Spectral Unmixing Cluster Validity Index for Multiple Sets of Endmembers

Derek T. Anderson, *Member, IEEE,* and Alina Zare, *Member, IEEE*

***Abstract*—A hyperspectral pixel is generally composed of a relatively small number of endmembers. Several unmixing methods have been developed to enforce this concept through sparsity promotion or piece-wise convex mixing models. Piece-wise convex unmixing methods often require as parameters the number of endmembers and the number of sets of endmembers needed. However, these values are often unknown in advance and difficult to estimate. In this article, a new cluster validity index for multiple sets of endmembers is developed. The proposed index is used to evaluate spectral unmixing results and identify optimal parameter sets for piece-wise convex unmixing methods. No other conventional cluster validity index is directly applicable or theoretically well-suited for the piece-wise convex model. Specifically, we focus on addressing cases in which endmembers may or may not be located in a dense region of the data. Additionally, we focus on cases in which hyperspectral data is well distributed within a convex cluster (not exhibiting significant holes or gaps). The proposed validity index is applied to both simulated and real hyperspectral data. Results show that the proposed method consistently selects the best parameter set.**

***Index Terms*—Cluster validity, endmember, hyperspectral, piece-wise convex, spectral variation, unmixing.**

I. Introduction

In RECENT years, many linear spectral unmixing methods for hyperspectral image analysis have been developed [1]–[6]. The standard mixing model used is the linear mixing model where the hyperspectral pixels are modeled as convex combinations of the endmember signatures [7]. The linear mixing model can be written as

$$\mathbf{x}_i = \sum_{k=1}^{M} p_{ik}\mathbf{e}_k + \epsilon_i \tag{1}$$

where M is the number of endmembers, ϵ_i is an error term, p_{ik} is the abundance of endmember k in pixel i, and $\mathbf{e}_k$ is the k^{th} endmember. The abundances of this model satisfy the following constraints:

$$p_{ik} \geq 0 \quad \forall k = 1, \ldots, M \tag{2}$$

$$\sum_{k=1}^{M} p_{ik} = 1. \tag{3}$$

Manuscript received August 30, 2011; revised November 16, 2011. Date of publication May 22, 2012; date of current version July 20, 2012.

D. T. Anderson is with the Department of Electrical and Computer Engineering, Mississippi State University, MS, 39762 USA (corresponding author, e-mail: anderson@ece.msstate.edu).

A. Zare is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, 65211 USA (e-mail: zarea@missouri.edu).

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.2012.2189556

Given input hyperspectral data, $\mathbf{X}$, endmember extraction and spectral unmixing methods estimate the spectral signatures of the endmembers, $\mathbf{E}$, and/or the values of the abundances (or proportions), $\mathbf{P}$, of each endmember in every hyperspectral pixel. Estimating endmembers under the linear mixing model amounts to estimating the spectral signatures whose convex hull enclose the input hyperspectral data.

Recently, work in sparse unmixing methods [8]–[11], piece-wise convex unmixing methods [12]–[14], and methods such as [15], [16] where each pixel can be modeled with a different numbers and types of endmembers and [17] where spatially-dependent endmember are extracted have addressed the issue that, often, each hyperspectral pixel is a mixture of a small number of endmembers from the complete set of endmembers which describe the full hyperspectral scene. In sparse unmixing, a single set of endmembers are used to describe an input hyperspectral scene and each pixel is, then, described by a sparse set of the entire set of endmembers in the scene. In other words, $\mathbf{x} = \mathbf{E}\mathbf{p} + \boldsymbol{\epsilon}$ where $\mathbf{x}$ is an input hyperspectral pixel, $\mathbf{E}$ is an $D \times M$ matrix containing M endmember spectra and $\mathbf{p}$ is the vector of proportion values such that $\mathbf{p}$ is a sparse vector containing a relatively small number of non-zero values. Piece-wise convex unmixing methods extend this idea by, not only using a sparse set of endmembers to describe each pixel, but also partitioning the pixels in a hyperspectral scene such that each partition is described using the same small set of endmembers. Therefore, these methods differ from traditional linear unmixing methods in that they unmix hyperspectral pixels using a small set of endmembers and, in the case of piece-wise convex unmixing, groups of hyperspectral pixels from a scene are constrained to be unmixed with the same small set of endmembers.

Consider the AVIRIS Indian Pines hyperspectral data set shown in Fig. 1 as motivation for the use of piece-wise convex unmixing methods [18]. This data was collected over an agricultural area in northern Indiana. Since the crops were in very early growth stages and the fields are composed of primarily soil with very little crop cover (roughly 5% crop cover) [19], this data is known to be highly mixed. The fields are primarily composed of soil and varying levels of residue from the previous year's crops. Therefore, the spectral signature of each pixel in a field is a mixture of soil, residue and crop endmembers. Fig. 2 displays a scatter plot of all of the Corn-min and Woods points in the image after applying principal components analysis. The corn-min pixels are shown in blue and the woods pixels are shown in green. Given that the corn-min pixels are composed of soil, corn crop, and previous crop residue, endmembers appropriate for these materials can be used to unmix the corn-min pixels and define a convex region to enclose the blue corn-min pixels in Fig. 2. The woods pixels in this image are mixtures of the many natural materials found in this portion of the scene. The corresponding endmembers for the woods pixels can be

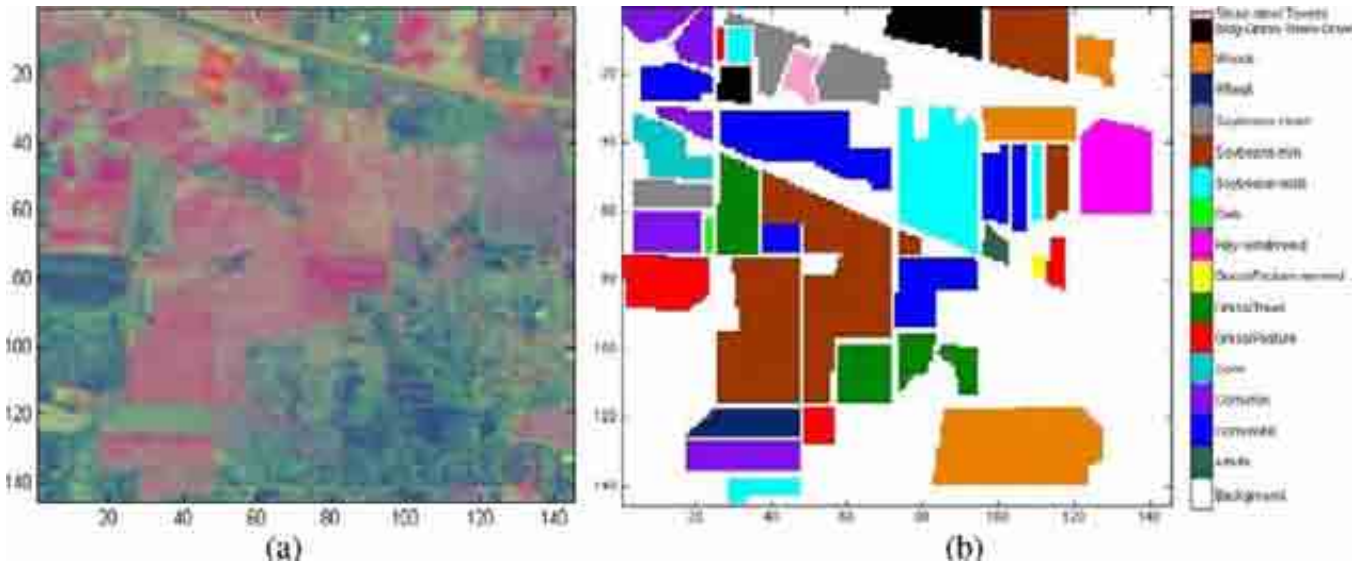


Fig. 1. June 1992 AVIRIS Indian Pines "Scene 4" data set. Image has 145 × 145 pixels with 220 spectral bands. (a) RGB figure generated using the first three principal components of the data where red is PC 1, green is PC 2 and blue is PC 3. (b) Groundtruth.

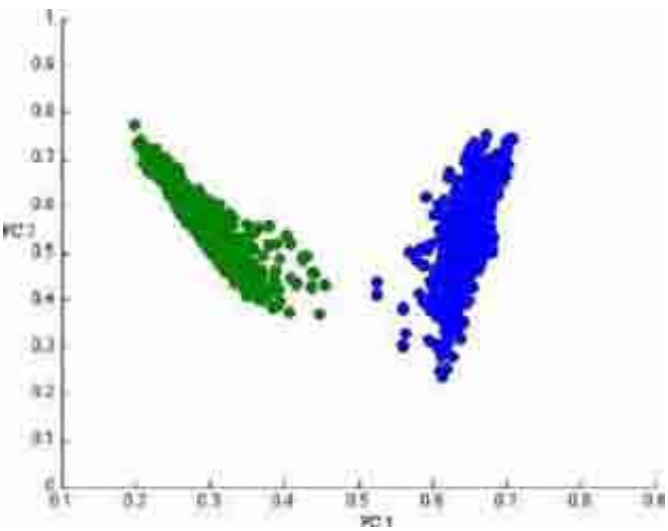


Fig. 2. Scatter plot of Corn-min (blue) and Woods (green) points from the AVIRIS data set after applying PCA and reducing to two dimensions.

estimated and used to define a convex region that encloses the green points in Fig. 2. As can be seen in the figure, these sets of points are not overlapping and should be unmixed with two sets of endmembers: one to enclose the corn-min pixels and one to enclose the woods pixels. If a single set of endmembers was used to represent all of this data (both corn-min and woods), proportion estimates for pixels from the Corn-min class may incorrectly have non-zero values for endmembers corresponding to Woods and vice versa. A piece-wise convex model addresses this issue by describing each highly-mixed and distinct convex region using a separate set of endmembers. Thus, for the data in the scatter plot shown in Fig. 2, two distinct sets of endmembers would be used to represent the data: one set of endmembers for the corn-min materials and one set of endmembers for the woods materials. Fig. 3 shows a scatter plot of the full Indian Pines data set which, as a whole, is non-convex and would be better represented with a piece-wise convex representation of the data. Many of the fields may be well represented with the same set of endmembers for soil, crop and residue and another set of endmembers can be use for grass, woods and trees, for example. Furthermore, interior endmembers cannot be recovered using methods based on the standard linear mixing model. For example, the Hay-windrowed pixels from this image create a corner (at roughly [0.4, 0.2] in the scatter plot) that likely corresponds to an endmember. The point lies within the convex hull defined by other endmembers in the scene. Unmixing methods which rely on the linear mixing model with a single set of endmembers cannot recover this Hay-windrowed endmember following PCA dimensionality reduction. However, methods based on a piece-wise convex representation are able to estimate an endmember that falls within the convex hull of a different set of endmembers.

Some piece-wise convex methods as in [12] require as parameters the number of endmembers and number of convex regions. These are difficult parameters to estimate. Several methods have been developed for estimating the number of endmembers [20]–[24]. However, these methods do not address the piece-wise convex model since these methods estimate the number of endmembers needed assuming a single convex region (i.e., a single set of endmembers). In this paper, a clustering validity index for multiple sets of endmembers is developed. This index can be used to evaluate piece-wise convex estimated endmember and proportion values. Using this index, the best set of parameters for a piece-wise convex endmember detection algorithm (e.g., the number of endmembers and the number of sets of endmembers) are estimated.

In the following, Section II provides a review of conventional cluster validity indices. Section III describes the proposed cluster validity index for evaluating piece-wise convex spectral unmixing results. Sections IV and V provide experimental results on simulated and real data. Finally, Section VII provides conclusions and future work.

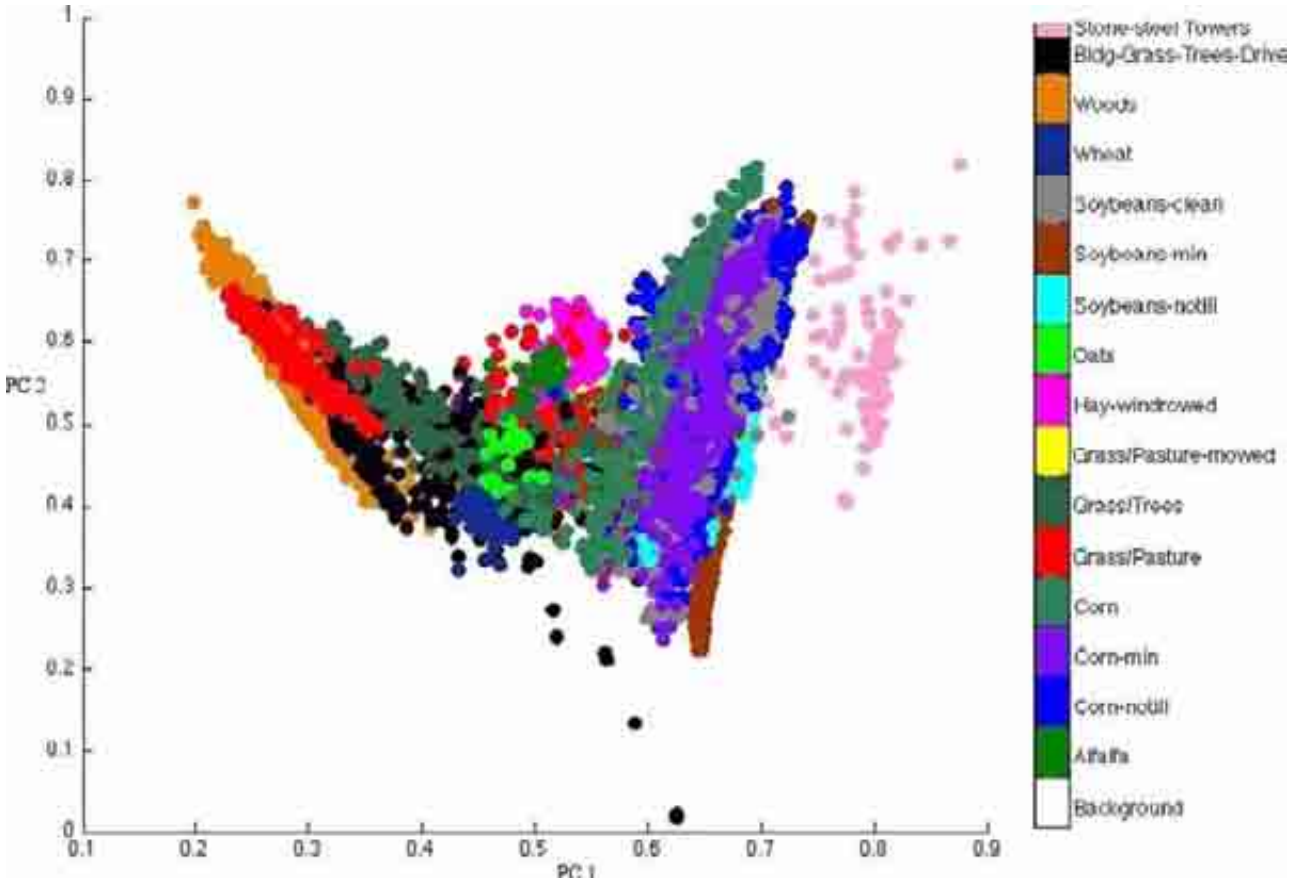


Fig. 3. First two principal components from PCA for AVIRIS. This shows that AVIRIS is not convex but, instead, appears piece-wise convex. Pixel colors correspond to ground truth labels. Points with the label background are colored white and cannot be seen. Note, convex data in high dimensions remain convex after reduction using PCA and convexity in a lower space is not known if the data is non-convex in the original dimensionality. Thus, if the data is non-convex after PCA, that data is non-convex also in the original dimensionality.

II. Conventional Cluster Validity

Clustering algorithms map an input data set, X, to a partition matrix, U, i.e., $X \subset \Re^d \mapsto M_{*cn}$, where c is the number of clusters, n is the number of objects and $*$ is h for the set of hard partitions (M_{hcn}), f for fuzzy (M_{fcn}) or p for the set of possibilistic partitions (M_{pcn}) [25]. For example, the set of possibilistic partitions is $M_{pcn} = \{U \in \Re^{cn} : u_{ik} \in [0,1] \forall i,k; 0 < \sum_{k=1}^{n} u_{ik} \forall i\}$. In addition, clustering algorithms produce a (terminal) set of parameters, e.g., $\Theta_i = \{\mu_1, \ldots, \mu_c, \Sigma_1, \ldots, \Sigma_c\}$ for the fuzzy c-means (FCM) clustering algorithm with Mahalanobis distance. Varying cluster algorithm parameters, such as the fuzzifier (q) in the FCM, or different initializations on the same input data can result in different clustering results. The question investigated in this article is which (U_i, θ_i) *best* explains and represents the (unknown) structure in our data set.

If clustering is guided by an objective function, at first glance, it seems like values of the objective function should suffice to choose the best solution. However, it is well known that even global extremum of many objective functions (e.g., J_1 for hard c-means) can lead to unrealistic partitions [26]. Moreover, some intuitively desirable properties wanted in a clustering result cannot be captured by any functional that is easily optimized. These are arguably two compelling reasons to add validation to the clustering process.

Let $CP = \{U_i : 1 \leq i \leq B\}$ denote B different *candidate partitions* that result from clustering a fixed data set X with one algorithm at various values of its parameters. In general, there are two approaches to identify the best U in CP: *validity indices* or *comparison indices*. First, *validity indices*, i.e., $v(U)$, can be computed for each U_i and used to evaluate the clustering results. The set of values $\{v(U) : U \in CP\}$ are used to identify the *best* U in CP, U^*, in one of the following two ways: (1) U^* optimizes v over CP (i.e., min or max); or (2)U^* is the antecedent or successor partition that essentially optimizes the derivative of v. Many call this second method the big-jump approach, and is usually confined to sequences of partitions that are hierarchically nested by algorithms such as the single-linkage method. Almost all optimization and big-jump indices are *internal criteria*, i.e., they use only internal information that is generated by the clustering process itself to select U^*. The second group of methods use *comparison indices*, i.e., $s(U, V)$, that compare pairs of partitions [25]. In this article, we present a new cluster validity index that can be used to discover the *best* set of parameters for a piece-wise convex spectral unmixing using multiple sets of endmembers.

Currently, there is no validity index designed (or well-suited) for piece-wise convex spectral unmixing. Most existing cluster validity indices focus on defining *good* clusters in terms of two properties: *well-separated* and *compact*. In general, well-sepa-

rated means a validity index prefers clusters to be far apart, i.e., a pair of clusters are not close to each other and have little to no overlap. Compact generally means that a method should prefer small and tight clusters, i.e., low within class/cluster scatter. For the piece-wise convex endmember-cluster validity index developed in this paper, interpretations for what well-separated and compact means in this context need to be developed. Prior to the development of the proposed piece-wise convex endmember-cluster validity index, a review of conventional cluster validity indices is provided.

A. Partition Coefficient and Classification Entropy Indices

Two classical validity indices are the partition coefficient, PC, and classification entropy, CE [27]. $PC(c)$ is defined as

$$PC(c) = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^2. \tag{4}$$

The PC was created to measure the amount of overlap between clusters. Disadvantages of the PC are that it is monotonic decreasing with c and it lacks any direct connection to properties of the data. $CE(c)$ is defined as

$$CE(c) = -\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}log(u_{ij}). \tag{5}$$

As shown by Bezdek, $0 \leq 1 - PC(c) \leq CE(c)$ for all probabilistic cluster partitions c. Thus, CE is essentially a measure of the amount of overlap of clusters, similar to PC. Also, the CE, like the PC, possess monotic tendency with c. While the PC and CE are classical (and relatively simple) indices, they are reported because they are examples of validity indices that rely solely on the partition matrix.

B. Davies-Bouldin Index

The Davies-Bouldin index is a (crisp, i.e., an element is assumed to belong to just one cluster) validity index that takes into account information from the clustering result as well as the input dataset [28]. The DBI incorporates the desire for compact and well-separated clusters by computing within cluster scatter and between cluster separation, respectively. The within cluster scatter for the i^{th} cluster is

$$S_{i,q} = \left(\frac{1}{|A_i|}\sum_{x\in A_i}\|x - v_i\|_2^q\right)^{q^{-1}} \tag{6}$$

where A_i is the set whose elements are the data points assigned to the ith cluster and v_i is the ith clusters prototype. Smaller $S_{i,q}$ values indicate more compact clusters.

The between cluster separation for clusters i and j is

$$d_{ij} = \|v_i - v_j\|. \tag{7}$$

The DBI is defined as

$$DBI(c) = \frac{1}{c}\sum_{i=1}^{c}R_i \tag{8}$$

where

$$R_i = \max_{j\in c, j\neq i}\left\{\frac{S_{i,q} + S_{j,q}}{d_{ij}}\right\}. \tag{9}$$

C. Xie and Beni Index

The Xie and Beni, XB, index is a (*fuzzy*) validity index that uses both the clustering algorithm's objective function value and a measure of separation between the clusters [29]. The XB index is computed as

$$XB(c) = \frac{\frac{J}{n}}{Sep_v} \tag{10}$$

where J is the objective function value, n is the number of data points, and $Sep_v = \min_{ij,i\neq j}\|v_i - v_j\|^2$ is a measure of separation between the clusters. The denominator assumes a single prototype representation where v_k is the cluster prototype for the k^{th} cluster. This index was later extended by Pal and Bezdek in [30].

D. Application of Conventional Cluster Validity Indices to Multiple Sets of Endmembers

In the piece-wise convex model, multiple sets of endmembers are used to describe a hyperspectral scene. Under this model, each set of endmembers can be viewed as defining a cluster. Data points can be associated with one or more sets of endmembers by having non-zero abundances for those endmember sets. The results obtained from the piece-wise convex unmixing algorithms in [12]–[14] are sets of endmembers,

$$\{\{\mathbf{e}_{11}, \mathbf{e}_{12}, \ldots \mathbf{e}_{1M}\}, \{\mathbf{e}_{21}, \mathbf{e}_{22}, \ldots \mathbf{e}_{2M}\}, \ldots, \{\mathbf{e}_{K1}, \mathbf{e}_{K2}, \ldots \mathbf{e}_{KM}\}\},$$

and a proportion vector for each data point,

$$[p_{i11}, p_{i12}, \ldots, p_{i1M}, p_{i21}, p_{i22}, \ldots, p_{iKM}],$$

where i indicates the data point, K is the number of sets of endmembers and M is the number of endmembers in each set. Thus, if a data point is associated with a set of endmembers (cluster), then anywhere from 1 to M proportion values associated with that set of endmembers will be greater than zero.

The piece-wise convex model differs from standard clustering models in that each cluster is represented using a set of endmembers (as opposed to a single cluster prototype). Furthermore, unlike conventional clustering models where each data point has a single membership value for each cluster, in the piece-wise convex model, each data point has a proportion value for each endmember in a set resulting in multiple proportion values per cluster. Due to these differences, conventional cluster validity models are not appropriate for the piece-wise convex model since conventional models assume a single membership per cluster for a data point where, under the piece-wise convex model, each data point has a proportion values for each endmember in a cluster. As discussed previously, conventional cluster validity indices prefer clusters that are well-separated and compact. Generally, clusters are determined to be well-separated by computing the distance

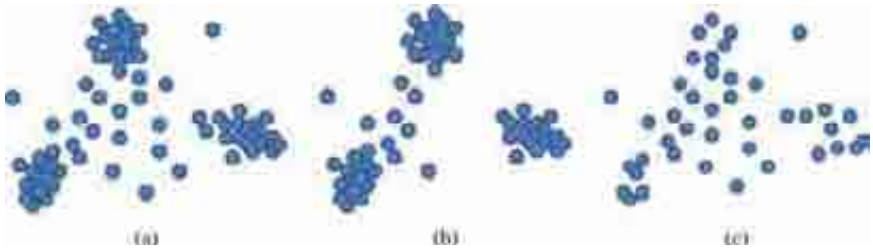


Fig. 4. We assert that scatter plot (a) is a single cluster with three endmembers in dense regions. This cluster exhibits both assumptions, dense endmember locations and a good distribution of samples across the cluster. Scatter plot (b) satisfies the dense endmember region assumption, but there are holes/gaps. How many clusters are there, one, two, etc.? Scatter plot (c) does not exhibit dense endmember regions, but samples are well distributed across the cluster. (a) Dense endmember regions. (b) Holes in convex cluster. (c) Good coverage.

between cluster prototypes (as is done in the DBI and the XB indices). However, in the piece-wise convex model, each cluster has multiple endmembers, not a single prototype. Given that each cluster has multiple prototypes, the separation can be computed between the clusters in a number of ways, e.g., minimum (maximum) linkage, a function of the distance between endmembers, summarizations of endmembers such as reduction to a single prototype, etc. In this work, the separation between clusters is found using Euclidean distance between the mean of the endmembers in each cluster.

The second property, compactness, is generally interpreted to mean that a technique should be designed to prefer tight clouds, e.g., small within class/cluster scatter. The most important question addressed in this article is, what does compact mean with respect to piece-wise convex endmember-based clusters? Many endmember extraction methods seek sets of endmembers that enclose a minimum volume while keeping a small reconstruction error between the input data and their estimates using the endmembers proportion values [31], [32]. Under the piece-wise convex model, sets of endmembers with extremely small volume can be obtained by covering the data with many endmember sets. In this work, the piece-wise convex cluster validity index trades off between minimum volume and a minimum number of needed endmembers sets.

Many other cluster validity indicies exist, e.g., Fukuyama and Sugeno index, Zahid *et al.* index, fuzzy hypervolume, etc. However, for similar reasons as already mentioned, none are necessarily well equipped to address endmember-based clustering. In the following section, we define an extension to the DBI for endmember clustering.

III. New Cluster Validity Index

The previous section identified a set of difficulties for conventional cluster validity indices with respect to multiple sets of endmembers for spectral unmixing. Namely, the interpretation of the concept *compact* in this mixing model and multiple endmember-based representations of clusters are not addressed in conventional cluster validity indices. The development of our validity index is based on the following two properties. Fig. 4 illustrates these different concepts.

1) Endmembers may be located in dense regions of the data. For example, each endmember could be represented using a Gaussian cloud to encapsulate the spectral variability of each endmember [e.g., Fig. 4(a) and (b)]. An index is needed that can address both endmembers representations which have variation and are located in a dense region of the data and those that do not [e.g., Fig. 4(c)]. This property addresses unmixing methods which allow for spectral variability of an endmember material as done in the use of endmember distributions as in [13], [14], the Normal Compositional Model [33], and others [34]–[36]. Currently, our index only addresses piece-wise convex results. However, since several recent endmember extraction methods in the literature estimate endmembers that incorporate spectral variability, our index allows for the property that endmembers may be located in dense regions of the data (i.e., surrounded by several pixels that have proportion values of 1 (or close to 1) for the same endmember).
2) Hyperspectral data within a convex cluster are *well* distributed. Specifically, clusters should not exhibit any significant *holes* or *gaps* [e.g., Fig. 4(a) and (c) versus Fig. 4(b)]. The set of points are not assumed to be uniformly distributed across a cluster. This property addresses the trade-off needed between minimum-volume and a minimum number of clusters to represent the data. If a cluster has significant holes, this shows that the materials may be physically but spatially unrelated and, so, no mixing occurs between these endmembers. In these cases, a cluster should be split.

An additional characterization of hyperspectral mixing would include highly mixed data with sparse endmember regions. The data for this characterization would be similar to Fig. 4(c) with endmembers far from the data. The proposed cluster validity index, however, prefers minimum volume fit over endmembers far from the data. This is a common assumption in many hyperspectral endmember extraction algorithms. No conventional cluster validity index sufficiently address these two properties (if at all). The PC and CE indices are overly simple, i.e., they only consider the terminal partition matrix (proportion matrix), and direct extensions of the DBI and XB (through the use of the proportion matrix and the mean of the endmembers for each cluster as the cluster representative) are also not well suited for endmember clustering. This is primarily due to their set of working assumptions, namely the interpretation of the concept *compact* which is exacerbated by the need to reduce the set of endmembers per cluster into a single point representation (e.g., mean of the endmembers for each cluster).

The proposed index is based on the DBI. The DBI cluster separation measure, d_{ij}, is still applied, where the set of endmembers for each cluster are collapsed into a single point,

e.g., the mean, and a distance formula between the cluster endmember means are used. However, the concept of *compact*, e.g., S_{ij}, in the DBI is replaced. An adjustable formula that measures the two earlier asserted assumptions: possible (but not necessary) dense endmember regions and no holes/gaps is applied. The proposed measure relies on the second property (well distributed) and allows for, but does not enforce, the first (dense endmember regions). Algorithm 1 is the first step in replacing S_i (*compact*).

Algorithm 1 *Error* within cluster i: The amount of holes/gaps in cluster i

1: Assign parameter K_1 where $0 < K_1 \leq n$ for the K-nearest neighbor algorithm (KNN)
2: Assign parameter T for the integer number of subdivisions to perform on cluster i
3: Subdivide cluster i and produce a set, S_{div}, of samples using Algorithm 2
4: **for** each element $s_j \in S_{div}$ **do**
5: Find the K_1 nearest neighbors of s_j with respect to X
6: Aggregate the K_1 nearest neighbor distances using an ordered weighted average (OWA) resulting in d_j^{in}
7: **end for**
8: Aggregate subdivision sample OWA scores, $\{d_1^{in}, ..., d_{|S_{div}|}^{in}\}$, using an OWA, resulting in S_i^{in}

Algorithm 1 measures the extent of *holes/gaps* in a cluster. That is, a sampling is induced in the cluster and holes are associated with a high score (*error*), thus large observed KNN-based distances. Algorithm 1 starts by subdiving the (convex) cluster using the recursive pair-wise subdivision defined in Algorithm 2.

Algorithm 2 Sampling of points in cluster i using recursive subdivision

1: Initilize the set W to the set of endmembers for cluster i
2: Set the number of subdivisions to M
3: **for** i = 1 to M **do**
4: Initilize $Q = \phi$
5: **for** each pair $(r, t) \in WxW$ **do**
6: Create a new sample p 1/2 way between samples r and t
7: **if** $p \notin W$ **then**
8: Add p to Q
9: **end if**
10: **end for**
11: Set $W = Q \cup W$
12: **end for**

It is important to note that clusters come in different shapes and sizes, even in the same data set. Therefore, *scale* versus *exact spacing* is of concern. That is, a method such as a Possision distribution of points uniformly spaced with constant distance across clusters is not used. Instead, Algorithm 2 subdivides a cluster a consistent number of times and samplings are proportionally intact at the different cluster scales.

Once the samples are generated for a cluster, the KNN algorithm is used to identify the K nearest points in X relative to each sampled point. These distances are aggregated using an ordered weighted average (OWA) [37]. An OWA is a specific instance of the Choquet fuzzy integral, where in the fuzzy measure sets of equal cardinality have equal fuzzy measure. The first step in the OWA is sorting, in descending order, of the values (distances), $c_1 \geq \cdots \geq c_n$, where (i) is the ith sorted index. The OWA also takes a set of weights, $w_1, \ldots, w_n$, that generally sum to one or the calculation is normalized by the sum of the weights. The first weight, w_1, is for the largest sorted value, c_1, the second weight, w_2, is for the second largest value, c_2, etc. The OWA is

$$OWA(\{w_1, \ldots, w_n\}, \{c_1, \ldots, c_n\}) = \sum_{i=1}^{n} (w_i c_{(i)}). \quad (11)$$

Using an OWA, one can specify a variety of familiar concepts such as the maximum ($w = (1, 0, \ldots, 0)$), minimum ($w = (0, \ldots, 0, 1)$), mean ($w = (1/n, \ldots, 1/n)$), median ($w = (0, \ldots, 1, \ldots, 0)$), etc.

Algorithm 1 calculates, for each induced sample point in a cluster, an OWA-based aggregation value of the K nearest neighbor distances as an *error* estimate. For clusters without holes or gaps, this value will be near 0 (i.e., the smaller the better). For the experiments performed in this article, the average *error*, i.e., $w = (1/n, \ldots, 1/n)$, was calculated per sample. Next, another OWA is used to aggregate these individual *error* estimates across the entire cluster. An OWA that assigns a weight of 0.5 to the largest error (distance), 0.25 to the next largest error, 0.15, 0.1 and 0 elsewhere is used herein. This OWA is a *soft max*. However, these parameters (weights) can be changed to model different concepts such as penalizing *any* gaps, *a few* gaps, *many* gaps, etc.

A second measurement is needed to make sure that the cluster is *large enough* and *spans* our input hyperspectral data using Algorithm 3. Algorithm 3 ensures that the input hyperspectral data is well-represented by the multiple sets of endmembers and proportion values.

Algorithm 3 *Error* outside of cluster i

1: Assign parameter K_2 where $0 < K_2 \leq n$ for the K-nearest neighbor algorithm (KNN)
2: Assign parameter T to an integer number of subdivisions to perform on cluster i
3: Subdivide cluster i and produce a set, S_{div}, of samples using Algorithm 2
4: **for** each element $x_k \in X$ **do**
5: Find the K_2 nearest neighbors for x_k from S_{div}
6: Aggregate the KNN distances using an OWA operator, resulting in d_k^{out}
7: **end for**
8: Aggregate the set of subdivision sample OWA scores, $\{d_1^{out}, ..., d_n^{out}\}$, taking into account the proportion (partition) matrix, U, resulting in S_i^{out} using Equation 12

The big difference between Algorithms 1 and 3 is the *direction* of the comparison. Algorithm 1 compares each sampled

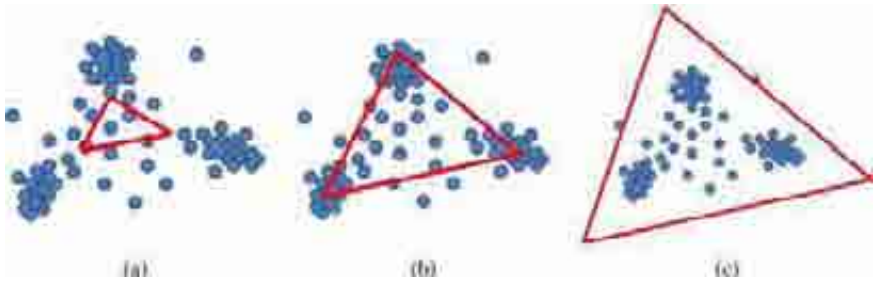


Fig. 5. The convex cluster in (a) is too small, that is it does not fit the data set well. The convex cluster in (b) is a better fit of the convex cluster to the data set. The convex cluster in (c) is too large, that is it is not a tight fit. Our proposed S_i^{in} calculation helps avoid the (c) scenario. Our proposed S_i^{out} calculation helps avoid the (a) scenario. (a) Too small of fit. (b) Good fit. (c) Too big of fit.

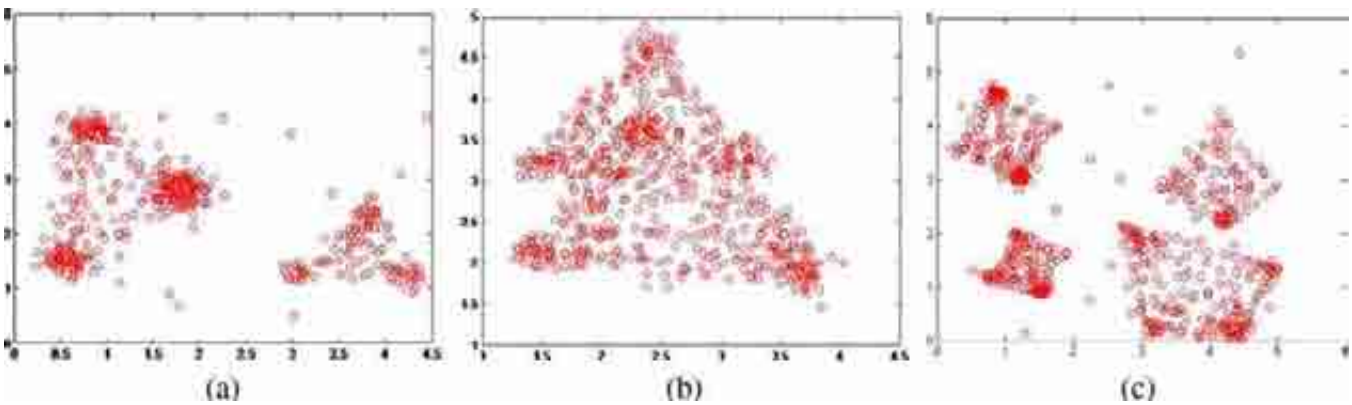


Fig. 6. Three simulated data sets hand crafted for analysis of the behavior of the proposed validity index. Data set 1 (a) has dense regions around endmembers. Data set 2 (b) shows overlapping clusters and data set 3 (c) is four clusters with four endmembers and no significantly dense regions around endmembers. This data is included to show that the proposed index can handle varying number of endmembers and clusters with a data set that can be visualized.

point in a cluster to X. Algorithm 3 compares each sample from X to the set of samples generated for each cluster. The combination of the two checks informs how well *fit* a cluster is to the data. For the experiments performed in this article, a $K_2 = 1$ (distance to the nearest sampled point) was selected. Therefore, no OWA aggregation is required. However, if a K_2 greater than 1 is desired, an OWA is recommended for aggregation as in Algorithm 1, i.e., average or soft max. In Algorithm 3, which results in S_i^{out}, step 8 is calculated as

$$S_i^{out} = \frac{\sum_{k=1}^{n} (d_k^{out} * U_{ik})}{\sum_{k=1}^{n} U_{ik}}. \tag{12}$$

This formula takes into account the distance and how much each point belongs (i.e., fuzzy or possibilistic membership degree) to the particular cluster under investigation.

Our modified DBI, referred subsequently as DBI', for multiple endmember-based clustering replaces R_i with R_i'.

$$R_i' = \max_{j \in c, j \neq i} \left\{ \frac{S_i^{in} + S_j^{in} + S_i^{out} + S_j^{out}}{d_{ij}} \right\}. \tag{13}$$

As in the original DBI, the goal is to minimize the numerator and maximize the denominator of DBI'. Fig. 5 illustrates our replacement of the concept 'compact' in the DBI.

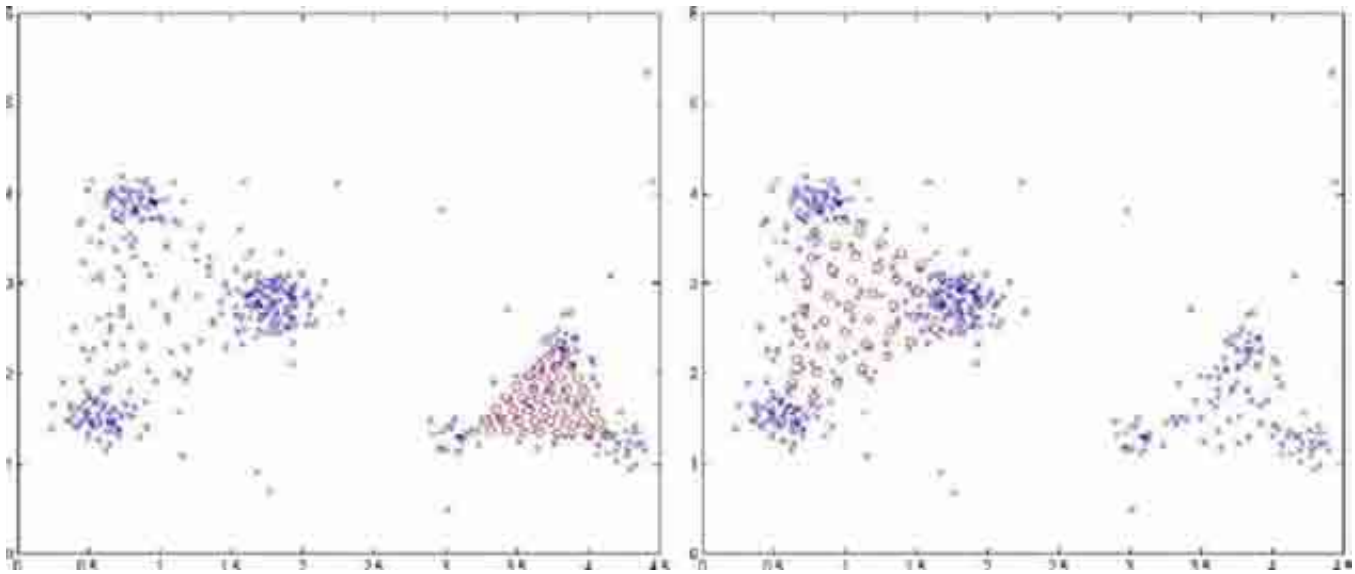
The following is in regard to computational efficiency. Our method requires one to run a clustering algorithm various times using different selections of parameters (e.g., different c). For large data sets, running a clustering algorithm multiple times (likely in search of multiple parameters) turns out to be the most expensive step in cluster validity. However, all cluster validity measures require a clustering algorithm to be run multiple times and these different executions can be carried out in parallel. We comment on efficiency as it relates to the proposed algorithms 1, 2 and 3. Algorithm 2 (sampling of points in a cluster) is just recursive subdivision. This is independent of the data set X, it can be carried out in parallel across clusters and it can also be performed in parallel at each level of subdivision. The most expensive step in our approach is the K-NN in algorithm 1. For each element in the subdivision set (generally a small number), the K_1 nearest neighbors from X (a large set) are identified. In algorithm 3 we also use the K-NN, however it is from each sample in X to the subdivision set. In both cases, the K-NN searches can be carried out in parallel. Additionally, computation of the OWA and the final index value are negligible. The efficiency of our algorithm resides with the K-NN, a common procedure used heavily in pattern recognition and large data sets.

IV. Experiments: Simulated Data Sets

Since clustering is unsupervised learning and not classification and, with real hyperspectral data, the true solution is generally unknown, simulated data sets were generated using known endmember values to test the proposed validity index.

TABLE I
VALIDITY INDEX RESULTS FOR DATA SETS 1–3

(a) Data Set 1				(b) Data Set 2				(c) Data Set 3			
Index	c	m	alpha	Index	c	m	alpha	Index	c	m	alpha
CE	2	4	0.1	CE	2	4	0.1	CE	2	4	0.1
DBI	4	2	0.7	DBI	4	2	0.7	DBI	3	2	0.7
XB	3	2	0.1	XB	3	2	0.1	XB	2	5	0.1
DBI'	2	3	0.4	DBI'	2	3	0.4	DBI'	4	4	0.4



(a) (b)

Fig. 7. Convex cluster subdivision results for the determined optimal parameter set of $c = 2$, $m = 3$ and $\alpha = 0.4$. (a) Subdivision of cluster 1. (b) Subdivision of cluster 2.

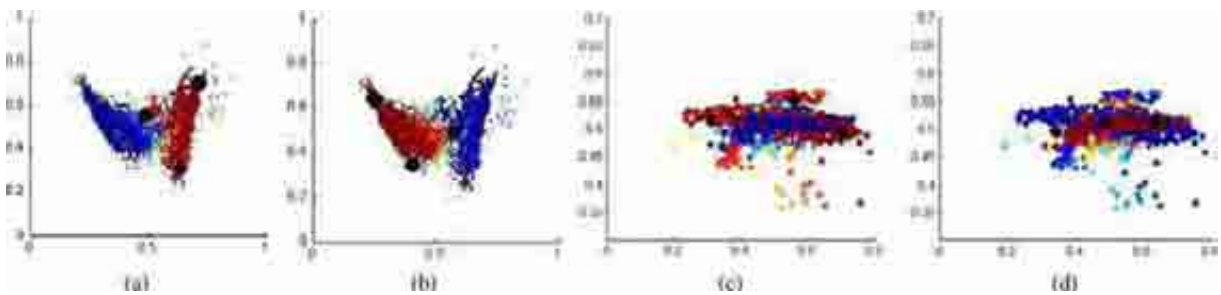


Fig. 8. Scatter plots of PCA projection of Indian Pines data set, location of the endmembers for the two clusters (large black dots), and mixture values (red is high and blue is low membership) with PFCMPCE algorithm parameters $c = 2$, $m = 3$, $\alpha = 0.1$. Result selected by the CE and DBI' indices. (a & b) The x- and y- dimensions are the first two principal components. (c & d) The x- and y- dimensions are the second and third PCs of the data. (a) Partition 1: Contains points from Corn, Soybean, Hay-windrowed classes. (b) Partition 2: Contains points from Woods, Grass classes. (c) Partition 1: Contains points from Corn, Soybean, Hay-windrowed classes. (d) Partition 2: Contains points from Woods, Grass classes.

A. Description of Simulated Data Sets

Three simulated data sets were generated to test and illustrate the proposed validity index. Fig. 6 displays the three two dimensional simulated data sets (scatter plots) that were generated. They were generated manually via a Matlab graphical user interface. Thus, a variety of interesting non-Gaussian convex cluster scenarios can be created and explored at will. Data set 1 contains 455 elements, data set 2 is of size 670 and data set 3 is of size 639. The size of these data sets is relatively small due to the manual process of creation. However, in the next section (real data set) our algorithm is run on a large data set. All three data sets were generated as two-dimensional data – no pre-processing steps were applied to the data. For all of the

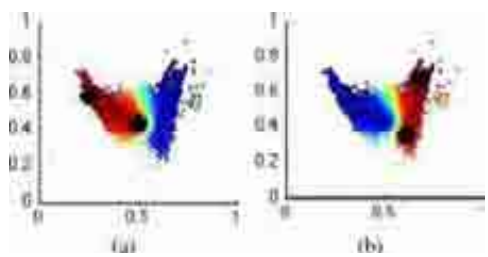


Fig. 9. Scatter plots of PCA projected (first two principal components) Indian Pines data, location of endmembers for the two clusters (black dots), and mixture values (red is high and blue is low membership) for PFCMPCE algorithm parameters $c = 2$, $m = 2$, $\alpha = 0.1$. Result found by DBI index. (a) Partition 1. (b) Partition 2.

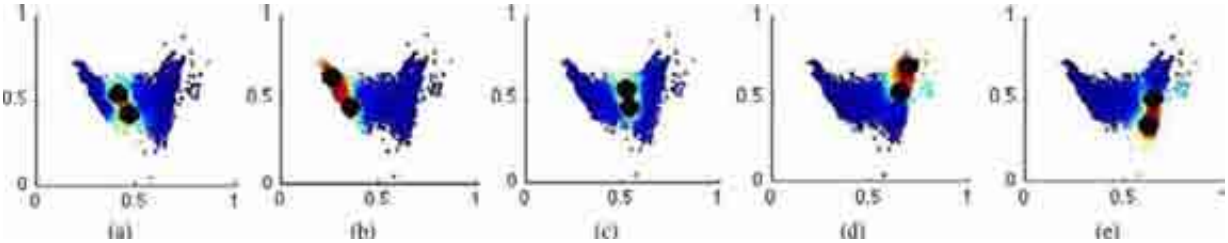


Fig. 10. Scatter plots of the PCA projected (first two principal components) Indian Pines data, location of endmembers for the two clusters (black dots), and mixture values (red is high and blue is low membership) for PFCMPCE algorithm parameters $c = 5$, $m = 2$, $\alpha = 0.1$. Result found by XB index. (a) Partition 1. (b) Partition 2. (c) Partition 3. (d) Partition 4. (e) Partition 5.

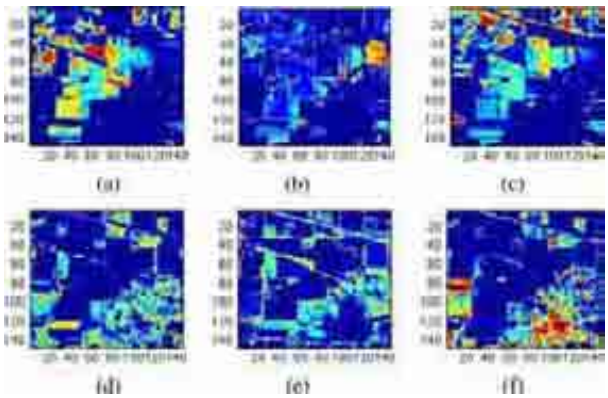


Fig. 11. Proportion maps of Indian Pines data set with PFCMPCE algorithm parameters of $c = 2$, $m = 3$, $\alpha = 0.1$. Result selected by CE and DBI' indices. Color scale of these maps range from 0 (dark blue) to 1 (red). (a) Soybean, Corn. (b) Haywin, Alfalfa. (c) Soybean, Corn. (d) Woods, Wheat. (e) Grass, Pasture. (f) Woods, Trees.

results shown, the piece-wise convex spectral unmixing algorithm, PFCMPCE, described in [12] was applied with varying parameters values. The parameter values for the number of clusters, c, the number of endmembers per cluster, m, and the regularization parameter, α, for trading off between volume and error were varied. The proposed cluster validity index, DBI' and the conventional cluster validity indices CE, DBI, and XB were applied to all of the spectral unmixing results across the range of parameter values. The results in this section list the parameter set that was selected by each validity index as providing the best result. As shown below, the proposed validity index consistently selects the best parameter set across all data sets whereas conventional cluster validity indices do not.

B. Data Set 1: Dense Regions in the Scatter Plot Around Endmembers

The first simulated data set is an example of dense regions at endmembers. However, only one cluster has dense regions around its endmembers. The other has its points distributed more uniformly across the cluster. The point of this first example is to show that the proposed measure can handle cases in which endmembers are or are not dense regions in space. The other validity indices are not well suited for this task. Table I(a) shows the results.

The set of algorithm parameters explored in generating Table I include; $c = \{2, 3, \ldots, 6\}$, $m = \{2, 3, 4, 5\}$ and $\alpha = \{.001, 0.1, 0.4, 0.7\}$. We used a non-optimized and non-parallel Matlab implementation of the proposed index

TABLE II
VALIDITY INDEX RESULTS FOR THE INDIAN PINES DATA SET

Index	c	m	alpha
CE	2	3	0.1
DBI	2	2	0.1
XB	5	2	0.1
DBI'	2	3	0.1

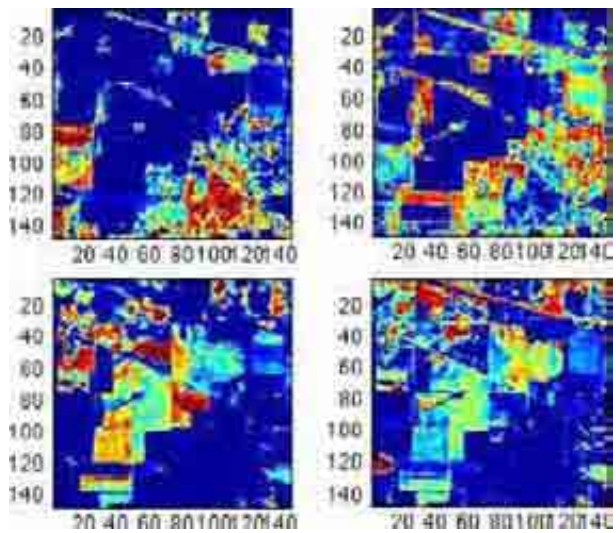


Fig. 12. Proportion maps of Indian Pines data set with PFCMPCE algorithm parameters of $c = 2$, $m = 2$, $\alpha = 0.1$. Result selected by the DBI index. Cp scale of these maps range from 0 (dark blue) to 1 (red).

and PFCMPCE clustering algorithm. These parameters were selected in part based on visualization of the PCA projected scatter plots. However, the number of combinations arising from this set of parameters covers a wide range of clustering outcomes (e.g., 6 clusters with 5 endmembers, 6 clusters with 2 endmembers, 2 clusters with 5 endmembers, etc.). On an empirical note, the index values were *degrading* (i.e., increasing or respectively decreasing based on the CV index used) outside of the optimal parameter set reported. The values reported in Table I are the parameters for the optimal solution according to the respective index, i.e., minimum for CE, minimum for DBI and DBI' and maximum for XB. As Table I(a) demonstrates, the proposed modified DBI' is the only index that finds the correct answer (two clusters with three endmembers each). The CE index finds the correct number of clusters, but not the right number of endmembers. The unmodified DBI prefers to separate the cluster with dense endmembers into three clusters and the cluster on the right into one. The XB does not find the

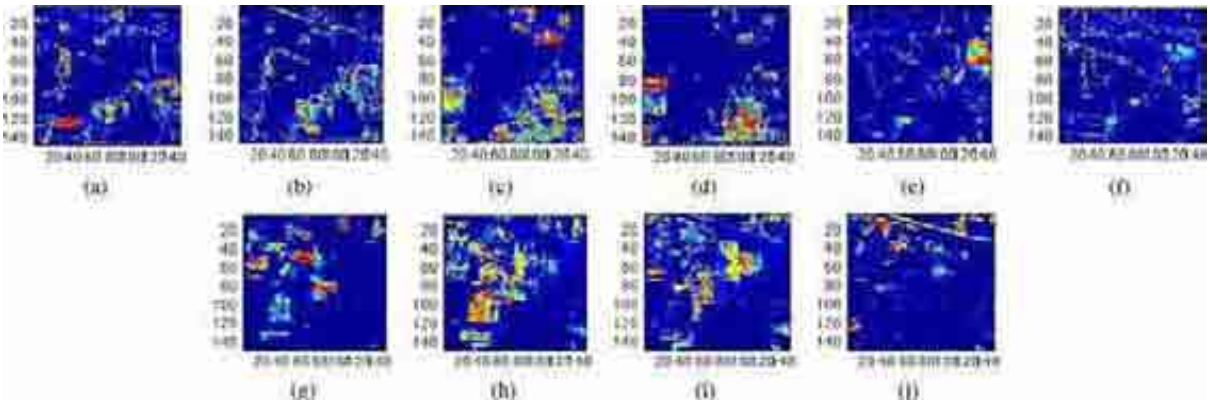


Fig. 13. Proportion maps of Indian Pines data set with PFCMPCE algorithm parameters of $c = 5$, $m = 2$, $\alpha = 0.1$. Result selected by the XB index. The color scale of these maps range from 0 (dark blue) to 1 (red).

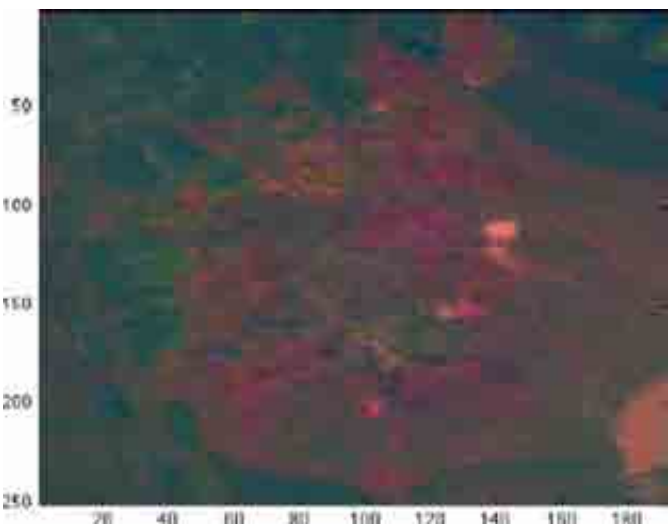


Fig. 14. Subset of the AVIRIS Cuprite data set shown as an RGB figure generated using the first three principal components of the data where (R) is PC 1, (G) is PC 2 and (B) is PC 3.

right solution either. Again, depending on if the endmembers exist in dense regions or not, the proposed index is able to address this behavior. Fig. 7 shows the subdivision result of the convex cluster for the optimal parameter set.

C. Data Set 2: Highly Overlapping Clusters

The second synthetic data set is an example of highly overlapping clusters, which is common in real-world natural scenes. Table I(b) are the results. Again, the new index, DBI', is the only one that finds the correct number of clusters.

D. Data Set 3: Four Clusters With Four Endmembers Each

The third data set is included in order to demonstrate that the index does indeed find the correct number of endmembers. In data set 3, both the number of endmembers and clusters is increased to four. Table I(c) shows the results.

Once again, the new index, DBI', finds the asserted correct number of clusters and endmembers where conventional cluster validity methods do not.

TABLE III
VALIDITY INDEX RESULTS FOR THE CUPRITE DATA SET

Index	c	m	alpha
CE	2	4	0.1
DBI	2	5	0.1
XB	2	2	0.1
DBI'	2	3	0.1

V. EXPERIMENTS CASE 2: AVIRIS INDIAN PINES DATA

In the previous section, the performance of the proposed index on a variety of different simulated data sets was explored. In this section, the indices are applied to the AVIRIS Indian Pines data set shown in Fig. 1.

A. Description of the AVIRIS Indian Pines Data Set

The June 1992 AVIRIS Indian Pines "Scene 4" data set was collected over the Indian Pines test site in an agricultural area of northern Indiana. The image is 145×145 pixels and has 220 spectral bands ranging from 0.4 to 2.5 microns. The spatial resolution is 20 m. The data contains approximately two-thirds agricultural land. These agricultural fields were imaged during early growth stages and have roughly 5% crop cover. Each of these fields have varying levels of residue from the previous year's crop; the "min," "notill," "clean," labels indicate the amount of crop residue remaining on the fields. Since many of these crops are crop-rotated, it is likely that the residue from the previous year's crops are not the same type of crop of which the field is labeled. Since each of these fields have very low crop cover and consists also of soil and residue, this data set is considered a highly-mixed data set.

Prior to applying piece-wise convex unmixing algorithms to this data, principal components analysis (PCA) was applied to reduce the dimensionality to the top 3 PC bands. PCA was applied such that the results could be visualized and manually evaluated. In the following, results are shown using scatter plots of the first two PC components of the Indian Pines data set. These provide a good representation of the results found using three PC components. Although results are shown here using PCA-reduced data, the validity index can be applied to hyperspectral data with full spectral resolution.

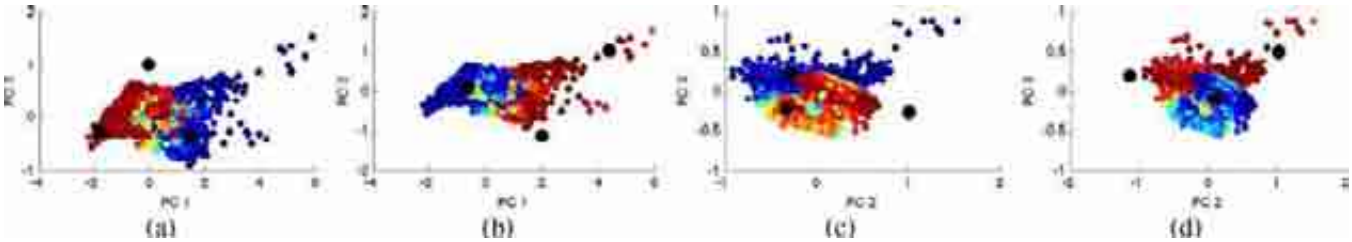


Fig. 15. Scatter plots of the PCA projection of the Cuprite data set, the location of the endmembers for the two clusters (large black dots), and the resulting mixture values (red is high membership and blue is low membership) with parameter values for the PFCMPCE algorithm set to $c = 2$, $m = 3$, $\alpha = 0.1$. Result selected by the DBI' indices. The x- and y- dimensions in these plots are the first two PCs or the second two PCs of the data set. (a) Partition 1. (b) Partition 2. (c) Partition 1. (d) Partition 2.

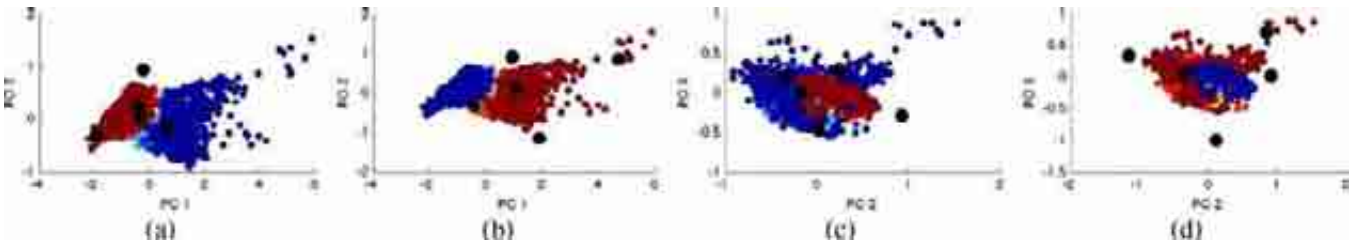


Fig. 16. Scatter plots of the PCA projection of the Cuprite data set, the location of the endmembers for the two clusters (large black dots), and the resulting mixture values (red is high membership and blue is low membership) with parameter values for the PFCMPCE algorithm set to $c = 2$, $m = 5$, $\alpha = 0.1$. Result selected by the DBI indices. The x- and y- dimensions in these plots are the first two PCs or the second two PCs of the data set. (a) Partition 1. (b) Partition 2. (c) Partition 1. (d) Partition 2.

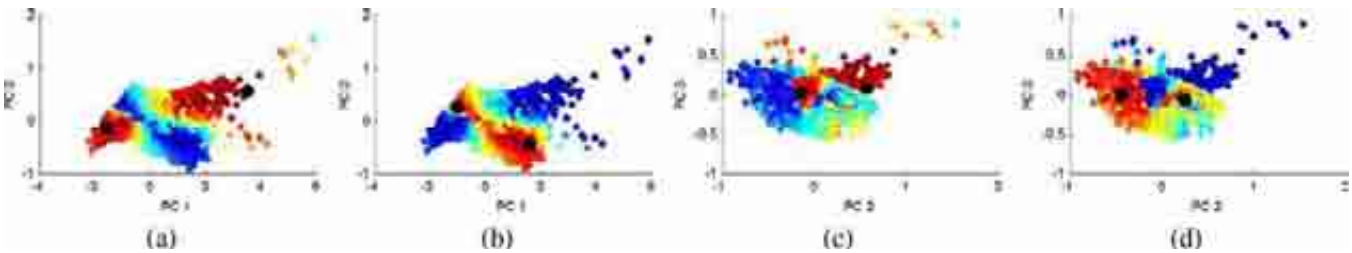


Fig. 17. Scatter plots of the PCA projection of the Cuprite data set, the location of the endmembers for the two clusters (large black dots), and the resulting mixture values (red is high membership and blue is low membership) with parameter values for the PFCMPCE algorithm set to $c = 2$, $m = 2$, $\alpha = 0.1$. Result selected by the XB indices. The x- and y- dimensions in these plots are the first two PCs or the second two PCs of the data set. (a) Partition 1. (b) Partition 2. (c) Partition 1. (d) Partition 2.

In the AVIRIS Indian Pines experiment, the parameters were varied from $c = \{2, 3, \ldots, 6\}$, $m = \{2, 3, 4, 5\}$ and $\alpha = \{.001, 0.1, 0.4, 0.7\}$. This data set contains a greater number of elements, 21,025 samples of dimension 220, versus the largest of the simulated data sets (i.e., 670 samples and dimension 2). Table II lists the cluster validity index results for the Indian Pines data set.

Figs. 8 and 11 display the results selected by the CE and DBI' indices. The piece-wise convex unmixing results are displayed by showing the endmembers and partitions selected by the validity index in the scatter plot (endmembers are shown in black and the color of each point indicates the corresponding pixel's membership in the corresponding convex region. Also the proportion maps associated with the selected endmember set are displayed). Figs. 9 and 12 display the piece-wise convex unmixing results selected by the DBI index. Figs. 10 and 13 display the piece-wise convex unmixing results selected by the XB index. Examining the scatter plots in Figs. 8–10, the best fit on the data is found in Fig. 8 and selected by the proposed DBI' index. When examining the proportion maps. the results in Fig. 8 separate regions in the data without over-segmentation as done in Fig. 10 or under-segmentation in Fig. 9. These results were evaluated by visually examining the location the selected endmembers among the PCA projected data points. Furthermore, these results were also evaluated by comparing the groundtruth map shown in Fig. 1 to the estimated proportion maps. Results that grouped Soybean and Corn, and Hay and Alfalfa, and Woods and Trees in the same convex region (and, in some cases, endmembers) were determined to appropriately partition the data without over segmentation. Also, consider the Hay-windrowed class discussed in Section I of this paper. The result chosen by the proposed index finds this interior endmember and the associated proportion map is shows in Fig. 11(b). A similar endmember is also estimated in results selected by the XB index as shown in the proportion map Fig. 13(e). We also showed, in the simulated data set experiments, that the CE index is not theoretically well-suited in general for a wide range of cluster scenarios. However, for the Indian Pines data set the CE index does find the correct number of clusters.

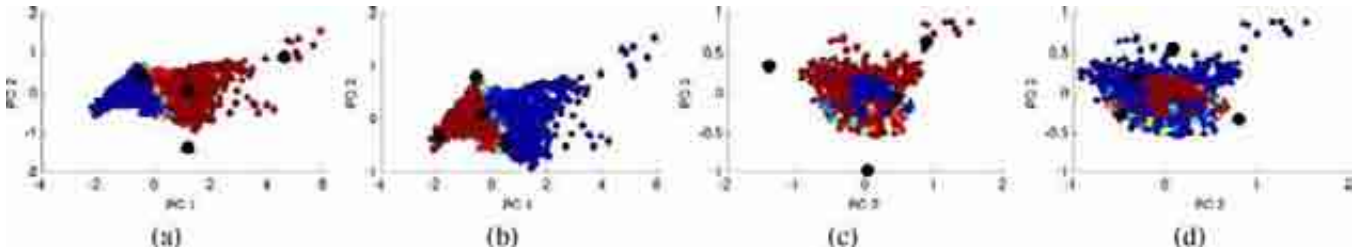


Fig. 18. Scatter plots of the PCA projection of the Cuprite data set, the location of the endmembers for the two clusters (large black dots), and the resulting mixture values (red is high membership and blue is low membership) with parameter values for the PFCMPCE algorithm set to $c = 2$, $m = 4$, $\alpha = 0.1$. Result selected by the CE indices. The x- and y- dimensions in these plots are the first two PCs or the second two PCs of the data set. (a) Partition 1. (b) Partition 2. (c) Partition 1. (d) Partition 2.

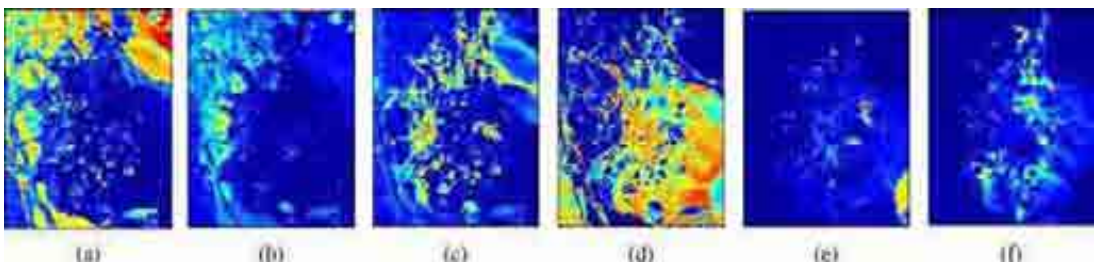


Fig. 19. Proportion maps of Cuprite data set with PFCMPCE algorithm parameters $c = 2$, $m = 3$, $\alpha = 0.1$, i.e., (P1,E2) is product map of partition 1 and endmember 2. Result selected by DBI'. Color scale of maps range from 0 (dark blue) to 1 (red). (a) (P1,E1). (b) (P1,E2). (c) (P1,E3). (d) (P2,E1). (e) (P2,E2). (f) (P2,E3).

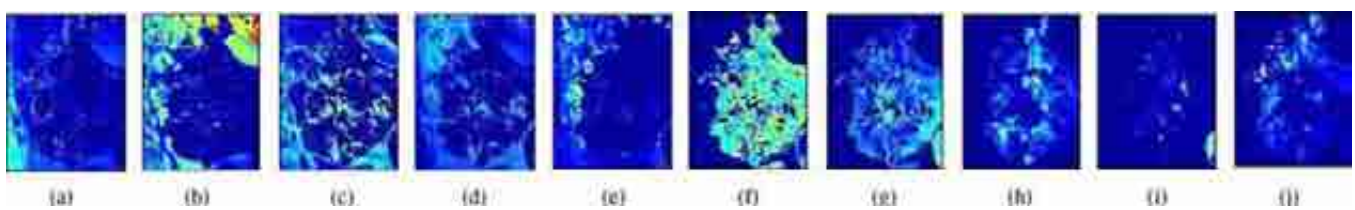


Fig. 20. Proportion maps of Cuprite data set with PFCMPCE algorithm parameters $c = 2$, $m = 5$, $\alpha = 0.1$, i.e., (P1,E2) is product map of partition 1 and endmember 2. Result selected by DBI. Color scale of maps range from 0 (dark blue) to 1 (red). (a) (P1,E1). (b) (P1,E2). (c) (P1,E3). (d) (P1,E4). (e) (P1,E5). (f) (P2,E1). (g) (P2,E2). (h) (P2,E3). (i) (P2,E4). (j) (P2,E5).

VI. Experiments Case 2: AVIRIS Cuprite Data

For sake of completeness, we have included a second real data set, a subset of the AVIRIS Cuprite data set (shown in Fig. 14). With respect to interpreting these clustering results, we have provided the scatter plots and proportion maps for the different CV indices analyzed herein.

A. Description of the AVIRIS Cuprite Data Set

The AVIRIS hyperspectral data scene was collected with the AVIRIS sensor over Cuprite, NV, a mining area in southern Nevada with little vegetation and visible mineral coverage. A mineral map of the area can be found in [38]. The same sub-image as described in [39] was used. This sub-image is 250×190 in pixels with 224 spectral bands.

Prior to applying piece-wise convex unmixing algorithms to this data, principal components analysis (PCA) was applied to reduce the dimensionality to the top 3 PC bands. In the following, results are shown using scatter plots of the first two PC components and the second two PC components of the Cuprite data set.

In the AVIRIS Cuprite experiment, the parameters were varied from $c = \{2, 3, \ldots, 6\}$, $m = \{2, 3, 4, 5\}$ and $\alpha = \{.001, 0.1, 0.4, 0.7\}$. Table III lists the cluster validity index results for the Cuprite data set.

Figs. 15 and 19 display the scatter plots with the associated endmembers and the proportion maps for each convex region selected by the DBI' indices. Figs. 16 and 20 display the piece-wise convex unmixing results selected by the DBI index. Figs. 17 and 21 display the piece-wise convex unmixing results selected by the XB index. Figs. 18 and 22 display the piece-wise convex unmixing results selected by the CE index. Examining the scatter plots in Figs. 15–18, the best fit on the data is found in Fig. 15 and it was selected by the proposed DBI' index. These results were evaluated by visually examining the location the selected endmembers among the PCA projected data points. In particular, one can see that the DBI' method gives relatively smooth proportion maps with both high and low values as opposed to the results selected by the DBI index which have proportion values grouped around 0.5 (light blues and greens instead of everything from red to blue). In comparison to the results selected by the XB index, both methods given smooth proportion maps with a large range of values. However, comparison of the scatter plots of the XB and DBI' results show that the proposed approach selects an endmember set that better surrounds the input data. Finally,

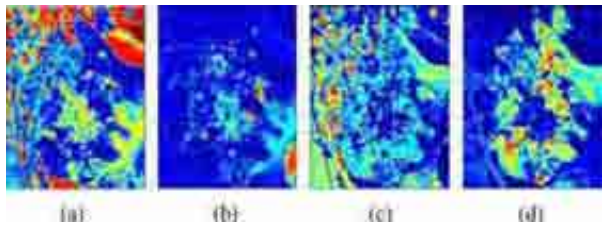


Fig. 21. Proportion maps of the Cuprite data set with PFCMPCE algorithm parameters $c = 2$, $m = 2$, $\alpha = 0.1$, i.e., (P1,E2) is product map of partition 1 and endmember 2. Result selected by the XB indices. The color scale of these maps range from 0 (dark blue) to 1 (red). (a) (P1,E1). (b) (P1,E2). (c) (P2,E1). (d) (P2,E2).

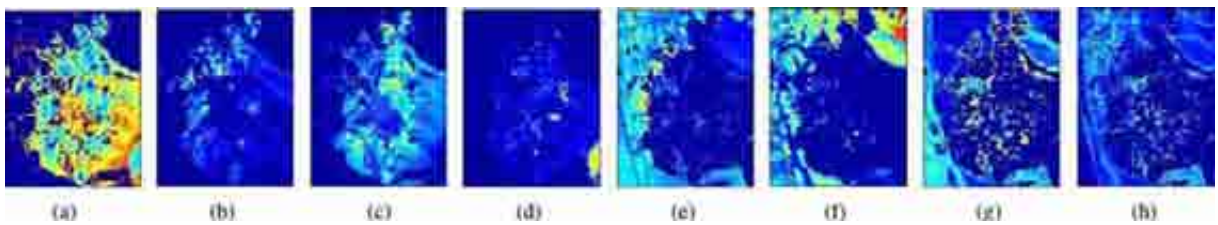


Fig. 22. Proportion maps of the Cuprite data set for PFCMPCE algorithm parameters $c = 2$, $m = 4$, $\alpha = 0.1$, i.e., (P1,E2) is product map of partition 1 and endmember 2. Result selected by the CE indices. The color scale of these maps range from 0 (dark blue) to 1 (red). (a) (P1,E1). (b) (P1,E2). (c) (P1,E3). (d) (P1,E4). (e) (P2,E1). (f) (P2,E2). (g) (P2,E3). (h) (P2,E4).

when examining the scatter plot results of the CE, it can be seen that unnecessary endmembers placed within a partition (rather than at a corner of the data), are included in the result. In comparison, the proposed DBI' index, selects a results with sufficient endmembers to surround the data without the inclusion of extraneous endmembers.

VII. Summary and Future Work

A cluster validity index for convex clusters was developed. The proposed index was applied towards spectral unmixing results using multiple sets of endmembers. Over a variety of piece-wise spectral unmixing results, the proposed index is able to identify appropriate parameter sets that effectively balance the number of sets of endmembers and the volume of the area encapsulated by each set of endmembers. The proposed index consistently identified the minimum number of compact clusters that do not enclose gaps or holes in the data. In contrast, conventional cluster validity indices do not apply to the piece-wise convex model.

Future work includes exploring other methods of measuring cluster separation by explicitly taking into account the entire set of endmembers per cluster, not a reduction of each cluster into a single endmember. Also, comparisons of different OWA operators for the aggregation of distance values will be conducted. Currently, the proposed index assumes multiple sets of endmembers based on the concept that hyperspectral pixels are usually composed of a small subset of the endmembers found in a scene. However, future work will include extending this index to evaluate single convex regions (one set of endmembers). Furthermore, future work will include investigating the trade-off between endmember variation and the number of endmembers per convex region. For example, some convex regions could be represented with one or a small number of endmembers with large variation or equally represented with several endmembers with small variation.

References

[1] J. Silván-Cárdenas and L. Wang, "Fully constrained linear spectral unmixing: Analytic solution using fuzzy sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 3992–4002, Nov. 2010.

[2] M. Shoshany, F. Kizel, N. Netanyahu, N. Goldshlager, T. Jarmer, and G. Even-Tzur, "An iterative search in end-member fraction space for spectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 706–709, Jul. 2011.

[3] A. Villa, J. Chanussot, J. Benediktsson, and C. Jutten, "Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 521–533, Jun. 2011.

[4] W. Xia, X. Liu, B. Wang, and L. Zhang, "Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2165–2179, Jun. 2011.

[5] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4112–4122, Nov. 2011.

[6] A. Plaza, Q. Du, Y.-L. Chang, and R. King, "High performance computing for hyperspectral remote sensing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. (JSTARS)*, vol. 4, no. 3, pp. 528–544, Sep. 2011.

[7] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, pp. 44–57, 2002.

[8] J. B. Greer, S. S. Shen and P. E. Lewis, Eds., "Sparse demixing," in *Proc. SPIE, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI*, Orlando, FL, May 2010, vol. 7695.

[9] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[10] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[11] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[12] A. Zare and P. Gader, "Piece-wise convex spatial-spectral unmixing of hyperspectral imagery using possibilistic and fuzzy clustering," presented at the IEEE Int. Conf. Fuzzy Systems, Taipei, Taiwan, Jun. 2011.

[13] A. Zare and P. Gader, "PCE: Piece-wise convex endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2620–2632, 2010.

[14] A. Zare and P. Gader, "An investigation of likelihoods and priors for Bayesian endmember estimation," presented at the 30th Int. Workshop Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Chamonix, France, Jul. 2010.
[15] P. Dennison and D. A. Roberts, "Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE," *Remote Sens. Environ.*, vol. 87, pp. 123–135, 2003.
[16] D. 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.
[17] K. Canham, A. Schlamm, A. Ziemann, B. Basener, and D. Messinger, "Spatially adaptive hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4248–4262, Nov. 2011.
[18] AVIRIS Free Standard Data Products California Institute of Technology, Jet Propulsion Lab., Pasedena, CA, Sep. 2004 [Online]. Available: http://aviris.jpl.nasa.gov/ql/list95.html
[19] A. Zare and P. Gader, "Hyperspectral band selection and endmember detection using sparsity promoting priors," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 256–260, Apr. 2008.
[20] D. Messinger, A. Ziemann, A. Schlamm, and B. Basener, "Spectral image complexity estimated through local convex hull volume," in *Proc. 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2010)*, University of Iceland, Reykjavik, Iceland, Jun. 2010, pp. 1–4.
[21] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[22] J. Broadwater and A. Banerjee, "A neyman-pearson approach to estimating the number of endmembers," in *Proc. 2009 IEEE Int. Geoscience and Remote Sensing Symp., IGARSS 2009*, Jul. 2009, vol. 4, pp. IV-693–IV-696.
[23] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.
[24] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
[25] D. T. Anderson, J. C. Bezdek, M. Popescu, and J. M. Keller, "Comparing fuzzy, probabilistic, and possibilistic partitions," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 906–918, 2010.
[26] J. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
[27] J. C. Bezdek, "Cluster validity with fuzzy sets," *J. Cybernit.*, vol. 3, pp. 55–72, 1974.
[28] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, pp. 224–227, 1979.
[29] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 841–847, 1991.
[30] N. R. Pal and J. C. Bezdek, "On cluster validity for fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, pp. 370–379, 1995.
[31] M. E. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. Applied Geologic Remote Sensing*, Vancouver, BC, Canada, 1999, pp. 337–344.
[32] 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.
[33] M. T. Eismann and D. Stein, "Stochastic mixture modeling," in *Hyperspectral Data Exploitation: Theory and Applications*, C. Chang, Ed. : Wiley, 2007, ch. 5.
[34] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.
[35] C. Bateson, G. Asner, and C. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.
[36] B. Somers, G. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Environ.*, vol. 115, pp. 1603–1616, 2011.
[37] R. Yager, "Applications and extensions of OWA aggregations," *Int. J. Man–Machine Studies*, vol. 37, pp. 103–122, 1992.
[38] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada," in *Summaries 3rd Annu. JPL Airborne Geosciences Workshop*, 1992, vol. 1, pp. 47–49.
[39] Aviris Cuprite Data Sub-Image. AVIRIS, Jan. 2012 [Online]. Available: http://www.lx.it.pt/ bioucas/code/cuprite_ref.zip

Derek T. Anderson (S'06–M'10) received the Ph.D. degree in electrical and computer engineering (ECE) in 2010 from the University of Missouri, Columbia.

He is currently an Assistant Professor in ECE at Mississippi State University. His research interests include pattern recognition, information fusion, image processing and computer vision, linguistic summarization of human activity from video and clustering.

Alina Zare (S'07–M'08) received the Ph.D. degree from the Department of Computer and Information Science and Engineering from the University of Florida, Gainesville, in December 2008.

She is currently an Assistant Professor in the Electrical and Computer Engineering Department, University of Missouri, Columbia. Her research interests include machine learning, Bayesian methods, sparsity promotion, image analysis, pattern recognition, hyperspectral image analysis, and human geography.