

https://dspace.org/

On the estimation and removal of noise in hyperspectral images

Holgate, Gavin

2016-01-19
http://hdl.handle.net/10539/19338

Downloaded from a DSpace repository. https://dspace.org/

On the Estimation and Removal of Noise in Hyperspectral Images

Gavin Holgate

Supervisor:

Prof Michael Sears

July 14, 2015

Acknowledgements

While it is quite impossible to thank every single person who has given me help, insight and support during the completion of this dissertation, I would like to take the opportunity to thank a few.

To my parents for their unquestioning support of my desire to take my knowledge a step further and explore the world of hyperspectral images.

To Deborah Haselau for constant support, encouragement and assistance in proof reading.

To Greg Mountford for your invaluable insights into the styles of my presentation and observations that led to a more refined and inclusive account of my work.

To Dr Asad Mahmood for technical help and invaluable input from an experienced perspective.

To all those whose research and work I have used and drawn insight from, no research can ever be done in isolation and it is a privilege to build on your great works and ideas.

Finally to Professor Michael Sears for guiding me through the entire experience. The roads we followed and ideas we tested, finding interesting results, knowledge and ideas worth sharing. I hope to work with you again.

Thank you!

Abstract

Hyperspectral images find application in many areas of modern society, we use them for land surveying, core sample analysis, in the conservation and forestry industries and many more.

A major problem in hyperspectral images is how to deal with noise. Many methods that analyse hyperspectral images either need clean images or accurate estimations of the noise statistics in the images. The goal of this dissertation is to present and compare methods for statistic estimation and noise removal. We use an artificial hyperspectral image to study some existing methods and develop some new ones based on existing methods, specifically the BM3D algorithm. We test methods that estimate the level of the noise present in an image, methods that estimate the structure of the noise and methods that remove noise. We analyse all the methods under an additive noise model and consider spectrally correlated and uncorrelated noise. Within our investigations we investigate different types of correlation. We will show the strengths that the various methods have and establish a way to approach treating a hyperspectral image with no information beyond the image itself.

Using our observations and insights from the experiments on the artificial data we analyse some radiance data from the AVIRIS instrument. We show that the additive signal independent part of the noise is small but not negligible. We also show some evidence for the structure of the noise in the AVIRIS instrument.

Contents

1 Introduction **1**

1.1 Hyperspectral Images 1

1.2 Noise in Hyperspectral Data 4

1.3 Notation 5

1.4 Abbreviations 6

1.5 Noise Model 7

2 Discussion regarding noise statistics estimation methods for hyperspectral data **8**

2.1 Meer's Method 8

2.2 Roger's Method 9

2.3 Regression estimator 10

2.4 Modified Linear Regression 11

3 Discussion regarding noise removal methods **13**

3.1 BM3D 13

3.1.1 Block Matching 13

3.1.2 Collaborative Filtering 13

3.2 BM4D 15

4 Proposed methods for noise removal **16**

4.1 Band by band BM3D (bBM3D) 16

4.2 PCA-BM3D 16

5 Proposed experiments **17**

5.1 Data 17

5.1.1 Artificial Data 17

5.1.2 Real Data 18

5.2 Experiments 19

5.2.1 Artificial Image 19

5.2.2 Real Data 22

6 Results **22**

6.1 Artificial data 22

6.1.1 Meer's Method 23

6.1.2 Roger's Method 26

6.1.3 Standard Regression 30

6.1.4 Modified Regression Method 37

6.1.5 BM3D Bandwise 50

6.1.6 PCA-BM3D 54

6.1.7 BM4D 57

6.1.8 Summary 61

6.1.9 Concluding Remarks 67

6.2 Real Data 68

6.2.1 First Pair of Spectrometers 68

6.2.2 Second Pair of Spectrometers 76

6.2.3 Signal to noise ratio in the instrument 83

7 Conclusion **85**

List of Figures

1 Hyperspectral data cube 2

2 AVIRIS spectral intensities 3

3 Illustration of block matching 14

4 Illustration of the BM3D algorithm. 15

5 Illustration of the PCA BM3D algorithm 17

6 Subsets of AVIRIS Cuprite used 19

7 Meers method estimate of noise variance -10^{-2} 23

8 Meers method estimate of noise variance -10^{-3} 24

9 Meers method estimate of noise variance -10^{-4} 24

10 Meers method estimate of noise variance -10^{-5} 25

11 Meers method estimate of noise variance -10^{-6} 25

12 Rogers method estimate of noise variance -10^{-2} 27

13 Meers method estimate of noise variance -10^{-3} 27

14 Meers method estimate of noise variance -10^{-4} 28

15 Meers method estimate of noise variance -10^{-5} 28

16 Meers method estimate of noise variance -10^{-6} 29

17 Standard regression method estimate of noise variance -10^{-2} 30

18 Standard regression method estimate of noise variance -10^{-3} 31

19 Standard regression method estimate of noise variance -10^{-4} 31

20 Standard regression method estimate of noise variance -10^{-5} 32

21 Standard regression method estimate of noise variance -10^{-6} 32

22 Standard regression method estimate of noise -10^{-2} 34

23 Standard regression method estimate of noise -10^{-3} 34

24 Standard regression method estimate of noise -10^{-4} 35

25 Standard regression method estimate of noise -10^{-5} 35

26 Standard regression method estimate of noise -10^{-6} 36

27 MR1 method estimate of noise variance -10^{-2} 38

28 MR1 method estimate of noise variance -10^{-3} 38

29 MR1 method estimate of noise variance -10^{-4} 39

30 MR1 method estimate of noise variance -10^{-5} 39

31 MR1 method estimate of noise variance -10^{-6} 40

32 MR3 method estimate of noise variance -10^{-2} 41

33 MR3 method estimate of noise variance -10^{-3} 42

34 MR3 method estimate of noise variance -10^{-4} 42

35 MR3 method estimate of noise variance -10^{-5} 43

36 MR3 method estimate of noise variance -10^{-6} 43

37 MR5 method estimate of noise variance -10^{-2} 44

38 MR5 method estimate of noise variance -10^{-3} 45

39 MR5 method estimate of noise variance -10^{-4} 45

40 MR5 method estimate of noise variance -10^{-5} 46

41 MR5 method estimate of noise variance -10^{-6} 46

42 BM3D method estimate of noise -10^{-2} 51

43 BM3D method estimate of noise -10^{-3} 51

44 BM3D method estimate of noise -10^{-4} 52

45 BM3D method estimate of noise -10^{-5} 52

46 BM3D method estimate of noise -10^{-6} 53

47 PCA-BM3D method estimate of noise -10^{-2} 54

48 PCA-BM3D method estimate of noise -10^{-3} 55

49 PCA-BM3D method estimate of noise -10^{-4} 55

50 PCA-BM3D method estimate of noise -10^{-5} 56

51 PCA-BM3D method estimate of noise -10^{-6} 56

52 BM4D method estimate of noise -10^{-2} 57

53 BM4D method estimate of noise -10^{-3} 58

54 BM4D method estimate of noise -10^{-4} 59

55 BM4D method estimate of noise -10^{-5} 59

56 BM4D method estimate of noise -10^{-6} 60

57 BM4D method sigma estimate sensitivity 60

58 Variance estimation method comparison for i.i.d. noise 62

59 Noise estimation method comparison for i.i.d. noise 62

60 Variance estimation method comparison for single noise 63

61 Noise estimation method comparison for single noise 64

62 Variance estimation method comparison for weak noise 65

63 Noise estimation method comparison for weak noise 65

64 Variance estimation method comparison for strong noise 66

65 Noise estimation method comparison for strong noise 67

66 First two spectrometers σ^2 according to Rogers method 69

67 First two spectrometers σ^2 according to the standard regression method . . . 70

68 First two spectrometers σ^2 according to the MR1 method 71

69 First two spectrometers σ^2 according to the MR3 method 71

70 First two spectrometers σ^2 according to the MR5 method 72

71 Noise determined by BM4D for first pair of spectrometers for the first subset 72

72 Noise determined by BM4D for first pair of spectrometers for the second subset 73

73 Noise determined by BM4D for first pair of spectrometers for the third subset 73

74 The variance of the "noise" removed by BM4D for the first pair of spectrometers for the first subset. 74

75 The variance of the "noise" removed by BM4D for the first pair of spectrometers for the second subset. 74

76 The variance of the "noise" removed by BM4D for the first pair of spectrometers for the third subset. 75

77 Correlation matrix estimates determined by MR20 for the three subsets in the first pair of spectrometers 75

78 Second two spectrometers σ^2 according to Rogers method 76

79 First two spectrometers σ^2 according to the standard regression method . . . 77

80 First two spectrometers σ^2 according to the MR1 method 78

81 First two spectrometers σ^2 according to the MR3 method 78

82 First two spectrometers σ^2 according to the MR5 method 79

83 The difference in the pixel values through each band for three pixels in the image for the first subset. 79

84 The difference in the pixel values through each band for three pixels in the image for the second subset. 80

85 The difference in the pixel values through each band for three pixels in the image for the third subset. 80

86 The variance of the "noise" removed by BM4D for the second pair of spectrometers for first the subset. 81

87 The variance of the "noise" removed by BM4D for the second pair of spectrometers for second the subset. 81

88 The variance of the "noise" removed by BM4D for the second pair of spectrometers for third the subset. 82

89 The correlation matrix estimates determined by MR20 for the three subsets in the second pair of spectrometers 82

90 Lab SNR measurements for AVIRIS 84

91 SNR estimates made using MR1 84

List of Tables

1 Bands used in the artificial image together with wavelength 18

2 MR1, MR3 and MR5 correlation estimates coloured noise 10^{-5} 47

3 MR1, MR3 and MR5 correlation estimates coloured noise 10^{-6} 48

4 MR1, MR3 and MR5 correlation estimates single noise 10^{-5} 48

5 MR1, MR3 and MR5 correlation estimates single noise 10^{-6} 48

6 MR1, MR3 and MR5 correlation estimates weak noise 10^{-5} 48

7 MR1, MR3 and MR5 correlation estimates weak noise 10^{-6} 49

8 MR1, MR3 and MR5 correlation estimates strong noise 10^{-5} 49

9 MR1, MR3 and MR5 correlation estimates strong noise 10^{-6} 49

1 Introduction

In this dissertation we will be investigating two principal problems encountered when considering noise in hyperspectral images. The first of these problems is determining the statistics of the noise. Knowing the statistics is often a necessary first step in hyperspectral image processing whether it be noise removal, pattern recognition or intrinsic dimension calculation. We investigate methods for determining noise assuming an additive spatially uncorrelated noise model. The second problem is that of noise removal, that is cleaning up images so the data can be more accurately processed and interpreted.

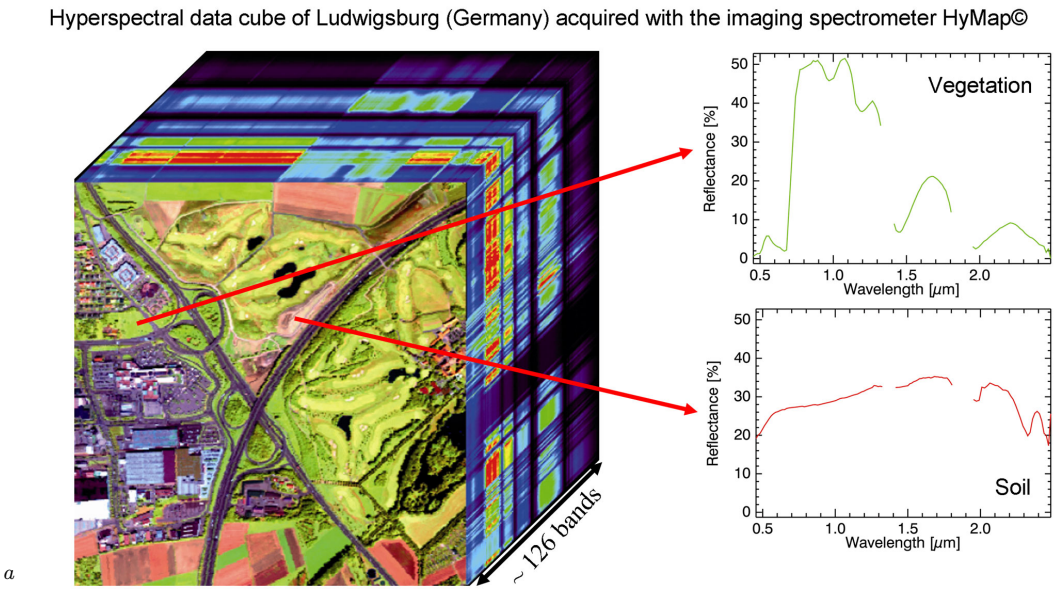
We will be comparing some methods that are well established for different types of noise. In addition we will explore methods that we propose based on the BM3D [1, 2] algorithm for estimating the noise statistics in grayscale images as well as a new spatial-spectral regression approach. We begin by discussing hyperspectral images and a spatially independent noise model. The problem is presented in the introduction and sections 2 and 3 serve as a literature review.

1.1 Hyperspectral Images

A hyperspectral image is a data set, usually taken over a large spectral range. The result is an image composed of many layers or bands (upwards of 100) analogous to RGB images which are composed of only three bands. Hyperspectral images typically record measurements which extend outside of the visible spectrum (stretching into the ultra violet range and long wave infra red). The AVIRIS (Airborne Visible/Infrared Spectrometer) spectrometer for example takes readings from 400nm to 2500nm over 224 bands [3]. Hyperspectral images exhibit a high amount of spectral redundancy as a result of adjacent bands being highly correlated [4].

Hyperspectral images play an important role wherever spectral analysis is used [3,5–8]. Many important minerals all have characteristic spectral profiles and the same is true for types of vegetation and soil [9]. figure 1 shows a visualisation of the hyperspectral data cube as well as spectral profiles for vegetation and soil.

Hyperspectral images play an important role in many industrial applications. They are exploited in the mining industry in the analysis of drill core samples where the spectral information is valuable in the analysis of the composition of the core sample [10]. They are also used in ground surveying and remote sensing. Hyperspectral cameras are fitted to planes and satellites and take pictures of the earth below. These images can then be analysed to find



a

Figure 1: Image showing a typical data cube representation of a hyperspectral image. The graphs represent the spectral profiles of vegetation and soil. These spectra would be reflected in the hyperspectral image as intensities in certain bands. The front most image in the block is a false colour composite of the image created by setting three bands of the image equal to red, green and blue channels and printing the colour image with them. This is why the picture appears to have an odd colour.

[a]Source: `http://www.dlr.de/eoc/en/DesktopDefault.aspx/tabid-6206/10515_read-23248/gallery-1/gallery_read-Image.60.20017`

fertile land and discover features that simply would not be discernible from regular imagery.

Hyperspectral images are often taken using different spectrometers (cameras) for different spectral ranges. The AVIRIS data for example is the combination of four spectrometers. In figure 2 we can see the ends of the spectrometers. Note also the lower signal in the spectrometers from 1250nm onwards (the second pair of spectrometers). This feature is attributed to the fact that the sun radiates as a blackbody and as such has a non-uniform emission spectrum.

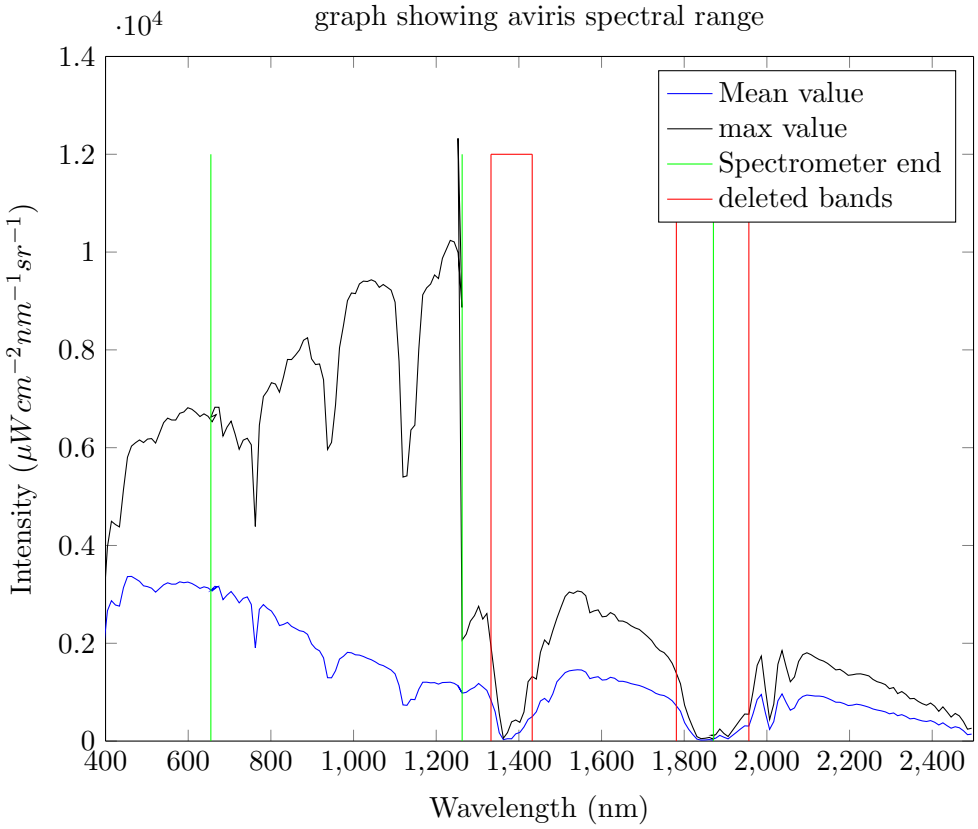


Figure 2: Graph of AVIRIS intensities by band number for a subsection of the radiance of the Cuprite scene. The mean intensity as well as the maximum intensity are shown for the sake of comparing spectral responses. The red blocks indicate bands that are "dead" resulting from atmospheric absorption causing no light to reach the spectrometer in those bands. These bands are ignored in analysis as they contain no useful information. The green lines show the ends of the individual spectrometers that make up the instrument.

1.2 Noise in Hyperspectral Data

A problem encountered in hyperspectral imagery, as with any imagery, is the presence of noise. Noise corrupts and masks the true signal, hiding information. Noise effects methods that attempt to extract endmembers from hyperspectral images or unmix them [11, 12].

Noise arises from various factors such as interference from the atmosphere, physical defects in the sensor, electrical interference and numerous other sources. It is often assumed that noise is spatially independent (additive noise [13]) as a simplifying assumption. In general the noise is correlated to the signal as well. Indeed the AVIRIS data has been shown to have significant signal dependency [14].

The noise that is independent of the signal is inherent to the sensor and depends on temperature. This noise is usually assumed to be stationary zero mean Gaussian noise. This noise is often referred to in the literature as instrument noise. While this noise is getting better in modern instruments it is still non-negligible.

The signal dependant part of the noise has a variance that is proportional to the magnitude of the signal. At each pixel in the sensor the camera catches photons of light over a particular integration time and outputs a magnitude based on the number of photons collected. Since the signal is dependant on the photons and the noise is dependent on the signal, this noise is referred to as photonic noise.

In aerial spectrometer readings, the camera is flown in a plane over the region that is being surveyed. Since we cannot fly as slowly as we would like the integration times for the pixels is short and therefore less photons reach the sensor. It should also be noted that as the number of bands in the instrument increase (i.e. as the bandwidth decreases), there will be less photons per band. These two facts would tend to make the photonic noise smaller whilst the instrument noise would remain the same.

We need ways to identify and remove noise from images without destroying data in the process or introducing artefacts or fictitious spectra. In the literature various techniques exist to attenuate noise in hyperspectral images.

The question is then: how can we remove the noise? This is not a trivial question since it is impossible to say definitively what is and is not actually noise and whether a feature is noise or signal. Due to the spatial resolution of the instruments, one pixel represents an area of the sample being scanned, and so pixels contain mixed spectral information from more than one material. This is particularly the case in aerial or satellite surveys which are typically

taken at high altitude so that a large area can be surveyed. In these images one pixel in the image can represent 400 square meters or more on the ground.

Spectral unmixing methods [15,16] work by assuming there are only so many pure materials in the image, each with its own unique spectrum across the bands. A comprehensive overview of unmixing approaches is given in [15]. Under these assumptions the image can be decomposed into a sparse representation. The image is then represented as abundances of each material in each pixel. This can then be moved back into the image space (that is the the full p bands). These methods carry the danger of losing information about trace materials (i.e. materials which have very little presence in the sample) and are typically highly sensitive to spectrally correlated noise [16–18].

Principal components are used in a similar fashion usually keeping only enough components to explain 95-99% of the variance. The assumption is that the signal accounts for the greatest variation in the image and so neglected components represent noise. Thus we discard or neglect components with very small presence in the image. Again information is lost since not all of the signal is strongly represented and even the low eigenvalue components will contain some signal in general.

Regression is another standard approach to denoise hyperspectral images. In general regression is carried out on a per pixel level by expressing every band as a combination of all the others. Regression can also be performed spatially using adjacent pixels in the band as well as corresponding pixels in adjacent bands. These methods generally assume additive spectrally uncorrelated noise and perform poorly when noise is correlated [16, 19].

1.3 Notation

Throughout this dissertation we will make use of the following notation conventions.
Any matrix will be denoted by a boldface capital (i.e. $\mathbf{A}$). Vectors will be represented by boldface lower case letters (i.e. $\mathbf{a}, \boldsymbol{\gamma}$,...) and are assumed to be column vectors. Row vectors will be indicated by ^{T}. Scalars will simply be written as lower case letters.
In addition the ith column of a matrix will be written as the matrix and a subscript (i.e. $\mathbf{A}_i$), the matrix whose i^{th} column has been removed will be written as $\mathbf{A}^{(i)}$ and the matrix whose i^{th} column and j^{th} row have been omitted will be $\mathbf{A}^{(i)(j)}$.
Special characters that will be used globally are:
p - the number of spectral bands in the image
$\mathbf{S}$ - the signal (i.e. the uncorrupted noise free image)

$\mathbf{X}$ - the data (observation) matrix
$\mathbf{C}$ - the correlation matrix of the noise
$\mathbf{\Delta}$ - the covariance matrix of the noise
σ_i^2 - the variance of the noise in the ith band equivelent to $\mathbf{\Delta}_{ii}$
$\mathbf{N}$ - the noise matrix (corruption) distributed as $N(0, \mathbf{\Delta})$
M - the number M of pixels in the image
δ_{ij} - the Kronecker delta function. Its value is 1 if $i = j$ and zero otherwise.
$x_{i,j,k}$ - the kth band of the pixel at location i, j.

In addition ̂will denote an estimate produced by a method of the quantity of interest whether it be the signal ($\hat{\mathbf{S}}$), noise variance ($\hat{\sigma}^2$) etc. and a bar will denote the average e.g. $\bar{\sigma}$ would represent the average variance in the image across all bands.

1.4 Abbreviations

Abbreviations used in this dissertation in text and particularly graph legends are laid out for easy reference.
BM3D - Block matching with collaborative 3D filtering denoising method (for grayscale images).
bBM3D - BM3D method applied bandwise (for hyperspectral image).
SNR - Signal to noise ratio.
MRn - The modified regression method with 2n off diagonals considered.
PCA-BM3D - Principal component augmented BM3D (for hyperspectral images).
BM4D - Block matching with collaborative 4D filtering denoising method (for hyperspectral images).
Rog - Roger's regression based method for variance estimation.
std - Standard regression method for variance estimation and image denoising.
Meer - Meer's method for variance estimation.
single - Noise correlated with only one neighbour on either side.
weak - noise "weakly correlated" with three adjacent neighbours on either side. The correlation matrix is a seven diagonal matrix with diagonal entries $[0.1, 0.3, 0.4, 1, 0.4, 0.3, 0.1]$.
strong - noise "strongly correlated" with three adjacent neighbours on either side. The correlation matrix is a seven diagonal matrix with diagonal entries $[0.3, 0.45, 0.65, 1, 0.65, 0.45, 0.3]$.
i.i.d. - independent and identically distributed noise.

1.5 Noise Model

In general the noise may be signal dependant [20] as well as additive. This corresponds to a multiplicative noise model. Methods exist [21–23] to determine whether the noise is signal dependant. We will, however, follow the commonly used approach and assume that the noise is independent of the signal. i.e. Assuming an image of M pixels, the noise model we will be considering is an additive noise model and can be expressed as follows:

$$\mathbf{x}_i = \mathbf{s}_i + \mathbf{n}_i.$$

The subscript i refers to the ith pixel in the image and has the range $1, ..., M$. $\mathbf{x}_i$ is the observed spectrum of the ith pixel i.e. x_i^j is the observed value of the jth spectral band for the ith pixel. $\mathbf{s}_i$ is the true signal value through all bands for the ith pixel, that is the image without noise i.e. s_i^j is the true value of the jth spectral band for the ith pixel. $\mathbf{n}_i$ is the noise through all bands for the ith pixel. We assume that $\mathbf{n}_i$ is zero mean Gaussian noise distributed according to $\mathbf{\Delta}$ where $\mathbf{\Delta}$ is the covariance matrix of the noise and defines how bad the noise is in each band (the variance) as well as the crosstalk it has with every other band (correlation).

There are three principal noise types of interest in the additive noise model described.

The first is uniform i.i.d white Gaussian noise, which corresponds to

$$\mathbf{\Delta} = \sigma^2 \mathbf{I} \tag{1}$$

where $\mathbf{I}$ is the identity matrix and σ^2 is the variance of the noise in each band.

The second is non-uniform (coloured) Gaussian noise, which corresponds to

$$\mathbf{\Delta}_{ij} = \begin{cases} \sigma_i^2 & i = j \\ 0 & i \neq j \end{cases} \tag{2}$$

where σ_i^2 is the variance of the noise in the ith band.

The third noise model is the general case of correlated zero mean Gaussian noise. In this case we write

$$\mathbf{\Delta} = \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{C} \mathbf{\Sigma}^{\frac{1}{2}} \tag{3}$$

where $\mathbf{C}$ is the correlation matrix of the noise and

$$\mathbf{\Sigma}_{ij} = \begin{cases} \sigma_i^2 & i = j \\ 0 & i \neq j \end{cases}$$

where σ_i^2 is the variance of the noise in the ith band. $\mathbf{C}$ is a positive definite symmetric matrix with ones down the main diagonal and off diagonal elements with values in the range $[-1, 1]$ [24]. Indeed any matrix fitting these criteria can be a correlation matrix. Typically the correlation matrix is taken to be a $(2n + 1)$-diagonal matrix, reflecting the reasonable assumption that the correlation is strong only between relatively close adjacent bands.

It is often assumed that noise correlation is negligible. It has however been shown [4, 11] that this is not the case. Spectrally correlated noise has a big impact on the effectiveness of many hyperspectral image processing applications as the methods may see the noise as signal. This is particularly true in the case of signal subspace identification and unmixing [18, 25]. Noise estimation and removal algorithms that use spectral information are often impacted by correlated noise.

2 Discussion regarding noise statistics estimation methods for hyperspectral data

Noise estimation (determining the statistics of the noise present in the image) is a key step in many image processing applications. In our case the statistic of interest is the covariance matrix. Below we detail some existing methods used to estimate the variance in hyperspectral images.

2.1 Meer's Method

Meer's method [27] is a method employed to estimate the noise variance of a grayscale image. The method works by producing variance estimates for a series of tessellations, comparing successive generations to refine the estimate. The largest power of 2 square subsection of the image is taken on which to perform the tessellations (i.e. a 2^nx2^n subset where $n \leq [log_2(x)]$ and x is the smallest dimension of the image).

The steps per tessellation are described as follows:

- The image is tessellated into 2^lx2^l blocks where l is the level of the algorithm and increases by 1 each tessellation.
- The variances for each of the blocks is computed and the smallest 4 are retained (except for the nth tessellation where there is only one number).

- Slippage tests are performed on the 4 numbers to eliminate outliers. These tests are not performed on the $(n-1)^{th}$ and n^{th} tessellations.
- The average of the remaining values is taken and stored. The ratios of these estimates are compared at every step and if the ratios fail a certain test the algorithm terminates and is not able to return an estimate.

The variance estimate for the image is then returned based on a large set of rules that take into account what the authors refer to as the "busyness" of the image which is an indication of the signal variance.

Since the method was designed for grayscale images it has no power to determine the correlation of the noise in hyperspectral images. The output of the method when applied to a hyperspectral image is an estimation of the variance of the noise present in each band.

This method is worth considering because it is a blind method, that is it requires no input other than the image itself.

2.2 Roger's Method

Roger's method [28] is a regression based estimator for the variance in a hyperspectral image. The method exploits both spectral correlation and spatial correlation in the data. It has been shown [19] that this method (referred to as SSDC) provides superior results over the other regression based estimators that the authors tested.

The method is carried out as follows:

- The image is considered one band at a time.
- The image is divided up into blocks.
- A regression is performed on each block to produce an estimate of the block. The regression is carried out on a per pixel basis using one adjacent pixel as well as the corresponding pixels in the two adjacent bands

$$r_{i,j,k} = ax_{i,j+1,k} + bx_{i,j,k-1} + cx_{i,j,k+1} + d - x_{i,j,k}. \tag{4}$$

  Of course the formula changes at the first and last band where the two nearest spectral neighbours are used (i.e. in one spectral direction only).

- The average of the residuals squared is considered an estimate of the variance of the noise in the block.
- All the estimates (from each block) are collected and formed into a homogeneous group. This is done by discarding the top and bottom 15% of the estimates which are considered outliers.
- The mean of the remaining estimates is taken as the estimate of the variance of the noise in the band under consideration.

It was found that eliminating the top and bottom 15% of the estimates had no significant impact on the estimation and as such is not implemented in order to reduce complexity. The block size recommended in [28] is 16x16 whereas [19] claimed a block size of 6x6 is optimal. We suspect that the block size that is optimal will depend on the spatial homogeneity of the image. The method makes the assumption that adjacent pixels are the same and this is erroneous if there is a high degree of spatial clutter. Block size will be one of the parameters considered in our experimental analysis of Roger's method.

We suspect that the dependence of this method on the two adjacent bands could make it very sensitive to correlation in the noise. This is because the correlation in the noise is expected to be strongest between adjacent bands.

This method outputs an estimate of the variance present in each band and gives no information about the correlation present in the image.

2.3 Regression estimator

When performing a multiple linear regression on a hyperspectral image one attempts to write every band in terms of the remaining bands. To do this $p-1$ regression coefficients need to be found for each band (where p is the number of bands in the image) that when multiplied by the remaining bands produces a best estimate in the least squares sense of the image. This can be expressed mathematically as follows.

We write $\mathbf{X}$, the data matrix, in one dimensional format with bands as columns and rows as pixels i.e. $\mathbf{X}$ is an $M \times p$ matrix. Then $\mathbf{X}_i$ is the ith band in the image and $\mathbf{X}^{(i)}$ is the image with the ith band omitted. Then the estimation $\hat{\mathbf{X}}_i$ of $\mathbf{X}_i$ is given by

$$\hat{\mathbf{X}}_i = \mathbf{X}^{(i)}\mathbf{b}^i$$

where $\mathbf{b}^i$ is the vector of the optimal regression coefficients corresponding to the ith band. When the image is reconstructed using the $\hat{\mathbf{X}}_i$ the new image $\hat{\mathbf{X}}$ is not exactly the same as the original image and the residual is taken to be the noise, that is

$$\hat{\mathbf{X}} - \mathbf{X} = \mathbf{N}.$$

Now the variance of this noise can be calculated as the estimates for σ_i^2.

Regression, however, performs unreliably in the case of correlated noise, this is because the correlation in the noise means that the noise can be represented to a greater extent as a linear combination of the noise present in adjacent bands. A way around this weakness and to a better estimate of the covariance matrix has recently been proposed [29].

2.4 Modified Linear Regression

For this approach [29] the observation is made that in performing the regression some noise is included in the estimation of the image. Under this observation $\mathbf{r}^i$, the ith residual, can be expressed as follows:

$$\begin{aligned} \mathbf{r}^i &= \mathbf{X}_i - \mathbf{X}^{(i)}\mathbf{b}^i \\ &= (\mathbf{S}_i + \mathbf{N}_i) - (\mathbf{S}^{(i)} + \mathbf{N}^{(i)})(\boldsymbol{\beta}^i + \boldsymbol{\gamma}^i) \end{aligned} \tag{5}$$

Here the optimal regression coefficient $\mathbf{b}^i$ has been broken up into a component "explaining" signal, $\boldsymbol{\beta}^i$, and a component "explaining" noise, $\boldsymbol{\gamma}^i$. Under the assumption of high redundancy in a hyperspectral image (i.e. the dimensions of the signal subspace is very much less than the number of bands) we can make the further assumption that $\boldsymbol{\beta}^i$ can be chosen to perfectly explain the signal. i.e.

$$\mathbf{S}_i = \mathbf{S}^{(i)}\boldsymbol{\beta}^i \tag{6}$$

using (6) we can write (5) as

$$\begin{aligned} \mathbf{r}^i &= \mathbf{S}_i + \mathbf{N}_i - \mathbf{S}^{(i)}\boldsymbol{\beta}^i - \mathbf{S}^{(i)}\boldsymbol{\gamma}^i - \mathbf{N}^{(i)}\boldsymbol{\beta}^i - \mathbf{N}^{(i)}\boldsymbol{\gamma}^i \\ &= \mathbf{N}_i - \mathbf{S}^{(i)}\boldsymbol{\gamma}^i - \mathbf{N}^{(i)}\mathbf{b}^i \end{aligned} \tag{7}$$

The last two terms in (7) represent the error made in estimating $\mathbf{N}_i$ by the residual alone.

Using these new expressions for the residuals we construct a method for estimating the noise covariance matrix $\boldsymbol{\Delta}$. Now we start by calculating $\mathbf{r}^{iT}\mathbf{r}^j$

$$\begin{aligned} \mathbf{r}^{iT}\mathbf{r}^j =& \mathbf{N}_i^T\mathbf{N}_j + \mathbf{b}^{iT}\mathbf{N}^{(i)T}\mathbf{N}^{(j)}\mathbf{b}^j + \boldsymbol{\gamma}^{iT}\mathbf{S}^{(i)T}\mathbf{S}^{(j)}\boldsymbol{\gamma}^j - \mathbf{b}^{iT}\mathbf{N}^{(i)T}\mathbf{N}_j \\ & - \mathbf{N}_i^T\mathbf{N}^{(j)}\mathbf{b}^j + \boldsymbol{\gamma}^{iT}\mathbf{S}^{(i)T}(\mathbf{N}^{(j)}\mathbf{b}^{\mathbf{j}} - \mathbf{N}_j) - (\mathbf{N}_i^T - \mathbf{b}^{iT}\mathbf{N}^{(i)T})\mathbf{S}^{(j)}\boldsymbol{\gamma}^j. \end{aligned} \tag{8}$$

Note that according to the model $\mathbf{S}$ and $\mathbf{N}$ are uncorrelated and since M is large when we divide (8) through by M any terms that look like $\mathbf{S}^T\mathbf{N}$ tend to zero. First we define a new notation to simplify calculations. We write $\tilde{\boldsymbol{\beta}}^i$ as the coefficient vector for $\boldsymbol{\beta}$, augmented with a 0 in position i and similarly for $\mathbf{b}$ and $\boldsymbol{\gamma}$. Note that $\tilde{\mathbf{b}}^{iT}\boldsymbol{\Delta}\tilde{\mathbf{b}}^j \equiv \mathbf{b}^i\boldsymbol{\Delta}^{(i)(j)}\mathbf{b}^j$ where $\boldsymbol{\Delta}^{(i)(j)}$ represents $\boldsymbol{\Delta}$ with the ith row and j^{th} column omitted. Also note $\boldsymbol{\Delta}_i$ is the ith column of $\boldsymbol{\Delta}$. So dividing (8) by M we get

$$\frac{\mathbf{r}^{iT}\mathbf{r}^j}{M} \approx \Delta_{ij} + \tilde{\mathbf{b}}^{iT}\boldsymbol{\Delta}\tilde{\mathbf{b}}^j - \tilde{\mathbf{b}}^{iT}\boldsymbol{\Delta}_j - \tilde{\mathbf{b}}^{jT}\boldsymbol{\Delta}_i + \tilde{\boldsymbol{\gamma}}^{iT}\mathbf{S}\tilde{\boldsymbol{\gamma}}^j \tag{9}$$

We neglect the last term in (9). This omission was found by the authors to be small in practice [29]. Noting that the covariance matrix is symmetric equation (9) gives us $\frac{1}{2}p(p+1)$ equations in the $\frac{1}{2}p(p+1)$ unknowns $\boldsymbol{\Delta}_{kl}$. There are $\frac{1}{2}p(p+1)$ non zero elements in the upper triangle of the covariance matrix including the main diagonal. We can rewrite equation (9) as

$$\frac{\mathbf{r}^{iT}\mathbf{r}^j}{M} = \sum_{r,s=1}^{p}[\hat{\Delta}_{rs}\delta_{ir}\delta_{js} - \tilde{b}_r^i\hat{\Delta}_{rs}\delta_{js} - \tilde{b}_r^j\hat{\Delta}_{rs}\delta_{is} + \tilde{b}_r^i\hat{\Delta}_{rs}\tilde{b}_s^j] \tag{10}$$

where δ is the Kronecker delta function. Furthermore, collecting terms, the coefficient of Δ_{kl} in the ij^{th} equation (10) is

$$a(i,j,k,l) = \delta_{ik}\delta_{jl} - \tilde{b}_k^i\delta_{jl} - \tilde{b}_k^j\delta_{il} + \tilde{b}_k^i\tilde{b}_l^j + (1-\delta_{lk})[\tilde{b}_l^i\tilde{b}_k^j - \tilde{b}_l^i\delta_{jk} - \tilde{b}_l^j\delta_{ik}] \tag{11}$$

with $i \geq j$ and $k \geq l$.

We write the equations (10) as a matrix equation

$$\mathbf{r} = \mathbf{A}\hat{\boldsymbol{\Delta}} \tag{12}$$

where $\mathbf{r}$ is the vector obtained by listing $\frac{1}{M}r^{iT}r^j$ by row, $\mathbf{A}$ is the matrix of coefficients obtained from (11), and $\hat{\Delta}$ is now written as a column vector taking one row at a time. In practice we make the assumption that the covariance matrix has limited non-zero diagonals which reduces the number of diagonals we have to consider. If, for example, we restrict our consideration to $2n+1$ diagonals we need only consider $\frac{1}{2}(n+1)(2p-n)$ equations in as many unknowns (The nth off diagonal has $p-n$ entries and the number of elements can then be found by using the arithmetic series formula.) This assumption is reasonable as we do not expect the noise in band 1 to be correlated to the noise in band 50. The fact that we force the estimation into $2n+1$-diagonal form also limits the error we accumulate from the omitted term in (9).

The authors claim that this formulation gives a refined estimate in the case of i.i.d. noise and an improved estimate in the case of correlated noise as well as an estimate of the correlation structure of the noise [29].

An important consideration in this method is the execution time required by the method. The value chosen for n determines the size of the matrix to be inverted so a potential trade-off between execution time and accuracy/structure determination warrants investigation.

3 Discussion regarding noise removal methods

3.1 BM3D

The BM3D algorithm is a new highly effective algorithm for attenuating noise in grayscale images [1,2]. This method makes use of the intrinsic spatial similarity in natural images. Extending this to hyperspectral images (in methods we describe later) may produce algorithms of superior quality and will hopefully produce methods which are less sensitive to correlation of the noise. The spatial assumptions made by this method may make it less effective when spatial homogeneity is absent and similar regions are not present in the image.

3.1.1 Block Matching

In natural images there is a high degree of similarity between image fragments. With this in mind, groups can be constructed of similar blocks. A block is a window of fixed size. Block matching is then carried out as follows:

- Choose a reference block in the image.
- Shift the window to a new block and compare its similarity to the reference block using some metric.
- If the blocks are measured to be similar, they are added to an array containing the reference block and all other similar blocks.

3.1.2 Collaborative Filtering

Once the blocks have been found and stacked in a group the assumption is made that the blocks are identical and the difference between them is due entirely to noise contained in the image. To deal with this, the group is transformed using the discrete cosine transformation

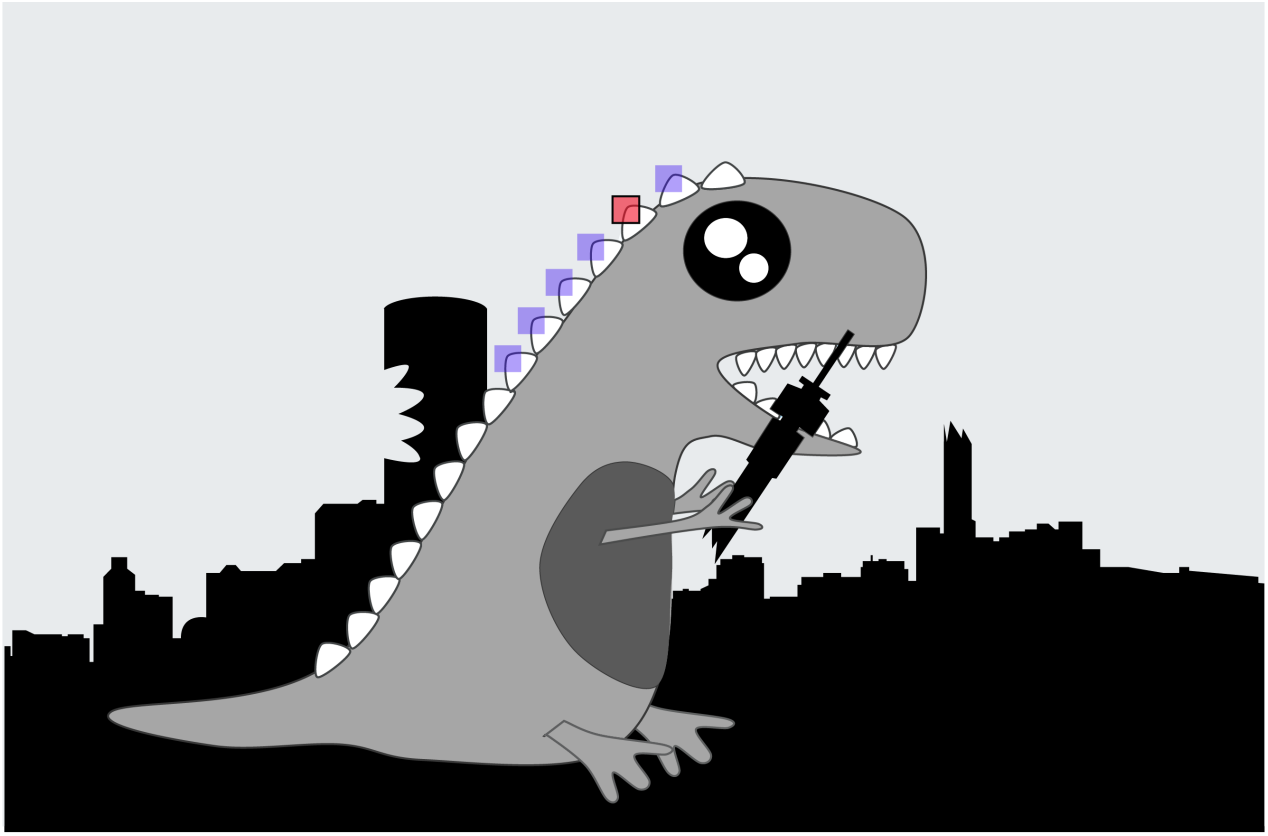


Figure 3: Illustration of block matching algorithm. The red block with the black border is the chosen reference block, the blue blocks are the ones deemed similar by the metric. Reference blocks are chosen periodically in the image so that every pixel of the image should be represented in at least one group. Note that the blue blocks do not extend on Godzilla's spines past the building even though they have the same structure. This is because the black of the building (which makes up about half the frame) makes the blocks too dissimilar to match.

and coefficients are discarded that account for the least data (smallest values). This effectively attenuates the noise in the blocks. The inverse transformation is then taken and estimates are obtained for each of the blocks in the group.

Since the blocks can, in general, overlap there are multiple estimates for each pixel once the algorithm has completed filtering on every group that has been found. These are then assigned weights and combined to produce an estimate of the image.

In the grayscale BM3D algorithm the filtering is carried out in two phases.

1. Hard threshholding

    - Block matching is carried out on the noisy image and groups are formed.
    - Collaborative filtering is performed using hard thresholding in transform domain.
    - Inverse transformation is performed and estimates of blocks are weighted and combined to produce image estimate.

2. Wiener filtering

    - Block matching is conducted on the image estimate produced in the hard threshholding phase.
    - The groups produced are also taken within the original noisy data.
    - Collaborative Wiener filtering is performed on the estimate group in transform space.
    - The Wiener coefficients found in the estimate data are used in the groups of the noisy data to filter the blocks in the transform domain.
    - The inverse transformation of the groups of filtered noisy data is taken and the blocks aggregated to produce the final clean image.

The process is shown simplified in figure 4. This algorithm serves as the base for the new methods proposed for hyperspectral images. It should be noted that this method takes an estimate of σ^2 as an input in order to establish the threshholds used in the method.

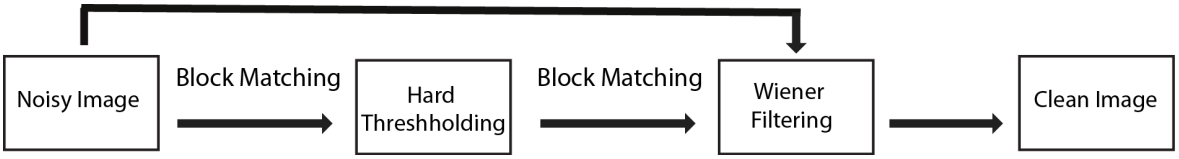


Figure 4: Illustration of the BM3D algorithm.

3.2 BM4D

The BM4D algorithm [30, 31] is a method developed by the same authors of BM3D specifically for multispectral/hyperspectral images. The process is essentially the same except that where the BM3D algorithm matches two dimensional blocks the BM4D algorithm matches 3 dimensional blocks. These data blocks are again a fixed size in the two spatial dimensions and fixed length down the spectral dimension.

The blocks are collected and stacked in a 4D array. The blocks are then transformed using separable transformations and again noise is attenuated in transform space. This is done first with hard-threshholding and the initial estimate is used to create better matches and obtain Wiener filtering coefficients which are again applied to the original image in transform space to produce the final estimate of the image. Again the spatial assumptions made by this method may make it less effective when spatial homogeneity is absent and similar regions are not present in the image.

This method, like the BM3D method, takes an estimate of σ^2 to establish the threshholding parameters.

4 Proposed methods for noise removal

4.1 Band by band BM3D (bBM3D)

This is a rudimentary implementation of the BM3D algorithm to hyperspectral images. The algorithm is simply applied to the hyperspectral image one band at a time. This method is costly in terms of execution time but is included because it may prove to be more effective than the other method in attenuating the noise. The inherent redundancy in the hyperspectral image has not been exploited.

Since this method uses no spectral information we expect that it will be insensitive to correlation in the noise.

4.2 PCA-BM3D

The idea of utilising principal component analysis together with the block matching algorithms has been used before in [32, 33].

In this formulation we exploit the intrinsic redundancy of hyperspectral images. Part of the BM3D algorithm was still implemented on each band. We process the image globally first using principal component analysis (PCA). The principal component analysis is conducted as follows:

- Principal components are taken for the image.

- The principal components accounting for a certain percentage of the variance in the image are kept and the rest is considered noise and discarded (the first few principal components describe the majority of the energy of the image and the lower energy PCs are believed to contain most of the noise [34]).
- We then reconstruct the image from the reduced basis.

This provides the first estimate required for the Wiener filtering part of the BM3D algorithm. We expect a huge reduction in time due to the fact that the block matching operation is only conducted once. The expectation is that this method will produce worse results but with more efficient time. We are concerned particularly by correlated noise which may be carried through the PCA step.

This process is shown simplified in figure 5

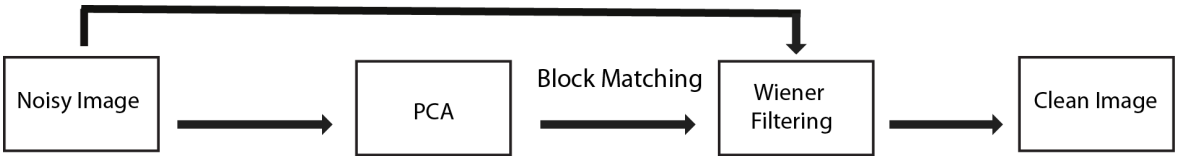


Figure 5: Illustration of the PCA BM3D Hybrid algorithm.

This method will also be investigated as a method to approximate the noise statistics in hyperspectral images. To accomplish this we will simply calculate the statistics of the "noise" that the method removes.

5 Proposed experiments

5.1 Data

5.1.1 Artificial Data

The artificial image was constructed as in [12, 19] from a 350x350 subset of the 1997 AVIRIS Cuprite reflectance data[1]. Firstly bad bands were removed, NFINDR [35] is then fed this data as well as an estimate of 20 for the intrinsic dimension. The endmembers determined by NFINDR are then used to demix the image and create an abundance map. These endmembers

[1]Source: `http://http://www.aviris.jpl.nasa.gov/data/free_data.html`

and the abundance map are used to construct the artificial image. This image because of its construction obeys the linear mixing model. We treat this image as noise free.

The bands used for the artificial image are given in table 1. The bands that are omitted are those we considered having poor signal or bad noise. In our analysis we only plot bands 3 - 170 because some methods had bad behaviour at the boundaries of the image making the graphs scale meaningless. We elected to plot all methods on the same domain so that comparison is easy.

Bands in Artificial image	Bands in AVIRIS	Wavelength Range (nm)
1 - 28	5 - 32	404 - 668
29 - 99	34 - 104	664 - 1323
100 - 133	116 - 149	1442 - 1771
134 - 181	171 - 218	1967 - 2436

Table 1: Table showing the construction of the artificial image. Bands were excluded as being too noisy or not containing enough signal. The wavelength is included for reference purposes.

5.1.2 Real Data

The real data we will be analysing is taken from the radiance data of the AVIRIS Cuprite scene from 11 June 2008[2]. We will take three subsets of the scene. The reason we take subsets from one scene is so that we have data which is acquired under the same conditions and on the same instrument. This has the advantage that (under the assumption of signal independent noise) we can use the consistency or lack thereof of the three estimates as an indicator of the accuracy.

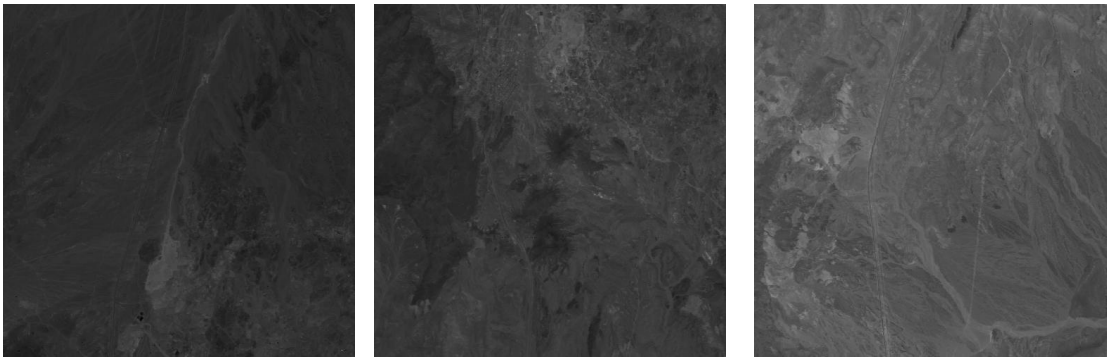
We ignore the following bands in our consideration:

- Bands 1-4: We consider these dead bands containing little useful data.
- Band 33: Considered a bad band with unacceptable signal to noise ratio.
- Bands 97-104: Short interval in the third spectrometer that is seperated from the other useful bands by noisy in the spectrometer.
- 105-128: Noisy or "dead" bands i.e. either black or containing unacceptable signal to noise ratio.

[2]Source: `http://http://www.aviris.jpl.nasa.gov/data/free_data.html`

- 150-170: Noisy or "dead" bands i.e. either black or containing unacceptable signal to noise ratio.
- 220-224: very low response as well as bad signal to noise ratio.

We treat the first two spectrometers separately from the last two. The reason we do this is because of the difference in spectral response - the first two spectrometers are brighter - as well as the large separation in wavelength when bad bands are omitted.



(a) First subset. (b) Second subset. (c) Third subset.

Figure 6: One band of subsections of the Cuprite scene taken from radiance data for experiments. Note that the third subset is brighter than the others showing spatial variation in the image.

5.2 Experiments

All experiments are performed on intel®Core^{TM} i5-2500K CPU 3.30GHz with 8GB RAM

5.2.1 Artificial Image

The artificial image will be analysed as our controlled experiment. We control two parameters in our experiments:

Firstly σ^2 - the global variance of the noise. This is essentially the severity of the noise and is of interest as we wish to study how the various methods hold up under different levels of noise. We will investigate σ^2 in the range $[10^{-6}, 10^{-2}]$ in five experiments. For the coloured

noise the diagonal of the covariance matrix will be a diagonal with entries chosen to mimic the colour estimated for the real AVIRIS data using the modified regression approach.

Secondly **C** - the correlation matrix. This represents the dependence of the noise in any given band on all others. This is of particular interest as spectral methods (principal component analysis, multiple regression etc.) often demonstrate dependencies on the correlation of the noise and usually perform poorly in its presence, mistaking noise for signal due to the correlation. We investigate three different correlation matrices at all noise levels:

- The identity matrix. This corresponds with i.i.d. and coloured uncorrelated noise and represents a baseline for method effectiveness.
- A three diagonal matrix with diagonals $[0.5, 1, 0.5]$. This corresponds to fairly strong correlation only between neighbouring bands. We expect this noise to impact Roger's method severely as it only uses immediate neighbours.
- A "weak" correlation matrix which is a seven diagonal matrix with diagonal entries $[0.1, 0.3, 0.4, 1, 0.4, 0.3, 0.1]$. This correlation is expected to impact all spectral methods by a noticeable amount.
- A "strong" correlation matrix which is a seven diagonal matrix with diagonal entries $[0.3, 0.45, 0.65, 1, 0.65, 0.45, 0.3]$. This is an extreme case and we believe it will demonstrate the stability of the methods very well.

All of these matrices are positive definite and symmetric and therefore can be correlation matrices.

While the methods explored have the same goals (statistic estimation or noise removal) they are also unique and present their own set of parameters to examine (whether it be block size in Roger's method or guesses at correlation structure in the modified regression method). These parameters will be examined and discussed in the sections pertaining to the method in question.

If interesting observations that warrant further exploration are made in our experiments then this will be dealt with in the result section of the particular method. We will describe the observation that prompted the interest, describe the further experimentation performed and present the results.

In our extended experiments larger correlation matrices may be required. If the correlation matrix does not meet the stronger criteria of being positive definite, we can, instead of

tinkering until we happen upon a positive definite matrix, simply correct the correlation matrix so that it is positive definite. A matrix is positive definite if and only if all of its eigenvalues are greater than zero. To this end we use a rank one correction method to increase any non-positive eigenvalues to small positive values. Suppose $\mathbf{A}$ is a matrix with negative eigenvalue λ and corresponding normalised eigenvector $\mathbf{x}$, and further let ϵ be a small positive number. The update we perform is then as follows

$$\mathbf{A}' = \mathbf{A} + (\epsilon - \lambda)\mathbf{x}\mathbf{x}^T.$$

$\mathbf{A}'$ remains symmetric if $\mathbf{A}$ is symmetric. There are two results that are important in the consideration as to whether the rank one correction is a feasible method and won't disturb the intended correlation shape too much. The first of these is to demonstrate that the method leaves the eigenbasis (that is the eigenvectors) unchanged and merely changes the eigenvalue in consideration.

Let λ be an eigenvalue with corresponding normalised eigenvector $\mathbf{x}$, further let ϵ be a small positive number

$$\begin{aligned}
\mathbf{A}'\mathbf{x} &= (\mathbf{A} + (\epsilon - \lambda)\mathbf{x}\mathbf{x}^T)\mathbf{x} \\
&= \mathbf{A}\mathbf{x} + \epsilon\mathbf{x}\mathbf{x}^T\mathbf{x} - \lambda\mathbf{x}\mathbf{x}^T\mathbf{x} \\
&= \lambda\mathbf{x} + \epsilon\mathbf{x} - \lambda\mathbf{x} && \mathbf{x}^T\mathbf{x} = 1 \\
&= \epsilon\mathbf{x}
\end{aligned}$$

Hence ϵ is an eigenvalue of $\mathbf{A}'$ with associated eigenvector $\mathbf{x}$. Secondly, we need to show that the other eigenvectors remain unchanged. If we let ξ be another eigenvalue of $\mathbf{A}$ with associated normalised eigenvector $\mathbf{y}$ then

$$\begin{aligned}
\mathbf{A}'\mathbf{y} &= (\mathbf{A} + (\epsilon - \lambda)\mathbf{x}\mathbf{x}^T)\mathbf{y} \\
&= \mathbf{A}\mathbf{y} + \epsilon\mathbf{x}\mathbf{x}^T\mathbf{y} - \lambda\mathbf{x}\mathbf{x}^T\mathbf{y} \\
&= \xi\mathbf{x} && \mathbf{x}^T\mathbf{y} = 0
\end{aligned}$$

Hence ξ is still an eigenvalue of the new matrix $\mathbf{A}'$ with associated eigenvector $\mathbf{y}$.

The methods discussed will be implemented and the noise remaining will be analysed to compare how effective the various noise removal methods are with different noise types and levels.

The residual noise is calculated as follows:

$$\mathbf{N}_{res} = \hat{\mathbf{S}} - \mathbf{S}.$$

The noise estimate is calculated as follows:

$$\hat{\mathbf{N}} = \mathbf{X} - \hat{\mathbf{S}}.$$

The matrix of interest for the assessment of the noise estimation methods is $\hat{\mathbf{N}}$. The variance of this matrix as well as its correlation matrix will be the important measures. The noise estimation methods' results will be compared directly with the statistics of the noise added. The time required by each method is also an important consideration and will be measured as well.

5.2.2 Real Data

The way the real data is processed will be informed by the results from the artificial data experiments. The course decided upon will be described when the results are presented.

6 Results

6.1 Artificial data

The results are presented first for each method for the artificial data.

The error in the variance estimate ($\hat{\sigma}^2 - \sigma^2$) is presented plotted against band number for the methods producing a noise statistic estimate. The variance of the noise remaining ($\mathbf{N}_{res}$) in the image is plotted against band number for the denoising methods. Finally the methods are compared to each other under each noise type and at each noise level.

You will notice there are spikes at the lower noise levels in most methods. These spikes correspond to features of the artificial image. They represent the boundaries of the spectrometers and bands that are slightly less bright in the short wave infra red side of the spectrum, the signal to noise ratio will then be smaller in these bands.

6.1.1 Meer's Method

Meer's method only gives an estimate of the variance in each band. In figures 7 - 11 we plot the results from our experiments.

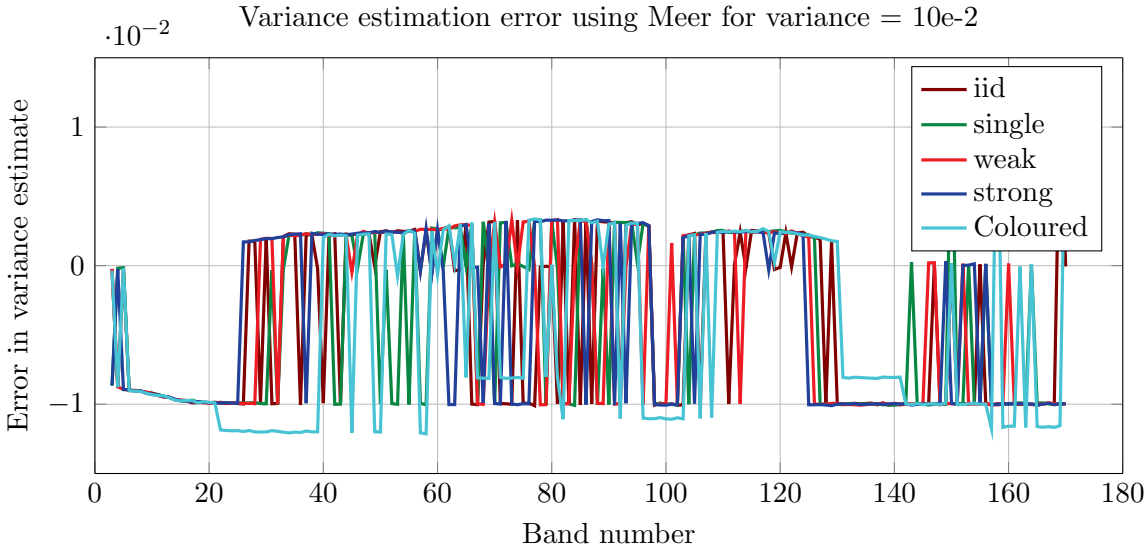


Figure 7: Meers method applied to noise with variance 10^{-2}. Note that the condition in Meer's method that decides whether the method thinks there is no noise present is triggered fairly often (indicated by an error of -10^{-2}). Otherwise the error in estimation is around 15%.

Meer's method has demonstrated acceptable proficiency at higher noise levels. It does however perform poorly at lower noise levels (variances less than 10^{-4}). The lower end of the noise levels that we tested are what we expect of modern instruments suggesting that Meer's method is not very useful in variance estimation in modern hyperspectral images.

The execution time for all noise types was of the order 120 seconds. The method could be parallelised treating each band simultaneously.

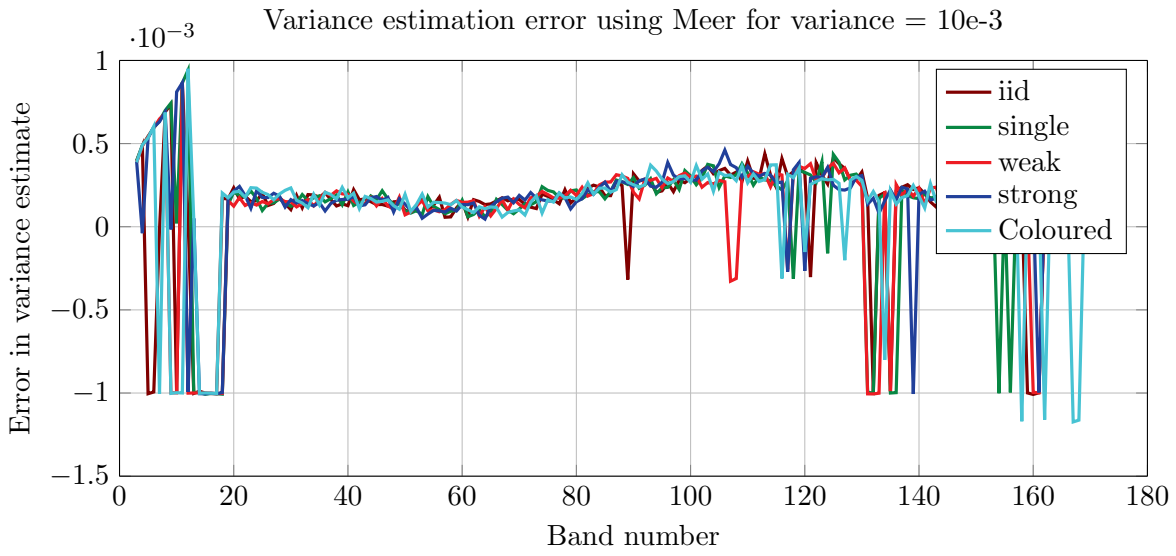


Figure 8: Meer's method applied to noise with variance 10^{-3}. Note again that the zero noise condition has been triggered fairly often but the error is otherwise around 15-25%.

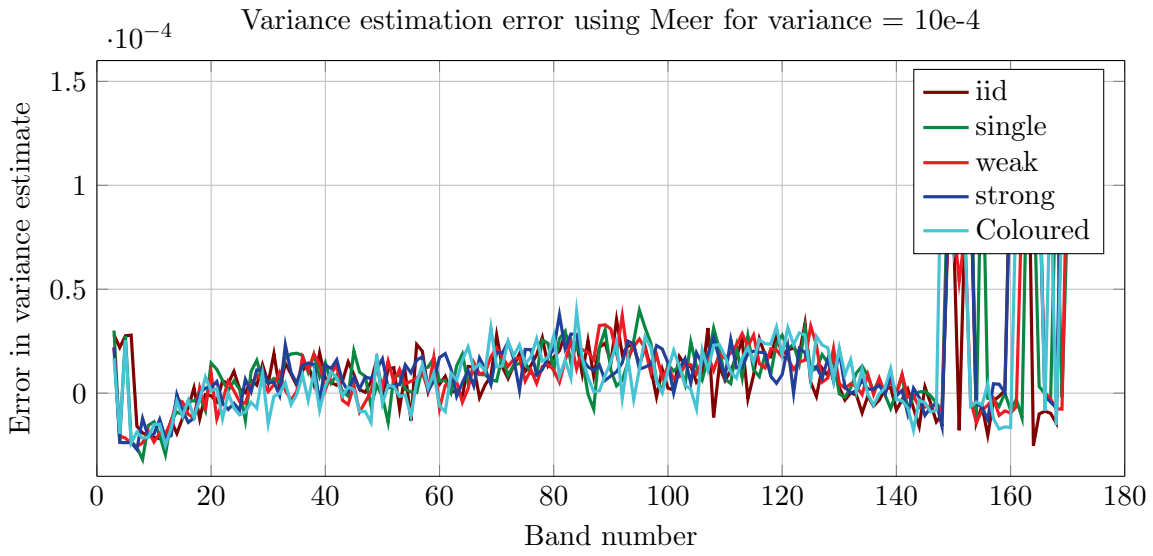


Figure 9: Meer's method applied to noise with variance 10^{-4}. The method is correct within an error of 20% of the actual noise level.

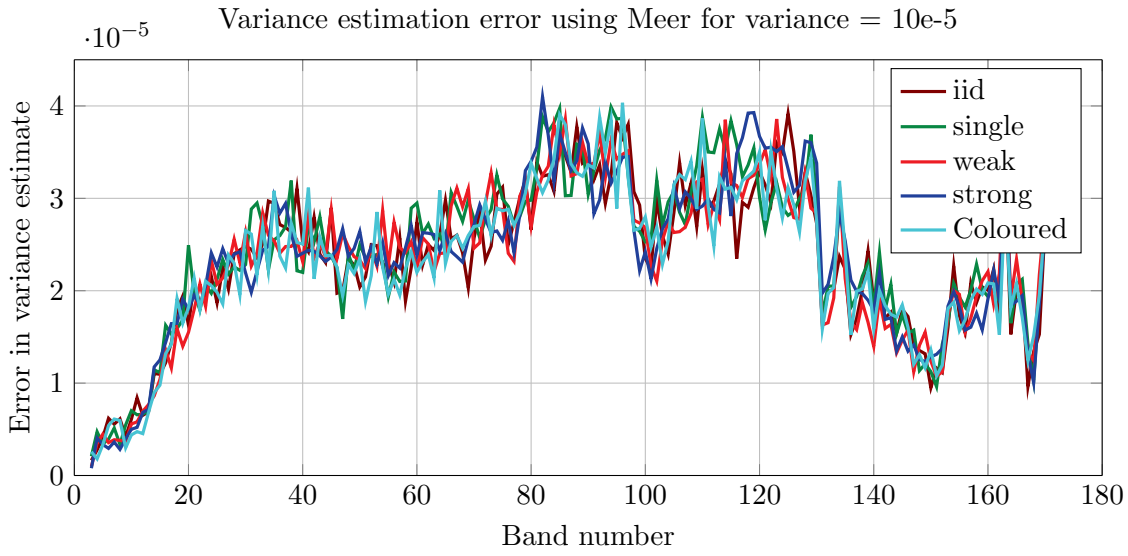


Figure 10: Meer's method applied to noise with variance 10^{-5}. The error in the estimation has become quite bad at this lower noise level. The method consistently overestimates the noise in the extremes by about 300%.

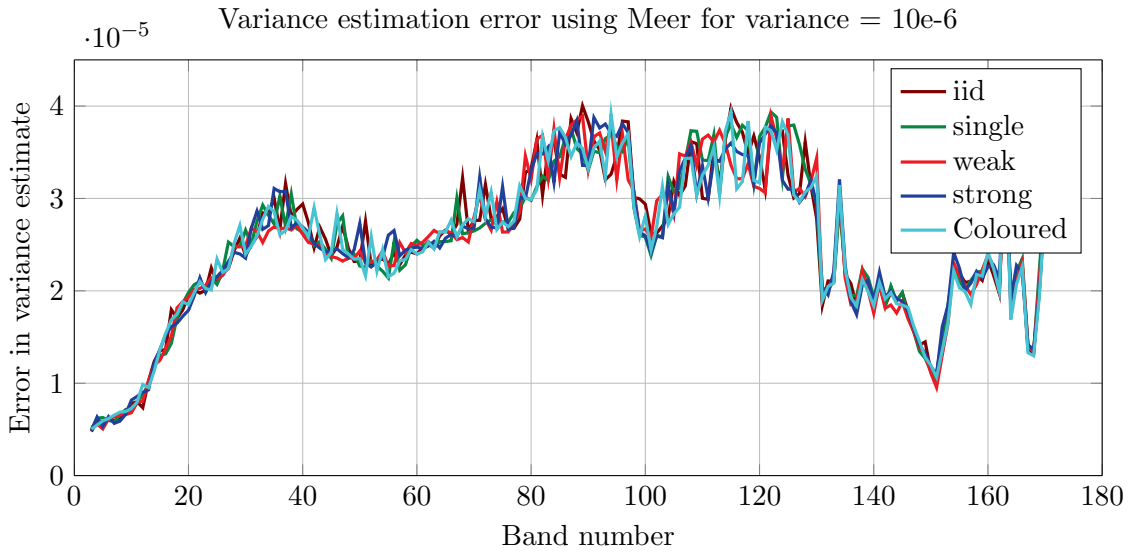


Figure 11: Meer's method applied to noise with variance 10^{-6}. The estimates of the variance are the same as those for the noise level 10^{-5} suggesting that Meer's method no longer sees the noise at all but is misreading signal variance as noise.

6.1.2 Roger's Method

Roger's method provides us with an estimate of the variance in each band. The method uses spectral information and, as such, we predicted that the correlated noise cases would produce a worse estimation than the i.i.d. and coloured noise cases.

We explored different block sizes for Roger's method and found that a block size of 8x8 was optimal. Larger block sizes led to overestimations of the noise. This is understandable as in the regression step more data points need to be explained simultaneously and the fact that the image is not smooth spatially, having borders and textures, means that they can be far apart from each other.

As we expected, the stronger correlation with the immediate neighbour in the single band correlation produced worse results than the weak three neighbour correlation in Roger's method as the method uses only 1 spectral neighbour in its regression (the values of the correlation coefficients are 0.5 vs 0.4 on the first diagonals). The results in the case of the strong three neighbour correlation are worse still (having a correlation coefficient of 0.65 on the immediate neighbour). Smaller block sizes took longer to process (more regression steps are required) and yielded worse results anyway.

Roger's method has demonstrated a dependancy on the correlation type as well as the noise level. The method performs best at high noise levels, however the error never exceeds 50% of the true noise variance at the lowest noise levels. Interestingly the error in the estimation in the weakly correlated case is smaller than the other noise types at the lowest noise levels. The trend has been that the method underestimates the noise variance in correlated noise and it would appear that the trend to overestimate noise of uncorrelated cases is counteracted by the error incurred under weak correlation. The method does give satisfactory results at all levels.

It should be noted that in modern hyperspectral images the noise is likely to be of the order $10^{-6} - 10^{-5}$ and be weakly correlated. Taking this into consideration our experiments suggest that Roger's method is a very dependable method for variance estimation in hyperspectral images since it has the smallest estimation error under these conditions.

The method takes of the order 40 seconds to to complete. The method is easily parallelisable and that has not been exploited in our code. The regression blocks could be treated simultaneously as well as all bands.

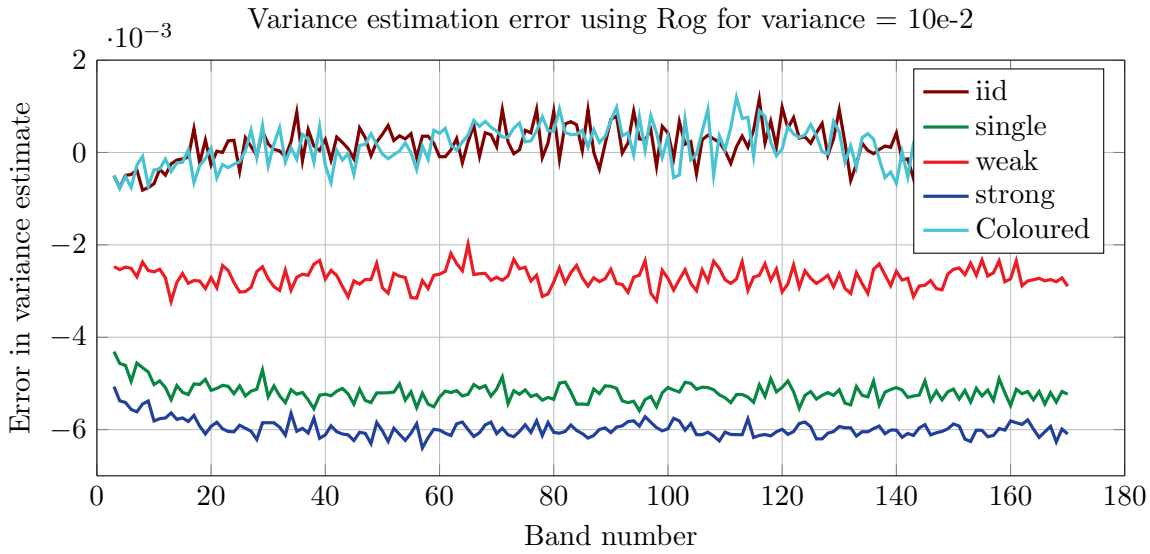


Figure 12: Roger's regression method applied to noise with variance 10^{-2}. We observe that the method estimates the noise correctly within an error of 5% in the case of i.i.d. and coloured noise. In the case of correlated noise the method underestimates the noise in all cases. The estimation of the noise variance in the case of the weakly correlated noise is off by 20%. In the case of the more strongly correlated noise types the estimation is off by 40-60%.

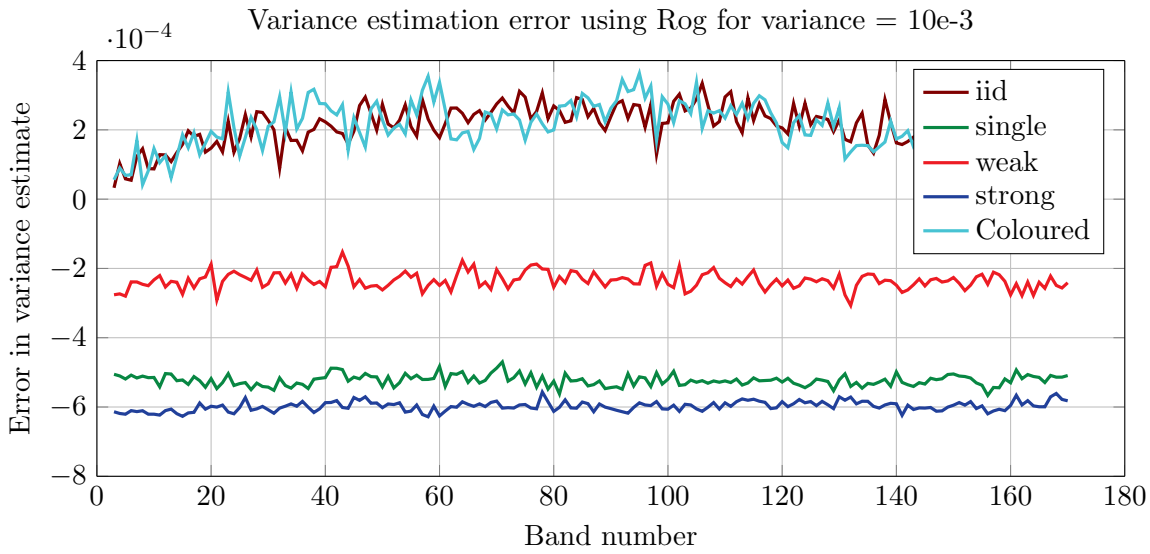


Figure 13: Roger's regression method applied to noise with variance 10^{-3}. The error in estimation of the noise in the uncorrelated cases is now out by 20% while the error in the correlated cases is more or less consistent.

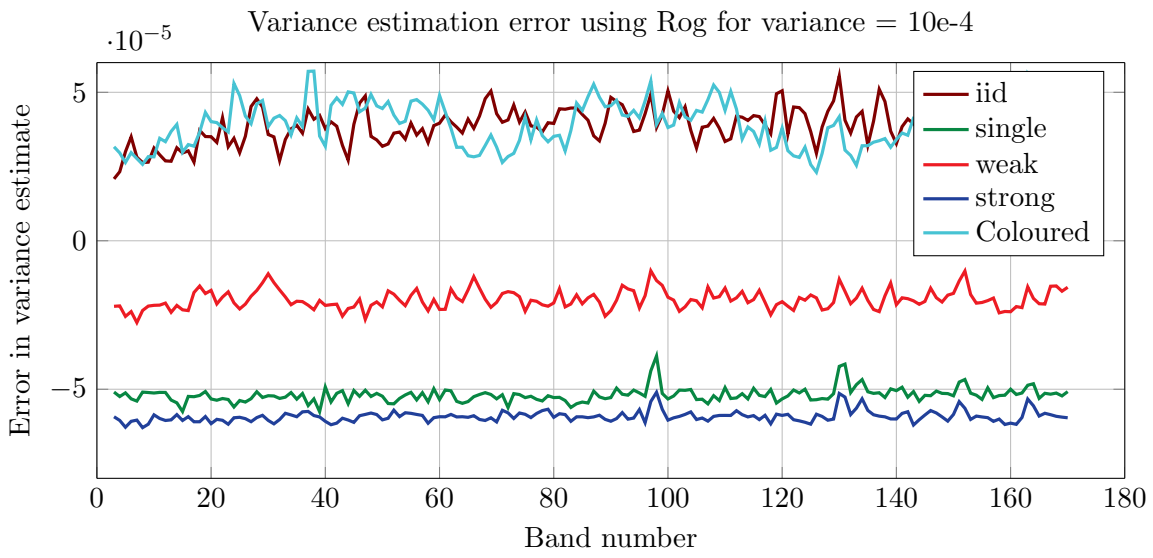


Figure 14: Roger's regression method applied to noise with variance 10^{-4}. The error in the estimate in the uncorrelated cases is again twice as large as the error seen at the higher noise level while the correlated cases remain in the same error range.

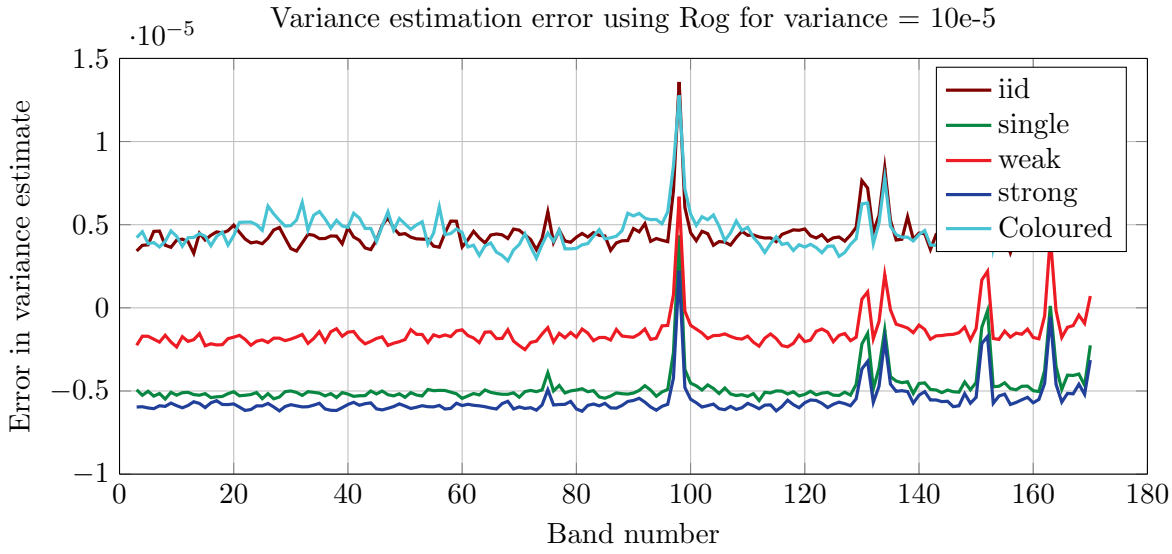


Figure 15: Roger's regression method applied to noise with variance 10^{-5}. The error in the uncorrelated cases has grown to 50% and again the error in the correlated cases is constant.

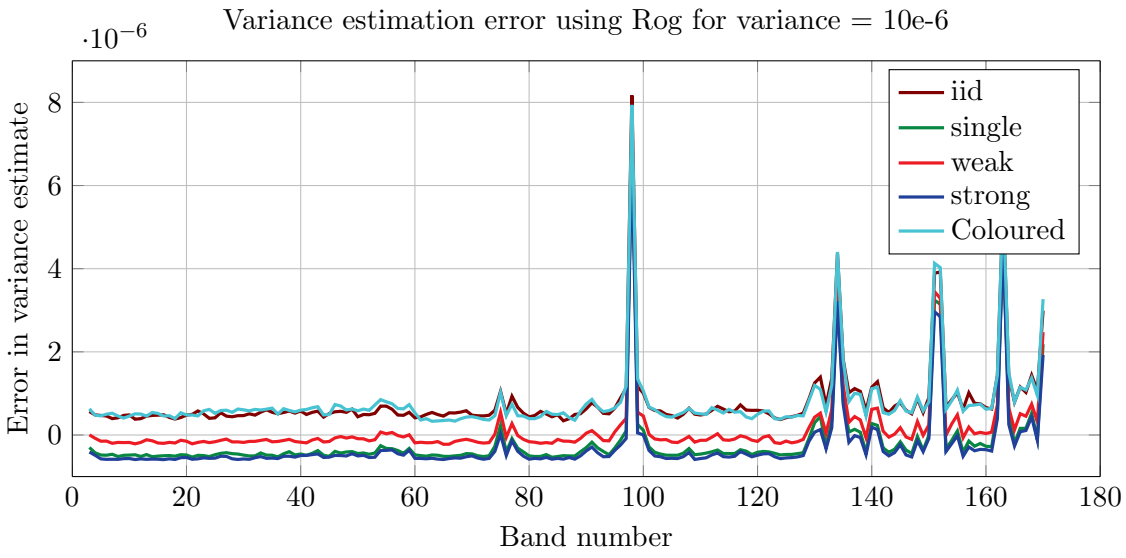


Figure 16: Roger's regression method applied to noise with variance 10^{-6}. The error in the i.i.d. case is now overestimated by 50%. The cases of correlated noise again remain in about the same error range.

6.1.3 Standard Regression

6.1.3.1 Variance Estimation The standard regression technique is an important method as it is completely blind. It only takes the image as an input to produce its estimates of both noise statistics and actual noise. The method has been used primarily as a statistic estimation tool. This method is based entirely on spectral information and as such suffers under correlated noise.

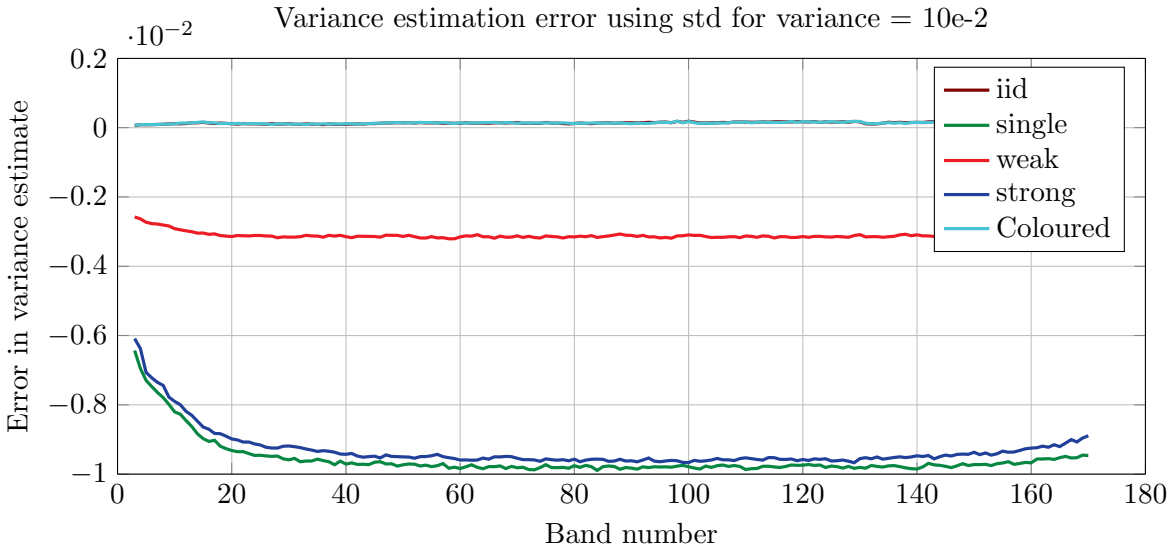


Figure 17: Standard regression method applied to noise with variance 10^{-2}. Notice how much the method estimate suffers under the presence of correlated noise.

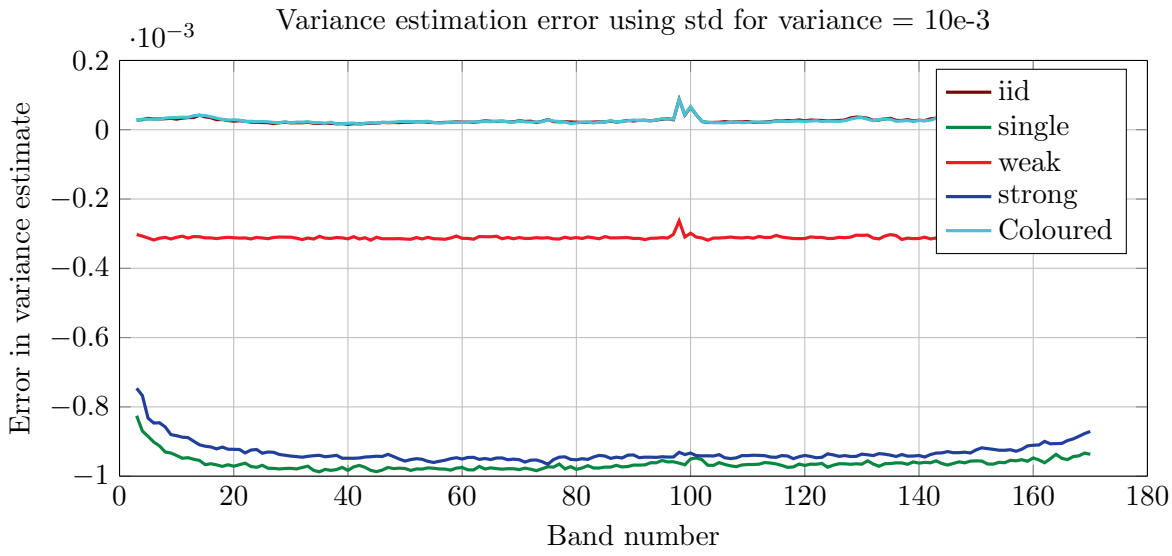


Figure 18: Standard regression method applied to noise with variance 10^{-3}. Notice how much the method estimate suffers under the presence of correlated noise.

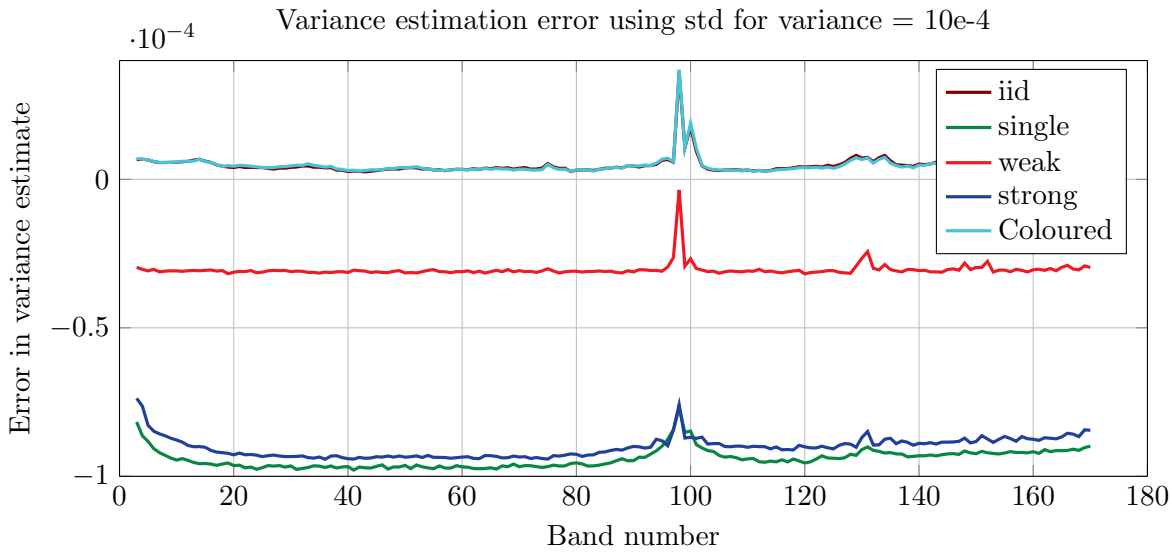


Figure 19: Standard regression method applied to noise with variance 10^{-4}. Notice how much the method estimate suffers under the presence of correlated noise.

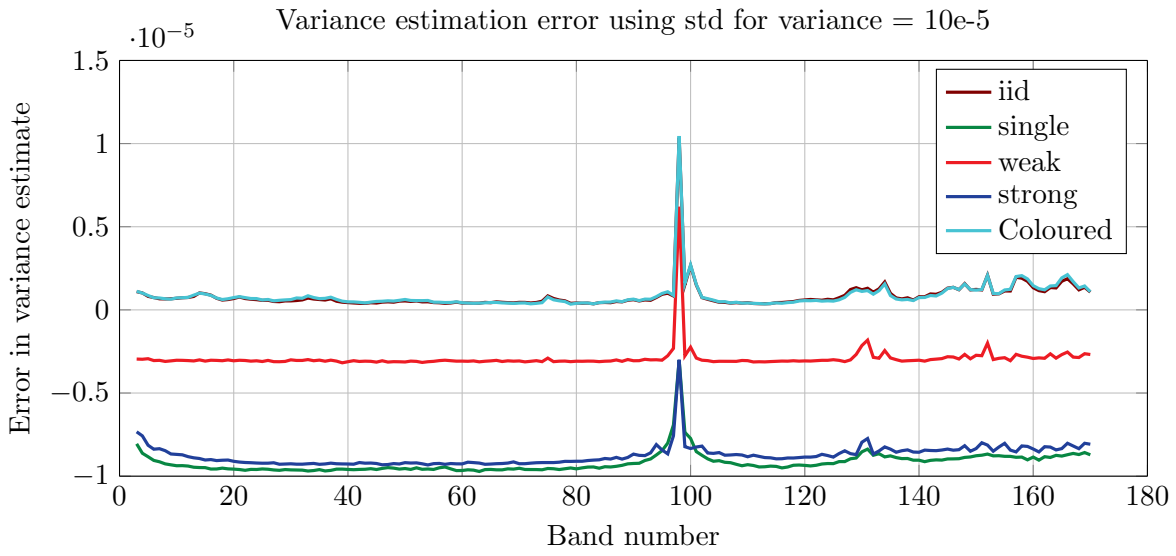


Figure 20: Standard regression method applied to noise with variance 10^{-5}. Notice how much the method estimate suffers under the presence of correlated noise.

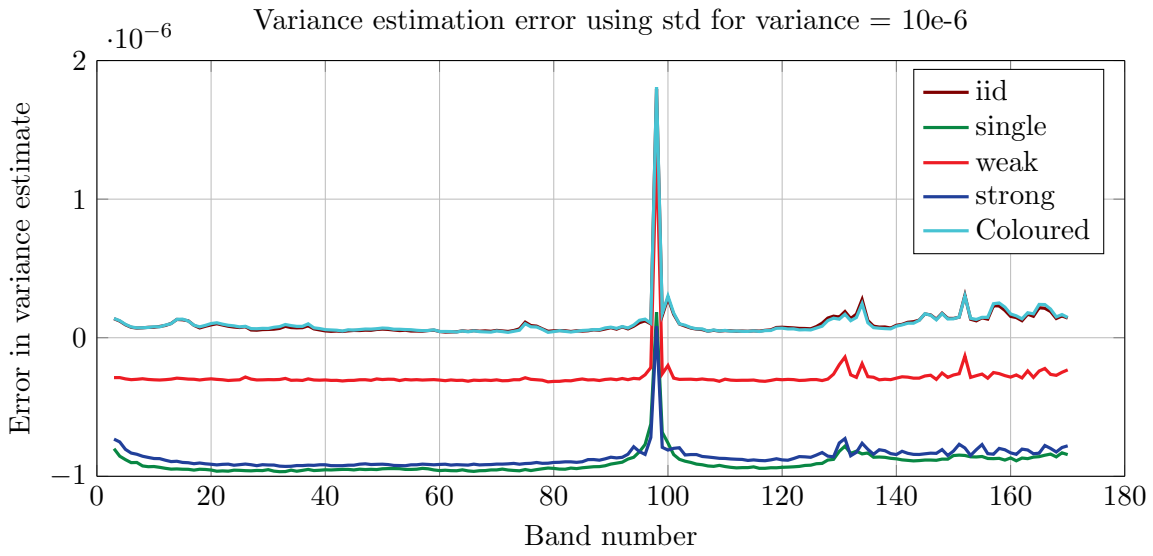


Figure 21: Standard regression method applied to noise with variance 10^{-6}. Notice how much the method estimate suffers under the presence of correlated noise.

The standard regression approach demonstrates a good deal of consistency in its results at all noise levels. It does perform slightly better at higher noise levels. The error in the presence of stronger correlation is significant. It is interesting to note that the large correlation coefficient in the immediate neighbouring bands in the single correlation case produces a greater error than the weakly correlated three neighbour noise and of the same order as the strongly correlated three neighbour noise.

The trend seen from figures 17 - 21 is obvious. The estimation error in the case of uncorrelated noise (the coloured and i.i.d.) is negligible and the method demonstrates no dependency on the colour of the noise, however, as soon as the noise is correlated the error jumps significantly. The method overestimates noise in the uncorrelated noise (indicated by a positive error) and underestimates correlated noise which is what we expect as correlated noise starts to look like signal to the method. The method error in the case of the weakly correlated noise is only of the order of 30% of the noise level which is acceptable. In the case of stronger correlation, however, the error jumps to 80-90% of the noise level. Interestingly the effect of single band correlation is more detrimental than the multi-neighbour correlations (even the strongly correlated noise which has higher off diagonal correlation coefficients).

6.1.3.2 Noise Removal As expected the method performs very well for uncorrelated noise due to the high redundancy in the image (indicating that the assumption that the image subspace is much smaller than the number of bands is justified) and the fact that noise in adjacent bands is hard to express in terms of the other bands. The method demonstrated no dependence on the colour of the noise. Interestingly the results for the weakly correlated noise were better than expected and even better than the case of single correlation.

We see again an obvious trend with good consistency in the performance at all noise levels. The method removes around 90% of the noise in the uncorrelated cases. As soon as noise is correlated it removes no better than 40% of the noise. Again we see the behaviour that the weakly correlated noise impairs the method less than the more correlated noise types and again the single neighbour correlation is the worst. Figures 22 - 26 show that the the method does perform slightly better at higher noise values.

The method took of the order of 350 seconds to complete. There is room for parallelisation as each band can be treated simultaneously which was not done in our code. The bonus to this method is that you get an estimate of the variance as well as an estimate of the actual noise simultaneously with no input or supervision.

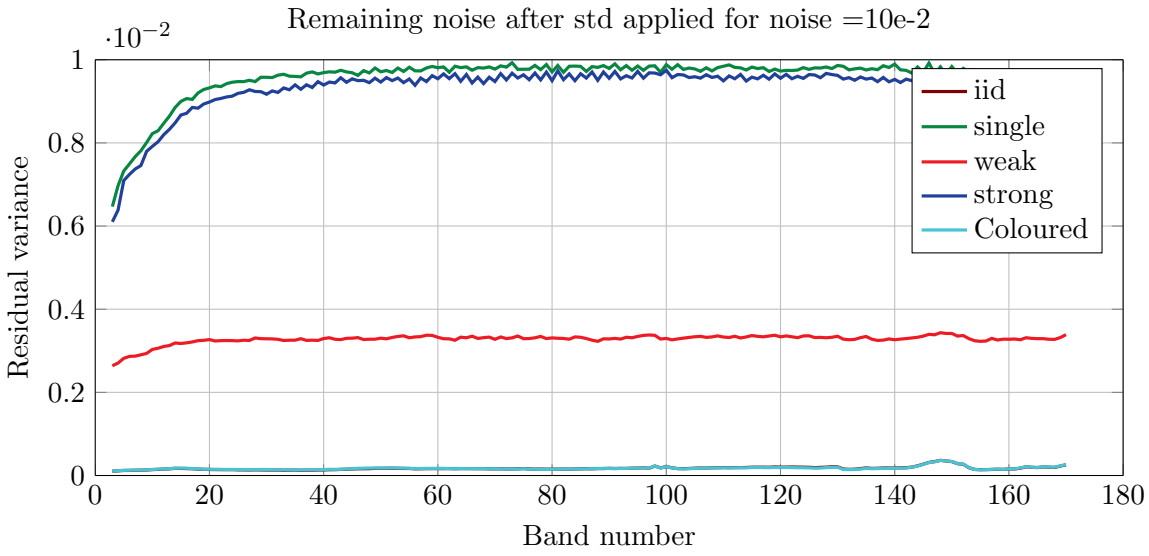


Figure 22: Standard regression method applied to noise with variance 10^{-2}. The method is extremely effective in the cases of uncorrelated noise and is adequate for weakly correlated noise.

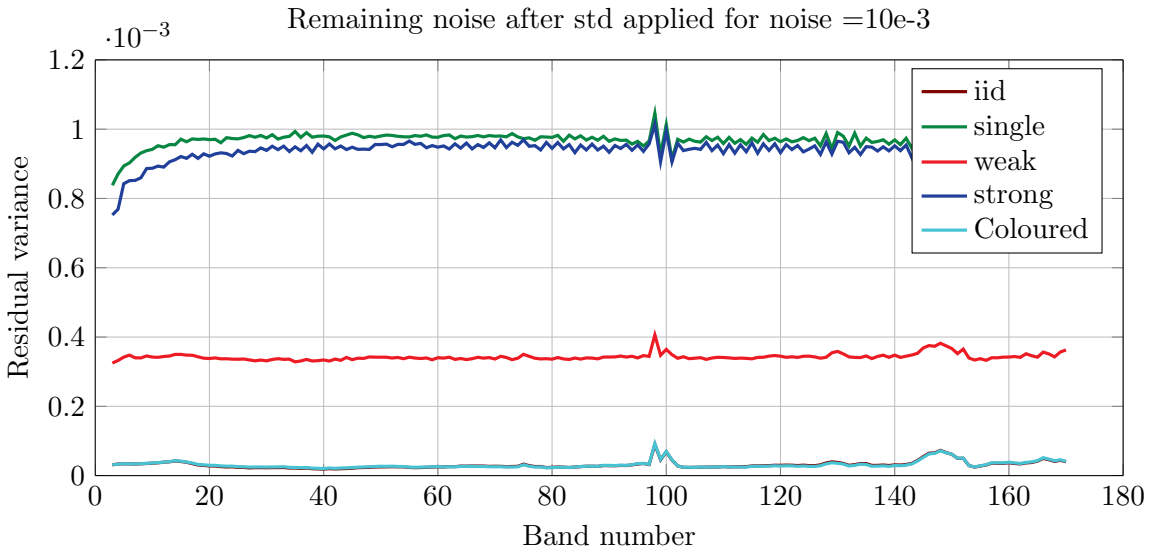


Figure 23: Standard regression method applied to noise with variance 10^{-3}. The method is extremely effective on the cases of uncorrelated noise and is adequate for weakly correlated noise. The cases of stronger correlation removed almost no noise.

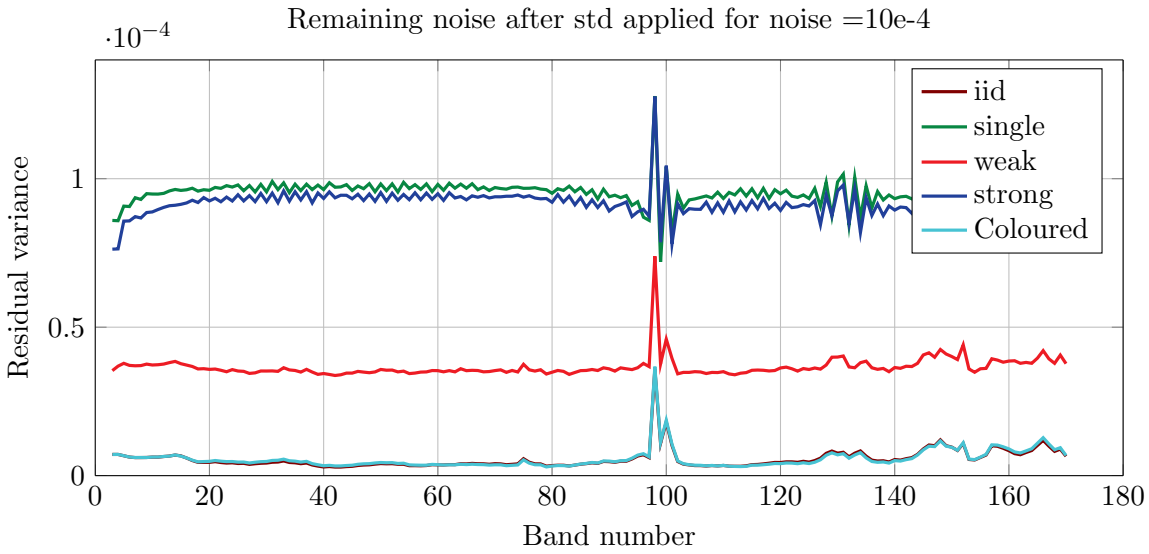


Figure 24: Standard regression method applied to noise with variance 10^{-4}. The method is extremely effective on the cases of uncorrelated noise and is adequate for weakly correlated noise. The cases of stronger correlation removed almost no noise.

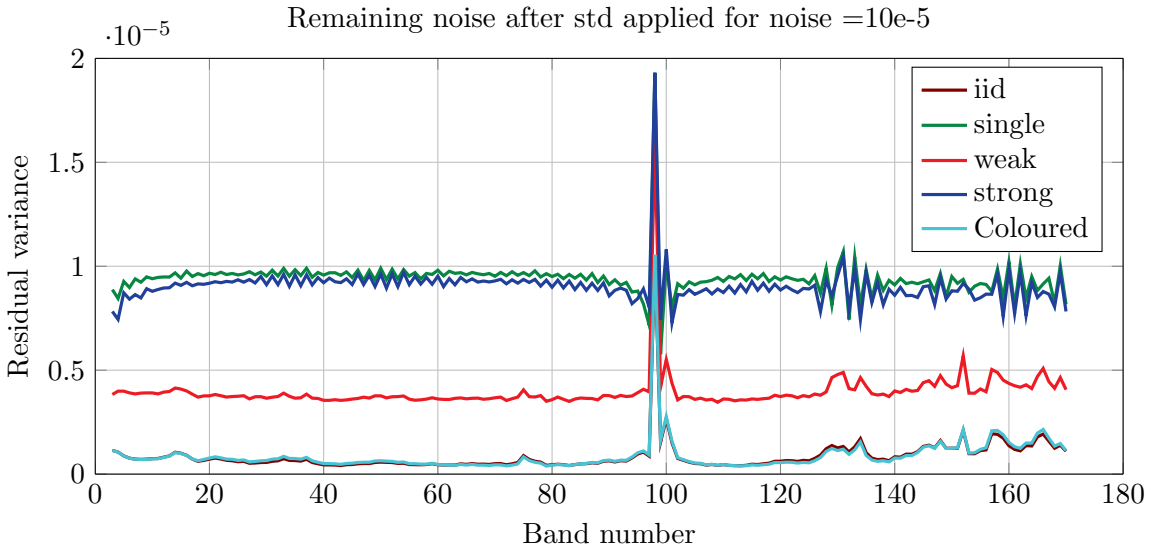


Figure 25: Standard regression method applied to noise with variance 10^{-5}. The method is extremely effective on the cases of uncorrelated noise removing over 80% of the noise and is adequate for weakly correlated noise removing about 60% of the noise. The cases of stronger correlation removed almost no noise.

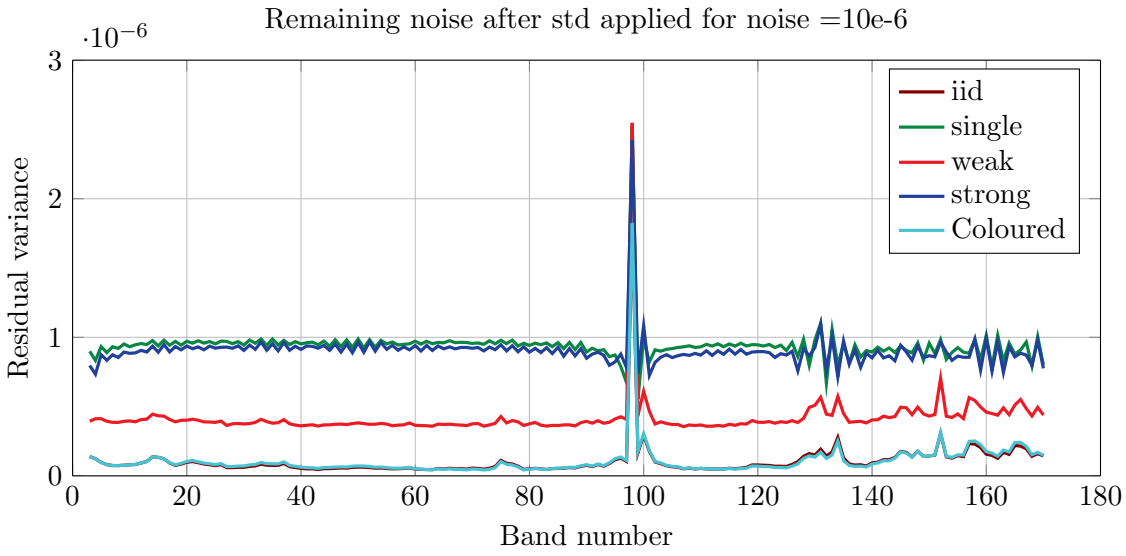


Figure 26: Standard regression method applied to noise with variance 10^{-6}. The method is extremely effective on the cases of uncorrelated noise removing over 80% of the noise and is adequate for weakly correlated noise removing about 60% of the noise. The cases of stronger correlation removed almost no noise.

6.1.4 Modified Regression Method

The modified regression method is in a sense not completely blind. It requires a guess as to the correlation structure present in the noise. As such we present the experiments for this method in 3 subsections. The purpose of these extended experiments is to see the effects of under and overestimation of the correlation matrix on the effectiveness of the method.

The execution times for the methods were very small (MR0 averaged about 0.5 seconds while MR5 was closer to 5 seconds). It should be noted that the standard regression does need to be performed first so the time from start to finish is also of the order of 350 seconds. The step in the method that takes the longest time is the inversion of the matrix **A** in equation (12).

6.1.4.1 MR1 In the MR1 case we assume that only one neighbour is correlated. That is we search for a tri-diagonal correlation matrix. This is expected to produce good estimates for the uncorrelated noise experiments as well as the single neighbour correlated noise. This is indeed the case as we see from figures 27 - 31. The method shows good consistency at all noise levels estimating the variance in the coloured and i.i.d. cases almost perfectly, as well as the single neighbour correlation. The MR1 implementation underestimates the weakly correlated noise by about 40% of the original variance and the strong correlation by about 90%. MR1 has levelled the results of single correlation to those for uncorrelated noise, a big improvement.

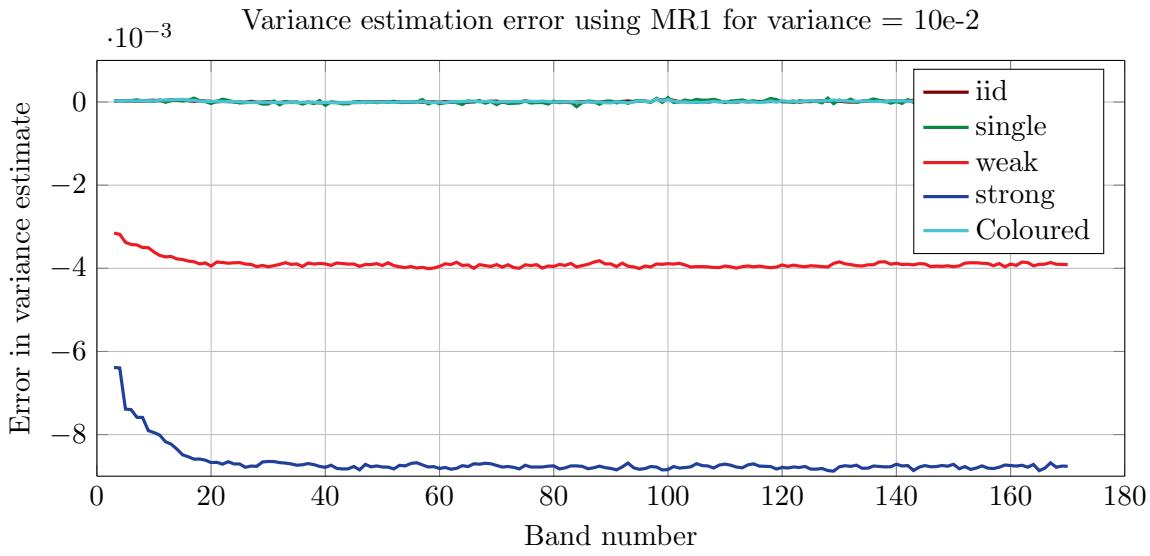


Figure 27: Modified regression method applied to noise with variance 10^{-2}. One neighbour correlation assumed.

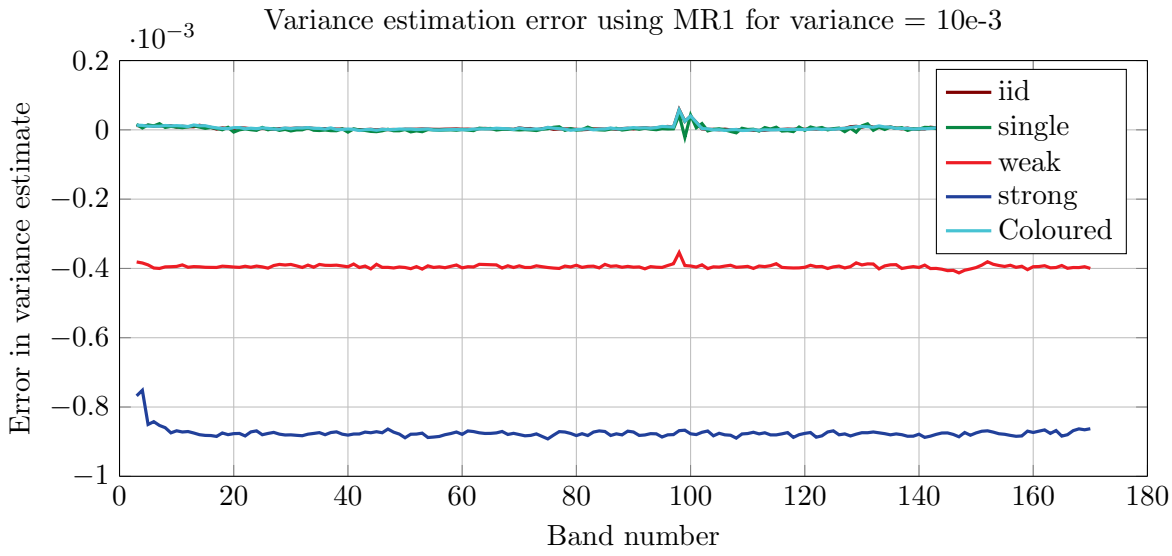


Figure 28: Modified regression method applied to noise with variance 10^{-3}. One neighbour correlation assumed.

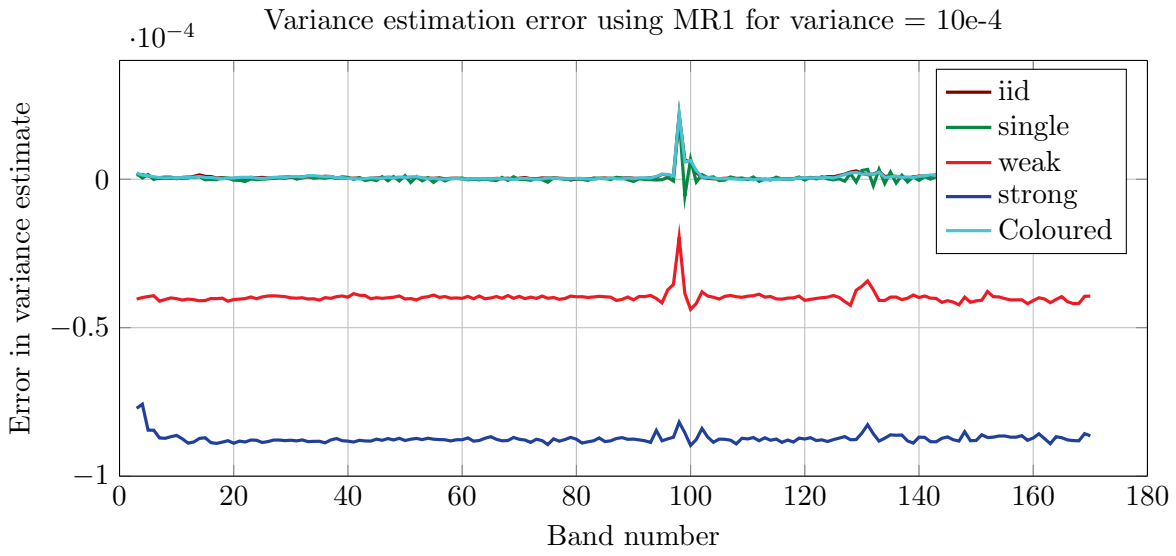


Figure 29: Modified regression method applied to noise with variance 10^{-4}. One neighbour correlation assumed.

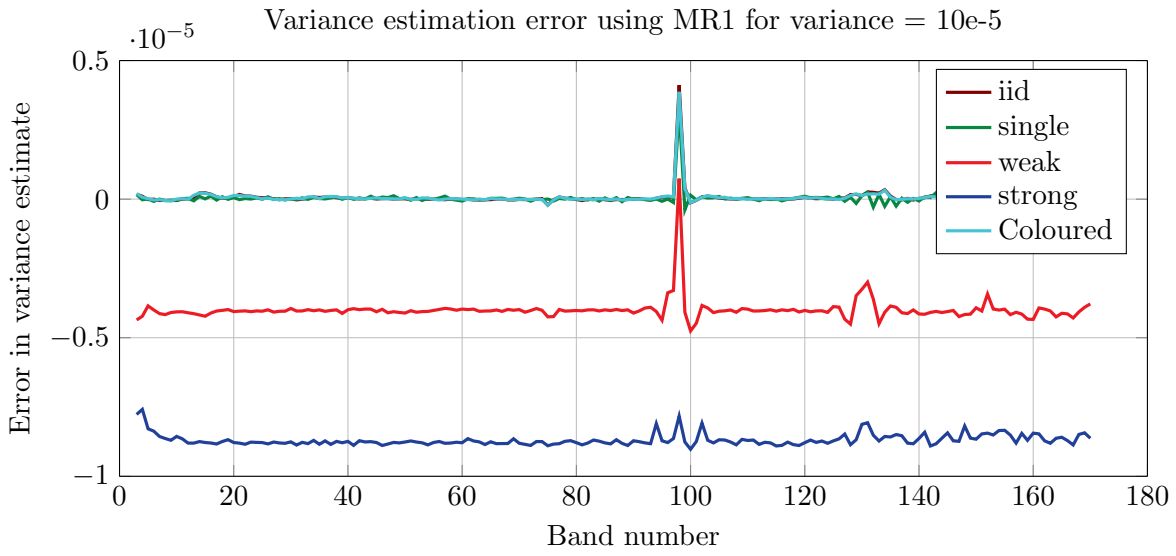


Figure 30: Modified regression method applied to noise with variance 10^{-5}. One neighbour correlation assumed.

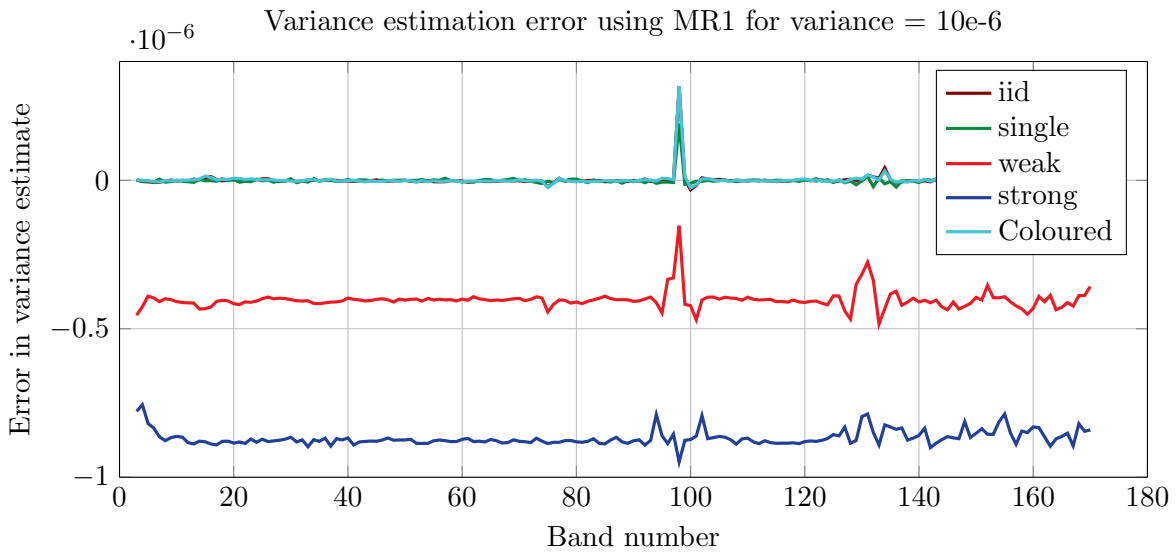


Figure 31: Modified regression method applied to noise with variance 10^{-6}. One neighbour correlation assumed.

6.1.4.2 MR3 In the MR3 case we assume that three neighbouring bands in each direction are correlated. That is we search for a seven-diagonal correlation matrix. This is expected to produce good estimates for the uncorrelated noise experiments as well as all the correlated cases as they are all seven-diagonal or less correlation matrices. This is indeed the case as we see from figures 32 - 36. The method with the seven-diagonal assumption has levelled all the noise levels to essentially the same error. The error at all noise levels is two orders of magnitude less than the original noise (around the 2-5% mark). This shows how sensitive the method is to a correct assumption of the correlation structure being such a great improvement on the MR1 for the three neighbour correlations (strong and weak).

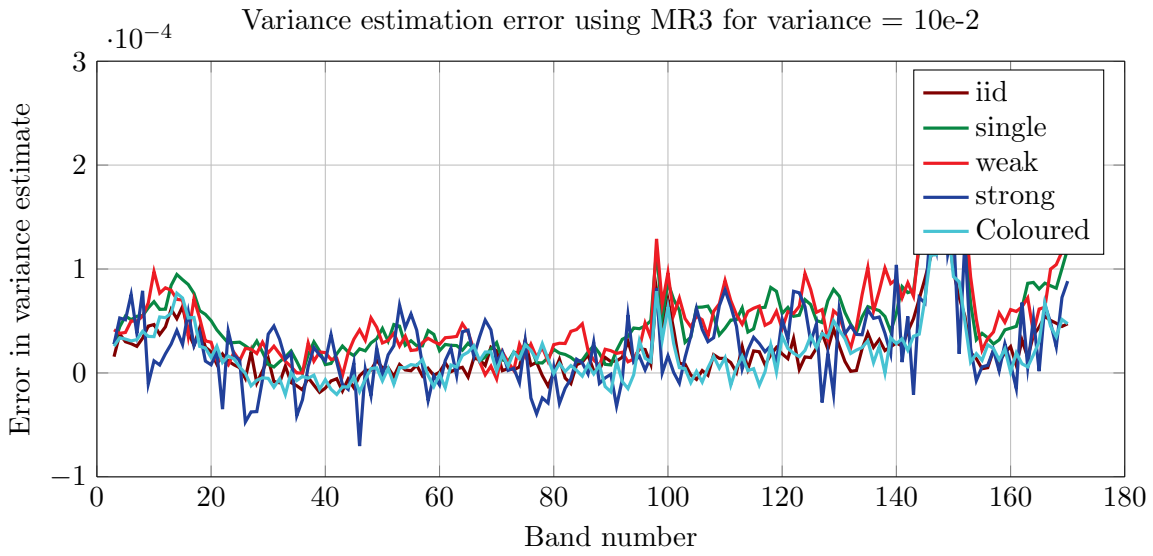


Figure 32: Modified regression method applied to noise with variance 10^{-2}. Three neighbour correlation assumed.

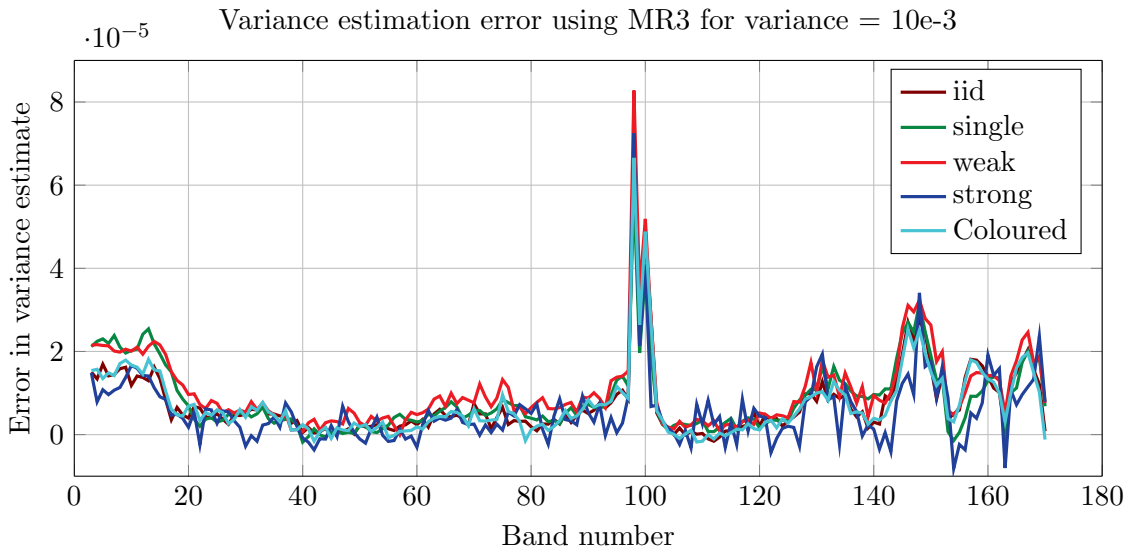


Figure 33: Modified regression method applied to noise with variance 10^{-3}. Three neighbour correlation assumed.

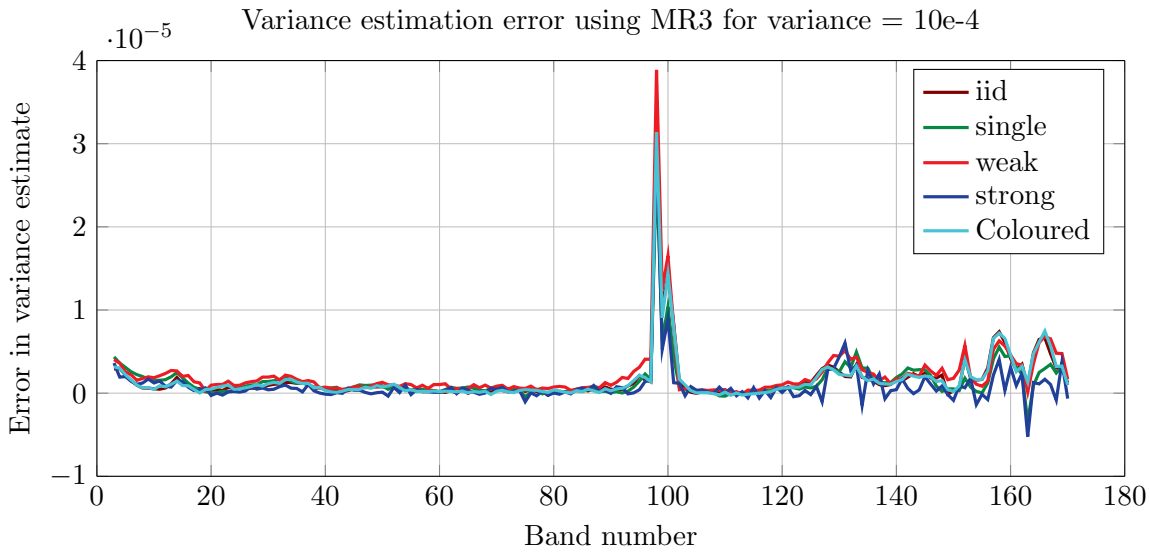


Figure 34: Modified regression method applied to noise with variance 10^{-4}. Three neighbour correlation assumed.

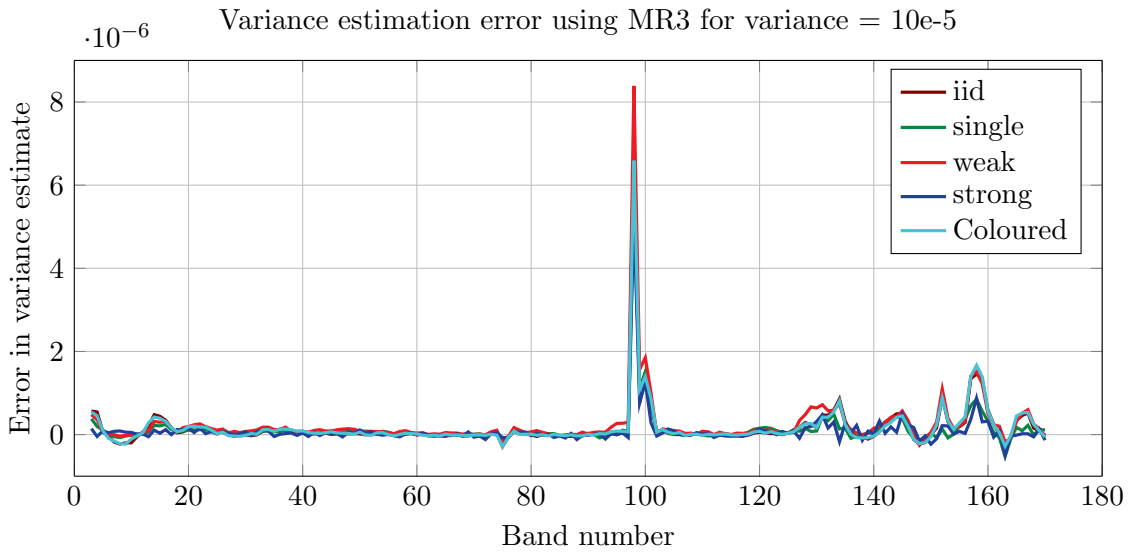


Figure 35: Modified regression method applied to noise with variance 10^{-5}. Three neighbour correlation assumed.

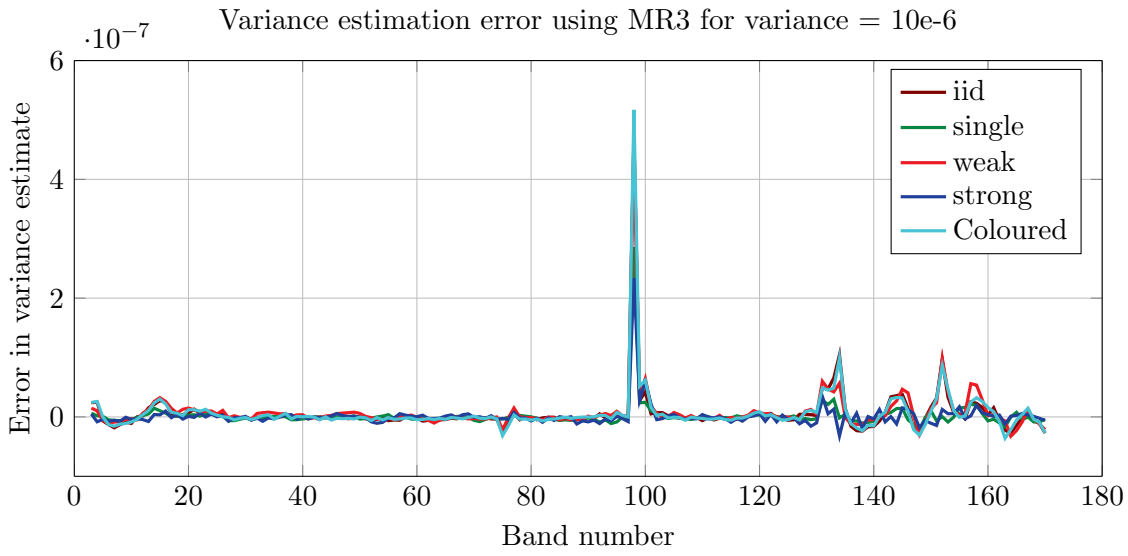


Figure 36: Modified regression method applied to noise with variance 10^{-6}. Three neighbour correlation assumed.

6.1.4.3 MR5 In the MR5 case we assume that five neighbouring bands in each direction are correlated. That is we search for a eleven-diagonal correlation matrix. This is expected to produce good estimates for the uncorrelated noise experiments as well as the all the correlated cases as they are all seven-diagonal or less correlation matrices. This is indeed the case as we see from figures 37 - 41. The method with the eleven-diagonal assumption has levelled all the noise levels to essentially the same error. The error at all noise levels is two orders of magnitude less than the original noise (almost non-existent). This shows how an overestimation of the correlation structure is less erroneous than an underestimation.

It should be noted that we experimented with correlation assumptions all the way up to a 101-diagonal matrix. The results indicate that the method does break down at greater overestimations allowing method errors to compound themselves. This breakdown occured at the lower noise levels and more for the correlated cases than the uncorrelated cases (which remained within the 5-10% range). Even so, the error was within 60-80%. The indication is clear: a severe overestimate of the correlation does little harm to the estimate while a small underestimate does a lot of harm to the estimate.

In addition we ran the experiment on a correlation matrix that was roughly 81-diagonal (a 81-diagonal matrix was used as a seed matrix and corrected using the rank 1 update method described earlier). The results were again consistent an overestimate produced better results than an underestimate and the method was stable even at such high correlations.

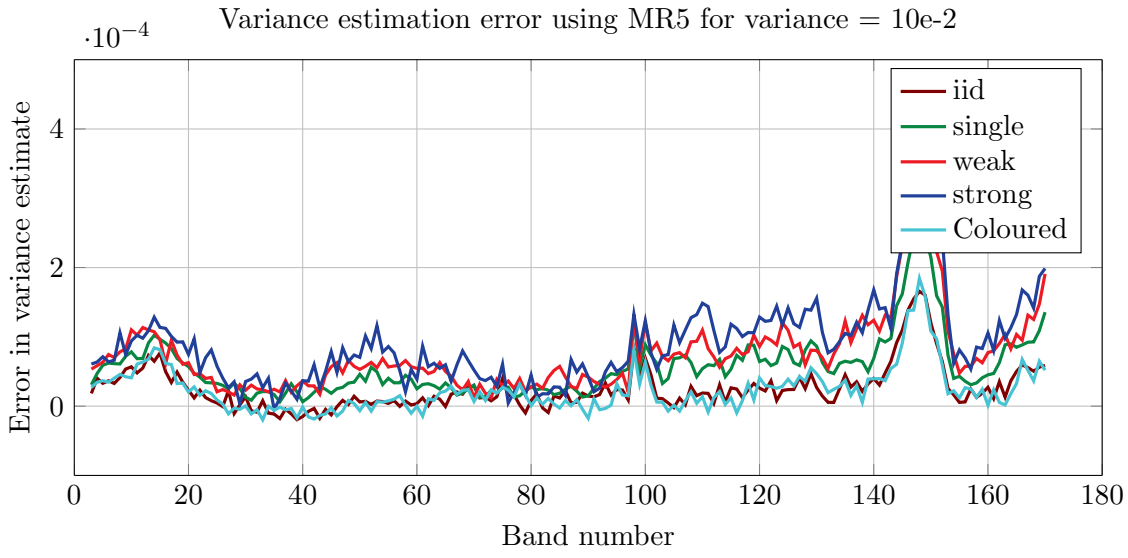


Figure 37: Modified regression method applied to noise with variance 10^{-2}. Five neighbour correlation assumed.

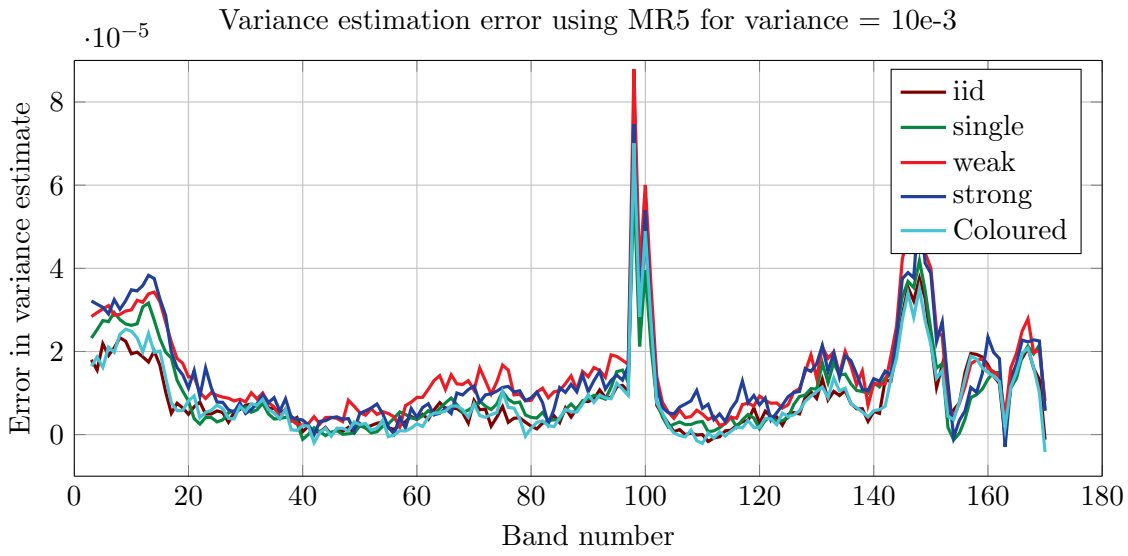


Figure 38: Modified regression method applied to noise with variance 10^{-3}. Five neighbour correlation assumed.

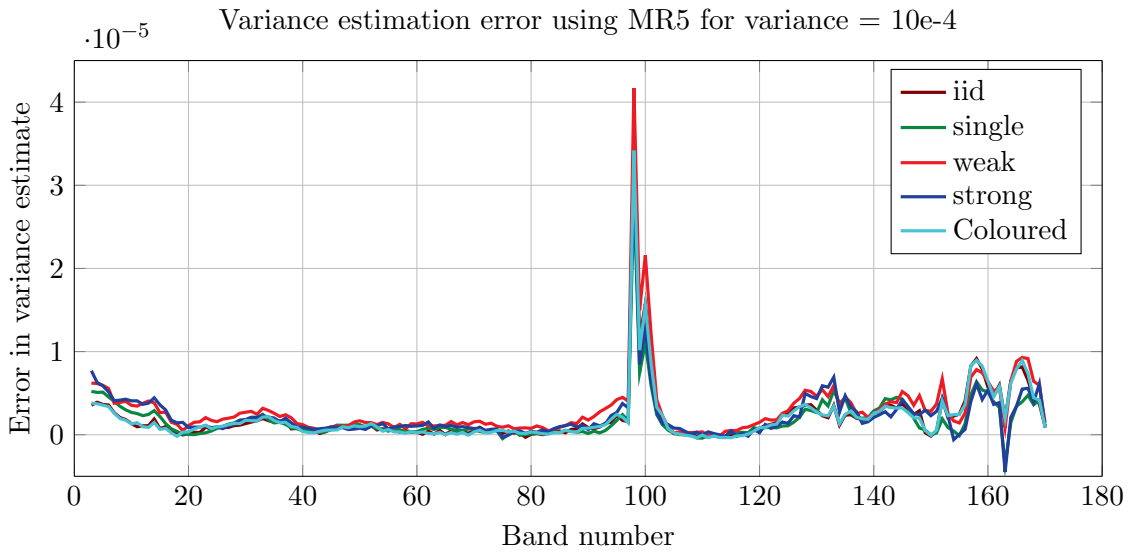


Figure 39: Modified regression method applied to noise with variance 10^{-4}. Five neighbour correlation assumed.

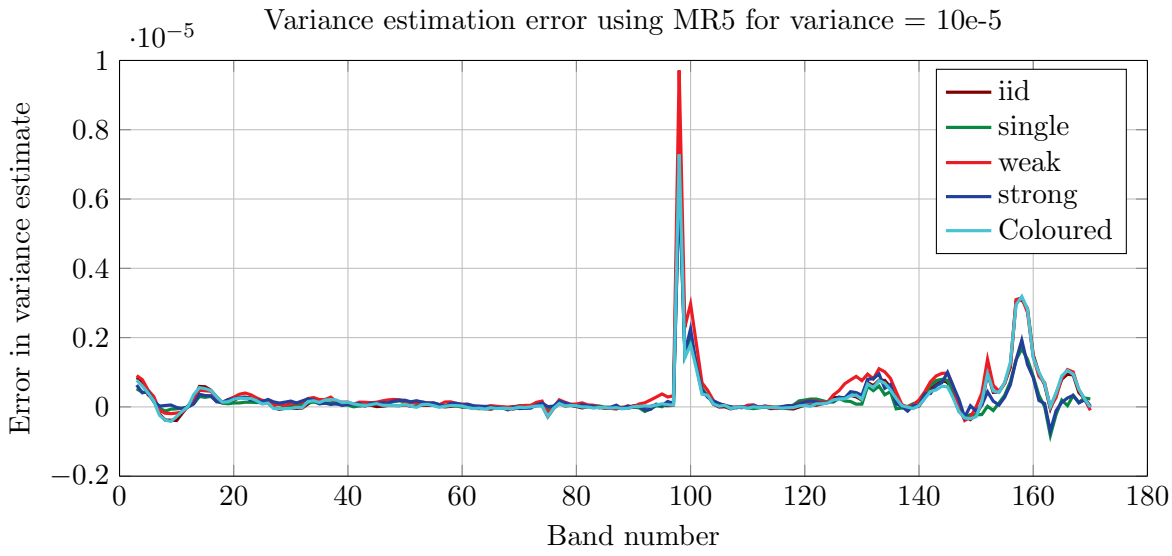


Figure 40: Modified regression method applied to noise with variance 10^{-5}. Five neighbour correlation assumed.

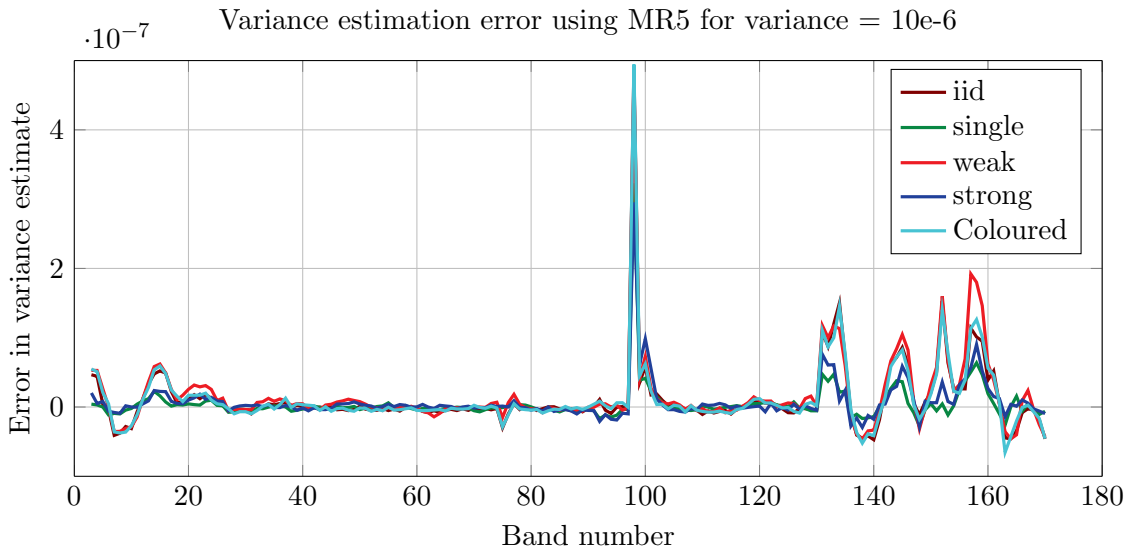


Figure 41: Modified regression method applied to noise with variance 10^{-6}. Five neighbour correlation assumed.

6.1.4.4 Correlation estimate To analyse the effectiveness of the modified regression method to determine the structure of the correlation matrix we make use of the following measures as defined in [19]:

The mean absolute error for the jth diagonal is given by

$$\text{MAE} = \frac{1}{K_j}\sum_{i=1}^{K_j}|\hat{\mathbf{r}}_i - \mathbf{r}_i| \tag{13}$$

where $\mathbf{r}$ is the diagonal of the correlation matrix under consideration and $\hat{\mathbf{r}}$ is its approximation produced by the method. K_j is the number of elements in the diagonal under consideration.

The standard deviation of the absolute error given by

$$\text{SDAE} = \left(\frac{1}{K_j}\sum_{i=1}^{K_j}\left(|\hat{\mathbf{r}}_i - \mathbf{r}_i| - \text{MAE}\right)^2\right)^{\frac{1}{2}} \tag{14}$$

The entries for the second through fifth column in the MR1 method are omitted since the method does not produce estimates for those diagonals (as such the MAE would be the mean correlation coefficient in that diagonal and the SDAE would be 0). The same is true for the MR3 method for the fourth and fifth diagonal.

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.0112	0.0242	-	-	-	-	-	-	-	-
MR3	0.0196	0.0354	0.0157	0.0233	0.0134	0.0203	-	-	-	-
MR5	0.0304	0.0519	0.0241	0.0348	0.0222	0.0337	0.0180	0.0271	0.0139	0.0274

Table 2: Coloured noise with variance 10^{-5}. Three MR methods comparing the strength of the correlation estimate.

The results from tables 2 and 3 show that the error in estimating the correlation structure in the case of uncorrelated noise increases as we increase the number of diagonals that we consider. The error is, however, very small of the order 10^{-2}. The trend also seems that the error grows as the noise level increases.

The results from tables 4 and 5 show how powerful the method is when we input the correct

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.0061	0.0173	-	-	-	-	-	-	-	-
MR3	0.0121	0.0229	0.0108	0.0216	0.0084	0.015	-	-	-	-
MR5	0.0228	0.0382	0.0181	0.0263	0.0161	0.025	0.0127	0.0242	0.0089	0.0244

Table 3: Coloured noise with variance 10^{-6}. Three MR methods comparing the strength of the correlation estimate.

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.0050	0.0218	-	-	-	-	-	-	-	-
MR3	0.0090	0.0250	0.0121	0.0231	0.0057	0.0097	-	-	-	-
MR5	0.0118	0.0200	0.0199	0.0286	0.0171	0.0227	0.0121	0.0150	0.0065	0.0133

Table 4: Single neighbour correlation with variance 10^{-5}. Three MR methods comparing the strength of the correlation estimate.

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.0037	0.0223	-	-	-	-	-	-	-	-
MR3	0.0048	0.0282	0.0054	0.0138	0.0031	0.0071	-	-	-	-
MR5	0.0061	0.0244	0.0083	0.0173	0.0068	0.0129	0.0051	0.0071	0.0029	0.0032

Table 5: Single neighbour correlation with variance 10^{-6}. Three MR methods comparing the strength of the correlation estimate.

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.2804	0.0316	-	-	-	-	-	-	-	-
MR3	0.0154	0.0396	0.0127	0.0368	0.0084	0.0168	-	-	-	-
MR5	0.0239	0.0394	0.0234	0.0343	0.0226	0.0295	0.0177	0.0250	0.0110	0.0259

Table 6: Weakly correlated noise with variance 10^{-5}. Three MR methods comparing the strength of the correlation estimate.

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.2917	0.0890	-	-	-	-	-	-	-	-
MR3	0.0114	0.0506	0.0090	0.0385	0.0052	0.0105	-	-	-	-
MR5	0.0157	0.0295	0.0161	0.0347	0.0144	0.0214	0.0112	0.0220	0.0072	0.0193

Table 7: Weakly correlated noise with variance 10^{-6}. Three MR methods comparing the strength of the correlation estimate.

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.1900	0.0639	-	-	-	-	-	-	-	-
MR3	0.0083	0.0419	0.0084	0.0387	0.0051	0.0168	-	-	-	-
MR5	0.0113	0.0293	0.0124	0.0312	0.0090	0.0170	0.0089	0.0157	0.0053	0.0177

Table 8: Strongly correlated noise with variance 10^{-5}. Three MR methods comparing the strength of the correlation estimate.

Diagonal	1st		2nd		3rd		4th		5th	
	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE	MAE	SDAE
MR1	0.1991	0.0780	-	-	-	-	-	-	-	-
MR3	0.0057	0.0406	0.0048	0.0250	0.0036	0.0108	-	-	-	-
MR5	0.0069	0.0306	0.0078	0.0262	0.0061	0.0147	0.0057	0.0105	0.0037	0.0044

Table 9: Strongly correlated noise with variance 10^{-6}. Three MR methods comparing the strength of the correlation estimate.

structure of the correlation matrix. The error in MR1 is only 2% of the correct correlation value. This error margin is very encouraging. However we do see that as we begin to overestimate the correlation structure (MR3 and MR5) the method error begins to grow for the estimation of the first diagonal. This error is still small, the greatest error is in MR5 at noise level 10^{-5} and that is only 6% of the true value.

The errors from the multi neighbour correlation noise types (Tables 6 - 9) show a very clear trend:

- An underestimation of the correlation structure produces a greater error than an overestimation of the correlation structure. This is apparent in the error from the greater error in MR1 in all cases than MR5.
- As the noise level increases the error becomes greater. The error is greater at noise level 10^{-5} than 10^{-6} in both correlation types.
- The greater the correlation coefficients the smaller the error in their estimation by the method. Clear from the difference from Tables 6 and 7 to Tables 8 and 9.

The strategy for noise estimation that these results suggest is clear. We should make a conservative guess of the correlation structure and rather (potentially) over estimate than underestimate. We will have to accept that the method error does get compounded the greater the overestimate, but it is clearly better than underestimating the structure.

Further experimentation with this method is warranted. We suggest testing the method with weaker correlation structures and also different types of correlation. Block correlation matrices are a type worth considering. The method should be updated to take block correlation matrices as initial guesses as well to decrease the potential crossover of over- and underestimation (some bands in block correlation matrices are correlated with a large number of other bands whilst others will only be correlated with a few other bands).

6.1.5 BM3D Bandwise

As this method uses no spectral information in its implementation we anticipated that it would not be adversely effected by correlation present in the noise. The method as presented by the authors for greyscale images showed a high aptitude for noise removal at high noise values. This was our observation in our experiments as well. The method did demonstrate some sensitivity to noise colour.

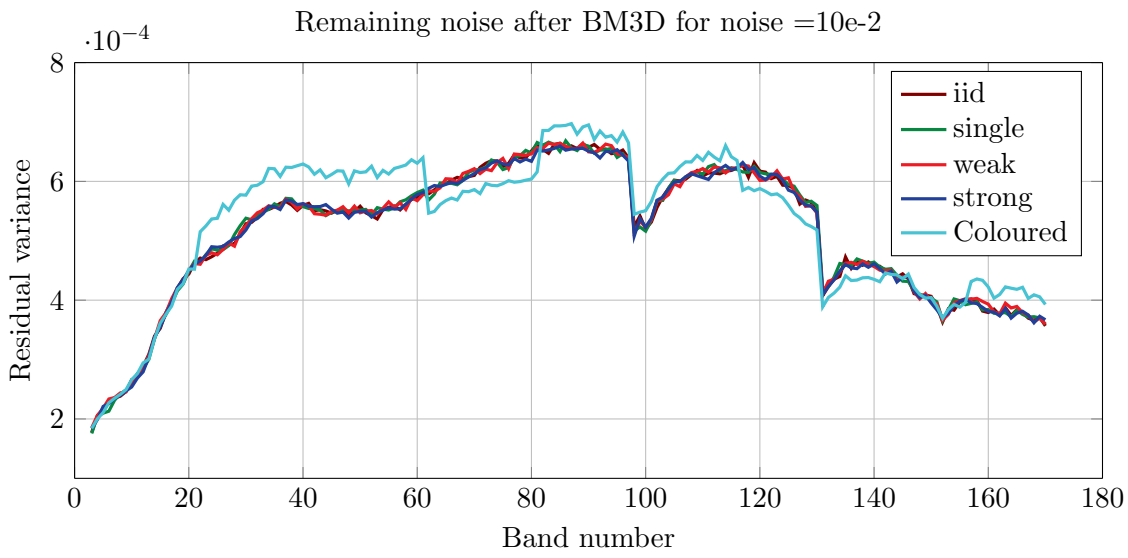


Figure 42: BM3D method applied to noise with variance 10^{-2}. We note that the error in estimation is independent on the correlation of the noise as expected. The residual noise is two orders of magnitudes less than the original noise.

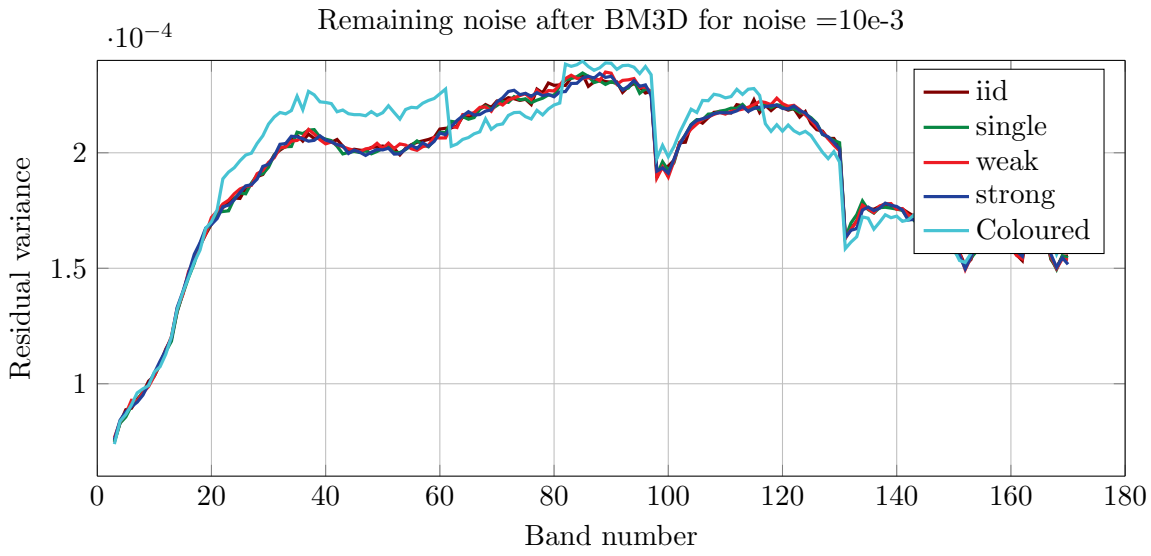


Figure 43: BM3D method applied to noise with variance 10^{-3}. The residual noise is an order of magnitude less with the method attenuating around 80% of the noise.

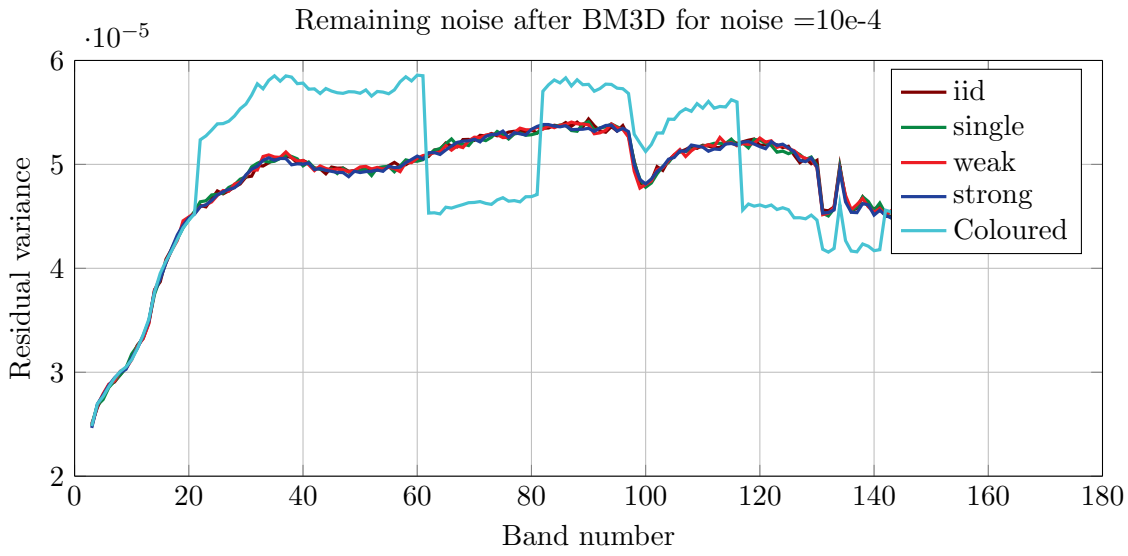


Figure 44: BM3D method applied to noise with variance 10^{-4}. The method has attenuated around least 50% of the noise.

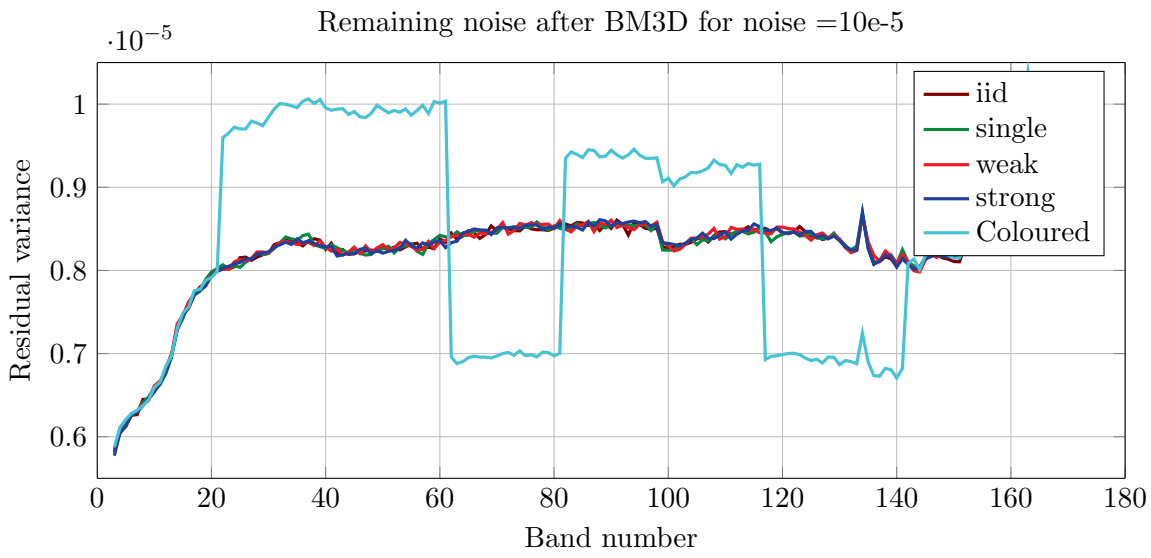


Figure 45: BM3D method applied to noise with variance 10^{-5}. The performance of the method has fallen significantly attenuating around 15% of the noise.

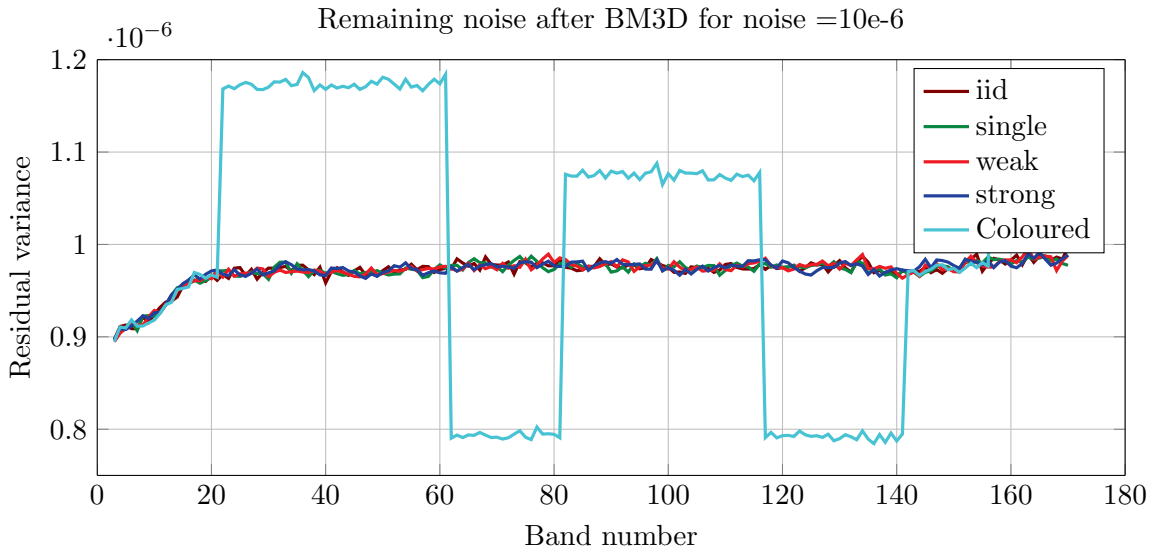


Figure 46: BM3D method applied to noise with variance 10^{-6}. The performance of the algorithm is now poor. The method has attenuated less than 10% of the noise and only about 6% on average.

The bandwise BM3D method has proven itself to be effective at noise levels greater than 10^{-5} for all noise types. It is interesting that the profile of the errors (the error through all the bands) has a consistent shape demonstrating some dependency on the underlying image (not all bands are equally bright and regions may appear homogeneous in one band but not in others corresponding to pixels with similar spectra). The noise remaining in the coloured noise case stands out more and more from the others, particularly visible in figure 46. We believe this distinction does not point to instability of the method (sensitivity to variance estimate) but is rather just because there is more or less noise in those particular bands.
The bBM3D algorithm takes of the order 950 seconds to complete. It should be noted that the method has possibility for parallelisation that has not been exploited in our experiments. This could be done by processing bands simultaneously or by grouping and processing blocks simultaneously.

6.1.6 PCA-BM3D

The PCA-BM3D method was expected to show sensitivity to noise correlation and this was indeed found to be the case. In the uncorrelated cases the initial principal component estimate produces a much clearer initial estimate so the block matching with Wiener filtering works well, producing a better final estimate. Like the bBM3D method the PCA-BM3D method demonstrates much better performance at higher noise levels - obvious in figure 47 where the residual noise is around 20% of the original noise in the case of correlated noise and about 5% in the uncorrelated cases. This residual noise ends up being about 80% at the lowest noise levels in figure 51. The residual noise for all noise types ends up being almost the same at the lowest noise levels.

The method took of the order 700 seconds to complete. It should be noted that the method has room for parallelisation in the block matching with Wiener filtering step.

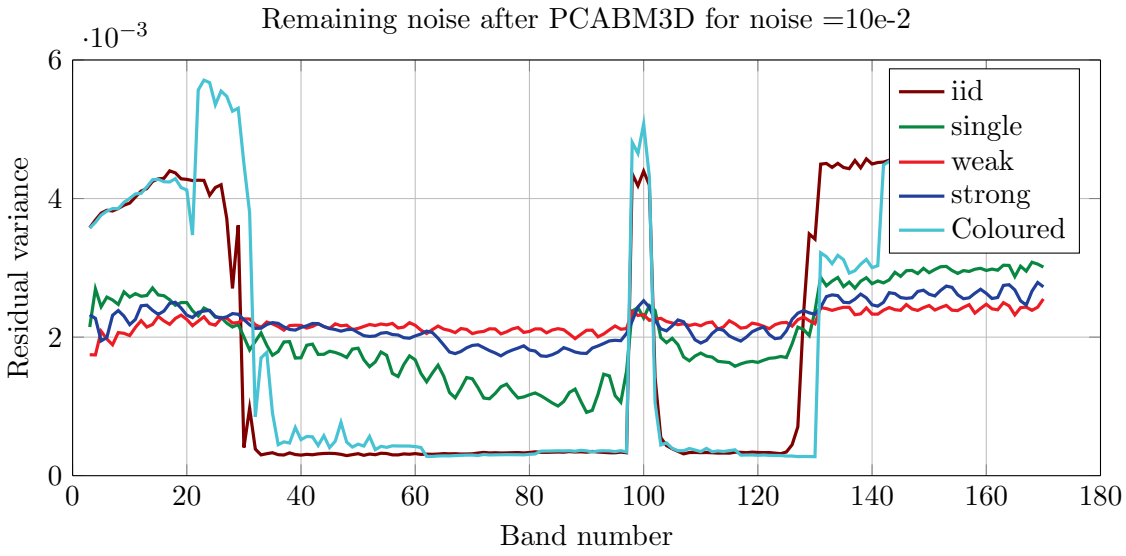


Figure 47: PCABM3D method applied to noise with variance 10^{-2}.

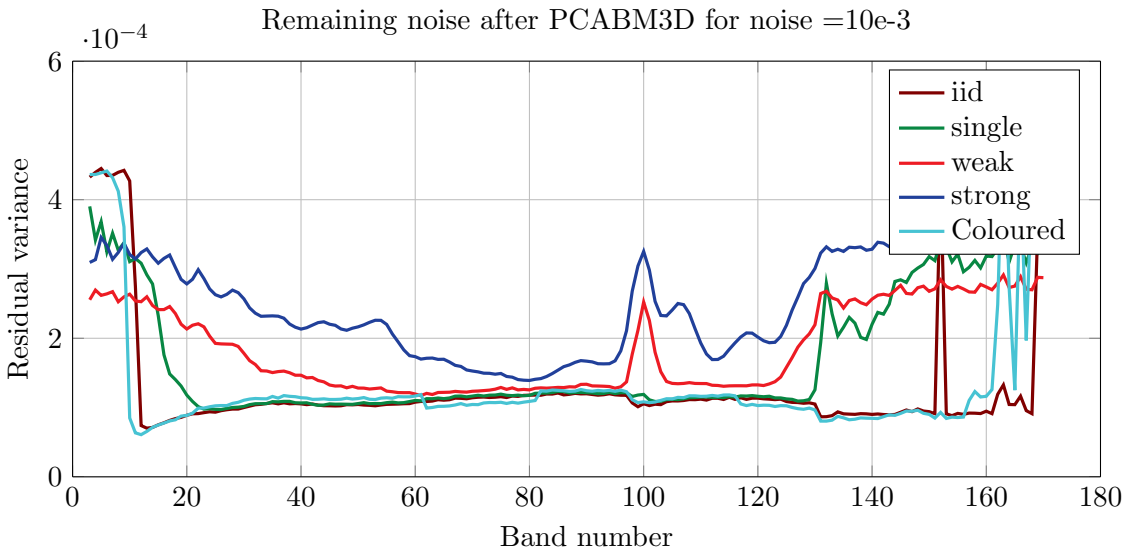


Figure 48: PCABM3D method applied to noise with variance 10^{-3}.

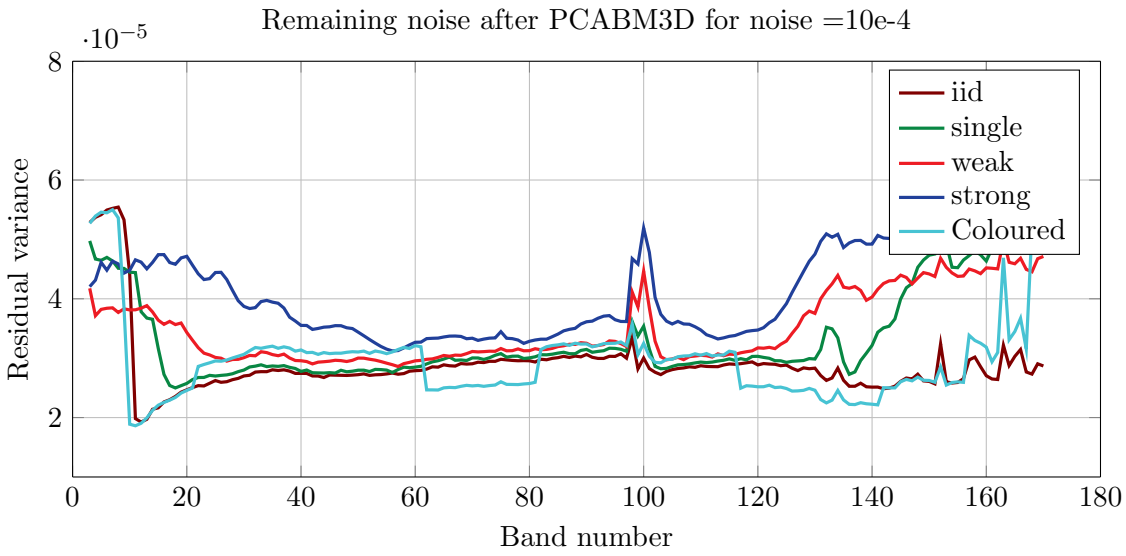


Figure 49: PCABM3D method applied to noise with variance 10^{-4}.

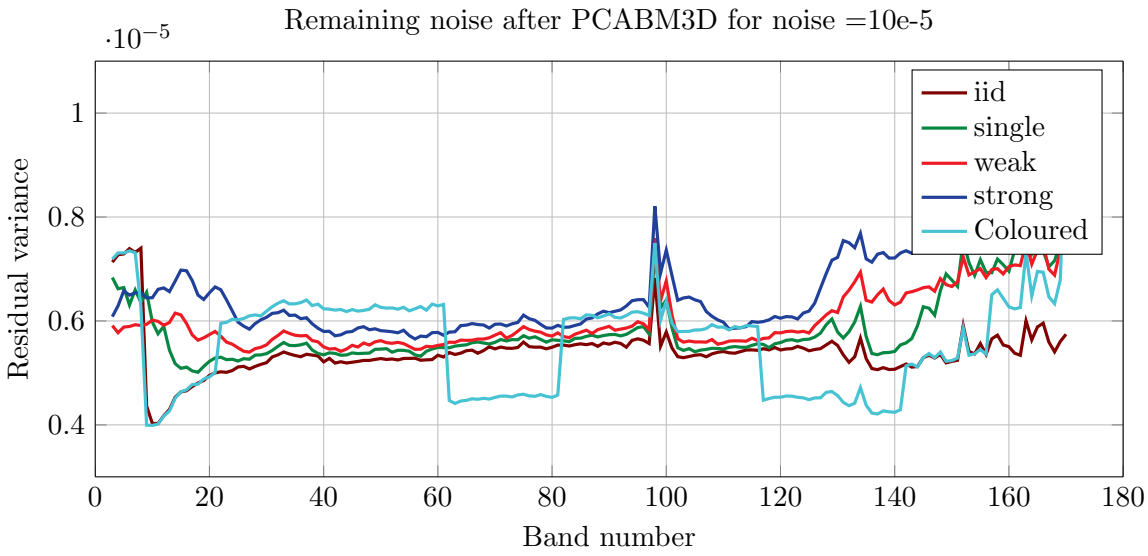


Figure 50: PCABM3D method applied to noise with variance 10^{-5}.

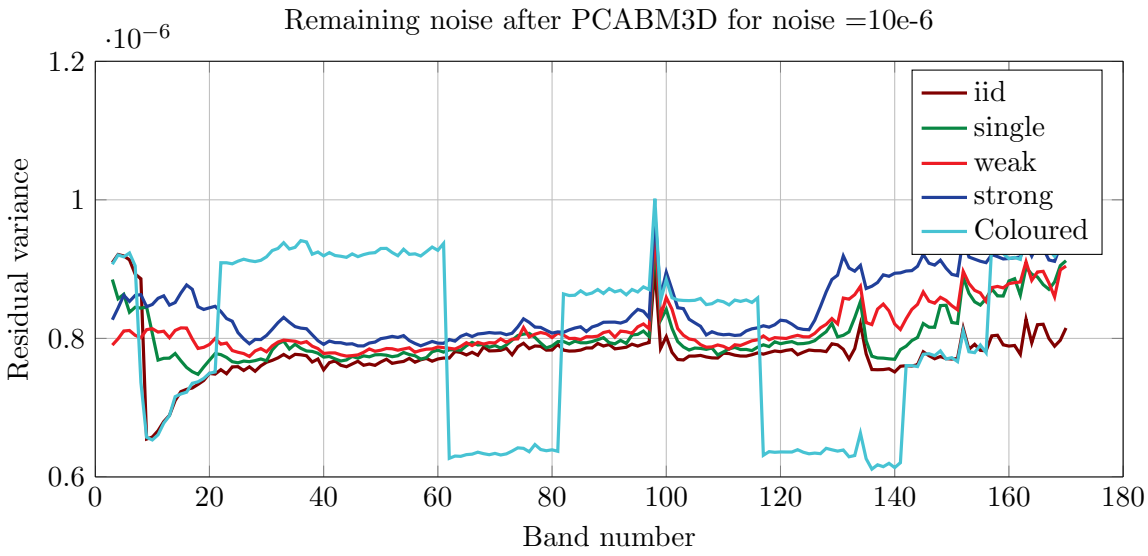


Figure 51: PCABM3D method applied to noise with variance 10^{-6}.

6.1.7 BM4D

6.1.7.1 Noise Removal As the BM4D algorithm uses spectral information in its implementation, we anticipated that noise correlation would have a negative impact on the efficacy of the method. This is evident in figures 52-56.

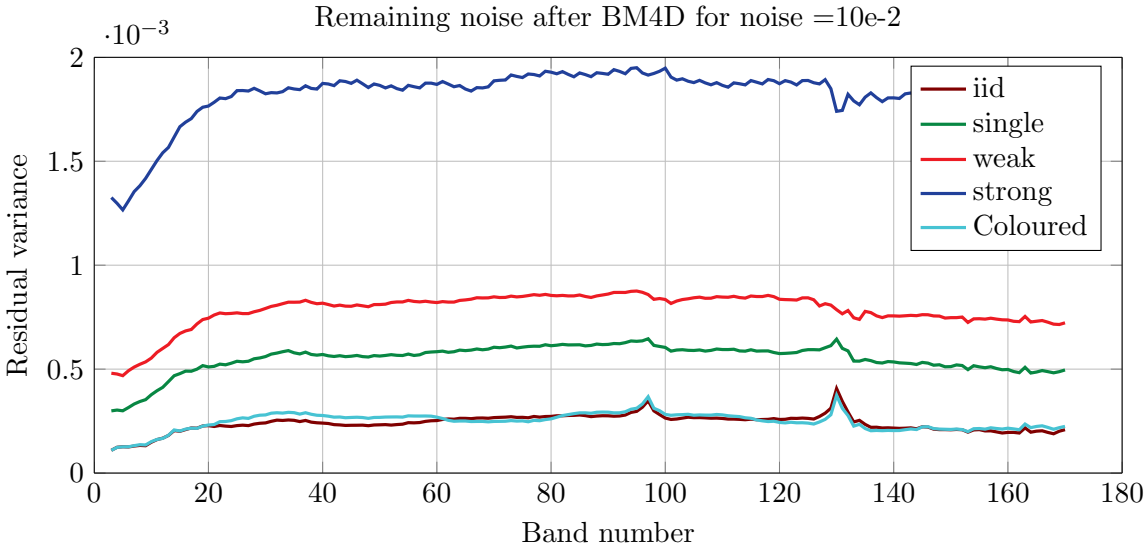


Figure 52: BM4D method applied to noise with variance 10^{-2}. The method has removed most of the noise from the image, the more correlated noise types had more noise remaining but still 85% of the noise has been removed.

The BM4D method has performed very well at all noise levels we tested even under correlated noise. The method demonstrates more consistent results than the other block matching algorithms, that is the method peaks at higher noise levels but stabilises and removes a consistent percentage of the noise regardless of noise level. An interesting observation is that the response is worse if the noise correlation stretches over multiple bands than if the correlation coefficient is higher but only neighbouring bands are correlated.

Since this method demonstrated the best performance of the denoising methods we did further experiments to find out how the method performs with an incorrect estimate of the variance. We used a ten percent overestimate and a 10 percent underestimate. The method demonstrated very little dependence on the accuracy of the estimate and actually performed better with an overestimate of the noise. We see this in figure 57, the graphs for the other noise types are not shown because they are essentially identical. The measure that we use in the graph is the same we as use in the summary section (equation 15).

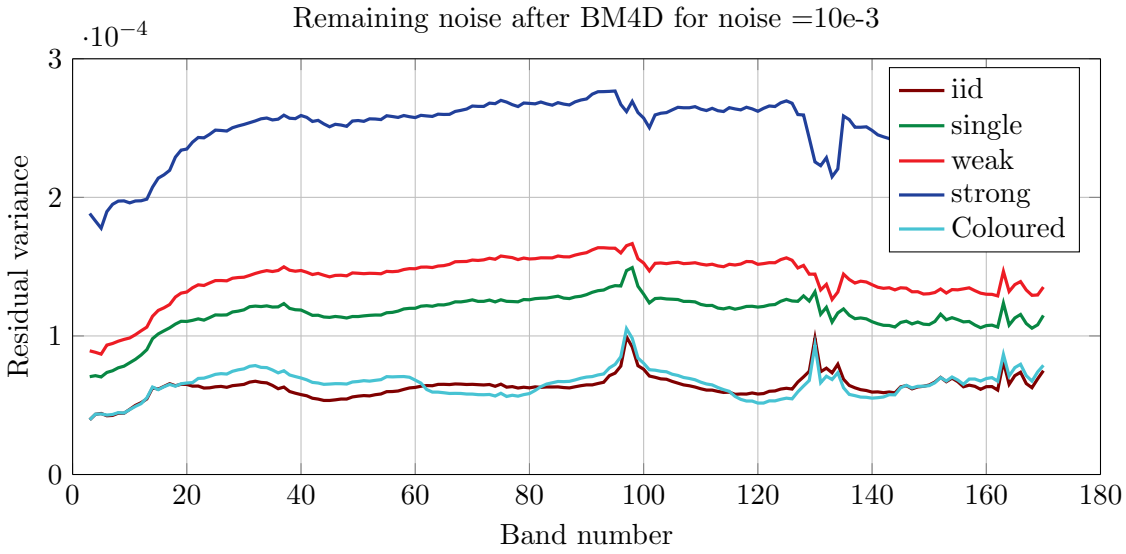


Figure 53: BM4D method applied to noise with variance 10^{-3}. The method has again reduced the noise present by a significant margin but has been less effective than it was at 10^{-2}. The uncorrelated noise has been reduced by about 90% while the strongly correlated noise has been reduced by 60%.

The method is however one of the most costly in terms of time taking of the order of 1350 seconds. There is room for parallelisation that has not been exploited, e.g. groups could be processed in parallel.

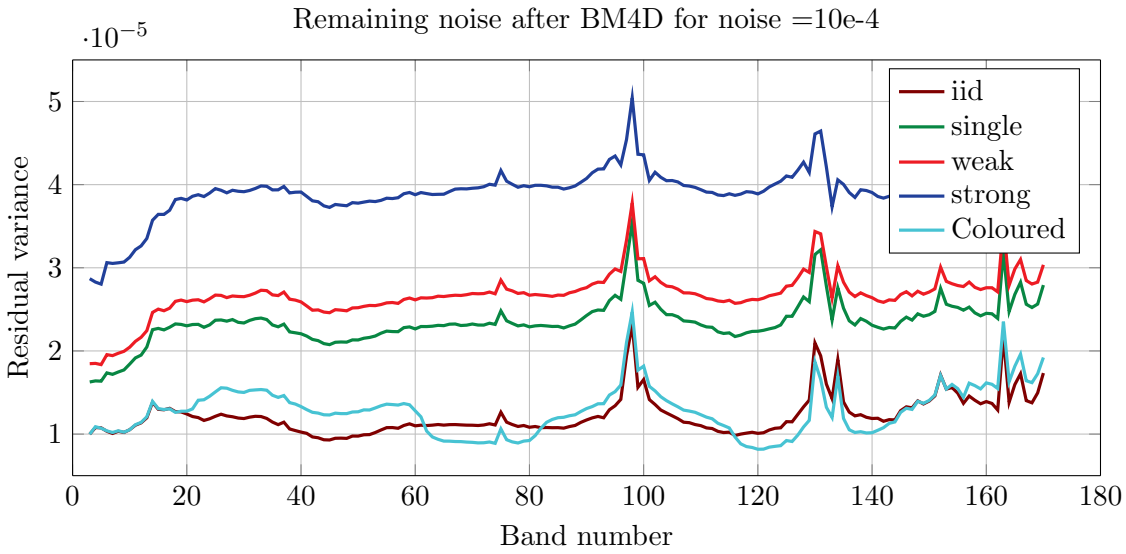


Figure 54: BM4D method applied to noise with variance 10^{-4}. Again the method has been less effective than at the higher noise values. The i.i.d. noise has been reduced by 90% while the strongly correlated noise has been reduced by 65%.

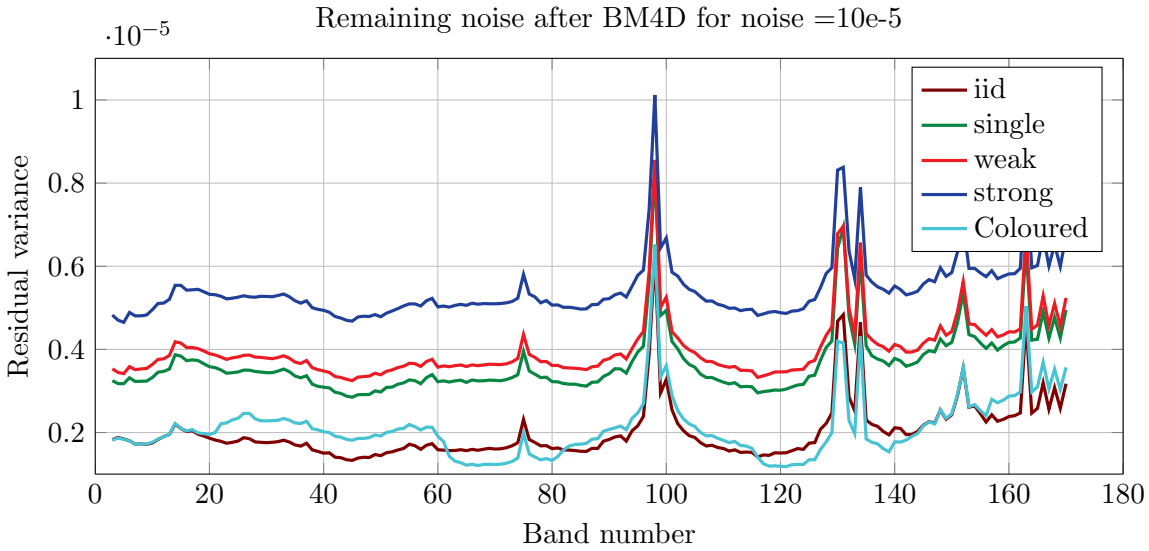


Figure 55: BM4D method applied to noise with variance 10^{-5}. Again the method has been less effective than at the higher noise values. The i.i.d. noise has been reduced by about 85% while the strongly correlated noise has been reduced by 50%. Spikes in the residual noise have begun to appear.

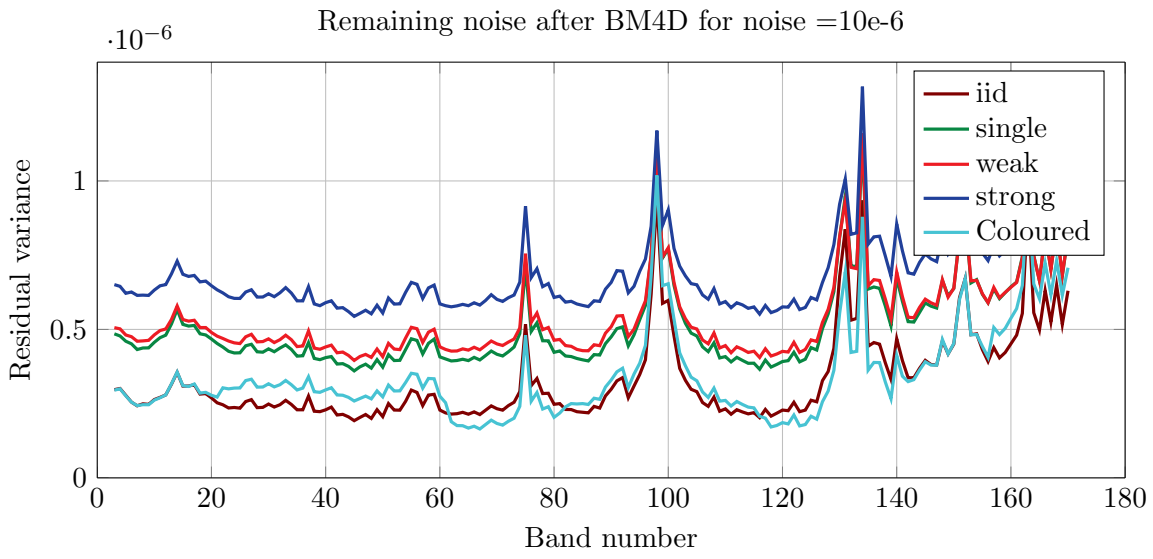


Figure 56: BM4D method applied to noise with variance 10^{-6}. Again the method has been less effective than at the higher noise values. The i.i.d. noise has been reduced by about 78% while the strongly correlated noise has been reduced by 40%. Spikes in the residual noise have become pronounced.



Figure 57: BM4D method applied to i.i.d. noise with over and underestimates of the variance. We see little effect with an estimate error of 10%.

6.1.8 Summary

In the following section we compare the methods directly against each other for each noise type at all noise levels. For each noise type the normalised root mean squared error for the variance estimate is given by:

$$\frac{1}{\bar{\sigma}^2}\sqrt{\frac{\sum_{i=1}^{p}(\hat{\sigma}_i^2-\sigma_i^2)^2}{p}} \tag{15}$$

where $\bar{\sigma}^2$ is the mean variance in the image(average of the covariance main diagonal).

This is plotted against the initial variance. For convenience the initial variance is expressed as a log scale so the graphs are more legible. The aim is to gain insight into which methods are best under which conditions. This will determine how we interpret the results of noise estimates for the real data. The first 2 data points (corresponding to $\sigma^2 = 10^{-5}$ and 10^{-6}) for Meer's method are omitted since they are extreme outliers and the method clearly fails at low noise levels.

For each noise type we plot the normalised average variance of the remaining noise after the denoising methods are applied given by

$$\frac{1}{p\bar{\sigma}^2}\sum_{i=1}^{p}\sigma_i^2 \tag{16}$$

again this is plotted on a log scale for the initial variance.

6.1.8.1 i.i.d. We see in figure 58 the trend that the residual based methods produce the best results. The standard residual has an error less than 0.1 at all noise levels. However the modified residual method produces an improvement to this estimate even when the correlation structure is overestimated (MR5). The best estimates are still given when the correlation structure is correctly estimated (MR0).

In figure 59 we see unsurprisingly that the standard regression method performs fairly well with a residual variance less than 40% of the original noise. Interestingly this estimate is better than the bBM3D and PCA-BM3D methods at all noise levels. The BM4D method however is the best at low noise levels (which we expect from modern instruments) by a fair margin. Another interesting observation we make from this graph is that the PCA-BM3D method has out performed the bBM3D method at all but the highest noise level (10^{-2}).

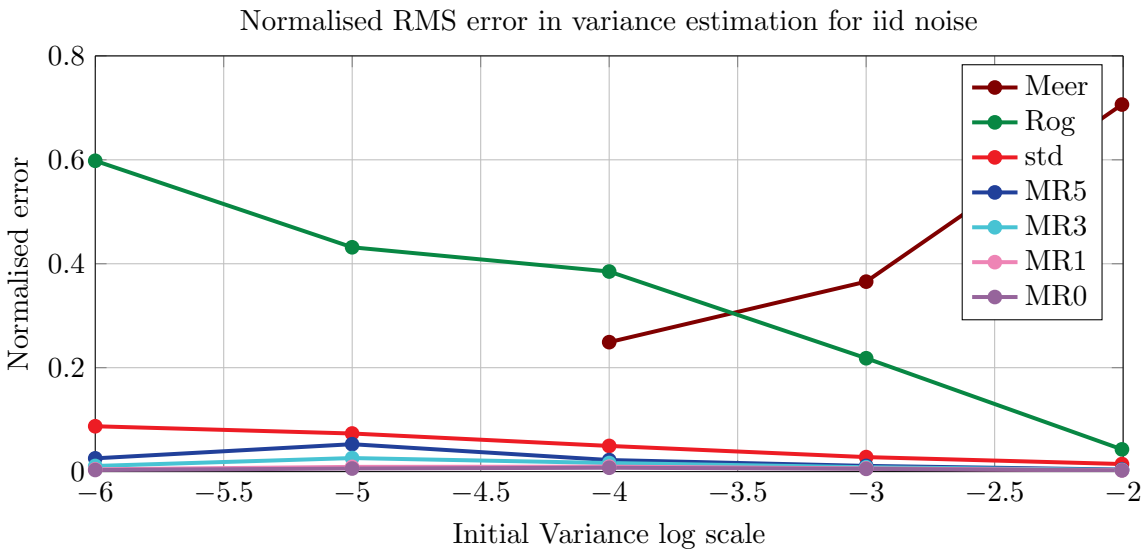


Figure 58: All of the noise statistic estimation methods are plotted on the same log axis of the initial variance. We see that the best estimate at all noise levels is given by the modified residual method under the assumption that the noise is uncorrelated.

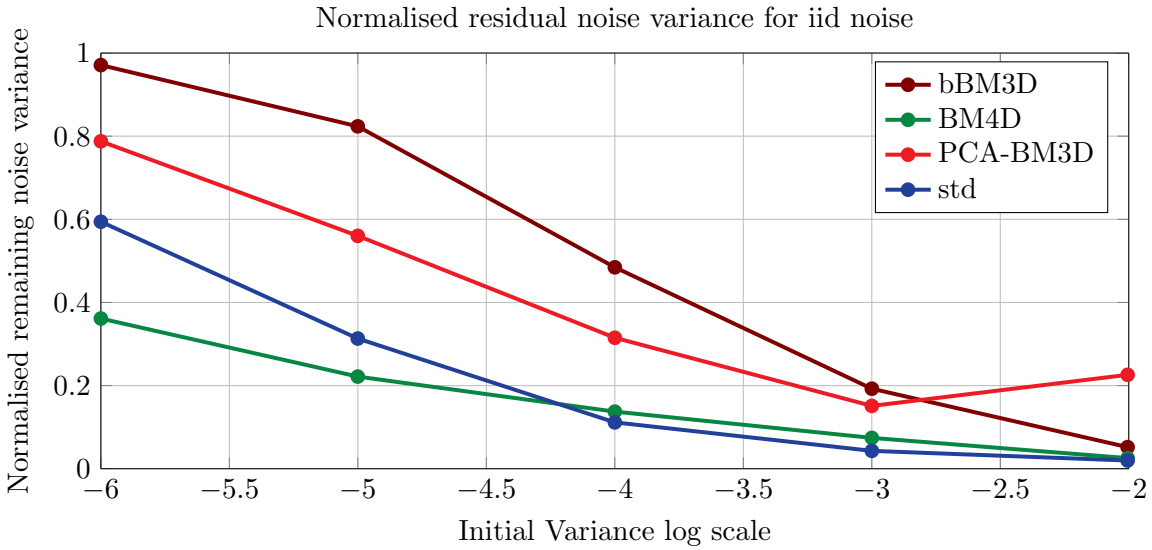


Figure 59: All of the noise removal methods are plotted on the same log axis of the initial variance for the case of i.i.d. noise. The graph shows that at low noise levels the BM4D method is the best while the standard regression method is better at higher noise levels.

6.1.8.2 Coloured The graphs for coloured uncorrelated noise are omitted because they are essentially identical to the i.i.d. case.

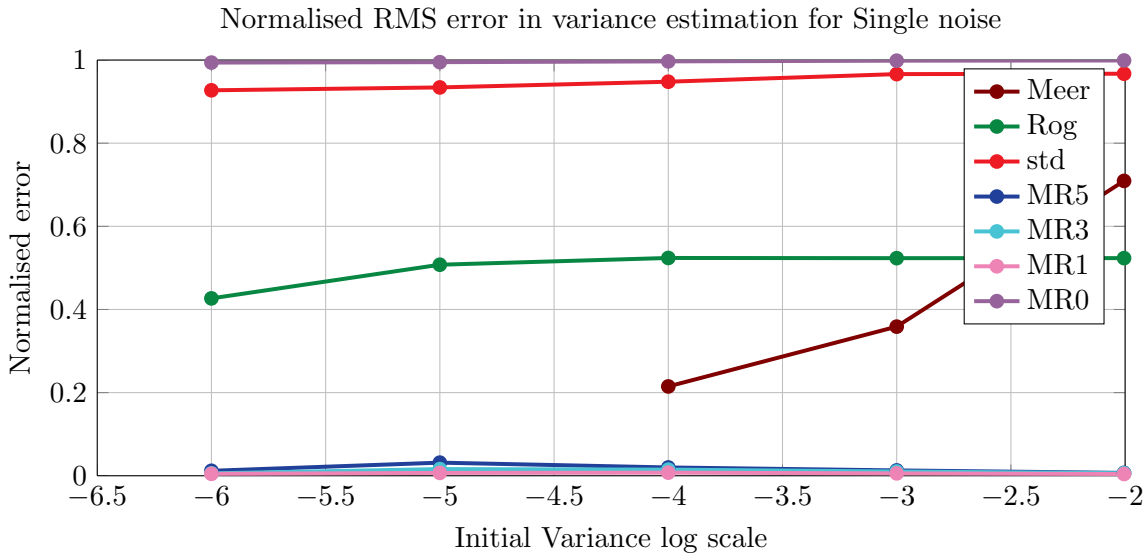


Figure 60: All of the noise statistic estimation methods are plotted on the same log axis of the initial variance for the case of single neighbour correlation. The modified regression method with the correct assumption of the correlation structure (MR1) produces the best estimate at all noise levels.

The results from figure 60 are clear. The modified regression method with the correct assumption of correlation structure (MR1) is the best at all noise levels followed (almost negligibly) by the modified regression method with an overestimation of the correlation structure (MR3 and MR5). Rogers method performs adequately as well with an error between 0.4 and 0.6. Unsurprisingly the standard regression method has performed poorly under the correlated noise and the modified residual method with an underestimation of the correlation structure (MR0) has produced a worse estimate still.

In figure 61 we see that the standard regression method is ineffective in the extreme and being unsuccessful at removing the noise. Interestingly we see again that the PCA-BM3D method has performed better than bBM3D even under the correlated noise present in the image contrary to our intuition. We see again that the best results were obtained with BM4D this time for all noise levels.

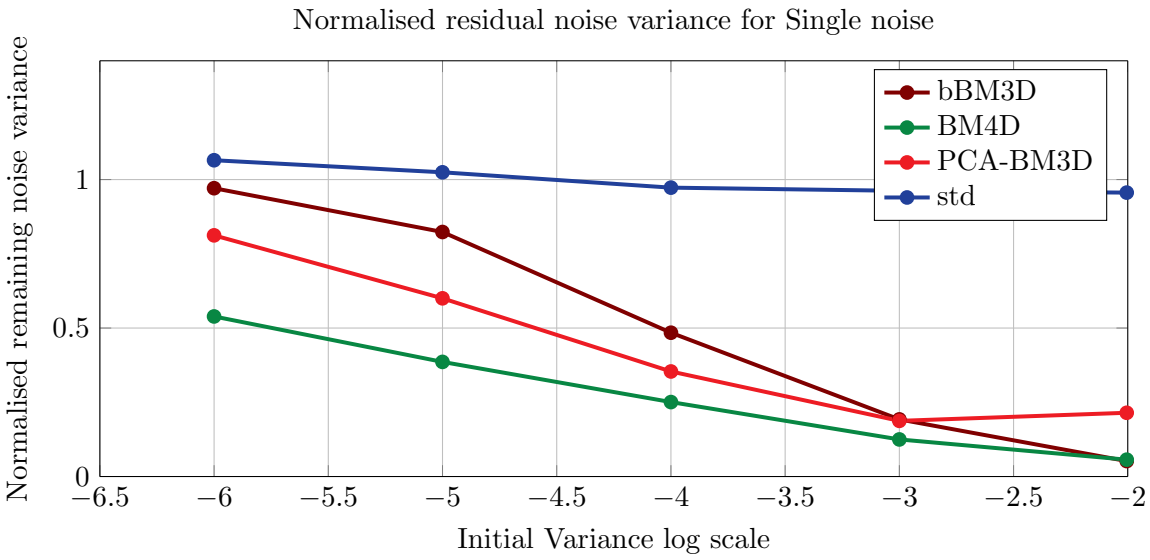


Figure 61: All of the noise removal methods are plotted on the same log axis of the initial variance for the case of single neighbour correlation. The BM4D method has produced the best result at all noise levels.

6.1.8.3 Weakly correlated The results in the estimation experiments (shown in figure 62) are similar to that of single neighbour correlation. There are some differences however. Roger's method has performed better (by about 30% according to the measure we have defined) since there is a smaller correlation coefficient in the neighbouring band. MR1 (as an underestimate of the correlation structure) now performs worse than the standard regression method but better than MR0. The best results are with MR3 (which corresponds to a perfect guess of the correlation structure) closely followed by MR5 (an overestimation of the correlation structure) showing again that an overestimate of the correlation structure is better than an underestimation of the correlation structure.

Again the results from the noise removal experiments show clear results in figure 63. The BM4D method gives the best image of the methods that we tested. The PCA-BM3D was again better than the bBM3D and the standard regression and lies in the same error range as the single neighbour correlation showing that the extended correlation doesn't make the image harder to denoise than stronger single neighbour correlation.

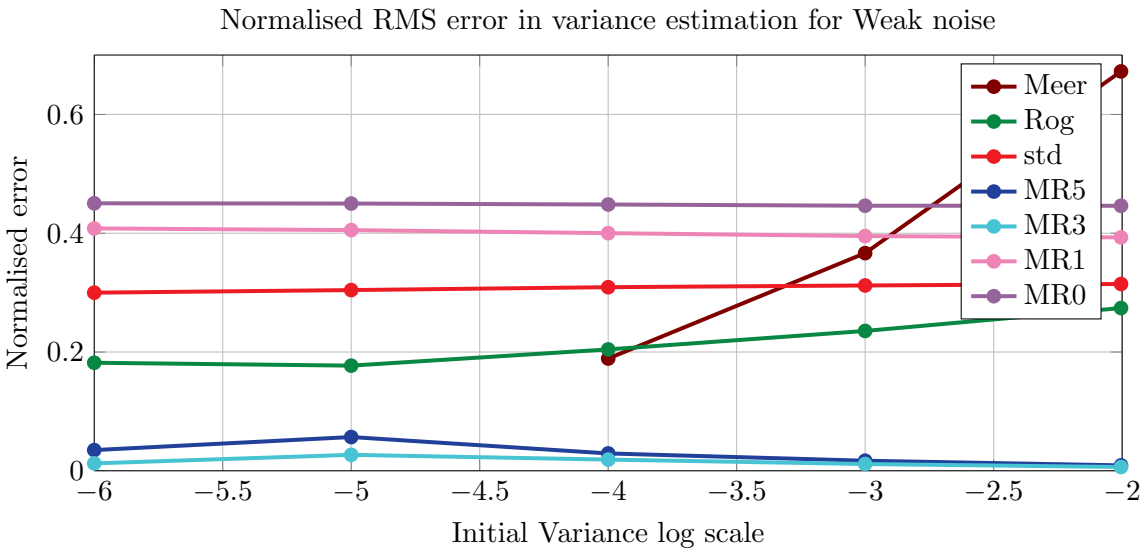


Figure 62: All of the noise statistic estimation methods are plotted on the same log axis of the initial variance for the case of weakly correlated noise. The best estimators are again the modified regression estimates. MR3 which is the correct correlation guess is slightly better than MR5.

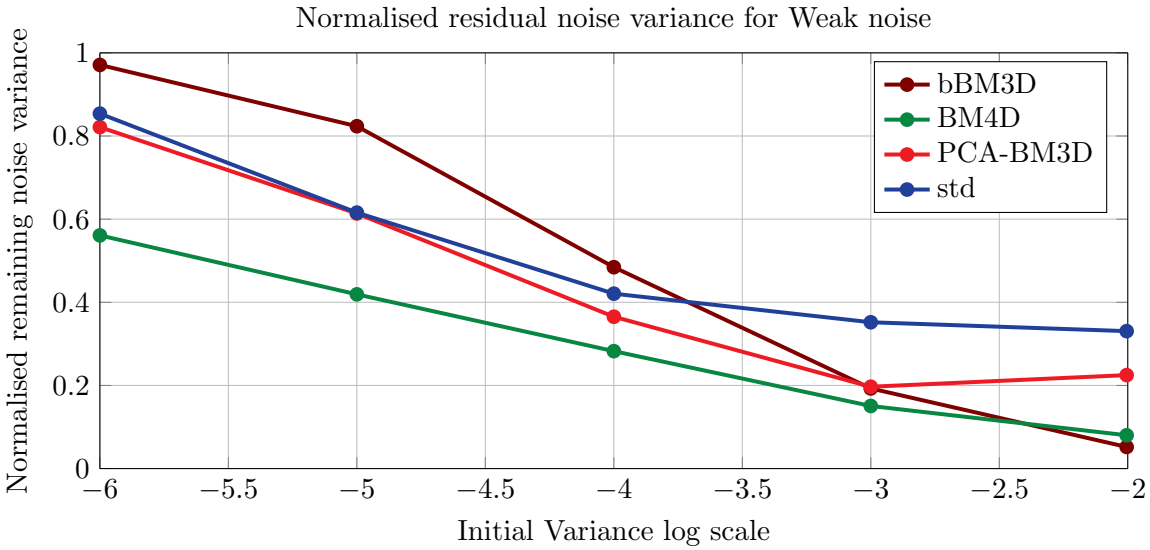


Figure 63: All of the noise removal methods are plotted on the same log axis of the initial variance for the case of weakly correlated noise. The BM4D method again has produced the best final image.

6.1.8.4 Strongly correlated The results in the estimation experiments (shown in figure 64) contrast how sensitive some of the methods are to correlation when we compare the results to that of weak correlation. The error in estimation for MR0, MR1 and the standard regression method is almost 100% of the initial variance. Roger's method now produces estimates that have errors of the order 60% of the noise (a figure more than twice as bad as in the weakly correlated noise). Meer's method demonstrated a similar performance which is unsurprising as it does not rely on any spectral information. Again the best results are with MR3 closely followed by MR5.

Again the results from the noise removal experiments show clear results in figure 65. The BM4D method gives the best image of the methods that we tested at the lower noise levels. The PCA-BM3D was again better than the bBM3D at the lower noise levels. The standard residual now produced what is clearly the worst estimate of the methods tested at all noise levels. At the highest noise levels (10^{-3} and 10^{-2}), the method which has no spectral dependence (bBM3D) produces the best result.

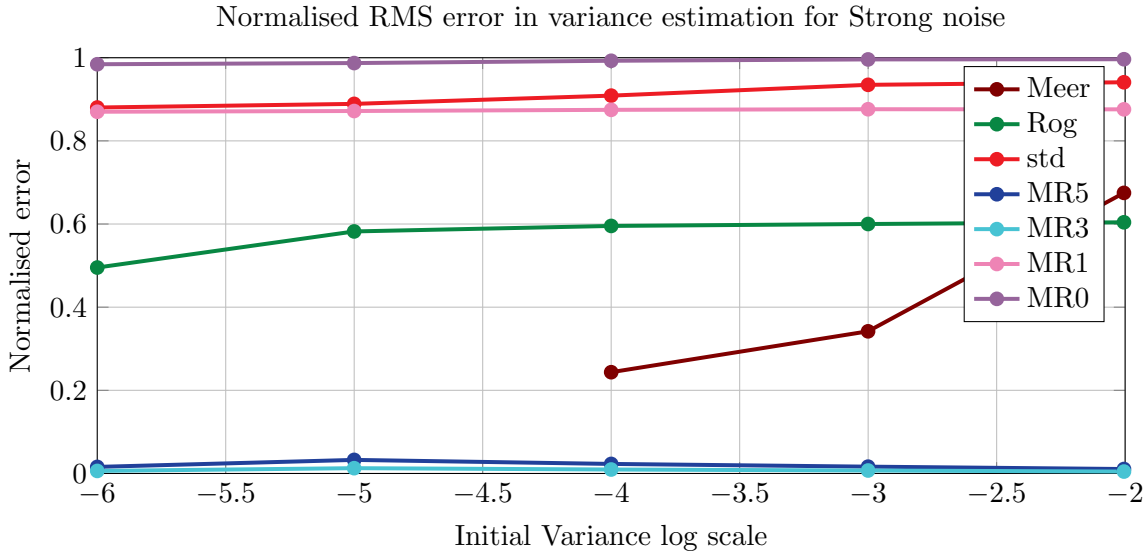


Figure 64: All of the noise statistic estimation methods are plotted on the same log axis of the initial variance.

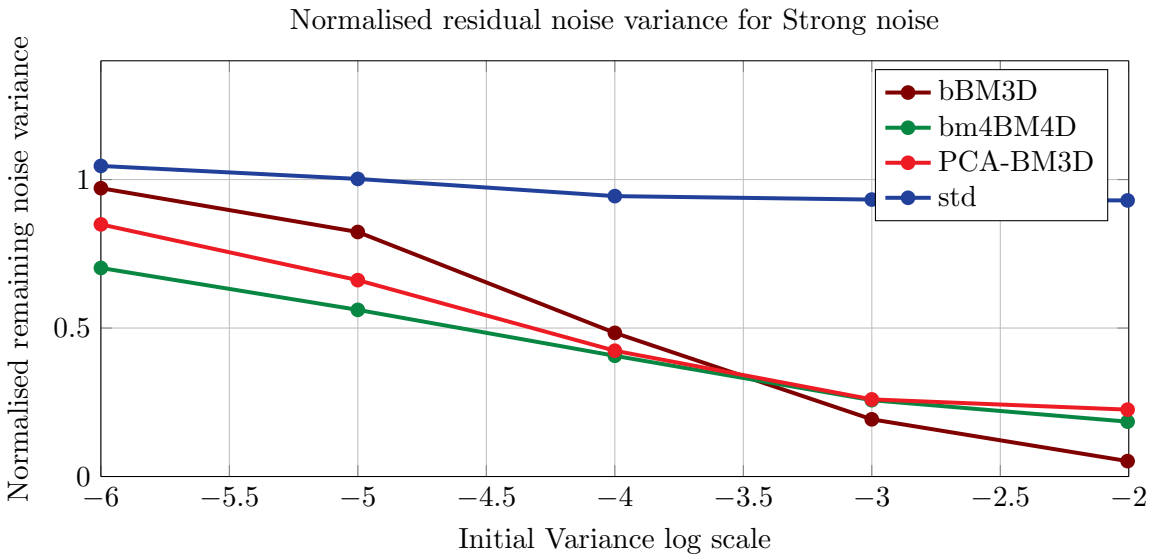


Figure 65: All of the noise removal methods are plotted on the same log axis of the initial variance.

6.1.9 Concluding Remarks

The results of our observations suggest a clear strategy for tackling the problem of noise in hyperspectral images. An initial estimate of the variance should be obtained using the modified regression method. It is not an unreasonable assumption that the noise in hyperspectral images is only correlated significantly between a few neighbouring bands; to this end we believe that in a close situation the results of MR5 could be treated as accurate. This being said we can examine the variances estimated for several correlation assumptions (MR3-MR20 say) to get a feel for the variance. As we increase the correlation assumption the estimate of the variance itself grows and at a certain point beyond the true correlation the estimate begins to grow rapidly. Looking for this acceleration point can give insight into what is a good guess to the correlation structure. We will use this approach in the real data processing. Multiple MR will be assessed and an estimate will be obtained after inspection.

The noise removal is a clearer answer. If the noise is estimated to be on the low end (10^{-5} and less), which we expect it to be in modern instruments, the BM4D method should be applied to obtain the best estimate regardless of correlation structure in the noise. The fact that the BM4D algorithm demonstrated little dependence on the accurate estimation of the variance is another reason to apply this method over others. Alternatively if time is a consideration in the data processing (i.e. if a large amount of data needs to be processed) the PCA-BM3D

method will provide the next best result in a reduced time period.

6.2 Real Data

Our observations about statistic estimation methods tells us that the modified regression approach produces the best results under a generous guess of the correlation structure (i.e. an overestimation of the number of correlated bands). As such we will proceed with the real data analysis by calculating the estimates of the covariance matrix using the modified residual method for neighbouring correlation estimates from 1 to 50. We will analyse the estimates produced to determine our best guess of the correlation structure and the noise level. To accomplish this we make use of the observation that the estimate of the variance grows sharply after a certain overestimation of the correlation structure. We will also compare the estimates to those produced by the standard regression method and Roger's method to aid in our decision.

Once we have determined our estimate of the variance we will feed the image into BM4D together with the estimate to produce an estimate of the true signal.

We follow this procedure for each of the three subsets of the Cuprite radiance scene.

6.2.1 First Pair of Spectrometers

We omit the graphs for the modified regression method cases with high correlation assumptions that we believe represent overestimates of the noise correlation.

A common observation in the estimates produced using all methods is that the first ten or so bands exhibit significantly higher variances than the remaining bands. If this is indeed the case it points to either a weakness in the sensor at the higher frequency side of the spectrum or noise that is dependent on the response in the sensor (i.e. signal dependent noise) since these bands have a greater response as seen in the mean value plot in figure 2.

We also see in the estimates that some of the methods (Roger's, MR1 and the standard regression methods) tend to estimate the noise to be greater in the subset with the greatest mean value (subset three) which also has the lowest signal variance. This again could point to noise that is dependant on the response in the instrument (signal dependency). This observation is not so prominent in the MR3 and MR5 experiments and we believe it may be method errors that depend on the brightness of the image. This variability should be tested

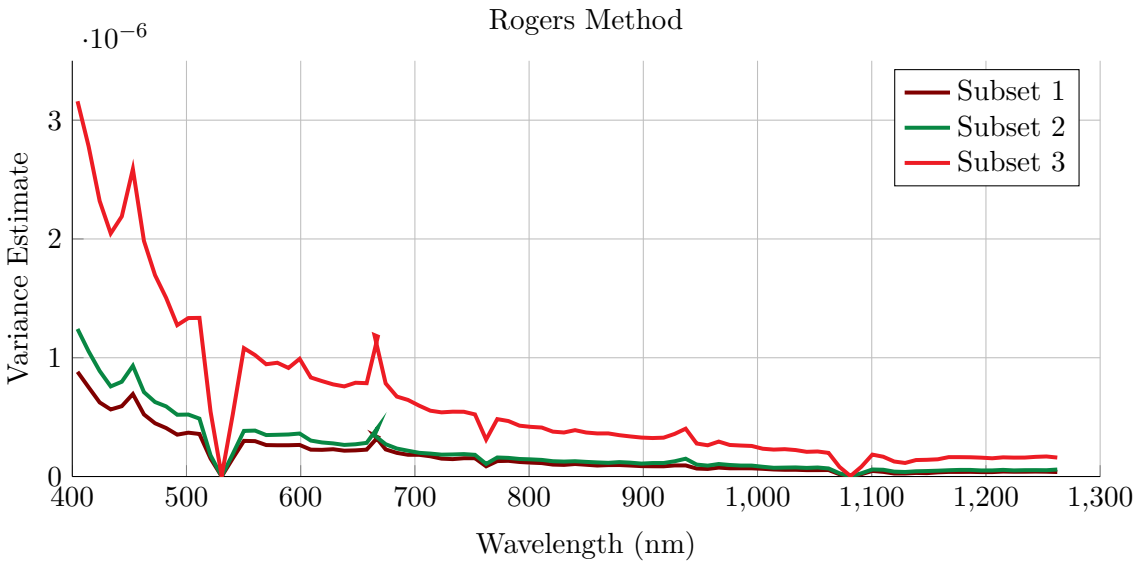


Figure 66: The variance per band as determined by Roger's method for all three subsets for the first 2 spectrometers. Roger's method estimates a greater noise variance for the third subset of the image, this subset is also the brightest of the three.

in future experiments on a greater set of artificial images with varying brightnesses.

Our observations at higher correlation assumptions (MR6 onwards) show very little change until MR10 when the variance estimates start to grow rapidly (the graph loses its fairly monotonic behaviour). From this observation we assume that the estimate produced by MR5 is a good guess and this is supported by Roger's method (which tended to overestimate the variance at low noise levels) and the standard regression method. This leads us to assume a cautious estimate of $0.5\text{x}10^{-6}$ for the variance in the image and use it to apply the BM4D algorithm.

In figures 71 - 73 we plot the difference between three pixels before and after processing with the BM4D algorithm. We see a very small difference in the pixels post processing. The difference between the individual pixels is also quite stable and random consistent with what we would expect from noise.

In figures 74 - 76 we plot the variance of the data removed with the BM4D algorithm (i.e. what BM4D deems to be the noise present in the image). The variance is less than our initial estimate and interestingly has a very similar profile across all the subsets which would suggest that it is either a property of the sensor or atmospheric interference.

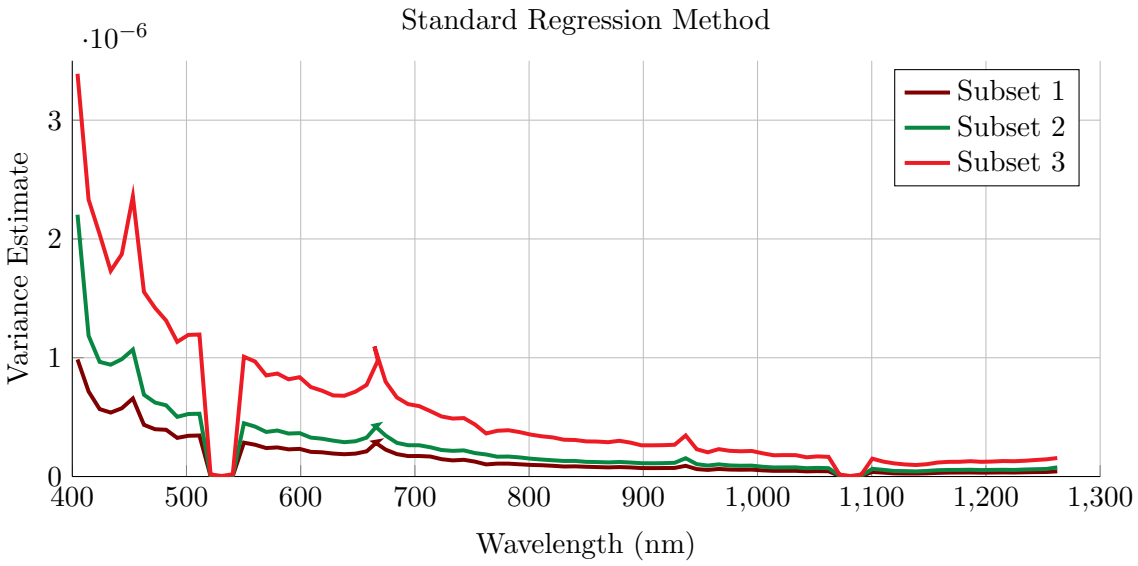


Figure 67: The variance per band as determined by the standard regression method for all three subsets for the first two spectrometers. The standard regression method estimates a greater noise variance for the third subset of the image. This subset is also the brightest of the three.

In figure 77 we draw the correlation matrix estimates determined by MR20. A white pixel corresponds to a 1 and black to a 0. We see that the structure of the correlation matrices is very similar for all three subsets. This shows a very clear indication of the structure of the noise in the AVIRIS instrument. The block matrix structure that is evident is not unexpected.

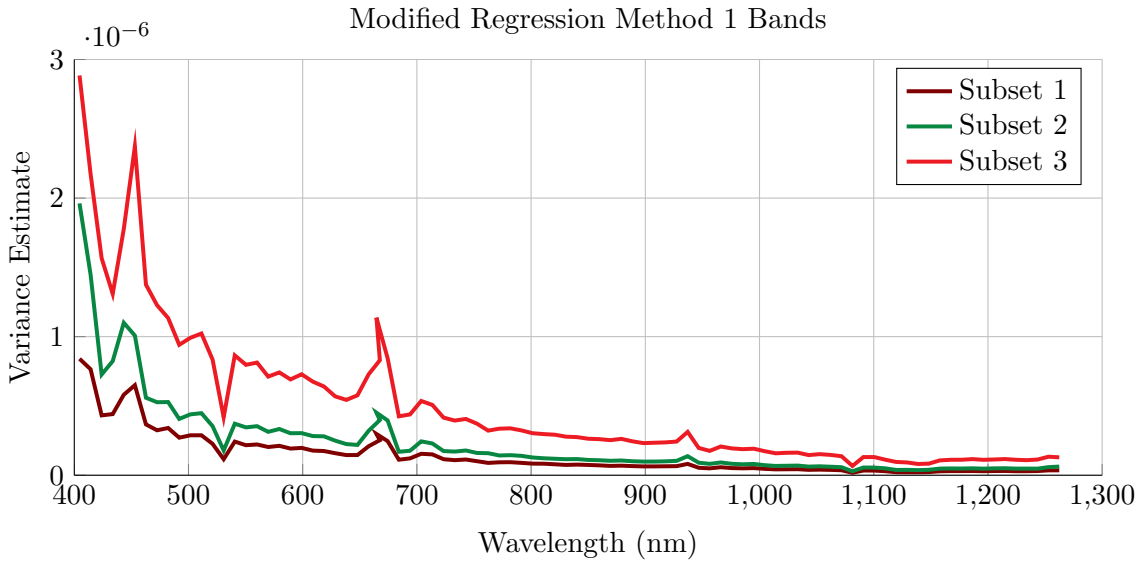


Figure 68: The variance per band as determined by the MR1 method for all three subsets for the first two spectrometers. The MR1 method estimates a greater noise variance for the third subset of the image. This subset is also the brightest of the three.

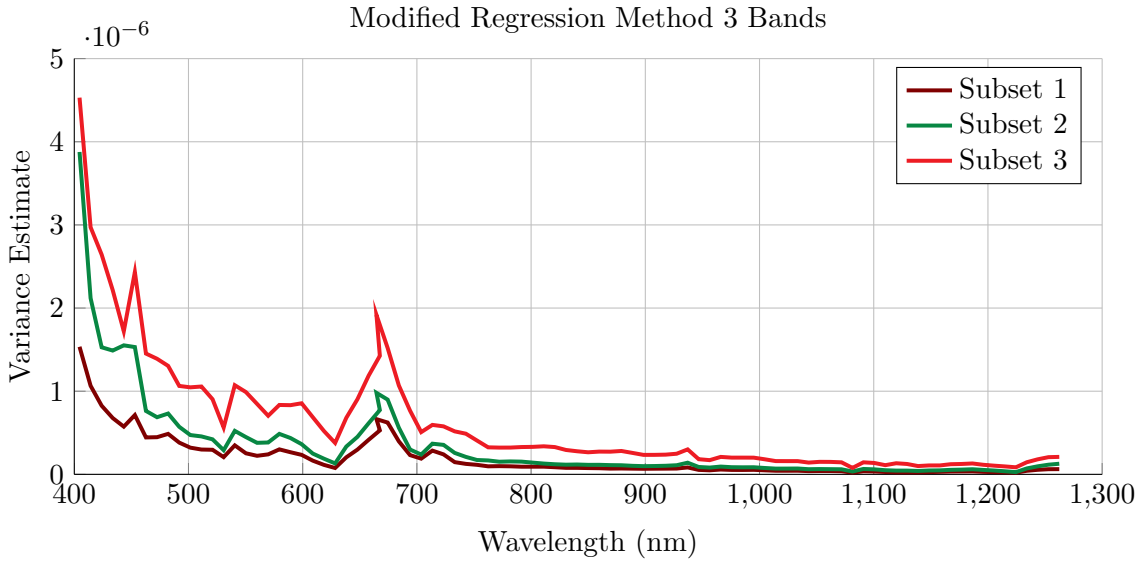


Figure 69: The variance per band as determined by the MR3 method for all three subsets for the first two spectrometers. The method shows a smaller separation in the variance estimates of the three subsets as compared to MR1, Roger's and the standard regression methods.

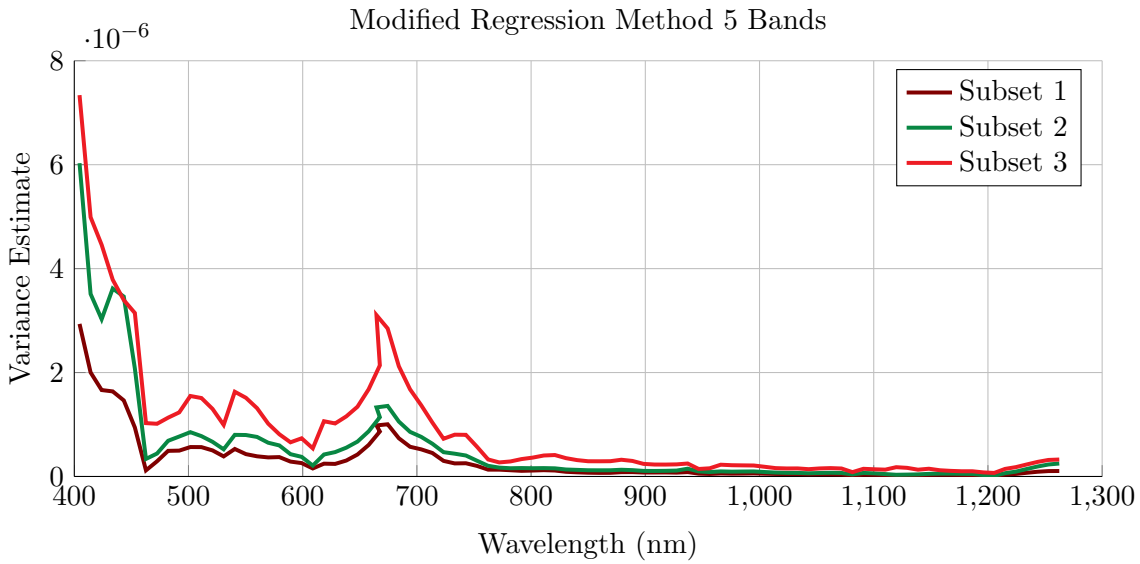


Figure 70: The variance per band as determined by the MR5 method for all three subsets for the first two spectrometers. The method shows a smaller separation in the variance estimates of the three subsets as compared to MR1, Roger's and the standard regression methods.

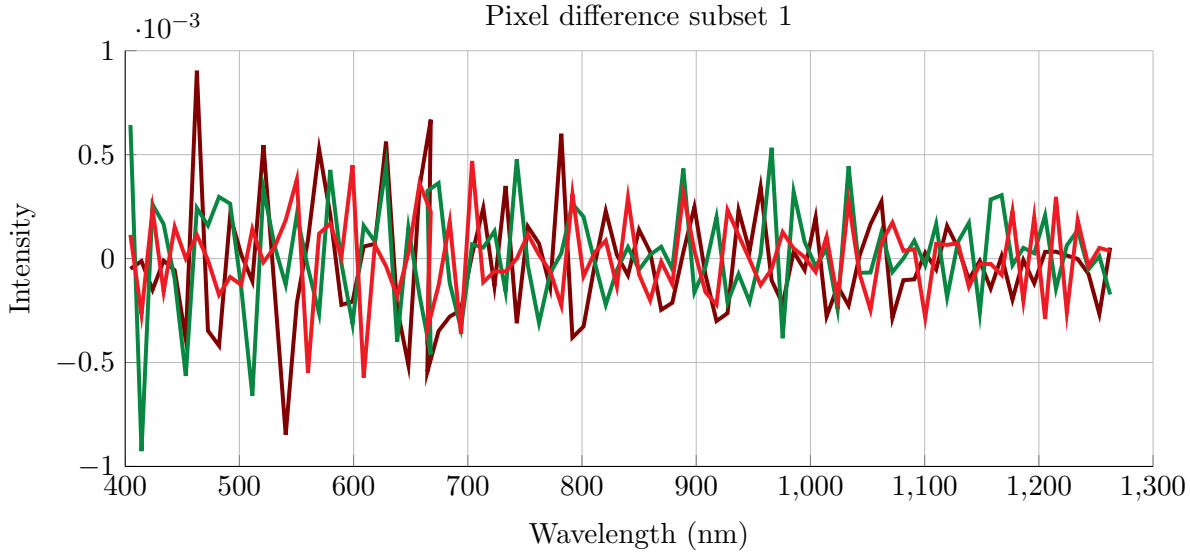


Figure 71: The difference in the pixel values through each band for three pixels in the image for the first subset.

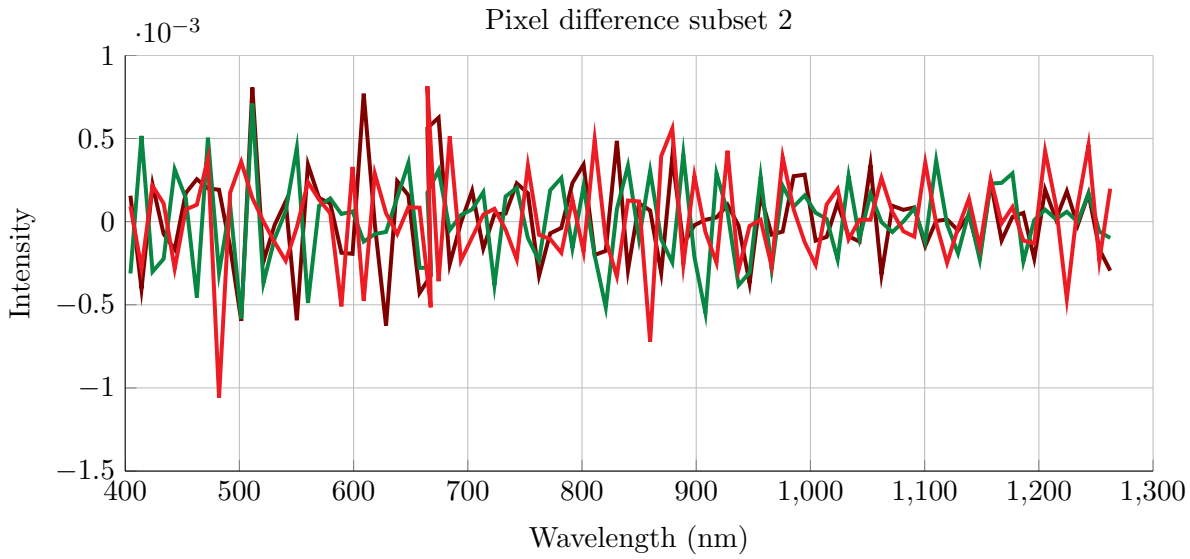


Figure 72: The difference in the pixel values through each band for three pixels in the image for the second subset.

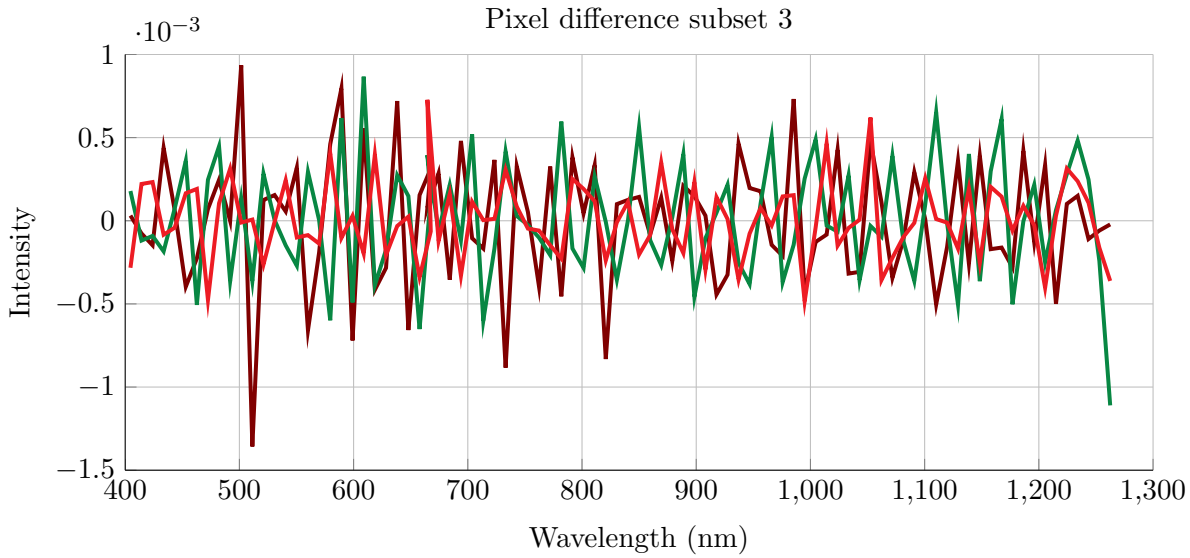


Figure 73: The difference in the pixel values through each band for three pixels in the image for the third subset.

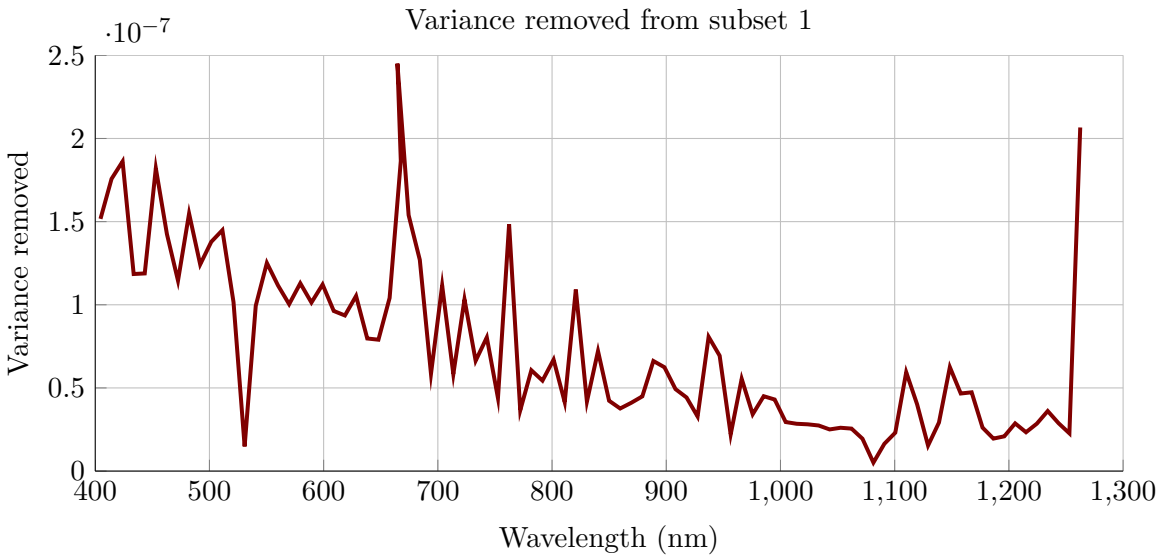


Figure 74: The variance of the "noise" removed by BM4D for the first pair of spectrometers for the first subset.

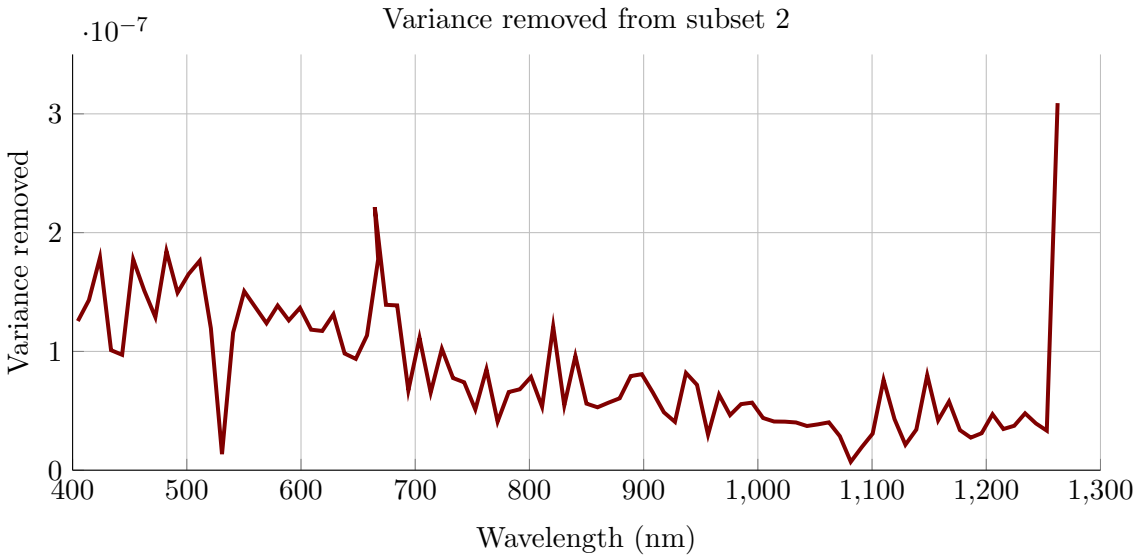


Figure 75: The variance of the "noise" removed by BM4D for the first pair of spectrometers for the second subset.

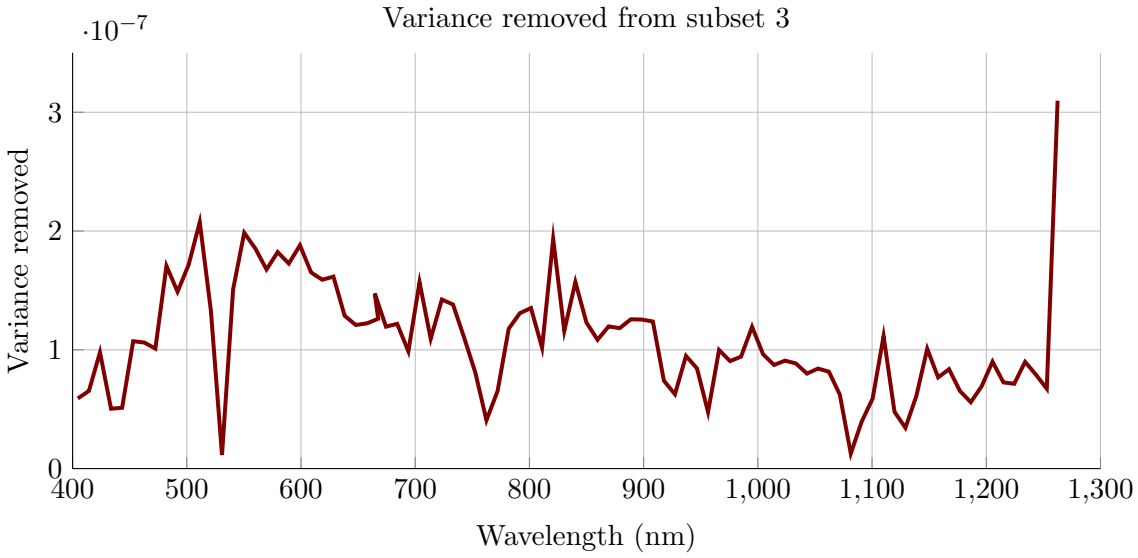
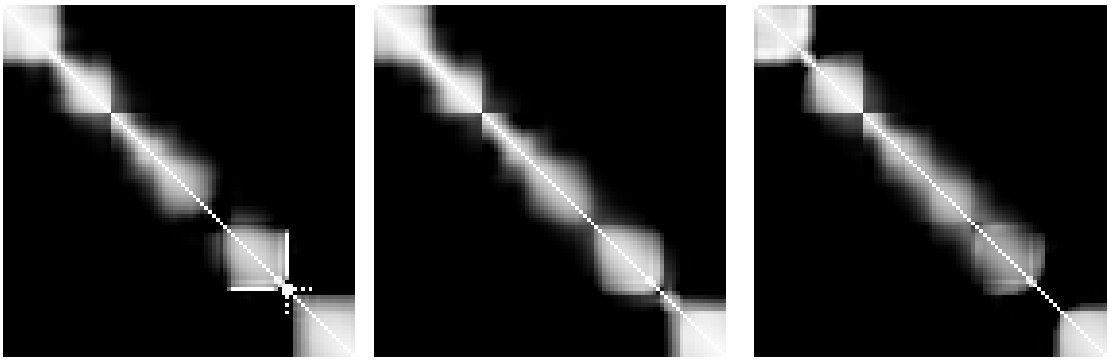


Figure 76: The variance of the "noise" removed by BM4D for the first pair of spectrometers for the third subset.



(a) First subset. (b) Second subset. (c) Third subset.

Figure 77: The correlation matrix estimates determined by MR20 for the three subsets in the first pair of spectrometers. The structure is similar for all subsets although the third subset is slightly less similar than the other two.

6.2.2 Second Pair of Spectrometers

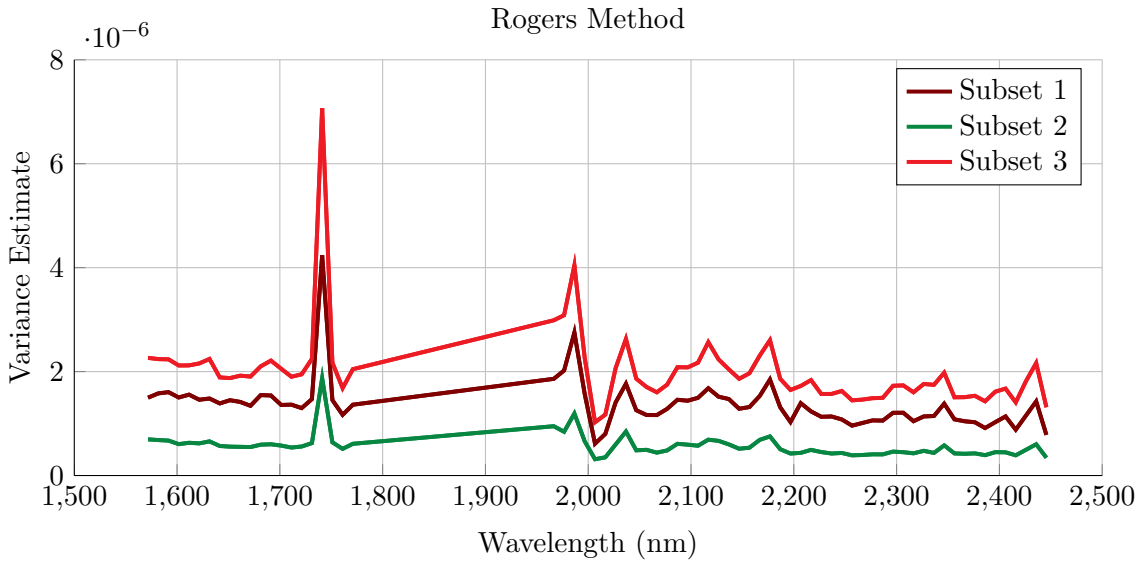


Figure 78: The variance per band as determined by Roger's method for all three subsets for the second 2 spectrometers.

In the second pair of spectrometers there is a clear trend in the results through all the methods. The noise in subset three is greater than the noise in subset one which is in turn greater than the noise in subset two. This points clearly to image dependant noise, whether it is simply response based (greater responses in the camera generate greater noise variances) or spatially dependent.

It is worth noting that the methods all agree that there is more noise in the second pair of spectrometers than in the first two.

Our observations at higher correlation assumptions (MR6 onwards) show very little change until MR10 when the variance estimates start to grow rapidly and inconsiderately (the graph loses its fairly monotonic behaviour). From this observation we assume that the estimate produced by MR5 is a good guess and this is supported by Roger's method (which tended to overestimate the variance at low noise levels) and the standard regression method. This leads us to assume a cautious estimate of $2\text{x}10^{-6}$ for the variance in the image.

In figures 83 - 85 we plot the difference between three pixels before and after processing with the BM4D algorithm. We see a very small difference in the pixels post processing. The difference between the individual pixels is also quite stable and random consistent with what

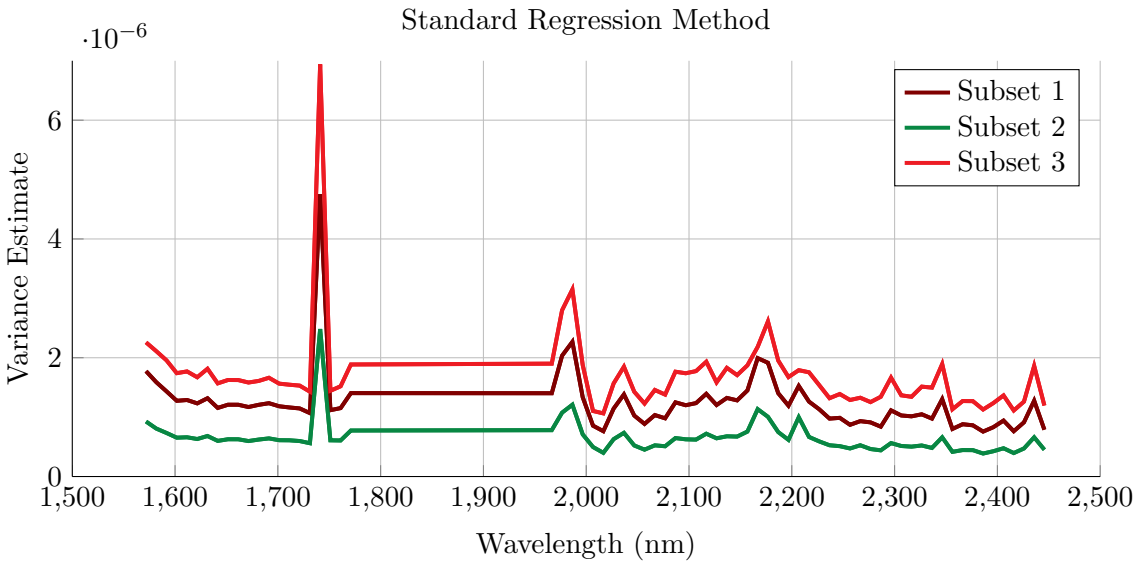


Figure 79: The variance per band as determined by the standard regression method for all three subsets for the second 2 spectrometers.

we would expect from noise.

In figures 86 - 88 we plot the variance of the data removed with the BM4D algorithm (i.e. what BM4D deems to be the noise present in the image). Again the variance is less than our initial estimate and interestingly has a very similar profile across all the subsets, which suggests that it is either a property of the sensor or atmospheric interference.

In figure 89 we draw the correlation matrix estimates determined by MR20. A white pixel corresponds to a 1 and black to a 0. We see that the structure of the correlation matrices is very similar for all three subsets. This shows a very clear indication of the structure of the noise in the AVIRIS instrument. The block matrix structure that is evident is not unexpected.

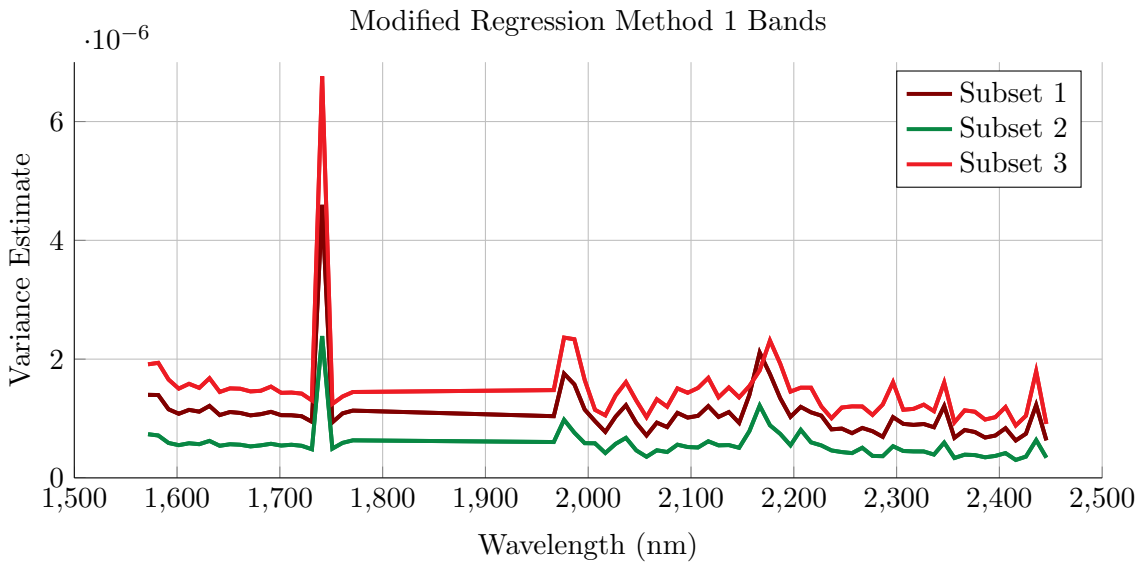


Figure 80: The variance per band as determined by the MR1 method for all three subsets for the second 2 spectrometers.

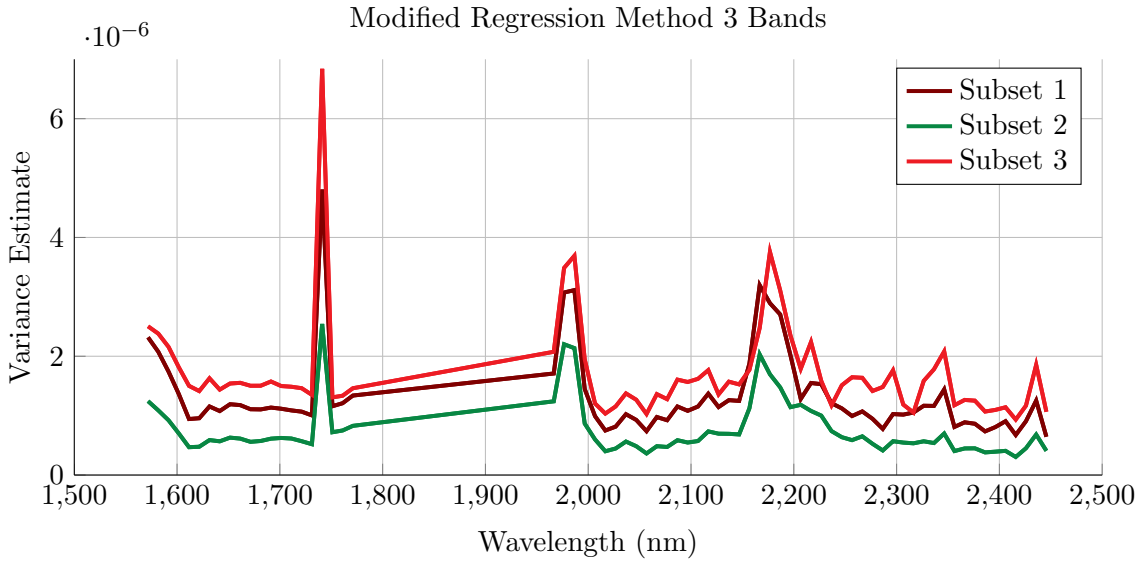


Figure 81: The variance per band as determined by the MR1 method for all three subsets for the second 2 spectrometers.

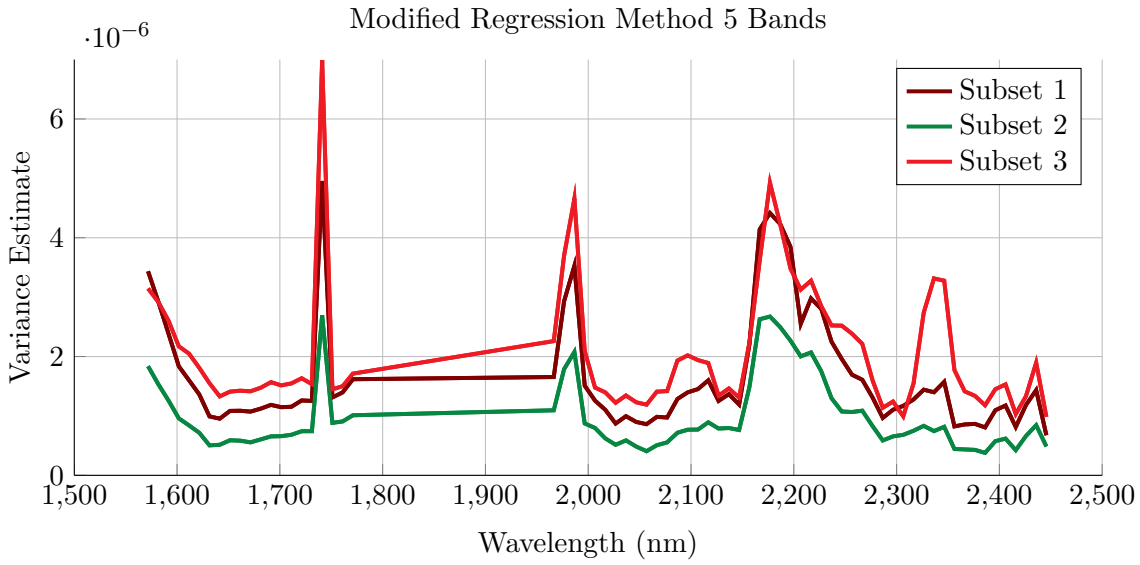


Figure 82: The variance per band as determined by the MR1 method for all three subsets for the second 2 spectrometers.

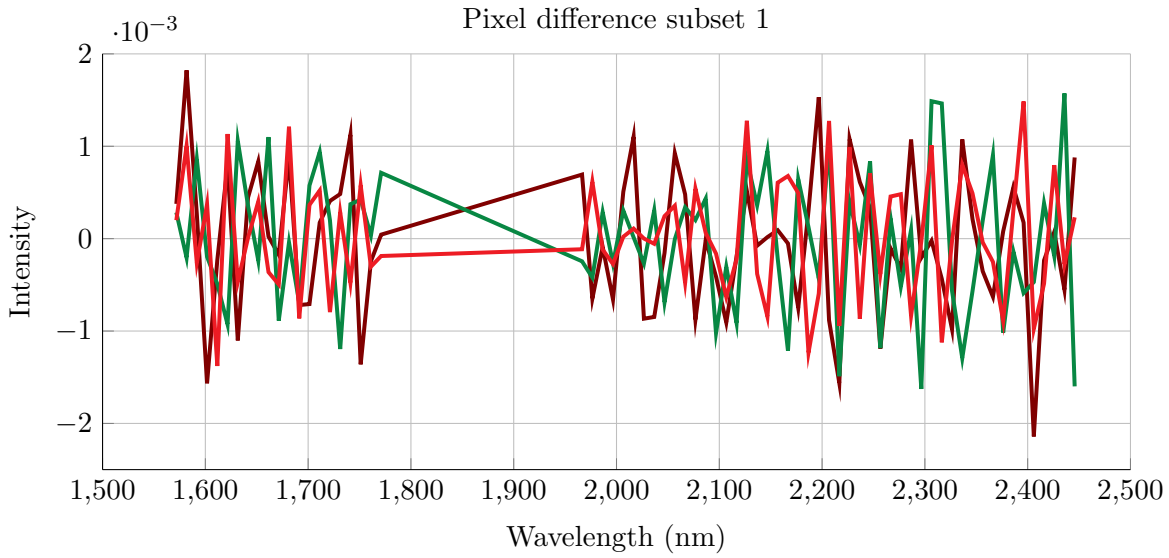


Figure 83: The difference in the pixel values through each band for three pixels in the image for the first subset.

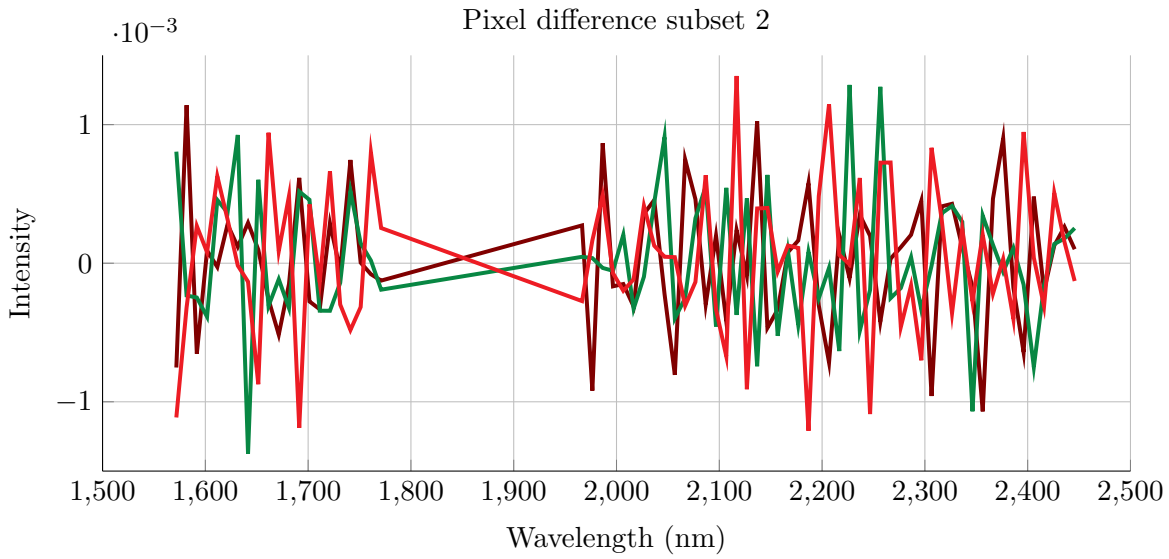


Figure 84: The difference in the pixel values through each band for three pixels in the image for the second subset.

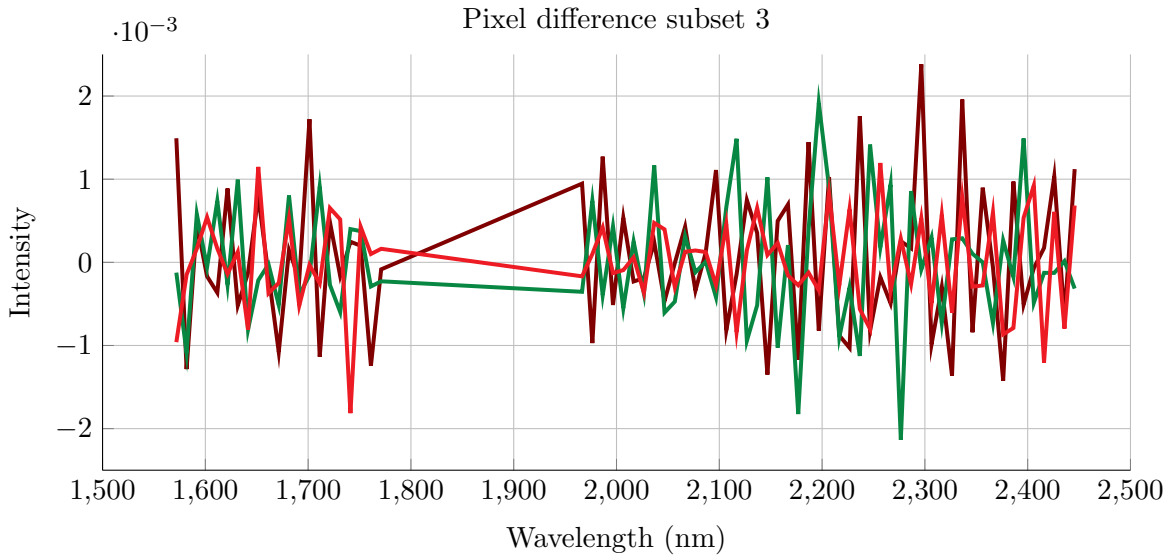


Figure 85: The difference in the pixel values through each band for three pixels in the image for the third subset.

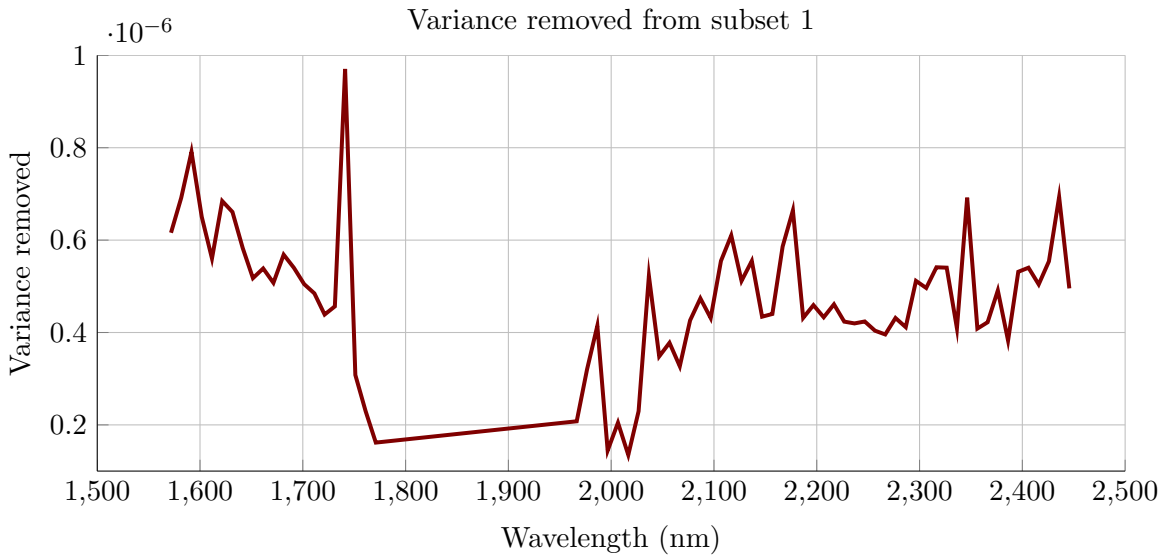


Figure 86: The variance of the "noise" removed by BM4D for the second pair of spectrometers for first the subset.

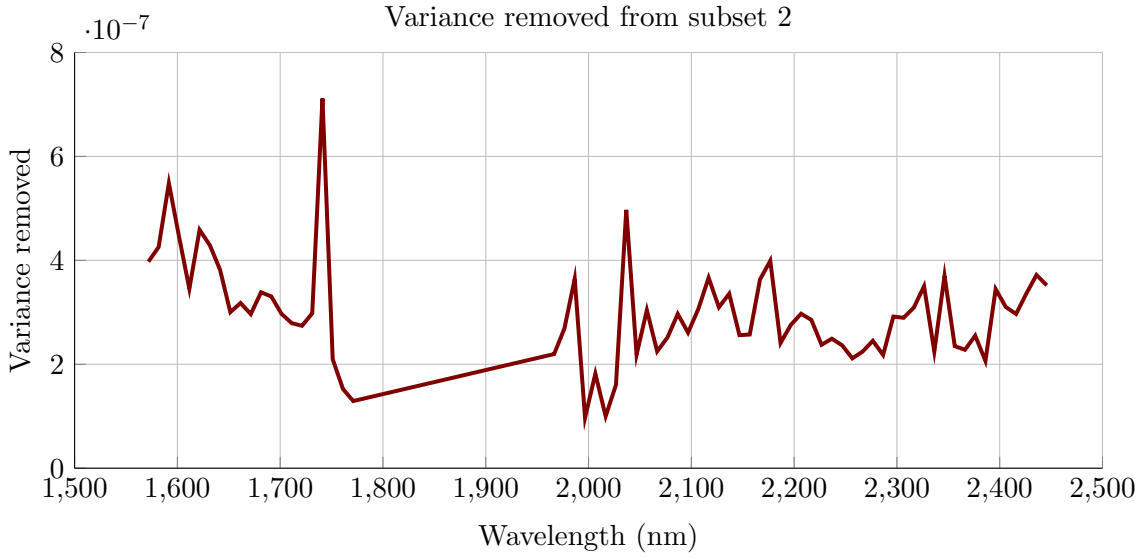


Figure 87: The variance of the "noise" removed by BM4D for the second pair of spectrometers for second the subset.

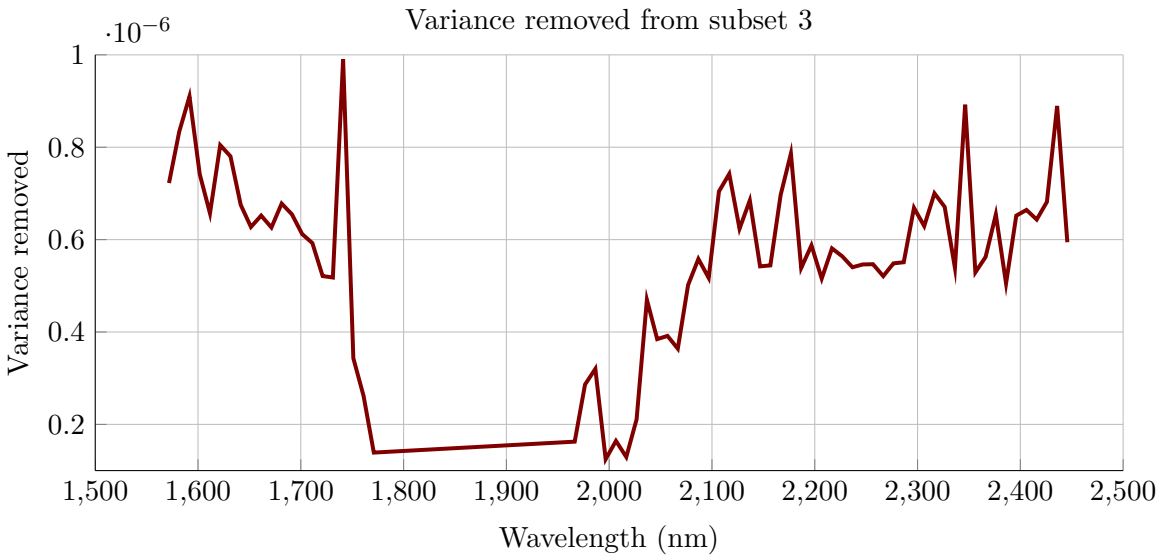
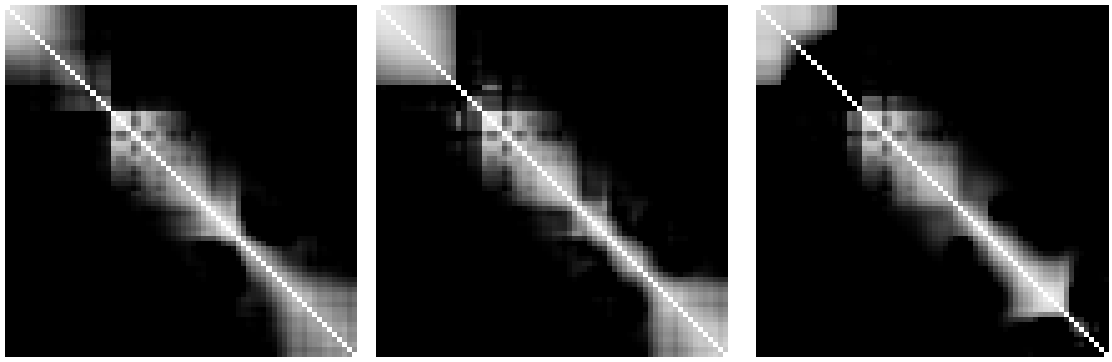


Figure 88: The variance of the "noise" removed by BM4D for the second pair of spectrometers for third the subset.



(a) First subset. (b) Second subset. (c) Third subset.

Figure 89: The correlation matrix estimates determined by MR20 for the three subsets in the second pair of spectrometers. The structure is similar for all subsets although the third subset is slightly less similar than the other two.

6.2.3 Signal to noise ratio in the instrument

We now present a comparison of the SNR in the instrument as determined by MR1 for all three subsets against laboratory SNR measurements. The lab measurements from [36] only cover the instrument as it was in 1997, however it is sufficient to show the typical profile of the AVIRIS instrument. The choice of MR1 is fairly arbitrary as all the methods produce a similar profile and nothing is gained by presenting all of them. We define SNR as

$$\mathbf{SNR} = \frac{Var(\mathbf{S})}{\sigma^2}$$

Since we do not know the true signal we make use of the assumption that the noise and the signal are uncorrelated. This means that

$$Var(\mathbf{X}) = Var(\mathbf{S}) + \sigma^2$$

and so we can write

$$\mathbf{SNR} = \frac{Var(\mathbf{X}) - \sigma^2}{\sigma^2}.$$

When we normalised our the radiance images for processing we divided all the intensity values by the greatest value in the image instead of $2^{15} - 1$ which effectively performed a contrast stretch to the subsets. The contrast stretch introduces a factor into the variance of the signal and will effect the value of the noise variance determined as well. To make the profiles of all the SNR estimates comparable we have scaled them. The graphs plotted in figure 91 have been scaled by the following factors: subset 1 by 1/9.6, subset 2 by 1/16 and subset 3 by 1/6.4.

The SNR plots from our experiments show agreement across the subsets on the SNR profile with some variation (as you would expect). Additionally if we take the scaling as a true scaling, we see that the trend in figure 90 of increasing SNR in the instrument is followed in our 2008 data. SNR has increased in all bands just as in the 1994 SNR compared to the 1997 SNR in figure 90.

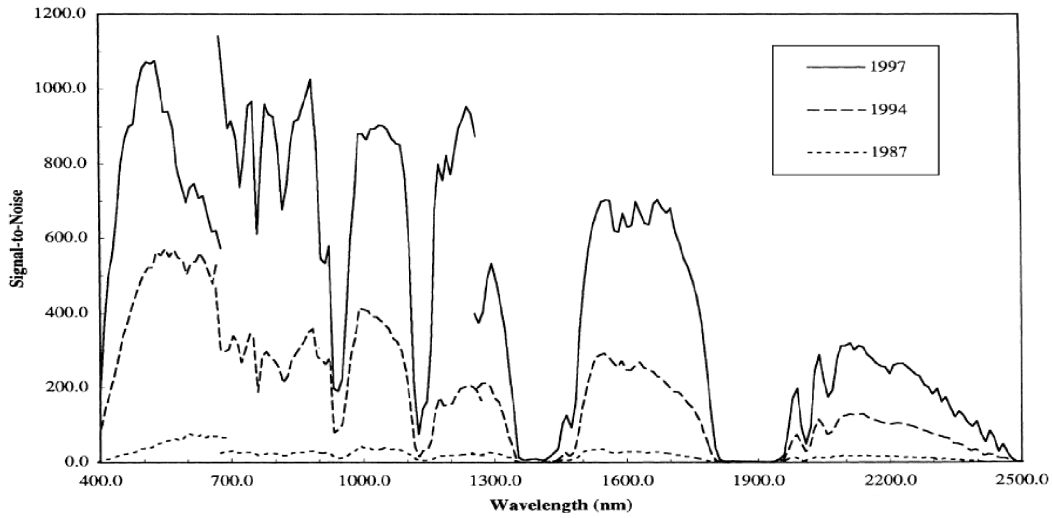


Figure 90: The SNR for the years 1987, 1994 and 1997 as measured in [36]. Graph taken from [36].

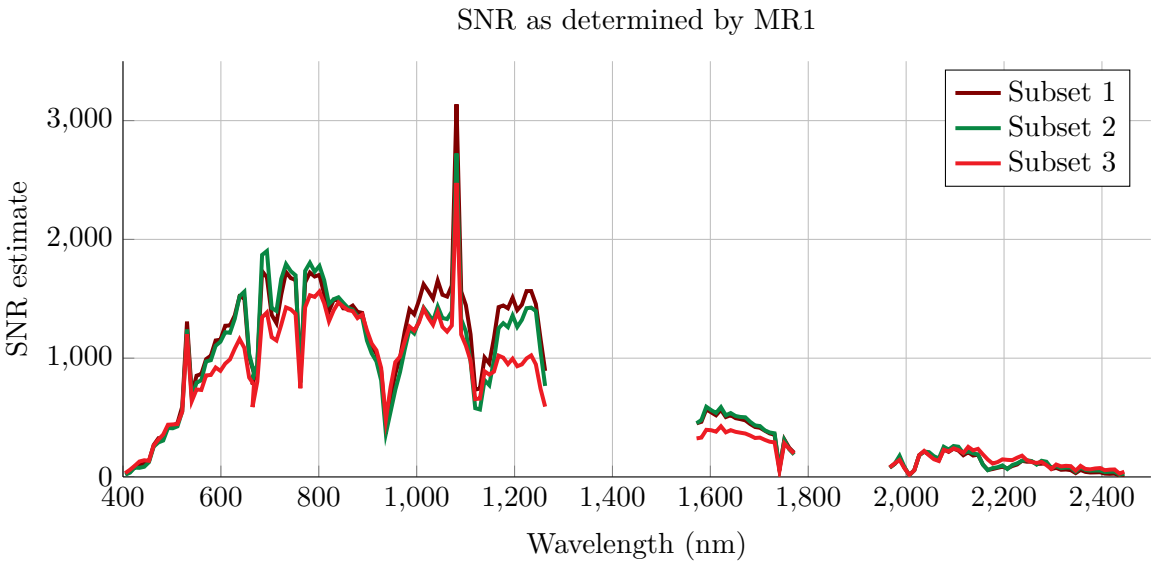


Figure 91: The SNR determined by MR1 for all three subsets. The graphs have been scaled to show that the profiles are all similar to each other and are comparable to the profile measured in figure 90.

7 Conclusion

We evaluated the performance of various noise statistic estimation methods and noise removal methods on an an artificial dataset. We added noise to the artificial data according to an additive noise model with different spectral correlations. We then used our observations to tackle a real dataset (AVIRIS Cuprite radiance). We performed our analysis on three subsets of the data to assess the consistency of the methods and therefore the accuracy of our estimations.

Our observations on the artificial data have produced a nice tidy scheme for dealing with hyperspectral imagery. Under appropriate conditions, the modified regression method proved superior in estimating the statistics of a hyperspectral image. It was found that an overestimate of the correlation extent of the noise produces a better estimate than an underestimate. This leads us to suggest that modified regression should be performed for multiple correlation estimates and the results compared to best estimate the noise statistics (based on the differences between successive MRn). Following estimation of σ^2, the BM4D method for denoising produced the best results. Again an overestimate for the input parameter (σ^2) produces a better result than an underestimate. If execution time is a concern, the PCABM3D method can be used instead of BM4D and still produce decent results.

Further research is suggested for the modified regression method. The modified regression method should be adapted to consider block matrix type correlation to avoid errors incurred by not considering significant correlations further away from the main diagonal than n, instead of the inclusion of every possibly insignificant term closer to the main diagonal than n. In these tests an iterative approach should be considered, with a first estimate of the correlation structure retrieved from simple diagonal matrix MRn and using these estimates to refine the guess for the correlation structure.

Additional research is also suggested on the determination of the SNR ratio in the AVIRIS instrument. The SNR should be measured in different scenes (not just different subsets of the same scene), and scaled to [0:1] without a contrast stretch.

Since we noted that in the case of our artificial data set the BM4D method proved superior, investigations are suggested on a greater set of test images that have different spatial properties (edges, homogeneous areas, more random surfaces, more spatial clutter etc.) to find out whether its performance is uniform across image types and, if it is not, how sensitive it is to this sort of variation. The method may perform poorly if the data is very smooth with few edges or if there are too many edges and the image is more cluttered.

Our observations on the AVIRIS radiance data leads us to make the following conclusions about the sensor. The AVIRIS data is very good before denoising has even been performed (with noise of the order 10^{-6}). Looking at the noise correlation matrix estimates we conclude that the structure of the correlation is that of a block matrix. The noise is also seen to be worse in the second pair of spectrometers which correspond to low energy wavelengths.

Observations on the AVIRIS radiance data suggest that we need to further test all the methods on a greater range of artificial images. Images that vary in intensity should be investigated in particular to better understand the observations that were made on the three subsets (the brightest subset demonstrated the lowest SNR). We need to be able to say something about the methods' sensitivities to variance in the signal itself, so as a starting point we suggest images with different signal variances and additive noise to create the same signal to noise ratio in all images. Secondly the set of images should be tested with the same noise variance (and consequently different signal to noise ratios). The results of these two experiments should provide insight into method sensitivities to signal variance.

Bibliography

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," in *Electronic Imaging 2006*, vol. 6064, pp. 606414—-606414, International Society for Optics and Photonics, 2006.

[3] R. Green and M. Eastwood, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 4257, no. 98, pp. 227—-248, 1998.

[4] D. Manolakis, R. Lockwood, and T. Cooley, "On The Spectral Correlation Structure of Hyperspectral Imaging Data," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS*, vol. 2, pp. II–581–II–584, July 2008.

[5] V. Mulder, S. de Bruin, M. Schaepman, and T. Mayr, "The use of remote sensing in soil and terrain mapping - A review," *Geoderma*, vol. 162, pp. 1–19, Apr. 2011.

[6] Y. Li, J. Shan, Y. Peng, and X. Gao, "Nondestructive assessment of beef-marbling grade using hyperspectral imaging technology," *New Technology of Agricultural Engineering (ICAE)*, pp. 779—-783, 2011.

[7] J. Foster, "Predicting tropical forest carbon from EO-1 hyperspectral imagery in Noel Kempff Mercado National Park, Bolivia," *Geoscience and Remote Sensing Symposium*, vol. 00, no. Figure 1, pp. 3108–3110, 2002.

[8] D. G. Goodenough, T. Han, J. S. Pearlman, A. Dyk, and S. Mcdonald, "Forest Chemistry Mapping with Hyperspectral Data," *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*, pp. 395–398, 2003.

[9] F. F. Sabins, "Remote sensing for mineral exploration," *Ore Geology Reviews*, vol. 14, pp. 157–183, Sept. 1999.

[10] F. A. Kruse, "Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra," *International Journal of Remote Sensing*, vol. 17, no. 9, pp. 1623–1632, 1996.

[11] M. Farzam and S. Beheshti, "Information theoretic assessment of correlated noise in hyperspectral signal unmixing," in *Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on*, pp. 938–942, May 2011.

[12] K. Cawse-Nicholson, A. Robin, and M. Sears, "Evaluation of bands containing spectrally correlated noise in hyperspectral imagery," in *IEEE Whispers*, 2013.

[13] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, P. Marcoionni, I. Pippi, and M. Selva, "Noise modelling and estimation of hyperspectral data from airborne imaging spectrometers," *Annals of geophysics*, vol. 49, no. February, 2006.

[14] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Local signal-dependent noise variance estimation from hyperspectral textural images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 469–486, 2011.

[15] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, pp. 354–379, Apr. 2012.

[16] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, pp. 2435–2445, Aug. 2008.

[17] K. Cawse, A. Robin, and M. Sears, "The effect of noise whitening on methods for determining the intrinsic dimension of a hyperspectral image," *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, 2011.

[18] K. Cawse-Nicholson, A. Robin, and M. Sears, "The Effect of Correlation on Determining the Intrinsic Dimension of a Hyperspectral Image," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 6, pp. 482–487, Apr. 2013.

[19] L. Gao, Q. Du, B. Zhang, W. Yang, and Y. Wu, "A comparative study on linear regression-based noise estimation for hyperspectral imagery," *Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 488 – 498, 2013.

[20] L. Gómez-Chova, L. Alonso, L. Guanter, G. Camps-Valls, J. Calpe, and J. Moreno, "Modelling spatial and spectral systematic noise patterns on CHRIS/PROBA hyperspectral data," in *Proc. SPIE*, vol. 6365, p. 63650Z, 2006.

[21] J. S. Lee and K. Hoppel, "Noise modeling and estimation of remotely-sensed images," in *Geoscience and Remote Sensing Symposium*, vol. 2, pp. 1005–1008, IEEE, 1989.

[22] N. Acito, M. Diani, and G. Corsini, "Signal-Dependent Noise Modeling and Model Parameter Estimation in Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 2957–2971, Aug. 2011.

[23] L. Alparone, M. Selva, L. Capobianco, S. Moretti, L. Chiarantini, and F. Butera, "Quality Assessment of Data Products from a New Generation Airborne Imaging Spectrometer," *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, vol. 2, no. 2, pp. 422–425, 2009.

[24] R. J. Harris, *A primer of multivariate statistics*. Lawrence Erlbaum Associates Publishers, 2001.

[25] M. Farzam and S. Beheshti, "Robust hyperspectral signal unmixing in the presence of correlated noise," *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 1361–1364, 2011.

[26] P. Curran and J. Dungan, "Estimation of signal-to-noise: a new procedure applied to AVIRIS data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 21, no. 5, 1989.

[27] P. Meer, J. Jolion, and A. Rosenfeld, "A fast parallel algorithm for blind estimation of noise variance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, pp. 216–223, Feb. 1990.

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

[29] A. Mahmood, A. Robin, and M. Sears, "Estimation of Correlated Noise in Hyperspectral Images," *Available: https://drive.google.com/file/d/0B3Q87Z1L4b7QcHRlcG1yM2Y2eUk/view?usp=sharing*, 2014.

[30] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "A Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 119–133, 2013.

[31] M. Maggioni and A. Foi, "Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation," in *IS&T/SPIE Electronic Imaging*, no. 213462, pp. 829600—-829600, International Society for Optics and Photonics, 2012.

[32] G. Chen, S.-E. Qian, and S. Gleason, "Denoising of hyperspectral imagery by combining PCA with block-matching 3-D filtering," *Canadian Journal of Remote Sensing*, vol. 37, no. 6, pp. 590–595, 2012.

[33] G. Chen, T. D. Bui, K. G. Quach, and S.-E. Qian, "Denoising Hyperspectral Imagery Using Principal Component Analysis and Block-Matching 4D Filtering," *Canadian Journal of Remote Sensing*, vol. 40, pp. 60–66, June 2014.

[34] G. Chen and S.-E. Qian, "Denoising of Hyperspectral Imagery Using Principal Component Analysis and Wavelet Shrinkage," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, pp. 973–980, Mar. 2011.

[35] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pp. 266–275, International Society for Optics and Photonics, 1999.

[36] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.