

# Systematic approach towards extracting endmember spectra from hyperspectral image using PPI and SMACC and its evaluation using spectral library

Arpit Aggarwal · R. D. Garg

Received: 7 February 2013 / Accepted: 9 December 2014 / Published online: 24 December 2014  
© Società Italiana di Fotogrammetria e Topografia (SIFET) 2014

**Abstract** This study focuses on systematic approach towards extraction of endmember spectra from hyperspectral image. The study demonstrates the effect of systematic preprocessing like atmospheric correction, radiometric correction (bad band and columns correction), and geometric correction on hyperspectral image. The study also focuses on the selection of the method for extracting endmember spectra depending upon the land-cover classes present in the study area. Two algorithms for extracting endmember spectra, i.e., pixel purity index (PPI) and sequential maximum angle convex cone (SMACC), have been used. To validate and perform comparative analysis of the two algorithms, spectral library has been created using field spectroradiometer and used as a reference to evaluate their performance. Visually, good results have been observed between extracted endmember spectra and reference library spectra after applying the rigorous preprocessing. To further analyze the two endmember extraction algorithms, spectral angle mapper (SAM) scores have been computed for various endmember spectral classes with respect to reference spectral library. The tabulated SAM scores for the endmembers of PPI and SMACC show that SMACC is more effective in extracting endmember spectra of vegetation classes while PPI is a more effective algorithm for roads and dry soil. It has been observed that systematic approach towards extracting endmember spectra from a hyperspectral image

should consist of proper preprocessing steps, ground validation with a reference spectral library, and most importantly the proper selection of algorithm as the performance of algorithms for extracting endmember spectra depends on the land-cover classes present in the study area.

**Keywords** Endmembers · Atmospheric correction · PPI · SMACC · Spectral library · SAM

## Introduction

### General

Hyperspectral remote sensing is an important tool for the observations of earth after development of imaging spectrometry. The spectral details provided by hyperspectral remote sensing are its main advantage. The spectral properties of land covers and surface material types differ in their spectral response. For spectral separation and mapping, different important spectral features are needed. The application of hyperspectral remote sensing is limited by the large diversity of materials and lack of knowledge of their spectral properties.

Space-borne hyperspectral data like Hyperion which covers spectral range from 400 to 2500 nm and has 242 continuous channels with 10-nm bandwidth is capable of discriminating most of the earth surface materials. Narrow bandwidth and numerous channels give much information about the site surface under investigation. Spectral images even having a large number of spectral channels never depict the true radiance of the surface due to the sensors and atmosphere.

---

A. Aggarwal (✉)  
Risk Management Solutions India, 617, Pocket-D, Dilshad Garden,  
New Delhi 110095, India  
e-mail: Arpit.coer@gmail.com

R. D. Garg  
Department of Civil Engineering, Geomatics Engineering Group,  
Indian Institute of Technology Roorkee, Roorkee,  
Uttarakhand 247667, India

## Endmember spectrum

An important problem in hyperspectral imaging processing is to decompose the mixed pixels into the materials that contribute to the pixel, endmember, and a set of corresponding fractions of the spectral signature in the pixel, abundances, and this problem is known as the unmixing problem. According to the definition, an endmember is an idealized pure signature of a class. Endmember extraction is one of the fundamental and crucial tasks in hyperspectral data exploitation (Mozaffar et al. 2008).

It has received considerable interest in recent years, with many researchers devoting their effort to develop algorithms for endmember extraction from hyperspectral data. An ultimate goal of an endmember extraction algorithm (EEA) is to find the purest form of each spectrally distinct material on a scene.

## Endmember extraction algorithms

### Overview

Most of the popular endmember extraction algorithms nowadays are based on geometric analysis of image data. Keshava and Mustard (2002) argued that the basic assumption for the geometric endmember extraction is the endmember spectra are pure spectra in the image which lie at the extreme ends of the volume occupied by the dataset. Winter (1999) proposed the N-FINDR algorithm for endmember extraction. The algorithm determines a simplex of largest volume, within the dataset, containing the maximum number of pixels. In order to refine the endmember estimate, the volume of the simplex is calculated by replacing each endmember by each pixel in the image. The procedure continues until there is no further replacement of endmembers.

Another algorithm for endmember extraction based on the geometric analysis of the data is Vertex component analysis (VCA) (Nascimento and Bioucas-Dias 2005). The algorithm starts by determining the subspace spanned by the endmembers using HySime and then projects the spectral vectors in a direction orthogonal to the determined subspace. The extreme ends of the projection correspond to the endmember spectra.

### Pixel purity index (PPI)

One of the most successful approaches has been the Pixel Purity Index or PPI, which is based on the geometry of convex sets (Green et al. 1988). PPI considers spectral pixels as vectors in an  $N$ -dimensional feature space. The algorithm

proceeds by generating a large number of random  $N$ -dimensional vectors, also called skewers, through the dataset. Every data point is projected onto each skewer, along which position it is pointed out. The data points which correspond to extreme values in the direction of a skewer are identified and placed on a list. As more skewers are generated, the list grows and the number of times a given pixel is placed on this list is also tallied (Boardman 1994).

The pixels with the highest tallies are considered the purest ones, since a pixel count provides a PPI. It is important to emphasize that the PPI algorithm does not identify a final list of endmembers. PPI was conceived not as a solution but as a guide; in this, comparing the pure pixels with target spectra from a library and successively projecting the data to lower dimensional spaces while endmembers are identified is proposed. There are several interactive software tools available to perform this task.

The original implementation of PPI proposes the use of unitary vectors as skewers in random directions of the  $N$ -dimensional space. This implementation may be improved by a careful selection of existing vectors to skew the dataset. Intelligent selection of skewers may result in a more efficient behavior of the algorithm. Some tools based on variations of PPI concepts have been proposed (Theiler et al. 2000).

### Sequential maximum angle convex cone (SMACC)

SMACC algorithm is based on a convex cone model for representing vector data. The endmembers are selected directly from the dataset. The algorithm for finding the endmembers is sequential: the convex cone model starts with a single endmember and increases incrementally in dimension. Abundance maps are simultaneously generated and updated at each step. A new endmember is identified based on the angle that it makes with the existing cone. The data vector which is making the maximum angle with the existing cone is chosen as the next endmember to add to enlarge the endmember set. The algorithm updates the abundances of previous endmembers and ensures that the abundances of previous and current endmembers remain positive or zero. The algorithm terminates when all of the data vectors are within the convex cone, to some tolerance. The method offers advantages for hyperspectral datasets where high correlation among channels and pixels can impair unmixing by standard techniques.

Mathematically, SMACC uses the following convex cone expansion for each pixel spectrum (endmember), defined as follows:

$$H(c, i) = \sum_k^N R(c, k)A(k, j)$$

Where

$i$  is the pixel index.

$j$  and  $k$  are the endmember indices from 1 to the expansion length,  $N$ .

$R$  is a matrix that contains the endmember spectra as columns.

$c$  is the spectral channel index.

$A$  is a matrix that contains the fractional contribution (abundance) of each endmember  $j$  in each endmember  $k$  for each pixel (Gruninger et al. 2004).

those who regularly use satellite and airborne remote-sensing data. ENVI provides comprehensive data visualization and analysis for images of any size and any type—all from within an innovative and user-friendly environment.

ENVI is written in interactive data language (IDL), a powerful structured programming language that offers integrated image processing. IDL is required to run ENVI and the flexibility of ENVI is largely due to IDL's capabilities. ENVI contains extensive support for carrying out processing of hyperspectral imagery.

## Study area and data used

The study area is Dehradun, the capital of Uttarakhand. The city lies at 30° 19' N and 78° 04' E and is located at an altitude of 640 m above mean sea level (MSL). Location of study area is shown in Fig. 1.

The Hyperion image of Dehradun was acquired on 25 December 2006. There are 242 bands with the spectral range of 355 to 2577 nm of 10 nm bandwidth. The data used in this study has 242 bands. Total 149 bands are calibrated out of 242 bands, the bands which are not calibrated are set to zero.

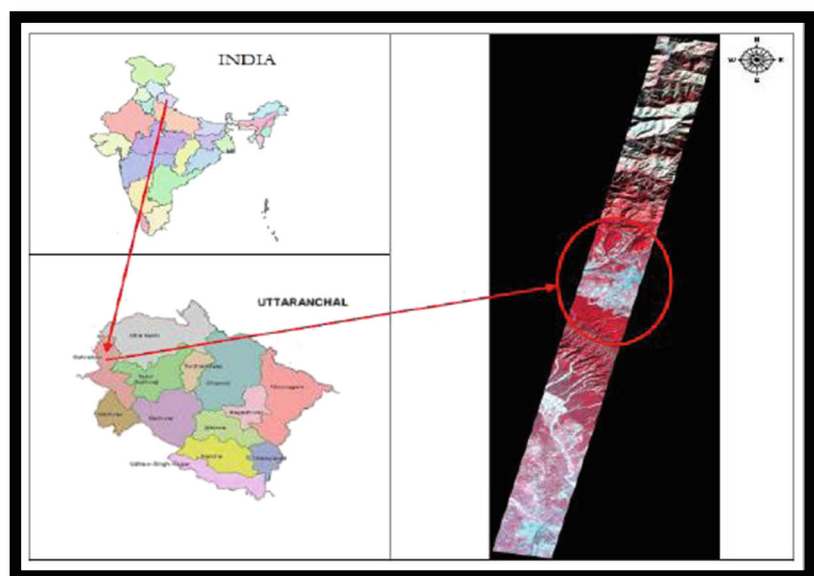
## Software used

The Environment for Visualizing Images (ENVI) 4.7 is a revolutionary image processing software. ENVI was designed to address the numerous and specific needs of

## Instrument used

An analytical spectral device (ASD), Field Spec®-Pro spectroradiometer, was used for measurements of surface reflectance to take samples. The ASD spectroradiometer is a portable array-based spectrometer consisting of a spectrometer unit, computer interface, and fiber optic probe. The instrument has two integrated radiometers covering 350 to 2500 nm. The spectroradiometer consists of one silicon photodiode array and two fast scanning thermoelectrically (TE) cooled spectrometers with a spectral resolution of 10 nm. The instrument was operated with 5° full-field-of-view (FFOV) fore optics. A laptop interface with the instrument allows real-time viewing of the spectrum recorded. The ASD instrument records the spectra in 2151 continuous bands. The spectral resolution of Hyperion data and the spectral range of ASD instrument are the same for the present study.

**Fig. 1** Study area and data used



## Methodology

The methodology adapted in the present study is for achieving the objectives which includes the following:

- Extraction of endmember spectra using PPI and SMACC algorithms on Hyperion image after fast line-of-sight atmospheric analysis of spectral hypercubes (FLAASH) atmospheric correction
- Construct a spectral library using ASD spectroradiometer
- Comparison of the ground-measured spectra collected from ASD spectroradiometer with that of atmospherically corrected image spectra
- Assess the performance of PPI and SMACC by evaluating SAM score between endmember spectra from the image and spectral library built

Figure 2 shows the complete methodology used for achieving the above-mentioned objectives.

### Removal of bad bands and columns

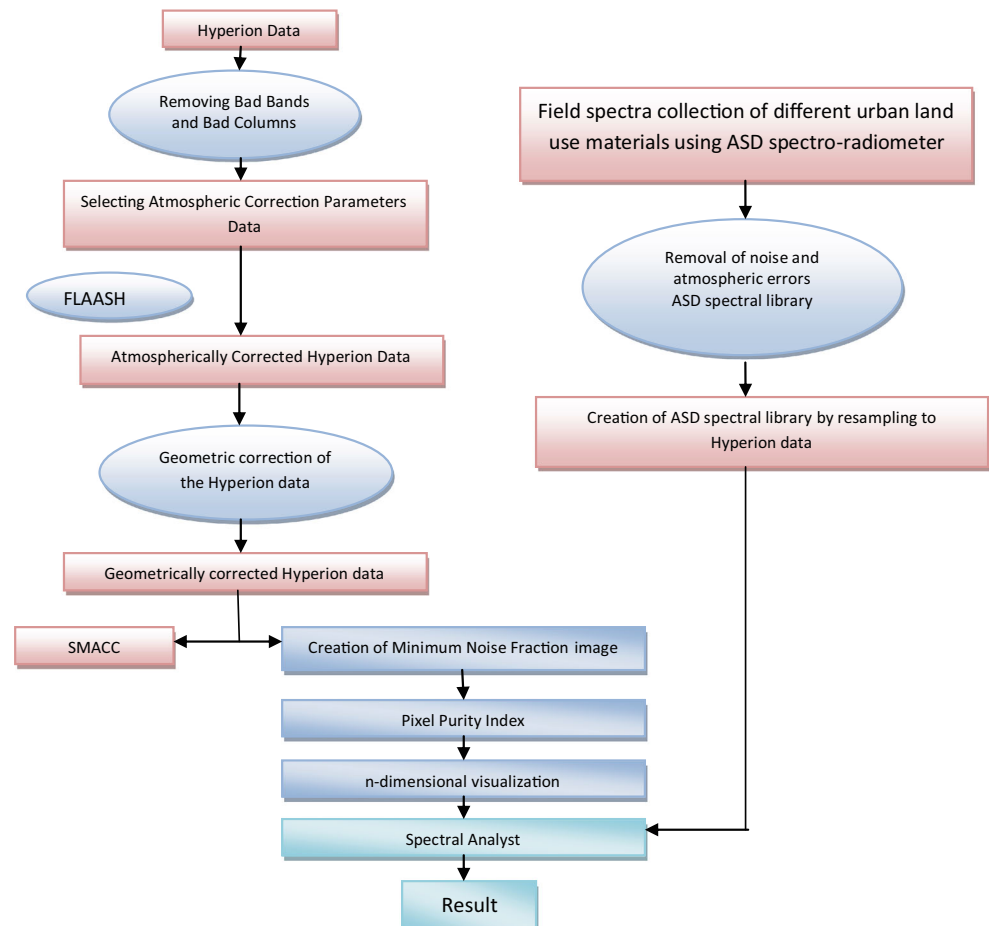
The Hyperion level 1 radiometric product used in the present study has 242 bands. However, only 149 bands

are calibrated to nonzero (band 8 to 57 for visible and near infrared (VNIR) and 77–224 for short-wave infrared (SWIR) region). The reason for not calibrating all 242 channels is low detector response. The bands that are not calibrated are set to zero. Before starting the actual processing, the bands with no data were identified and removed, so as to get spectral subset image with 149 bands. The image bands were finally resized into 149 bands by excluding bands, having no information, bands having negative values for the wavelengths, and bands falling in the water absorption range. Table 1 lists the bands removed from the dataset and the corresponding reason for eliminating it.

In a push-broom sensor, a poorly calibrated detector in either VNIR or SWIR arrays leaves high-frequency errors (“vertical stripes”) on the image bands. In Hyperion, striping pixels have been classified in four categories (Han and Goodenough 2002):

1. Continuous with a typical digital number (DN) value
2. Continuous with low DN values as compared to adjacent column
3. Intermittent with a typical DN value
4. Intermittent with lower DN values

**Fig. 2** Methodology



**Table 1** Bands removed

Bands	Details
1–7	No information
9–11	Negative values
68, 69	No information
70, 71, 72	Negative values
73, 74	Negative values
76,80,81	Negative values
83,85, 86, 88	No information
120–130	No information
165–181	No information
185, 186	Negative values
Above 224	No information

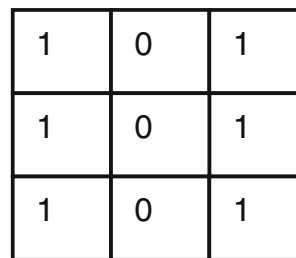
The first two categories of stripes are the most extreme type as they contain very little or no valid data about the ground feature. In the level 1R product, these stripes are left unmodified, allowing the users to handle or replace the pixels as per the requirement. In order to facilitate extraction of calibrated spectra from Hyperion dataset, it is significant to carefully balance for the striping in the dataset. In the present study, the bad columns were identified visually to avoid enforcing severe change in the spectra. A total of 36 bad columns were identified in 13 VNIR bands of the dataset. SWIR bands were found devoid of visible stripes.

Table 2 lists the bad columns identified in the respective bands. A 3×3 mean filter was applied on the bad columns, without taking the values of the bad columns. Figure 3 shows the mean filter for replacing the bad columns. Figure 4a, b illustrates the effect of mean filter in replacing bad columns of a band.

After removal of bad bands and bad columns, the 242 band images have been reduced to 149.

**Table 2** Bad columns identified

Bands	Bad column
8	6, 68, 114, 125, 132, 141
9	6, 68, 114, 119, 125, 246
10	6, 68, 114, 199
11	6, 68,114,199
12	6, 114
13	114
15	114
27	47
28	47, 114
31	114
54	13, 17, 20, 32, 39
55	13, 17, 20, 32, 33, 39
56	13, 17, 20, 32, 33, 39
57	8, 13, 17, 20, 33, 36, 38



**Fig. 3** Mean filter for bad columns

Atmospheric correction

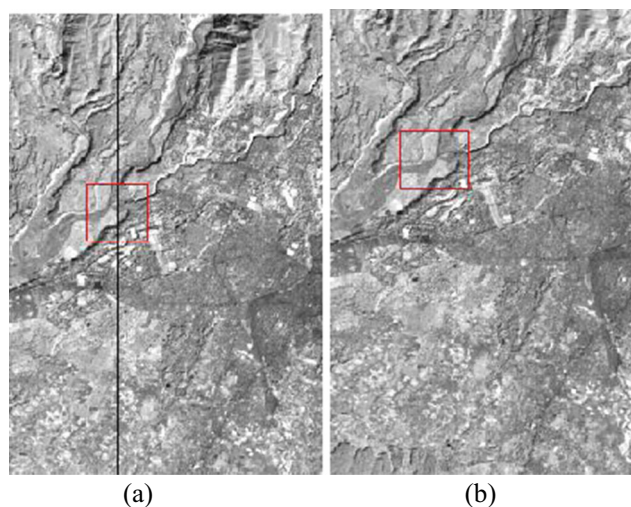
After the removal of the bad bands and destriping, the resized 149 bands were corrected for atmospheric errors using the FLAASH model of ENVI’s atmospheric correction module. The Hyperion image of Dehradun area, due to its time of acquisition, is badly affected by atmospheric error such as haze. Thus, atmospheric corrections of Hyperion images are required for the reduction of the atmospheric influence on the reflectance and to filter out the target reflectance cleanly from the mixed signal, using FLAASH wavelengths ranging through visible infrared and short-wave infrared can be corrected for atmospheric errors.

The different parameters which were applied for running the FLAASH model on Hyperion image are listed in Table 3.

Geometric correction

Geo-referenced Landsat ETM+ data (UTM/WGS 84 datum, zone 43 N) is used as the base-referenced data for the co-registration of the Hyperion data in the image to image registration.

For pixel to pixel matching of the two images, image to image registration was important. The root mean square error was tried to be kept as low as possible. First-order polynomial transformation was applied as the area was not very large to do



**Fig. 4** a Bad columns. b Column after mean filtering

**Table 3** Parameters for FLAASH

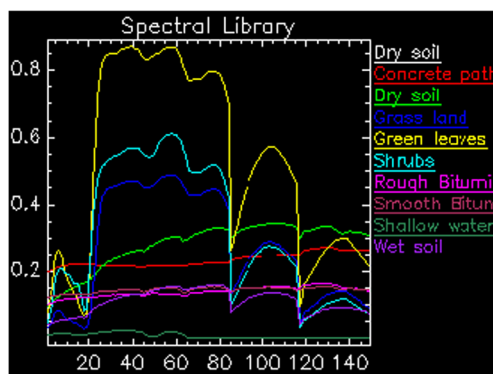
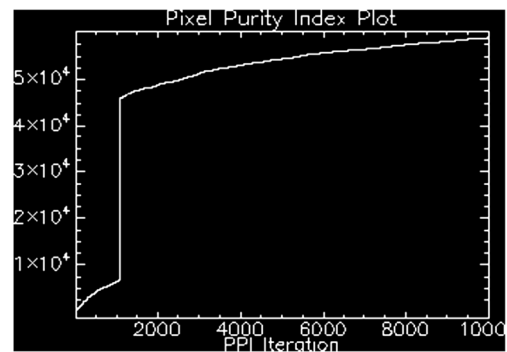
Parameters	
Sensor altitude	705 km
Spatial resolution	30 m
Radiometric resolution	16 bits
VNIR range	0.45–1.35 $\mu\text{m}$
IFOV (mrad)	0.043
Number of rows	256
Number of columns	3128
Swath	7.2 km
SWIR range	1.40–2.48 $\mu\text{m}$

the registration; however, four well-distributed GCPs were taken to do the registration as any misregistration could lead to erroneous results later in the research. In order to avoid spectral interpolation, nearest neighborhood resampling method was used. The correctness of the registration was also further verified by using different tools like swiping and flickering.

#### Building the spectral library

The spectra of different classes were collected using ASD spectroradiometer. The process of building the spectral library can be divided into three steps:

1. Instrument calibrations: A certain amount of electrical current is generated by thermal electrons within the ASD and always added to the incoming photons of light during spectra collection. This adversely affects the spectra collection and has to be removed. This process is known as “Dark Current Correction.” Spectral data collection requires instrument calibration using a reference panel (white reference) provided along with the instrument. During the white reference collection, a reference 100 % line is available to the user to check the status of the instrument performance. White reference collection

**Fig. 5** Spectral library plots**Fig. 6** Pixel purity index plot

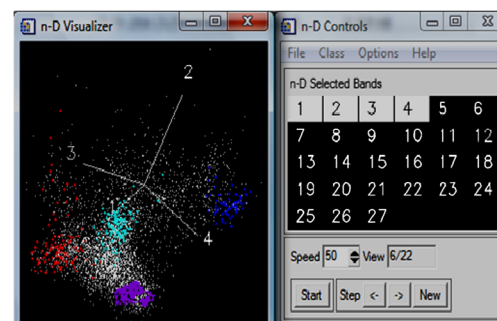
- includes dark current correction and was repeated every 30 min during the collection of sample spectra. This minimizes the effects of the changing lighting conditions on the recorded spectra.
2. Spectral data collection: The 10 different spectra were collected from ASD spectroradiometer belonging to the major land cover and earth surface materials present in the study area. All the spectral measurements have been collected during noontime to avoid the impact of illumination changes on the spectral responses. A spectral library of pure earth surface materials such as urban, vegetation, soil etc. is created for the Dehradun City and is taken up for further processing.
- (3) Resampling of the spectral data: Since ASD collects the data in 2151 continuous bands, it was needed to resample it to the Hyperion wavelengths as later comparison between the two was required.

Figure 5 shows the final spectral library developed for different land covers.

#### Endmember extraction using PPI and SMACC

Collecting endmember using PPI technique was accomplished in three steps:

1. Dimensionality reduction using MNF (minimum noise fraction) algorithm
2. Applying the PPI to MNF images

**Fig. 7** N-dimensional visualizer

### 3. Visualizing the results of PPI in n-dimensional visualize of ENVI and selecting the endmember cloud

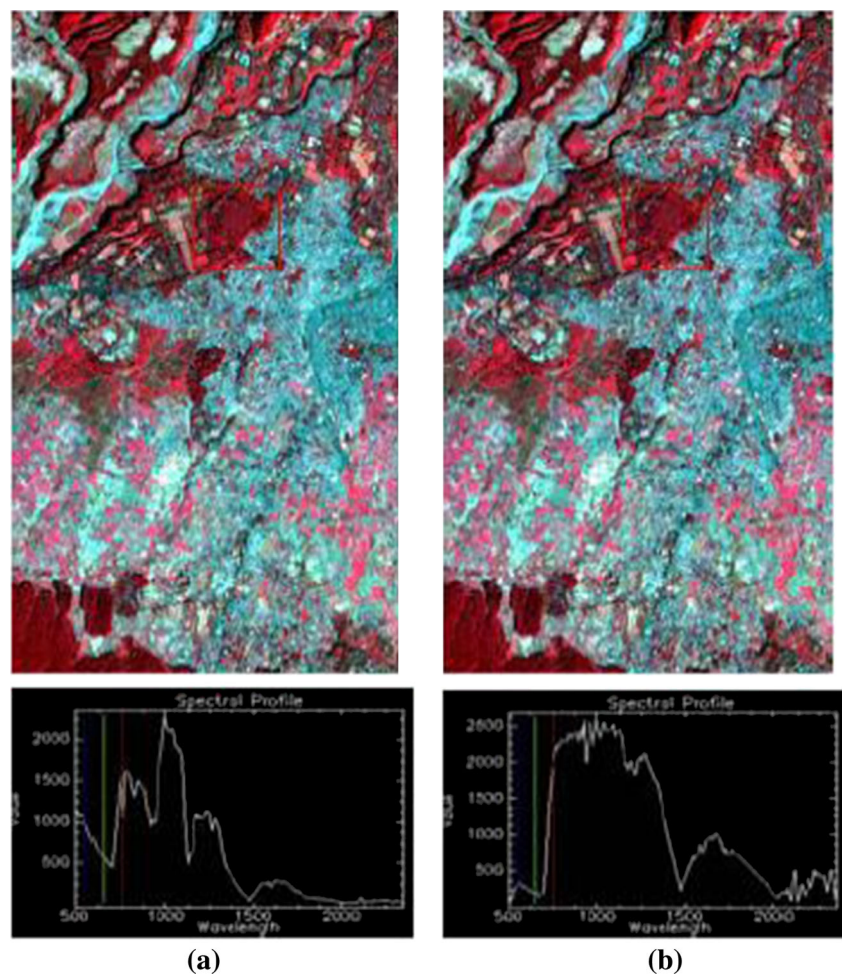
Minimum noise fraction (MNF) transformation was used for reduction of the dimensionality of the data. It applies principal component analyses (PCAs) to produce eigen images used for noise reduction and to provide useful data. The MNF analysis provides an image that displays minimum noise data with lighter color areas (generally) containing the least amount of noise or variation within a pixel. The MNF provides a useful filtering step in determining the location of the pure pixels of the material of interest. A forward MNF transformation was applied to the data and 24 out of 149 bands were noise free enough for classification purposes.

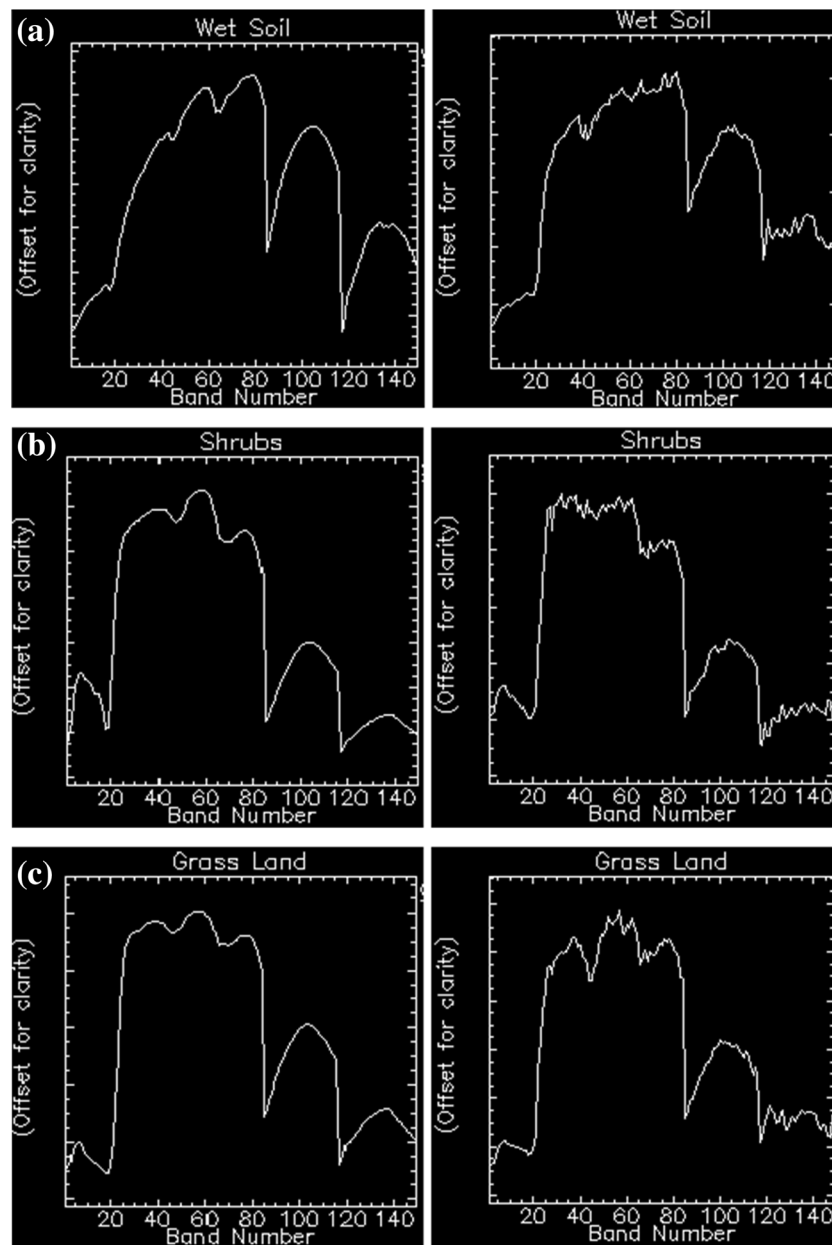
After MNF transformation, the resulting useful data was then processed using a PPI where the threshold value was 2.5 and processed through 10,000 iterations. The PPI has been widely used for endmember extraction to find the most “spectrally pure” pixels in the MNF image. The object is to find pixels that are dominated

by the material of interest as opposed to pixels that are mixed with many other species or land covers. The PPI process looks at where pixels fall within an n-dimensional data cloud. The more pure a pixel, the more it is likely to fall within a distinct dimension within the data cloud. A user-specified threshold helps to determine how much of the data will be considered a hit. Those pixels with the most hits are considered the most pure. Figure 6 shows the pixel purity index plot.

The n-dimensional visualizer was used for endmember determination of these spectrally purest pixels. The endmembers were automatically collected and used in the classification of the reflectance data. This procedure generates clouds of points related to the pixels in n-dimensional space defined by the MNF components. This tool allows manipulating the clouds providing a better positioning to discriminate different spectral groups (as shown in Fig. 7). This procedure promotes a better definition of the groups and the distinction can be manually made using an interactive drawing tool. Thus, this method consists in a classification process where the

**Fig. 8** Spectral profile and image **a** before and **b** after atmospheric correction





**Fig. 9** **a** Wet soil. **b** Shrubs. **c** Grassland. **d** Green leaves. **e** Bituminous road. **f** Dry soil

analyst defines the classes. The set of the selected points can be isolated in groups and analyzed by statistical processes.

SMACC is an endmember extraction approach that selects the endmembers directly from the hyperspectral data. After applying the atmospheric and geometric correction on Hyperion data, SMACC is applied. SMACC selects a number of desired endmembers spectra and their corresponding abundances with some tolerance. The algorithm for finding the endmembers is sequential: the convex cone model starts with a single endmember and increases incrementally in dimension. Abundance maps are simultaneously generated and updated at each step. A new endmember is identified based on

the angle that it makes with the existing cone. The data vector which is making the maximum angle with the existing cone is chosen as the next.

## Results and discussions

### Atmospheric correction results

After running the FLAASH model, the haziness in the image is minimized to a certain level and the features are sharpened



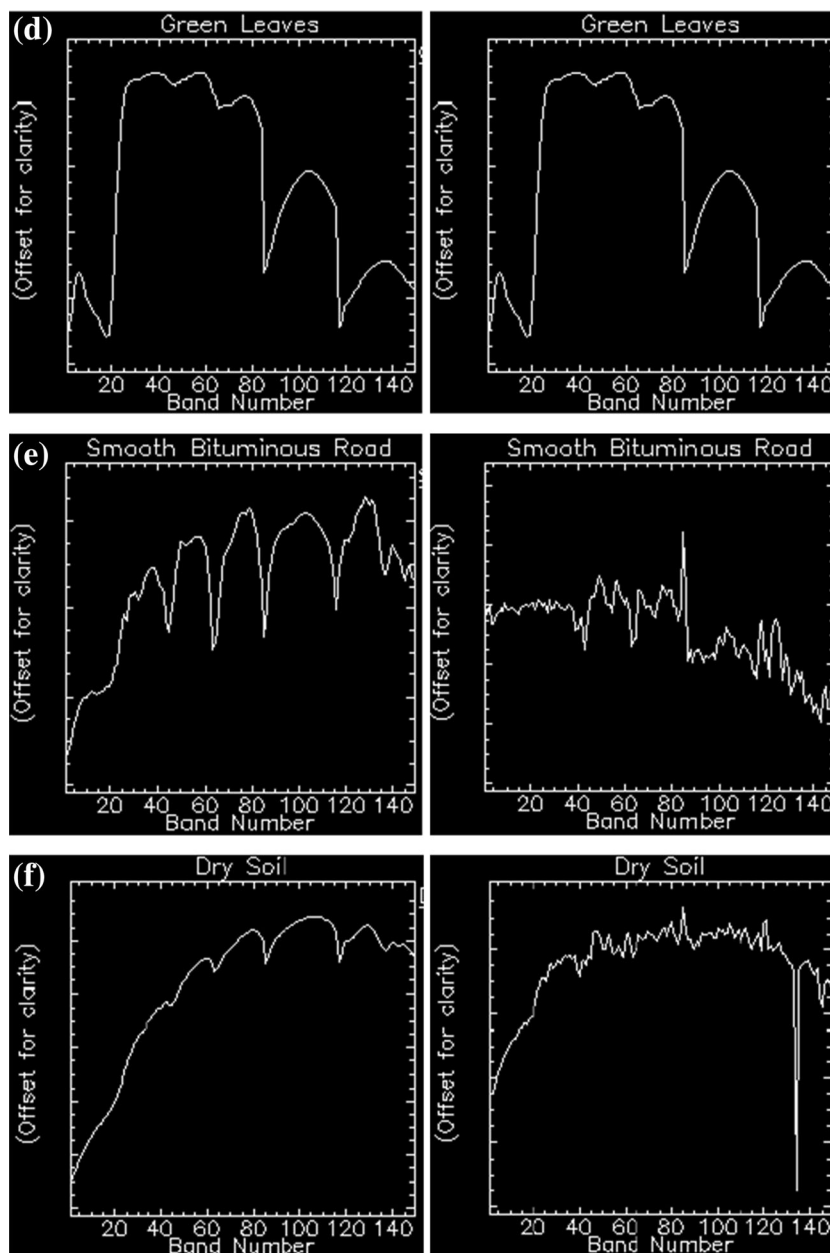


Fig. 9 (continued)

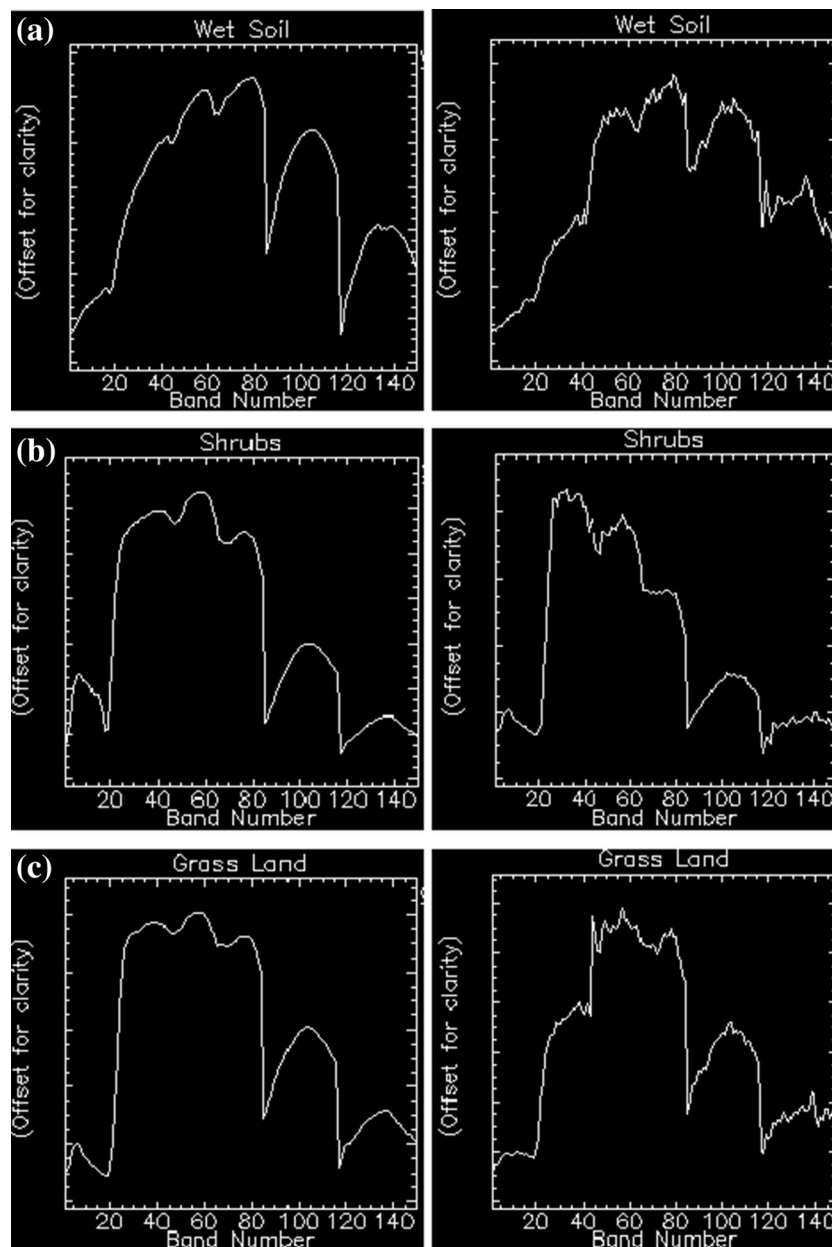
with increased brightness. The features like river, agricultural land, urban areas etc. are looking clear in the image after correction than before. In Fig. 8, there is a difference in the feature seen before and after running FLAASH model. The suppressed spectral response at some wavelengths has improved after atmospheric correction.

#### Result of endmember extraction algorithms

The endmember spectra extracted from the Hyperion image have been compared with the reference spectral

library created for six classes. Figure 9a–f shows the endmember spectra extracted using PPI (right) and the reference spectra (left).

Similar comparison has been carried out for endmember spectra extracted using SMACC (Fig. 10a–f). Figures clearly illustrate dip in the reflectance curve at band 85 and 115 due to water absorption. Dip at band 20 in vegetation classes is due to chlorophyll absorption. As roads are barely visible on the image, their spectra are poorly correlated to reference spectrum.



**Fig. 10** a Wet soil. b Shrubs. c Grassland. d Green leaves. e Bituminous road. f Dry soil

### Comparison of PPI and SMACC algorithms

For comparison of the two techniques of endmember collection, SAM (Kruse et al. 1993) scores have been evaluated, which is a widely used measure for assessing spectral similarity. It is similar to angle in multidimensional space. Lesser SAM score indicates more spectral similarity between reference spectrum and extracted endmember spectrum and hence better performance of the endmember extraction algorithm (EEA).

Table 4 summarizes the SAM scores for spectra extracted from PPI and SMACC with respect to reference spectra.

It can be observed from the comparative analysis that SMACC algorithm performs better in comparison to PPI in extracting endmember spectra of vegetation class, while PPI algorithm performed better for dry soil and bituminous roads.

### Conclusions and recommendations

The process of extracting endmember spectra from preprocessed Hyperion image yields good results from the existing algorithms of PPI and SMAAC.

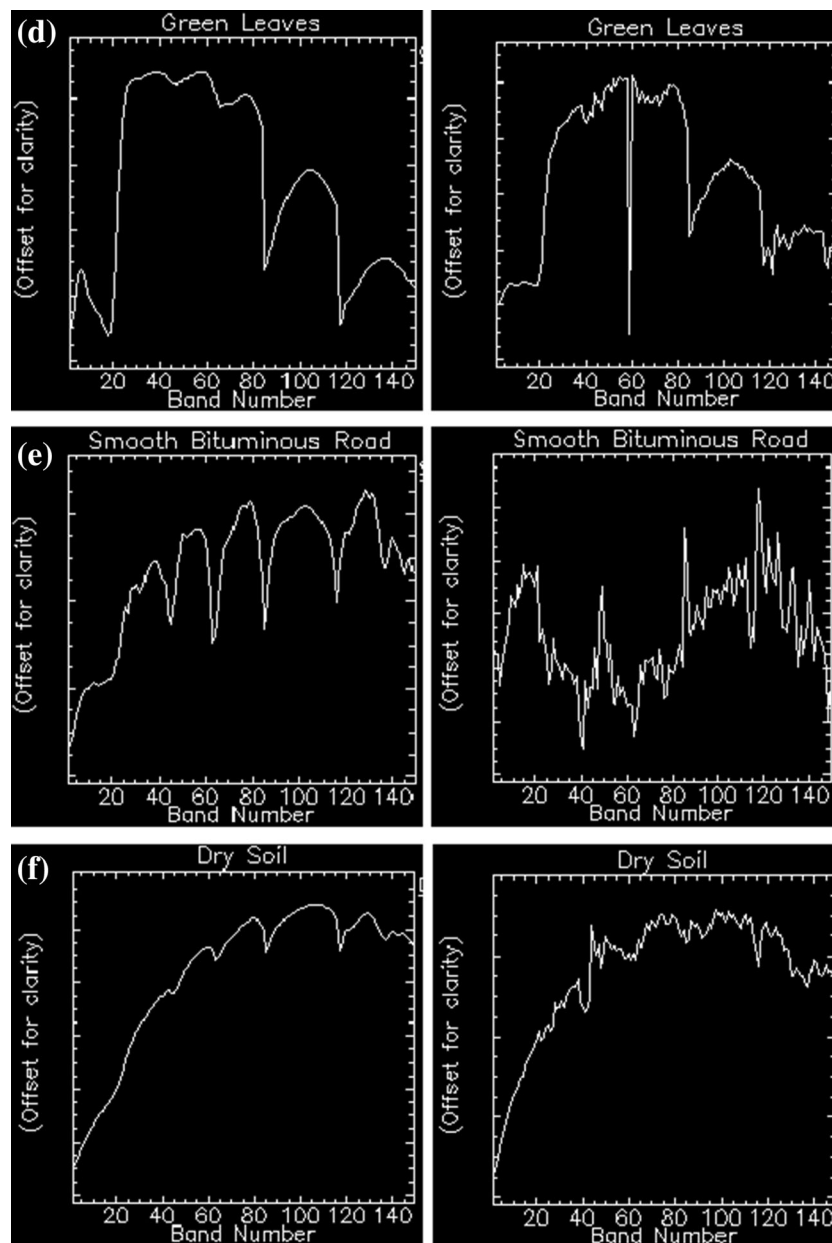


Fig. 10 (continued)

**Table 4** Comparative analysis using SAM scores

Class	SAM score	
	PPI	SMACC
Dry soil	0.809	0.874
Wet soil	0.883	0.803
Grass land	0.910	0.790
Green leaves	0.809	0.785
Shrubs	0.874	0.793
Smooth bituminous road	0.858	0.911

Atmospheric and radiometric corrections play a vital role in extracting the true endmember spectra. Atmospheric errors like haziness in the scene can cause considerable errors in the final endmembers.

It has been observed that spectral profile of pixels improves to a great extent after atmospheric corrections.

It has been observed that SMACC performs better for vegetation class in the present study area as compared to PPI, while converse is true for the case of roads and dry soil. Field observation gives a clear picture about the algorithm to be used for extracting the endmembers. So, before selecting an algorithm for endmember extraction in a given study area,

comparison using field observations is recommended. It can be seen from the study that choice of endmember extraction algorithm depends on the type of materials present in the study area or the target material for which endmember spectrum needs to be extracted. Further investigation is needed to check the performance of these algorithms in the presence of other materials and land cover type.

## References

- Boardman JW (1994) Geometric mixture analysis of imaging spectrometry data. *Proc Int Geosci Remote Sens Symp* 4:2369–2371
- Green AA, Berman M, Switzer P, Craig MD (1988) A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans Geosci Remote Sens* 26(1):65–74
- Gruninger J, Ratkowski AJ, Hoke ML (2004) The sequential maximum angle convex cone (SMACC) endmember model. *Proceedings SPIE, Algorithms for Multispectral and Hyper-spectral and Ultraspectral Imagery*, Vol. 5425-1, Orlando, Florida
- Han T, Goodenough D (2002) Detection and correction of abnormal pixels in hyperion images. *IEEE International Geoscience and Remote Sensing Symposium*. pp. 1327-1330
- Keshava N, Mustard J (2002) Spectral unmixing. *Signal Processing Magazine*. *IEEE* 19(1):44–57
- Kruse FA, Lefkoff AB, Boardman JB (1993) The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*. pp. 145-163
- Mozaffar MH, Zoj Valadan MJ, Sahebi MR, Rezaei Y (2008) Vegetation endmember extraction in hyperion image. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B7. Beijing
- Nascimento J, Bioucas-Dias J (2005) Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans Geosci Remote Sens* 43(4):898–910
- Theiler J, Lavenier DD, Harvey NR, Perkins JJ, Szymanski JJ (2000) Using blocks of skewers for faster computation of pixel purity index. *SPIE Proc* 4132:61–71
- Winter ME (1999) N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. *Proceedings of SPIE*, pp. 266-275, 1999