consistently gave the lowest ID estimates, and this method was sensitive to outliers. RMT with band removal often produced the highest estimate for K of all methods.

All methods were tested on the three spatial subsets that were acquired by two different sensors. Moffet Fields contained vegetation and all methods determined different ID values for each image, which may be expected as the images were not acquired at the same time. In Cuprite, when AVIRIS and Hyperion were compared, RMT (with correlated and bad bands removed) determined very similar ID values for the two images when compared on the same spectral subset. When AVIRIS and SpecTIR were compared for Cuprite, NSP gave nearly consistent results for both images, and corresponded to the RMT result for AVIRIS, but the small image size of the SpecTIR subset may have caused an underestimation for RMT. Overall, RMT performed very well in these tests.

When subsets of large images were considered, RMT was the only method to show significantly lower estimates of K for small subsets, further reinforcing the sensitivity of RMT to rare substances as seen in Chapter 7. Conversely, RMT did also require a certain number of pixels for the method to produce accurate results. This was also seen in the synthetic tests.

Recall that while correlated noise posed a problem for the noise approximation methods (Chapter 4), RMT as a method was not sensitive to this (Chapter 7). Because RMT does not depend on a particular noise approximation method, this leads us to believe that with the correct noise approximation method, RMT could be accurate even in the presence of correlated noise. Methods exist to remove correlated bands, and in this thesis only a fixed thresholding method was evaluated, simply to show the benefit to removing

most of the correlated bands. Better methods for dealing with the correlated noise may be considered at a later stage.

9. Conclusion

In this thesis a new method, which we refer to as RMT, has been developed to determine the intrinsic dimension of an image. This method uses techniques in Random Matrix Theory, and builds on existing methods so that they may be applicable to hyperspectral imagery.

RMT with i.i.d. noise

The original RMT method for finding the ID of a hyperspectral dataset assumed i.i.d. and uncorrelated noise (noise that is chosen from the same distribution $\mathcal{N}(0, \sigma^2)$ in each band of a hyperspectral image). This method performed well on synthetic sets where the noise assumptions were met, but proved to be very sensitive to these assumptions. The original RMT performed significantly better than well known methods AIC and MDL, which make the same noise assumptions, but it did not perform well on the real hyperspectral images that were tested, due to the fact that these images contained non-i.i.d. and correlated noise.

RMT with whitened noise

Whitening methods transform the noise so that it is i.i.d. and uncorrelated. The original RMT assumptions are then met. In synthetic tests, this method was still sensitive to non-i.i.d. and correlated noise, and application to real images showed improved results, especially when using Meer's spatially based method for noise approximation. Due to the synthetic tests, this method was not considered consistent enough for application to hyperspectral images.

RMT with non-i.i.d. and correlated noise

Due to the limitations of the original RMT method, a new evaluation criterion was derived so that the new method allows for non-i.i.d. and correlated noise, which enables application to hyperspectral images. Noise was considered as the sum of a mean noise variance and a matrix containing variation around that mean, so that the mean noise

variance followed the same behaviour as i.i.d. noise. The variation around this mean was treated separately in two mathematical propositions, which showed that noise and signal eigenvalues remain separable when the noise is non-i.i.d. and correlated, and a new threshold to separate these eigenvalues was derived.

RMT was then tested on synthetic data to determine if it was dependent on any variable. It required an image size of at least 6,000 pixels, and RMT showed a decrease in the ID estimate for small subsets of real images, although this may also be affected by the presence of rare substances in the scene. This limit is much smaller than the size of a standard hyperspectral image. RMT was also compared with existing state of the art methods, namely HySime and NSP. RMT performed better than both of these in images with high noise variance, and images that contain rare substances, i.e. substances that make up only a fraction of a single pixel. RMT performed comparably to HySime and NSP in most other tests.

RMT requires only an estimate of the noise covariance matrix. It does not require a noise approximation for every pixel, like HySime, and it does not require a user-determined threshold, like NSP. Three methods for estimating the noise are discussed in this thesis — two statistical methods and one spatially based method. The statistical methods performed badly when noise was correlated and tended to underestimate the noise in this case. RMT is sensitive to underestimation of noise, so correlated bands must be removed if these statistical methods are to be used. Preferably, a noise approximation method should be used that does not assume uncorrelated noise. The spatially based method was not sensitive to correlation, but it did require spatially homogeneous regions in the image in order to accurately estimate the noise variance.

RMT was also applied to twelve real images, including ones acquired by satellite, airborne and land-based sensors. The images contained minerals and vegetation, some contained correlated noise, and a variety of spatial and spectral resolutions were considered. Three images were acquired over the same spatial area by more than one sensor. These were compared for consistency: Moffet Fields contains large areas of vegetation, which we would not consider to be consistent at different times; in Cuprite (AVIRIS and Hyperion) RMT was the only method to give comparable results for both images; and in Cuprite (AVIRIS and SpecTIR) NSP gave consistent results, and RMT matched this result in the larger image, while the smaller contained only $\sim 5,000$ pixels and ID was underestimated, probably due to image size. Overall, RMT performed well in these tests.

One of the advantages of RMT is that is not dependent on any noise approximation method, and this may be chosen according to the requirements of the user. If the noise

is correlated, then a method must be used that is accurate in such circumstances; if the noise is uncorrelated and a simple statistical method will suffice, then RMT returns the result in the lowest computational time compared to other methods.

9.1. Open questions

Some open questions remain, which may lead to further work.

Noise approximation methods that are accurate in the presence of correlated noise

Correlated noise has been shown to be one of the most important influences on determining the ID of a hyperspectral image. Correlated noise occurs in many real images, and while RMT as a method was not affected by this, the statistical noise approximation methods severely underestimated the noise when it was correlated. Meer's noise approximation method was also not reliable, since it requires spatially homogeneous regions in the image in order to accurately estimate the noise. Newer noise approximation methods that are accurate in the presence of correlated noise should be tested with RMT on real images, and we expect to see realistic results without the need to remove bands from the dataset. Should this not be successful, better band removal methods should be considered to remove the correlated bands before applying RMT.

A noise estimation method that is worth testing is to approximate the noise in each pixel as the residual of a smooth curve fit (e.g. a cubic spline). Another method is to assume that the scene may be represented as a sum of multivariate Gaussian distributions, and the noise covariance matrix may be estimated using the Expectation Maximization algorithm. These approaches may be more appropriate to real world hyperspectral data than the methods usually discussed in the literature, which are used in this thesis.

Determining the cause of noise correlations

The cause of this correlated noise should also be better understood. AVIRIS and Hyperion share many common sensor characteristics such as band width, spectral range and spectral overlap; so the cause of the high correlation seen in Hyperion must be due to another sensor characteristic (for instance, Hyperion is a pushbroom sensor, while AVIRIS is a whiskbroom) or pre-processing techniques. The negative correlation seen in SpecTIR may also be due to pre-processing, but should also be adequately explained. The designers of hyperspectral sensors and pre-processing algorithms will benefit from this understanding, so that datasets may, if possible, be produced that are free from correlated noise.

Intra-class variation of endmembers

In real images certain substances, especially vegetation, may produce spectral responses that are not fixed throughout the image. For example, two trees that are of the same species may contain different numbers of leaves, which would result in slightly different spectra. RMT should be evaluated in the case of variable endmembers that are centered around some mean value.

Synthetic images that adequately model real images

The synthetic dataset should also be evaluated so that it can more accurately approximate a real image. Shadow played a large part in the endmember extraction, with VCA even separating the shadow endmember into several classes. This and other variables should be better understood so that tests on synthetic images can accurately determine the performance of ID approximation methods.

ID estimates that are dependent on the endmember extraction method

The differences in NFINDR and VCA show that the correct number of endmembers for accurate unmixing may be dependent on the endmember extraction method that is used, and certain extraction methods may be preferred for a specific image type. Several endmember extraction methods should be considered for multiple hyperspectral images, to determine the optimal number of endmembers that should be extracted for each method. It would be interesting to compare this number to the ID determined by RMT, NSP and HySime.

Noise that is dependent on the signal

Modern sensors in particular may be affected by signal-dependent noise, as well as Gaussian noise. The ID may not be accurately estimated in this case. The effects of this type of noise should be considered for ID estimation methods, and some of the newly developed methods for estimating the parameters of the signal-dependent noise should be evaluated with RMT.

Bibliography

[1] C-I. Chang. *Hyperspectal Data Exploitation: Theory and Application.* Wiley, 2007.

[2] S.K. Jain. *Lossless compression of hyperspectral images.* ProQuest, 2006.

[3] D. Bannon. Hyperspectral imaging: Cubes and slices. *Nature Photonics*, 3:627–629, 2009.

[4] R. P. Gupta. *Remote Sensing Geology.* Springer, 2003.

[5] A. Banerjee, H. Ren, J. Jensen. Extended linear hyperspectral mixing models. *Proc. of SPIE*, 5268:288–301, 2004.

[6] D. Gillis, *et al*. A generalised linear mixing model for hyperspectral imagery. *Proc. of SPIE*, 6966:1–11, 2008.

[7] J.M. Bioucas-Dias, A. Plaza. Hyperspectral unmixing: geometrical, statistical and sparse regression-based approaches. In *Proceedings of SPIE*, volume 7830, 2010.

[8] J.M. Bioucas-Dias, J.M.P. Nascimento. Estimation of signal subspace on hyperspectral data. In *Proceedings of SPIE*, volume 5982, pages 191–198, 2005.

[9] Chein-I Chang, Qian Du. Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery. *IEEE Transactions of Geoscience and Remote Sensing*, 42 (3), 2004.

[10] C-C. Wu, W. Liu, C-I. Chang. Exploration of methods for estimation of number of endmembers in hyperspectral imagery. *Proc. of SPIE*, 7(43):1–11, 2006.

[11] P. Bajorski. Does Virtual Dimensionality work in hyperspectral images? In *Proceedings of SPIE*, volume 7334, 2009.

[12] A. Schlamm, D. Messinger, W. Basener. Geometric estimation of the inherent dimensionality of single and multi-material clusters in hyperspectral imagery. *Journal of Applied Remote Sensing*, 3, 2009.

[13] D.C. Heinz, C-I. Chang. Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):529–545, 2001.

[14] S. Kritchman, B. Nadler. Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, 94:19–32, 2008.

[15] H-M. Park, *et al*. Subband-based blind signal separation for noisy speech recognition. *Electronic letters*, 35:2011–2012, 1999.

[16] F.A. Kruse. Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra. *International journal of remote sensing*, 17(9):1623–1632, 1996.

[17] F.A. Kruse. Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra. *International Journal of Remote Sensing*, 17(9):1623 – 1632, 1996.

[18] G.P. Asner, K.B. Heidebrecht. Imaging spectroscopy for desertification studies: comparing AVIRIS and EO-1 Hyperion in Argentina drylands. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1283 – 1296, 2003.

[19] J.-P. Ardouin, J. Levesque, T.A. Rea. A demonstration of hyperspectral image exploitation for military applications. In *10th International Conference on Information Fusion*, pages 1 – 8, 2007.

[20] H. Horwitz, R. Nalepka, P. Hyde, J. Morgenstern. Estimating the proportions of objects within a single resolution element of a multispectral scanner. In *Proceedings of the 7th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan*, pages 1307–1320, 1971.

[21] J. A. Richards, X. Jia. *Remote Sensing Digital Image Analysis: An Introduction.* Springer, 2006.

[22] R. A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing.* Academic Press, 2007.

[23] G.M. Foody, T.A. Warner, M.D. Nellis. *The SAGE handbook of remote sensing.* SAGE, 2009.

[24] Kali Charan Sahu. *Textbook of Remote Sensing and Geographical Information Systems*. Atlantic, 2007.

[25] R. Richter. Atmospheric/Topographic Correction for Satellite Imagery. Technical report, DLR, Germany, 2011. www.atcor.de.

[26] R. Richter. Atmospheric/Topographic Correction for Airborne Imagery. Technical report, DLR, Germany, 2011.

[27] K. Stamnes, S.-C. Tsay, W. Wiscombe, K. Jayaweera. Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered method. *Applied Optics*, 27:2502–2509, 1988.

[28] T.E. Avery, G.L. Berlin. *Fundamentals of remote sensing and airphoto interpretation*. Macmillan, 1992.

[29] N. Acito, M. Diani, G. Corsini. Subspace based striping noise reduction in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (4):1325–1324, 2011.

[30] J.S. Pearlman. Hyperion Validation Report. Technical report, Boeing, 2003.

[31] EO1 home page. edcsns17.cr.usgs.gov/eo1/.

[32] EnMAP home page. www.enmap.org.

[33] DLR EnMAP page. http://www.dlr.de/caf/en/desktopdefault.aspx/tabid-5514/9217_read-17744/.

[34] Kaufmann *et al*. Environmental Mapping and Analysis Program. In *ISIS*, 2007.

[35] Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond. http://cce.nasa.gov/pdfs/HYSPIRI.pdf.

[36] O. Mutanga, J. van Aardt, L. Kumar. Imaging spectroscopy (hyperspectral remote sensing) in southern Africa: an overview. *South African Journal of Science*, 105:193 – 198, 2009.

[37] AVIRIS home page. aviris.jpl.nasa.gov.

[38] SpecTIR home page. www.spectir.com.

[39] HyVista home page. www.hyvista.com.

[40] F.A. Kruse. Comparison of AVIRIS and Hyperion for Hyperspectral Mineral Mapping. In *11th JPL Airborne Geoscience Workshop*, 2002.

[41] M. Herald, D.A. Roberts. Spectral characteristics of asphalt road aging and deterioration: implications for remote sensing applications. *Applied Optics*, 44(20):4327–4334, 2005.

[42] S.S. Rajapakse, S. Khanna, M.E. Andrew, S.L. Ustin, M. Lay. Identifying and classifying water hyacinth (*Eichhornia crassipes*) using the HyMap sensor. In *Proceedings of SPIE*, 2006.

[43] N. Acito, M. Diani, G. Corsini. Hyperspectral signal subspace identification in the presence of rare signal components. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4):1940–1954, 2010.

[44] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19:716–723, 1974.

[45] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[46] G. Schwartz. Estimating the dimension of a model. *Annals of Statistics*, 2:461–464, 1978.

[47] M. Wax and T. Kailith. Detection of signals by information criterion. *IEEE Trans. Acoustic, Speech and Signal Processing*, 33:387–392, 1985.

[48] R.E. Malinowski. Determination of the number of factors and experimental error in a data matrix. *Analytical Chemistry*, 49 (4):612–617, 1977.

[49] H.T. Wu, J.F. Yang, F.K. Chen. Source number estimators using transformed Gerschgorin radii. *IEEE Tras. on Signal Process.*, 43(6):1325–1333, 1995.

[50] M.T. Eismann and D. Stein. Stochastic mixture modeling. In C.-I. Chang, editor, *Hyperspectral data exploitation: theory and applications*. New York: Wiley, 2007.

[51] O. Eches, N. Dobigeon and J. Tourneret. Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm. *IEEE Journal of Selected Topics in Signal Processing*, 4(3), 2010.

[52] O. Kuybeda, D. Malah, M. Barzohar. Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors. *IEEE Transactions on Signal Processing*, 55(12):5579–5592, 2007.

[53] N. Acito, M. Diani, G. Corsini. A new algorithm for robust estimation of the signal subspace in hyperspectral images in the presence of rare signal components. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11):3844–3856, 2009.

[54] R.E. Roger. Principal components transform with simple, automatic noise adjustment. *International Journal of Remote Sensing*, 17 (14):2719–2727, 1996.

[55] C.-I. Chang, W. Xiong, W. Liu, M.-L. Chang, C.-C. Wu, C. Chen. Linear spectral mixture analysis based approaches to estimation of virtual dimensionality in Hhperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):3960–3979, 2010.

[56] C.-I. Chang, W. Xiong, H.-M. Chen, J.-W. Chai. Maximum orthogonal subspace projection approach to estimating the number of spectral signal sources in hyperspectral imagery. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):504–520, 2011.

[57] J.M. Bioucas-Dias, J.M.P. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46 (8):2435–2445, 2008.

[58] S. Kritchman, B. Nadler. Non-Parametric Detection of the Number of Signals: Hypothesis Testing and Random Matrix Theory. *IEEE Transactions on Signal Processing*, 57 (10):3930–3941, 2009.

[59] I.M. Johnstone. High dimensional statistical inference and random matrices. In *Proceedings of the International Congress of Mathematics*, pages 307–333, 2006.

[60] I.M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327, 2001.

[61] J. Baik, J. Silverstein. Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models. *Journal of Multivariate Analysis*, 97:1382, 2006.

[62] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, 36 (6):2757–2790, 2008.

[63] 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. In *IEEE Transactions on Geoscience and Remote Sensing*, volume 26 (1), pages 65–74, 1988.

[64] M. Iordache, J.M. Bioucas-Dias, A. Plaza. Hyperspectral unmixing with sparse group lasso. In *IGARSS*, 2011.

[65] J.W. Boardman. Analysis, understanding and visualization of hyperspectral data as convex sets in n-space. *SPIE*, 2480:23–36, 1995.

[66] F. Chaudhry, C.-C. Wu, W. Liu, C.-I. Chang, A. Plaza. Pixel purity index-based algorithms for endmember extraction from hyperspectral imagery. In *Recent advances in hyperspectral signal and image processing*. Transworld research network, 2006.

[67] M. Winter. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. *Proceedings of SPIE*, 3753:266–275, 1999.

[68] J. Nascimento and J. Bioucas Dias. Vertex Component Analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43 (4):898–910, 2005.

[69] C-I. Chang, C-C. Wu, W-M. Liu, C-Y. Ouyang. A new growing method for simplex-based endmember extraction algorithm. *IEEE transactions on geoscience and remote sensing*, 44(1):2804–2819, 2006.

[70] C.-I. Chang, D.C. Heinz. Constrained Subpixel Target Detection for Remotely Sensed Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1144–1159, 2000.

[71] R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *Proceedings of Micro Machine and Human Science*, pages 39–43, 1995.

[72] G. Swayze, R. Clark, S. Sutley, A. Gallagher. Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada. In *Summaries of the 3rd Annual JPL Airborne Geosciences Workshop*, pages 47–49, 1992.

[73] P. Meer, J-M. Jolion, A. Rosenfeld. A fast parallel algorithm for blind estimation of noise variance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (3):216–223, 1990.

[74] M.L. Uss, B. Vozel, V.V. Lukin, K. Chehdi. Local Signal-Dependent Noise Variance Estimation from Hyperspectral Textural Images. *IEEE Journal of Selected Topics in Signal Processing*, 5 (3):469 – 486, 2011.

[75] J. Meola, M.T. Eismann, R.L. Moses, J.N. Ash. Modeling and estimation of signal-dependent noise in hyperspectral imagery. *Applied Optics*, 50 (21):3829–3846, 2011.

[76] N. Acito, M. Diani, G. Corsini. Signal dependent noise modeling and model parameter estimation in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (8):2597–2971, 2011.

[77] B. Clarke, E. Fokoué, H.H. Zhang. *Principles and theory for data mining and machine*. Springer, 2009.

[78] B.C. Levy. *Principles of signal detection and parameter estimation*. Springer, 2008.

[79] Z. Wang. *A framework for the analysis and evaluation of optical imaging systems with arbitrary response functions*. ProQuest, 2008.

A. Publications resulting from this work

- K. Cawse-Nicholson, S. Damelin, A. Robin, M. Sears. A parameter free approach for determining the intrinsic dimension of a hyperspectral image using Random Matrix Theory. Submitted to *IEEE Transactions on Image Processing* in July 2011.

- K. Cawse-Nicholson, A. Robin, M. Sears, The effect of spectrally correlated noise on noise estimation methods for hyperspectral images. *Workshop in Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2012.

- K. Cawse, A. Robin, M. Sears. The effect of noise whitening on methods for determining the intrinsic dimension of a hyperspectral image. *Workshop in Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2011.

- K. Cawse, M. Sears, A. Robin, S. Damelin, K. Wessels, F. van den Bergh, R. Mathieu. Using random matrix theory to determine the number of endmembers in a hyperspectral image. *Workshop in hyperspectral image and signal processing: evolution in remote sensing*, 2010