The effect of dimensionality reduction on signature-based target detection for hyperspectral remote sensing

Sivert Bakken, Milica Orlandic, Tor Arne Johansen

Sivert Bakken, Milica Orlandic, Tor Arne Johansen, "The effect of dimensionality reduction on signature-based target detection for hyperspectral remote sensing," Proc. SPIE 11131, CubeSats and SmallSats for Remote Sensing III, 111310L (30 August 2019); doi: 10.1117/12.2529141

SPIE.

Event: SPIE Optical Engineering + Applications, 2019, San Diego, California, United States

The Effect of Dimensionality Reduction on Signature Based Target Detection for Hyperspectral Remote Sensing

Sivert Bakken[a], Milica Orlandic[b], and Tor Arne Johansen[a]

[a]Centre for Autonomous Marine Operations and Systems (NTNU-AMOS), Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[b]Department of Electronic Systems, Norwegian University of Science and Technology, 7491 Trondheim, Norway

ABSTRACT

Target detection is one of the more popular applications of hyperspectral remote sensing. To enhance the detection rate, it is common to do preprocessing to reduce the effects of noise and other forms of undesired interference with the observed spectral signatures. In current earth observing systems, in particular small satellite systems, data rate limitations can make the utilization of sensors with high spectral dimensionality undesirable and even unobtainable.

In this paper, the effect of different methods for dimensionality reduction and noise removal has been observed on multiple classical methods for signature matched target detection often used in hyperspectral imaging. The dimensionality reduction differs from resampling in the sense that the original spectral range and resolution can be restored via a linear transformation.

This paper suggests that by combining dimensionality reduction and target detection, the resulting data cube has a reduced dimensionality and suppressed undesired effects. The ability to correctly detect spectral phenomena has improved while also achieving reduce data volume. Combining dimensionality reduction and target detection can also reduce the number of computational operations needed in later stages of processing, when operating on the projected space. The observed effects are demonstrated by using simulated and real-world hyperspectral scenes. The real-world scenes are from well-calibrated sensors e.g. AVIRIS, ROSIS, and Hyperion, of classified agricultural and urban areas. The simulated scene is generated using the ASTER library.

1. INTRODUCTION

Recent advances in remote sensing enable hyperspectral data to be collected with systems that use considerably less physical volume and weight than more traditional remote sensing systems.[1,2] This new generation of electro-optic remote sensing systems is not only significantly smaller, but also able to achieve the same or better performance as traditional systems. Constellations of high-quality remote sensing satellites has become more attainable when combined with the advances of smallsat platforms,[3] i.e. satellite platforms that are of low mass and size. In this way, it has become possible to achieve a higher level of robustness for current and future remote sensing earth observing programs, all at a lower economic and temporal cost.[2,3]

One of the major challenges for such smallsat remote sensing systems is the limitations in communication data rate capabilities, especially so for hyperspectral remote sensing systems.[3] In this paper, the effect of dimensionality reduction as a method for compression, and in turn on target detection is investigated. The effectiveness of lossy compression algorithms has been evaluated by comparing the restored scene with the original scene.[4] This is limited in the sense that the original scene cannot be void of noise or other unwanted artifacts from the sensor, which does not give a definitive answer to whether the lossy compression algorithms reduce noise or degrades the information in the signal. To answer this question, this paper investigates at the effect of dimensionality reduction on signature based target detection for hyperspectral remote sensing, and uses the results as a basis for discussion. This approach does not quantify the ability to restore the original

Further author information: (Send correspondence to Sivert Bakken)
Sivert Bakken: E-mail: sivert.bakken@ntnu.no, Telephone: +47 41 61 45 65

CubeSats and SmallSats for Remote Sensing III, edited by Thomas S. Pagano, Charles D. Norton, Sachidananda R. Babu, Proc. of SPIE Vol. 11131, 111310L
 · doi: 10.1117/12.2529141

data, but rather quantify the effect of the lossy compression on signature based target detection, both through the F_1-score[5] for target detection performance and visibility for robustness.[6]

The data shows that a high compression ratio can be achieved by exploiting structures found in the first and second statistical moments. Given the application of target detection, the projected subspace of the data provides for most cases an improved performance with respect to both the F_1-Score and the visibility. As the data can be analyzed in its compressed state, this can in turn significantly reduce the number of operations required to perform further onboard processing in a smallsat system, whilst at the same time having reduced the total data volume that needs to be stored and transmitted.

1.1 Notation

All hyperspectral images are inherently three dimensional cubes with two spatial dimensions and one spectral dimension. However, in all the subsequent sections a two dimensional matrix representation of the hyperspectral data is used, and the relationship is given in equation (1) and (2), where each $\mathbf{x}_{i,j}$, and subsequently $\mathbf{x}_i$, is a vector of the spectra with a sub-script re-indexation from equation (2) to equation (3). Here the original data set, prior to any dimensionality reduction, is given as $\mathbf{X}_o$

$$\mathbf{X}_{Cube} = \begin{bmatrix} \mathbf{x_{1,1}} & \mathbf{x_{1,2}} & \cdots & \mathbf{x_{1,i}} \\ \mathbf{x_{2,1}} & \mathbf{x_{2,2}} & \cdots & \mathbf{x_{2,i}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x_{j,1}} & \mathbf{x_{j,2}} & \cdots & \mathbf{x_{j,i}} \end{bmatrix} \tag{1}$$

$$\mathbf{X}_{Matrix} = \begin{bmatrix} \mathbf{x_{1,1}} & \cdots & \mathbf{x_{1,i}} & \mathbf{x_{2,1}} & \cdots & \mathbf{x_{2,i}} & \cdots & \mathbf{x_{j,1}} \cdots & \mathbf{x_{j,i}} \end{bmatrix} \tag{2}$$

$$\mathbf{X}_o = \begin{bmatrix} \mathbf{x}_1, & \mathbf{x}_2, & \mathbf{x}_3, & \cdots, & \mathbf{x}_i, & \cdots, & \mathbf{x}_N \end{bmatrix} \tag{3}$$

2. THEORY

This section briefly describes the methods and algorithms used in the analysis. In-depth description of the theory applied can be found in the references.[7]

2.1 Dimensionality Reduction

In classification or target detection applications, high dimensional data such as a hyperspectral image can be used. The high-dimensional data contains a higher number of observed variables than potential classes to classify or targets to detect in the image. This is a source of variation and redundancy in hyperspectral images. Some of the information in the image is thus redundant for analysis, due to this strong correlation. Dimensionality reduction is a common approach to handle the high dimensionality in hyperspectral imaging. In the analysis that follows, four classic methods for dimensionality reduction are explored and analyzed in the context of target detection.

2.1.1 Principal Component Analysis

Principal Component Analysis (PCA) is a method to reduce the dimensionality by utilizing the correlation of different variables in a set.[8] Within PCA the principal components are ordered according to variance, such that the first components carries more information with respect to the full data space than the later components. If the variables or spectral signal under observation carry additive independent normally distributed noise, PCA is an optimal method for noise filtering.

The eigenvalues σ^2 corresponding to the principal components are computed as follows:

$$\begin{aligned} \mathbf{m} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \\ \Sigma_o &= \frac{1}{N}\mathbf{X}_o\mathbf{X}_o^T - \mathbf{m}\mathbf{m}^T \\ \det(\Sigma_o - \sigma^2\mathbf{I}) &= 0 \end{aligned} \tag{4}$$

where Σ_o is the covariance matrix computed from $\mathbf{X}_o$, $\mathbf{I}$ is an identity matrix with dimensions equal to the number of spectral channels, and σ^2 is an eigenvalue. Then, the corresponding unit eigenvectors V_{PCA} are ordered according to the magnitude of the eigenvalues, such that

$$\Sigma_o \mathbf{V}_{PCA} = \mathbf{V}_{PCA}\mathbf{D} \tag{5}$$

where $\mathbf{D}$ is a diagonal matrix consisting of the eigenvalues, ordered according to the magnitude with the highest eigenvalues as the first element, and $\mathbf{V}_{PCA}$ is a unitary matrix consisting of the eigenvectors to fulfill $\Sigma_o \mathbf{v}_j = \sigma_j^2 \mathbf{v}_j$. From this, the transformed subspace can be represented with the columns of $\mathbf{Z}_{PCA}$ as basis vectors, where the dimensionality reduction is performed by selecting a subset of the eigenvectors related to the largest eigenvalues, in the following way

$$\mathbf{Z}_{PCA} = \mathbf{V}_{PCA}^T \mathbf{X}_o \tag{6}$$

where $\mathbf{X}_o$ is, as defined in equation (3), a vector of a hyperspectral signal or a matrix of the observed data with pixels in each column and spectral information in the rows.

2.1.2 Maximum Noise Factoring

Maximum Noise Factoring (MNF) is a dimensionality reduction method that accounts for the source of the noise. Whereas PCA is only dependent on the variance in the data, MNF sorts the principal components based on the estimated signal to noise ratio. The underlying assumption made in the MNF transform is that the noise is additive, but not normally distributed white noise, expressed as

$$\Sigma_o = \Sigma_s + \Sigma_n \tag{7}$$

where Σ_s is the covariance of the signal and Σ_n is the covariance of the noise. As the MNF seeks to maximize the eigenvalues with respect to Signal to noise ratio and decorrelate the covariance matrix, it can be performed as a two-step PCA. In the analysis two different approaches for noise estimation were used. In equation (8), the noise vector set $\mathbf{X}_n$ is computed by taking the difference between neighboring pixels in a spatial area of uniform spectral content,[9] later denoted as GMNF. In equation (9) the noise vector set $\mathbf{X}_n$ is computed by taking the difference between neighboring pixels in the spectral dimension projected into the space of the spectral band in question,[10] later denoted as BMNF.

$$\mathbf{X}_n = \begin{bmatrix} \mathbf{x}_1 - \mathbf{x}_2, & \mathbf{x}_2 - \mathbf{x}_3, & \cdots, & \mathbf{x}_{N-1} - \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{n}}_1, & \hat{\mathbf{n}}_2, & \cdots, & \hat{\mathbf{n}}_{N-1} \end{bmatrix} \tag{8}$$

$$\begin{aligned} \mathbf{X}_o^T &= \mathbf{Y} \\ \mathbf{Y}_{(-k)} &= \begin{bmatrix} \mathbf{y}_1, & \mathbf{y}_2, & \cdots, & \mathbf{y}_{k-1}, & \mathbf{y}_{k+1}, & \cdots, & \mathbf{y}_L \end{bmatrix} \\ \tilde{\mathbf{n}}_k &= \mathbf{y}_{(k)} - \mathbf{Y}_{(-k)} \left(\mathbf{Y}_{(-k)}^T \mathbf{Y}_{(-k)} \right)^{-1} \mathbf{Y}_{(-k)}^T \mathbf{y}_{(k)} \\ \mathbf{X}_n^T &= \begin{bmatrix} \tilde{\mathbf{n}}_1, & \tilde{\mathbf{n}}_2, & \cdots, & \tilde{\mathbf{n}}_L \end{bmatrix} \quad \forall k \in \{1, 2, \cdots, N\} \end{aligned} \tag{9}$$

L is the number of spectral channels, $\mathbf{X}_n$ is in turn used to compute the noise covariance matrix Σ_n from equation (10). The singular value decomposition is used for numerical stability and reduced computational time for data sets with a higher number of samples than variables.[11] The singular value decomposition is given is this paper with matrices $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, with $\mathbf{S}$ being a diagonal matrix with entries sorted according to the magnitude of the eigenvalue, $\mathbf{V}$ being an orthogonal square matrix with dimensions equal to the number of spectral channels, and $\mathbf{U}$ being an orthogonal square matrix with dimensions equal to the number of pixels. With $\mathbf{X}_o$ as the original data set and $\mathbf{X}_w$ as the noise-whitened data set, where the the noise-whitening transformation with a whitening matrix expressed in equation (10) as $\mathbf{U}_n \mathbf{S}_n^{-1/2}$, yields the whitened random vector $\mathbf{X}_w$ with unit diagonal covariance and zero mean, further expressed as follows

$$\Sigma_n = \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n^T \qquad \mathbf{X}_w = \mathbf{X}_o \mathbf{U}_n \mathbf{S}_n^{-1/2} = \mathbf{U}_w \mathbf{S}_w \mathbf{V}_w^T \tag{10}$$

From this, the transformation matrix $\mathbf{V}_{MNF}$ can be expressed as

$$\mathbf{V}_{MNF} = \mathbf{U}_n \mathbf{S}_n^{-1/2} \mathbf{V}_w. \tag{11}$$

Hence, the transformed subspace can be represented as $\mathbf{Z}_{MNF}$, where the dimensionality reduction is performed by selecting a subset of the singularvector corresponding to the largest singularvalue, in the following way

$$\mathbf{Z}_{MNF} = \mathbf{V}_{MNF}^T \mathbf{X}_0 \tag{12}$$

where $\mathbf{X}$ is a vector or matrix to be projected into the subspace.

2.1.3 Independent Component Analysis

Independent Component Analysis (ICA) attempts to decompose a multivariate signal into independent non-Gaussian signals, i.e. a decomposition that provides statistical independence between the estimated components. Blind source separation, or ICA, of a mixed signal can separate the different signal sources well when the statistical independence assumption is correct. Whilst PCA and MNF are computed based on first and second order statistical moments when estimating the subspace, ICA utilizes higher order statistical moments. The underlying assumption can conversely be stated as follows:

$$\mathbf{Z}_{ICA} = \mathbf{V}_{ICA}^{-1} \mathbf{X}_o \tag{13}$$

where $\mathbf{V}_{ICA}^{-1}$ is the separating matrix with the pre-selected number of components or assumed number of unique spectral signatures, and $\mathbf{Z}_{ICA}$ is the transformed data set with values per pixel corresponding to the computed abundance of a derived spectral signature. In $\mathbf{V}_{ICA}^{-1}$ every row can be regarded as an independent signal, e.g. the spectral signature of a material. The separating matrix $\mathbf{V}_{ICA}^{-1}$ is computed iteratively over the entire data set. In this paper, Joint Approximation Diagonalization of Eigen-matrices (JADE) is used to compute the components or independent signals, due to its convergence properties when compared with other approaches.[12] It should be noted that for computation of the components, as opposed to using a pre-computed projection matrix from a relevant dataset, the JADE ICA approach can be more computationally demanding than other approaches such as FastICA.[12,13] JADE ICA separates observed mixed signals into latent source signals, i.e. underlying structures not directly observed, by exploiting kurtosis, the fourth order statistical moment, computed as shown in equation (14). Kurtosis is a measure of how Gaussian the observed data distribution is, and is used for defining independence between the source signals. Thus, JADE ICA seeks an orthogonal rotation of the observed mixed vectors to estimate source vectors which possess high values of excess kurtosis. Kurtosis is defined as follows:

$$\mathbf{Kurt}[\mathbf{X}] = \mathbf{E}\left[\left(\frac{\mathbf{X} - \mu}{\sigma}\right)^2\right]. \tag{14}$$

2.2 Target Detection

In this paper, four popular classical signature based target detection algorithms are used in the analysis.[7,14] This subset of target detection algorithms exploits a priori information about the desired target. There is an assumption that the target signature is available and has been normalized to fit the scene. These target detection algorithms can also be considered as linear processes that utilizes the information provided by first and second order statistics.[2,7]

More advanced nonlinear target detection methods are able to exploit the statistical moments of a higher order than traditional methods. This may in some cases yield better performance, but it demands a higher level of a priori knowledge of a specific application, and more training data to be able to map the solution space. This makes it difficult to claim whether a given target detection algorithm is superior to other alternatives.[15] The use of classical target detection algorithms may be sufficient for a given application, both in terms of robustness and target detection performance.

2.2.1 Spectral Angle Mapper

Spectral Angle Mapper (SAM) is a detection algorithm based on the following hypotheses:

$$\mathbf{H}_0 : \mathbf{x} = \mathbf{b} \qquad \mathbf{H}_1 : \mathbf{x} = \alpha\mathbf{s} + \mathbf{b} \tag{15}$$

where $\mathbf{x}$ is a vector in the data set, $\mathbf{b}$ is a is the background and clutter noise, α is a parameter accounting for uncertainty in the desired signal strength, and $\mathbf{s}$ is the desired signal. The SAM operator can be expressed as in equation (16) and a derivation of the operator can be found in.[7]

$$r_{SAM}(\mathbf{x}) = \frac{(\mathbf{s}^T\mathbf{x})^2}{(\mathbf{s}^T\mathbf{s})(\mathbf{x}^T\mathbf{x})} \tag{16}$$

2.2.2 Adaptive Cosine Estimator

Adaptive Cosine Estimator (ACE) is a detection algorithm based on the following hypotheses:

$$\mathbf{H}_0 : \mathbf{x} = \mathbf{b} \qquad \mathbf{H}_1 : \mathbf{x} = \alpha\mathbf{s} + \beta\mathbf{b} \tag{17}$$

where $\mathbf{x}$ is a vector in the data set, $\mathbf{b}$ is the background and clutter noise, $\mathbf{s}$ is the desired target signal, and α and β is parameters accounting for uncertainty in the signal strengths. The ACE operator can be expressed as

$$r_{ACE}(\mathbf{x}) = \frac{(\mathbf{s}^T\Sigma_o^{-1}\mathbf{x})^2}{(\mathbf{s}^T\Sigma_o^{-1}\mathbf{s})(\mathbf{x}^T\Sigma_o^{-1}\mathbf{x})}. \tag{18}$$

2.2.3 Constrained Energy Minimization

Constrained Energy Minimization (CEM) is a detection algorithm based on minimization of the projected background energy, where the energy can be expressed as in equation (19) leading to the expression found in equation (20), with the solution found in equation (21). A derivation of the operator can be found in the references.[7]

$$E = \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}^T\mathbf{x}_i\mathbf{x_i}^T\mathbf{h} = \mathbf{h}^T\mathbf{R}_o\mathbf{h} \qquad \mathbf{R}_o = \frac{1}{N}\mathbf{X}_o\mathbf{X}_o^T \tag{19}$$

$$\min_{\mathbf{h}} \mathbf{h}^T\mathbf{R}_o\mathbf{h} \qquad \text{s.t.} \quad \mathbf{h}^T\mathbf{s} = 1 \tag{20}$$

$$r_{CEM}(\mathbf{x}) = \frac{(\mathbf{s}^T\mathbf{R}_o^{-1}\mathbf{x})^2}{(\mathbf{s}^T\mathbf{R}_o^{-1}\mathbf{s})} \tag{21}$$

2.2.4 Orthogonal Subspace Projection

Orthogonal Subspace Projection (OSP) is a detection algorithm based on the hypotheses given in (22).

$$\mathbf{H}_0 : \mathbf{x} = \mathbf{B}\beta + \mathbf{n} \qquad \mathbf{H}_1 : \mathbf{x} = \alpha\mathbf{s} + \mathbf{B}\beta + \mathbf{n} \tag{22}$$

where $\mathbf{x}$ is a vector in the data set, $\mathbf{B}$ represent the background subspace, β is the background basis coefficient vector, $\mathbf{n}$ is independent normally distributed noise with zero mean, α is a parameter accounting for uncertainty in the desired signal strength, and $\mathbf{s}$ is the desired target signal. The OSP operator can be expressed as in equation (23) and a derivation of the operator can be found in the references.[7] In this paper, the N-FINDR endmember extraction algorithm as described in[16] is used to estimate the background subspace $\mathbf{B}$, to separate it from the desired signal. The desired signal is identified as the spectral signature closest to the sought after signature based on SAM and removed to make up the background subspace $\mathbf{B}$.

$$r_{OSP}(\mathbf{x}) = \mathbf{s}^T(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\mathbf{x} = \mathbf{s}^T\mathbf{P}_{\mathbf{B}}^{\perp}\mathbf{x} \tag{23}$$

3. DATA SET DESCRIPTION

An overview of the real-world and simulated data sets used in the analysis is given. All real-world scenes are publicly available.[17] A short description of how the simulated data sets is provided. True color images for all the real world data sets are given in Figure 1, with the image showing the distribution of endmembers.

Gray space represents unclassified pixels, but still used in model creation. In addition, all performance metrics are computed on the full scene. For each scene used in the analysis Table 1 gives sensor, spatial extent of the pixels, the number of classes in each set, the spectral range and resolution of each sensor. Furthermore, a short description of the characteristics of each data set is given, and the different compression ratios with respect to dimensionality reduction are given in Table 2. As training samples for all the dimensionality reduction methods and target detection methods, 30% of the pixels in a given data set were randomly selected as training samples. No manually selected portion of the scene was used to compute $\mathbf{X}_n$, for either of the MNF transforms. The poorer noise estimation in the Pavia data set is due to the in-homogeneity of the scene, compared to the more homogeneous scenes e.g. Salinas.

None of the pixels in the data sets presented here are explicitly mutually exclusive, i.e. there can be more than one material in a given pixel.

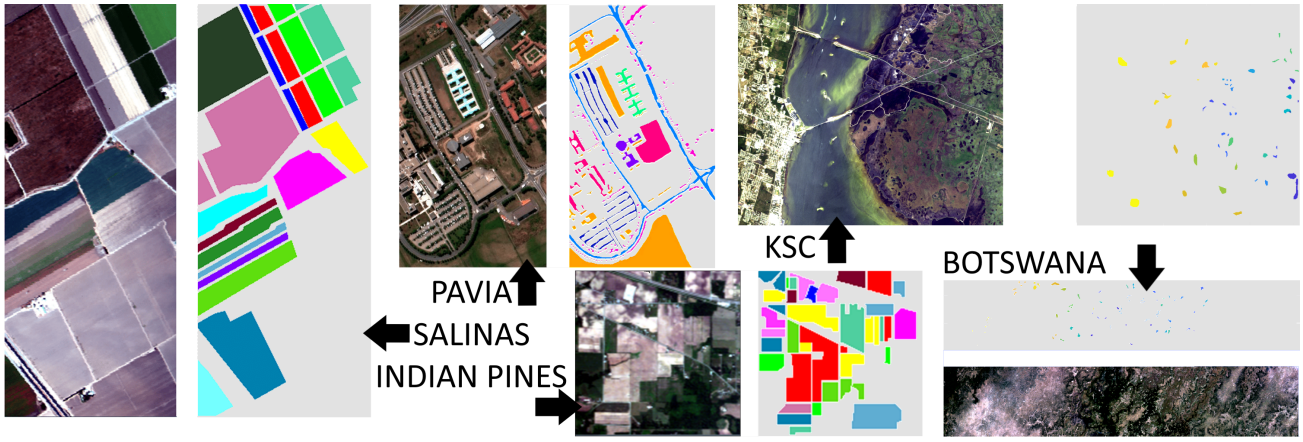

Figure 1. True color images for all the real world data sets The accompanying image shows the distribution of endmembers. Grey areas is unclassified pixels, which is still used in model creation and performance testing.

Data set	Sensor	Image format	Endmembers	Range (nm)	Resolution (nm)	Reference
Indian Pines	AVIRIS	145x145	16	[400,2500]	10	[18,19]
Salinas	AVIRIS	512x217	16	[400,2500]	10	[19]
Pavia	ROSIS-3	610x340	9	[430, 860]	4	[20]
KSC	AVIRIS	512x614	13	[400,2500]	10	[19]
Botswana	Hyperion	1476x256	14	[400, 2500]	10	[21]
Simulated Scenes	ASTER	100x100	10	[400,925]	4	[22]

Table 1. The data sets used in the analysis and important parameters.

3.1 Indian Pines

The Indian Pines scene was gathered by the Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) sensor with 224 spectral reflectance bands. The test site in North-western Indiana and consists of 145 by 145 pixels. The scene is a subset of a larger one and contains two-thirds agriculture and one-third natural perennial vegetation e.g. forest.[17,18]

3.2 Salinas

The Salinas scene was collected by the AVIRIS sensor over Salinas Valley, California. The scene has a spatial resolution of 3.7-meter pixels, which can be considered a high spatial resolution. The area consists of 512 by 217 pixels and includes vegetables, bare soils, and vineyard fields.[17,19]

3.3 Pavia Centre

The Pavia Centre Scene is acquired by Reflective Optics System Imaging Spectrometer (ROSIS) during a flight campaign over Pavia in northern Italy. The scene has a spatial resolution of 1.3-meter pixels, and 102 of the 115 spectral reflectance bands in the wavelength range of 430*nm* to 860*nm* are made available in the data set. The other bands were removed due to high absorption in the atmosphere.[17,20]

3.4 Kennedy Space Center

The Kennedy Space Centre Scene (KSC) is a site in Florida. The scene has a spatial resolution of 18-meter pixels. The publicly available data set has removed some bands to mitigate the effects of bad detectors, calibration errors, and other anomalies. The subset consists of 176 of the 224 bands, and the classified pixels provided represents various land cover types that occur.[17,19]

3.5 Botswana

The Botswana scene is acquired using Hyperion, a payload on the NASA EO-1 satellite. The acquired scene is from the Okavango Delta. The Hyperion sensor on EO-1 acquires had a spatial resolution of 30-meter pixels, and 224 spectral reflectance bands originally. The publicly available data set has removed bands to mitigate the effects of bad detectors, calibration errors, and other anomalies, and the remaining subset consists of 145 of the 224 bands available.[17,21]

Data set \Components	50	45	40	35	30	25	20	15	10	5
Indian Pines	4.48	4.98	5.60	6.40	7.50	8.96	11.20	14.93	22.40	44.80
Salinas	4.48	4.98	5.60	6.40	7.50	8.96	11.20	14.93	22.40	44.80
Pavia	2.06	2.29	2.58	2.94	3.43	4.12	5.15	6.87	10.30	20.60
KSC	3.52	3.91	4.40	5.03	5.87	7.04	8.80	11.73	17.60	35.20
Botswana	2.90	3.22	3.63	4.14	4.83	5.80	7.25	9.67	14.50	29.00
Simulated Scenes	2.40	2.67	3.00	3.43	4.00	4.80	6.00	8.00	12.00	24.00

Table 2. Compression ratios for different number of components after dimensionality reduction. The compression ratio is only dependent on number of components. Computed as CR = spectral channels / components

3.6 Simulated Synthetic Hyperspectral Scenes

The ASTER spectral library[22] was used as a basis for the spectral signatures in the simulated scene. From the spectral library 10 spectral signatures were selected to make up the different synthetically generated images. The spatial scene was created using *Hyperspectral Imagery Synthesis toolbox*,[23] using the Matern covariance function with $\theta_1 = 0.5$ and $\theta_2 = 0.5$, and the spectral signatures to generate a noise free images with a reasonable spatial distribution of the materials. 10 different synthetic scenes were generated and used to get the average performance statistics.

Noise is added to the synthetic scenes to simulate undesired effects and artifacts according to the following relationship[10]

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{n}_i \tag{24}$$

Where $\mathbf{x}_i$ is the spectral signature taken from the ASTER spectral library and $\mathbf{n}_i$ is subject to

$$SNR \equiv 10\log_{10}\frac{\mathbb{E}[\mathbf{x}_i^T\mathbf{x}_i]}{\mathbb{E}[\mathbf{n}_i^T\mathbf{n}_i]} \tag{25}$$

Where the SNR is set to 50 and the variance of the noise signal $\mathbf{n}_i$, as a function of spectral channel i, is given as

$$\sigma_i^2 = \sigma^2\frac{e^{-\frac{(i-L/2)^2}{2\eta^2}}}{\sum_{j=1}^{L} e^{-\frac{(j-L/2)^2}{2\eta^2}}} \tag{26}$$

Where L is the number of spectral channels, which gives a $\sigma^2 \approx 9.73 \times 10^{-07}$ and a σ_i as shown in Figure 9, with $\eta = 18$. Thus the generated noise signal have different characteristics for different wavelengths, making it colored noise rather than white noise.

4. METHODS

The analysis performed in this paper differs from earlier comparisons of dimensionality reduction and target detection performance.[4,6,14,24] Rather than using the receiver operating characteristic (ROC) curve[5] to determine the performance, a combination of the F_1-score[5] and visibility[6] is used.

The original data sets[17] has not been processed prior to dimensionality reduction or target detection. Earlier publications[19,25] have stated that it is not necessary to calibrate scenes captured with AVIRIS to reflectance before applying hyperspectral processing methods, due to the high SNR of the sensor and a relatively flat quantum efficiency function i.e. a similar sensitivity over the captured electromagnetic spectrum. Following this approach, the combination of DR and TD presented in this paper are trained on radiance spectra and operate directly on the measured radiance for each pixel in a given scene or data set. The same approach is used for the scenes acquired from Hyperion and ROSIS, and for the simulated scene. The scope of this paper does not include the effects of atmospheric compensation, as this is not a part of the discussed level of data processing.

In the analysis of the real-world scenes, the dominating endmember of a pixel is set as ground truth. Thus it is not accounted for pixels that were not mutually exclusive i.e. consisting of more than a single endmember. This ground truth is used as a true positive for the signature sought after, whereas other pixels are marked as a true negative. The mixture relationship information of each pixel was not available from the real-world data sets[17] used in the analysis. As a result, only one endmember is associated with a single pixel.

4.1 Performance Metrics

In this paper, the effects of dimensionality reduction on signature based target detection for hyperspectral imaging are investigated. Traditionally, SNR and ROC have been used as a performance metrics in dimensionality reduction and target detection, respectively. This is a suitable approach when investigating a specific case.

The ROC is suitable to determine the optimal threshold for a specific application, but the different ROC curves for different endmembers are not suitable when looking at the average performance for several endmembers in a given scene, due to the curve not necessarily following the same trajectory for comparing cases. ROC is ill-suited when the number of cases grows large, and the analysis given here investigate more than 2000 such cases.

SNR measures the level of restoration to the original data, which will be contaminated by noise. A lower SNR score can be due to loss of information carrying data or removal of sensor artifacts and other undesired effects, but the exact source of the lower SNR score is often unclear. Some earlier work[4] adds noise to the scene to control that variable, but this does not solve the issue of determining the effect on the original noise.

To circumvent the limitations of SNR and ROC, it is proposed to take into consideration the complete chain, i.e. lossy compressing the data and performing exploitative analysis on the compressed image, in order to determine the performance. The analysis attempts to show how well the combination of dimensionality reduction and target detection performs and how how robust the performance is on the average case for different scenes and cases using the F_1-score and Visibility.

The analysis will determine the robustness and performance on the average case for different scenes using the F_1-score and Visibility. The target detection problem can be regarded as a binary classification problem, so the F_1-score[5] is used as a measure of the performance of a given algorithm. The F_1-score is calculated as

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = 2 \times \frac{\text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}} \tag{27}$$

where precision or positive predictive value (PPV) is the number of true positives divided by the sum of all positives, and recall or true positive rate (TPR) is the number of true positives divided by the sum of the true positives and the false negatives. This relationship is common to use in binary classification, but not as common to use in the hyperspectral remote sensing domain. In the analysis the optimal F_1-score is chosen, i.e. the highest resulting F_1-score for all possible threshold values.

Visibility,[6] a measure of robustness for target detection algorithms, is also computed. A higher separability between target and background gives an easier to define detection threshold, and by extension improved suppression of undesirable false alarms i.e. a higher level of robustness. The evaluation metric visibility used as a measure of separability is given as

$$vis = \frac{|\bar{T}_t - \bar{T}_b|}{max(T) - min(T)}. \tag{28}$$

where $\bar{T}_t$ is the average response value or probability of target for target pixels, $\bar{T}_b$ is the average response value or probability of target for non-target pixels, $max(T)$ and $min(T)$ are the maximum and minimum probability of target in the scene, normalizing the visibility value. The maximum score of visibility is 1, and the lowest score is 0. By including visibility results, the scenarios where the F_1-score is high for a very narrow set of possible thresholds are avoided.

5. RESULTS

Results from the real world data are given in section 5.1, focusing on the aforementioned performance metrics. The simulated data is given in section 5.2 showcases the restoration properties in terms of SNR, and the degradation of computing the DR transformation matrices from a subset of data.

5.1 Results From Real World Data

The subsequent plots give the averages for the F_1-score and visibility, across all available endmembers. Averages for each DR method per scene is given in Figures 4, 5, 6 and 7. Averages across all scenes are given in Figure 2. For all the figures the results for ACE is given in blue, CEM in red, OSP in yellow, and SAM in purple. Figure 3 is a more visual example of how the dimensionality reduction can affect the performance in terms of F_1-score and visibility.

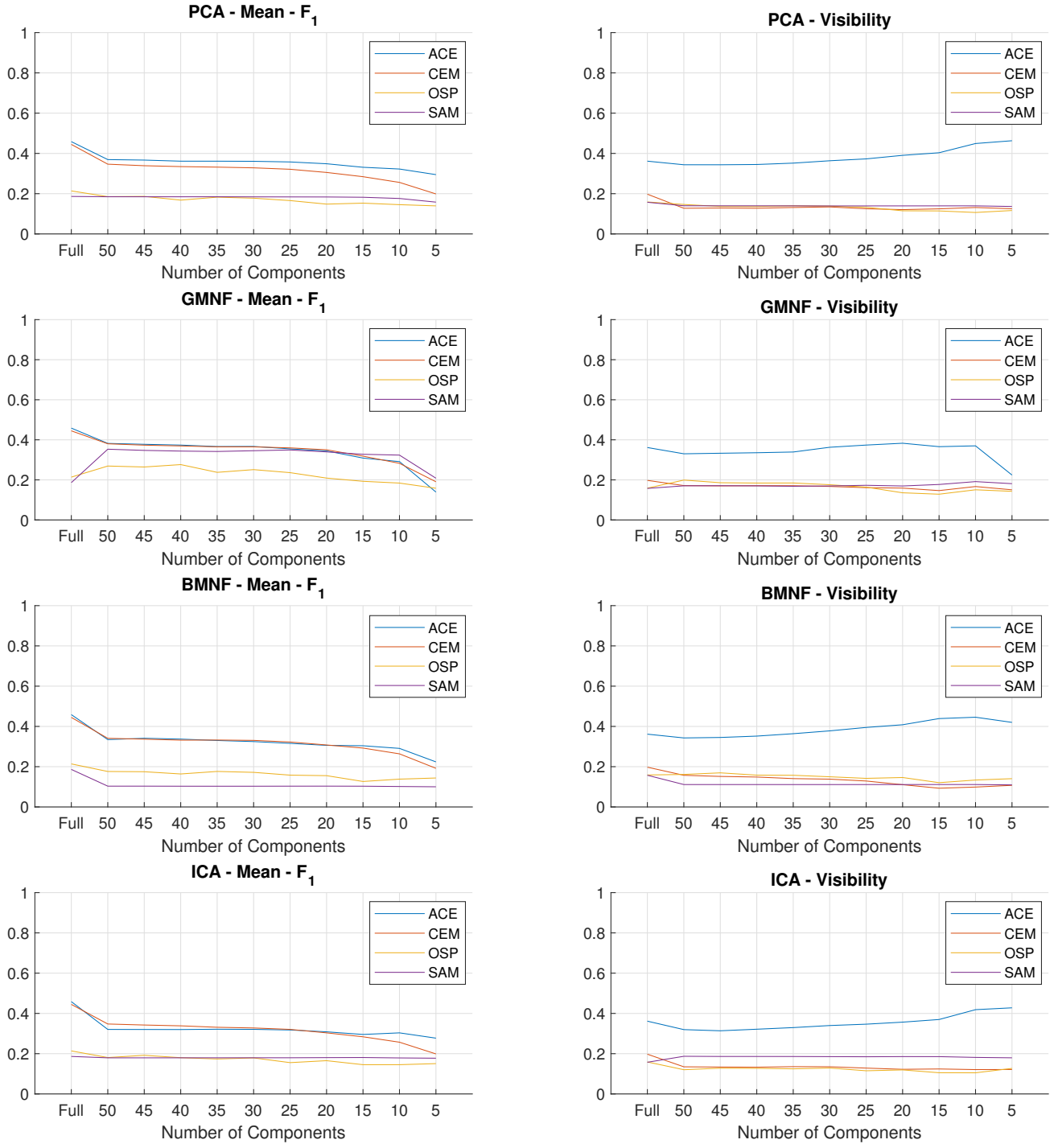

Figure 2. Averages across all endmembers for all scenes for all DR transforms.

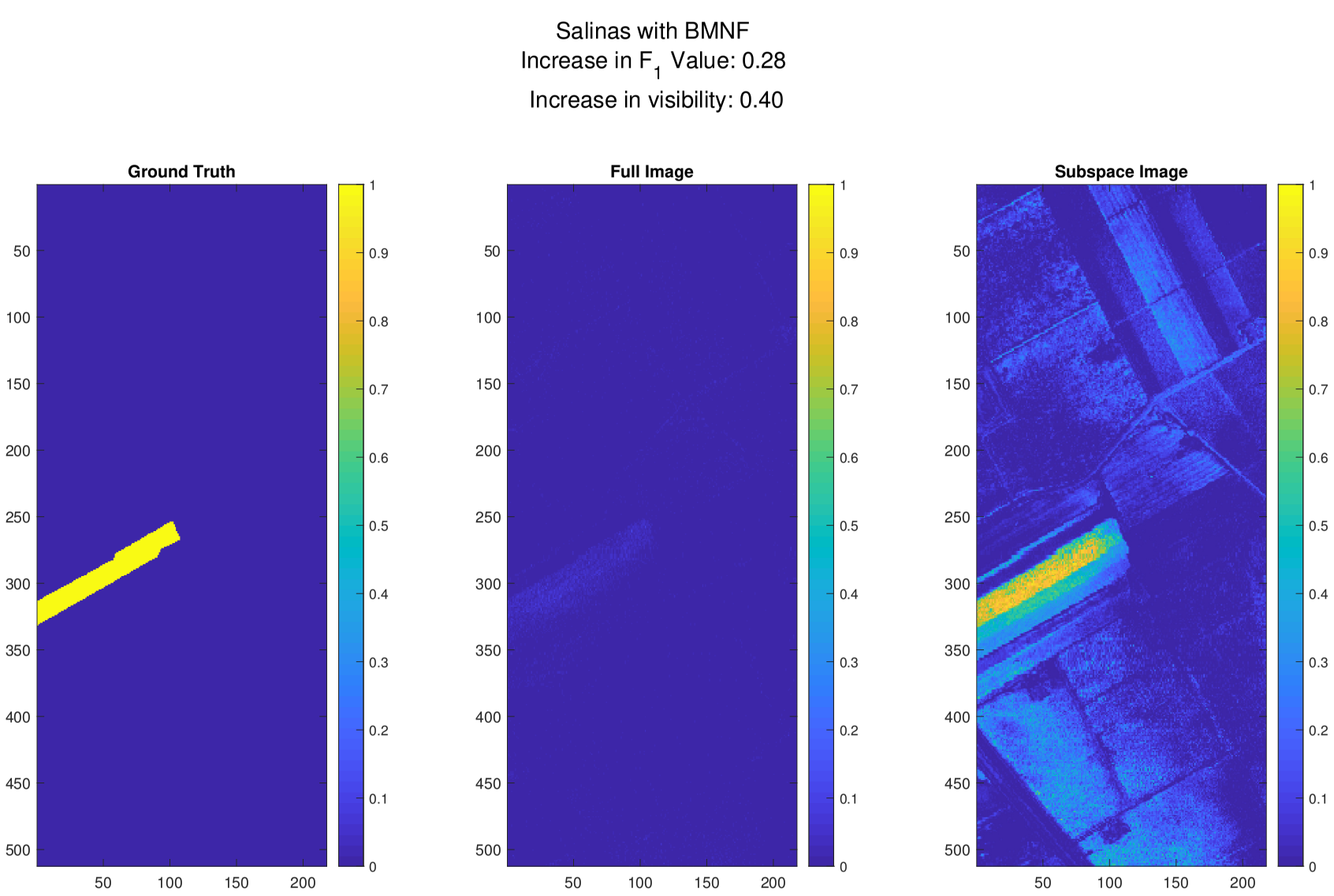

Figure 3. Example of the probability intensities computed by target detection. From left to right, the probability image of the ground truth, ACE target detection on the full spectral range, and ACE target detection on the BMNF subspace using 15 components.

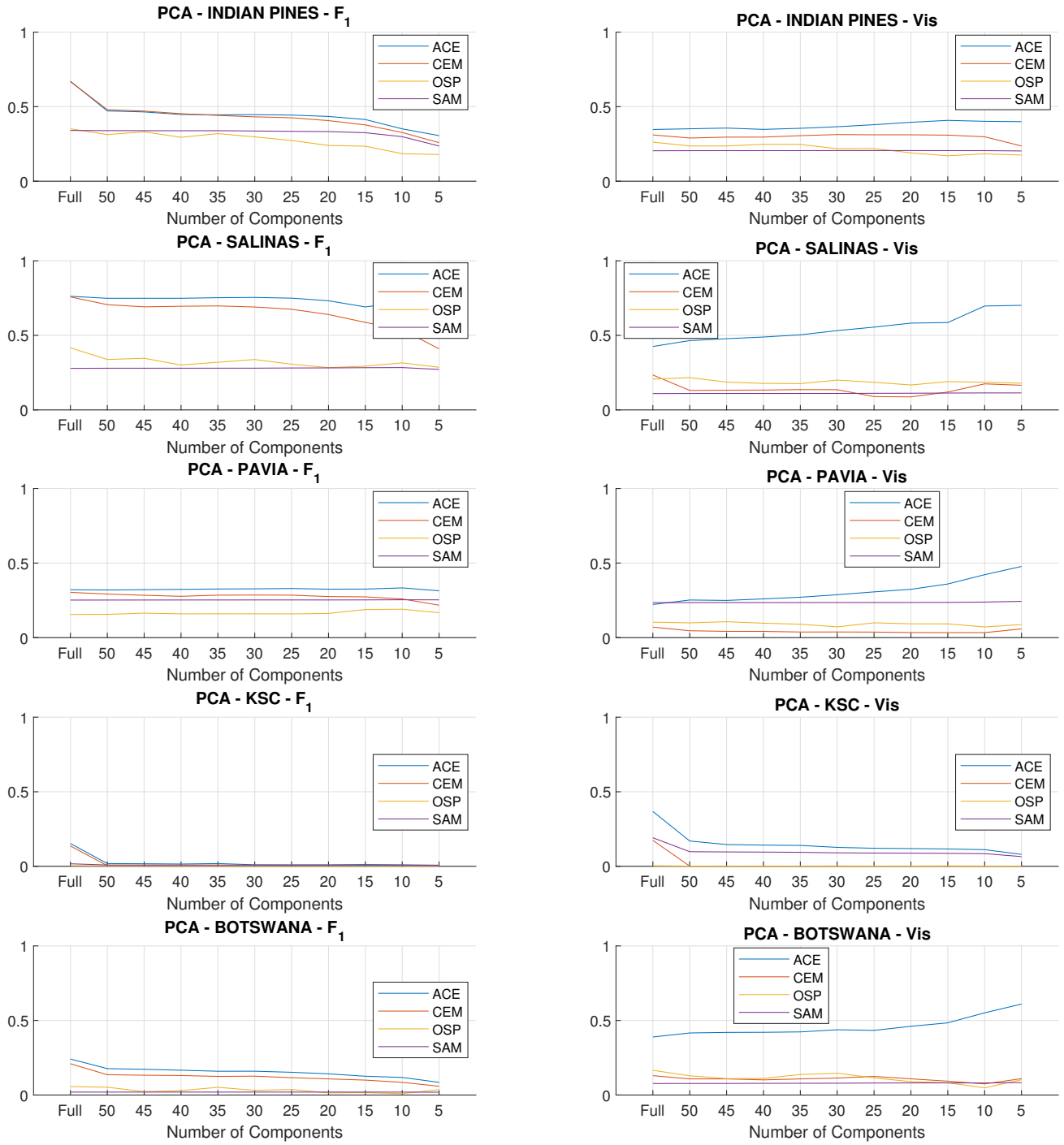

Figure 4. Average performance for all endmembers per scene using the PCA transform.

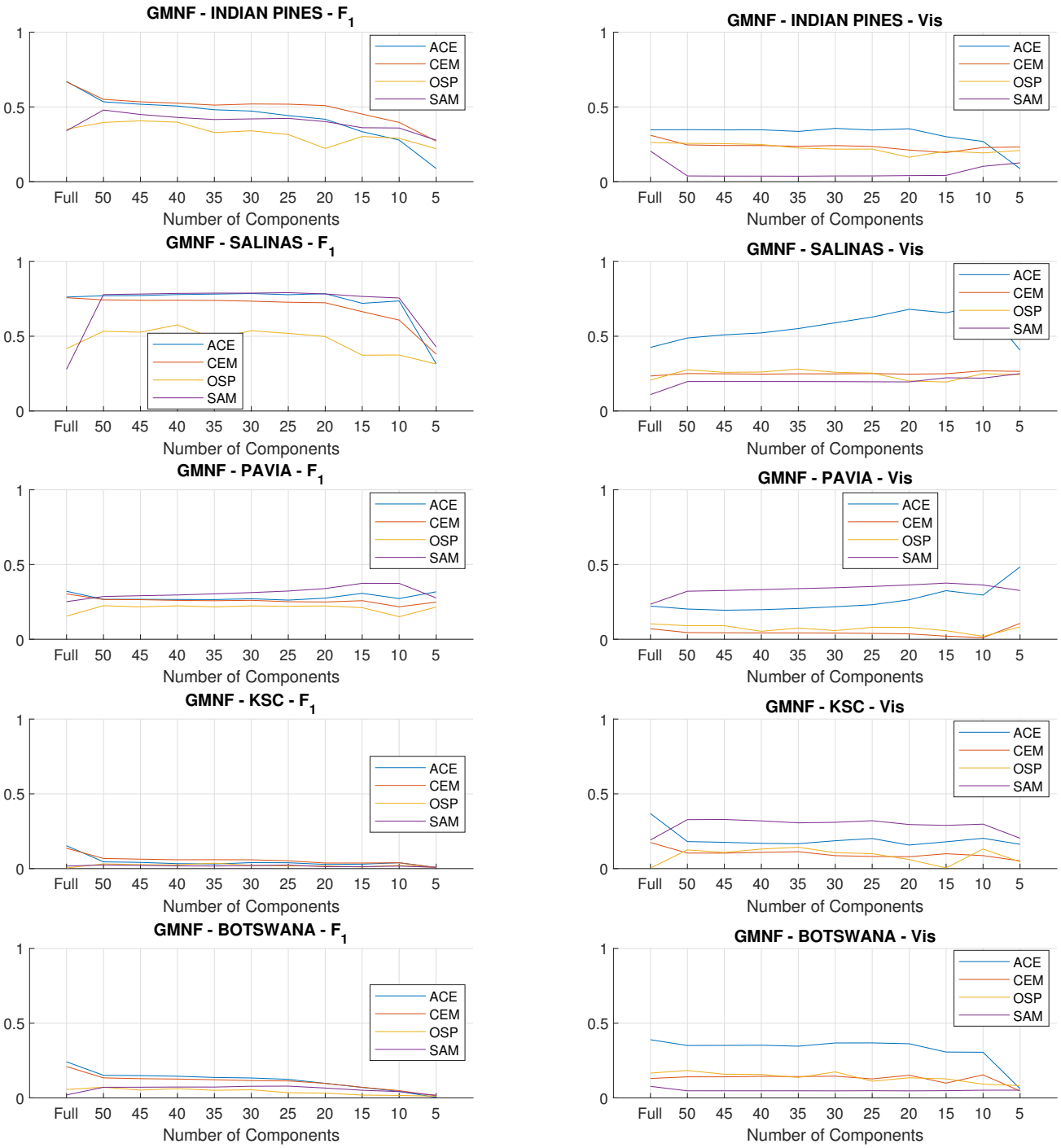

Figure 5. Average performance per scene using the MNF transform with noise estimation given in equation (8).

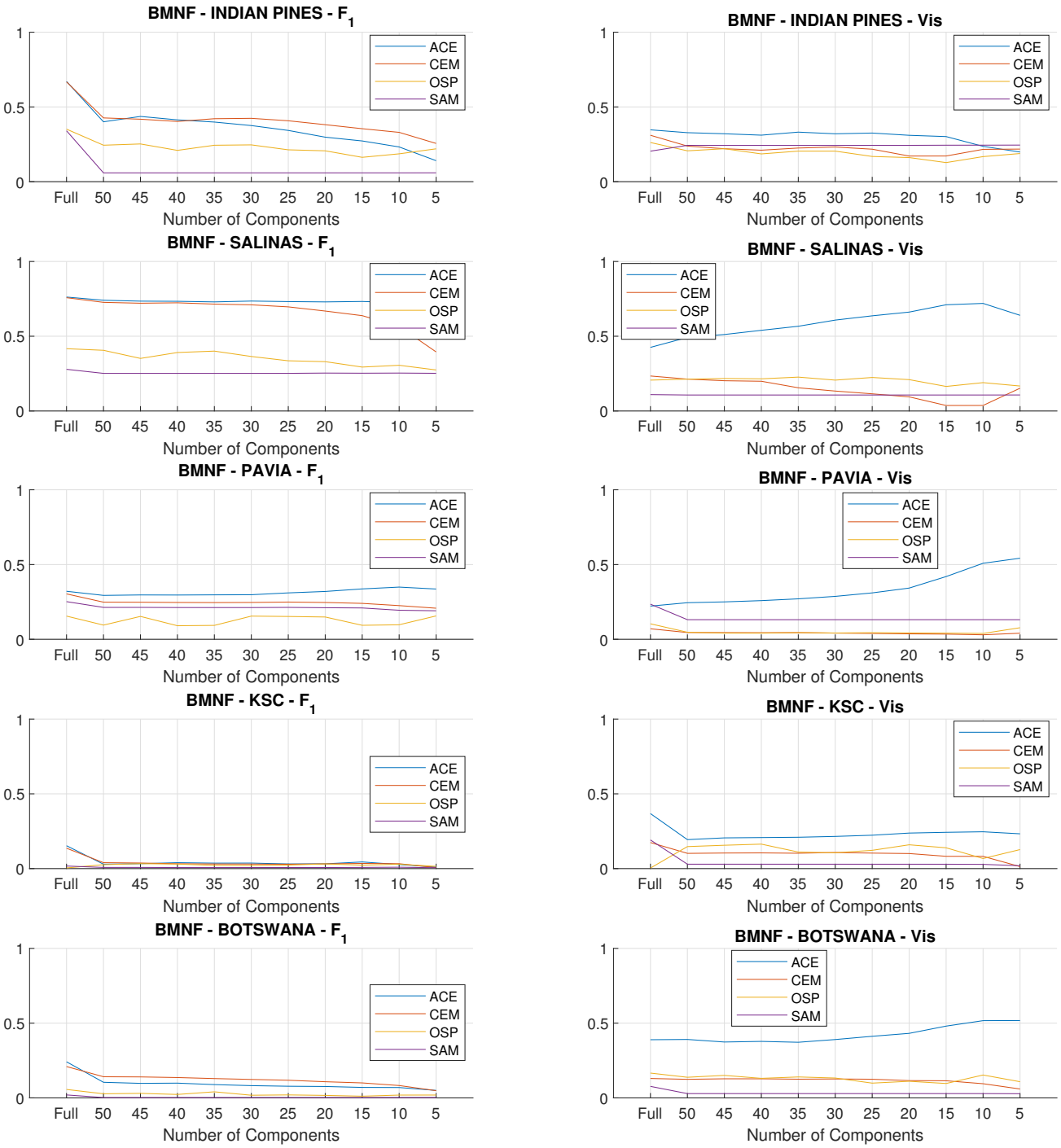

Figure 6. Average performance per scene using the MNF transform with noise estimation given in equation (9).

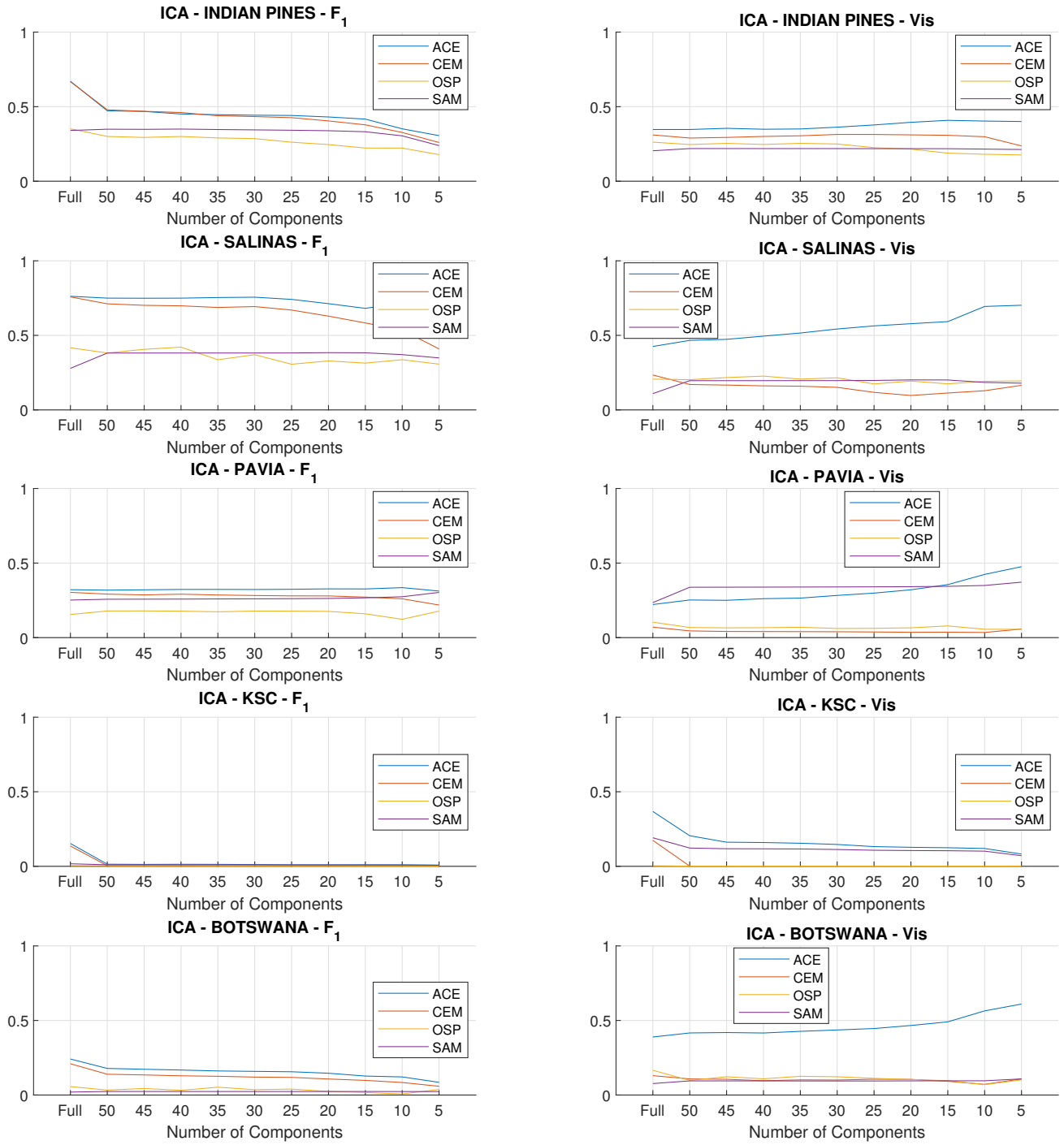

Figure 7. Average performance for all endmembers per scene using the ICA transform.

5.2 Results From Simulated Data

Figure 8 gives the restoration performance in terms of actual SNR of the data before noise is applied. That is, the SNR of the data when compared to the spectral signatures of the synthetically generated image before adding noise and the data re-projected back to the original space via the transformation matrix, with the number of components used in the transformation matrix given along the x-axis. In addition a comparison of the restoration performance on the full data set using components computed using all the available data and 30% randomly sampled pixels is given in Figure 8.

Figure 9 gives the noise characteristics of the noise applied in the simulated data, both in terms of standard deviation as a function of wavelength and a randomly selected noise vector.

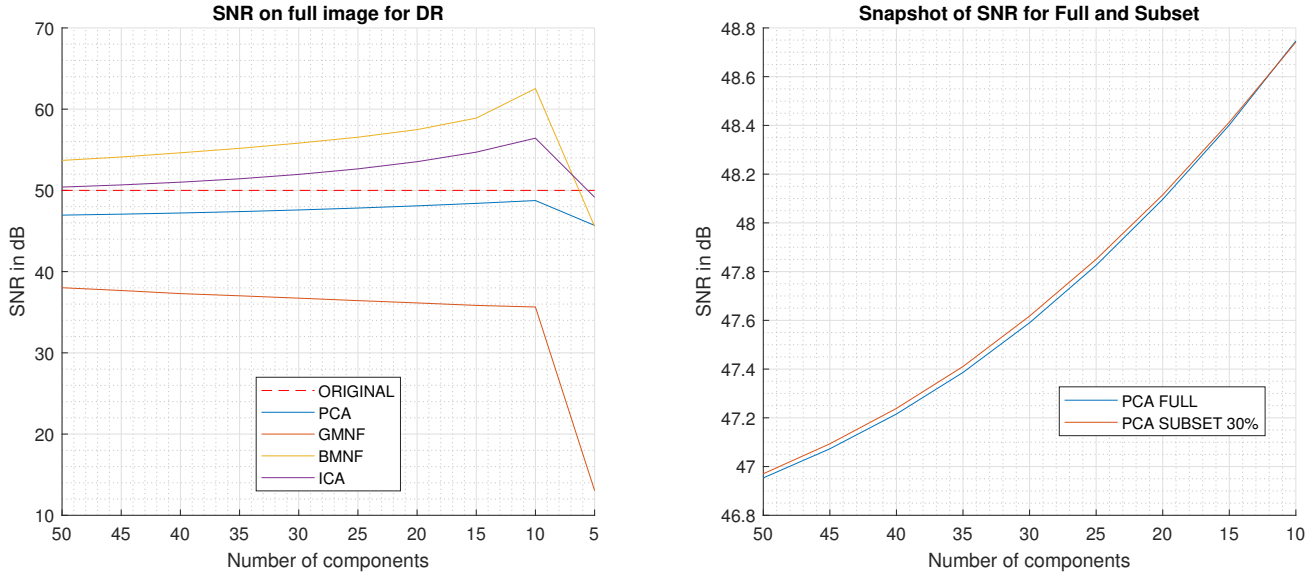

Figure 8. SNR for simulated case with known noise-free signal for all DR methods.

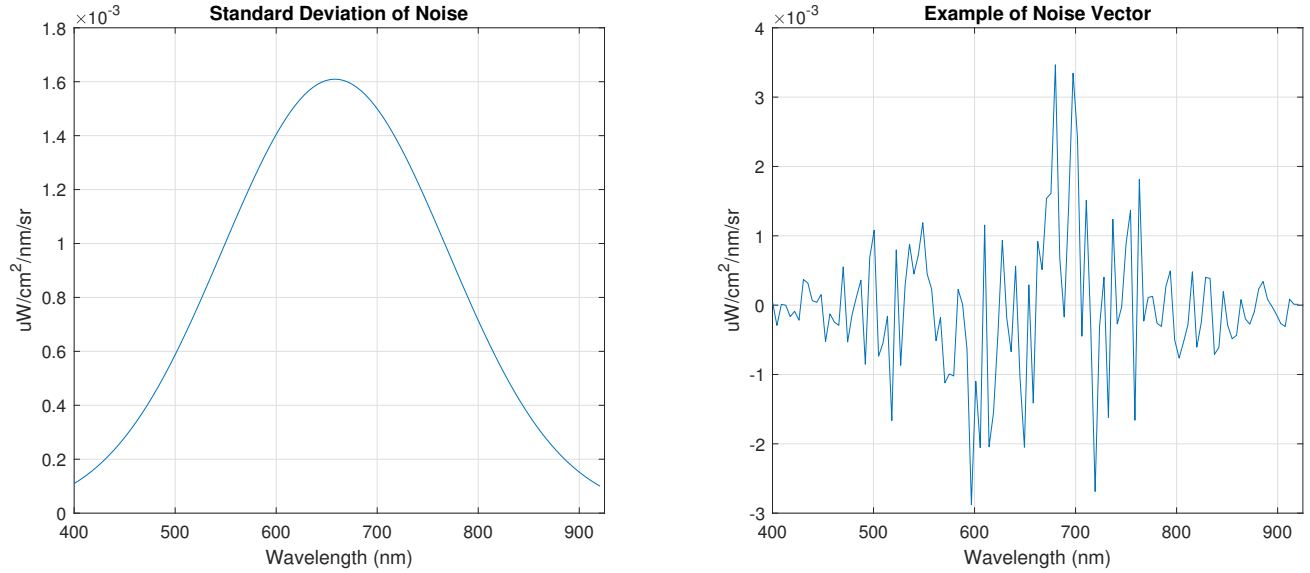

Figure 9. Noise characteristics of simulated data.

6. DISCUSSION AND CONCLUSION

The target detection F_1-score on average tend to be lower when going from the full space to 50 components, as seen in Figure 2 . On average, the number of classes in the real-world data sets is 13, and when accounting for unclassified areas, the virtual dimensionality[10] i.e. the number of unique materials, in the real-world data sets should exceed 10 or 15. In section 5.1 the effect of this can be seen in the decreased F_1-score when using less components than the assumed virtual dimensionality. Reducing the number of components further from 50 does not seem to affect the F_1-score significantly up until the assumed virtual dimensionality is reached.

As seen in Figure 2 the data can be analyzed in the subspace without significant loss in terms of target detection visibility on average, and in some cases the visibility tend to increase with the reduced number of components. The overall visibility performance found for the ACE detector corresponds well with the favorable results found in other publications comparing detection algorithms.[6,14,15,24,26]

Both GMNF and BMNF perform well on the Salinas scene, where the assumption of homogeneous areas are well met, as seen in Figure 5 and 6. When compared to their respective results for the heterogeneous KSC scene, the performance decreases. An example of good case for the BMNF transform is given in Figure 3. This indicates that a more suitable noise model can improve both the detection rate and visibility of a given target or application. An inaccurate noise model for the data can degrade the performance. However, a non-ideal noise estimation does not make the overall performance degrade very significantly. From the average statistics per scene, it is evident that when the assumptions made to create the DR model are erroneous, a more simplistic approach e.g. PCA, seen in Figure 4, is just as suitable if not more suitable than the more complex dimensionality reduction methods in terms of target detection in the subspace. The more complex JADE ICA method, seen in Figure 7, did not yield a favorable performance when compared to the aforementioned PCA or MNF. However, as seen in Figure 9 the ICA and BMNF transform are able to restore the true signal in the synthetic simulated data better than PCA. On average ICA and BMNF is able to remove noise when the number of components used are more than or equal to 10, the virtual dimensionality that the number of classes used in the synthetization would suggest.

The simplicity of implementation should also be taken into consideration. For the dimensionality reduction methods surveyed, the JADE ICA transform is by far the most complex and computationally demanding, whilst the PCA is the simplest. The MNF transforms could be regarded as a double PCA, a whitening step, and derivation of components. The most computationally demanding and complex of the surveyed target detection algorithms is the OSP detector, which gives relatively poor results. The ACE and CEM detector are of similar complexity and give similar results, but the ACE detector tends to be more robust in terms of visibility. Lastly, the SAM detector is the simplest detector and gives especially promising results when combined with the GMNF transform.

The reduced performance seen in Figure 9, shows that there will be a negligible change in restoration performance when computing the DR transformation matrix using only 30 percent of the available data as opposed to all the available data in terms of SNR. Thus, a transformation matrix computed with only 30 percent of the data should still be able to yield a good analysis of the scene, both in terms of restoration to an original noise reduced scene, and in terms of target detection performance in the subspace. As computing the DR transformation matrix can be very computationally expensive, this is a promising result showing data it is not necessarily much to gain form using all the available data. This result has two potential pathways for remote sensing smallsats; if the transformation matrix is computed on-board it is not needed to account for all the available data, and the performance decrease from using a pre-computed transformation matrix can in some cases be negligible i.e. if the a priori assumptions is a good representation of the scene.

All the discussed spectral dimensionality reduction methods give a compression ratio than can be easily computed, as shown in Table 2. The spectral dimensionality reduction can conducted as a preliminary compression step prior to lossless compression methods e.g. JPEG2000, CCSDS123.[27] The information lost in the dimensionality reduction will consist of noise in the statistical sense, as seen in Figure 8, and later lossless compression stages can ensure a lower total data volume for the resulting data cube.

If the DR transformation matrix is computed based on what is expected, rather than the whole scene, this could save both time and power for a given remote sensing smallsat system. The change in performance seen in this analysis when using only 30 percent of the scene to compute the DR transformation matrix is negligible. How well the DR transformation would be able to project a given scene will be dependent on the specific application, but the results given here indicate that the change in performance on average could in many applications be acceptable, and in some cases beneficial.

6.1 Future Work

An adaptive approach to statistically estimate an optimal number of components used for dimensionality reduction[7, 10] could be beneficial to investigate. With such an approach the number of components selected does not need to be dependent on a satellite operators' input, but rather the statistical relationships in the data. It is currently not investigated if the state-of-the-art methods used to estimate an optimal number of components i.e. virtual dimensionality for dimensionality reduction are optimal with respect to the performance metrics given in this paper.

In the work presented here the noise model for the MNF DR methods is computed on a per-pixel basis, thus having the underlying assumption that all pixels will exhibit comparable noise characteristics. For many electro-optical systems this is an erroneous assumption. A better noise model, based on a sensor model and accounting for the different sensitivities of the focal plane array, as opposed to a per scene model or per pixel model, could make the MNF DR transformation matrix further suppress undesired noise or other artifacts from the data.

As it is possible to conduct analysis or data exploration in the subspace, it should be investigated how an application specific analysis is affected by being performed in the subspace. Performing an analysis in the subspace could save the number of operations needed, and thus reducing the computational strain, potentially increasing the throughput as an edge computing agent, such as smallsats for Remote Sensing.

ACKNOWLEDGMENTS

This work was supported by the Norwegian Research Council through the Centre of Autonomous Marine Operations and Systems (NTNU AMOS) (grant no. 223254), the MASSIVE project (grant no. 270959).

REFERENCES

[1] Sigernes, F., Syrjäsuo, M., Storvold, R., Fortuna, J., Grøtte, M. E., and Johansen, T. A., "Do it yourself hyperspectral imager for handheld to airborne operations," *Optics express* **26**(5), 6021–6035 (2018).

[2] Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J., "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine* **1**(2), 6–36 (2013).

[3] Toorian, A., Diaz, K., and Lee, S., "The cubesat approach to space access," in [*2008 IEEE Aerospace Conference*], 1–14, IEEE (2008).

[4] Luo, G., Chen, G., Tian, L., Qin, K., and Qian, S.-E., "Minimum noise fraction versus principal component analysis as a preprocessing step for hyperspectral imagery denoising," *Canadian Journal of Remote Sensing* **42**(2), 106–116 (2016).

[5] Powers, D. M., "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *Journal of Machine Learning Technologies* **2**(1), 37–63 (2011).

[6] Jin, X., Paswaters, S., and Cline, H., "A comparative study of target detection algorithms for hyperspectral imagery," in [*Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*], **7334**, 73341W, International Society for Optics and Photonics (2009).

[7] Eismann, M. T., "Hyperspectral remote sensing," SPIE Bellingham (2012).

[8] Pearson, K., "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**(11), 559–572 (1901).

[9] Green, A. A., Berman, M., Switzer, P., and Craig, M. D., "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on geoscience and remote sensing* **26**(1), 65–74 (1988).

[10] Bioucas-Dias, J. M. and Nascimento, J. M., "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing* **46**(8), 2435–2445 (2008).

[11] Wu, W., Massart, D., and De Jong, S., "The kernel pca algorithms for wide data. part i: theory and algorithms," *Chemometrics and Intelligent Laboratory Systems* **36**(2), 165–172 (1997).

[12] Westad, F. and Kermit, M., "Cross validation and uncertainty estimates in independent component analysis," *Analytica chimica acta* **490**(1-2), 341–354 (2003).

[13] Hyvärinen, A., Karhunen, J., and Oja, E., "Independent component analysis, adaptive and learning systems for signal processing, communications, and control," *John Wiley & Sons, Inc* **1**, 11–14 (2001).

[14] Manolakis, D., Truslow, E., Pieper, M., Cooley, T., and Brueggeman, M., "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Processing Magazine* **31**(1), 24–33 (2014).

[15] Manolakis, D., Lockwood, R., Cooley, T., and Jacobson, J., "Is there a best hyperspectral detection algorithm?," in [*Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery XV*], **7334**, 733402, International Society for Optics and Photonics (2009).

[16] Winter, M. E., "N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in [*Imaging Spectrometry V*], **3753**, 266–276, International Society for Optics and Photonics (1999).

[17] "Hyperspectral remote sensing scenes - grupo de inteligencia computacional (gic)." http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (1 2019). (Accessed on 01/11/2019).

[18] Baumgardner, M. F., Biehl, L. L., and Landgrebe, D. A., "220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3," *Purdue University Research Repository* **10**, R7RX991C (2015).

[19] Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M., et al., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris)," *Remote sensing of environment* **65**(3), 227–248 (1998).

[20] Mueller, A. A., Hausold, A., and Strobl, P., "Hysens-dais/rosis imaging spectrometers at dlr," in [*Remote Sensing for Environmental Monitoring, GIS Applications, and Geology*], **4545**, 225–236, International Society for Optics and Photonics (2002).

[21] Folkman, M. A., Pearlman, J., Liao, L. B., and Jarecke, P. J., "Eo-1/hyperion hyperspectral imager design, development, characterization, and calibration," in [*Hyperspectral Remote Sensing of the Land and Atmosphere*], **4151**, 40–52, International Society for Optics and Photonics (2001).

[22] Baldridge, A., Hook, S., Grove, C., and Rivera, G., "The aster spectral library version 2.0," *Remote Sensing of Environment* **113**(4), 711–715 (2009).

[23] Grupo de Inteligencia Computacional, Universidad del Pas Vasco / Euskal Herriko Unibertsitatea (UPV/EHU), S., "Hyperspectral imagery synthesis (eias) toolbox." http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Imagery_Synthesis_tools_for_MATLAB (2019 07). (Accessed on 07/08/2019).

[24] Manolakis, D., Marden, D., Shaw, G. A., et al., "Hyperspectral image processing for automatic target detection applications," *Lincoln laboratory journal* **14**(1), 79–116 (2003).

[25] Harsanyi, J. C. and Chang, C.-I., "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Transactions on geoscience and remote sensing* **32**(4), 779–785 (1994).

[26] Manolakis, D., Pieper, M., Truslow, E., Cooley, T., Brueggeman, M., and Lipson, S., "The remarkable success of adaptive cosine estimator in hyperspectral target detection," in [*Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*], **8743**, 874302, International Society for Optics and Photonics (2013).

[27] Fjeldtvedt, J., Orlandić, M., and Johansen, T. A., "An efficient real-time fpga implementation of the ccsds-123 compression standard for hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(10), 3841–3852 (2018).