MODIS Data Mining to Map Burned Areas

C. QUINTANO[a,1], A. STEIN[b], W. BIJKER[b], and A. FERNÁNDEZ-MANSO[c]
[a] *University of Valladolid, Electronic Technology Department, Spain*
[b] *International Institute for Geo-Information Science and Earth Observation (ITC), Earth Observation Science Department, The Netherlands*
[c] *University of León, Agrarian Engineering Department, Spain*

Abstract. Estimation of areas burned by forest fires in Mediterranean countries is achieved by applying an image mining method to MODerate-resolution Imaging Spectroradiometer (MODIS) satellite data. The proposed image mining based method needs a post-fire image as input and involves three steps: modelling as a sum of Gaussian functions, kernel based smoothing, and thresholding. The optimal stopping rule of the modelling iterative algorithm, the optimal smoothing parameter, and the optimal threshold have been identified by the κ statistic. This statistic measures the accuracy of the burned areas estimation by relating the estimation area with the burned area perimeters measured by Global Positioning System (GPS). Two indexes, specifically designed for burned area identification and adequate to distinguish burned areas have been used: the Burned Area Index (BAI) and the Normalized Burn Ratio (NBR). A Z-test allowed us to compare the accuracy of these estimates with the accuracy obtained by classifying directly the input images. Results show that image mining based methods allowed a higher accuracy of the burned areas estimation. We conclude that MODIS data mining is valid to accurately map burned areas.

Keywords. Burned areas, MODIS, image mining

Introduction

The use of remote sensing data and methods for mapping burnt areas has a long history in the European Mediterranean region [1]-[4]. It is able to provide temporal and spatial coverage of damage assessment without costly and intense fieldwork. The resulting information is suitable for integration into a Geographic Information System (GIS) that allows the storage and processing of large volumes of spatial data.

Data from the National Aeronautics and Space Administratiuon (NASA) MODerate resolution Imaging Spectrometer (MODIS) has been broadly used to map burnt areas [4]-[9]. MODIS is a sensor onboard of Terra (morning) and Acqua (afternoon) satellites with a daily coverage and more than 30 narrow bands at wavelengths from the visible to the thermal infrared and at a variable spatial resolution (250 m-1000 m).

Concerning the remote sensing methods dealing with fire scars mapping from reflectance satellite data, three groups can be distinguished: 1) single image analysis (unitemporal perspective), in which one image acquired after the fire event is available; 2) change detection techniques (multitemporal perspective), in which pre-fire and post-fire conditions are combined to discriminate changes; and 3) time series analysis, in which a fairly complete set of pre-fire and post-fire scenes are used. Single image analysis is commonly used for the processing of medium to high or very high spatial resolution satellite imagery, and any of the classification methods of remotely sensed data can be used for mapping burnt areas [10].

1 Corresponding Author: University of Valladolid, Electronic Technology Department, Francisco Mendizabal, 1, 47014-Valladolid, Spain; E-mail: menchu@tele.uva.es

Use of spectral indices is widespread among different inputs to the classifier algorithm. Vegetation indices such as the Normalized Difference Vegetation Index (NDVI) whose estimation typically involves data form red and near-infrared (NIR) bands have been commonly used to derive vegetation properties but also to discriminate and map burned areas [2],[11]-[13]. According to [14], burned vegetation results in a drastic reduction in NIR reflectance and an increase in shortwave infrared (SWIR) reflectance in most environment and fire regimes. Spectral indices that integrate data from NIR and SWIR bands and are specifically designed for burned land discrimination have been recently developed and tested. Novel spectral indices include the Burned Area Index for MODIS data (BAIM) [15] and the Normalized Burn Ratio (NBR) [16]-[17]. Other studies have investigated the utility of principal components analysis [18], texture analysis [19] and spectral mixture analysis [20].

Image mining, and more specifically spatial image mining, is a relatively new field of study that uncovers patterns or statistically significant structures in spatial data [21]. It allows for detecting and mapping features that show a spatial pattern. Although there are almost no references relating spatial image mining to burned areas by forest fires; in [22] the authors showed that the proposed image mining algorithm successfully obtained accurate estimations of burned areas from satellite imagery. Spatial mining algorithm comprises three steps: 1) modelling, expressed by applying Gaussian distributions on individual grid cells with deviating values; 2) pre-classification consisting of applying a smoothing kernel; and 3) thresholding with the objective to classify the image into burned and unburned classes.

In [22], we showed that image mining based on modelling and smoothing, produces more accurate burned area estimations than conventional methods based on thresholding an NDVI image. The accuracy level was measured by means of κ statistic. The objective of this research is to validate the spatial image mining method to map burned areas from the BAIM and NBR indexes derived from MODIS that are specifically designed to do so.

1. Study area and dataset

The study area is the Barbanza county in Galicia (Northwest Spain) with an area of 29 000 km^2). The Galician landscape mosaic is heterogeneous due to the high degree of fragmentation caused by human activity (64% of the total area is forestry land). Most fires occur between June and September, being the dry summer period. During the fire season in 2006 in Galicia, 939 km^2 (540 km^2 forested) was burned. A total of 1651 fires occurred from 4 to 15 August, mainly caused intentionally and favored by strong winds and a dry, hot summer [23].

Two types of data were used: 1) the Global Positioning System (GPS) based perimeter of burned areas in the Barbanza county; and 2) MODIS surface reflectance data. Both data suffered some preprocessing. The contiguous burned areas on the GPS-based burned area perimeter vector file were grouped and only burned areas larger than 0.5 km^2 were considered. The preprocessed GPS-based vector file consisted of 15 burned areas, and the total considered burned area was 150 km^2.

Next, the MYD13Q1 (vegetation indexes, 16-day, 250 m) MODIS product from 21 August to 5 September was windowed to the study area, re-projected to the Universal Transverse Mercator (UTM) system, WGS84-29n, and co-registered with the preprocessed GPS based burned area perimeter. This product has a spatial resolution of 250 m and includes four reflectance bands (blue, red, NIR, SWIR) and two vegetation indexes (NDVI and Enhanced Vegetation Index–EVI-).

2. Method

The proposed image mining algorithm consists of three independent modules: modelling (M), smoothing (S), and thresholding (T) that are explained in detail in the next subsections. To obtain a burnt area estimation from the input image, four alternatives have been considered:

1) Thresholding the input image (T). This estimation method does not include an image mining based method. Therefore, it serves as a reference to be compared to other burnt area estimations identified below.
2) Thresholding the modelled input image (M+T).
3) Smoothing the image, followed by thresholding the smoothed input (S+T).
4) Modelling the input, followed by smoothing and thresholding (M+S+T).

2.1. Calculation of BAIM and NBR

The BAIM index is estimated using the Equation (1):

$$BAIM = \frac{1}{(\rho_{CNIR} - \rho_{NIR})^2 + (\rho_{CSWIR} - \rho_{SWIR})^2} \tag{1}$$

where ρ_{CNIR} and ρ_{CSWIR} are the NIR and SWIR reference reflectance values, respectively, and ρ_{NIR} and ρ_{SWIR} are the pixel reflectance values in the same bands. Following [15], the reference reflectance values were: ρ_{CNIR} = 5 percentile of burned NIR distribution and ρ_{CSWIR} = 95 percentile of burned SWIR distribution.

The NBR index is obtained by applying the Equation 2 to the preprocessed MODIS image:

$$NBR = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}} \tag{2}$$

where ρ_{NIR} and ρ_{SWIR} are the NIR and SWIR reflectance values, respectively.

2.2. Modelling module (M)

The modelling module (M) models the input image (I) as a sum of *N* weighted Gaussian functions (Equation 3), each of them representing a small part of the burned area. Their characteristics, number and weight are determined by using a image mining iterative algorithm based on the two coordinates x_{i1} and x_{i2} of the i^{th} observation location in two perpendicular directions, following [24].

$$I(x_1, x_2) \cong \sum_{i=1}^{N} w_i f_i(x_{i1}, x_{i2}) \tag{3}$$

2.3. Smoothing module (S)

A kernel smoother uses weights that smoothly decrease when moving away from the pixel under consideration. The Gaussian kernel function used has a weight function based on the Gaussian density function (Equation 4) and assigns weights to points that decrease exponentially with the squared Euclidian distance from the pixel under consideration, χ_0:

$$K_\lambda(x_0, x) = \frac{1}{\lambda} \exp\left(-\frac{\|x - x_0\|^2}{2\lambda} \right) \tag{4}$$

The parameter λ corresponds to the variance of the Gaussian density, and controls the width of the neighbourhood. A large value of λ implies a low variance as the kernel averages more observations, but it also implies a larger bias, whereas a small value of λ implies a relatively large variance and a small bias. Its optimal value was defined experimentally at the accuracy assessment stage.

2.4. Thresholding module (T)

The classification of the input image into "burnt" and "unburnt" allow obtaining a burnt area estimation. In this study, a pixel of the input image will be assigned to the "burnt" class if its value is higher than a pre-defined threshold. Following [2], instead of absolutes values, relative thresholds based on the mean value of the image to be thresholded and on its standard deviation have been applied. As happened before, the optimal threshold was defined at the accuracy assessment stage.

2.5. Accuracy assessment

As recommended by [25], confusion matrices and the κ statistic were used to measure the accuracy of the achieved burned area estimations. In addition the overall accuracy (OA), producer's accuracy (PA) (omission error) and user's accuracy (UA) (commission error) were considered as well. As the ground truth image, a rasterized map resulting from the preprocessed GPS-based vector file was used. A stratified random disproportional sampling strategy was employed. All pixels corresponding to the burned areas in the ground truth image were considered and the 50% of the pixels corresponding to the unburned areas were randomly selected.

In addition, we used a leave-one-out cross-validation strategy [26]. In our study, this strategy meant that of the 15 burnt areas, 14 were used as training data to identify the optimal smoothing parameter and the optimal threshold that maximize the κ statistic, and one was retained as a validation area. This process was repeated 15 times, with each of the 15 burned areas used exactly once as the validation area. Finally, the optimal pair of smoothing parameters and thresholds were defined as the mode of the optimal values obtained for each of the validated burned areas (κ statistic higher or equal to 0.6); and the definitive estimation of the whole burned area in the Barbanza district was calculated using this optimal pair of values [22].

A two-sided *Z* test was applied to the individual κ statistics associated to these definitive estimations to know if there are significant differences between two estimations. Note that $z_c = 1.96$ at the 95% confidence level, and that the null hypothesis H_0: $(\kappa_1-\kappa_2) = 0$ is rejected when $Z > z_c$.

3. Results and discussion

Table 1 displays the highest κ statistics obtained by applying each alternative to the original NIR band and to the spectral indexes. Note that the κ statistics obtained when (S+T) and (M+S+T) alternatives were applied are almost equal ((S+T) lightly superior) and higher than the κ statistics obtained by using (T) or (M+T) alternatives. Regarding the inputs, NIR band, EVI and BAIM resulted in more accurate estimations, whereas NDVI and NBR showed the lowest accuracy. The smoothing parameters and threshold value used to obtain the more accurate burned area estimations for each input are displayed in Table 2 (OA, PA and UA are also presented).

Table 1. The highest κ statistics obtained by applying each alternative to each input (bold values show the highest κ statistic obtained for each input).

Input	Alternative			
	(T)	(M+T)	(S+T)	(M+S+T)
NIR	0.66	0.64	**0.69**	0.68
NDVI	0.57	0.40	**0.61**	0.60
EVI	0.67	0.64	**0.68**	0.67
BAIM	0.64	0.66	**0.67**	0.67
NBR	0.57	0.51	**0.61**	0.58

The Z statistics obtained when comparing the burned area estimations achieved applying the different alternatives to each input are shown in table 3. For the alternatives (T) and (S+T) (the ref-

erence alternative and the alternative with the highest κ statistics, respectively), significant differences were observed when NDVI or NBR were used as inputs, with the κ statistic being equal to 0.61 in both cases. Noticeably no significant differences were observed when BAIM was used as input. All the inputs except BAIM showed significant differences between (M+T) and (S+T).

Table 2. Information about the highest accurate burned area estimation (μ: mean, σ: standard deviation).

Input	Optimal alternative	Optimal λ	Optimal threshold	K statistic	σk	OA	PA	UA
NIR	(S+T)	1.50	μ+ 0.25*σ	0.69	0.0105	0.84	0.84	0.84
NDVI	(S+T)	1.75	μ+ 0.25*σ	0.61	0.0115	0.80	0.80	0.80
EVI	(S+T)	0.75	μ+ 0.25*σ	0.68	0.0107	0.84	0.84	0.84
BAIM	(S+T)	1.25	μ+ 0.25*σ	0.67	0.0109	0.84	0.84	0.84
NBR	(S+T)	2	μ+ 0.25*σ	0.61	0.0117	0.81	0.80	0.81

Table 3. Z statistics (bold values show Z > 1.96) comparing the four alternatives: (T): thresholding, (M+T): modeling and thresholding, (S+T): smoothing and thresholding, and (M+S+T): modeling, smoothing and thresholding.

Alternative 1	Alternative 2	Input				
		NIR	NDVI	EVI	BAIM	NBR
(T)	(M+T)	1.02	**9.51**	1.60	1.21	**3.43**
(T)	(S+T)	1.91	**2.00**	1.19	1.70	**2.27**
(T)	(M+S+T)	1.56	1.65	0.12	1.12	0.48
(M+T)	(S+T)	**2.93**	**11.51**	**2.80**	0.48	**5.75**
(M+T)	(M+S+T)	**2.58**	**11.14**	1.73	0.08	**3.92**
(S+T)	(M+S+T)	0.36	0.34	1.07	0.57	1.80

The simultaneous analysis of Tables 1 and 3 allows identifying the combination of input and alternative with the highest κ statistic that showed significant differences compared to the reference alternative (T). Such combination is obtained when using NDVI or NBR as inputs of the (S+T) alternative (k = 0.61). Visualizing both burnt area estimations (see Figure 2) the estimation based on NDVI is preferred because the NBR-based estimation presents an excessive smoothing.

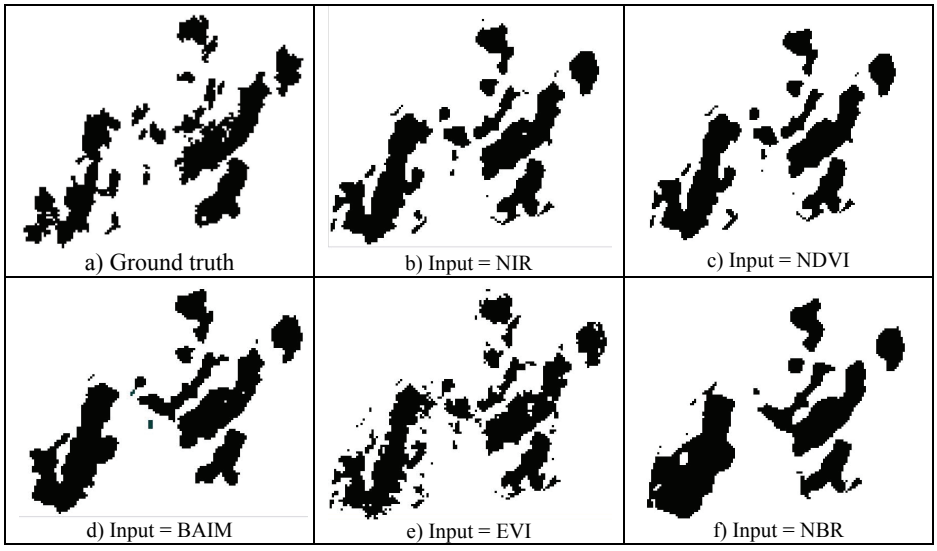

Figure 2. Burnt area estimations based on thresholding the smoothed input (S+T alternative with the parameters showed in Table 2).

Regarding the alternative that allowed more accurate estimations our result is in accordance with other authors (see [22] and [27]) that stated that smoothing used as a pre-classifier leads for higher accurate burned area estimations. Concerning the inputs used NBR stands as the more adequate index to estimate burnt areas, which is in accordance with several studies (e.g. [28]-[29]) that show how NBR highlights the severity of burned areas. NDVI performance is, however, unusual.

.

4. Conclusions

This study assesses the utility of image mining algorithms to map burned areas in Mediterranean countries from MODIS data. Specifically the smoothing algorithm has been used as a preclassifier. The results show that image mining algorithms allow burnt areas estimation with higher accuracy than conventional methods such as thresholding.

The burnt area estimation obtained by using the NBR index as input to the (S+T) alternative was more accurate than the burnt areas estimation obtained by thresholding the input image (T). In addition, the Z test of the κ statistics showed significant differences between this image mining algorithm and the reference procedure (just thresholding).

We conclude that MODIS data mining is a valid methodology to accurately map burned areas.

Acknowledgments

The first and last authors acknowledge the financial support of the Ministry of Education of the Castilla y León regional government (Research project: VA098A08).

References

[1] E. Chuvieco, R.G. Congalton, Application of remote sensing and geographic information systems to forest fire hazard mapping, *Remote Sensing of Environment* **29** (1989) 147-159.

[2] A. Fernández, P. Illera, J.L. Casanova, Automatic mapping of surfaces affected by forest fires in Spain using AVHRR NDVI composite image data, *Remote Sensing of Environment* **60** (1997) 153-162

[3] A. DeSantis, E. Chuvieco, P.J. Vaughan, Short-term assessment of burn severity using the inversion of PROSPECT and GeoSail models, *Remote Sensing of Environment* **113** (2009) 126–136.

[4] S. Merino-de-Miguel, M. Huesca, F. González-Alonso, MODIS reflectance and active fire data for burn mapping and assessment at regional level, *Ecological Modelling* **221** (2010) 67–74

[5] E. A.C.L. Sá, J.M.C. Pereira, M.J.P. Vasconcelos, J.M.N. Silva, N. Ribeiro; A. Awasse, Assessing the feasibility of sub-pixel burned area mapping in miombo woodlands of northern Mozambique using MODIS imagery, *International Journal of Remote Sensing* **24** (2003) 1783 – 1796.

[6] Chuvieco, S. Opazo, W. Sione, H. DelValle, J. Anaya, C. DiBella, I. Cruz, L. Manzo, G. López, N. Mari, F. González-Alonso, F. Morelli, A. Setzer, I. Csiszar, J. Ander-Kanpandegi, A. Bastarrika, R. Libonati, Global burned-land estimation in Latin America using MODIS composite data, *Ecological Applications*, **18** (2008) 64–79.

[7] X. Cao, J. Chen, B. Matsushita, H. Imura; L. Wang, An automatic method for burn scar mapping using support vector machines, *International Journal of Remote Sensing* **30** (2009) 577 – 594.

[8] L. Boschetti, D. Roy, P. Barbosa, R. Boca; C. Justice, A MODIS assessment of the summer 2007 extent burned in Greece, *International Journal of Remote Sensing* **29** (2008) 2433 – 2436.

[9] C.O. Justice, L. Giglio, S. Korontzi, J. Owens, J.T. Morisette, D. Roy, J. Descloitres, S. Alleaume, F. Petitcolin, and Y. Kaufman, The MODIS fire products, *Remote Sensing of Environment* **83** (2002) 244–262.

[10] J. San-Miguel- Ayanz, J.M.C. Pereira, R. Boca, P. Strobl, J. Kucera and A. Pekkarinen, Forest fires in the European Mediterranean region: mapping and analysis of burned areas, in E. Chuvieco (ed.) *Earth Observation of wildland fires in Mediterranean Ecosystems (189-204),* Springer-Verlang, Berlin Heidelberg, 2009.

[11] J.M.C, Pereira, A comparative evaluation of NOAA/AVHRR vegetation indexes for burned surface detection and mapping, *IEEE Transactions on Geoscience and Remote Sensing* **37** (1999) 217-226.

[12] E. Chuvieco, M.P. Martín, A. Palacios, Assessment of different spectral indices in the red–near-infrared spectral domain for burned land discrimination, *International Journal of Remote Sensing* **23** (2002) 5103–5110.

[13] C. Domenikiotis, N.R. Dalezios, A. Loukas, M. Karteris, Agreement assessment of NOAA/AVHRR NDVI with Landsat TM NDVI for mapping burned forested areas, *International Journal of Remote Sensing* **23** (2002) 4235-4246.

[14] L.B. Lentile, Z.A. Holden, A.M.S. Smith, M.J. Falkowski, A.T. Hudak, P. Morgan, S.A. Lewis, P.E. Gessler, N.C. Benson, Remote sensing techniques to assess active fire characteristics and post-fire effects. *International Journal of Wildland Fire* **15** (2006) 319–345.

[15] P. Martín, I. Gómez, E. Chuvieco, Performance of a burned –area index (BAIM) for mapping Mediterranean burned scars from MODIS data, *Proceedings of the 5th International Workshop on Remote Sensing and GIS Applications to forest fire management: fire effects assessment*, Zaragoza –Spain- (2005) 193-197.

[16] C.H. Key, N.C. Benson, The Normalized Burn Ratio (NBR): A Landsat TM Radiometric Measure of Burn severity. USDA (Bozeman, Mont), (1999) (available at http://nrmsc.usgs.gov/research/ndbr.htm).

[17] D.P. Roy, L. Boschetti, S.N. Trigg, Remote sensing of fire severity: assessing the performance of the Normalized Burn Ratio, *IEEE Transactions on Geoscience and Remote Sensing Letters* **3** (2006) 112-116.

[18] F.J. García-Haro, M.A Gilabert, J. Meliá, Monitoring fire-affected areas using Thematic Mapper data. *International Journal of Remote Sensing* **22** (2001) 533–549.

[19] A.M.S. Smith, M.J. Wooster, A.K. Powell, D. Usher, Texture-based feature extraction: application to burn scar detection in Earth Observation satellite imagery. *International Journal of Remote* **23** (2002) 1733–1739.

[20] C. Quintano, A. Fernández-Manso, O. Fernández-Manso, Y.E Shimabukuro, Mapping burned areas in Mediterranean countries using Spectral Mixture Analysis from a unitemporal perspective. *International Journal of Remote Sensing* **27** (2006) 645–662.

[21] J. Shawe.Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, Cambridge, 2004.

[22] C. Quintano, A. Stein, W. Bijker, A. Fernández-Manso, Pattern validation for data mining of burned area objects from MODIS satellite images. *International Journal of Remote Sensing* (in press)

[23] Xunta de Galicia, Informe sobre a vaga de incendios forestais do mes de Agosto de 2006, Consellería do medio rural -Dirección Xeral do Montes e Industrias Forestais-, Galicia –Spain-, 2006.

[24] U. Rajasekar, W. Bijker, A. Stein, A., Image mining for modelling of forest fires from Meteosat images. *IEEE Transactions on Geoscience and Remote Sensing* **45** (2007) 246-253.

[25] R.G. Congalton, K. Green, *Assessing the Accuracy of Remotely Sensed Data. Principles and Practices*, CRC Press, Taylor & Francis Group, Boca Raton, 2009

[26] C. Goute, Note on free lunches and cross-validation, *Neural Computation* **9** (1997) 1245-1249.

[27] C. Quintano, E. Cuesta, Improving satellite image classification by using fractional type convolution filtering, *International Journal of Applied Earth Observation and Geoinformation*, (in press).

[28] A.E. Cocke, P.Z. Fule, J.E. Crouse, Comparison of burn severity assessments using Differenced Normalized Burn Ratio and ground data, *International Journal of Wildland Fire* **14** (2005) 189–198.

[29] D.R. Roy, L. Boscheti, S.N. Trigg, Remote sensing of fire severity: assessing the performance of the normalized burn ratio*, IEEE Geoscience and Remote Sensing Letters* **3** (2006) 112–116.