

Peer Reviewed Paper openaccess

A comparison of common factor-based methods for hyperspectral image exploration: principal components analysis, maximum autocorrelation factors (MAF), minimum noise factors (MNF) and maximum difference factors (MDF)

Neal B. Gallagher

Eigenvector Research, Inc., 300 Bella Strada Lane, Manson, WA 98831, USA

ContactN.B. Gallagher: nealg@eigenvector.com<https://orcid.org/0000-0003-3446-2820>

Principal components analysis (PCA), maximum autocorrelation factors (MAF), minimum noise factors (MNF) and maximum difference factors (MDF) models are common factor-based models used for analysis of hyperspectral images. The models can be posed as maximisation problems that result in a symmetric eigenvalue problem (SEP) for each model. The SEPs allow a simple theoretical comparison of the models using a PCA metaphor with MAF, MNF and MDF describable as weighted PCA models. The examples show that the different methods captured different signals in the images that can be examined individually or combined synergistically allowing for additional modelling and extended visualisation. MDF is a factor-based edge detection model that not only allows for additional visualisation but the opportunity to identify and exclude (or highlight) edge signal in the images. An example shows that models can also be used synergistically for finding and elucidating anomalies. In the example, MDF showed the highest sensitivity of the models studied for anomaly detection.

Keywords: maximum autocorrelation factors, minimum noise factors, maximum difference factors

Introduction

Hyperspectral images are multivariate in nature and principal components analysis (PCA),^{1,2} MAF,^{3,4} minimum noise factors (MNF)⁵ and maximum difference factors (MDF)⁶ are complementary multivariate analysis algorithms that

can be used to explore hyperspectral images.^{7,8} The differences between these factor-based approaches can be understood by comparing the different objective functions presented in the Theory section. In each case, the derived

CorrespondenceN.B. Gallagher: nealg@eigenvector.com**Received:** 2 August 2022**Revised:** 6 August 2022**Accepted:** 8 August 2022**Publication:** 16 August 2022**doi:** 10.1255/jsi.2022.a6**ISSN:** 2040-4565**Citation**

N.B. Gallagher, "A comparison of common factor-based methods for hyperspectral image exploration: principal components analysis, maximum autocorrelation factors (MAF), minimum noise factors (MNF) and maximum difference factors", *J. Spectral Imaging* 11, a6 (2022).

<https://doi.org/10.1255/jsi.2022.a6>

© 2022 The Author

This licence permits you to use, share, copy and redistribute the paper in any medium or any format provided that a full citation to the original paper in this journal is given, the use is not for commercial reasons and you make no changes.



model results in a generalised eigenvalue problem (GEP) and a complementary symmetric eigenvalue problem (SEP). From this perspective, the MAF, MNF and MDF models can be described as weighted PCA models. The description is intended to be easily extended by practitioners desiring to define their own modelling objective.

In PCA, factors are obtained that capture maximum signal in the image. These factors can be used to create “scores” images that explore the major signal in the image. In contrast, MAF and MNF finds factors that capture maximum spatially autocorrelated signal in the image. As a result, the MAF and MNF factors focus on variance found in contiguous spatial patches in an image. To accomplish this task, MNF and MDF define less correlated spatial structure using first derivative operators and for the purposes of this paper, MNF uses a first difference and MDF uses a first central difference as described in detail below. In contrast, MDF is a complimentary approach to MAF and finds factors that have maximum spatial difference in the image. MDF can be considered a multivariate edge detection algorithm and MDF scores can be combined with more traditional edge detection algorithms used for greyscale images. Scores from each model are typically inspected separately, but it is shown that the scores can be combined for a synergistic exploratory analysis of hyperspectral images.

The theory section provides a comparison of the mathematical objective for each algorithm and provides information on model derivation. Four example images are used to contrast specific differences in the methods. The first and fourth examples show the broad differences between PCA, MAF, MNF and MDF and how scores from the different methods can be used synergistically for visualisation. The second example also demonstrates method differences but shows distinct differences between MAF and MNF even though they only differ slightly in the algorithm as described in the theory section. The second example also shows how MDF might be used to process an image to exclude edge signal within an image allowing for subsequent modelling to ignore mixed pixels during calibration. The third example shows differences in sensitivity for anomaly detection for PCA, MAF and MDF. MDF showed the highest sensitivity for the image studied. Each example is also discussed in the Supporting Information (SI).

Theory

Hyperspectral images are often described as $M_x \times M_y \times N$ data “cubes” \mathbf{X} that are M_x pixels by M_y pixels measured at N spectral channels (e.g., wavelengths, frequencies or

mass). For ease of nomenclature, and subsequent model descriptions, the image can be matricised to a $M \times N$ data matrix \mathbf{X} where $M = M_x M_y$, and a row of \mathbf{X} is given by the $1 \times N$ vector \mathbf{x}_m for $m = 1, \dots, M$.⁶

In the sub-sections that follow, the maximisation objective will be defined for PCA, MAF, MNF and MDF. In each case, the GEP will be derived. Additionally, a variable transform will provide the complementary SEP. Each step of the derivations is intended to allow a theoretical comparison of the algorithms.

Principal components analysis (PCA)

PCA is ubiquitous to multivariate analysis^{1,2} and is only briefly described here to allow comparison to MAF and MDF. In PCA, it is proposed that scores, \mathbf{t} , are given by $\mathbf{t} = \mathbf{X}\mathbf{p}$ where \mathbf{p} are the principal component loadings. The objective is to maximise $\mathbf{t}^T\mathbf{t}$ for $\mathbf{p}^T\mathbf{p} = 1$. The PCA maximisation objective can be written as

$$O(\mathbf{p}) = \max_{\mathbf{p}} \left(\frac{\mathbf{p}^T \mathbf{X}^T \mathbf{X} \mathbf{p}}{\mathbf{p}^T \mathbf{p}} \right). \quad (1)$$

The next step is to take the derivative with respect to \mathbf{p} and rearrange it to give the GEP given by

$$\mathbf{X}^T \mathbf{X} \mathbf{p} = \lambda \mathbf{p}. \quad (2)$$

(Additional details of the derivation are shown in the next sub-section for MAF.) The GEP is rarely shown for PCA and is provided here to allow for comparisons to MAF and MDF that follow. From Equation 2, it is trivial to obtain the more commonly shown SEP for PCA given by

$$\mathbf{X}^T \mathbf{X} \mathbf{p} = \lambda \mathbf{p}. \quad (3)$$

where \mathbf{p} is the eigenvector and the eigenvalue is given by $\lambda = \mathbf{p}^T \mathbf{X}^T \mathbf{X} \mathbf{p}$.

Some of the properties of the PCA decomposition follow. The rank of \mathbf{X} is less than or equal to the number of pixels, M , or variables, N : $R = \text{rank}(\mathbf{X}) \leq \min(M, N)$. Each principal component (PC), \mathbf{p}_r for $r = 1, \dots, R$, is a “factor” in the PCA model. Typically, the eigenvalues are distinct and the loadings can be collected into the orthogonal matrix $\mathbf{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_R]$ where $\mathbf{P}^T \mathbf{P} = \mathbf{I}_R$, and the scores can be collected into the matrix $\mathbf{T} = [\mathbf{t}_1 \ \mathbf{t}_2 \ \dots \ \mathbf{t}_R]$ where the columns of \mathbf{P} are orthogonal ($[\mathbf{t}_k]^{-1} \mathbf{t}_k^T [\mathbf{t}_r]^{-1} = 1$ for $k=r$ and for $k \neq r$). The columns of \mathbf{P} and \mathbf{T} are orthogonal due to the fact that Equation 3 is a SEP. The orthogonality of \mathbf{P} is discussed in most standard books on linear algebra and provides desirable mathematical and computational properties as well as useful properties for model interpretation.

For hyperspectral images with large numbers of pixels it is expected that, mathematically, R will equal N . However, due to redundancy in the image the latter PCs (those PCs with larger r) will be associated with random noise in the image and PCA models are typically truncated to retain fewer PCs in the model. For a PCA model with K factors with $K < R$, the loadings matrix is given by $\mathbf{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_K]$ and the model of \mathbf{X} is given by

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (4)$$

where the approximation $\hat{\mathbf{X}} = \mathbf{TP}^T$ is rank K and the residuals \mathbf{E} correspond to noise signal (\mathbf{E} is of rank $R - K$). K is often referred to as the chemical rank or pseudo-rank of \mathbf{X} and is typically less than the mathematical rank ($K < R$) but can often be significantly smaller ($K \ll R$). Therefore, PCA can be considered a lossy data compression technique and, due to the partitioning of noise into \mathbf{E} , PCA can also be considered a noise-filtering technique.

Maximum autocorrelation factors (MAF) and minimum noise factors (MNF)

In MAF, it is proposed that scores, \mathbf{t} , are given by $\mathbf{t} = \mathbf{X}\mathbf{w}$ where the weights \mathbf{w} are the MAF factors.³⁻⁵ The objective is to maximise $\mathbf{t}^T\mathbf{t}$ for the maximisation objective given by

$$O(\mathbf{w}) = \max_{\mathbf{w}} \left(\frac{\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w}} \right) \quad (5)$$

where \mathbf{D}_1 is the spatial first derivative operator.^{3,4} For the purposes of this paper, MAF and MNF⁵ are differentiated by defining \mathbf{D}_1 slightly differently: for MAF, \mathbf{D}_1 corresponds to the first central difference operator (see Equation A.1 of Reference 6) and in MNF \mathbf{D}_1 corresponds to the first difference operator (see Equation A.2 of Reference 6), but it is noted that \mathbf{D}_1 can be generalised to other first difference operators. For example, any Savitzky–Golay⁹ first derivative operator might be used. As in PCA, the numerator corresponds to signal in the image but the denominator now replaces \mathbf{I} with $\mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X}$ that corresponds to the first spatial derivative on the image in both the horizontal (left/right) and vertical (up/down) image directions. In practice, the operator must account for edges in the matricised image to avoid end-effects in the calculation of first differences. In the derivation of the MAF model, the derivative of Equation 5 with respect to \mathbf{w} is

$$\frac{dO(\mathbf{w})}{d\mathbf{w}} = \frac{(\mathbf{w}^T \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w}) \mathbf{X}^T \mathbf{X} \mathbf{w} - (\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}) \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w}}{(\mathbf{w}^T \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w})^2} \quad (6)$$

where terms in parentheses on the right-hand side, $(\)$, are scalars. Setting the result in Equation 6 to zero gives

$$(\mathbf{w}^T \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w}) \mathbf{X}^T \mathbf{X} \mathbf{w} - (\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}) \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w} = 0$$

Rearranging provides the GEP for MAF:

$$\mathbf{X}^T \mathbf{X} \mathbf{w} = \lambda \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w} \quad (7)$$

where the eigenvalue λ is given by the Rayleigh quotient

$$\lambda = \frac{\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w}}$$

For simplicity in notation, define $\Sigma = \mathbf{X}^T \mathbf{X}$ and $\Sigma_1 = \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X}$. Equation 7 is then given by

$$\Sigma \mathbf{w} = \lambda \Sigma_1 \mathbf{w} \quad (8)$$

and $\lambda = \mathbf{w}^T \Sigma \mathbf{w} / \mathbf{w}^T \Sigma_1 \mathbf{w}$.

A variable transformation is used to obtain the corresponding SEP. Starting with Equation 8 and defining $\Sigma_1 = \Sigma_1^{0.5} \Sigma_1^{0.5}$, the transformation is given by the following steps

$$\Sigma \Sigma_1^{0.5} \Sigma_1^{0.5} \mathbf{w} = \lambda \Sigma_1^{0.5} \Sigma_1^{0.5} \mathbf{w},$$

$$\Sigma_1^{-0.5} \Sigma \Sigma_1^{-0.5} \Sigma_1^{0.5} \mathbf{w} = \lambda \Sigma_1^{0.5} \mathbf{w}, \text{ and finally}$$

$$\Sigma_1^{-0.5} \Sigma \Sigma_1^{-0.5} \mathbf{p} = \lambda \mathbf{p} \quad (9)$$

where $\mathbf{p} = \Sigma_1^{0.5} \mathbf{w}$ and $\mathbf{w} = \Sigma_1^{-0.5} \mathbf{p}$ and, as is conventional for PCA, $\mathbf{p}^T \mathbf{p} = 1$. Equation 9 can be interpreted as PCA of \mathbf{X} whitened, or de-weighted, by the first spatial derivative, i.e. PCA of the matrix $\mathbf{X} \Sigma_1^{-0.5}$. The eigenvectors of the MAF model can be collected into an orthogonal $N \times K$ matrix $\mathbf{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_K]$ where $\mathbf{P}^T \mathbf{P} = \mathbf{I}_K$. However, $\mathbf{W} = \Sigma_1^{-0.5} \mathbf{P}$ is not generally expected to be orthogonal and, in general, $\mathbf{w}^T \mathbf{w} \neq 1$. The scores $\mathbf{T} = \mathbf{X} \mathbf{W} = \mathbf{X} \Sigma_1^{-0.5} \mathbf{P}$ have orthogonal columns such that $\mathbf{T}^T \mathbf{T}$ is diagonal.

Maximum difference factors (MDF)

In MDF, it is proposed that scores, \mathbf{t} , are given by $\mathbf{t} = \mathbf{D}_1 \mathbf{X} \mathbf{w}$ where the weights \mathbf{w} are the MDF factors.⁶ The objective is to maximise $\mathbf{t}^T \mathbf{t}$ for the maximisation objective given by

$$O(\mathbf{w}) = \max_{\mathbf{w}} \left(\frac{\mathbf{w}^T \mathbf{X}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{X}^T \mathbf{D}_2^T \mathbf{D}_2 \mathbf{X} \mathbf{w}} \right) \quad (10)$$

where \mathbf{D}_1 and \mathbf{D}_2 are the spatial first and second derivative operators, respectively. The MDF signal is the first spatial derivative in the image, the denominator is the second spatial derivative operator. Following a similar

development for MAF given above, the MDF GEP is given by

$$\Sigma_1 \mathbf{w} = \lambda \Sigma_2 \mathbf{w} \quad (11)$$

and the corresponding SEP is given by

$$\Sigma_2^{-0.5} \Sigma_1 \Sigma_2^{-0.5} \mathbf{p} = \lambda \mathbf{p}. \quad (12)$$

Equation 12 can be interpreted as PCA of $\mathbf{D}_1 \mathbf{X}$ de-weighted by variance in the second derivative, i.e. PCA of the matrix $\mathbf{D}_1 \mathbf{X} \Sigma_2^{-0.5}$. The eigenvectors of the MDF model can be collected into an orthogonal $N \times K$ matrix $\mathbf{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_K]$ where $\mathbf{P}^T \mathbf{P} = \mathbf{I}_K$. However, $\mathbf{W} = \Sigma_2^{-0.5} \mathbf{P}$ is not generally expected to be orthogonal and generally $\mathbf{w}^T \mathbf{w} \neq 1$. The scores $\mathbf{T} = \mathbf{D}_1 \mathbf{X} \mathbf{W} = \mathbf{D}_1 \mathbf{X} \Sigma_2^{-0.5} \mathbf{P}$ have orthogonal columns such that $\mathbf{T}^T \mathbf{T}$ is diagonal. A major difference between MDF compared to PCA and MAF is that the scores are calculated vertically (UD) and horizontally (LR) resulting in a doubling of the size of the scores for an image:

$$\mathbf{T}^T = \begin{bmatrix} \mathbf{T}_{UD}^T & \mathbf{T}_{LR}^T \end{bmatrix}$$

Thus, in MDF it makes intuitive sense to examine individual UD and LR scores images to account for lighting effects. However, when the interest is identifying edges it may make more sense to examine the mean Hotelling's T^2 statistic for the two directions. Hotelling's T^2 and Q residuals were calculated according to procedures outlined in Jackson,¹ and Wise and Gallagher.¹⁰

Summary of PCA, MAF/MNF and MDF models

Table 1 summarises the models derived in the preceding sections. In each case, the matricised scores are rearranged back to the image plane for inspection. The table shows a symmetry between the methods and that all three can be interpreted though a PCA metaphor for weighted signal.

Algorithm for MAF and MDF

The QZ algorithm is a direct method often used to solve non-Hermitian GEPs.¹¹ In this work the problems are Hermitian and non-complex, and the algorithm solved the SEPs using the singular value decomposition (SVD). The advantage of this approach is that the different factorisations can be described and compared using the well-understood PCA metaphor with real, orthogonal scores and loadings. However, the inverses in the proposed SEPs may not exist or be ill-conditioned and to account for this

potential problem a regularisation procedure was used (also described in Reference 12). For a symmetric matrix Σ , the SVD is given by

$$\Sigma = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \quad (13)$$

where \mathbf{V} is an orthogonal matrix of eigenvectors ($\mathbf{V}^T \mathbf{V} = \mathbf{I}$) and $\mathbf{\Lambda}$ is a diagonal matrix of non-negative eigenvalues, λ_n for $n = 1, \dots, N$. The regularisation replaces λ_n with d_n given by

$$\lambda_n \rightarrow d_n = \lambda_n + \frac{\lambda_1 / N_c}{1 + N_c^2 \lambda_n^2 \lambda_1^{-2}} \quad (14)$$

where λ_1 is the first eigenvalue of Σ and N_c is an input parameter that describes the maximum condition number for the regularised Σ such that

$$\lim_{\lambda_n \rightarrow 0} d_n \rightarrow \lambda_1 / N_c.$$

The regularised inverse of $\Sigma^{0.5}$ is given by

$$\Sigma^{-0.5} = \mathbf{V} \mathbf{D}^{-0.5} \mathbf{V}^T \quad (15)$$

where \mathbf{D} is a diagonal matrix with elements given by d_n for $n = 1, \dots, N$. Matrix inverses were regularised using $N_c = 10^4$ in this work.

For each model, the percent variance captured by each factor is based on the scores. In PCA the percent variance captured by each factor is calculated for \mathbf{X} . For MAF and MNF it is for $\mathbf{X} \Sigma_1^{-0.5}$ and for MDF it is for $\mathbf{D}_1 \mathbf{X} \Sigma_2^{-0.5}$.

When modelling all pixels in an image, MAF and MDF exclude the outer pixels at the image boundary when calculating the spatial derivatives to avoid end-effects in the Savitzky–Golay filter.⁹ However, all pixels are included when calculating scores. When pixels are excluded in an image (e.g., due to defects, artefacts or undesirable signal that should not “contaminate” the model), the appropriate surrounding pixels are also excluded when calculating spatial derivatives to avoid end-effects (e.g., see https://www.researchgate.net/publication/338518012_Savitzky-Golay_Smoothing_and_Differentiation_Filter).

Examples

Three examples are shown below. Example 1 compares results from the different algorithms for an infrared image of an Excedrin tablet. This example demonstrates the differences in signal captured and also how scores from the different methods can be combined

Table 1. PCA, MAF/MNF and MDF model comparisons.

Algorithm	GEP	SEP	Scores	Eigenvalue
PCA	$\Sigma \mathbf{p} = \lambda \mathbf{I} \mathbf{p}$	$\Sigma \mathbf{p} = \lambda \mathbf{p}$	$\mathbf{t} = \mathbf{X} \mathbf{p}$	$\lambda = \mathbf{p}^T \Sigma \mathbf{p} / \mathbf{p}^T \mathbf{I} \mathbf{p} = \mathbf{p}^T \Sigma \mathbf{p}$
MAF	$\Sigma \mathbf{w} = \lambda \Sigma_1 \mathbf{w}$	$\Sigma_1^{-0.5} \Sigma \Sigma_1^{-0.5} \mathbf{p} = \lambda \mathbf{p}$	$\mathbf{t} = \mathbf{X} \Sigma_1^{-0.5} \mathbf{p}$	$\lambda = \mathbf{w}^T \Sigma \mathbf{w} / \mathbf{w}^T \Sigma_1 \mathbf{w}$
MDF	$\Sigma_1 \mathbf{w} = \lambda \Sigma_2 \mathbf{w}$	$\Sigma_2^{-0.5} \Sigma_1 \Sigma_2^{-0.5} \mathbf{p} = \lambda \mathbf{p}$	$\mathbf{t} = \mathbf{D}_1 \mathbf{X} \Sigma_2^{-0.5} \mathbf{p}$	$\lambda = \mathbf{w}^T \Sigma_1 \mathbf{w} / \mathbf{w}^T \Sigma_2 \mathbf{w}$

\mathbf{P} is orthogonal, the columns of \mathbf{T} are orthogonal, \mathbf{W} is not generally orthogonal.

synergistically. Example 2 is a popular AVIRIS image of Indian Pines often used to demonstrate the performance of classification models. Here, the results show that although MAF and MNF are only slightly different they can capture different signal. The example also shows how MDF can be used to find and eliminate edge signal that might interfere with the calibration of classification models, and that the scores can be concatenated for further analysis. Example 3 is a Landsat 8 image of north central Washington state in USA. This example shows how the methods can be used for anomaly detection in the images and that detected signals can be used synergistically to help elucidate the source of detected anomalies. A fourth example shows results for an energy dispersive spectroscopy image and demonstrates significant differences. Additional information on the examples can be found in the SI.

Auto-contrasting is used for many of the images shown below to enhance visualisation. Auto-contrasting consists of mean-centring the scores and saturating at ± 2.5 standard deviations prior to displaying in an image.

Example 1. Excedrin tablet

An infrared image of an Excedrin tablet that contains aspirin, acetaminophen, caffeine and microcrystalline cellulose was studied. The tablet was measured with a tuneable laser in the mid-infrared from 1800 cm^{-1} to 800 cm^{-1} over an approximate 2 mm square area. The image cube was 218×208 pixels by 250 wavenumbers. The image can be downloaded at <https://eigen-vector.com/resources/data-sets/> and was provided by Agilent (<https://www.agilent.com>), and was discussed in Reference 13. For each algorithm shown, the data were mean-centred prior to modelling.

Figure 1 shows scores images for the first two factors (PCs 1 and 2) for PCA, MAF, MNF and the mean MDF score (mean of the vertical, UD, and horizontal, LR, scores). [PCs 3 and 4 are shown in the SI.] MAF and MNF show similar spatial structure for this image. MAF and MNF also show significant spatial structure on PC 1 and 2, but PCA shows less spatial structure on PC1—the

algorithms clearly capture different signal. PC 2 for PCA captures spatial structure similar to MAF's PC 1. The MDF scores show different forms of edge information on PC 1 and 2 because, like PCA and MAF, MDF is factor-based and can be considered a "factor-based edge detection" algorithm. The MDF images provide complementary information for image visualisation and exploration. It is interesting to note that the PCA loadings are the most different for the different decomposition methods. The fraction variance captured by PC 1 for MAF and MDF are much smaller than for PCA—recall that the variances are calculated for the weighted PCA for MAF and MDF and that the correlation structure can be significantly modified due to the weighting (i.e., the data appear less correlated for MAF and MDF than for PCA).

Figure 2 shows composite images for the Excedrin image. The left-hand set of four images includes all pixels in the modelling calculation and the right-hand set of four images has edge pixels, identified by the mean MDF Hotelling's T^2 , excluded from the modelling step. (The excluded edges correspond to pixels with high mean MDF T^2 values separating contiguous patches in the image.) Although the excluded pixels were projected onto the model, and could be visualised on the right-hand, the excluded edge pixels have been masked and are not displayed (they are seen as dark pixels). Noting the results in Figure 1 suggests that PCA factors 2, 3 and 4 correspond best to the signal in the other methods for factors 1, 2 and 3. Therefore, for ease of comparison, the PCA scores are shown for factors 2, 3 and 4 for PCA and 1, 2 and 3 for MAF, MNF and MDF. It is interesting to note that the right-hand images appear to capture slightly different information in the scores than the left-hand images as indicated by the different colours in the false colour images. The bottom-right image in the left-hand images shows a composite of the scores on PCA PC 1 (red), MDF PC 2 UD (green) and LR (blue) as an example of using the scores from different algorithms synergistically.

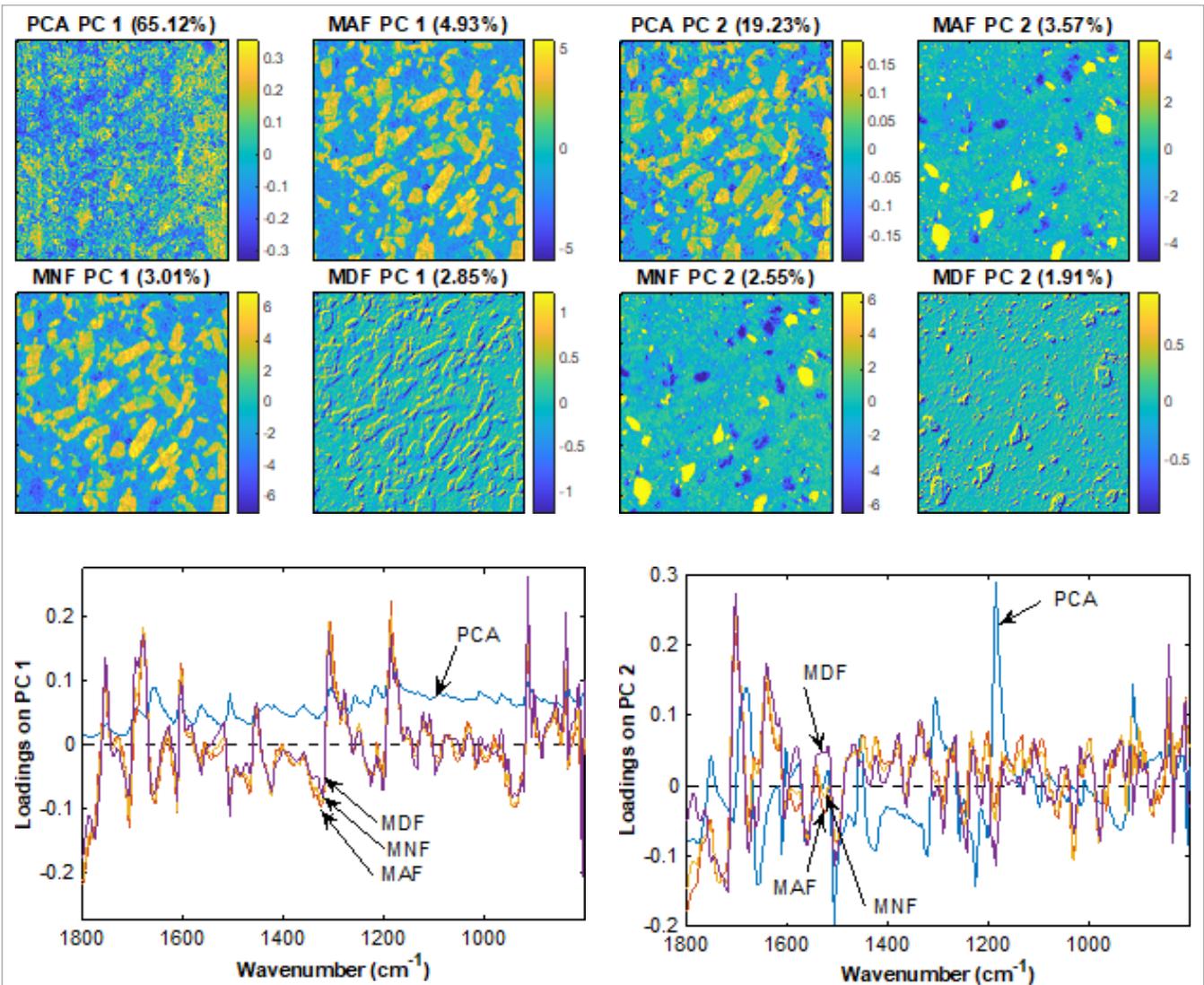


Figure 1. PCA, MNF, MAF and mean MDF scores on PC 1 (top, left) and 2 (top, right) for the Excedrin image [auto-contrasted]. Bottom: corresponding loadings plots.

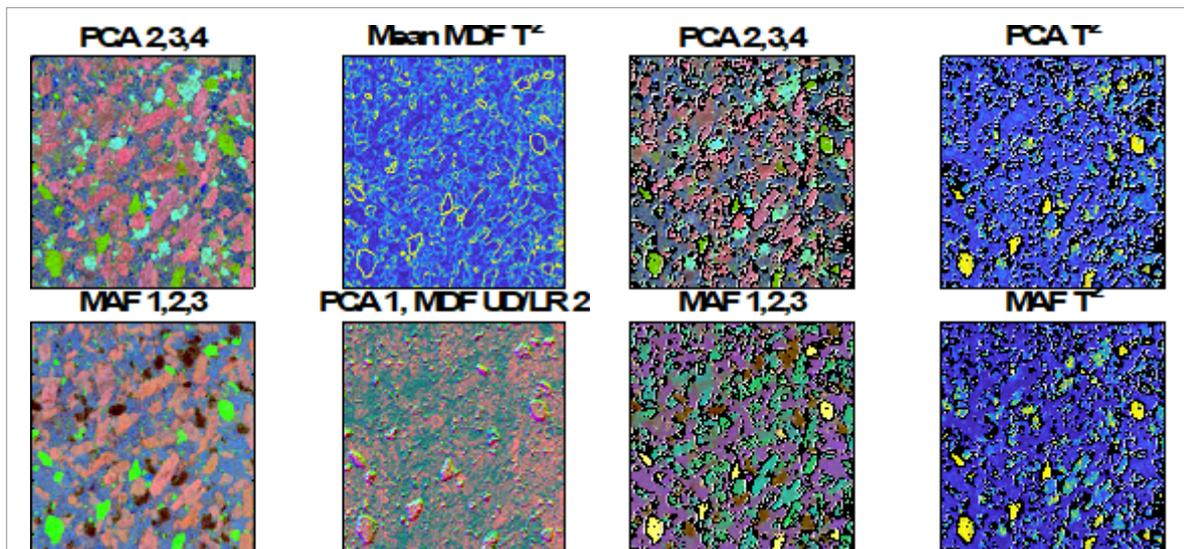


Figure 2. Composite images for the Excedrin image. The left set of four images includes all pixels. The right set of four images has edge pixels identified by MDF excluded [auto-contrasted].

Example 2. Indian Pines

Comparison of PCA, MAF and MDF for an AVIRIS image

The Indian Pines data set¹⁴ is often used to demonstrate the performance of classification algorithms. The image is 145×145 pixels measured at 220 bands from 400nm to 2500nm over several agriculture fields. This example shows how the image decomposition techniques might be used to prepare the image for classification and how scores from the different decompositions can be concatenated and used for classification. The image data were preprocessed using 1-norm normalisation followed by mean-centring.

Figures 3 and 4 show the scores and loadings on PC 1 to 4 respectively for the four decomposition methods. As

expected, PCA and MAF capture slightly different signals and MDF captured edge signal. It is interesting to note that the MNF image on PC 1, that uses a first difference for D_1 , looks significantly sharper than the MAF image that uses a first central difference. Additionally, MAF and MNF capture different signal on PCs 2–4. This shows that the definition of the D_1 can impact the MAF/MNF results.

Figure 5 shows results where pixels with high Q residuals (unusual signal, Figure SI.12) and high MDF Hotelling's T^2 were excluded from the modelling step. Pixels with high Q are often excluded because this signal can diminish model performance. For this study, pixels with Q greater than the 99% confidence limit for each decomposition method were excluded. It is also difficult

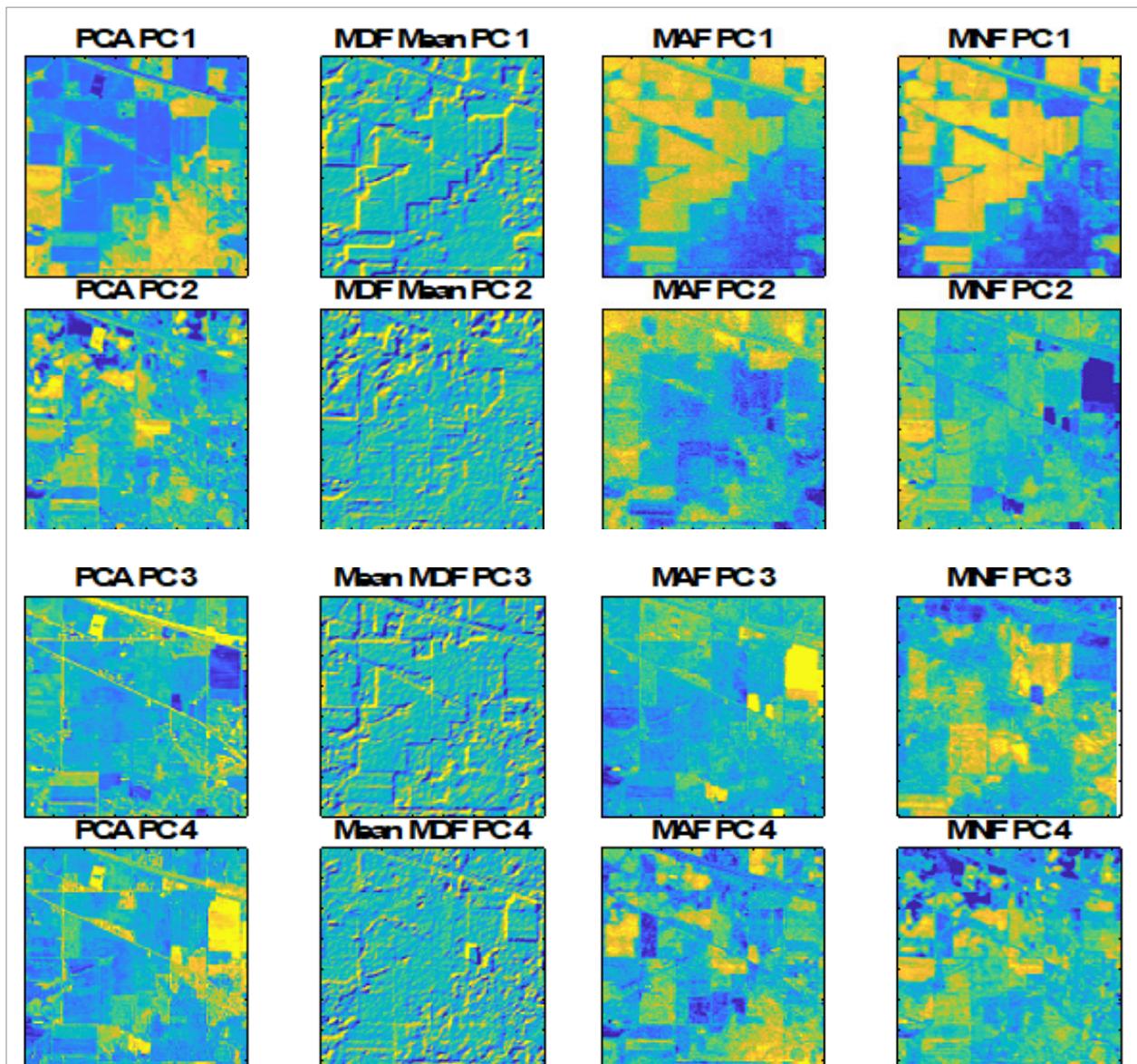


Figure 3. PCA, MDF, MAF and MNF mean scores for the Indian Pines image [auto-contrasted].

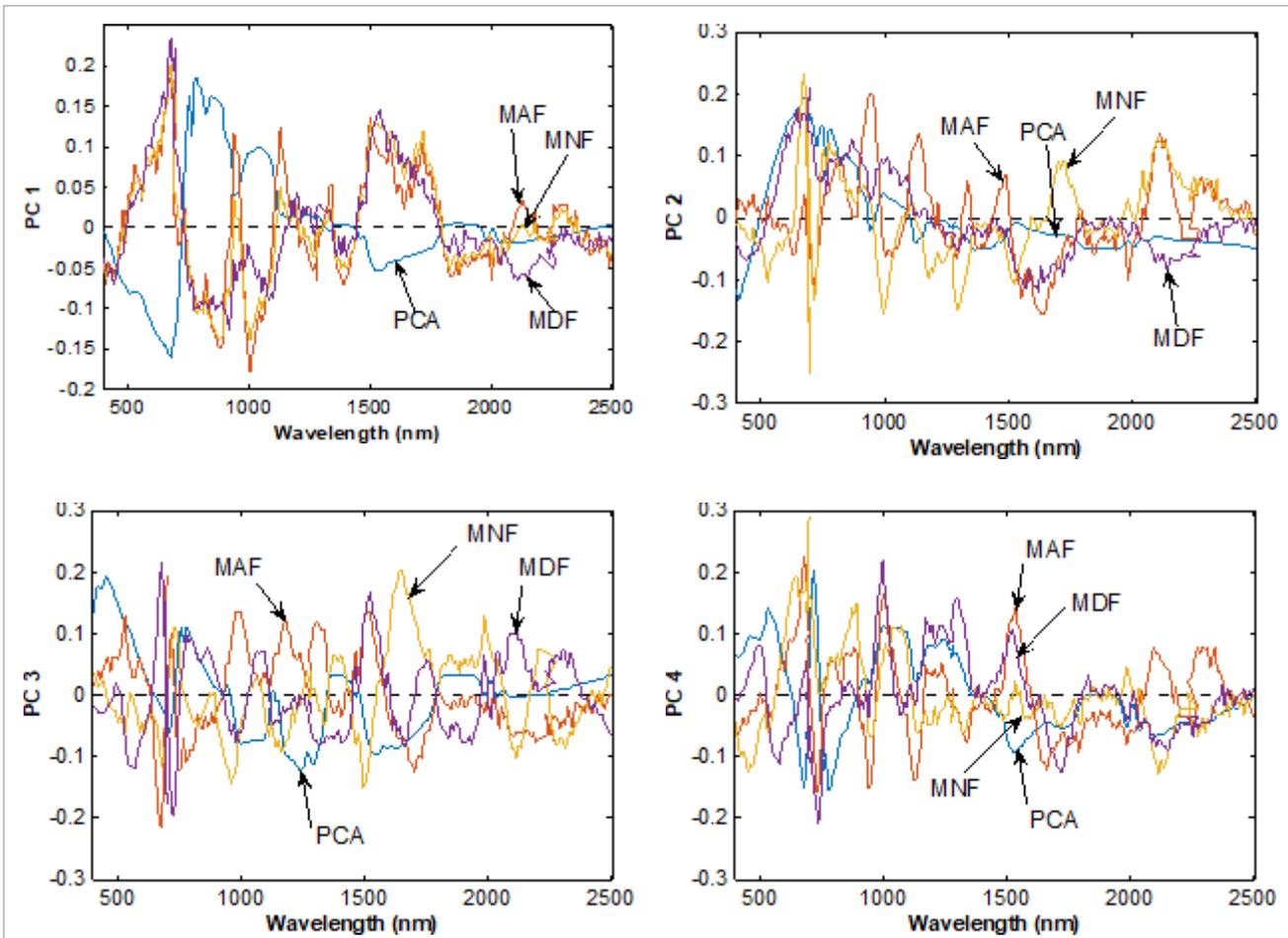


Figure 4. Loadings for PCA, MDF, MAF and MNF mean scores for the Indian Pines image.

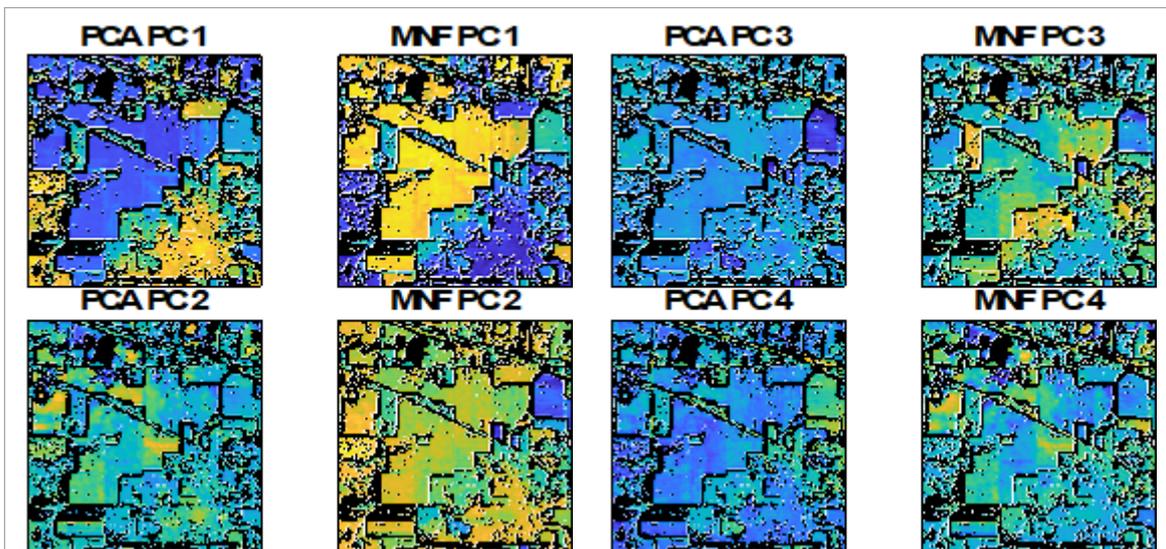


Figure 5. PCA and MNF scores for the AVIRIS Indian Pines image with high Q and edges excluded [auto-contrasted].

to provide quality ground truth edge signal (mixed pixels) in images—especially for images obtained in remote and standoff sensing. The result is that edge signal makes characterisation of classification performance more difficult. For this work, edge signal identified using MDF Hotelling's T^2 was excluded: MDF UD and LR pixels with Hotelling's $T^2 > 10$ were removed prior to additional analysis. PCA and MNF modelling results after these “data cleaning” steps are shown in Figure 5. With the exception of an apparent sign ambiguity, it appears that PC 1 for PCA and MNF capture similar signal with differences between PCA and MNF more apparent in the higher PCs. Classification models can take advantage of the differences in captured signal between the methods by concatenating scores into a consensus model. For example, the results for PCA of the autoscaled concatenated scores from PCA, MAF and MNF are shown in Figure 6. The loadings plot for PC 1 on the consensus model (Figure 6 right) shows high correlation between the PCA, MAF and MNF scores on PC 1. The observation of high loadings with opposite signs for PCA and MNF is consistent with the interpretation given above of a sign ambiguity between the PC 1 images in Figure 5. The analysis of concatenated scores is another example of synergistic use of the different models.

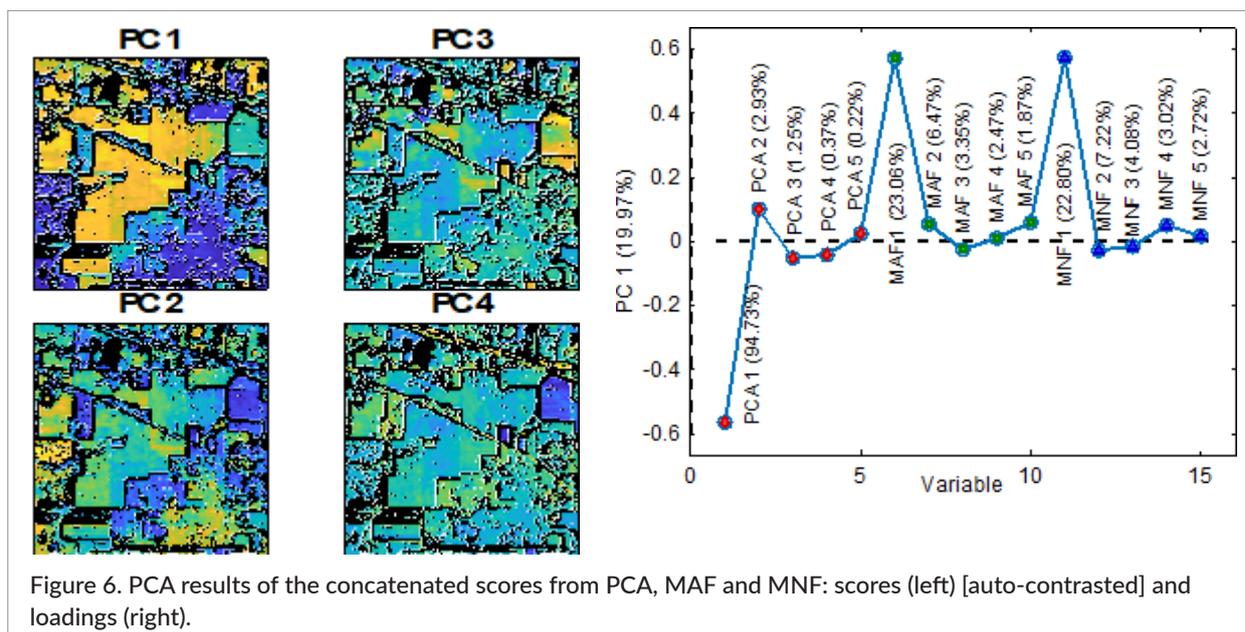
Example 3. Landsat 8 image

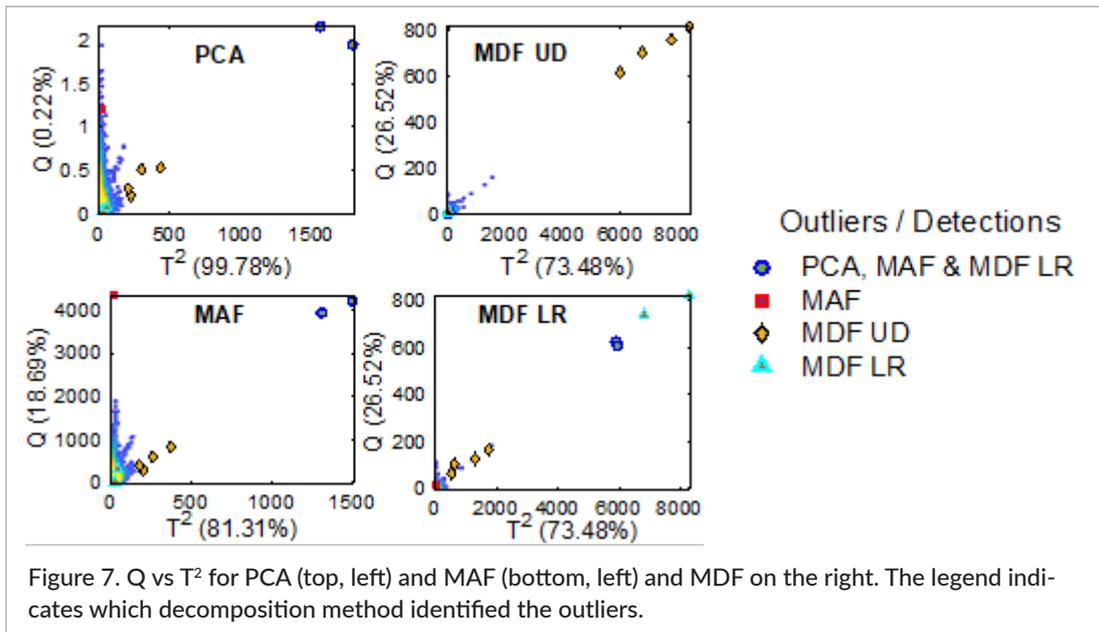
A section of a Landsat 8 [image LC08_L1TP_045027_20170621_20170630_01] corresponding to the north-west end of Lake Chelan, WA was analysed. The image was courtesy of the US Geological Survey FILE_DATE =

2017-06-30T11:25:51Z (DATE_ACQUIRED = 2017-06-21, SCENE_CENTER_TIME = 18:49:12.7536510Z) and is a portion of the same image shown on the cover of https://www.impopen.com/vi-toc/V_IASIM-2018. The image is 900 pixels wide by 700 pixels high. Eight bands at an approximate 30m spatial resolution were included in the analysis (see Table SI.3 in the SI).

The interest in this example is anomaly detection, but small anomalies (e.g., with only a few pixels) can be difficult to identify from the image directly. Instead, plots of Q versus T^2 can be useful as shown in Figure 7 using autoscaling preprocessing and keeping four PCs for each model. In this case, PCA, MAF and MDF LR identified the same two outlier pixels while MDF LR identified two additional outlier pixels. MDF UD identified an additional four outlier pixels. MAF identified an unusual pixel (red square marker in Figure 7) that was on the far left-hand side of the image, while the remaining eight outliers were in close proximity to each other suggesting they are related. Of these eight pixels, PCA and MAF caught the same two outliers while MDF caught an additional six pixels showing that the methods are sensitive to different types of spatial signal in the image and are complementary in identifying an anomaly in the image.

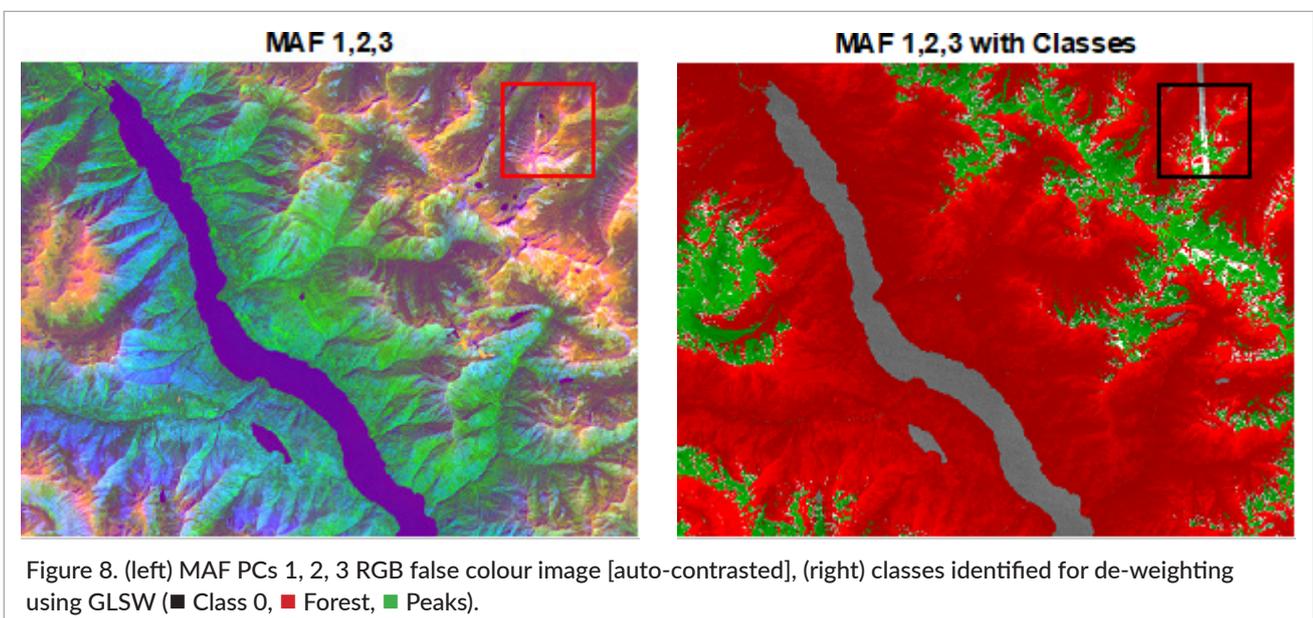
Once an anomaly is identified, other preprocessing can be used to enhance the anomalous signal and potentially elucidate its source. An example using autoscaling and generalised least squares (GLSW) decluttering^{12,15,16} (de-weighting) is shown in Figure 8. In this image, the blue pixels in Figure 8 (left) correspond to water and the decluttering was based on the identified classes shown in Figure





8 (right): Forest and Peaks. GLSW did not use mean-centring of the classes and used a regularisation parameter of $\alpha=0.002$. It is important to note that pixels associated with the anomaly were not included in the GLSW de-weighting because this is the signal to be enhanced in an effort to elucidate the cause of the anomaly. The result is that GLSW de-weighted the Forest and Peaks signal prior to modelling. Rectangles in the upper-right in the images of Figure 8 indicate the location of the anomaly shown in Figure 9. The scores on MDF LR PC 4 (Figure 9) show very strong signal seen as extreme white and black pixels

indicated by the arrow (this image was saturated at ± 8). Pixels in the images represent an area of approximately $30\text{m} \times 30\text{m}$, and the MDF LR image suggests that the anomaly was due to an aircraft flying from top-to-bottom in the image with a contrail trailing towards the top of the image. Although the extreme pixels were also seen in the MDF UD image (not shown), the contrail was not clearly apparent. However, the contrail was clearly seen on PC 4 in the MDF LR image; MDF was more sensitive in the LR image than the UD image. The difference in sensitivity was likely due to differences in lighting in the image. MDF was



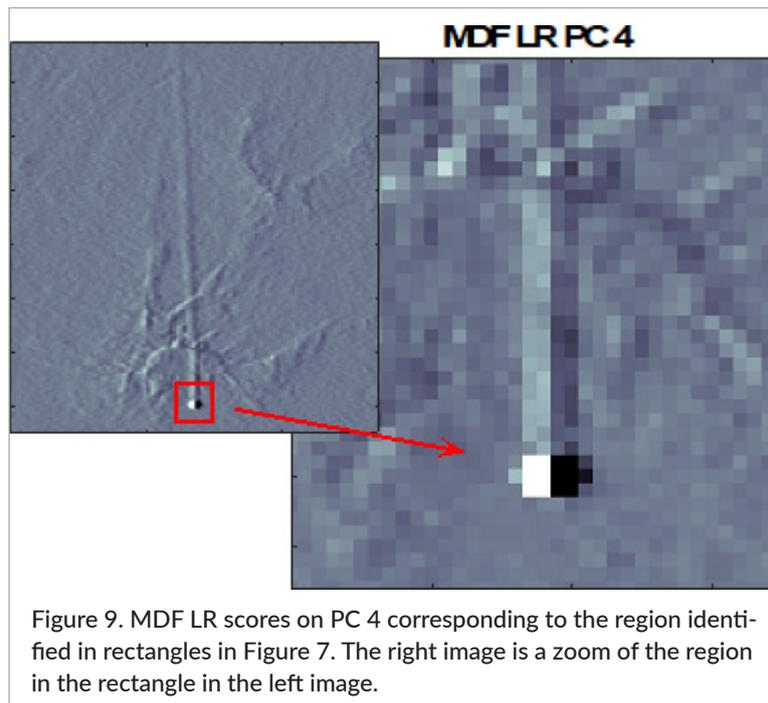


Figure 9. MDF LR scores on PC 4 corresponding to the region identified in rectangles in Figure 7. The right image is a zoom of the region in the rectangle in the left image.

also more sensitive than MAF and PCA. This enhanced sensitivity was attributed to the fact the MDF factors were based on local variance in the signal in contrast to global variance used in PCA and MAF.

Example 4. Comparison of PCA, MAF and MDF for an EDS image

In this example, six types of wires made of different alloys were embedded in epoxy. The sample was then cut and the exposed ends were measured with energy dispersive spectroscopy (EDS)¹⁷ and discussed in Reference 6. The wires were arranged in rows with the top row 100% Ni, Row 2 36% Ni, 64% Fe, Row 3 70% Cu, 30% Zn, Row 4 16% Cr, 84% Fe, Row 5 13% Mn, 4% Ni, 83% Cu, and the bottom row 100% Cu.

Figures 10 and 11 shows scores images and loadings, respectively, for PCA, MAF and MDF for four factors. The observed spatial patterns for PCA and MAF are similar but the colouring (yellow is high and blue is low) is different indicating that the two algorithms capture different signal. It is interesting that the scores for PCs 3 and 4 for both PCA and MDF appear to be focused on a similar signal (similar location) in the images while that spatial statistics for MAF on these factors is significantly different. In summary, the results show significant differences between the methods. Figure 10 (bottom right) shows synergistic visualisation on PC 4 for PCA and MDF.

Conclusions

Principal components analysis (PCA), maximum auto-correlation factors (MAF), minimum noise factors (MNF) and maximum difference factors (MDF) can all be posed as a maximisation problem that ultimately yields a symmetric eigenvalue problem (SEP) for each model. The SEPs allow a simple comparison of the models for analysis of hyperspectral images using a PCA metaphor with MAF, MNF and MDF describable as weighted PCA. The examples showed how the different methods captured different signals in the images and the scores from each method can be combined synergistically allowing for additional modelling and extended visualisation. MDF is a factor-based edge detection model that not only allows for additional visualisation but the opportunity to identify and exclude (or highlight) edge signal in the images. The third example showed that models can also be used synergistically for finding and elucidating anomalies. In the example, MDF showed the highest sensitivity of the models studied for anomaly detection.

Acknowledgements

The author would like to thank Robert T. Roginski and Donal O'Sullivan for reviewing this article.

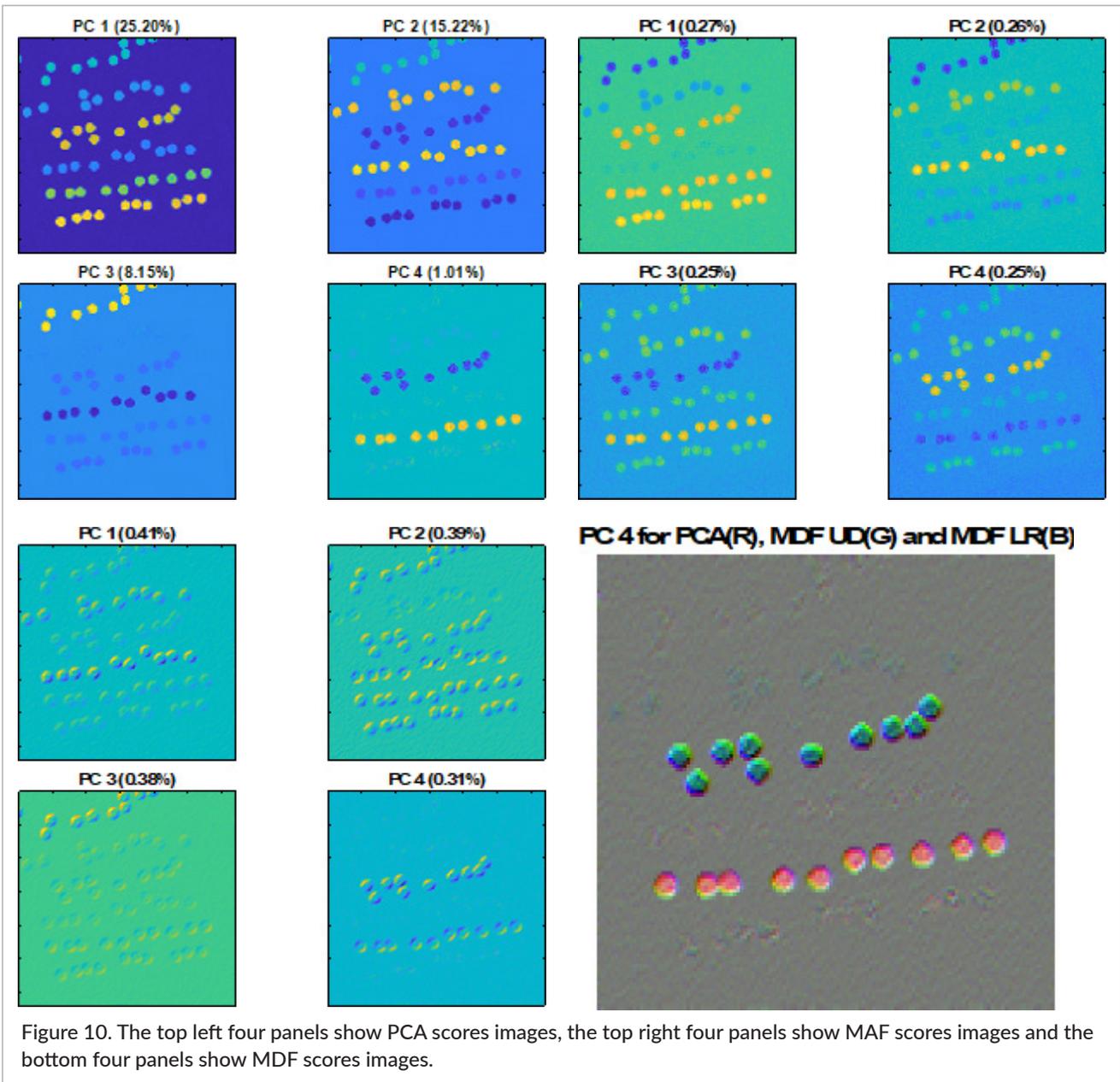


Figure 10. The top left four panels show PCA scores images, the top right four panels show MAF scores images and the bottom four panels show MDF scores images.

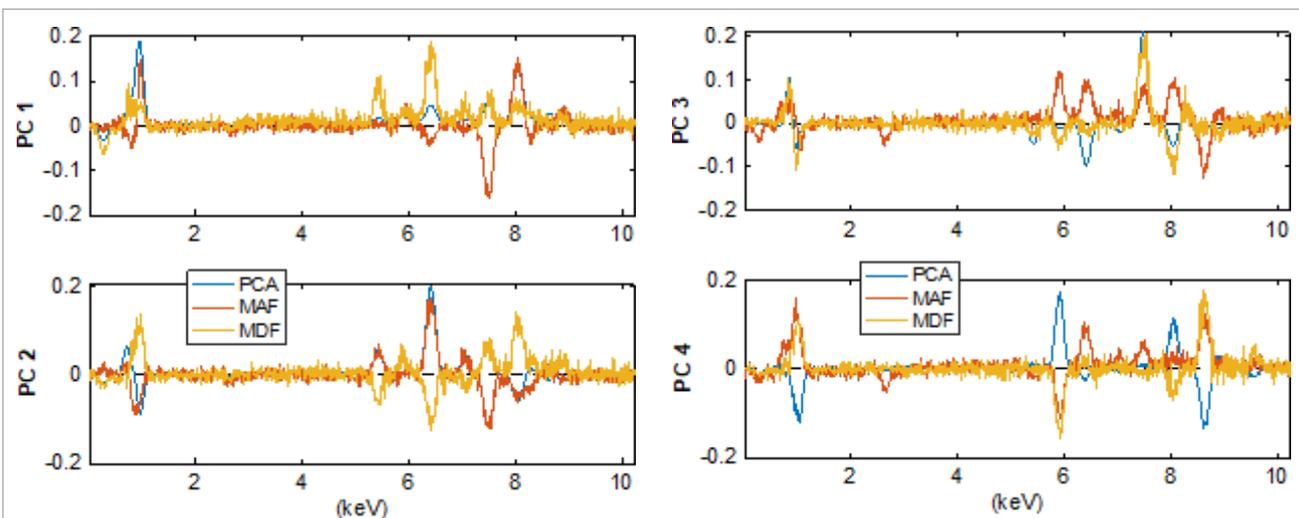


Figure 11. Left: loadings for PCs 1 and 2 for the three decompositions. Right: PCs 3 and 4.

References

1. J.E. Jackson, *A User's Guide to Principal Components*. John Wiley & Sons, New York, NY (1991). ISBN: 0-471-62267-2.
2. R. Bro and A.K. Smilde, "Principal components analysis", *Anal. Methods* **6**, 2812–2831 (2014). <https://doi.org/10.1039/c3ay41907j>
3. A.A. Green, M. Berman, P. Switzer and M.D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal", *IEEE T. Geosci. Remote* **26(1)**, 65–74 (1988). <https://doi.org/10.1109/36.3001>
4. R. Larsen, "Decomposition using maximum autocorrelation factors", *J. Chemometr.* **16(8–10)**, 427–435 (2002). <https://doi.org/10.1002/cem.743>
5. M. Berman, A. Phatak and A. Traylen, "Some invariance properties of the minimum noise fraction transform", *Chemometr. Intel. Lab. Syst.* **117**, 189–199 (2012). <https://doi.org/10.1016/j.chemo-lab.2012.02.2005>
6. N.B. Gallagher, J.M. Shaver, R. Bishop, R.T. Roginski and B.M. Wise, "Decompositions using maximum signal factors", *J. Chemometr.* **28(8)**, 663–671 (2014). <https://doi.org/10.1002/cem.26347>
7. J.M. Amigo (Ed.), *Hyperspectral Imaging, Vol 32 (Data Handling in Science and Technology)*, 1st Edn. Elsevier, Amsterdam, The Netherlands (2019). ISBN: 978-0-444-63978-3
8. H.F. Grahn and P. Geladi (Eds), *Techniques and Applications of Hyperspectral Image Analysis*. John Wiley & Sons, UK (2007). ISBN: 978-0-470-01087-7.
9. A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data by simplified least squares procedures", *Anal. Chem.* **36(8)**, 1627–1639 (1964). <https://doi.org/10.1021/ac60214a047>
10. B.M. Wise and N.B. Gallagher, "The process chemometrics approach to chemical process monitoring and fault detection", *J. Proc. Cont.* **6(6)**, 329–348 (1996). [https://doi.org/10.1016/0959-1524\(96\)00009-1](https://doi.org/10.1016/0959-1524(96)00009-1)
11. C.B. Moler and G.W. Stewart, "An algorithm for generalized matrix eigenvalue problems", *SIAM J. Numer. Anal.* **10(2)**, 241–256 (1973). <https://doi.org/10.1137/0710024>
12. T.A. Blake, J.F. Kelly, N.B. Gallagher, P.L. Gassman and T.J. Johnson, "Passive standoff detection of RDX residues on metal surfaces via infrared hyperspectral imaging", *Anal. Bioanal. Chem.* **395(2)**, 337–348 (2009). <https://doi.org/10.1007/s00216-009-2907-5>
13. T.G. Avval, B. Moeini, V. Carver, N. Fairley, E.F. Smith, J. Baltrusaitis, V. Fernandez, B.J. Tyler, N. Gallagher and M.R. Linford, "The often-overlooked power of summary statistics in exploratory data analysis: comparison of pattern recognition entropy (PRE) to other summary statistics and introduction of divided spectrum-PRE (DS-PRE)", *J. Chem. Inf. Model.* **61**, 4173–4189 (2021). <https://doi.org/10.1021/acs.jcim.1c00244>
14. M.F. Baumgardner, L.L. Biehl and D.A. Landgrebe, *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*. Purdue University Research Repository (2015). <https://doi.org/10.4231/R7RX991C>, <https://aviris.jpl.nasa.gov/aviris/index.html>
15. H. Martens, M. Høy, B.M. Wise, R. Bro and P.B. Brockhoff, "Pre-whitening of data by covariance-weighted pre-processing", *J. Chemometr.* **17(3)**, 153–165 (2003). <https://doi.org/10.1002/cem.780>
16. N.B. Gallagher, "Classical least squares for detection and classification", in *Hyperspectral Imaging, Vol. 32 (Data Handling in Science and Technology)*, 1st Edn, Ed by J.M. Amigo. Elsevier, Amsterdam, The Netherlands, pp. 231–246 (2019). ISBN: 9780444639776, <https://doi.org/10.1016/B978-0-444-63977-6.00011-0>
17. M.R. Keenan, "Multivariate analysis of spectral images composed of count data", in *Techniques and Applications of Hyperspectral Image Analysis*, Ed by H.F. Grahn and P. Geladi. John Wiley & Sons, UK, pp. 89–126 (2007). ISBN: 978-0-470-01087-7