Unsupervised hyperspectral band selection by combination of unmixing and sequential clustering techniques

Sarra Ikram Benabadji, Moussa Sofiane Karoui, Khelifa Djerriri, Issam Boukerch, Nezha Farhi and Mohammed Amine Bouhlala
Centre des Techniques Spatiales, Arzew, Algeria

ABSTRACT
Selecting the decisive spectral bands is a key issue in unsupervised hyperspectral band selection techniques. These methods are the most popular ways for dimensionality reduction of original data. A compact data representation without compromising the physical information and optimizing the separation between different materials are the main objectives of such selection processes. In this work, a hyperspectral band selection approach is proposed based on linear spectral unmixing and sequential clustering techniques. The use of these two specific techniques constitutes the main novelty of this investigation. The proposed approach operates in different successive steps. It starts with extracting material spectra contained in the considered data using an unmixing method. Then, the variance of extracted spectra samples is calculated at each wavelength, which results in a variances vector. This one is segmented into a fixed number of clusters using a sequential clustering strategy. Finally, only one spectral band is selected for each segment. This band corresponds to the wavelength at which a maximum variance value is obtained. Experiments on three real hyperspectral data demonstrate the superiority of the proposed approach in comparison with four methods from the literature.

Introduction
Earth observation based on remote sensing hyperspectral imagery has been widely used in recent years for better understanding and managing our natural resources and our environment. Hyperspectral imaging sensors simultaneously acquire spectral bands with wavelengths ranging from the visible spectrum to the infrared region (Landgrebe, 2002). Obtained data are useful to several remote sensing applications, such as image classification and target detection. Indeed, these data increase the opportunity for better identification of spectrally unique materials (Chang, 2007).

The high number of spectral bands can result in an excessive computational complexity of hyperspectral data analysis techniques. More precisely, the obtained detailed spectra lead to an increase in the dimensionality of the data that brings the “Hughes phenomenon” for the classification process (Hughes, 1968; Kim & Landgrebe, 1991). This phenomenon significantly impacts the performance of supervised classifiers (Jimenez & Landgrebe, 1998). Thus, reducing the dimensionality of these data is a simpler way to address such problems.

Feature extraction and feature selection are two ways to perform dimensionality reduction. Feature extraction methods (Imani & Ghassemian, 2014; Marpu et al., 2012; Mojaradi, Moghaddam, Zoej, & Duin, 2009; Yang, Yu, & Kuo, 2010) transform the spectral bands of a hyperspectral image to a low dimensional feature through projection techniques. Although these projections preserve the maximum information of the original data, the obtained features lack physical interpretability. On the contrary, feature selection (also referred to as band selection) methods (Das, Ghosh, & Ghosh, 2014; Feng, Itoh, Parente, & Duarte, 2017; Guo, Damper, Gunn, & Nelson, 2008; Martinez-Uso, Pla, Sotoca, & Garcia-Sevilla, 2007; Yang, Liu, Brizzzone, Guan, & Du, 2013; Zhan, Hu, Xing, & Yu, 2017; Zhu, Huang, Li, Tang, & Liang, 2017) choose some representative spectral bands that preserve the physical information of the original data. These band selection methods are, in some situations, preferable than feature extraction techniques, especially when some materials need to be recognized using spectral libraries, which contain physical information about these materials. Also, band selection methods can be very useful to design new sensors by an initial choice of decisive spectral bands, which allow a maximum discrimination between different materials.

Depending on the availability of labeled reference data, band selection can be performed with supervised or unsupervised ways. Supervised methods (Backer, Kempeneers, Debruyne, & Scheunders, 2005;
Bruzone & Persello, 2009; Shi, Shen, & Liu, 2003; Yang, Du, Su, & Sheng, 2011) require the collection of reliable labeled samples, which is a difficult and tedious task. A supervised mutual-information-based band selection approach is proposed in Guo, Gunn, Damper, & Nelson (2006). Also, an approach that uses a spectral separability index is reported in Yin, Wang, & Zhao (2010). In Yang et al. (2011), a supervised band selection method, which uses known spectra without examining original spectral bands of the considered data, is considered.

Supervised band selection methods can generally achieve better selection than unsupervised ones, and that by considering some performance criteria such as the overall accuracy of the classification of selected spectral bands. However, unsupervised approaches (Amri, Benabadji, & Karoui, 2017; Chang, Du, Sun, & Althouse, 1999; Du & Yang, 2008; Rashwan & Dobigeon, 2017; Xu, Shi, & Pan, 2017; Zhang, Ma, & Gong, 2017) may be more suitable in applicative scenarios since they do not require any labeled data to perform the band selection process. Many of these unsupervised methods are based on the clustering strategy. For example, a hierarchical-clustering-based band selection approach is introduced in Martinez-Uso et al. (2007), which consists in grouping spectral bands to minimize the intra-cluster variance while maximizing the inter-cluster variance. Also, in Cariou, Chehdi, & Moan (2011), an original approach that splits initial spectral bands into disjoint clusters is reported. Finally, an unsupervised approach based on band clustering by split-and-merge stages is introduced in Rashwan & Dobigeon (2017).

In this paper a new unsupervised hyperspectral band selection approach is proposed, wherein linear spectral unmixing and sequential clustering techniques are combined. The designed approach operates in two principal stages. The first one consists of extracting, from the considered data, pure material spectra (also called endmember), while the second stage aims at sequentially clustering the variances vector obtained from the extracted spectra. This allows to directly taking into consideration the nature of classes contained in an imaged scene. The used clustering strategy results in sequentially clustered wavelengths. From each cluster, only one decisive spectral band is selected in an optimal sense described in the next section.

It should be noted here that the proposed method can be seen as an approach for identifying spectral regions, which allow the separability between different materials present in a given remote sensing image. For example, in Feilhauera et al. (2013), a data-driven approach, which considers intercorrelations between neighboring wavelengths in a correlogram of a field spectrometer data, is reported. Local minima of correlation values are chosen as boundaries of desired spectral regions. Other approaches, in the same field, exist and interested readers can refer to Feilhauera et al. (2013) or to the references cited therein.

The remainder of this paper is structured as follows. In section 2, the proposed approach is presented. Section 3 consists of test results with three real hyperspectral data sets. In that section, the obtained results are compared with those obtained by four methods from the literature. Finally, section 4 concludes this paper.

Proposed hyperspectral band selection approach

As stated in Section 1, the proposed hyperspectral band selection approach combines unmixing and sequential clustering techniques. The proposed method operates in different successive stages:

The unmixing stage

Since the selected bands must discriminate between endmembers that are ideally characterized by their spectra, the first stage consists of extracting these spectra by an unsupervised Linear Spectral Unmixing (LSU) technique (Bioucas-Dias et al., 2012). Indeed, and principally due to the coarse spatial resolution of hyperspectral sensors, mixed pixels may occur in the acquired images. These ones prevent direct recognition of endmembers without advanced processing to unmix these spectra. LSU is one of the most used techniques for processing considered data (Bioucas-Dias et al., 2012). It aims at linearly unmixing these data into a set of endmember spectra, and a collection of associated abundance fractions. It is worth remembering here that in the mathematical data model, used in the LSU techniques, each observed nonnegative radiance/reflectance pixel spectrum is supposed to be a linear mixture of the nonnegative radiance/reflectance endmember spectra contained in that pixel. As a result, the entire observed nonnegative radiance/reflectance hyperspectral image $X \in \mathbb{R}^{N \times P}$ can be modeled, in matrix form, as (Bioucas-Dias et al., 2012)

$$X = AS$$

$$= \begin{bmatrix}
a_{11} & \cdots & a_{1L} \\
\vdots & \ddots & \vdots \\
a_{N1} & \cdots & a_{NL}
\end{bmatrix}
\begin{bmatrix}
s_{11} & \cdots & s_{1P} \\
\vdots & \ddots & \vdots \\
s_{L1} & \cdots & s_{LP}
\end{bmatrix}$$

(1)

where each row of the matrix $X$ represents one spectral band of the considered hyperspectral image. The $P$ pixels of this image are restructured as a one-dimensional array. $N$ represents the number of spectral bands of that image. Each column of the matrix $A \in \mathbb{R}^{N \times L}$ represents one endmember spectrum and...
The variance measures stage

At each wavelength \( \lambda_i \) (\( i = 1\ldots N \)), the variance \( \text{var}_{\lambda_i} \), of the extracted spectra samples \( (a_{i1} \ldots a_{iL}) \), is calculated as follows

\[
\text{var}_{\lambda_i} = \frac{1}{L} \sum_{j=1}^{L} (a_{ij} - m_i)^2 \text{ with } i = 1 \ldots N
\]  

(2)

where \( m_i \) is the mean of the above spectra sample values.

Then, the variances vector \( v \) is obtained and defined as

\[
v = (\text{var}_{\lambda_1}, \text{var}_{\lambda_2}, \ldots, \text{var}_{\lambda_N})
\]  

(3)

The potential spectral bands may be those that correspond to wavelengths where obtained variances reach maximum values. Indeed, in this case the differences between the samples of the endmember spectra are significant, and thus a better discrimination is possible between endmembers for a classification purpose. However, this strategy is not sufficient, because it can result in the selection of very close spectral bands, that is a highly correlated subset of spectral bands, which certainly leads to misclassification results. As a consequence, the first step of the proposed band selection approach may be coupled with a clustering technique applied on the computed variances vector \( v \). This way is also not promising for choosing significant spectral bands, for the reason that standard clustering techniques can lead to disjoint segments, which belong to the same cluster. Thus, the fact of choosing a spectral band that belongs to a cluster where the variance is maximum, can lead to the neglect of significant spectral bands in certain wavelengths of the considered electromagnetic spectrum. In this case, it is possible to consider each segment as a separate cluster, and select a spectral band for each one. However, in this situation the number of selected spectral bands may be higher than the number previously fixed.

**The sequential clustering stage**

The computed variances vector \( v \) is segmented to a predefined number \( K \) of clusters with a sequential clustering algorithm (Trahanias & Skordalakis, 1989). This way allows the selection of significant spectral bands, i.e. where the differences between the samples of the extracted endmember spectra are significant, corresponding to wavelengths spread across the considered electromagnetic spectrum. In the proposed spectral band selection approach, the Warped K-Means (WKM) (Leiva & Vidal, 2013) sequential clustering algorithm is used in order to segment the computed variances vector \( v \). The WKM is a \( K \)-means-based clustering algorithm that imposes a hard sequential constraint, which is implicitly embedded, via the considered wavelengths, in the computed variances vector \( v \). Therefore, this algorithm leads to the segmented variances vector \( v_{\text{seg}} \) defined as

\[
v_{\text{seg}} = (\text{seg}_1, \text{seg}_2, \ldots, \text{seg}_K)
\]  

(4)

where

\[
\text{seg}_1 = [\text{var}_{\lambda_{k_1}}, \ldots, \text{var}_{\lambda_{k_{t_1}}}], \quad \text{seg}_2 = [\text{var}_{\lambda_{k_{t_1}}}, \ldots, \text{var}_{\lambda_{k_{t_2}}}], \ldots, \text{seg}_K = [\text{var}_{\lambda_{k_{t_{K-1}}}}, \ldots, \text{var}_{\lambda_{k_{t_K}}}] \text{ with } k_1, k_2, \ldots \text{ are integers that are constrained by: } 1 < k_1 < k_2 < \ldots < K.
\]

It should be noted here that this clustering strategy leads to the same sequential wavelength clusters structure.

**The selection stage**

For each obtained segment \( \text{seg}_t \) (\( t = 1\ldots K \)), only one spectral band is selected. This one, denoted \( s_{\lambda_t} \), corresponds to the wavelength where the variance \( \text{var}_{\lambda_t} \), which belongs to the considered segment \( \text{seg}_t \), reaches a maximum value.

The whole proposed hyperspectral band selection approach is described by the following pseudo-code (Table 1).

**Experimental results**

**Data sets**

Here, the performance of the proposed band selection approach is evaluated on three real radiance hyperspectral data sets.

The first hyperspectral image is acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor on the Kennedy Space Center (KSC), Florida, USA. This 512 × 614 pixel image is characterized by 176 spectral bands and a spatial resolution of 18 m. This image is available with labeled samples ground truth including 13 materials that represent the various land cover types that occur in this scene (scrub, willow swamp, cabbage palm/oak hammock, slash pine, oak/broadleaf hammock,
The proposed method pseudo-code.

| Input | Initial hyperspectral image with \( N \) spectral bands. |
|-------|---------------------------------------------------------|
| 1.   | **Unmixing stage**                                      |
| 1.1. | Estimate the number \( L \) of endmembers by using the HySime method. |
| 1.2. | Extract the \( L \) endmember spectra by using the VCA method. |
| 2.   | **Variance measures stage**                             |
| 2.1. | Calculate the variances vector \( v \) from the extracted \( L \) endmember spectra by using (3). |
| 3.   | **Sequential clustering stage**                         |
| 3.1. | Sequentially segment the variances vector \( v \) to predefined number \( K \) of clusters by using the WKM method. |
| 4.   | **Selection stage**                                     |
| 4.1. | Select only one spectral band, where the variance is maximum, for each obtained cluster. |

**Output:** Reduced image with \( K \) spectral bands (\( K < N \)).

hardwood swamp, graminoid/spartina/cattail/salt marsh, mud flats and water). The reader is referred to (Ham, Chen, Crawford, & Ghosh, 2005) for more details on this hyperspectral image.

The second image is also gathered by the AVIRIS sensor over the land of Indian Pines, in North-Western, Indiana, USA. This 145 \( \times \) 145 pixel image, with a spatial resolution of 20 m, contains 185 spectral bands. This image is also available with labeled samples including 16 materials, which principally represent different cereal classes (corn, corn-min/notill, soybeans-min/notill/clean, oats, wheat) and vegetation. The reader is referred to (Jimenez & Landgrebe, 1999) for more information on this second used image.

The last hyperspectral image is acquired by the spaceborne Hyperion sensor on the Okavango Delta, Botswana. In total 14 ground truth materials with labeled samples are available for this 256 \( \times \) 1476 pixel image that is characterized by 145 spectral bands and a spatial resolution of 30 m. The 14 materials represent, in the considered area, the land-cover types in seasonal swamps, occasional swamps, and drier woodlands. These materials reflect the impact of flooding on vegetation. The reader is referred to (Ham et al., 2005) for more details on this last used hyperspectral image.

It is clear that the used images contain some specific classes that can be confused in the classification task. Thus, the use of these data sets allows a good evaluation of the proposed technique.

**Design of experiments**

In this paper, the proposed unsupervised approach is evaluated by a supervised hyperspectral pixel classification in comparison with four literature methods, including the supervised Rough set and fuzzy C-means (Rough-FCM) (Shi et al., 2003), the supervised approach based on Rough Set (Patra, Modi, & Bruzzone, 2015), the unsupervised wards linkage strategy using divergence (WaLuDi) (Martinez-Uso et al., 2007), and the unsupervised wards linkage strategy using mutual information (WaLuMI) (Martinez-Uso et al., 2007). It should be noted here that these two well-known unsupervised band selection methods are the most techniques used in the literature. Also, the classification, using all the spectral bands, is considered in the conducted experiments.

The desired number of spectral bands to be selected is not known in advance and certainly varied from an image to another. In the conducted experiments, different numbers of selected spectral bands are considered ranging from 5 to 40 with a step size of 5.

As stated above, and after running the considered band selection algorithms, the supervised multiclass Support Vector Machine (SVM) classifier (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998) is used to evaluate the effectiveness of selected spectral bands. This classifier is implemented with the Radial Basis Function (RBF) kernel. The spread of the RBF kernel and the regularization parameter of the SVM classifier are derived, for each subset of selected spectral bands and for each data set, by using a grid search strategy. For each data set, the half of the available labeled ground truth samples are randomly selected for the learning of the used classifier. The accuracies are evaluated on the remaining samples by using Overall Accuracy (OA) and \( Kappa \) measures. Ten runs are performed to reduce the random effect of the results and the means of them are reported.

**Results and discussion**

Since the proposed approach uses the VCA unmixing technique in its first stage, the extracted endmember spectra from the considered data set are given in the following figures (Figures 1–3).

The above figures clearly demonstrate that the used data sets contain some specific classes, with similar radiance values in several spectral bands, which can be confused in the classification task.

In the next figures, the obtained variances vectors, computed from the above spectra, are given (Figures 4–6).

These last figures give an initial idea on the spectral bands where the discrimination between the classes, contained in observed scenes, is possible.

Also, and as examples, the results of the WKM sequential clustering algorithm applied on the above variances vectors with 10 segments (in order to select 10 decisive spectral bands) are given in the next figures. These figures show the obtained clusters with different colors, and the red stars correspond to the maximum variance of each segment. The indices of these stars are those of the spectral bands to be selected (Figures 7–9).
The following tables report the mean overall classification accuracy measures, as well as the mean Kappa accuracies obtained for both the data sets.

Table 2 shows the obtained results for the KSC data set. This table reports that when selecting only 5 spectral bands, the proposed approach is capable to attain 86.56% mean overall accuracy, while the WaLuDi, the WaLuMI, the Rough-FCM and the Rough Set methods, with the same number of selected spectral bands, achieve 82.63%, 80.67%, 81.18% and 85.03% mean overall accuracies respectively. This table also shows that the proposed approach achieves better mean overall accuracy when selecting only 20 spectral bands than when using all the available spectral bands. Also, Table 2 reports that the results are improved by increasing the number of selected spectral bands. It can also be seen that the proposed
approach is able to achieve better results, for the different number of selected spectral bands, than the used literature methods, which confirms the efficiency of the proposed band selection approach.

In an analog manner, Table 3 highlights, for the Indian Pines data set, the same points as the previous, but with significantly better performances for the proposed approach compared to the used literature methods. In particular, when selecting only 5 spectral bands, the proposed band selection approach is able to achieve 78.08% mean overall accuracy, while the used literature methods attain mean overall accuracies between 60.07% and 67.19% with the same number of selected spectral bands. For the same data set, and compared to the use of all spectral bands, the proposed approach is able to achieve better mean overall accuracy by using only 25 spectral bands.
Concerning the obtained results by using the Botswana data set, Table 4 shows that when selecting small numbers of spectral bands, the proposed approach produces higher classification accuracy as compared to the literature techniques. For higher numbers of selected bands, the proposed approach is slightly less effective than the other used methods (in particular the supervised Rough set approach). These results (i.e. for higher numbers of selected bands) can be explained by the type of materials, and consequently their spectra, present in the considered data set. Also, and as stated in the introduction section, these results can be explained by the fact that the Rough set method is supervised while the proposed one is unsupervised.
Finally, these tables show that the mean Kappa accuracies, for both data sets, follow the same trends as the mean overall classification accuracies.

It should be noted here that the proposed approach gives, globally, very satisfactory results although the presence of the described specific classes in the used data sets. Also, the proposed approach is very easy to implement (since it combines well-known techniques) and it is very fast: for an example, its execution time, for all data sets when considering ten selected spectral bands, is less than 3 s by using an Intel Core i5 CPU processor running at 2.50 GHz, with a memory capacity of 4 GB.

### Conclusion

In this paper, a new method was proposed for unsupervised hyperspectral band selection. The designed approach combines linear spectral unmixing and sequential clustering techniques.

The proposed method starts with a linear spectral unmixing technique that extracts material spectra contained in the considered remote sensing hyperspectral data. After that, the variance of extracted spectra samples is calculated at each wavelength resulting in a variances vector. Then, the obtained vector is segmented with a sequential clustering technique. Finally, only one spectral band is selected for each segment. This band corresponds to the wavelength where a maximum of variance is obtained.

The designed approach along with four other literature methods were applied to three real hyperspectral data sets for comparison purpose. The evaluation, of the obtained results, was performed by using a supervised hyperspectral pixel classification. Experimental results show that the proposed method yields very satisfactory results. Indeed, for the first two considered data sets, this original approach outperforms the tested methods from the literature for all selection scenarios. For the last considered data set, the proposed method also outperforms the tested

### Table 2. The mean overall classification accuracies (%) and the mean kappa accuracies (KSC data set).

| No of selected bands | OA | Kappa | OA | Kappa | OA | Kappa | OA | Kappa | OA | Kappa |
|----------------------|----|-------|----|-------|----|-------|----|-------|----|-------|
| 5                    | 82.63 | 806 | 80.67 | 783 | 81.18 | 790 | 85.03 | 833 | 86.56 | 850 |
| 10                   | 85.71 | 841 | 88.83 | 876 | 87.52 | 861 | 90.51 | 894 | 90.70 | 896 |
| 15                   | 89.04 | 878 | 92.25 | 914 | 92.22 | 913 | 93.01 | 922 | 93.12 | 923 |
| 20                   | 89.65 | 885 | 93.15 | 924 | 92.32 | 914 | 93.73 | 930 | 94.35 | 937 |
| 25                   | 90.64 | 896 | 93.66 | 929 | 93.63 | 929 | 94.39 | 937 | 94.70 | 941 |
| 30                   | 91.62 | 907 | 94.23 | 936 | 94.41 | 938 | 94.66 | 941 | 94.94 | 944 |
| 35                   | 92.17 | 913 | 94.52 | 939 | 94.97 | 944 | 94.93 | 944 | 95.09 | 945 |
| 40                   | 93.70 | 930 | 95.11 | 946 | 95.11 | 946 | 95.23 | 947 | 95.42 | 949 |
| All bands            | OA   | Kappa | 93.47 | .927 |

### Table 3. The mean overall classification accuracies (%) and the mean kappa accuracies (Indian Pines data set).

| No of selected bands | OA | Kappa | OA | Kappa | OA | Kappa | OA | Kappa | OA | Kappa |
|----------------------|----|-------|----|-------|----|-------|----|-------|----|-------|
| 5                    | 60.07 | .529 | 60.09 | .531 | 63.92 | .579 | 67.19 | .618 | 78.08 | .748 |
| 10                   | 74.35 | .704 | 77.16 | .737 | 76.01 | .724 | 77.53 | .742 | 85.92 | .839 |
| 15                   | 83.47 | .811 | 79.67 | .766 | 82.77 | .803 | 83.02 | .806 | 86.58 | .846 |
| 20                   | 85.31 | .832 | 84.99 | .828 | 83.94 | .816 | 84.85 | .826 | 86.77 | .849 |
| 25                   | 87.42 | .856 | 86.10 | .841 | 85.65 | .836 | 86.73 | .848 | 88.60 | .869 |
| 30                   | 87.74 | .860 | 87.42 | .856 | 86.41 | .845 | 87.41 | .856 | 89.17 | .876 |
| 35                   | 89.47 | .880 | 89.12 | .876 | 88.15 | .865 | 89.14 | .876 | 89.66 | .882 |
| 40                   | 89.83 | .884 | 89.49 | .880 | 88.71 | .871 | 89.66 | .882 | 90.02 | .886 |
| All bands            | OA   | Kappa | 88.54 | .869 |

### Table 4. The mean overall classification accuracies (%) and the mean kappa accuracies (Botswana data set).

| No of selected bands | OA | Kappa | OA | Kappa | OA | Kappa | OA | Kappa | OA | Kappa |
|----------------------|----|-------|----|-------|----|-------|----|-------|----|-------|
| 5                    | 80.01 | .783 | 86.30 | .852 | 81.32 | .798 | 89.63 | .888 | 90.00 | .892 |
| 10                   | 91.20 | .905 | 91.35 | .906 | 92.19 | .915 | 93.44 | .929 | 93.87 | .933 |
| 15                   | 93.66 | .931 | 92.72 | .921 | 93.22 | .927 | 94.70 | .943 | 93.83 | .933 |
| 20                   | 95.19 | .948 | 94.05 | .936 | 95.24 | .948 | 95.30 | .949 | 94.37 | .939 |
| 25                   | 95.96 | .956 | 94.94 | .945 | 95.68 | .953 | 96.22 | .959 | 94.46 | .941 |
| 30                   | 96.80 | .965 | 95.55 | .952 | 96.18 | .959 | 96.87 | .966 | 94.21 | .939 |
| 35                   | 97.20 | .970 | 96.54 | .962 | 96.88 | .966 | 97.57 | .974 | 94.89 | .943 |
| 40                   | 97.46 | .973 | 96.93 | .967 | 97.13 | .969 | 97.80 | .976 | 95.17 | .948 |
| All bands            | OA   | Kappa | 95.68 | .953 |

EUROPEAN JOURNAL OF REMOTE SENSING  
37
literature approaches when selecting small numbers of spectral bands. For higher numbers of selected bands, the proposed approach is slightly less effective than one of the tested methods.

The proposed method is easy to implement, and can undoubtedly contribute to the dimensionality reduction of remote sensing hyperspectral data. It also can contribute to the design of future sensors by an initial choice of decisive spectral bands allowing a maximum discrimination between different materials.

Future extensions of this work will particularly consist in automatically determining the optimal number of clusters by using clustering validity indices.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**References**

Amri, M.B., Benabadji, S.I., & Karoui, M.S. (2017). Unsupervised hyperspectral band selection by sequential clustering. International Conference on Watermarking and Image Processing (ICWIP), Paris, France.

Backer, S.D., Kempeners, P., Debruyn, W., & Scheunders, P. (2005). A band selection technique for spectral classification. *IEEE Geoscience and Remote Sensing Letters*, 2(3), 319–323. doi:10.1109/LGRS.2005.848511

Bioucas-Dias, J.M., & Nascimento, J.M.P. (2008). Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), 2435–2445. doi:10.1109/TGRS.2008.918089

Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 354–379. doi:10.1109/JSTARS.2012.2194696

Bruzzone, L., & Persello, C. (2009). A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability. *IEEE Transactions on Geoscience and Remote Sensing*, 47(9), 3180–3191. doi:10.1109/TGRS.2009.2019636

Cariou, C., Chehdi, K., & Moan, S.L. (2011). BandClust: An unsupervised band reduction method for hyperspectral remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 8(3), 565–569. doi:10.1109/LGRS.2010.2091673

Chang, C.-I. (2007). Hyperspectral data exploitation: Theory and applications. Hoboken, NJ: Wiley.

Chang, C.-I., Du, Q., Sun, T.-L., & Alhouse, M.L.G. (1999). A joint band prioritization and band decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6), 2631–2641. doi:10.1109/36.803411

Das, A., Ghosh, S., & Ghosh, A. (2014). Band elimination of hyperspectral imagery using partitioned band image correlation and capacitory discrimination. *International Journal of Remote Sensing*, 35(2), 554–577. doi:10.1080/01431161.2013.871392

Du, Q., & Yang, H. (2008). Similarity-based unsupervised band selection for hyperspectral image analysis. *IEEE Geoscience and Remote Sensing Letters*, 5(4), 564–568. doi:10.1109/LGRS.2008.200619

Feilhauer, H., Thonfeldb, F., Faude, U., He, K.S., Rocchini, D., & Schmidtlein, S. (2013). Assessing floristic composition with multispectral sensors—A comparison based on monotemporal and multiseasonal field spectra. *International Journal of Applied Earth Observation and Geoinformation*, 21, 218–229. doi:10.1016/j.jag.2012.09.002

Feng, S., Itoh, Y., Parente, M., & Duarte, M.F. (2017). Hyperspectral band selection from statistical wavelet models. *IEEE Transactions on Geoscience and Remote Sensing*, 55(4), 2111–2123. doi:10.1109/TGRS.2016.2636850

Guo, B., Damper, R.L., Gunn, S.R., & Nelson, J.D.B. (2008). A fast separability-based feature-selection method for high-dimensional remotely sensed image classification. *Pattern Recognition*, 41(5), 1653–1662. doi:10.1016/j.patcog.2007.11.007

Guo, B., Gunn, S.R., Damper, R.L., & Nelson, J.D.B. (2006). Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 3(4), 522–526. doi:10.1109/LGRS.2006.878240

Ham, J., Chen, Y., Crawford, M.M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 492–501. doi:10.1109/TGRS.2004.842481

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, 13(4), 18–28. doi:10.1109/5254.708428

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1), 55–63. doi:10.1109/TIT.1968.1054102

Imani, M., & Ghassemian, H. (2014). Band clustering-based feature extraction for classification of hyperspectral images using limited training samples. *IEEE Geoscience and Remote Sensing Letters*, 11(8), 1325–1329. doi:10.1109/LGRS.2013.2292892

Jimenez, L.O., & Landgrebe, D.A. (1998). Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 28(1), 39–54. doi:10.1109/5326.661089

Jimenez, L.O., & Landgrebe, D.A. (1999). Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6), 2653–2667. doi:10.1109/36.803413

Kim, B., & Landgrebe, D.A. (1991). Hierarchical classifier design in high-dimensional numerous class cases. *IEEE Transactions on Geoscience and Remote Sensing*, 29(4), 518–528. doi:10.1109/36.135813

Landgrebe, D.A. (2002). Hyperspectral image analysis. *IEEE Signal Processing Magazine*, 19(1), 17–28. doi:10.1109/79.974718

Leiva, L.A., & Vidal, E. (2013). Warped K-means: An algorithm to cluster sequentially-distributed data. *Information Sciences*, 237, 196–210. doi:10.1016/j.ins.2013.02.042

Marpu, R.R., Pedergnana, M., Dalla Mura, M., Peeters, S., Benediktsson, J.A., & Bruzzone, L. (2012). Classification of hyperspectral data using extended attribute profiles
based on supervised feature extraction. *International Journal of Image and Data Fusion*, 3(3), 269–298. doi:10.1080/19479832.2012.702687

Martínez-Uso, A., Pla, F., Sotoca, J.M., & García-Sevilla, P. (2007). Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12), 4158–4171. doi:10.1109/TGRS.2007.904951

Mojaradi, B., Moghaddam, H.A., Zoej, M.J.V., & Duin, R.P. W. (2009). Dimensionality reduction of hyperspectral data via spectral feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 2091–2105. doi:10.1109/TGRS.2008.2010346

Nascimento, J.M.P., & Bioucas-Dias, J.M. (2005). Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), 898–910. doi:10.1109/TGRS.2005.844293

Patra, S., Modi, P., & Bruzone, L. (2015). Hyperspectral band selection based on rough set. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10), 5495–5503. doi:10.1109/TGRS.2015.2424236

Rashwan, S., & Dobigeon, N. (2017). A split-and-merge approach for hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1378–1382. doi:10.1109/LGRS.2017.2713462

Shi, H., Shen, Y., & Liu, Z. (2003). Hyperspectral bands reduction based on rough sets and fuzzy c-means clustering. *Instrumentation and Measurement Technology Conference (IMTC 2003)*, 2, 1053–1056.

Trahanias, P., & Skordalakis, E. (1989). An efficient sequential clustering method. *Pattern Recognition*, 22(4), 449–453. doi:10.1016/0031-3203(89)90053-8

Xu, X., Shi, Z., & Pan, B. (2017). A new unsupervised hyperspectral band selection method based on multiobjective optimization. *IEEE Geoscience and Remote Sensing Letters*, 14(11), 2112–2116. doi:10.1109/LGRS.2017.2753237

Yang, C., Liu, S., Bruzzone, L., Guan, R., & Du, P. (2013). A feature-metric-based affinity propagation technique for feature selection in hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 10(5), 1152–1156. doi:10.1109/LGRS.2012.2233711

Yang, H., Du, Q., Su, H., & Sheng, Y. (2011). An efficient method for supervised hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, 8(1), 138–142. doi:10.1109/LGRS.2010.2053516

Yang, J.M., Yu, P.T., & Kuo, B.C. (2010). A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1279–1293. doi:10.1109/TGRS.2009.2031812

Yin, J., Wang, Y., & Zhao, Z. (2010). Optimal band selection for hyperspectral image classification based on inter-class separability. Proceeding Symposium on Photonics and Optoelectronics, Chengdu, 1–4. doi:10.1109/SOPO.2010.5504325

Zhan, Y., Hu, D., Xing, H., & Yu, X. (2017). Hyperspectral band selection based on deep convolutional neural network and distance density. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2365–2369. doi:10.1109/LGRS.2017.2765339

Zhang, M., Ma, J., & Gong, M. (2017). Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 773–777. doi:10.1109/LGRS.2017.2681118

Zhu, G., Huang, Y., Li, S., Tang, J., & Liang, D. (2017). Hyperspectral band selection via rank minimization. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2320–2324. doi:10.1109/LGRS.2017.2763183