Experimental Approach to the Selection of the Components in the Minimum Noise Fraction

Umberto Amato, Rosa Maria Cavalli, Angelo Palombo, Stefano Pignatti, and Federico Santini

Abstract—An experimental method to select the number of principal components in minimum noise fraction (MNF) is proposed to process images measured by imagery sensors onboard aircraft or satellites. The method is based on an experimental measurement by spectrometers in dark conditions from which noise structure can be estimated. To represent typical land conditions and atmospheric variability, a significant data set of synthetic noise-free images based on real Multispectral Infrared and Visible Imaging Spectrometer images is built. To this purpose, a subset of spectra is selected within some public libraries that well represent the simulated images. By coupling these synthetic images and estimated noise, the optimal number of components in MNF can be obtained. In order to have an objective (fully data driven) procedure, some criteria are proposed, and the results are validated to estimate the number of components without relying on ancillary data. The whole procedure is made computationally feasible by some simplifications that are introduced. A comparison with a state-of-the-art algorithm for estimating the optimal number of components is also made.

Index Terms—Image enhancement, image processing, image restoration, noise, remote sensing.

I. INTRODUCTION

MINIMUM NOISE fraction (MNF) is a linear transform used to estimate the actual dimension of an image, remove noise, and reduce processing time [1]. MNF projects input data (a multidimensional image) into a new space that is split into two parts: The first one is associated to the signal; the second one is dominated by noise affecting the image. Using only the first subspace of data, an improvement of the image is produced in terms of a reduced dimension and an increase of the signal to noise ratio (SNR). One of the main problems in this respect is the choice of the boundary between the two subspaces.

Most applications do not use an objective (i.e., driven by data only) criterion for choosing the number of MNF components to be retained. Sometimes an empirical threshold is set; sometimes completely visual criteria are resorted based on the behavior of the singular values of suitable matrices. As an example, Zhang et al. [2] utilize the first 28 MNF bands with eigenvalues greater than two for the analysis of Airborne Visible/Infrared Imaging Spectrometer images in a study of the vegetation stress; Burke et al. [3] use a visual evidence to separate clear MNF bands from the other ones dominated by noise in the coastal characterization by means of Advanced Land Imager Hyperion images; Li et al. [4] use a visual inspection to select the number of components in the land cover mapping. In other circumstances, in [5], it is preferred to overestimate the number of retained components to avoid loss of information or to assign the number of components depending on the accuracy of the results of the classification as in [6]. We also mention [7]–[12] for the application of MNF to other contexts relying on subjective criteria. See [13] for an application of MNF when some information is known on the noise structure. The objective estimate of the optimal number of components is an active field of research, because, in many circumstances, the visual analysis of the eigenvalues is not sufficient to permit a reliable estimate; in addition, it is the only solution in near real-time applications. For these reasons, several objective criteria have been developed in the literature for the estimate of the optimal number of components mostly connected with the Principal Component Analysis (PCA). The most classical ones are the well-known Akaike information criterion [14] and the Bayesian information criterion (see [15], e.g., for a clear introduction). They are based on the principle of the maximum likelihood of the model underlying the PCA, but they include a penalization term that encourages the choice of a parsimonious model (i.e., low number of components). These criteria are nowadays generally believed to be too rough for real applications, so that some more advanced methodologies have been developed. In this respect, we mention the use of the Gershgorin radii [16], which rely on special mathematical properties of eigenvalues of matrices. One of the most recent and effective methodologies for estimating the optimal number of components is the profile likelihood introduced in [17], which can be naturally extended to MNF.

In this paper, we propose a new experimental methodology to estimate the boundary between the signal and noise spaces when images come from the Multispectral Infrared and Visible Imaging Spectrometer (MIVIS) sensor onboard aircraft. To this purpose, some MIVIS images were used. The methodology is based on a first step where synthetic noise-free images are produced; these images are obtained by a mixing procedure based on spectral libraries and real MIVIS images. Then, noise is added to these images; noise is estimated through a specific measurement by MIVIS inside a dark chamber. In this way, we have a set of “true” images unaffected by noise, which we can use as a reference, and a set of corresponding images...
affected by noise. By iterative application of MNF (with different number of retained components) to the noisy images and comparing the retrieved spectra with the corresponding ones unaffected by noise, we can infer the number of optimal components (OCs). Then, relationships between the number of OCs and some prognostic parameters that can be extracted from the images are investigated, in order to yield an objective (i.e., data driven) criterion for the choice of the number of OCs for any real image acquired by MIVIS. These relationships hold effective for the whole period of activity of the sensor, unless deep changes in its radiometric characteristics occur; in which case, it is needed to repeat the procedure. Of course, the whole methodology can be extended to any sensor.

II. MNF

The MNF is based on the projections of multivariate input data (image) into a new space where the SNR is highest. Let us consider the vector $Z_i \equiv (z_1, z_2, \ldots, z_N)$ (dropping pixel index on the vector components not to heavy notation) that represents the spectrum of the $i$th pixel of the multispectral image having $N$ bands. We assume that signal and noise are uncorrelated according to the following model

$$Z_i = S_i + N_i \quad (1)$$

where $S_i$ and $N_i$ are signal and noise components, respectively [13], [18]. By introducing matrices $C_Z$ and $C_N$, covariance matrices of image and noise, respectively, we seek for a transform $A$ such that

$$A^T C_N A = I$$

where $I$ is identity matrix of order $N$. Matrix $A$ can be obtained via the following:

$$A = B \Lambda^{-1/2}$$

with $B$ being the matrix that diagonalizes $C_N$ and $\Lambda$ being the diagonalized matrix

$$B^T C_N B = \Lambda.$$  

Then, the covariance matrix of the image $C_Z$ is transformed by $A$, getting

$$\tilde{C}_Z = A^T C_Z A.$$  

The matrix $D^T$ that diagonalizes $\tilde{C}_Z$ is found as

$$D^T \tilde{C}_Z D = \Delta \quad (2)$$

with $\Delta$ being the resulting diagonal matrix; finally, the transform matrix of MNF is given by

$$H = AD.$$  

As shown by Green et al. [13], Lee et al. [19], Xu and Gong [20], and Roger [21], applying this transform to a vector $Z_i$

$$\tilde{Z}_i = H^T Z_i$$

we obtain a new vector $\tilde{Z}_i$, where transformed bands (superbands) are ordered according to their noise fraction. MNF is based on the assumption that a signal mostly concentrates on the first superbands, whereas noise is spread over all superbands with unit variance. Therefore, in the whole space, it is possible to isolate a subspace of dimension $p$ where the variance of the SNR is highest. This result can be accomplished by keeping the first $p$ bands by equating to zero the remaining $N - p$ (smallest in terms of variance contribution) components of the spectrum $\tilde{Z}_i$

$$\tilde{Z}_i = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_p, 0, \ldots, 0). \quad (3)$$

Once the $p$ superbands containing the signal have been selected, it is possible to reproject the spectrum into the originary space; in this way, noise contained in the zeroed $N - p$ superchannels has been removed

$$Z_i^{\text{extr}} = H^{-T} \tilde{Z}_i \quad (4)$$

with $Z_i^{\text{extr}}$ being the retrieved spectrum. For this study, the matrix $C_Z$ was estimated from the data

$$C_z = \frac{1}{Q} \sum_{i=1}^{Q} (Z_i - \bar{Z})(Z_i - \bar{Z})^T$$

where $Q$ is the number of pixels of the image and $\bar{Z}$ represents the average value over the pixels.

Matrix $C_N$ has been estimated from values of the dark current acquired by spectrometer for each scanning line. The dark current has been measured on a blackbody at a temperature around 300 K; therefore, it is a good representation of the instrumental noise in the range of 400–2500 nm

$$C_N = \frac{1}{K} \sum_{i=1}^{K} (N_i - \bar{N})(N_i - \bar{N})^T$$

where $K$ is the number of scan lines acquired in dark conditions, $N_i$ is the $i$th spectrum of the dark acquisition, and $\bar{N}$ represents the average over the $K$ dark spectra.

III. SYNTHETIC IMAGES

The proposed method is based on the use of synthetic images to define a criterion for the selection of the number of components to retain in the MNF inverse transformation. Simulated images have to satisfy the assumption (1) that signal and noise are additive, hypothesis underlying MNF. Therefore, the two components have been built independently. Noise has been obtained by gathering an image in darkness conditions on the calibration bench. Signal has been obtained by building synthetic images by the selection of appropriate endmembers in the linear mixing technique.

A. Linear Mixing

In order to produce the synthetic images, a linear combination of spectra inside a suitable library has been considered. The assumption is that the spectrum of a composite material is a linear combination of one or more pure materials (endmembers)
which it is composed of, coefficients of the combination being proportional to the corresponding abundance [22].

This can be described by the following:

\[
\begin{align*}
y_1 &= a_1 x_1^1 + a_2 x_1^2 + \cdots + a_m x_1^m + \varepsilon_1 \\
y_2 &= a_1 x_2^1 + a_2 x_2^2 + \cdots + a_m x_2^m + \varepsilon_2 \\
\vdots & \quad \vdots \\
y_N &= a_1 x_N^1 + a_2 x_N^2 + \cdots + a_m x_N^m + \varepsilon_N 
\end{align*}
\]

where \( Y \equiv (y_1, y_2, \ldots, y_N) \) represents the spectrum of the composite material, \( X^i \equiv (x_1^i, x_2^i, \ldots, x_N^i) \) is the spectrum of the \( i \)th endmember, \( m \) is the number of endmembers, \( E \equiv (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N) \) is noise and the part of spectrum not represented by endmembers, and \( A \equiv (a_1, a_2, \ldots, a_m) \) represents the vector of the \( m \) abundances. In practice, since \( E \) is unknown, a solution is sought to the following system of equations:

\[
\begin{align*}
y_1 &= a_1 x_1^1 + a_2 x_1^2 + \cdots + a_m x_1^m \\
y_2 &= a_1 x_2^1 + a_2 x_2^2 + \cdots + a_m x_2^m \\
\vdots & \quad \vdots \\
y_N &= a_1 x_N^1 + a_2 x_N^2 + \cdots + a_m x_N^m. \quad (5)
\end{align*}
\]

In all practical cases, this system of equations does not have an analytic solution. Therefore, an approximate solution is sought by linear least squares method (LS) [23].

Usually, a constraint is imposed \( 0 \leq a_i \leq 1, \ i = 1, \ldots, m \), accounting for the fact that abundances cannot be negative or greater than one. In our study, no restriction has been applied, since our main interest lies in the solution that best approximates spectrum \( Y \), not limiting search to linear combinations that are meaningful from a physical point of view.

### B. Endmember Selection

Synthetic images have been obtained using a spectral library composed of 1170 spectra. Spectra of minerals and some of vegetation have been taken from the ENVI 4.2 spectral library [on their own extracted from the spectral libraries of the following Laboratories: U.S. Geological Survey, Johns Hopkins University, Jasper Ridge Biological Reserve of the Stanford University, and Jet Propulsion Laboratory (http://speclib.jpl.nasa.gov)]. Other spectra of urban areas, vegetation, and sea water have been obtained in laboratory and during measurement campaigns of the Airborne Laboratory for Environmental Research of the Italian National Research Council; other spectra concerning sea and lake surfaces have been obtained by simulation via Hydrolight, which is a radiative transfer model, for several concentrations of algeas and sediment.

To set up the synthetic images reflecting the real operational condition of sensors, eight reflectance images acquired on different scenarios (urban, industrial, rural, and coastal) have been used as reference. For each real image, a corresponding synthetic image (i.e., not affected by noise) has been obtained by linear mixing procedure. For each pixel or cluster of pixels, the \( m \) endmembers of (5) have to be selected from within the spectral library. If the spectral library contains \( P \) spectra, there are \( P!/(m!(P-m)!) \) different combinations that are possible, which is unfeasible (besides being unnecessary) on any computer. For this reason, we apply a forward step selection procedure as follows: For each spectrum of the library \( \ell = 1, \ldots, P \), we estimate a distance \( \delta_\ell \) from the spectrum of the pixel \( Y \equiv (y_1, y_2, \ldots, y_N) \) according to

\[
\delta_\ell = \frac{1}{N} \sqrt{\sum_{j=1}^{N} [(y_j - a_1 x_{\ell,j} - a_2 x_{\ell,j}) \tau_j SNR_j]^2}
\]

where \( N \) is the number of spectral bands, \( \tau_j, j = 1, \ldots, N \), is the transmittance of a standard atmosphere for band \( j \), and \( SNR_j \) is the SNR for the band \( j \). In this way, we give a lower weight to the noisy spectral bands or those located in a spectral range where atmosphere is scarcely transparent. The first endmember \( \ell_1 \) will be chosen as the one that gives minimum distance \( \delta_1 := \arg\min_\ell \delta_\ell \). Once we have chosen the first endmember, we choose the second one \( \ell_2 \) by minimizing

\[
\delta_\ell = \frac{1}{N} \sqrt{\sum_{j=1}^{N} [(y_j - a_1 x_{\ell_1,j} - a_2 x_{\ell_2,j}) \tau_j SNR_j]^2}
\]

where the abundances are the solution to the following system of equations for each \( \ell = 1, \ldots, P, \ell \neq \ell_1 \):

\[
\begin{align*}
y_1 &= a_1 x_{\ell_1,1} + a_2 x_{\ell_1,1} \\
y_2 &= a_1 x_{\ell_1,2} + a_2 x_{\ell_1,2} \\
\vdots & \quad \vdots \\
y_N &= a_1 x_{\ell_1,N} + a_2 x_{\ell_1,N}.
\end{align*}
\]

Having fixed \( \ell_1 \) and \( \ell_2 \), we proceed solving the following system for the third endmember:

\[
\begin{align*}
y_1 &= a_1 x_{\ell_1,1} + a_2 x_{\ell_1,1} + a_3 x_{\ell_1,1} \\
y_2 &= a_1 x_{\ell_1,2} + a_2 x_{\ell_1,2} + a_3 x_{\ell_1,2} \\
\vdots & \quad \vdots \\
y_N &= a_1 x_{\ell_1,N} + a_2 x_{\ell_1,N} + a_3 x_{\ell_1,N}.
\end{align*}
\]

The procedure is interrupted when the desired number of endmembers is reached. In this work, the number of endmembers has been chosen as eight. In principle, it would be possible to pick different endmembers for each pixel of the image, but this would be too heavy from the computational point of view. Therefore, in order to reduce the computational cost, an a priori unsupervised classification of the images has been worked out. We have used isodata classification [24], choosing a number of classes in the range from 20 to 30, so that all pixels were sorted in homogeneous groups from the spectral point of view. Then, an average spectrum for each class has been computed, and the endmember stepwise search procedure previously described was applied on these average spectra. In this way, we obtained eight endmembers for each of the classes that compose the whole image. Finally, for each single pixel, (5) has been solved, by using the eight endmembers of the cluster to which the pixel belongs, by linear least squares, yielding the
Fig. 1. Example of synthetic images used as reference images for the proposed methodology: (a) real MIVIS image acquired on a coastal zone of Central Italy (Follonica), (b) the corresponding synthetic image, and (c) the distribution of RMSE between the two images for all pixels and all bands.

coefficients \( A = (a_1, a_2, \ldots, a_8) \) that better retrieve each pixel of the image.

IV. TRAINING DATA SET

Using the procedure previously described, eight reflectance synthetic images have been obtained. As an example, Fig. 1 shows a real image (a) acquired on a coastal zone of Central Italy (Follonica) and a synthetic one (b); the distribution of the root mean square error (RMSE) between the real and synthetic images is also shown (c). The synthetic images have a noise level that can be considered negligible. Therefore, they represent the noise-free images in the MNF procedure. Fig. 2 shows a comparison between the real image taken by the spectrometer MIVIS (band 88, 2.44 \( \mu \text{m} \)) and the corresponding synthetic image obtained by simulation.

### A. MIVIS Sensor

MIVIS is a scanning spectrometer flying on aircraft. It measures radiance spectrum in 102 bands ranging from the visible (VIS) to the thermal infrared (TIR); see Table I for some details concerning MIVIS. Each scanning line is composed of 755 pixels that are digitized at 12 bit per pixel. The main geometric parameters of the MIVIS acquisition system are [25] the following:

1) total field of view: \( \text{FOV} = 71.06^\circ \);
2) instantaneous field of view: \( \text{IFOV} = 2.0 \text{ mrad} \);
3) sample rate (angular step) = 1.64 mrad;
4) scan rotation frequency: 25, 16.7, 12.5, 8.3, and 6.25 Hz.

MIVIS has been used to estimate the noise. To measure the sensor sensitivity, before each acquisition, MIVIS is equipped with a test bench for the calibration (Fig. 3) composed of a box that isolates it from external light, a barium sulphate panel, and two calibration lamps. A record of 1000 scanning lines is made in dark conditions. It represents the term \( N_i \) of (1).
B. MIVIS Synthetic Images

As MNF transformation is applied to RAW data, in order to determine the signal and noise dimensions, it is necessary to obtain training images in RAW format from the corresponding reflectance synthetic images. In order to retrieve the simulated reflectance synthetic MIVIS images, from the reflectance image, and radiometric correction has been applied to the eight reflectance synthetic images. In order to retrieve the simulated images from the corresponding RAW format, in order to have a significant number of different conditions, 99 atmospheric scenarios have been obtained, one including the noise of the sensor and the other one where noise is not present. Thus, the optimal number of OCs that minimizes RMSE has been derived. It represents, for each image, the number of OCs to retain in the MNF procedure. In order to develop a quick method to practically figure out the number of components to be retained in the MNF application in real cases, we searched a possible operativity conditions of the airborne spectrometer. For each radiation image, by (7) and (8), two raw images have been obtained, one including the noise of the sensor and the other one where noise is not present. Thus, two sets of 792 DN images have been obtained.

MNF procedure has been applied in a sequential way to each image affected by noise, selecting in (3) and (4) in succession 2, 3, …, 42, 43, 55, 70, 91 principal components, for a total of 45 runs per image. For each of the 792 × 45 runs, RMSE has been computed between the image not affected by noise and the cleaned image. Fig. 4 shows the behavior of RMSE versus the number of retained components in correspondence of some values of SZA, flight height (H), and VIS. It is possible to note that RMSE always shows a minimum within the range 2–43 of the components. As it could have been expected, the number of OCs increases with the amount of radiance, due to the fact that the SNR of the images increases. Analyzing all the 792 plots, the optimal number of retained components that minimizes RMSE has been derived. It represents, for each image, the number of OCs to retain in the MNF procedure. In order to develop a quick method to practically figure out the number of components to be retained in the MNF application in real cases, we searched a possible relationship between the number of OCs and some prognostic parameters that can be extracted from the images themselves. To this purpose, the following parameters have been considered:

1) sum of the eigenvalues of the covariance matrix of the images;
2) the first eigenvalue of the covariance matrix of the images;
3) mean integrated radiance of the image, integrated on the wavelength and averaged on the pixels.

Figs. 5–7 show plots of the three extracted prognostic parameters as a function of the number of OCs for each of the
TABLE II
VALUE OF THE PARAMETERS OF THE FIT EQUATION FOR THE THREE
PROGNOSTIC PARAMETERS. THE VALUE OF THE $R^2$ THAT
QUANTIFIES THE GOODNESS OF FIT IS ALSO REPORTED

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Fit</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of eigenvalues</td>
<td>$f^1(x) = 0.1839x^{0.3838}$</td>
<td>0.8777</td>
</tr>
<tr>
<td>First eigenvalue</td>
<td>$f^2(x) = 0.2818x^{0.3537}$</td>
<td>0.8653</td>
</tr>
<tr>
<td>Average integrated radiance</td>
<td>$f^3(x) = 986.03x^{0.5391}$</td>
<td>0.8212</td>
</tr>
</tbody>
</table>

where $x$ is the number of OCs and $n$ is an index of the three investigated parameters. Results for the three parameters are shown in Table II. The sum of the eigenvalues and the first eigenvalue show the best (and similar) fit. Anyway, all three parameters seem to predict fairly well the number of OCs to be retained in MNF. Other prognostic parameters were investigated in this work, but they are not shown here for the sake of brevity.

A. Profile Likelihood

In order to compare the proposed methods with the state-of-the-art algorithms, we consider the computation of the profile likelihood from the scree plot introduced in [17]. The scree plot orderly displays the eigenvalues of the PCA analysis, and it aims at estimating the gap or the elbow in the corresponding curve that is commonly associated to the optimal number of components. Profile likelihood outperforms other competitor methods for estimating the optimal number of principal components both with respect to accuracy and computational time, due to its simplicity. In addition, it is not strictly tailored to the PCA analysis, but it is based on the general assumption of reducing the dimensionality of data via a generic dimension reduction methodology that produces a relative measure of the importance of each component, giving rise to an ordering of the resulting coordinates. For this reason, its application to the MNF framework is straightforward. Let $\hat{\delta} \equiv (\delta_1, \ldots, \delta_N)$ be the diagonal elements of the matrix $\Delta$ resulting from (2). Profile likelihood relies on the assumption that when a gap or elbow exists at a certain position of the sequence $\hat{\delta}$, for example, $\bar{k}$, the sets $S_1 \equiv \{\delta_1, \ldots, \delta_{\bar{k}}\}$ and $S_2 \equiv \{\delta_{\bar{k}+1}, \ldots, \delta_N\}$ belong to different populations. The profile log-likelihood function $l(\bar{k})$ is defined as

$$l(\bar{k}) = \sum_{i=1}^{\bar{k}} \log G(\delta_i; \mu_1, \sigma) + \sum_{i=\bar{k}+1}^{N} \log G(\delta_i; \mu_2, \sigma)$$  (10)

where it is assumed that populations underlying the samples $S_1$ and $S_2$ obey the Gaussian density function $G$, so that $G(z; \mu, \sigma)$ is its evaluation in the point $z$, and $\mu$ and $\sigma$ are the corresponding mean value and standard deviation, respectively, (we defer to [17] for full details of the method). In (10), $\mu_1$ and $\mu_2$ are estimated from the samples $S_1$ and $S_2$, respectively, whereas $\sigma$ is estimated from the whole sample $S_1 \cup S_2$. The estimate of the optimal number of components is simply found in correspondence with the maximum of the profile likelihood with respect to $\bar{k}$.  

792 images. For each parameter, a fit has been searched by LS, relying on the following function:

$$f^\mu(x) = ax^b$$  (9)
Fig. 8. RMSE distribution for the three considered prognostic parameters: Sum of the eigenvalues, the first eigenvalue, and mean integrated radiance. The solid line represents the RMSE distribution of noisy image.

B. Accuracy Evaluation

To quantify the effectiveness of the fits (9) in estimating the number of OCs, the distribution of RMSE between the image cleaned by MNF with the number of components predicted by (9) and the corresponding noise-free image has been computed. As reference, the RMSE ($\Gamma^m$) related to the number of components chosen in an optimal way (by selecting the minimum value of the RMSE curves shown in Fig. 4), which is $\Gamma^\text{OC}$, and to the images not cleaned by MNF, which is $\Gamma^\text{noise}$, is also computed. Fig. 8 shows histograms of $\Gamma^m$ for the three considered prognostic equations. They are similar to each other and also similar to the one obtained by choosing the number of principal components in an optimal way ($\Gamma^\text{OC}$). Histogram based on the profile likelihood criteria is also shown.

In order to have a further evaluation of the feasibility of (9), we computed the cumulative function of $\Gamma^m$ [28]

$$g^n(z) = \frac{\int \Gamma^m(z')dz'}{\int \Gamma^m(z')dz'}. \quad (11)$$

It represents the fraction of images having an RMSE lower than a certain value $z$ (Fig. 9). It is clear that the sooner $g(z)$ reaches the value one, the lower the error. In the ideal case, all images have null RMSE and $g^n(z) = I_0(z)$, with $I_0(z)$ being the indicator function

$$I_0(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0. \end{cases}$$

Therefore, as a reference, we computed the area $A$ above $g^n(z)$ and below $I_0(z)$. Formally, it is obtained as

$$A = \int_0^\infty (I_0(z) - g(z)) dz. \quad (12)$$

Fig. 9 shows a plot of $g$ and the area $A$ for the original noisy images, in correspondence of all the methods considered in estimating the number of MNF components (optimal, via (9), and the three reported criteria) and by the profile likelihood method (criterion (10) as developed in [17]). It is clear that the criteria of the first eigenvalue and the sum of eigenvalues outperform all the other ones. In particular, the profile likelihood is ranked lowest among all. An analysis made with box plots (not reported here for the sake of brevity) also shows that profile likelihood is the only method not able to correct some outliers that are present in the full data set of images.

VI. CONCLUSION

We have developed a procedure to estimate the number and the corresponding list of principal components to retain in the MNF transform. Even though somewhat complex from the computational point of view, it allows one to accurately separate the noise and signal components of the image. Some objective (i.e., data driven) criteria have been also proposed for the estimate of the number of optimal principal components that do not rely on any ancillary information (sum of eigenvalues of the image covariance matrix, first eigenvalue of the same matrix, and average radiance over the full image). The methodology is based on estimating noise structure by dark current acquisitions. This can be accomplished easily for sensors onboard aircraft. For satellite sensors, when no shutter is available, it is possible to use dark current data acquired for the radiometric calibration.

ACKNOWLEDGMENT

The authors would like to thank an anonymous reviewer for his useful comments that helped to improve this paper.

REFERENCES


Umberto Amato received the M.Sc. degree in physics from the University of Napoli, Naples, Italy, in 1981.

He is currently the Director of Research with the Napoli branch of the Institute for Applications of Calculus of the Italian National Research Council. His interests include the development of statistical methods for nonparametric regression, classification, and dimension reduction. They are applied to problems arising in remote sensing (analysis of images measured by sensors onboard aircraft or satellites) and medicine (magnetic resonance imaging).

Rosa Maria Cavalli received the M.Sc. degree in geology from the University of Roma “La Sapienza,” Rome, Italy.

She is currently a Researcher with the Institute for the Atmospheric Pollution, Italian National Research Council, Rome. In particular, she is responsible for the Airborne Laboratory for Environmental Research functional section. Her present research activity is in the development in different sectors of the hyperspectral airborne and satellite remote sensing, particularly covering data validation, data calibration and correction, the study and monitoring of the spectral characteristics of the atmosphere and of the land and sea vegetation, target detection, and archaeology. This activity is carried out in cooperation with other national and international research groups with whom she is involved in several projects.

Angelo Palombo received the M.Sc. degree in physics from the University of Rome, “La Sapienza,” Rome, Italy, in 2002.

He is currently with the Institute of Methodologies for Environmental Analysis, Italian National Research Council, Potenza, Italy. His research interests focus on the techniques of noise reduction and the radiometric and geometric correction of the hyperspectral images.

Stefano Pignatti received the M.Sc. degree in geology from the University of Rome “La Sapienza,” Rome, Italy, in 1988.

Since 1995, he has been a Researcher with the Institute of Methodologies for Environmental Analysis, Italian National Research Council, Potenza, Italy. His interests include data calibration and analysis of hyperspectral remote sensing data in the field of environmental applications.

Federico Santini received the M.Sc. degree in physics from the University of Roma “La Sapienza,” Rome, Italy, in 1998 and the Ph.D. degree in environmental engineering from the University of Basilicata, Potenza, Italy, in 2007.

Currently, he is a Researcher with the functional section at the Airborne Laboratory for Environmental Research, Institute for Atmospheric Pollution, Italian National Research Council (CNR), Rome. His present research activity is in the development of physical models for remotely sensed data interpretation on optically complex water bodies. This activity is carried out in cooperation with CNR Marine Science Institute, CNR Institute of Methodologies for Environmental Analysis, the San Marco Project Research Centre of the University “La Sapienza” of Rome, and the Southampton Oceanography Centre of Southampton, U.K.