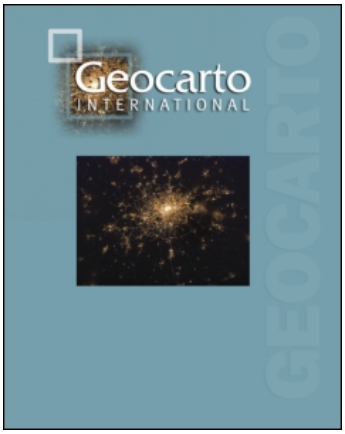

Geocarto International
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t759156373

Noise-signal index threshold: a new noise-reduction technique for generation of reference spectra and efficient hyperspectral image classification

K. N. Kusuma[a]; D. Ramakrishnan[a]; H. S. Pandalai[a]; G. Kailash[b]
[a] Department of Earth Sciences, Indian Institute of Technology, Powai, Mumbai, India [b] Department of Electrical Engineering, Stony Brook University, NY, USA

Accepted uncorrected manuscript posted online: 23 August 2010

First published on: 16 September 2010

To cite this Article Kusuma, K. N. , Ramakrishnan, D. , Pandalai, H. S. and Kailash, G.(2010) 'Noise-signal index threshold: a new noise-reduction technique for generation of reference spectra and efficient hyperspectral image classification', Geocarto International, 25: 7, 569 — 580, First published on: 16 September 2010 (iFirst)
To link to this Article: DOI: 10.1080/10106049.2010.510582
URL: http://dx.doi.org/10.1080/10106049.2010.510582

Geocarto International
Vol. 25, No. 7, November 2010, 569–580

Taylor & Francis
Taylor & Francis Group

Noise-signal index threshold: a new noise-reduction technique for generation of reference spectra and efficient hyperspectral image classification

K.N. Kusuma[a], D. Ramakrishnan[a]*, H.S. Pandalai[a] and G. Kailash[b]

[a]*Department of Earth Sciences, Indian Institute of Technology, Powai, Mumbai 400076, India;* [b]*Department of Electrical Engineering, Stony Brook University, NY 11794, USA*

(*Received 18 February 2010; final version received 20 July 2010*)

Reference spectra of terrestrial targets are usually collected using field spectro-radiometers for mineral abundance mapping and target detection. These spectra often have noise that masks characteristic absorption and reflection features and affects the efficiency of material mapping. This work aims at obtaining an empirical technique for reduction of high-frequency noise from field spectra. The proposed noise correction technique uses a 'normalized' measure R_n, where $R_n = (L_n - F_n)/L_n$ for each band (n) calculated from field and laboratory spectra of test material, with F_n and L_n being the depth of the absorption feature in field and laboratory spectra, respectively. On the basis of the assumption of the constancy of this ratio in neighbouring bands, an empirical algorithm that approximates the ratio R_n of a noisy band to the corrected ratio of an adjacent band is used to obtain the noise-corrected field spectra. The classification accuracy increases significantly when noise reduced field spectra are used as reference spectra.

Keywords: reflectance spectra; noise; empirical correction; classification; hyperspectral remote sensing

1. Introduction

A state-of-the-art imaging spectrometer collects reflectance data that are both spatially and spectrally contiguous. This provides an efficient means of obtaining information on terrestrial materials and their abundance. With the advent of NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and EO-1-Hyperion sensor systems, laboratory-like imaging spectra of a wide variety of Earth's surface materials such as rocks, soils, plants, snow, ice, water and artificial materials are made possible. This allows comparison of field and laboratory spectra with the pixel spectra of the satellite image and thereby helps in the direct identification of materials including surface minerals and their abundance (Goetz *et al*. 1985, Kruse *et al*. 1993, Rowan *et al*. 2004, Van der Meer 2004), atmospheric constituent gases (Marion *et al*. 2004, Black and Guo 2008), vegetation species discrimination (Tsai *et al*. 2007) and water quality (Brando and Dekker 2003).

*Corresponding author. Email: ramakrish@iitb.ac.in

ISSN 1010-6049 print/ISSN 1752-0762 online

DOI: 10.1080/10106049.2010.510582
http://www.informaworld.com

For target mapping, reference spectra of terrestrial target objects are generally collected using field or laboratory spectro-radiometers. Similar to their satellite-sensor counterparts, the field spectra are also vulnerable to degradation by wavelength-specific and non-specific noise arising from several parameters such as atmospheric water vapour, scattering, instrument performance and faulty data acquisition (Schmidt and Skidmore 2004, Ramsey *et al*. 2005). Green and Conel (1995) envisaged the deleterious effects of increase in water vapour and aerosol content on ground leaving reflectance of low reflecting materials. From their work it is evident that increase in atmospheric water vapour and aerosol content can reduce the effective solar irradiance in 2000–2500 nm regions. Water absorption bandwidths such as 900–980 nm; 1351–1434 nm and 1798–1961 nm are generally affected by specific noise. Besides these regions, the 2350–2500 nm band width is commonly affected by noise due to low solar irradiance.

Such noise affects the fundamental and overtone absorption features of minerals rich in iron (900–1000; 1800–2000 nm), Al—OH (2200 nm), Mg—OH (2300–2400 nm) and carbonates (1900, 2350, 2500 nm), thereby reducing the efficiency of rock and mineral mapping. Thus, pre-processing of the field-spectra of target objects is ideal before they are used for mapping and abundance-estimation (Schaepman and Dangel 2000, Liu *et al*. 2006).

One of the methods of minimizing noise in reference spectra is to collect the spectral signatures of the target materials using instruments having higher signal-to-noise ratio in moisture-free laboratory conditions with longer integration time. However, such ultra-fine reference spectra seldom match the satellite-acquired pixel spectra because of inherent heterogeneity associated with field conditions. This necessitates reducing the atmospheric- and the instrument-related noise to generate a reasonable field reference spectrum. Prevailing methods of noise management in field-spectra include either removal of the noisy segment or noise reduction by application of signal processing techniques such as Fourier transforms, Savitzky–Golay local polynomials, Gaussian functions, wavelet decomposition and artificial neural networks (Savitzky and Golay 1964, Schmidt and Skidmore 2004, Shafri and Mather 2005). Though these techniques reduce noise effectively, they can also affect the characteristic spectral features of materials such as central wavelength and absorption strengths of different materials. This affects their identification and abundance-estimation. In this communication, an empirical method is proposed which has not only the potential of reducing the noises from the field spectra, but also maintains the characteristic spectral features.

2. Methodology

2.1. Spectral data generation

The methodology adopted involves collection of field and laboratory reflectance measurements, processing of spectra and mapping of the target materials in the hyperspectral image data cube. Bi-hemispherical-conical reflectance of target materials such as laterites, bauxites and lateritic–bauxites were measured (in 350–2500 nm wavelength range) using the ASD-Fieldspec3 portable field spectro-radiometer. The spectra were collected from a nadir-looking optical setup. The instrument was calibrated using spectralon diffused reflectance panel (Labsphere, North Sutton, USA) and dark current.

Reflectance of the samples collected from the corresponding field locations was measured in the laboratory by keeping the same angular parameters. Processing of the spectra for noise-reduction was carried out using Matlab routines written for the above purpose and detailed in the next subsection.

Before matching the reference and the pixel spectra, the satellite data needed to be corrected for atmospheric effects and spectral shift. The EO-1 Hyperion satellite radiance data of the study area were pre-processed for atmospheric effects using the FLAASH module (Matthew *et al*. 2000) built into the ENVI (Ver. 4.3) software and converted to corresponding surface reflectance values. The data were collected over a part of the Konkan area in the Savitri basin (N17°45′–18°25′ and E73°00′–73°40′) that consists of lateritized terrain, underlain by Deccan basalt. Typical rock-types include basalt, laterite, lateritic–bauxite and bauxite. The details of parameters used for the atmospheric correction are given in Table 1.

The spectral angle mapping (SAM) technique (Boardman 1993, Kruse *et al*. 1993) was used as the mapping method to match the reference spectra with the pixel spectra. The accuracy of mapping was estimated with the aid of field data collected from 50 locations.

2.2. *Noise removal and generation of reference spectra*

The intensity of the spectra of samples under solar and laboratory illumination acquired at 10 nm bandwidth are termed here as F_n and L_n respectively with n indicating the wavelength/band number (n = 1–2150). The irradiance curves of solar and laboratory illuminations were recorded at 1 nm bandwidth and compared after normalization for illumination using the spectralon panel. The two normalized spectra differ with respect to intensities of illumination over all wavelengths, but the patterns of the two spectra are more or less similar (Figure 1(A)). Figure 1(B), shows the plot of the difference $(L_n - F_n) = D_n$, against wavelength n. It is seen that the difference is not constant across bands. On plotting D_n vs. L_n for chosen ranges of data (Figures 2(A)–(C)), it is observed that D_n is linearly correlated with L_n, the reflectance values of the lab spectrum in all wavelength regions such as visible, near infra red (NIR) and short-wave infra red (SWIR) having very high R^2 values (0.81–0.99). This strongly suggests that D_n is proportional to L_n (or $\frac{D_n}{L_n} = k$). The overall linear relationship between D_n and L_n may be due to the overall effect of instrument gain for higher intensities of illumination. At water absorption bands and SWIR regions, the bipolar noises present in the field spectra reduce the R^2 values (0.6–0.8).

Table 1. Parameters used to correct the atmospheric effects in hyperion images of the investigated area.

Atmospheric model	Tropical
Aerosol model	Rural
Water vapour absorption feature	1135 nm
Initial visibility	20 kms
Retrieved atmospheric water vapour	1.4 cms
Spectral shift	2–3 nm in visible and NIR regions 6 nm in SWIR region
Correction for adjacency effect	Yes
Spectral polishing	Yes

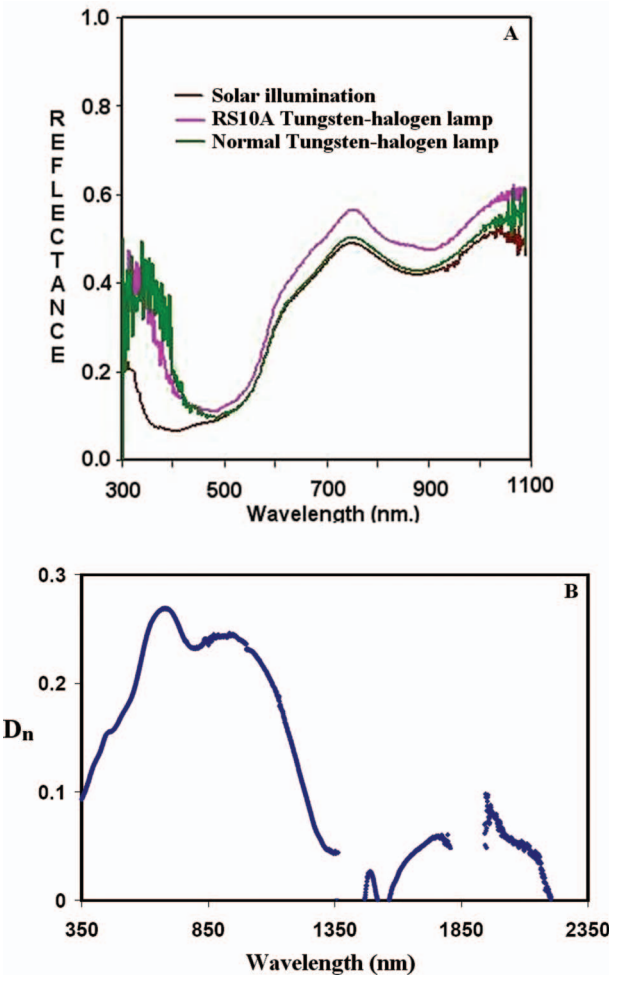

Figure 1. (A) Spectra of laterite collected using different illumination sources. (B) D_n vs. wavelength (n) plot.

This apparently linear relationship is drastically disturbed in some bands, and it is inferred that this disturbance is due to noise. The value D_n/L_n $(=R_n)$ for each band number is random due to inherent noise in each band and the value of R_n computed for each band is an estimate of the constant k. When noise-levels are low, the values of R_n and R_{n+1} in adjacent bands are nearly constant. On the other hand, when there is a sudden change in the value of R_{n+1} relative to R_n in consecutive bands, it is inferred that this change is due to noise. This noise may be attributed to random effects (such as atmospheric or instrument effects). As it is desirable to correct the field spectra for such noise, a suitable algorithm was devised for correction and is described below.

The algorithm may proceed either from band number 1 to band number 2150 or in reverse from band number 2150 to band 1. The first procedure is outlined below. The ratio $\frac{[R_{n+1}-R_n]}{R_n}$ is computed and whenever the value of this ratio exceeds a

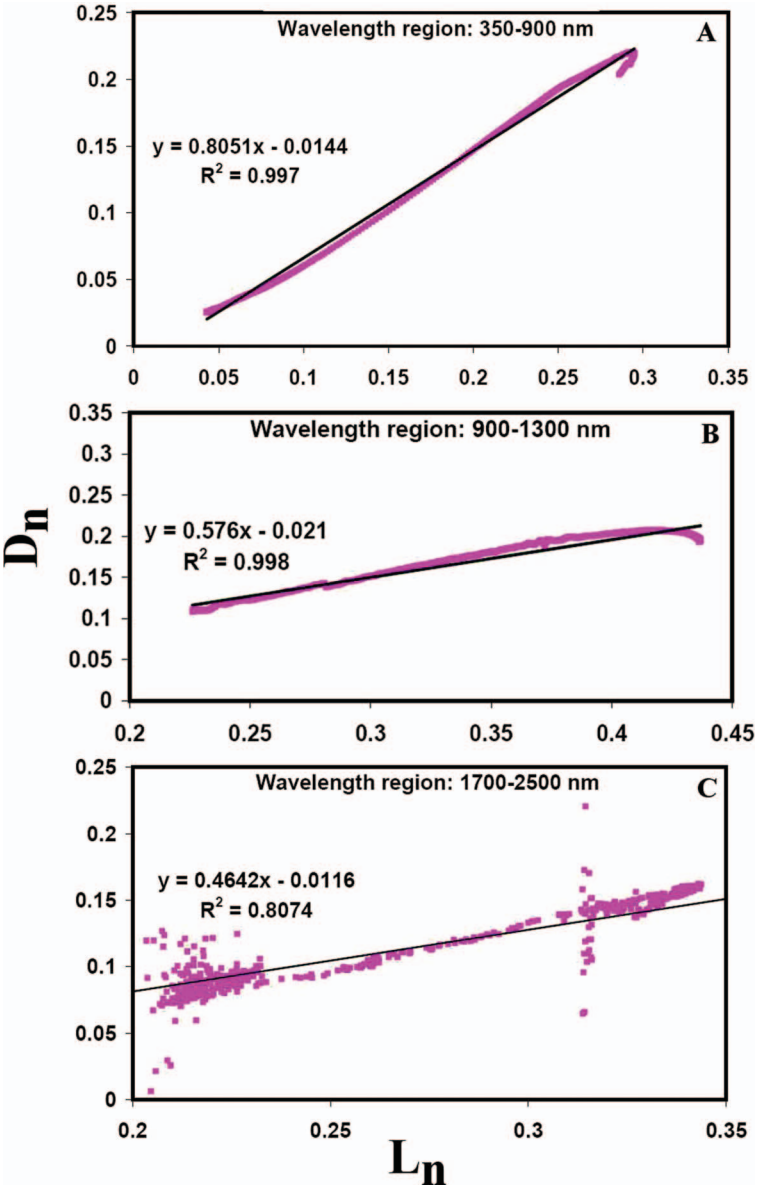

Figure 2. D_n vs. L_n plot depicting the linear relationship.

threshold value, R_{n+1} is replaced in the equation $R_{n+1} = \frac{L_{n+1} - F_{n+1}}{L_{n+1}}$ by R_n (or $R_n^{'}$, its corrected value). In general one may write $R_n^{'} = R_n$ if no correction occurs and then the appropriate value of the field spectrum for band number $n+1$ is calculated using the relation

$$F_{n+1} = L_{n+1} - (L_{n+1} \times R_n^{'}) \tag{1}$$

While proceeding sequentially towards higher band numbers, it is assumed that the value of R_n or its corrected value $R_n^{'}$ (which may have been obtained in the

preceding step), is a better estimate of k than R_{n+1}. The converse would be true when proceeding from a higher band number towards a lower band number.

To start with, the value of R_1, the ratio in the first band of the spectrum, is checked to verify whether it is within acceptable limit. This is done by comparing it with the local mean computed over a few bands, typically 3 to 10 in case of high-frequency noises. The local mean is used as an estimate of the value of R_0 as instrumental drift and other errors may result in the drift of the mean value of R_n over the entire bandwidth being studied. In this study, an average of three bands is found to be optimum to obtain an average value designated as $R_0 = \frac{1}{3}\sum_{i=1}^{3} R_i$. The value of R_0 is used to compute $|\frac{R_1 - R_0}{R_0}|$ to check whether R_1 is acceptable on the basis of a threshold value (T) selected for the purpose. A value of the ratio that is higher than the chosen threshold is considered as arising due to high-frequency noise. This leads to correction of R_1 by approximating it with R_0 and computation of a corrected value of F_n as given in Equation (1). The corrected ratio is given the notation $R_1^{'}$. If no correction is required, $R_1^{'}$ is taken as equal to R_1. The algorithm proceeds successively to the next channel and compares the ratio R_{n+1} with $R_n^{'}$ and computes the relative change given by the ratio $\left(R_{n+1} - R_n^{'}\right)/R_n^{'}$. This ratio is herein termed the 'noise-signal index' as it is a measure of the relative noise-to-signal ratio of adjacent bands. The critical assumption here is that $D_n/L_n = k$, which is a measure of the noise-to-signal ratio, is nearly constant for adjacent bands.

An attempt was made to understand the effect of threshold values on efficiency of noise-correction. For this purpose, the noise-signal index threshold was progressively increased ($\Delta T = 0.04$) from 0.01 to 0.55 for spectra corresponding to laterite, lateritic–bauxite and bauxite (Figures 3 (A)–(E)). Figure 3(A) is the noise ridden field spectra with characteristic bipolar noises in $-$OH (1.4 μm), H$-$OH (1.9 μm), FeO$-$OH (2.41 μm), Mg$-$OH (2.44 μm) and CO_3 (2.35, 2.5 μm) absorption regions. Figure 3(B) corresponds to the spectra of same materials collected in controlled, laboratory conditions with the least noise in all wavelength regions. When noise-to-signal indexing were carried out using these spectra, at lower thresholds (Figure 3(C)), the noise are significantly reduced. But, the spectral pattern has some similarity with the laboratory spectra, which is undesirable. At threshold values more than 0.2, noise starts to appear and increases with higher values of threshold (Figure 3(E)). After carefully analysing the spectra, the threshold range 0.11–0.13 is considered optimum for noise-reduction. At this threshold range, both the objectives, i.e. noise-reduction and preserving the field spectral pattern are achieved (Figure 3(D)).

3. Results

3.1. Noise-corrected reference spectra and image analysis

The noise-corrected field-spectra, herein designated as reference-spectra of laterite, lateritic–bauxite and bauxite were generated by the above technique with a noise-signal index threshold of 0.13. The root mean square error between the field- and the reference-spectra, a measure of goodness of fit, varies between 0.01 and 0.2. This clearly indicates least disturbance to spectral characters of reference-spectra while restoring the spectral features in the noise-affected field spectra.

These reference-spectra were compared subsequently with the satellite pixel spectra for mapping the materials of interest. For this purpose, SAM was used.

The SAM technique directly compares the image spectra to reference spectra given by

$$\alpha = \mathrm{Cos}^{-1}\left(\frac{\sum_{i=1}^{nb} t_i r_i}{\left(\sum_{i=1}^{nb} t_i^2\right)^{1/2}\left(\sum_{i=1}^{nb} r_i^2\right)^{1/2}}\right) \quad (2)$$

where, i is the band number, nb is the total number of bands, t_i is the amplitude of pixel spectrum and r_i is the amplitude of reference spectrum.

Since the SAM algorithm uses only the vector direction, it is insensitive to illumination/intensity-related differences. In the present case, the spectral angles of 0.30, 0.20 and 0.10 radians were used as thresholds for bauxite, lateritic–bauxite and

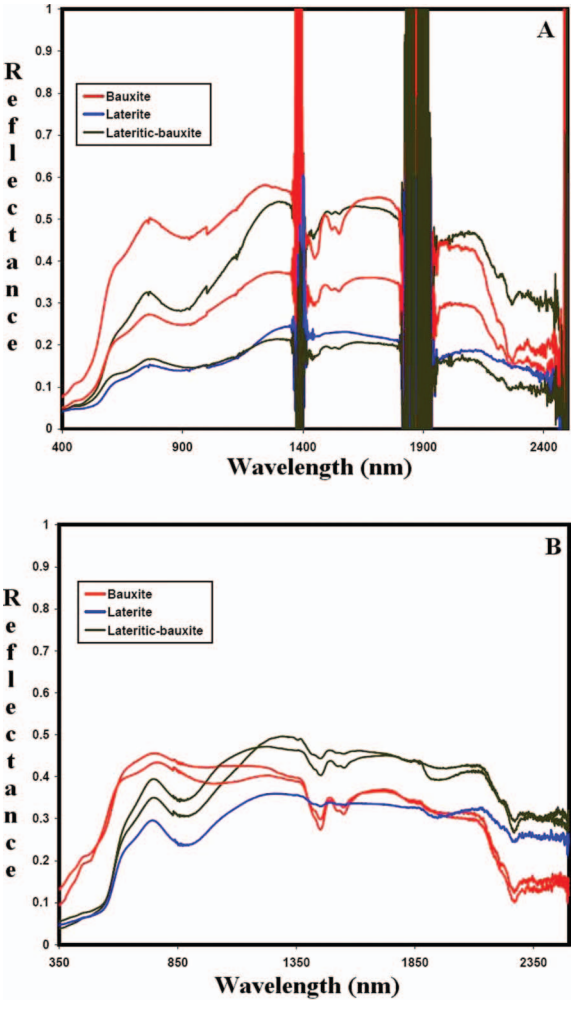

Figure 3. Spectra of laterite, lateritic–bauxite and bauxite collected at (A) field and (B) laboratory conditions. (C) and (D) Noise-corrected spectra of laterite, lateritic–bauxite at minimum (0.01) and optimum (0.13) noise-signal index thresholds respectively. (E) Noise-corrected spectra of laterite, lateritic–bauxite and bauxite at maximum (0.21) noise-signal index threshold.

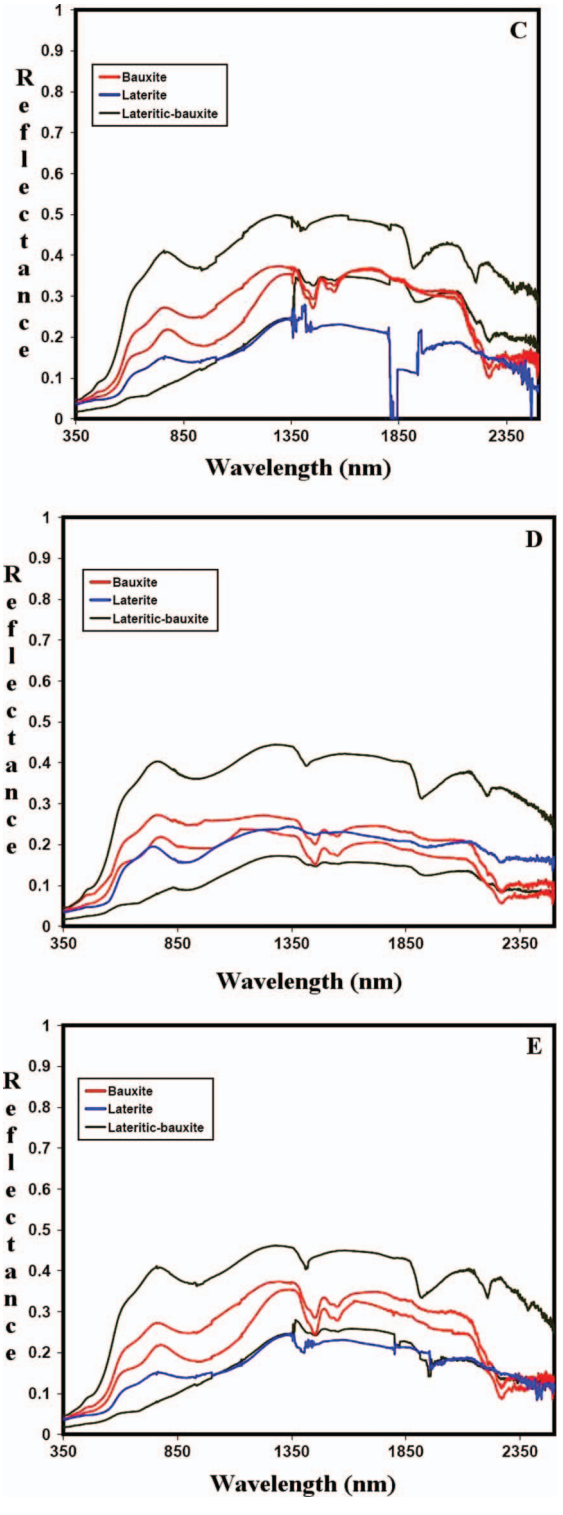

Figure 3. (*Continued*).

laterite, respectively. These optimum values were arrived at after attempting several combinations of spectral angles ranging between 0.1 and 0.4 for all three classes. It was observed that these optimized spectral angles for each class alone gave the best possible classification accuracy. It is evident from the classification results of SAM (Figure 4) that the classification done by using the noisy spectra could not discriminate between the laterite and lateritic–bauxites, whereas the same could be achieved with noise-corrected reference-spectra. The change detection statistics (Table 2) indicate that when classification was done by using noise-corrected reference-spectra, subtle differences among the three classes such as bauxite, lateritic–bauxite and laterite became more apparent. It is perceptible that the bauxite class, mapped by using noisy spectra, includes 6.5% of lateritic–bauxite and 22.1% of other classes (unclassified). Similarly, the lateritic–bauxite class is found to incorporate 13.7% of laterite, 0.8% of bauxite and 0.3% of other classes. The user's and the producer's classification accuracies were also estimated following the procedure of Congalton (1991) by comparing the classification result with ground-truth information obtained at 50 locations for both methods of classification

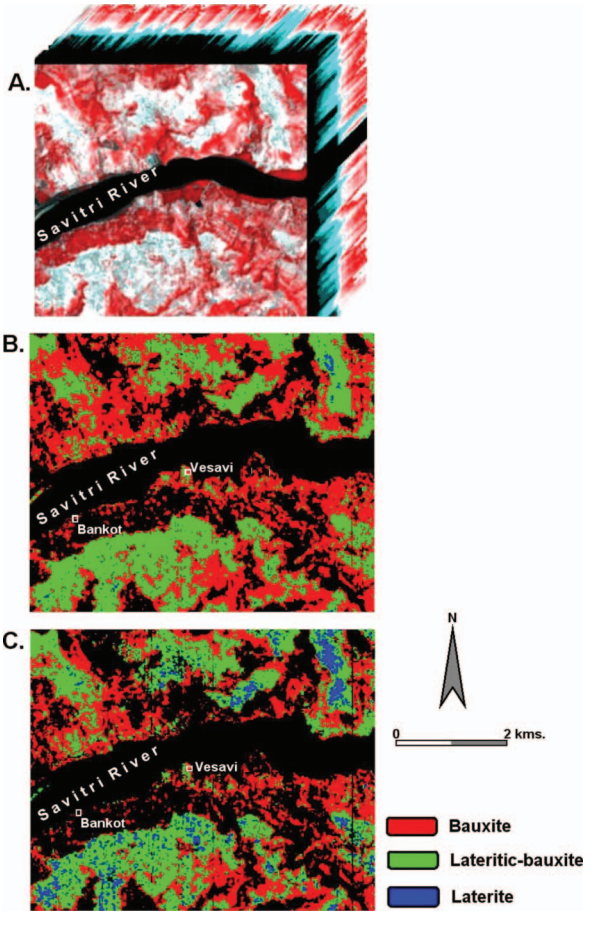

Figure 4. (A) Hyperion imagery cube, (B) SAM-classified output using noisy spectra, and (C) noise-corrected spectra.

(Table 3). The overall accuracy of classification significantly improved for all three classes, when noise-corrected spectra were used.

4. Discussion and conclusion

In this study, it is observed that the field spectra are usually found to contain noises in SWIR region. However, when measurements are made in the laboratory conditions using a tungsten halogen source lamp ($\lambda_{max} = 1200$ nm), the spectra are noiseless. These indicate to the relevance of atmospheric water vapour and dust particles in decreasing the effective solar irradiance in 2000–2500 nm regions. With lower irradiance and associated reflectance, it is often difficult to resolve the incoming signals from noises. Noise in the wavelength regions of interest often hampers spectral absorption features. It is evident from the classification results (Tables 2 and 3) that the use of noise-ridden field spectra reduce the efficiency of spectral discrimination among laterite, lateritic–bauxite and bauxite and hence, the accuracy of classification. This occurs because of the following reasons.

(1) Presence of noise at critical wavelengths, in this case at 940–950, 1349–1475 and 2225–2500 nm, affects the spectral pattern and hence the efficiency of discrimination among the classes.
(2) Spectral-matching algorithms like SAM are a measure of the cumulative match between the pixel- and the reference-spectra at all wavelengths. Hence, the presence of noise at any given wavelength obviously leads to deterioration of the overall match-score.

Table 2. Change detection statistics (in percentage) between SAM classified images using noisy spectra (n) and noise-corrected spectra (nc).

	Unclassified (n)	Bauxite (n)	Lateritic–bauxite (n)	Laterite (n)
Unclassified (nc)	99.6	22.1	0.3	0.0
Bauxite (nc)	0.4	71.4	0.8	0.0
Lateritic–bauxite (nc)	0.0	6.5	85.2	0.0
Laterite (nc)	0.0	0.0	13.7	100
Class total	100	100	100	100

Table 3. Accuracy assessment of the image classification based on correlation with observations made in the field at many locations.

Class	Accuracy	SAM classified map (noisy spectra)	SAM classified map (noise-free spectra)
	Overall	47%	66.67%
	κ	0.15	0.52
Laterite	Producers accuracy	0	86%
	Users accuracy	0	55%
Lateritic–bauxite	Producers accuracy	39%	44%
	Users accuracy	47%	80%
Bauxite	Producers accuracy	68%	75%
	Users accuracy	52%	100%

When the relative and the absolute accuracy of classification between noise-free and noisy-spectra is compared, the latter produces inferior results for all three classes. The work thus demonstrates the significance of noise-correction in field spectra in improving classification accuracy. The proposed noise-signal threshold index could be advantageous over Fourier transforms, Savitzky–Golay local polynomials, and Gaussian functions, in terms of its efficiency in portraying the original spectral pattern. This property of the proposed method is attributed mainly to selective reconstruction of spectral features in noisy regions, with the aid of laboratory spectra, and maintaining the original field spectral characters in noise-free regions. Other methods not only affect the entire spectra but also affect the absorption/reflection features and shift characteristic wavelength positions. Although the necessity of noise-free laboratory spectra is a limitation to this study, the efficiency of spectral restoration in noisy regions and associated improvements in classification accuracy certainly outweigh this constraint. This study clearly indicates that as long as the noise-free laboratory-spectra and the noisy field-spectra are similar, the proposed technique can be effectively used even in situations where within-class spectral variability exists.

Acknowledgements

The authors are thankful to Professor S.S. Major, IIT, Bombay and Dr. R. N. Sahoo, IARI, New Delhi for instrumentation support. Constructive criticism from two unknown reviewers is highly appreciated for their help in improving the quality of the manuscript. D.R. acknowledges the Department of Science and Technology, Government of India for financial support.

References

Black, S.C. and Guo, X., 2008. Estimation of grassland CO_2 exchange rates using hyperspectral remote sensing techniques. *International Journal of Remote Sensing*, 29 (1), 145–155.

Boardman, J.W., 1993. Automated spectral unmixing of AVIRIS data using convex geometry concepts: in summaries. *In: 4th JPL airborne geoscience workshop*, Vol. 1. Washington, DC: JPL Publications, 93–126.

Brando, V.E. and Dekker, A.G., 2003. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Transactions on Geoscience and Remote Sensing*, 41 (6), 1378–1387.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, 35–46.

Goetz, A.F.H., *et al*., 1985. Imaging spectrometry for earth remote sensing. *Science*, 228, 1147–1153.

Green, R.O. and Conel, J.E., 1995. Movement of water vapor in the atmosphere measured by an imaging spectrometer at Rogers Dry Lake, CA. *In*: *Summaries of the* 5th annual JPL airborne earth science workshop, JPL Publication 96–4, Vol. 1. Pasadena, CA: Jet Propulsion Laboratory, 87–90.

Kruse, F.A., Lefkoff, A.B., and Dietz, J.B., 1993. Expert system-based mineral mapping in northern Death Valley, California/Nevada using the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, 44, 309–336.

Liu, D., *et al*., 2006. Effects of sensor noise in spectral measurements on chlorophyll – a retrieval in Nanhu lake of Changchun, China. *Journal of Electromagnetic Waves and Applications*, 20 (4), 547–557.

Marion, R., Michel, R., and Faye, C., 2004. Measuring trace gases in plumes from hyperspectral remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 42 (4), 854–864.

Matthew, M.W., *et al*., 2000. Status of atmospheric correction using a MODTRON4 based algorithm. *In*: *SPIE Proceeding, Algorithms for multispectral, hyperspectral, and ultraspectral imagery VI*, Orlando, FL: SPIE, 4049, 199–207.

Ramsey, E., *et al*., 2005. Generation and validation of characteristic spectra from EO1 Hyperion image data for detecting the occurrence of the invasive species, Chinese tallow. *International Journal of Remote Sensing*, 26 (8), 1611–1636.

Rowan, L.C., Simpson, C.J., and Mars, J.C., 2004. Hyperspectral analysis of the ultramafic complex and adjacent lithologies at Mordor, NT, Australia. *Remote Sensing of Environment*, 91, 419–431.

Savitzky, A. and Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36 (8), 1627–1639.

Schaepman, M.E. and Dangel, S., 2000. Solid laboratory calibration of non-imaging spectroradiometer. *Applied Optics*, 39 (21), 3754–3764.

Schmidt, K.S. and Skidmore, A.K., 2004. Smoothing vegetation spectra with wavelets. *International Journal of Remote Sensing*, 25 (6), 1167–1184.

Shafri, H.Z.M. and Mather, P.M., 2005. Wavelet shrinkage in noise removal of hyperspectral remote sensing data. *American Journal of Applied Sciences*, 2 (7), 1169–1173.

Tsai, F., Lin, E.K., and Yoshino, K., 2007. Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species. *International Journal of Remote Sensing*, 28 (5), 1023–1039.

Van der Meer, F., 2004. Analysis of spectral absorption features in hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 5, 55–68.