DETECTION OF HOTSPOTS IN NOAA/AVHRR IMAGES USING PRINCIPAL COMPONENT ANALYSIS AND INFORMATION FUSION TECHNIQUE

R. S. Gautam[*], D. Singh, A. Mittal
{rsgcsdec@iitr.ernet.in, dharmfec@iitr.ernet.in, ankumfec@iitr.ernet.in}
Department of Electronics and Computer Engineering,
Indian Institute of Technology Roorkee, INDIA
Roorkee - 247 667

ABSTRACT:

India accounts for the world's greatest concentration of coal fires which cause several devastating environmental effects. Only Jharia Coal Field (JCF) in Jharkhand (India) contains nearly half of subsurface mine fires (hotspots) in Indian coalfields. Therefore attention is required in this direction for mapping, monitoring and detecting these hotspots. Operational satellite images can be very efficient, effective and economic tool for this purpose. Present paper deals with the potential application of operational satellite images to detect hotspots in Jharia region and proposes an algorithm for the same by employing Principal Component Analysis (PCA) and information fusion technique on NOAA/AVHRR images of Jharia region. PCA is an efficient and effective technique for finding patterns in data of high dimensions and a very powerful tool that provides a new set of images that are linear combinations of the original spectral band images. The algorithm consists of two steps: (1) application of PCA on multi-channel information along with vegetation index information obtained from NOAA/AVHRR image to obtain principal components, and (2) fusion of information obtained from principal component 1 and 2 to classify image pixels as either hotspots or non-hotspots. Results obtained with the proposed algorithm are compared with the results obtained by ground survey and a good agreement is obtained between observed and predicted hotspots.

1. INTRODUCTION

The Jharia coalfield in Jharkhand (India) is an exclusive storehouse of prime coke coal in the country. Most of the fires take place due to spontaneous heating of coal and cause a local rise in the surface temperature, which depends on various mining, geological and coal factors. Satellite images can be one of the best solutions for this problem which offer a cost effective and time saving technology for mapping various geo-environmental features. Several researchers have proposed methodologies to detect hotspots using satellite images acquired by different optical sensors, such as Along Track Scanning Radiometer (ATSR) (Arino *et al.*, 1999), Moderation Resolution Imaging Spectroradiometer (MODIS) (Li *et al.*, 2004), LANDSAT (Brustet *et al.*, 1991), Geostationary Operational Environmental Satellite (GOES) (Prins *et al.*, 1994), Defense Meteorological Satellite Program – Optical Linescan (DMSP/OLS) (Fuller *et al.*, 2000) and Advanced Very High Resolution Radiometer (AVHRR) aboard the National Oceanic and Atmospheric Administration's (NOAA) polar orbiting satellites (Flannigan *et al.*, 1986, Kaufman *et al.*, 1990, Kennedy *et al.*, 1994, Flasse *et al.*, 1996, Nakayama *et al.*, 1999, Boles *et al.*, 2000).

High resolution satellite images such as Landsat TM are quite expensive and offer limited spatial coverage and revisit frequency. In addition, these satellites exhibit strong limitations in acquiring cloud free images. Use of operational satellite images can provide an effective and alternative way to overcome these limitations. Therefore present paper utilizes the potential of NOAA/AVHRR images to detect hotspots in Jharia region. NOAA/AVHRR images have following advantages: (1) These images provide a good balance in spatial and temporal resolutions, (2) NOAA AVHRR images have good temporal coverage (two daily images), (3) AVHRR data provides information over a large geographical area with potentially more cloud free scenes for hotspot detection than other sensors images, such as Landsat TM images and have good sampling frequency and (4) The images are freely acquired, simplifying the generation of long temporal series of images.

Several Techniques such as Fuzzy Logic, Neural Network etc. are being used for hotspot detection in satellite images (Li *et al.*, 2001; Moore *et al.*, 2001; Bardossy and Samaniego, 2002). These techniques have several limitations. They are dependent on training conditions and take very long time to get trained. In this proposed algorithm Principal Component Analysis (PCA) technique has been used for detecting hot spots from the NOAA/AVHRR images of Jharia region. An advantage of PCA over other techniques (i.e. Fuzzy Logic, Neural Network) is that most of the information within all the bands (represented by the variance) can be compressed into a much smaller number of bands with little loss of information. This procedure greatly reduces the computer processing time. It allows us to extract the low-dimensional subspaces that capture the main linear correlation among the high-dimensional image data. This facilitates viewing the explained variance or signal in the available imagery, allowing both gross and more subtle features in the imagery to be seen (Barnsley *et al.*, 1997; Agassi and Yosef, 1998). Until now PCA is continuously being applied to high resolution satellite images for different purposes whereas the novelty of the proposed algorithm lies in the fact that hotspots are detected by employing PCA to low resolution satellite image.

---

[*] Corresponding author.

Objective of the present paper is to propose and evaluate an innovative method to detect hotspots from operational satellite data which is freely available with good temporal resolution, and provide useful information about spatial allocation of hotspots which in turn can help to manage and monitor these hotspots.

2. STUDY AREA AND DATA SET

2.1 Study area

The study was carried out in Jharia coalfield in Jharkhand (India) bounded by Latitude (N) $22^0 00'00''$ to $24^0 00'0''$ and Longitude (E) $85^0 00'00''$ to $87^0 00'00''$. Jharia Coalfield is located about 250 km NW of Kolkata and about 1150 km SE of Delhi and is the richest coal bearing area in India. This area contains high grade coal and large number of mine fires which have been burning for several decades. These fires lead to degradation of land and do not allow any vegetation to grow in the area.

2.2 Satellite data

Satellite data used for hotspot detection was AVHRR onboard the NOAA satellite series level 1B. These images have convenient coverage of the entire study area. NOAA/AVHRR image comprises five spectral bands: visible (ch.1, 0.63 μm), near-infrared (ch.2, 0.83 μm), mid-infrared (ch.3, 3.75 μm), and thermal (ch.4-5, 10-12 μm). The ground resolution of AVHRR is 1.1 kilometer. For our study NOAA/AVHRR data were collected in morning from 7-9 am for the month of February, in between the period of year 1995-2005. The type of each data is AVHRR- Local Area Coverage (LAC).

2.3 Auxiliary data

The results of the proposed algorithm were validated and characterized by using hotspot information provided by BCCL (Bharat Coking Coal Limited), India. These reports are based on ground observations and include the place and location of the hotspots in Jharia region. Jharia coalfield stretches between latitudes $23^0 38'00''$N and $23^0 50'00''$N and longitudes $86^0 07'00''$E and $86^0 30'00''$E (Prakash and Gupta, 1999).

3. METHODOLOGY

Flow diagram of the proposed methodology to detect hotspots is shown in Figure 1. First of all preprocessing of raw NOAA/AVHRR data takes place to correct geometric distortions, calibrate the data radiometrically and eliminate the noise and clouds that are present in the data. After preprocessing raw AVHRR data, in the first step of the algorithm, PCA is employed on multi-channel information i.e. band 1 (R_1), band 2 (R_2), along with NDVI (Normalized Difference Vegetation Index) and MSAVI (Modified Soil Adjusted Vegetation Index) information to obtain principal components which are linear combination of the original multi-channel band images. After that in the second step, information from first two components is fused and based upon this fused information along with NDVI information, it is determined that whether pixels belong to hotspot class or not.

Vegetation indices play very important role in hotspot detection and thus can't be ignored. These indices are numerical indicators which indicate the amount of vegetation on given
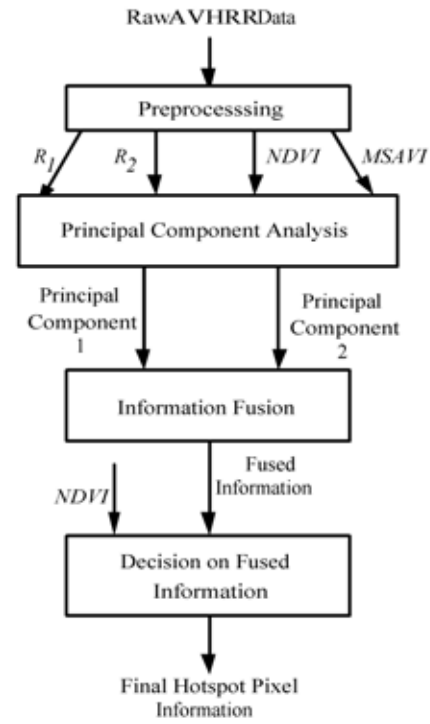

Figure 1. Flow diagram of the proposed algorithm.

area of interest. Higher the vegetation index value, higher will be the probability of healthy green vegetation on that area. In this paper information from two vegetation indices have been incorporated named as NDVI and MSAVI, along with multi-band information to detect hotspots in AVHRR images. These vegetation indices are defined as (Singh *et al.*, 2005)

$$NDVI = \frac{R_2 - R_1}{R_2 + R_1} \qquad (1)$$

$$MSAVI = \frac{((2R_2+1) - ((2R_2+1)^2 - 8(R_2 - R_1))^{0.5})}{2} \qquad (2)$$

where R_1 and R_2 are reflectances in channel 1 and 2 respectively.

Following section describes the methodology of principal component analysis of multi-channel information with some brief theory of it.

3.1 Principal component analysis (PCA)

PCA involves a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components. The principal component transform expresses the input digital numbers in the original bands in terms of the new principal component (PC) axes. In this technique the different spectral images are transformed into some few principal components contained almost the total variance of original images. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts

for as much of the remaining variability as possible. For an n-dimensional dataset, n principal components can be produced. In order to create the PC axes it is necessary to calculate the length of the PC axes and their direction. These are computed by determining the eigenvalues (length) and eigenvectors (direction) from the correlation matrix. The process of how correlation matrix is computed and how principal components are obtained from the correlation matrix is explained step by step as following:

Step 1: First symmetric correlation matrix is computed from 4-dimensional data i.e. band 1 (R_1), band 2 (R_2), *NDVI*, and *MSAVI*. In this correlation matrix all the columns are standardized.

Step 2: Now eigenvalues and eigenvectors are computed from the symmetric correlation matrix. These eigenvectors are orthogonal to each other, thus the data can be represented in terms of these perpendicular eigenvectors and more importantly, they provide with information about the patterns in the data.

Step 3: Once eigenvectors are found from the correlation matrix, next step is to order them by eigenvalues in descending order. This gives the components in order of significance. Here decision is made to ignore the components of lesser significance. If some components are left out, final data set will have less dimensions than the original.

Step 4: Now in the final step, a *feature vector* is formed by taking the chosen eigenvectors in the rows, with the most significant eigenvector in the top. This *feature vector* is then multiplied with the original data matrix and the *final data* is obtained. *Final data* gives the original data solely in terms of the chosen vectors.

Basically, the original data is transformed so that it can be expressed in terms of the patterns between them. PCA analysis
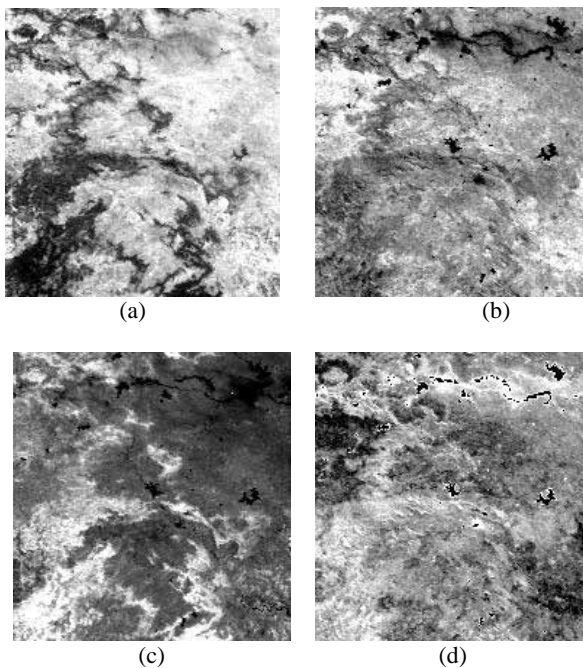

Figure 2. Original input images of (a) band 1, (b) band 2, (c) NDVI and (d) MSAVI.
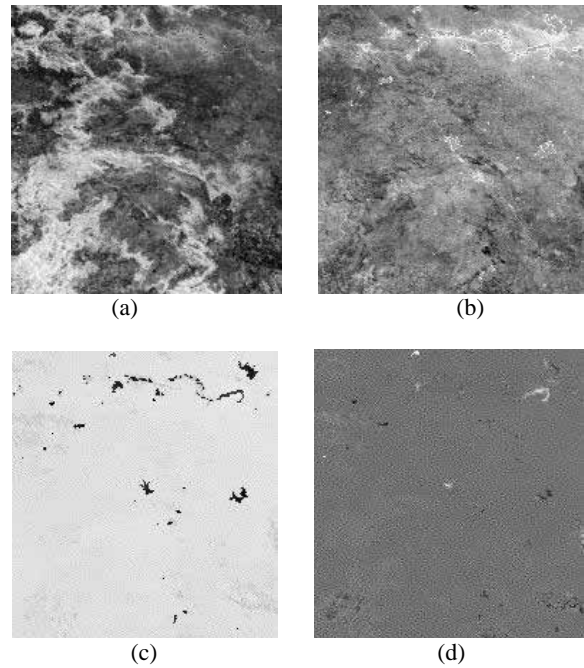

Figure 3. Principal component images corresponding to the images in Figure 2.

gives the original band images in terms of the differences and similarities between them and identifies statistical patterns in the data.

Figure 2 shows the four original input images corresponding to band 1, band 2, *NDVI*, and *MSAVI* of NOAA/AVHRR data of Jharia region respectively. Figure 3 shows the result of the principal component analysis in the form of four principal component images corresponding to the images in figure 1. It is clear that most of the contrast detail is contained in the first two component images, and it decreases rapidly from there. The reason of this can easily be explained by looking at the eigenvalues. Table 1 shows the 4 eigenvalues corresponding to the four principal component images shown in figure 2. First two eigenvalues are quite large in comparison to the others. Since eigenvalues are the variances and the variance is a measure of contrast, so it is not unexpected that the images which correspond to the higher eigenvalues will exhibit significantly higher contrast.

3.2 Information fusion

After applying PCA on 10 years of NOAA/AVHRR data, a thorough analysis was performed on principal component images of the Jharia region. Information contained in first two principal components was fused and the statistical analysis was carried out to calculate median and standard deviation for each component image of every AVHRR data. Based upon the observations on these statistical parameters following candidate

Eigenvalue 1	Eigenvalue 2	Eigenvalue 3	Eigenvalue 4
9018	3596	589	63

Table 1. Eigenvalues corresponding to the four principal component images shown in Figure 3.

range of the component values was decided for marking the pixels as hotspots.

$$PC1 = m_{PC1} \text{ to } m_{PC1} + std_{PC1} \qquad (3)$$

$$PC2 = m_{PC2} - n*std_{PC2} \text{ to } m_{PC2} - (n\text{-}3)*std_{PC2} \qquad (4)$$

Where m_{PC1} is median and std_{PC1} is standard deviation of PC1. Similarly m_{PC2} and std_{PC2} are median and standard deviation of PC2 respectively. n is a constant and n>3. Results were observed for different ranges of PC2 by changing the value of coefficient multiplied by std_{PC2} in upper limit of the candidate range specified in (4) and (n-3) was found to be the best option which could only define the actual hotspot range accurately. In addition, *NDVI* value was again considered to eliminate the possibility of detecting water pixels as hotspot pixels, as for water pixels *NDVI* is always negative.

4. RESULTS AND DISCUSSION

Proposed hotspot detection algorithm was applied to 10 years of NOAA/AVHRR images of Jharia region. The raw AVHRR data was first preprocessed and then PCA was employed to band 1, band 2, NDVI and MSAVI of these preprocessed AVHRR data in order to transform them in principal components as discussed in section 3. Since most of the information is contained in first two component images as they correspond to largest eigenvalues, therefore instead of considering all four components only first two component images were stored.

Once the first two principal component images were obtained, the information contained in first two principal components was fused and the statistical parameters in term of median and standard deviation were calculated for both of the principal components in order to determine hotspot pixels. Median is the middle of the distribution which is more informative and is less sensitive to extreme values that make median a better measure than mean. Standard deviation tells how tightly a set of values is clustered around the average (median) of those same values. It's a measure of dispersal, or variation, in a group of numbers.

Pixels falling in the candidate range (4) and (5) were marked as hotspot pixels. *NDVI* value was used to eliminate the possibility of detecting water pixels as potential hotspots pixels as *NDVI* is always negative for water pixels. Since, the limits used for the hotspot candidate ranges (4) and (5) make use of only statistical parameters and no absolute threshold has been incorporated for defining these ranges, the decision which is made to specify whether particular pixel belongs to hotspot class or not, is adaptive in nature and works successfully for most of the AVHRR images.

Figure 4(a) – 4(d) show the results obtained from the algorithm for the AVHRR data taken in year 1995, 1996, 2004, and 2005 respectively. Left column represents the original AVHRR images of Jharia region and right column represents the corresponding resulted images obtained by the application of proposed hotspot detection algorithm. Hotspots are shown as white pixels in the resulted images. In the proposed method, some false interpretations were found due to the spatial and temporal heterogeneity of the region's environments. To

4(a).1   4(a).2

4 (b).1   4(b).2
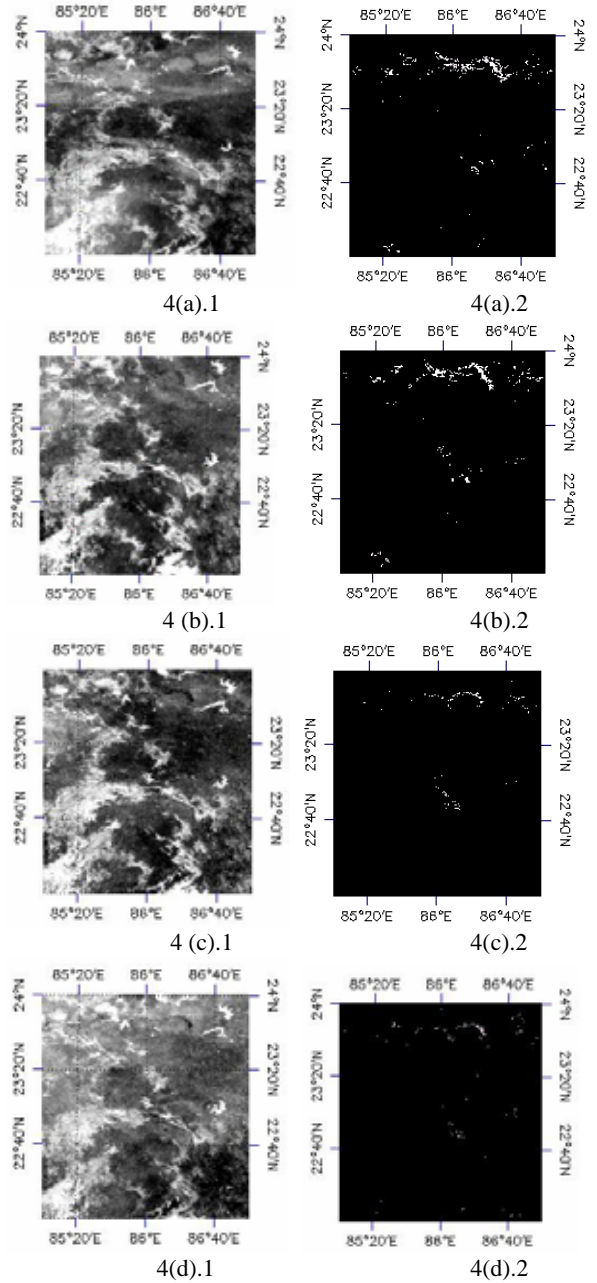
4 (c).1   4(c).2

4(d).1   4(d).2

Figure 4. Results of the proposed hotspot detection algorithm applied on NOAA/AVHRR data of Jharia region collected in year (a) 1995, (b) 1996, (c) 2004 and (d) 2005 respectively.

evaluate the performance of the proposed algorithm for hotspot detection, two metrics were computed defined as

$$\text{Detection accuracy} = \frac{\text{correctly reported hotspots}}{\text{total hotspots that exist}} \qquad (5)$$

$$\text{False Alarm Rate} = \frac{\text{Incorrectly detected hotspots}}{\text{Total no. of pixels} - \text{Total hotspots that exist}}$$

$$(6)$$

S.N.	Year	No. of hotspots detected successfully	Detection Accuracy (%)	False Alarm Rate (%)
1	1995	20	80	1.109673
2	1995	21	84	1.214899
3	1996	22	88	0.92794
4	1996	22	88	1.099888
5	1997	23	92	0.633216
6	1998	23	92	0.517494
7	1999	19	76	0.207653
8	2003	19	76	0.078924
9	2003	19	76	0.053437
10	2004	24	96	0.240633
11	2004	22	88	0.158686
12	2004	20	80	0.095113
13	2005	23	92	0.361352
14	2005	21	84	0.086505
15	2005	21	84	0.297419

Table 2. Performance of the proposed hotspot detection algorithm evaluated for AVHRR images of Jharia region for the period of year 1995-2005 (Total no. of existing hotspots are 25).

Table 2 shows the performance of our algorithm applied for the AVHRR images for the period from 1995 to 2000. It is seen that for all images, algorithm has detected at least 19 hotspots correctly out of 25 existing hotspots. So detection accuracy achieved by proposed algorithm on all images is consistently higher than 76% and maximum detection accuracy achieved is 96%. Similarly false alarm rate was always less than 1.3% which is not so much significant. This performance shows that the proposed simple algorithm detects the hotspots accurately as well as efficiently.
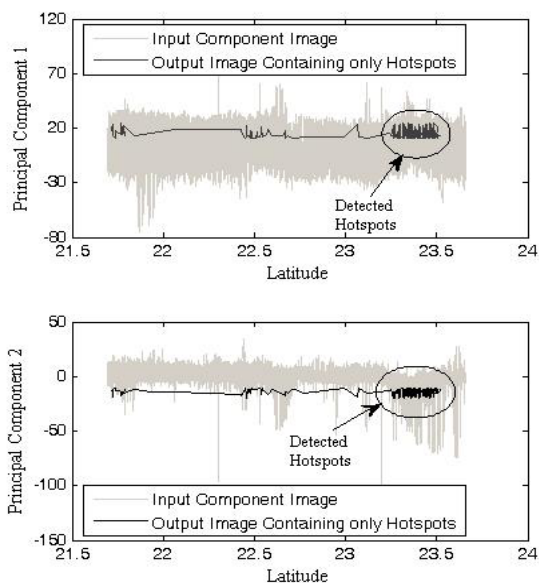

Figure 5. Plot of input principal component 1 and 2 and corresponding output image containing only hotspots, against latitude information.

Figure 5 shows the plot of principal component 1 and 2 values against latitude information for both input image and corresponding output images obtained after applying the proposed hotspot detection algorithm. In this figure hotspot pixels can easily be distinguished from rest of the pixels and are clustered in the ellipse.

5. CONCLUSION

An efficient and innovative technique was proposed for detecting hotspot pixels in NOAA/AVHRR images of Jharia Coalfield (India). The proposed algorithm was applied to 10 years of NOAA/AVHRR images of region of interest and the algorithm was found to detect hotspots successfully. The NOAA/AVHRR data has been used because of its good temporal resolutions and free image acquisition. PCA analysis was carried out on multi-channel information of AVHRR images and then information obtained from first two principal components was fused to make decision on hotspot pixels. In most of the AVHRR images the proposed algorithm detected the hotspots successfully. The detection accuracy achieved by the proposed algorithm was consistently higher than 76% and maximum detection accuracy achieved was 96%.

REFERENCES

Agassi, E., and Yosef, N. B., 1998. The effect of the thermal infrared data on principal component analysis of multi-spectral remotely-sensed data. *International Journal of Remote Sensing*, vol. 19, no. 9, pp. 1683-1694.

Arino, O., and Rosaz, 1999. 1997 world ATSR fire atlas. *Proceedings of Conference on Remote Sensing and Forest Monitoring*, Warsaw Agriculture University, pp. 606-615.

Bardossy, A., and Samaniego, L., 2002. Fuzzy rule-based classification of remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 2, pp. 362-374.

Barsnley, M. J., Allison, D., and Lewis, P., 1997. On the information content of multiple view angle (MVA) images. *Intermational Journal of Remote Sensing*, vol. 18, no. 9, pp. 1937-1960.

Boles, S. H., and Verbyla, D. L, 2000. Comparison of three AVHRR based fire detection algorithms for interior Alaska. *Remote Sensing of Environment*, vol. 72, pp. 1-16.

Brustet, J. M., Vickos, J. B., Fontan, J., Podaire, A., and Lavenu, F., 1991. Characterization of active fires in west Africa savannas by analysis of satellite data: Landsat Thematic Mapper. *Global Biomass Burning*, edited by Levine, J. S., MIT Press, Cambridge, Mass., pp. 53-60.

Flannigan, M. D., and Haar, T. H. V., 1986. Forest fire monitoring using NOAA satellite AVHRR. *Canadian Journal of Forest Research*, vol. 16, pp. 975-982.

Flasse, S. P., and Ceccato, P., 1996. A contextual algorithm for AVHRR fire detection. *International Journal of Remote Sensing*, vol. 17, pp. 419-424.

Fuller, D. O., and Fulk, M., 2000. Burned area in Kalimantan, Indonesia mapped with NOAA-AVHRR and Landsat TM

imagery. *International Journal of Remote Sensing*, vol. 21, pp. 181-187.

Kaufman, Y. J., Tucker, C. J., and Fung, I., 1990. Remote sensing of biomass burning in the tropics. *Journal of Geophysical Research*, vol. 95, pp. 9,927-9,939.

Kennedy, P. J., Belward, A. S., and Gregoire, J.-M., 1994. An improved approach to fire monitoring in West Africa using AVHRR data. *International Journal of Remote Sensing*, vol. 15, pp. 2235-2255.

Li, R.-R., Kaufman, Y. J., Hao, W. M., Salmon, J. M., and Gao, B.-C., 2004. A technique for detecting burn scares using MODIS data. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 1300-1308.

Li, Z., Khananian, A., Fraser, R. H., and Cihlar, J., 2001. Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to AVHRR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 9, pp. 1859-1870.

Moore, T. S., Campbell, J. W., and Feng, H., 2001. A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 8, pp. 1764-1776.

Nakayama, M., Maki, M., Elvidge, C. D., and Liew, S. C., 1999. Contextual algorithm adapted for NOAA-AVHRR fire detection in Indonesia. *International Journal of Remote Sensing*, vol. 20, pp. 3415-3421.

Prakash, A., and Gupta, R. P., 1999. Surface fires in Jharia coalfield, India – their distribution and estimation of area and temperature from TM data. *International Journal of Remote Sensing*, vol. 20, no. 10, pp. 1935-1946.

Prins, E. M., and Menzel, W. P., 1994. Trends in south American biomass burning detected with the GOES VISSR radiometer atmospheric sounder from 1983 to 1991. *Journal of Geophysical Research*, vol. 99, pp. 16,719-16,735.

Singh, D., Costa, G. A., Meirelles, M. S., Herlin, I., Berroir, J. P., and Silva, E. F., 2005. A methodology to support environmental degradation monitoring and analysis using AVHRR data. Anais XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, Brasil, INPE, pp. 2941-2948.