

# EXPLICIT SIGNAL TO NOISE RATIO IN REPRODUCING KERNEL HILBERT SPACES

Luis Gómez-Chova<sup>1</sup>, Allan A. Nielsen<sup>2</sup> and Gustavo Camps-Valls<sup>1</sup>

<sup>1</sup>Image Processing Laboratory (IPL). Universitat de València, Spain.

<sup>2</sup>DTU Space - National Space Institute. Technical University of Denmark.

## ABSTRACT

This paper introduces a nonlinear feature extraction method based on kernels for remote sensing data analysis. The proposed approach is based on the minimum noise fraction (MNF) transform, which maximizes the signal variance while also minimizing the estimated noise variance. We here propose an alternative kernel MNF (KMNF) in which the noise is explicitly estimated in the reproducing kernel Hilbert space. This enables KMNF dealing with non-linear relations between the noise and the signal features jointly. Results show that the proposed KMNF provides the most noise-free features when confronted with PCA, MNF, KPCA, and the previous version of KMNF. Extracted features with the explicit KMNF also improve hyperspectral image classification.

**Index Terms**— Kernel methods, signal to noise ratio, kernel principal component analysis, kernel minimum noise fraction, feature extraction

## 1. INTRODUCTION

Feature extraction methods are used to create a subset of new features by linear or nonlinear combinations of the existing ones. Linear feature extractors have been extensively used for remote sensing data analysis. Among these methods, an approach that is becoming progressively more popular is the minimum noise fraction (MNF) transform [1], which extends principal component analysis (PCA) by maximizing the signal variance while also minimizing the estimated noise variance. In recent years, kernel methods have emerged as an excellent tool to develop nonlinear feature extraction methods [2]. The kernel MNF (KMNF) is the standard kernelization of the canonical MNF in which noise is estimated in the original input space and then both signal and noise are transformed via suitable mappings endorsed with the reproducing kernel property [3]. In this paper, we propose an alternative KMNF formulation in which the noise is explicitly estimated in the reproducing kernel Hilbert space. This simplifies the formulation and enables KMNF dealing with non-linear relations between the noise and the signal features jointly.

This paper has been partially supported by the Spanish Ministry for Science under projects AYA2008-05965-C04-03 and CSD2007-00018.

## 2. NONLINEAR FEATURE EXTRACTION WITH KERNELS

This section presents a kernel method for the nonlinear extraction of features that maximizes the signal to noise ratio. We propose a formulation in which the noise is estimated directly in the kernel space and then a transform that maximizes the signal covariance while minimizes the estimated noise covariance is found. First, a brief introduction of standard linear and kernel feature extraction methods is given. Then the novel KMNF formulation is presented.

Notationally, we are given a set of  $n$  training feature vectors  $\mathbf{x}_i \in \mathbb{R}^N$  in the input space (i.e.  $N$  spectral channels or bands). This can be also expressed using matrix notation,  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$ , where  $^\top$  denotes matrix transposition,  $\tilde{\mathbf{X}}$  indicates the centered version of  $\mathbf{X}$ , and  $\mathbf{C}_{xx} = \frac{1}{n} \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$  represents the empirical covariance matrix of the input data. In this context, linear feature extraction can be carried out by projecting the data into the subspace characterized by the projection matrix  $\mathbf{U}$ , of size  $N \times n_p$ , so that the  $n_p$  extracted features of the original data are given by  $\tilde{\mathbf{X}}' = \tilde{\mathbf{X}}\mathbf{U}$ .

### 2.1. Minimum Noise Fraction (MNF)

The principal component analysis (PCA) projects linearly the input data onto the directions of largest input variance [4]. Therefore, to perform PCA, one has to solve:

$$\begin{aligned} \text{PCA: } \quad \mathbf{U} &= \arg \max_{\mathbf{U}} \{ \text{Tr}(\mathbf{U}^\top \mathbf{C}_{xx} \mathbf{U}) \} \\ &\text{subject to } \mathbf{U}^\top \mathbf{U} = \mathbf{I}, \end{aligned} \quad (1)$$

where  $\mathbf{I}$  is the identity matrix of size  $n_p \times n_p$ . This can also be expressed, using Lagrange multipliers, as the eigenvalue problem  $\mathbf{C}_{xx} \mathbf{u}_i = \lambda_i \mathbf{u}_i$  (or singular value decomposition of  $\mathbf{C}_{xx}$ ), which yields a set of sorted eigenvalues  $\{\lambda_i\}_{i=1}^{n_p}$  ( $\lambda_i \leq \lambda_{i+1}$ ) and the corresponding eigenvectors  $\{\mathbf{u}_i\}_{i=1}^{n_p}$ . The variance of the projected data  $\tilde{\mathbf{X}}'$  equals the eigenvalues.

The main limitation of PCA is that it does not consider the characteristics of the noise present in the input vectors. PCA simply performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance of the original data distribution and assumes that the noise variance is low corresponding to the last eigenvectors. Therefore, there

is no guarantee that the directions of maximum variance will not be affected by the variance of the data noise.

Assuming that we had access to the signal  $\mathbf{S}$  and the noise  $\mathbf{N}$ , the ideal objective would be to maximize the signal to noise ratio (SNR), i.e. the ratio between the signal and the noise variances for all the features:

$$\text{SNR: } \mathbf{U} = \arg \max_{\mathbf{U}} \left\{ \text{Tr} \left( \frac{\mathbf{U}^\top \mathbf{C}_{ss} \mathbf{U}}{\mathbf{U}^\top \mathbf{C}_{nn} \mathbf{U}} \right) \right\} \quad (2)$$

subject to  $\mathbf{U}^\top \mathbf{C}_{nn} \mathbf{U} = \mathbf{I}$ .

However, neither the signal nor the noise covariance matrices,  $\mathbf{C}_{ss}$  and  $\mathbf{C}_{nn}$ , are known.

The MNF transform assumes that dataset  $\mathbf{X}$  can be ideally split in  $\mathbf{X} = \mathbf{S} + \mathbf{N}$  and that the signal and the noise are mutually orthogonal  $\mathbf{S}^\top \mathbf{N} = \mathbf{N}^\top \mathbf{S} = \mathbf{0}$ . Then, maximizing the SNR is equivalent to minimizing the noise fraction,  $\text{NF} = 1/(\text{SNR}+1)$ :

$$\text{MNF: } \mathbf{U} = \arg \max_{\mathbf{U}} \left\{ \text{Tr} \left( \frac{\mathbf{U}^\top \mathbf{C}_{xx} \mathbf{U}}{\mathbf{U}^\top \mathbf{C}_{nn} \mathbf{U}} \right) \right\} \quad (3)$$

subject to  $\mathbf{U}^\top \mathbf{C}_{nn} \mathbf{U} = \mathbf{I}$ ,

which gives rise to the generalized eigenproblem  $\mathbf{C}_{xx} \mathbf{u}_i = \lambda_i \mathbf{C}_{nn} \mathbf{u}_i$ . It is worth noting that, in this case, since  $\mathbf{U}^\top \mathbf{C}_{nn} \mathbf{U} = \mathbf{I}$ , the eigenvalues are equal to the data variance and to the  $\text{SNR}+1$  in the projected space.

However, the main problem of MNF is that of obtaining a good estimation of the noise covariance matrix  $\mathbf{C}_{nn} = \frac{1}{n} \mathbf{N}^\top \mathbf{N}$ . For remote sensing images, the noise is usually obtained as the difference between the actual pixel value and a reference 'clean' value,  $\mathbf{N} = \mathbf{X} - \mathbf{X}_r$ . The reference signal  $\mathbf{X}_r$  is estimated from its neighborhood assuming that the signal is spatially smoother than the noise (e.g. taking as reference the mean of the values in a spatial neighborhood).

## 2.2. Kernel Minimum Noise Fraction (KMNF) methods

The previous method assumes that the best extracted features,  $\tilde{\mathbf{X}}'$ , explaining data distribution and minimizing the noise variance have a linear relation with the original data matrix,  $\tilde{\mathbf{X}}$ . However, in many situations this linearity assumption is not satisfied, and nonlinear feature extraction is needed to obtain acceptable performance. In this context, *kernel methods* are a promising approach, as they constitute an excellent framework to formulate nonlinear versions from linear algorithms [2, 5]. In this section, we describe the proposed kernel MNF (KMNF) formulations.

Notationally, consider a nonlinear function  $\phi(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathcal{H}$  that maps the input data into some *kernel space* of very large or even infinite dimension,  $\{\phi(\mathbf{x}_i)\}_{i=1}^n$ . The data matrix for performing the linear feature extraction (PCA or MNF) in  $\mathcal{H}$  is now given by  $\tilde{\Phi} = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_n)]^\top$ . As before, the centered versions of this matrix is denoted by  $\tilde{\tilde{\Phi}}$ . The projection of the input data will be given by  $\tilde{\tilde{\Phi}}' = \tilde{\tilde{\Phi}} \mathbf{U}$ , where the projection matrix  $\mathbf{U}$  is now of size  $\dim(\mathcal{H}) \times n_p$ . Note,

that the input covariance matrix in  $\mathcal{H}$ , which is usually needed by the different methods, becomes of size  $\dim(\mathcal{H}) \times \dim(\mathcal{H})$  and cannot be directly computed. However, making use of the *representer's theorem* [5], we can introduce  $\mathbf{U} = \tilde{\tilde{\Phi}}^\top \mathbf{A}$  into the formulation, where  $\mathbf{A} = [\alpha_1, \dots, \alpha_{n_p}]$  and  $\alpha_i$  is an  $n$ -length column vector containing the coefficients for the  $i$ th projection vector, and the maximization problem can be reformulated solely in terms of the kernel matrix. Note that, in these kernel feature extraction methods, the projection matrix  $\mathbf{U}$  in  $\mathcal{H}$  might not be explicitly calculated, but the projections of the input data can be obtained. Therefore, the extracted features for a new input pattern  $\mathbf{x}_*$  are given by:

$$\begin{aligned} \tilde{\tilde{\Phi}}'(\mathbf{x}_*)^\top &= \tilde{\tilde{\Phi}}(\mathbf{x}_*)^\top \mathbf{U} = \tilde{\tilde{\Phi}}(\mathbf{x}_*)^\top \tilde{\tilde{\Phi}}^\top \mathbf{A} \\ &= [\tilde{K}(\mathbf{x}_*, \mathbf{x}_1), \dots, \tilde{K}(\mathbf{x}_*, \mathbf{x}_n)] \mathbf{A}, \end{aligned} \quad (4)$$

which is expressed in terms of the inner products in the centered feature space that, as in all kernel methods, can be computed via a positive semidefinite Mercer kernel function  $\tilde{K}(\mathbf{x}_i, \mathbf{x}_j) = \tilde{\tilde{\Phi}}(\mathbf{x}_i)^\top \tilde{\tilde{\Phi}}(\mathbf{x}_j)$ .

As in the linear case, the aim of KMNF is to find directions of maximum signal to noise ratio of the input data projected in  $\mathcal{H}$ , which can be obtained by replacing  $\tilde{\mathbf{X}}$  by  $\tilde{\tilde{\Phi}}$  in (3), i.e. by replacing  $\mathbf{C}_{xx}$  by  $\tilde{\tilde{\Phi}}^\top \tilde{\tilde{\Phi}}$  and  $\mathbf{C}_{nn}$  by  $\tilde{\tilde{\Phi}}_n^\top \tilde{\tilde{\Phi}}_n$ :

$$\text{KMNF: } \mathbf{U} = \arg \max_{\mathbf{U}} \left\{ \text{Tr} \left( \frac{\mathbf{U}^\top \tilde{\tilde{\Phi}}^\top \tilde{\tilde{\Phi}} \mathbf{U}}{\mathbf{U}^\top \tilde{\tilde{\Phi}}_n^\top \tilde{\tilde{\Phi}}_n \mathbf{U}} \right) \right\} \quad (5)$$

subject to  $\mathbf{U}^\top \tilde{\tilde{\Phi}}_n^\top \tilde{\tilde{\Phi}}_n \mathbf{U} = \mathbf{I}$ .

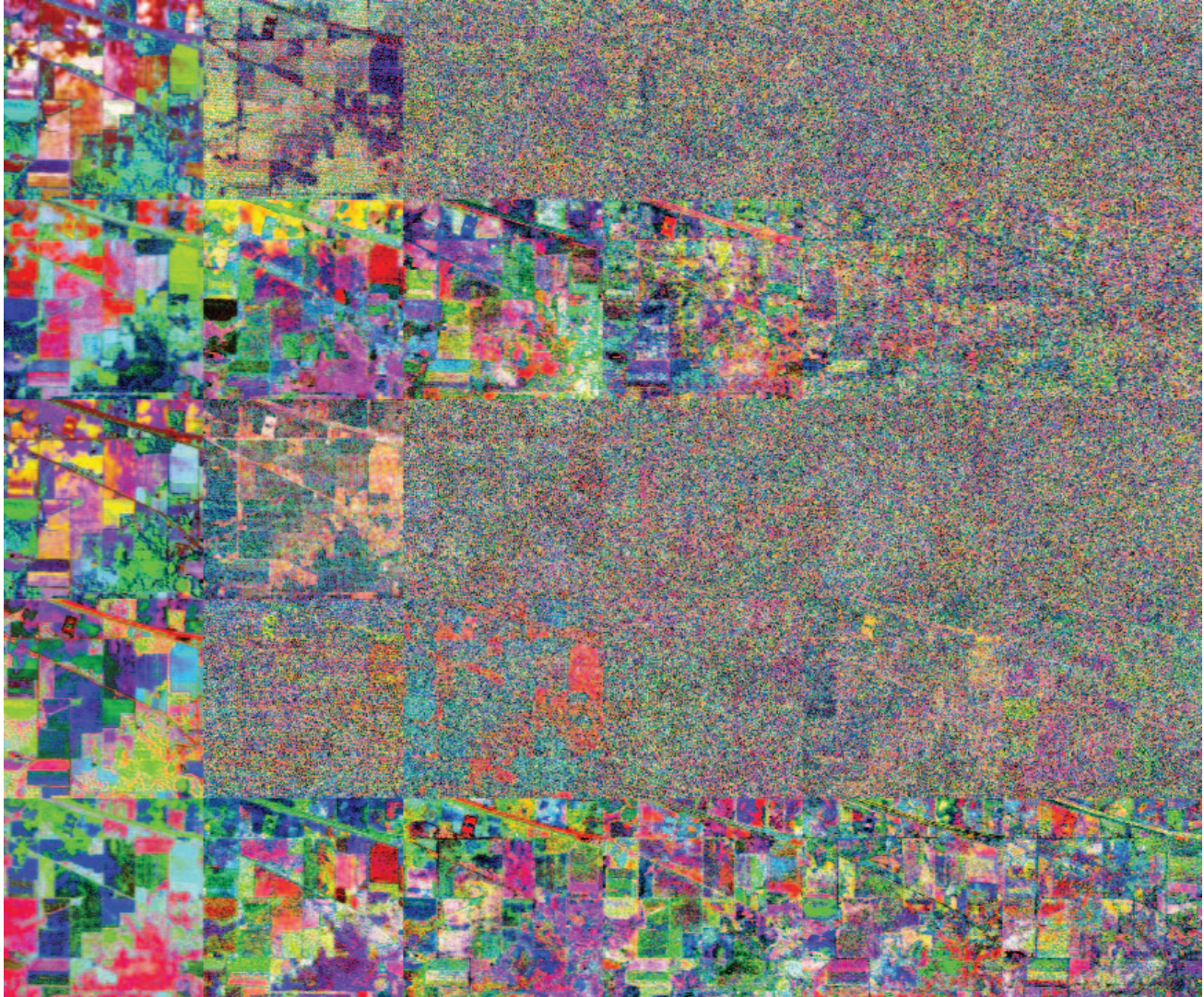
Making use of the representer's theorem one can introduce  $\mathbf{U} = \tilde{\tilde{\Phi}}^\top \mathbf{A}$  into the previous formulation

$$\begin{aligned} \text{KMNF: } \mathbf{A} &= \arg \max_{\mathbf{A}} \left\{ \text{Tr} \left( \frac{\mathbf{A}^\top \tilde{\tilde{\Phi}}^\top \tilde{\tilde{\Phi}} \tilde{\tilde{\Phi}}^\top \mathbf{A}}{\mathbf{A}^\top \tilde{\tilde{\Phi}}_n^\top \tilde{\tilde{\Phi}}_n \tilde{\tilde{\Phi}}_n^\top \mathbf{A}} \right) \right\} \\ &= \arg \max_{\mathbf{A}} \left\{ \text{Tr} \left( \frac{\mathbf{A}^\top \tilde{K}_{xx}^2 \mathbf{A}}{\mathbf{A}^\top \tilde{K}_{xn} \tilde{K}_{nx} \mathbf{A}} \right) \right\} \quad (6) \\ &\text{subject to } \mathbf{A}^\top \tilde{K}_{xn} \tilde{K}_{nx} \mathbf{A} = \mathbf{I}, \end{aligned}$$

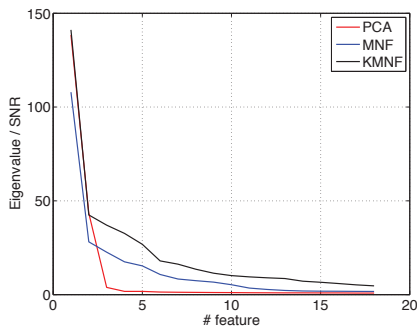
where we have defined the symmetric centered kernel matrix  $\tilde{K}_{xx} = \tilde{\tilde{\Phi}} \tilde{\tilde{\Phi}}^\top$  containing the inner products between any two points in the kernel space, and the non-symmetric kernel  $\tilde{K}_{xn} = \tilde{\tilde{\Phi}} \tilde{\tilde{\Phi}}_n^\top = \tilde{K}_{nx}^\top$  containing the inner products between the data and the noise in the kernel space. The solution to the above problem can be obtained from the generalized eigenproblem

$$\tilde{K}_{xx}^2 \alpha_i = \lambda_i \tilde{K}_{xn} \tilde{K}_{nx}^\top \alpha_i. \quad (7)$$

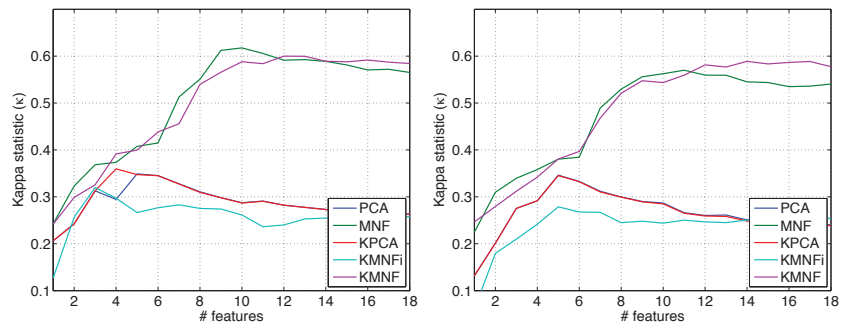
The problem of estimating the noise to compute  $\tilde{K}_{xn}$  arises. The approach in [3] estimated the noise in the input space as explained before,  $\mathbf{N} = \mathbf{X} - \mathbf{X}_r$ , and the signal-to-noise kernel is then computed as  $\tilde{K}_{xn} = \tilde{\tilde{\Phi}} \tilde{\tilde{\Phi}}_n^\top$  with  $\tilde{\tilde{\Phi}}_n^\top = [\tilde{\phi}(\mathbf{n}_1), \dots, \tilde{\phi}(\mathbf{n}_n)]$ .



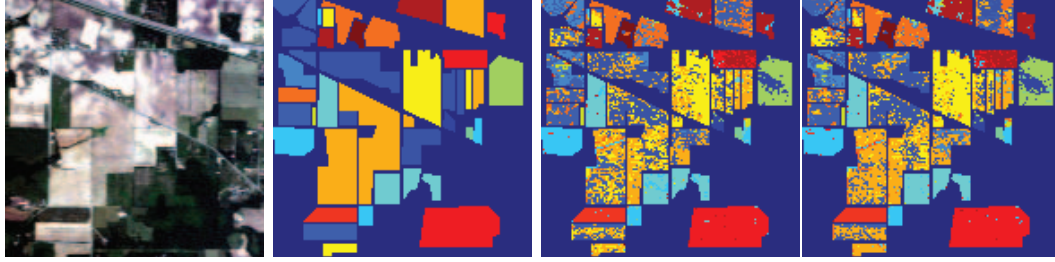
**Fig. 1.** Extracted features from the original image. From top to bottom: PCA, MNF, KPCA, standard KMNF, and explicit KMNF in the kernel space for the first 18 principal components. From left to right: each subimage shows the RGB composite of 3 components ordered in descending importance.



**Fig. 2.** Eigenvalues (variance and SNR) of the transformed data.



**Fig. 3.** Classification accuracy (kappa statistics,  $\kappa$ ) as a function of the number of used features extracted with different methods. (a) Original hyperspectral image. (b) Image corrupted with multiplicative random noise (10%).



**Fig. 4.** From left to right: (a) RGB composite of the hyperspectral image (bright bare soils and dark vegetated crops); (b) Ground Truth of the 16 land-cover classes; and (c) and (d) classification maps using the MNF and KMNF features, respectively.

The previous approach has a clear shortcoming. Note that in the previous KMNF formulation, two kind of kernels need to be computed thus dealing with objects of different nature and hence, ideally one should tune different kernel hyperparameters for each one of them. This implies that, by using different kernels, one is mapping signal and noise to different feature spaces and hence the extracted eigenvalues have no longer the meaning of SNR. We propose here to estimate the noise explicitly in the Hilbert space defining  $\tilde{\Phi}_n = \tilde{\Phi} - \tilde{\Phi}_r$  that results in a noise kernel  $\tilde{\mathbf{K}}_{xn} = \tilde{\Phi}\tilde{\Phi}_n^\top = \tilde{\Phi}(\tilde{\Phi} - \tilde{\Phi}_r)^\top = \tilde{\Phi}\tilde{\Phi}^\top - \tilde{\Phi}\tilde{\Phi}_r^\top = \tilde{\mathbf{K}}_{xx} - \tilde{\mathbf{K}}_{xr}$ . For example, if the reference is obtained as the average of the 4-connected neighboring pixels, then  $\tilde{\mathbf{K}}_{xn} = \tilde{\mathbf{K}}_{xx} - 1/4 \sum \tilde{\mathbf{K}}_{xr_{i,j}}$ . Then, this signal-to-noise kernel is used in the generalized eigenproblem (7). This formulation has clear advantages: 1) since both  $\mathbf{X}$  and  $\mathbf{X}_r$  are the same type of data the kernel function used to compute  $\tilde{\mathbf{K}}_{xx}$  and  $\tilde{\mathbf{K}}_{xr}$  might be the same; 2) the number of free parameters is thus lower; 3) data and noise are mapped in the same Hilbert space and thus the obtained eigenvalues can be interpreted as data variance and as the SNR in the projected space; and 4) it is possible to deal with non-linear relations between the noise and the signal and with non-additive noise.

### 3. EXPERIMENTAL RESULTS

This section is devoted to the design and application of different feature extraction approaches based on maximization of the signal variance through some linear and nonlinear transforms. The classical 220-bands AVIRIS hyperspectral image taken over Indiana’s Indian Pine test site in June 1992 is used in the experiments. The image is  $145 \times 145$  pixels, contains 16 crop types classes, and a total of 10366 labeled pixels. This image is a classical benchmark to validate model accuracy and constitutes a very challenging classification problem because of the strong mixture of the classes’ signatures.

In this work, we transform all the 220 original bands into a lower dimensional space of 18 features. It is worth noting that 20 bands covering the region of water absorption are really noisy, thus allowing us to analyze the robustness of the different feature extraction methods to real noise. The significance of the results for the different methods is analyzed in different ways:

1. Visual inspection of the extracted features in descending order of relevance (Fig. 1). The proposed approach provides more noise-free features.
2. Analysis of the eigenvalues of the transformed data, which represent signal variance for PCA and SNR for MNF and KMNF. The proposed approach provides the highest SNR (Fig. 2).
3. Land cover classification accuracy as the number of used extracted features increases. The best feature extraction methods are the linear MNF and the proposed KMNF (Fig. 3a). The proposed KMNF method outperforms MNF when the image is corrupted with non additive noise (Fig. 3b).
4. Visual inspection of the classification maps obtained with a linear discriminant analysis (LDA) classifier using the best sets of extracted features. The proposed approach provides more spatially homogeneous land cover maps than the other methods (Fig. 4).

### 4. CONCLUSIONS

This paper presented a kernel method for nonlinear feature extraction that maximizes the SNR in remote sensing images. The method has good theoretical and practical properties for extracting features in noisy situations.

### 5. REFERENCES

- [1] 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 Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1998.
- [2] G. Camps-Valls and L. Bruzzone, Eds., *Kernel methods for Remote Sensing Data Analysis*, Wiley & Sons, UK, Dec 2009.
- [3] A. A. Nielsen, “Kernel maximum autocorrelation factor and minimum noise fraction transformations,” *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 612–624, Mar. 2011.
- [4] I. T. Jolliffe, *Principal Component Analysis*, Springer, 1986.
- [5] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.