Extraction of Endmembers From Hyperspectral Images Using A Weighted Fuzzy Purified-Means Clustering Model

Linlin Xu, *Member, IEEE*, Alexander Wong, *Member, IEEE*, Fan Li, and David A. Clausi, *Senior Member, IEEE*

***Abstract*—Hyperspectral endmembers are the spectra of pure materials that are responsible for generating the mixed pixels in hyperspectral images (HSIs). Hyperspectral endmember extraction (HEE) is essentially an inverse problem, where the unknown endmembers are inferred from the spectral measurements. Efficient extraction of endmembers in HSI relies on a well-defined generative model that captures key factors in HSI generation process, such as the clustering effect in the spatial domain and the noise heterogeneity effect in the spectral domain. This paper presents a weighted fuzzy purified-means (WFP-means) clustering model for HEE, where the endmembers are modeled as mean vectors of individual classes, and the fractional contributions of individual endmembers, called abundances, are treated as soft class membership. Accordingly, an endmember is estimated as the weighted mean of purified pixels in HSI, while the abundances are estimated as the nonnegative regression coefficients. In contrast to a mixed pixel that consists of multiple endmembers, a "purified pixel" is due to a single endmember. The introduction of the concept of "purified pixels" into the fuzzy clustering model leads to an elegant optimization scheme. Moreover, the proposed model accounts for the noise variance heterogeneity issue, which is essential for achieving unbiased abundance estimation. The proposed method is tested on both simulated and real HSI, in comparison with several other HEE methods. The results demonstrate that the proposed method compares favorably with respect to the referenced methods in terms of both endmember and abundance estimation.**

***Index Terms*—Endmember extraction, fuzzy C-means, K-means, spectral unmixing, weighted fuzzy purified means.**

I. Introduction

HYPERSPECTRAL image (HSI) captured by imaging spectrometer has been widely used in a variety of applications, due to the very high spectral resolution. Nevertheless, the observed spectral pixel value in HSI always involves the spectral contribution of multiple materials within the ground instantaneous field of view of imaging instrument [1]. The determination of the spectra of pure materials, usually known as endmembers, is essential for hyperspectral unmixing (HU) that aims to estimate for each pixel in HSI the fractional contributions of individual endmembers, called the abundances of endmembers [2], [3]. In the last two decades, many different HU and hyperspectral endmember extraction (HEE) approaches have been proposed [3], including 1) geometrical-based approaches, e.g., vertex component analysis (VCA) [4], MVC-NMF [5], the simplex identification via variable splitting and augmented Lagrangian (SISAL) [6], N-FINDR [7], pixel purity index (PPI) [8], [9], iterative constrained endmembers (ICE) [10], and the minimum volume simplex analysis (MVSA) [11]; 2) statistical Bayesian approaches, e.g., [12]–[16]; and 3) sparse representation-based approaches, e.g., [17]–[20]. However, effective HU/HEE remains unresolved issue, because it faces important challenges [3], [21], [22]. In the following, we will highlight several aspects of the issue that have not been sufficiently addressed.

This is still a lack of statistical generative clustering model that accounts for the data generation mechanism of mixed pixels. In HSI, mixed pixels belonging to the same class tend to assume similar abundance pattern of endmembers, while those of different classes tend to admit different material compositions. In order to build effective data inversion model, some approaches take advantage of this source heterogeneity effect. For example, in [23]–[27], multiple convex regions in HSI are simultaneously modeled by a fuzzy clustering approach. In these approaches, each class defined by a convex region corresponds to a land cover type. Therefore, a class is required to contain multiple endmembers, and different classes may share some common endmembers. However, in HU and HEE, endmembers are treated as individual spatial processes. A more fundamental approach is thereby to model each endmember in HSI as a class. Moreover, since a mixed pixel contains the contributions of multiple endmembers, it should be allowed to belong to multiple classes based on the soft class membership. In this way, the class composition of a pixel is directly linked to the abundance pattern of endmembers on this pixel. Therefore, unsupervised HEE can be interpreted as a soft clustering issue. By estimating both the mean vectors of classes and the soft class membership for each pixels, both the endmembers and the abundances can be estimated.

Although pure pixels in HSI is essential for HU and HEE, there have not been enough efforts to explicitly estimate pure pixels from mixed pixels. Many geometrical approaches, e.g., VCA [4], N-FINDR [7], and PPI [8], [9], are based on the

Manuscript received November 25, 2014; revised June 10, 2015; accepted June 18, 2015. Date of publication August 05, 2015; date of current version February 09, 2016. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), in part by the Canadian Space Agency (CSA), and in part by the Canada Research Chairs program.

The authors are with the Department of Systems Design Engineering, University of Wateroo, Waterloo, ON N2L3G1, Canada (e-mail: l44xu@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2015.2450499

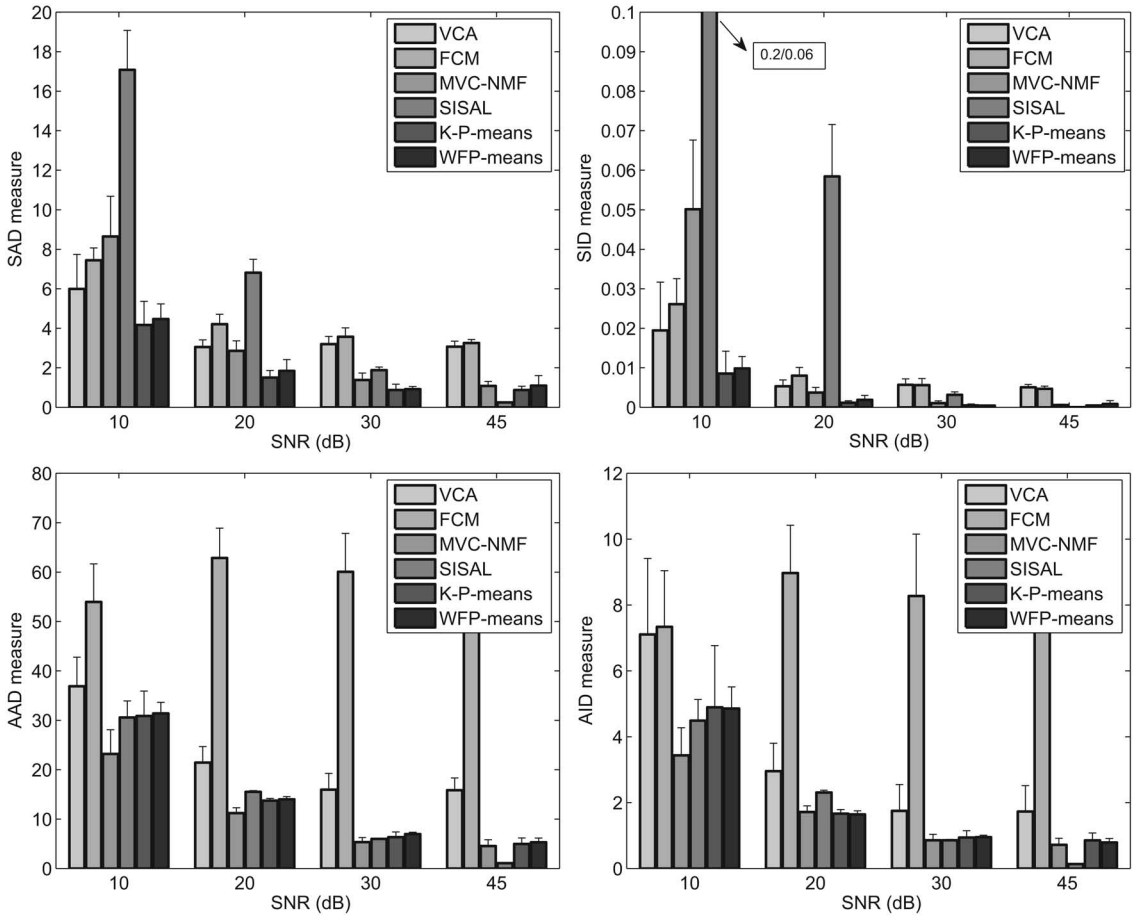


Fig. 1. Statistics (i.e., SAD, SID, AAD, and AID) achieved by different methods over different noise levels measured by SNR. The text "0.2/0.006" in the figure indicates that the mean value and standard deviation of SISAL when SNR $= 10$ are, respectively, 0.2 and 0.06. For all statistics, smaller values indicate better results. The proposed WFP-means algorithm achieves comparable results with the other advanced methods, e.g., K-P-means and MVC-NMF. The WFP-means and K-P-means are less sensitive to noise level variation than the other methods, i.e., SISAL and MVC-NMF. FCM achieves the lowest performance based on overall evaluation.

assumption that there is at least one pure pixel in HSI for each endmember. However, this assumption is not true in many datasets. In order to address HSI without pure pixels, the minimum volume approaches, e.g., MVC-NMF [5] and ICE [10], estimate endmembers by identifying a simplex of minimum size containing the data. However, these approaches could not achieve their optimal performance when the pixels in HSI are highly mixed [3]. The statistical approaches do not rely on the pure pixel assumption, because they model the endmembers and abundances using proper prior distributions. However, they require large computational cost, when dealing with practical data of large size. With respect to these difficulties, one alternative approach is to explicitly estimate the "purified pixels" from mixed pixels in HSI. A "purified pixel" of a particular endmember is defined as the residual after removing the spectral contribution of the other endmembers from the mixed pixel. If the purified pixels that are due to a particular endmember can be identified, the estimation of this endmember is an easy task, which, e.g., can be achieved by replacing the endmember with the mean value of purified pixels. This approach is investigated in this paper to achieve an efficient HEE model. Based on our previous research on the K-P-means algorithm [28], we propose a fuzzy version of K-P-means for addressing this issue.

The noise variance heterogeneity effect has not been fully addressed in the literature. HSI is inevitably contaminated by noise during the data acquisition process. Since the spectral signatures of different pure materials are highly correlated, the HU approaches are sensitive to noise [3]. Therefore, the success of HU and HEE approaches depends on their effectiveness in resisting the noise effect. Some geometric approaches, e.g., MVC-NMF [5], suppress noise beforehand using PCA. Some other approaches project data onto signal subspace to reduce noise influence. The statistical approaches deal with noise by explicitly modeling noise distribution. However, most of these methods assume that different bands in HSI contain the same degree of noise. This is a strong assumption and not true in most cases, considering the different imaging mechanism of different bands and the presence of junk bands [29]. Therefore, it is necessary to design HEE approach that is capable of addressing noise variance heterogeneity.

Motivated by the above discussions, the paper therefore presents a weighted fuzzy purified-means (WFP-means) clustering model for HEE. The proposed model interprets HEE as a soft clustering problem, where each class corresponds to an endmember, and the soft class memberships correspond to the abundances of endmembers. This interpretation is further

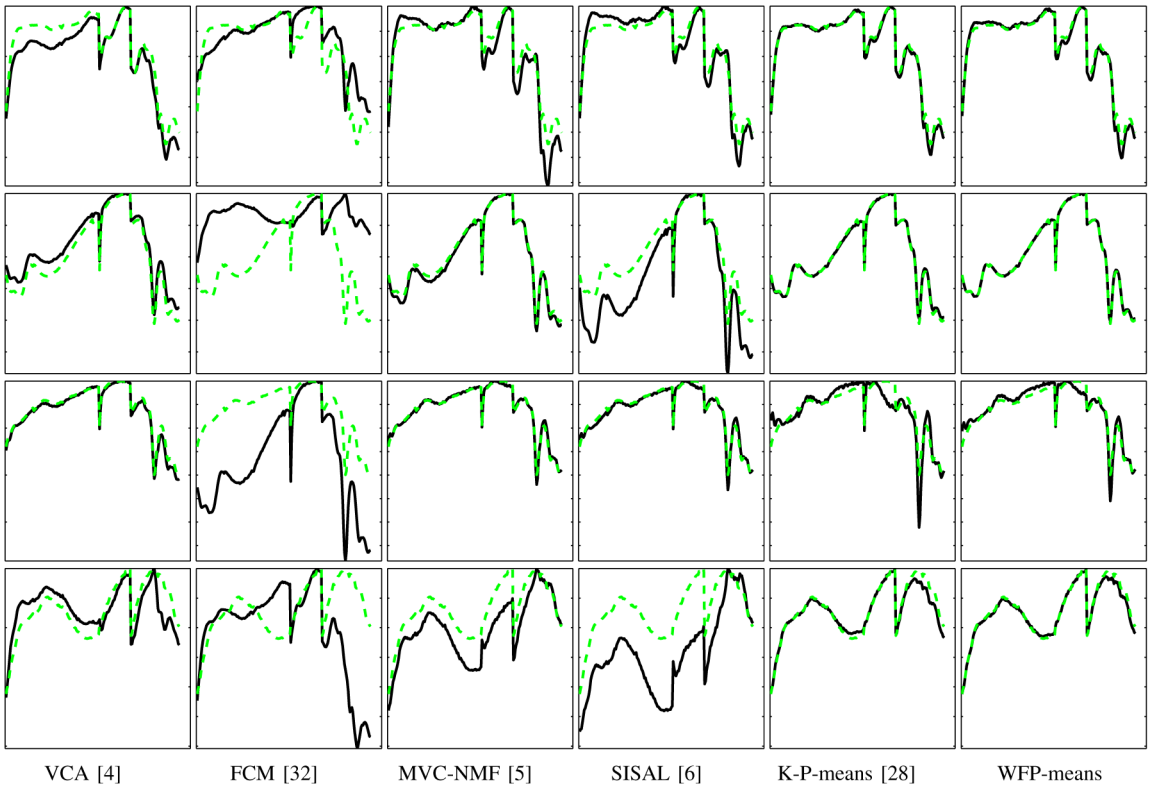


Fig. 2. True endmembers (green line) and estimated endmembers (black line) achieved by different methods. Each column displays the four endmembers of a method by row. WFP-means and K-P-means yield endmembers that are most similar to the true values. The endmembers in the bottom row achieved by SISAL and MVC-NMF do not match the true endmember very well. FCM fails to determine the endmembers from HSI.

combined with the ideas of purified pixels estimation and the heterogeneous noise adjustment for efficient HEE. For model optimization, the endmembers are estimated as the weighted means of the purified pixels in the same class, called "purified means," while the abundances of endmembers at a pixel are estimated as the coefficients of weighted nonnegative regression model. The proposed WFP-means is iteratively optimized by alternating between the estimation of endmembers and the estimation of abundances. The contributions of this paper lie in the following aspects.

1) A novel statistical generative fuzzy-membership clustering model is introduced for HEE. In the proposed WFP-means algorithm, the simultaneous modeling of endmembers, abundances, and heterogenous noise constitutes a sound modeling framework. Since the proposed model captures the data generation mechanism described in linear spectral mixture model (LSMM), the abundances and endmembers can be elegantly integrated into a fuzzy clustering framework. Moreover, the proposed model is compact comparing with other statistical approaches that relies on additional model definitions and parameters.
2) The noise variance heterogeneity effect is addressed. Since different spectral bands tend to have different noise levels, efficient HEE models need to accommodate the noise heterogeneity effect, and thereby resist noise influence on endmember estimation. The proposed WFP-means model is capable of incorporating the statistical description of noise characteristics. In model optimization, a weighted least square regression is adopted for abundance estimation, in order to account for the noise heterogeneity effect.
3) The modeling of endmembers as weighted expectations of purified pixels is introduced. The proposed WFP-means model is based on the concept of "purified pixels." Comparing with a "mixed" pixel in HSI that is usually treated as a linear combination of endmembers, a "purified" pixel is due to the sole contribution of a particular endmember. A "purified" pixel for an endmember can be obtained by removing the contributions of the other endmembers from the mixed pixel. The resulting purified pixels can be treated as realizations of the endmember subject to different weights and random noise. Therefore, given the purified pixels, the endmember can be estimated as the weight mean of the purified pixels.

Comparing with our previously proposed K-P-means algorithm that also adopted the purified means idea, WFP-means algorithm proposed in this paper provides necessary and important improvements and interpretational advantages. First, WFP-means addresses the noise heterogeneity effect, which is ignored in the K-P-means algorithm. The experiments demonstrate that WFP-means can better reduce the bias caused by noise variance variation. Second, WFP-means does not rely on the discrete/hard class membership by identifying dominant endmembers, but use the soft class membership of all pixels for estimating the endmembers. Therefore, WFP-means is more capable of identifying the endmember that does not admit predominant presence on any pixel in the image. Third, according to WFP-means, LSMM innately defines a clustering model,

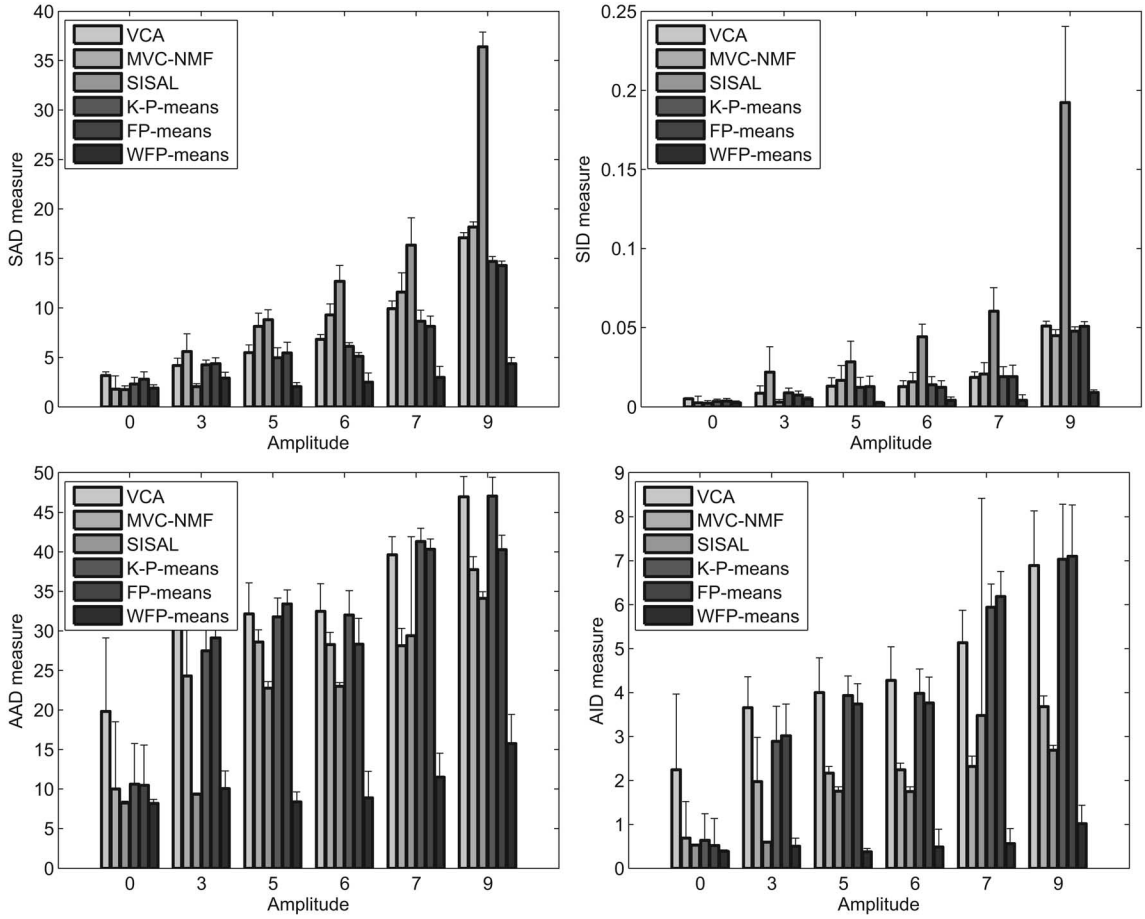


Fig. 3. Statistics (i.e., SAD, SID, AAD, and AID) achieved by different methods over different degrees of noise variance heterogeneity measured by the amplitudes. For all statistics, smaller value indicates better result. Overall, WFP-means produces the lowest values on all measures over all amplitude values. Moreover, WFP-means is insensitive to the changing of amplitude values, indicating that WFP-means is capable of resisting the noise variance heterogeneity effect. The statistic values of all other method increase fast with the increase of amplitudes. In particular, SISAL is most sensitive method, in terms of SAD and SID.

which provides further insights into the identified problems. Finally, WFP-means provides an alternative purified mean approach for estimating the endmembers, which is proved by simulated experiment to be able to achieve comparable performance with K-P-means, when the noise heterogeneous effect in HSI is ignored.

The rest of this paper is organized as follows. Section II introduces the proposed methodology. Section III conducts experiments on both simulated and real HSI. Section IV concludes the study.

II. Methodology

A. Problem Formulation

This paper addresses HEE based on LSMM with Gaussian noise. LSMM is a widely used approach to identify and quantify materials in remote sensing imagery [30]. In LSMM, the spectral signature of a mixed pixel at position i, denoted by $\boldsymbol{x}_i$, can be expressed as a linear combination of the spectra of K pure materials (i.e., endmembers), weighted by fractional abundances

$$\boldsymbol{x}_i = \sum_{k=1}^{K} \boldsymbol{a}_k s_{ik} + \boldsymbol{n}_i \tag{1}$$

where $\boldsymbol{a}_k$ is the kth endmember, s_{ik} is the abundance of $\boldsymbol{a}_k$ at pixel i, which is usually required to satisfy the sum-to-one and nonnegativity constraints

$$\sum_k s_{ik} = 1 \quad \text{and} \quad \forall\, s_{ik} \geq 0 \tag{2}$$

and $\boldsymbol{n}_i$ is the noise vector at pixel i, which satisfies a zero-mean Gaussian distribution

$$p(\boldsymbol{n}_i) = \frac{1}{\sqrt{(2\pi)^d\,|\boldsymbol{\Lambda}|}} exp\left(-\frac{1}{2}\boldsymbol{n}_i^T \boldsymbol{\Lambda}^{-1}\boldsymbol{n}_i)\right) \tag{3}$$

where $\boldsymbol{\Lambda}$ is the noise covariance matrix and d is the number of spectral bands. Based on (1) and (3), the conditional distribution of $\boldsymbol{x}_i$ can be expressed as

$$p(\boldsymbol{x}_i|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^d\,|\boldsymbol{\Lambda}|}} \cdot exp\left(-\frac{1}{2}\left(\boldsymbol{x}_i - \sum_k \boldsymbol{a}_k s_{ik}\right)^T \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{x}_i - \sum_k \boldsymbol{a}_k s_{ik})\right)\right) \tag{4}$$

where $\boldsymbol{\theta}$ includes $\{\boldsymbol{a}_k\}$, $\{s_{ik}\}$, and $\boldsymbol{\Lambda}$.

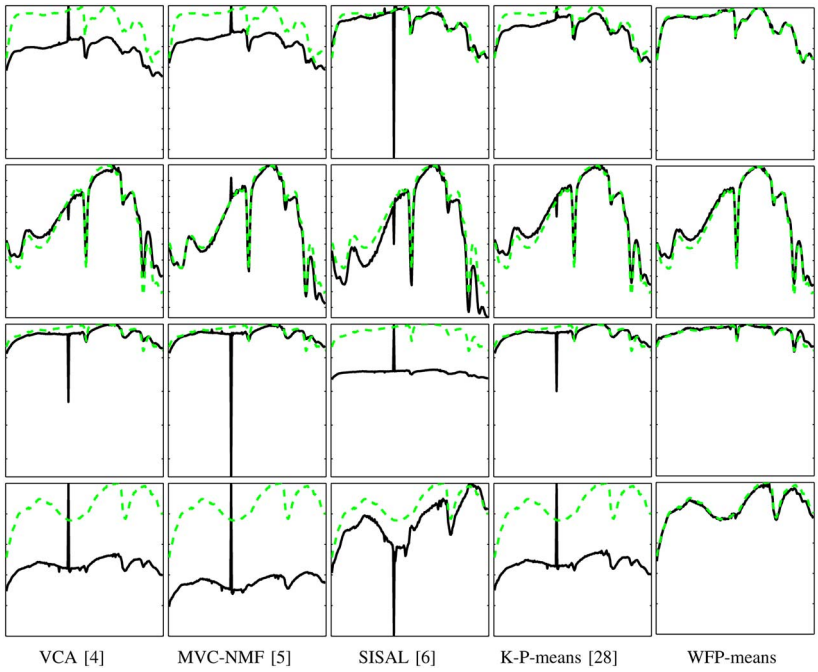


Fig. 4. True endmembers (green line) and estimated endmembers (black line) achieved by different methods. WFP-means produces spectra that match the true endmembers very well. In contrast, some spectra produced by other methods have needle-like features that cause them to deviate from the true endmembers.

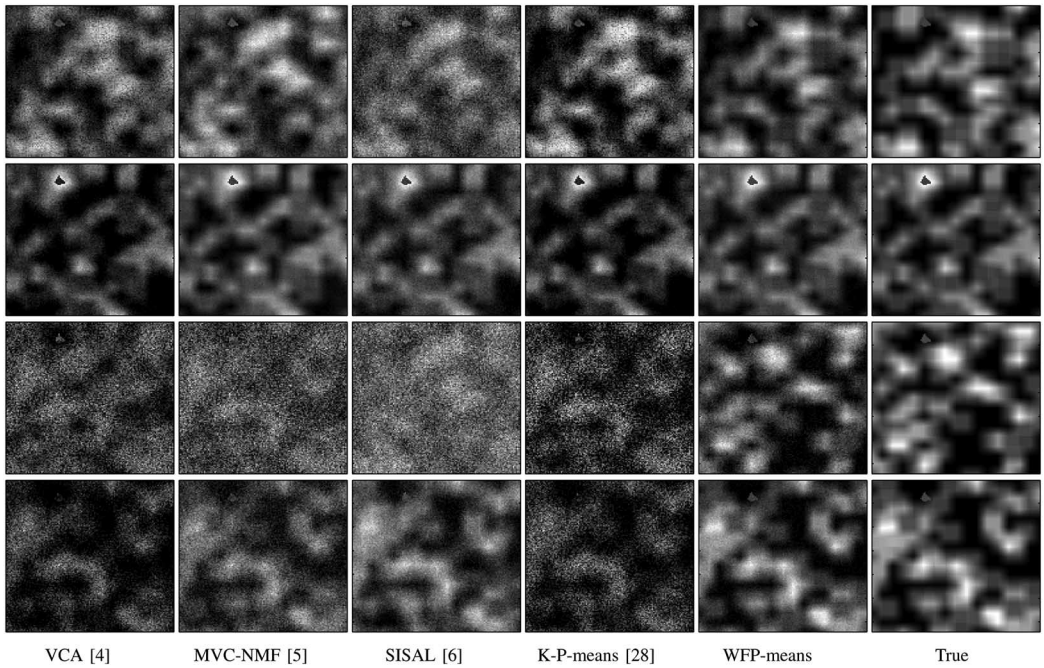


Fig. 5. Abundance maps obtained by different methods. Each column displays maps of four endmembers achieved by a method. It demonstrates that WFP-means produces abundance maps that are most similar to the true maps. It contains less artifacts than the other methods, which fail to produce informative and smooth endmember maps in some cases, especially in the third row.

According to the above model definition, HEE intends to estimate $\{\boldsymbol{a}_k\}$ and $\{s_{ik}\}$ from $\{\boldsymbol{x}_i\}$, which is formulated as a maximum a posteriori (MAP) problem

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{5}$$

This is an ill-posed inverse problem, since the number of unknown parameters is larger than the number of observations. Nevertheless, this model can be solved by the expectation maximization (EM) algorithm, which treats $\{\boldsymbol{a}_k\}$ as missing observations and alternates until convergence the estimation of $\{s_{ik}\}$ for given $\{\boldsymbol{a}_k\}$ and the update of $\{\boldsymbol{a}_k\}$, for given $\{s_{ik}\}$.

In probabilistic models, prior distributions $p(\boldsymbol{\theta})$ can be utilized to serve as regularization, in order to achieve meaningful data inversion. This paper follows the MAP strategy by proposing an WFP-means model. Nevertheless, instead of explicitly modeling the probabilistic distributions of unknown

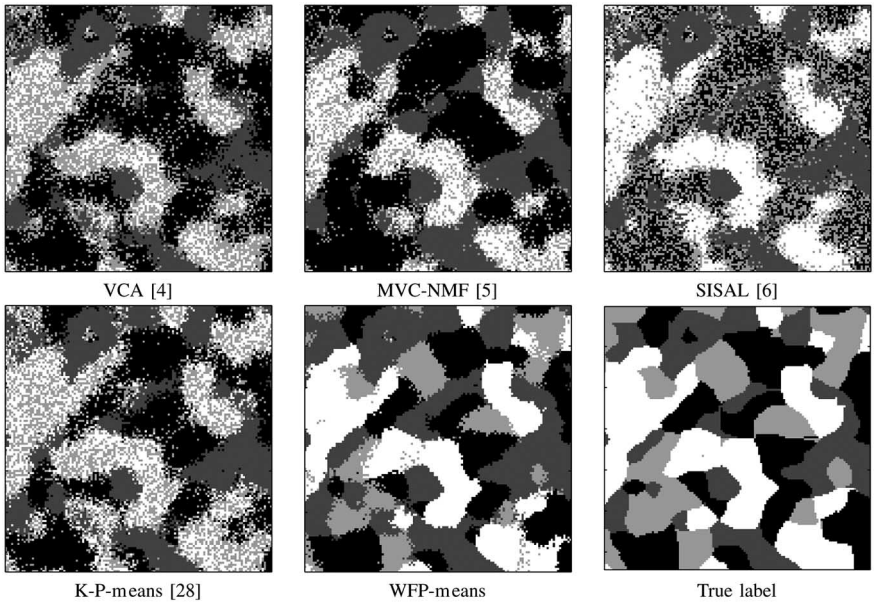


Fig. 6. Clustering labels obtained from abundances estimated by different methods. It indicates consistent results with Fig. 5. WFP-means produces a smooth map, where different classes are correctly delineated. The other methods produce maps with many artifacts, and they all fail to identify some classes correctly.

parameters, $\boldsymbol{\theta}$ is treated as nonrandom variable, and prior knowledge is incorporated by imposing some constraints on $\boldsymbol{\theta}$. In the proposed model, the following three important issues are addressed for efficient HEE.

1) The mixture/clustering effect among HSI pixels is utilized. The LSMM innately defines a clustering model, where each endmember $\boldsymbol{a}_k$ corresponds to an individual class. In LSMM, since abundances of endmembers at a pixel $\{s_{ik}|\ k = 1, 2, \ldots, K\}$ are nonnegative, they can be treated as the "soft" class membership [31], [32]. As a result, the class composition of a HSI pixel is directly connected with the abundance pattern of endmembers at that pixel. Therefore, HEE can be interpreted as a clustering issue, where the mean vectors $\{\mathbf{a}_k\}$ of classes and class membership of pixels $\{s_{ik}\}$ are iteratively optimized by alternating between the estimation of $\{s_{ik}\}$ and the update of $\{\boldsymbol{a}_k\}$. Based on this interpretation, we present a WFP-means clustering model for HEE. We will show in Section II-C, the differences and connection between the proposed model and popular clustering algorithms, e.g., K-means and fuzzy c-means (FCM).
2) The noise variance heterogeneity effect in spectral domain is considered. The noise distribution is characterized by $\boldsymbol{\Lambda}$. Since the correlation among spectral bands in $\boldsymbol{x}_i$ can be largely captured by $\sum_{k=1}^{K} \boldsymbol{a}_k s_{ik}$, $\boldsymbol{\Lambda}$ is a diagonal matrix

$$\boldsymbol{\Lambda} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{bmatrix} \tag{6}$$

where σ_j^2 is the noise variance of the jth band. In HSI, noise levels of different bands $\{\sigma_j^2\}$ are not necessarily equal, due to varying physical properties of different spectral bands and the existence of junk bands [29]. In Section II-D, this noise variance heterogeneity effect is addressed by adopting a weighted nonnegative least squares (WNNLS) algorithm to solve the proposed WFP-means model.
3) Purified pixels are separated from the mixed pixels for endmember estimation. A purified pixel for an endmember is defined as the spectral residual after removing the contribution of all the other endmembers from the mixed pixel. According to the LSMM in (1), a mixed pixel can be expressed as a weighted summation of endmembers based on their abundances. Accordingly, the individual contribution of endmembers can be separated, if the abundance and endmembers are known. Pixels, after being purified, can be treated as the realization of an endmember subject to different weights and random noise. Therefore, the endmember can be estimated as the weighted average of purified pixels.

By considering the above factors, the proposed WFP-means model constitutes an elegant solution for HEE and HU, without the need to resort to complex prior distributions and other modeling requirements. It is capable of estimating both endmembers and abundances iteratively in an adaptive manner. Moreover, it is especially effective for addressing HSI with band-dependent noise.

B. Existing Models

1) K-Means: The objective function of K-means clustering model can be formulated as

$$\arg\min_{\theta} \sum_{i=1}^{n} \left\| \boldsymbol{x}_i - \sum_{k=1}^{K} \boldsymbol{a}_k s_{ik} \right\|_2 \tag{7}$$

$$\text{s.t. } \forall s_{ik} = 0 \text{ or } 1, \text{ and } \sum_{k} s_{ik} = 1 \tag{8}$$

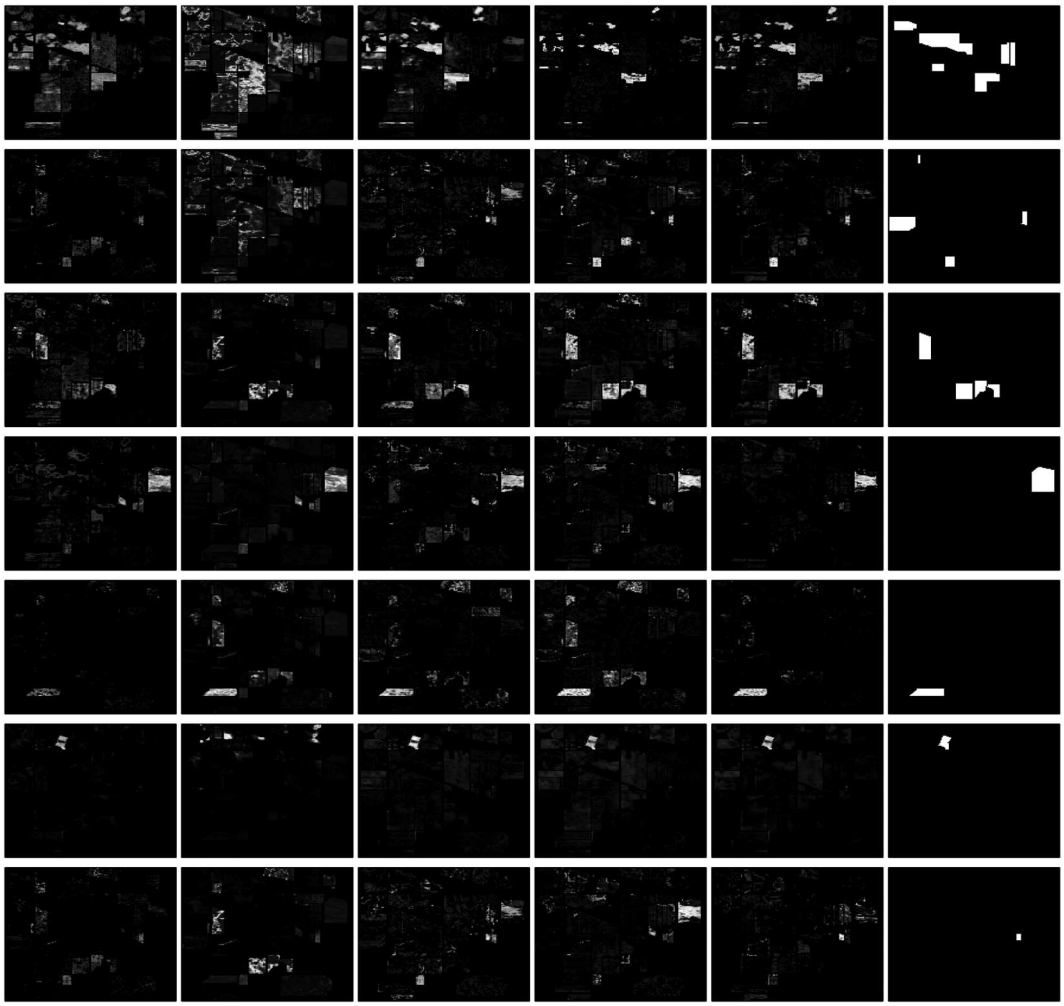


Fig. 7. Abundance maps obtained by different methods. Each column displays seven endmembers associated, respectively, with seven ground truth classes in Indian Pines image. Overall, WFP-means produces abundance maps that best correlate with the ground truth. Each abundance map of WFP-means highlight a class in ground truth, indicating that WFP-means has identified the correct endmembers, considering that different ground truth classes tend to be dominated by different endmembers. The observation that WFP-means sometimes highlights part of the ground truth area is also reasonable considering that a ground truth class tends to involve multiple endmembers, with each one dominating a different part of the ground truth class. Other methods fail to highlight the ground truth in the last row, and tend to produce more bright areas than the ground truth in other rows.

where $\|\cdot\|_2$ is the Euclidean norm, and $\boldsymbol{\theta}$ represents unknown parameters, including $\{\boldsymbol{a}_k\}$ and $\{s_{ik}\}$. Given this objective function, the K-means algorithm repeats two steps until convergence, i.e., estimate $\{s_{ik}\}$ based on $\{\boldsymbol{a}_k\}$ and update $\{\boldsymbol{a}_k\}$ using $\{s_{ik}\}$

$$s_{ik} = \delta(k - \hat{t}), \text{ with } \hat{t} = \arg\min_t \|\boldsymbol{x}_i - \boldsymbol{a}_t\| \tag{9}$$

$$\boldsymbol{a}_k = \frac{\sum_{i=1}^n s_{ik}\boldsymbol{x}_i}{\sum_{i=1}^n s_{ik}} \tag{10}$$

where $\delta(\cdot)$ is the Dirac delta function. Based on the above formulation, K-means is a "hard" membership clustering algorithm. It is not suitable for HEE, because the binary constraint on $\{s_{ik}\}$ implies that all pixels in HSI are pure pixels.

2) Fuzzy c-Means: As a soft clustering algorithm, FCM aims to optimize the following objective function [31], [32]:

$$\arg\min_\theta \sum_{i=1}^n \sum_{k=1}^K s_{ik}^m \|\boldsymbol{x}_i - \boldsymbol{a}_k\|_2 \tag{11}$$

$$\text{s.t. } 0 \le s_{ik} \le 1 \text{ and } \sum_k s_{ik} = 1 \tag{12}$$

where $1 < m < \infty$ determines the degree of fuzzyness. Similar to K-means, FCM iteratively alternates the estimation of $\{s_{ik}\}$ using (13) and the update of $\{\boldsymbol{a}_k\}$ using (10)

$$s_{ik} = \frac{1}{\sum_{j=1}^K \left(\frac{\|\boldsymbol{x}_i - \boldsymbol{a}_k\|_2}{\|\boldsymbol{x}_i - \boldsymbol{a}_j\|_2}\right)^{\frac{2}{m-1}}}. \tag{13}$$

Due to the soft constraints on s_{ik}, FCM produces fuzzy class membership, allowing a pixel to have mixed class composition. Nevertheless, estimating $\boldsymbol{a}_k$ using mixed pixels $\boldsymbol{x}_i$ as conducted by FCM in (10) is problematic. Since a mixed pixel in HSI involves the spectral contribution of multiple pure materials, the averaging of mixed pixels will lead to biased estimation of any endmembers involved. In order to address this problem, the proposed WFP-means model estimates $\boldsymbol{a}_k$ using purified pixels $\{\boldsymbol{y}_i\}$ that are solely due to the contribution of $\boldsymbol{a}_k$, as shown in Section II-E. Another problem with FCM is that it is difficult to obtain the optimal value of the fuzzyness parameter m. With a suboptimal parameter setting, FCM produces biased abundance estimation. In contrast, the proposed WFP-means model can avoid this problem by estimating abundance using constraint weighted least squares approach, where the degree of fuzzyness is adaptively determined.

C. WFP-Means

The proposed WFP-means model has the following objective function:

$$\arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(\boldsymbol{x}_i - \sum_k \boldsymbol{a}_k s_{ik}\right)^T \boldsymbol{\Lambda}^{-1} \left(\boldsymbol{x}_i - \sum_k \boldsymbol{a}_k s_{ik})\right) \tag{14}$$

$$\text{s.t. } \forall s_{ik} \geq 0. \tag{15}$$

The objective function of WFP-means has several differences with that of K-means. First, it imposes nonnegative constraint, instead of binary constraint on s_{ik}, thereby allowing pixels to have fuzzy/soft class membership. This characteristic helps build the connection between soft membership of classes and the abundance of endmembers, and enables WFP-means to model the HEE issue. Note that here we discard the use of sum-to-one constraint on s_{ik}, because we empirically find that it does not increase the HEE performance, but increases greatly the computation time. The second difference with classical K-means model is that WFP-means incorporates noise covariance matrix $\boldsymbol{\Lambda}$ defined in (6) to address the noise heterogeneity effect.

WFP-means also differs essentially with FCM in objective functions. In (14), WFP-means tries to minimize the difference between $\boldsymbol{x}_i$ and $\sum_k \boldsymbol{a}_k s_{ik}$, while FCM in (11) intends to minimize the weighted difference between $\boldsymbol{x}_i$ and $\boldsymbol{a}_k$. Therefore, given $\{s_{ik}\}$, WFP-means promotes $\boldsymbol{a}_k$ that is capable of providing a sound representation of $\boldsymbol{x}_i$ when working jointly with other $\{\boldsymbol{a}_j | j \neq k\}$, while FCM promotes $\boldsymbol{a}_k$ that alone represents $\boldsymbol{x}_i$ well. As a result, in HEE, where pixels are highly mixed, FCM is less efficient for separating different sources. In contrast, WFP-means is more efficient for HEE, because it innately reflects the data generation mechanism described by LSMM.

Following the EM algorithm, the proposed WFP-means is solved in an iterative manner, by alternating between the estimation of $\{s_{ik}\}$ based on $\{\boldsymbol{a}_k\}$ and the update of $\{\boldsymbol{a}_k\}$ based on $\{s_{ik}\}$. The former is essentially a nonnegative least squares optimization (NNLS) issue. To reduce the bias caused by the noise heterogeneity effect, in Section II-D, we introduce a weighted NNLS approach based on the estimated band-dependent noise variances.

For the estimation of $\{\boldsymbol{a}_k\}$, in contrast to K-means and FCM that rely on mixed pixels $\{\boldsymbol{x}_i\}$, we introduce in Section II-E, a novel approach utilizing "purified" pixels. In this approach, the abundance $\{s_{ik}\}$ is used to obtain purified pixels for each endmembers. Finally, an endmember is estimated as the weighted average of its purified pixels.

D. Abundance Estimation Via WNNLS

According to the EM algorithm, one essential step is to estimate $\{s_{ik}\}$ given $\{\boldsymbol{a}_k\}$, which is an NNLS issue, and is typically formulated as [33]–[35]

$$\arg\min_{\{s_{ik}\}} \left\| \boldsymbol{x}_i - \sum_k \boldsymbol{a}_k s_{ik} \right\|_2 \quad \text{s.t. } \forall\, s_{ik} \geq 0. \tag{16}$$

Based on this above formulation, a popular approach for NNLS is an active-set method, which was proposed by Lawson and Hanson in [33] and modified by Bro and De Jong for fast computation [35].

However, (16) assumes that different bands in HSI admit the same noise level, which is not necessarily true due to the varying physical properties of different spectral bands. Since least squares approaches are sensitive to noise distributions, failure to address this noise heterogeneity issue will lead to biased estimation of $\{s_{ik}\}$. To achieve accurate estimation of abundance, we want to trust more on informative bands with lower noise level. Therefore, in NNLS, instead of assigning equal weight to all bands, a sound approach is to assign smaller weight to a band with high noise level, and a larger weight to a band with low noise level. This can be achieved by using weights that are inversely proportional to noise variances.

A WNNLS approach is therefore presented, with the following objective function:

$$\arg\min_{\{s_{ik}\}} \left(\boldsymbol{x}_i - \sum_k \boldsymbol{a}_k s_{ik}\right)^T \boldsymbol{\Lambda}^{-1} \left(\boldsymbol{x}_i - \sum_k \boldsymbol{a}_k s_{ik})\right) \quad \text{s.t. } \forall s_{ik} \geq 0 \tag{17}$$

where $\boldsymbol{\Lambda}$ is defined in (6). Since $\{\sigma_j^2\}$ in $\boldsymbol{\Lambda}$ are unknown, they need to be estimated in WNNLS. One way is to iterative estimate noise variance during the model optimization process, such as in iterative weighted least squares algorithm. Here, however, we estimate noise variances before optimizing the model, by identifying a homogeneous area in HSI. Then, σ_j^2 is estimated as the variance of the pixels values in the homogeneous area of the jth band. This approach assumes that the true pixel values in homogeneous ares are the same, and therefore the variation is mainly caused by noise. This approach has been widely adopted in the image denoising literature, because it is robust and can reduce the computation cost due to iterative noise variance estimation.

Given the estimated $\hat{\mathbf{\Lambda}}$, the objective function can be reformulated as

$$\arg\min_{\{s_{ik}\}} \left\| \hat{\mathbf{\Lambda}}^{-0.5}\mathbf{x}_i - \sum_k \hat{\mathbf{\Lambda}}^{-0.5}\boldsymbol{a}_k s_{ik} \right\|_2 \quad \text{s.t. } \forall s_{ik} \geq 0. \tag{18}$$

It means that WNNLS can be achieved by adjusting $\boldsymbol{x}_i$ and $\boldsymbol{a}_k$ using $\hat{\mathbf{\Lambda}}$, and then feeding the ordinary NNLS approach with adjusted quantities. The complete algorithm of WNNLS based on fast active set algorithm [35] is summarized in Algorithm 1.

Algorithm 1. WNNLS

Input: $\boldsymbol{x}_i$, $\mathbf{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_K)$, $\mathbf{\Lambda}$
Output: $\boldsymbol{s} = (s_{i1}, s_{i2}, \ldots, s_{iK})^T$
Initialization: $\boldsymbol{x}_i = \hat{\mathbf{\Lambda}}^{-0.5}\boldsymbol{x}_i$ and $\mathbf{A} = \hat{\mathbf{\Lambda}}^{-0.5}\mathbf{A}$, two complementary indices sets $P = \varnothing$ and $Z = \{1, 2, \ldots, K\}$, $\boldsymbol{s} = \mathbf{0}$, $\boldsymbol{w} = \mathbf{A}^T(\boldsymbol{x} - \mathbf{A}\boldsymbol{s})$

```
while Z ≠ ∅ and max(w^Z > tol1) do
    t = index of max(w^Z) in w
    add t to P, and remove t from Z
    g^P = [(A^T A)^P]^{-1}(A^T b)^P
    g^Z = 0
    while min(g^P) < tol1 do
        α = min(s_k/(s_k − g_k) for i in P)
        s = s + α(g − s)
        Q = indices in s where abs(s^P) < tol1
        add Q to Z
        remove Q from P
        g^P = [(A^T A)^P]^{-1}(A^T b)^P
        g^Z = 0
    end while
    s = g
    w = A^T(x − As)
end while
```

Note: $\mathbf{X}^Y$ is restricted to the row and column variables of $\mathbf{X}$ that are included in the indices set of Y;

E. Endmember Estimation Via Weighted "Purified" Means

Another step in EM algorithm is the estimation of $\{\boldsymbol{a}_k\}$ given $\{s_{ik}\}$, according to the following equation:

$$\boldsymbol{a}_k = \frac{\sum_{i=1}^n \boldsymbol{y}_i^k s_{ik}}{\sum_{i=1}^n s_{ik}} \tag{19}$$

where $\boldsymbol{y}_i^k$ is the ith purified pixel that is solely due to the kth endmember, which can be obtained by removing the contributions of all endmembers except the kth one from the ith pixel in HSI

$$\boldsymbol{y}_i^k = \boldsymbol{x}_i - \sum_{t \neq k} \boldsymbol{a}_t s_{it}. \tag{20}$$

Therefore, comparing with K-means and FCM that use mixed pixels $\{\boldsymbol{x}_i\}$ in (10) for endmember estimation, the proposed WFP-means model estimates the kth endmember using the purified pixels $\{\boldsymbol{y}_i^k\}$ due to the kth endmember. Since mixed pixels $\{\boldsymbol{x}_i\}$ in HSI contain the contribution of multiple endmembers, using them to estimate a single endmember will inevitably be disturbed by the presence of the other endmembers. By performing weighted average, FCM adjusts the involvement of a mixed pixel $\boldsymbol{x}_i$ based on the soft membership it assumes on an endmember. Pixels that are of higher association with an endmember will be assigned larger weights in the estimation of this endmember. However, the mixed effect just mentioned cannot be resolved, regardless of the magnitude of the weight being assigned. Therefore, using weighted average of mixed pixels as estimate of endmember is essentially biased.

In the proposed WFP-means model, however, the disturbance effect caused by the existence of other endmembers is precluded by removing the contributions of the other endmembers from the mixed pixels. Since the resulting purified pixels are realizations of an endmember subject to different weight values and random noise, using them for performing weighted average leads to meaningful estimation of the endmember. A similar approach has been adopted by a K-P-means algorithm [28]. However, in contrast to K-P-means estimates $\boldsymbol{a}_k$ as the mean value of only those pixels in the kth class dominated by the kth endmember, WFP-means estimates $\boldsymbol{a}_k$ as weighted mean of all the pixels in the image regardless of their hard/discrete class labels. The new approach, as implemented in WFP-means, can be treated as soft version of that in K-P-means, and is able to handle the case when an endmember does not have dominant abundances on any pixels in the HSI.

F. Summary of Complete Algorithm

A summary of the proposed WFP-means algorithm is presented in Algorithm 2. Given an HSI, the time complexity of WFP-means depends primarily on the number of iterations before convergence. In step 7, the estimate of $\boldsymbol{a}_k$ relies on other endmembers $\{\boldsymbol{a}_t | t \neq k\}$ for calculating the purified pixels. In the mth iteration, $\{\boldsymbol{a}_t | t \neq k\}$ that are used for estimating $\boldsymbol{a}_k$ may have already been updated in the current iteration. All the updated endmembers in the current iteration are used for estimating $\boldsymbol{a}_k$, in order to increase the speed of convergence.

Algorithm 2. WFP-Means

Input: $X = \{\boldsymbol{x}_i\}$, K, $\mathbf{\Lambda}$, $iter$, i.e., the maximum number of iterations, and $tol2$, i.e., the tolerance for stopping criterion
Output: $\mathbf{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_K)$, and $\mathbf{S} = \{\boldsymbol{s}_i\}$ $where$ $\boldsymbol{s}_i = (s_{i1}, s_{i2}, \ldots, s_{iK})^T$
Initialization: $\mathbf{A}$=VCA$(\mathbf{X}, K)$, where VCA is the algorithm in [4]

```
1: while ||A^m − A^{m−1}||_F > tol2 and m < iter do
2:     for i = 1, 2, ..., n do
3:         s_i=WNNLS(x_i, A^m, Λ), using Algorithm 1
4:     end for
5:     for k = 1, 2, ..., K do
6:         calculate {y_i^k | for i = 1, 2, ..., n} using (20)
7:         calculate a_k using (19)
```

8: **end for**
9: $\mathbf{A}^m = \{\boldsymbol{a}_k | \ for\ k = 1, 2, \ldots, K\}$
10: **end while**
Note: $\|\cdot\|_F$ is the Frobenius norm

III. Experiments and Discussion

The proposed WFP-means algorithm is tested on both simulated and real HSI, in comparison with several other popular HEE methods. The methods used for comparison include FCM [31], [32], VCA [4], MVC-NMF [5], SISAL [6], and K-P-means [28]. The selection of these methods is based on the availability of the codes, as well as their relevance to our work. Three experiments are conducted. In the first experiment, different methods are tested on a simulated HSI with isotropic noise, which assumes the same variance on different bands. The second experiment is also performed on simulated image, but with heterogeneous noise that admits varying noise variances on different bands. The last experiment is based on the real Indian Pines HSI.

In all experiments, we empirically set the parameters of WFP-means to be $iter = 30$, $tol2 = 10^{-5}$. The $tol1$ parameter in WFP-means is set to be $tol1 = 10^{-8}$ in the first two experiments, and $tol1 = 10^{-7}$ in the last experiment. The WFP-means and K-P-means algorithms are set to have five replicates to adopt the one with the lowest representational error. The parameters setting of all the other algorithms are applied according, respectively, to the authors' suggestions. Since VCA and SISAL only estimate the endmembers, we use the the NNLS algorithm implemented in [35] to extract the abundance.

In the simulated studies, since the true values of endmembers and abundances are known, some numerical measures can be acquired to evaluate the performance of methods. Here, we adopt four widely used measures for performance assessment [5]. The spectral angle distance (SAD) is defined as

$$SAD = \cos^{-1}\left(\frac{\boldsymbol{a}^T\hat{\boldsymbol{a}}}{\|\hat{\boldsymbol{a}}\|\ \|\boldsymbol{a}\|}\right) \tag{21}$$

where $\hat{\boldsymbol{a}}$ and $\boldsymbol{a}$ are, respectively, the estimated and the true endmembers. Smaller SAD value means shorter distance between the estimated endmember and the true endmember, thereby indicates better performance in terms of endmember estimation. The spectral information divergence (SID) is expressed as

$$SID = D\left(\frac{\boldsymbol{a}}{\hat{\boldsymbol{a}}}\right) + D\left(\frac{\hat{\boldsymbol{a}}}{\boldsymbol{a}}\right) \tag{22}$$

where $D(\frac{\boldsymbol{x}}{\boldsymbol{y}})$ measures the relative entropy between $\boldsymbol{x}$ and $\boldsymbol{y}$ [36]. Similar to SAD, smaller SID value indicates better endmember estimation. The abundance angle distance (AAD) and abundance information divergence (AID) are obtained by replacing endmember vector $\boldsymbol{a}$ with abundance vector $\boldsymbol{s}$. Smaller AAD and AID values imply better abundance estimation.

A. Experiment 1: Test on Simulated HSI With Isotropic Noise

In the first experiment, a 256×256 sized image with mixed pixels of four endmembers randomly selected from the USGS digital spectral library is simulated, following the procedure reported in [5]. Using the four endmembers, mixed pixels are created by first dividing the entire image into 8×8 sized homogeneous blocks of one of the four endmembers, then degrading the blocks by applying a spatial low pass filter of 7×7. To further increase mixing degree, the remaining relatively pure pixels with 80% or larger single abundance are forced to take equal abundances on all endmembers.

Zero-mean Gaussian noise is added to further degrade the image. In this experiment, different bands in HSI are contaminated by noise of the same variance. In order to examine the robustness of different methods to noise levels, we simulated images with four different signal-to-noise ratio (SNR), i.e., SNR $= 10$, 20, 30, and 45. For each SNR, we produce 10 images with different noise realizations to reduce the bias that could be caused by randomness.

Fig. 1 displays the statistics of different methods over varying noise levels. It indicates that the proposed WFP-means algorithm achieves comparable results with the other advanced methods, e.g., K-P-means and NMF. Specifically, WFP-means produces slightly higher SAD and SID values than K-P-means, which seems to yield the smallest SAD and SID values in most cases. However, in terms of AAD and AID, WFP-means and K-P-means perform comparably. The WFP-means and K-P-means are less sensitive to noise level variation than the other methods, i.e., SISAL and MVC-NMF. SISAL performed very well when the noise level is low, e.g., SNR $= 30$ and 45. However, its performance deteriorates quickly from SNR $= 20$ to SNR $= 10$, especially in terms of SAD and SID. MVC-NMF indicates similar patterns. FCM achieves the lowest overall performance. It always produces large bias in abundance estimation.

Fig. 2 shows the endmembers estimated by different methods when SNR $= 20$, chosen randomly from the ten noise realizations when SNR $= 20$. WFP-means and K-P-means yield endmembers that are most similar to the true values. The endmembers in the bottom row achieved by SISAL and MVC-NMF do not match the true endmember vary well. FCM fails to determine the endmembers from HSI.

B. Experiment 2: Test on Simulated HSI With Heterogeneous Noise

The same method in Experiment 1 is adopted for generating HSI. However, instead of using the same SNR for all bands, different SNRs are assigned to different bands to simulate the noise heterogeneity effect. The SNR values used for simulation are estimated from the benchmark Indian Pines image, introduced in Section III-C, following the procedures described in Section II-D. Denote the estimated SNR vector that has been centralized and normalized by $\mathbf{q}$, then the simulated SNR $\mathbf{r}$ can be obtained according to the following rule:

$$\boldsymbol{r} = \alpha\boldsymbol{q} + \beta \tag{23}$$

where α is the amplitude that determines the magnitude of fluctuation of band-dependent SNR, and β is the center value that

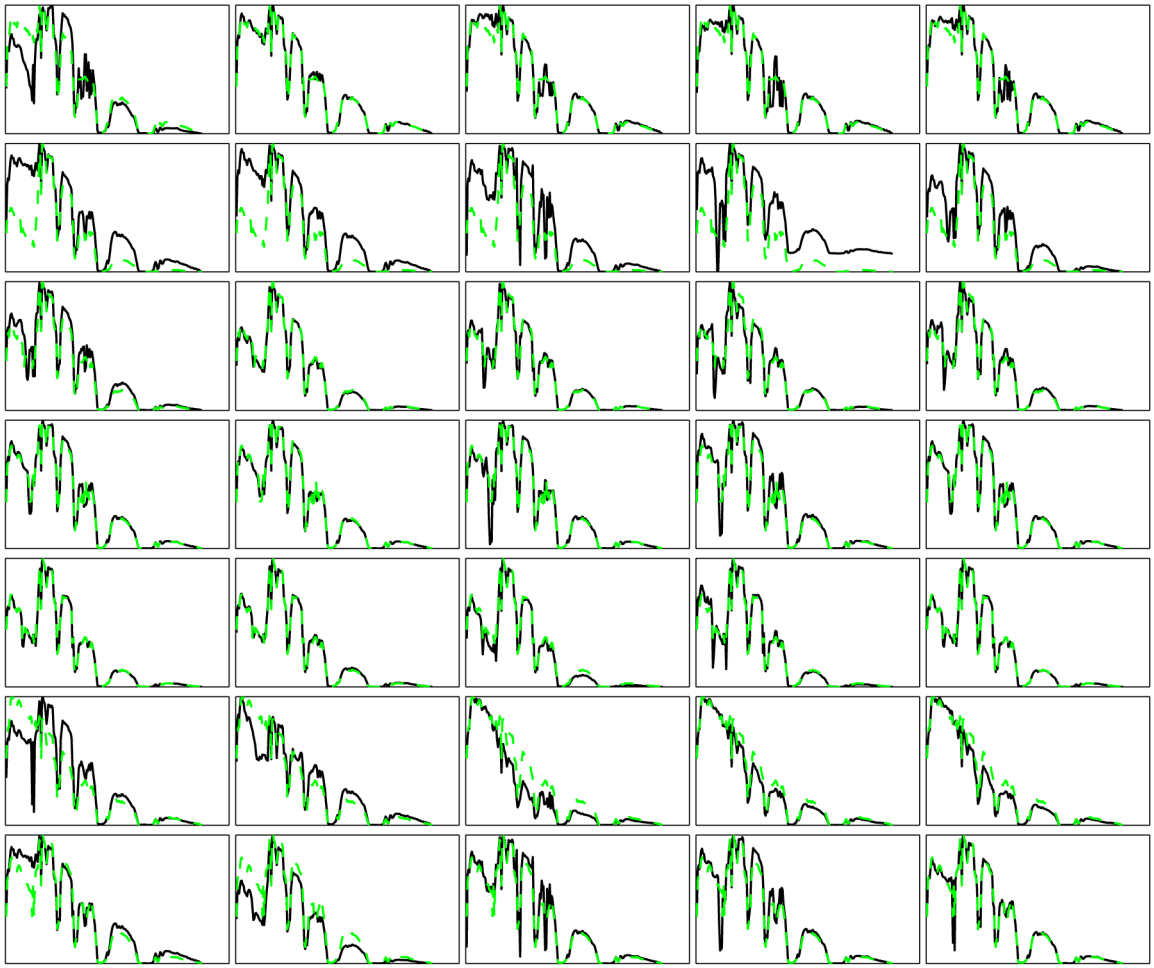


Fig. 8. These endmembers associated with the abundance maps in Fig. 7. Green lines are the the mean spectra of the ground-truth classes in Indian Pines image, and black lines are extracted endmembers by different methods. It indicates consistent results with Fig. 7, i.e., the endmembers extracted by WFP-means correlate with the mean spectra of ground-truth classes better than those of the other methods especially on the second row, leading to consistent conclusions drawn from Fig. 7.

defines the overall SNR of all bands. Therefore, by changing α, noise variance heterogeneity effect of varying degrees can be simulated. Bigger α value leads to stronger noise heterogeneity effect. Here, we use fixed $\beta = 20$, but six different α values, i.e., $\alpha = 0, 3, 5, 6, 7, 9$.

Fig. 3 shows the statistics achieved by different methods on the second simulated HSI. As we can see, WFP-means performs better than FP-means, demonstrating the importance and benefit of addressing the noise heterogeneous effect. Considering that K-P-means and FP-means perform similarly in Fig. 3, we only keep the results of K-P-means in the other figures.

Fig. 3 shows the statistics of different methods over different α values. A variant of WFP-means, called FP-means, which ignores the noise heterogeneous effect by discarding the use of the weighting scheme, is also included. Overall, WFP-means produces the lowest values on all measures over all amplitude values. The fact that WFP-means outperforms FP-means demonstrates the importance and benefit of addressing the noise heterogeneous effect. Moreover, WFP-means is insensitive to the changing of amplitude values, e.g., it produces the most consistent statistic values over the six amplitude levels, indicating that WFP-means is capable of resisting the noise variance heterogeneity effect. The statistic values of all other methods increase faster with the increase of amplitudes. SISAL is most sensitive method in terms of SAD and SID.

Fig. 4 plots the endmembers achieved by different methods when $\alpha = 7$. WFP-means produces spectra that match the true endmembers very well. In contrast, some spectra produced by other methods have needle-like features that cause them to deviate from the true endmembers.

Fig. 5 displays the abundance maps associated with the endmembers in Fig. 4. The results are randomly picked up from the 10 noise realizations when amplitude equals 5. It demonstrates that WFP-means produces abundance maps that are most similar to the true maps. It contains less artifacts caused by noise than the other methods, which fail to produce informative and smooth endmember maps in most cases, especially in maps of the third row.

Fig. 6 shows the hard membership, i.e., labels, obtained from the abundances estimated by different methods. It indicates consistent results with Fig. 5. WFP-means produces a

smooth map, where different classes are correctly delineated. The other methods produce maps with many artifacts, and all fail to identify some classes correctly.

C. Experiment 3: Test on Real HSI

In this experiment, all methods are tested on the benchmark Indian Pines image, which was captured by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over a vegetation area in northwestern Indiana, USA, with a spatial resolution of 20 m, consisting of 145×145 pixels of 16 ground-truth classes and 220 spectral bands.

Most methods use the number of endmembers as input, which, however, is unknown in the Indian Pines dataset. Considering that there are 16 ground-truth classes in the dataset, and that some classes may involve more than 1 endmember, each method is set to extract 20 endmembers from the dataset. We empirically found that using other values around 20 does not significantly influence the comparison results. The SISAL method is not used here because it fails to produce seemingly meaningful endmembers.

In this experiment, since the true endmembers are unknown, we rely primarily on the ground-truth information to evaluate the performances of different methods. Fig. 7 displays the abundance maps obtained by different methods, where each row displays an endmember associated with a ground-truth class. From top to bottom, the associated classes in Indian Pines image are, respectively, 1) corn-no-till; 2) grass/pasture; 3) grass/trees; 4) haywindrowed; 5) wheat; 6) stone-steel-tower; and 7) grass/pasturemowed. For each ground-truth class in Fig. 7, we looked through all the 20 endmembers of a method, and visually determine the endmember that best matches the ground-truth. Overall, WFP-means produces abundance maps that have the best correlation with the ground-truth classes. Considering that different ground-truth classes in HSI tend to be dominated by different endmembers, we would expect a particular endmember to have high abundance on a particular ground-truth class, but less abundance on the other classes. Accordingly, the observation that the abundance maps obtained by WFP-means associate better with the ground-truth classes than those of the other methods might indicate a better performance of WFP-means than the other methods. Considering the complexity of material composition in ground-truth classes, it is however not surprising to see some irregular patterns, e.g., the abundance maps of WFP-means sometimes highlight only parts of ground-truth class instead of the whole class, and sometimes highlight slightly other classes besides the main class.

Fig. 8 indicates consistent results with Fig. 7, i.e., the endmembers extracted by WFP-means correlate with the mean spectra of ground-truth classes better than those of the other methods, leading to consistent conclusion drawn from Fig. 7.

IV. Conclusion

This paper presented a novel fuzzy membership clustering model for HEE, where the endmembers are treated as individual classes, and the abundances of endmember are defined as the soft class membership. Therefore, HEE is interpreted as a clustering problem, which is solved by iteratively alternating the estimation between the estimation of abundances and the update of endmembers. The proposed model is further enhanced by utilizing the purified pixels estimated using abundance and endmember information. The endmembers are estimated as the mean values of purified pixels, leading to an elegant optimization solution. Moreover, the noise variance heterogeneity effect is addressed in the proposed model by explicitly modeling the noise distribution. The proposed WFP-means algorithm is tested on both simulated and real HSI, in comparison with several other popular methods. The results demonstrate that WFP-means produces better endmember and abundance estimation than most referenced methods, especially in the scenario of heterogeneous band-dependent noise.

Acknowledgment

The authors would like to thank the anonymous reviewers for their comments and suggestions.

References

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[2] A. Plaza, P. Martínez, R. Pérez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 650–663, Mar. 2004.

[3] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[4] J. M. Nascimento and J. M. Bioucas Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[5] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[6] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Proc. 1st Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS'09)*, 2009, pp. 1–4.

[7] M. E. Winter, "N-findr: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Int. Symp. Opt. Sci. Eng. Instrum.*, 1999, pp. 266–275.

[8] J. W. Boardman *et al.*, "Automating spectral unmixing of aviris data using convex geometry concepts," in *Proc. Summ. 4th Annu. JPL Airborne Geosci. Workshop*, 1993, vol. 1, pp. 11–14.

[9] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of aviris data," in *Proc. JPL airborne earth sci. workshop*, 1995, vol. 1, pp. 23–26.

[10] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images: Learning from earth's shapes and colors," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[11] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS'08)*, 2008, vol. 3, pp. III-250–III-253.

[12] S. Moussaoui, C. Carteret, D. Brie, and A. Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods," *Chemom. Intell. Lab. Syst.*, vol. 81, no. 2, pp. 137–148, 2006.

[13] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, 2009.

[14] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[15] M. Arngren, M. N. Schmidt, and J. Larsen, "Unmixing of hyperspectral images using bayesian non-negative matrix factorization with volume prior," *J. Signal Process. Syst.*, vol. 65, no. 3, pp. 479–496, 2011.
[16] J. M. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 863–878, Mar. 2012.
[17] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
[18] D. M. Rogge, B. Rivard, J. Zhang, and J. Feng, "Iterative spectral unmixing for optimizing per-pixel endmember sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 12, pp. 3725–3736, Dec. 2006.
[19] M.-D. Iordache, A. Plaza, and J. Bioucas-Dias, "On the use of spectral libraries to perform sparse unmixing of hyperspectral data," in *Proc. 2nd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS'10)*, 2010, pp. 1–4.
[20] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.
[21] A. Linear, "Foreword to the special issue on spectral unmixing of remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4103–4110, Nov. 2011.
[22] M. Parente and A. Plaza, "Survey of geometric and statistical unmixing algorithms for hyperspectral images," in *Proc. 2nd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS'10)*, 2010, pp. 1–4.
[23] A. Zare and P. Gader, "Piece-wise convex spatial-spectral unmixing of hyperspectral imagery using possibilistic and fuzzy clustering," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ)*, 2011, pp. 741–746.
[24] O. Bchir, H. Frigui, A. Zare, and P. Gader, "Multiple model endmember detection based on spectral and spatial information," in *Proc. 2nd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS'10)*, 2010, pp. 1–4.
[25] A. Zare, O. Bchir, H. Frigui, and P. Gader, "A comparison of deterministic and probabilistic approaches to endmember representation," in *Proc. 2nd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS'10)*, 2010, pp. 1–4.
[26] A. Zare and P. Gader, "PCE: Piecewise convex endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2620–2632, Jun. 2010.
[27] A. Zare, O. Bchir, H. Frigui, and P. Gader, "Spatially-smooth piece-wise convex endmember detection," in *Proc. 2nd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS'10)*, 2010, pp. 1–4.
[28] L. Xu, J. Li, A. Wong, and J. Peng, "K-p-means: A clustering algorithm of k "purified" means for hyperspectral endmember estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1787–1791, Oct. 2014.
[29] A. C. Zelinski and V. K. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp. (IGARSS'06)*, 2006, pp. 387–390.
[30] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[31] J. T. Kent and K. V. Mardia, "Spatial classification using fuzzy membership models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 5, pp. 659–671, Sep. 1988.
[32] G. Foody and D. Cox, "Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions," *Remote Sens.*, vol. 15, no. 3, pp. 619–631, 1994.
[33] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, vol. 161. Philadelphia, PA, USA: SIAM, 1974.
[34] D. Chen and R. J. Plemmons, "Nonnegativity constraints in numerical analysis," in *Proc. Symp. Birth Numer. Anal.*, 2009, pp. 109–140.
[35] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *J. Chemom.*, vol. 11, no. 5, pp. 393–401, 1997.
[36] C.-I. Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, May 2000.

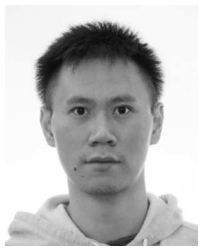
Linlin Xu (M'14) received the B.Eng. and M.Sc. degrees in geomatics engineering from China University of Geosciences, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in geography from the University of Waterloo, Waterloo, ON, Canada.

He is currently working as a Postdoctoral Fellow with Vision and Image Processing Laboratory, Systems Design Engineering, University of Waterloo. His research interests include hyperspectral and SAR image processing.

Alexander Wong (M'05) received the B.A.Sc. degree in computer engineering, the M.A.Sc. degree in electrical and computer engineering, and the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005, 2007, and 2010, respectively.

He is currently an Assistant Professor with the Department of Systems Design Engineering, University of Waterloo. He has authored refereed journal and conference papers, as well as patents, in various fields, such as computer vision, graphics, image processing, multimedia systems, and wireless communications. His research interests include image processing, computer vision, pattern recognition, and cognitive radio networks, with a focus on biomedical and remote sensing image processing and analysis such as image registration, image denoising and reconstruction, image super-resolution, image segmentation, tracking, and image and video coding and transmission.

Dr. Wong was the recipient of an Outstanding Performance Award, an Engineering Research Excellence Award, an Early Researcher Award from the Ministry of Economic Development and Innovation, a Best Paper Award by the Canadian Image Processing and Pattern Recognition Society, and the Alumni Gold Medal.



Fan Li received the B.S. degree in geographical information system from Sun Yat-sen University, Guangzhou, China, and the M.E. degree in pattern recognition and intelligent system from Wuhan University, Wuhan, China. He is currently pursuing the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada.

His research interests include the application of machine learning and computer vision techniques to the remote sensing filed.

David A. Clausi (S'93–M'96–SM'03) received the BA.Sc., MA.Sc., and Ph.D. degrees in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1990, 1992, and 1996, respectively.

After completing the Ph.D., he worked in the medical imaging field with Mitra Imaging Inc., Waterloo, ON, Canada. He started his academic career as an Assistant Professor in Geomatics Engineering with the University of Calgary, Calgary, AB, Canada, in 1997. He returned to his alma mater in 1999 and was awarded tenure and became an Associate Professor in 2003. He is an Active Interdisciplinary and Multidisciplinary Researcher. He has authored refereed journal and conference papers on remote sensing, computer vision, algorithm design, and biomechanics. His research interests include the automated interpretation of synthetic aperture radar (SAR) sea ice imagery, in support of operational activities of the Canadian Ice Service.

Dr. Clausi has received numerous scholarships, conference paper awards, and two teaching excellence awards.