Spectral mixture analysis to assess post-fire vegetation regeneration using Landsat Thematic Mapper imagery: accounting for soil brightness variation

S. VERAVERBEKE*1,2, B. SOMERS3, I. GITAS4, T. KATAGIS4, A. POLYCHRONAKI4
AND R. GOOSSENS1

1 Department of Geography, Ghent University, Krijgslaan 281 S8, BE-9000 Ghent, Belgium, rudi.goossens@ugent.be
2 Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, US-91109 Pasadena, USA, Sander.S.Veraverbeke@jpl.nasa.gov
3 Centre for Remote Sensing and Earth Observation Processes (TAP), Flemish Institute for Technological Research (VITO), Boeretang 200, BE-2400 Mol, Belgium, ben.somers@vito.be
4 Laboratory of Forest Management and Remote Sensing, Aristotle University of Thessaloniki, PO BOX 248, GR-54124, Greece, igitas@for.auth.gr thkatag@for.auth.gr anpolych@for.auth.gr

*Corresponding author, phone 18183540278, fax 18183545148

Abstract
Post-fire vegetation cover is a crucial parameter in rangeland management. This study aims to assess the post-fire vegetation recovery three years after the large 2007 Peloponnese (Greece) wildfires. Post-fire recovery landscapes typically are mixed vegetation-substrate environments which makes Spectral Mixture Analysis (SMA) a very effective tool to derive fractional vegetation cover maps. Using a combination of field and simulation techniques this study aimed to account for the impact of background brightness variability on SMA model performance. The field data consisted out of a spectral library of in situ measured reflectance signals of vegetation and substrate and 78 line transect plots. In addition, a Landsat Thematic Mapper (TM) scene was employed in the study. A simple SMA, in which each constituting terrain feature is represented by its mean spectral signature, a multiple endmember SMA (MESMA) and a segmented SMA, which accounts for soil brightness variations by forcing the substrate endmember choice based on ancillary data (lithological map), were applied. In the study area two main spectrally different lithological units were present: relatively bright limestone and relatively dark flysch (sand-siltstone). Although the simple SMA model resulted in reasonable regression fits for the flysch and limestones subsets separately (coefficient of determination $R^2$ of respectively 0.67 and 0.72 between field and TM data), the performance of the regression model on the pooled dataset was considerably weaker ($R^2 = 0.65$). Moreover, the regression lines significantly diverged among the different subsets leading to systematic over-or underestimations of the vegetative fraction depending on the substrate type. MESMA did not solve the endmember variability issue. The MESMA model did not manage to select the proper substrate spectrum on a reliable basis due to the lack of shape differences between the flysch and limestone spectra,. The segmented SMA model which accounts for soil brightness variations minimized the variability problems. Compared to the simple SMA and MESMA models, the segmented SMA resulted in a higher overall correlation ($R^2 = 0.70$), its regression slope and intercept were more similar among the
different substrate types and its resulting regression lines more closely resembled the expected one-one line. This paper demonstrates the improvement of a segmented approach in accounting for soil brightness variations in estimating vegetative cover using SMA. However, further research is required to evaluate the model's performance for other soil types, with other image data and at different post-fire timings.

**Keywords:** fire; vegetation recovery; Landsat Thematic Mapper; Spectral Mixture Analysis; MESMA; segmentation

### 1 Introduction

Wildfires are a determining disturbance in almost all terrestrial ecosystems (Dwyer et al., 1999; Bond and Keeley, 2005; Riaño et al., 2007). They partially or completely consume the protective vegetation and organic litter cover, which can destabilize surface soils on steep slopes (Shakesby and Doerr, 2006). Shortly after the fire, infiltration significantly decreases whereas surface erosion increases due the bares soil's elevated exposure to raindrop impact and surface run-off. What is more, biomass burning instigates abrupt changes in ecological processes and carbon fluxes (Epting and Verbyla, 2005; Amiro et al., 2006). After the fire event a more gradual regeneration process is generally initiated (Viedma et al., 1997; van Leeuwen, 2008). Post-fire recovery rates depend on fire severity (Díaz-Delgado et al., 2003), soil properties (Bisson et al., 2008), post-fire meteorological conditions (Henry and Hope, 1998; van Leeuwen et al., 2010) and ecotype (Viedma et al., 1998; Veraverbeke 2010a, 2011, Lhermitte et al. 2011). In fire-adapted sclerophyllous shrub lands, for example, recovery only takes a few years (Viedma et al., 1997; Pausas and Verdu, 2005) whereas in boreal forests recovery lasts several decades (Nepstad et al., 1999). The carbon sequestration by regenerating plants partly compensates the fire's emissions and thus importantly influences the net changes caused by fire (Amiro et al., 2006; Randerson et al., 2006). Vegetation recovery is thus the main factor in limiting the damage of fire and its consequences. The assessment of
post-fire vegetation regeneration is of crucial importance for the understanding of the environmental impacts of fire and to support sustainable rangeland management after fire. In comparison with labor-intensive field work, the synoptic nature of remote sensing systems offers a time-and cost-effective means to fulfill this duty.

The remote sensing of post-fire vegetation recovery has a long tradition in the use of the Normalized Difference Vegetation Index (NDVI) (a.o. Viedma et al., 1997; Díaz-Delgado et al., 2003; van Leeuwen 2008; Clemente et al., 2009; Lhermitte et al., 2011) because of the well established relationship between the index and above-ground biomass in a wide range of ecosystems (Carlson and Ripley, 1997; Henry and Hope, 1998; Cuevas-González et al., 2009). At moderate resolution scale Landsat data typically are the standard of choice. A plethora of studies demonstrated the utility of Landsat NDVI to assess post-fire vegetation dynamics (a.o. Viedma et al., 1997; Díaz-Delgado et al., 2003; McMichael et al., 2004; Malak and Pausas, 2006; Clemente et al., 2009). These studies were restricted to a limited number of images. Some other studies, however, used low resolution time series to monitor regeneration processes. Cuevas-González et al. (2009), for example, monitored post-fire forest recovery in Siberia using Moderate Resolution Imaging Spectroradiometer (MODIS)-derived NDVI data, while van Leeuwen et al. (2010) conducted a similar study in three different study areas (Spain, Israel and USA). At the expense of spatial detail, these studies offer the advantage of image acquisition with high temporal frequency (van Leeuwen et al., 2010; Veraverbeke et al., 2011). Including the temporal dimension, however, often hampers the differentiation between post-fire effects and seasonal dynamics (Veraverbeke et al., 2010a, Lhermitte et al., 2011).

The post-fire environment typically consists of a mixture of vegetation and substrate. Thus, monitoring post-fire regeneration processes essentially poses a sub-pixel issue at the resolution of most operational satellite systems such as Landsat. A number of image analysis
techniques accommodating mixing problems exist (Atkinson et al., 1997; Arai, 2008) with
Spectral Mixture Analysis (SMA) being the most common technique utilized in many
applications (a.o. Roberts et al., 1998; Asner and Lobell, 2000; Riaño et al., 2002; Roder et
al., 2008; Somers et al. 2010ab). SMA effectively addresses this issue by quantifying the sub-
pixel fraction of cover of different endmembers, which are assumed to represent the spectral
variability among the dominant terrain features. A major advantage of SMA is its ability to
detect low cover fractions, something which remains difficult with the traditional vegetation
indices (VIs) approach (Henry and Hope, 1998; Elmore et al., 2000; Rogan and Franklin,
2001). Moreover, SMA directly results in quantitative abundance maps, without the need of
an initial calibration based on field data as with VIs (Somers et al. 2010a, Vila and Barbosa,
2010). With regards to post-fire effects, rather few studies employed SMA to monitor post-
fire vegetation responses (Riaño et al., 2002; Roder et al., 2008; Sankey et al., 2008; Vila and
Barbosa, 2010). Although results of these studies were consistent, they were all restricted to
simple linear SMA models in which only one spectrum was allowed for each endmember. As
a consequence, the performance of these SMA models often appeared to be suboptimal
(Roder et al., 2008; Vila and Barbosa, 2010) because these models did not incorporate the
natural variability in scene conditions of terrain features inherent in remote sensing data
(Asner, 1998). To overcome this variability effect a number of solutions have been presented
(Asner and Lobell, 2000; Zhang et al. 2004, 2006; Somers et al. 2010b). Multiple endmember
SMA (MESMA), as presented by Roberts et al. (1998), probably is the most widely used
technique to reduce the variability effects. In this model natural variability is included by
allowing multiple endmembers for each constituting terrain feature. These endmember sets
represent the within-class variability (Somers et al., 2009a) and MESMA models search for
the most optimal endmember combination by reducing the residual error when estimating
fractional covers (Asner and Lobell, 2000). Rogge et al. (2006), however, clearly
demonstrated that reducing the residual error by applying MESMA not always results in the selection of the most appropriate endmember spectrum. An initial segmentation of the area prior to the unmixing process in order to retain areas which reveal a high similarity in the spectral properties of a certain endmember has been presented as a sound and computationally efficient solution to address this issue (Rogge et al., 2006).

In this context, we aim to map vegetation abundance three years after the large 2007 Peloponnese (Greece) wildfires using Landsat Thematic Mapper (TM) imagery. We contrast traditional simple SMA with one spectrum for each endmember with two approaches who account for the natural variability in substrates. The first approach is MESMA while the second method is a segmented SMA in which ancillary information (lithological map) is used to force the endmember selection. Using a combination of field and simulation techniques the accuracy of the MESMA and segmented SMA is assessed and compared to the traditional simple SMA.

2 Methodology

2.1 Study area

The study focuses on the recovery of several large burned areas situated at the Peloponnese peninsula, in southern Greece (36°50’-37°40’ N, 21°30’-22°20’ E) (Figure 1). The fire scars date from the 2007 summer. These fires were the worst natural disaster of the last decades in Greece, both in terms of human losses and the extent of the burned area. Elevations range between 0 and 2404 m above sea level. Limestone sediments cover most of the mountainous inland. Also significant outcrops of flysch sediments occur (Institute of Geology and Mineral Exploration, 1983; Higgins and Higgins, 1996). Flysch sediments are dominated by sandstone with finer siltstone and clay (Institute of Geology and Mineral Exploration, 1983; Higgins and Higgins, 1996). The hilly and mountainous inland is covered with shallow and gravelly soils (European Commission, 2005). The climate is typically Mediterranean with hot, dry summers
and mild, wet winters. For the Kalamata meteorological station (37°4’ N, 22°1’ E) the average annual temperature is 17.8 °C and the mean annual precipitation equals 780 mm (Hellenic National Meteorological Service, www.hnms.gr, accessed 20 December 2010). The fires consumed more than 175 000 ha, which merely consisted of shrub land and pine forest (Veraverbeke et al., 2010a). Black pine (*Pinus nigra*) is the dominant conifer species. The shrub layer is mainly characterised by *Quercus coccifera*, *Q. frainetto*, *Erica arborea* and *Arbutus unedo*. Perennial grasses cover significant parts of the ground. These grasses reveal summer-dormancy and are not photosynthetically active during the Mediterranean summer drought (Volaire and Lelievre, 2010). Mediterranean-type shrub lands are highly resilient to burning due to both obligate seeder and resprouter fire-adapted strategies. They regenerate in a couple of years (Trabaud, 1981; Capitaino and Carcaillot, 2008) in a so called autosuccession process (Hanes, 1971). Conversely, the recovery of the forests is considerably slower and can take up to several decades (Viedma et al., 1997; van Leeuwen et al, 2010).

**FIGURE 1 HERE**

### 2.2 Field data

#### 2.2.1 Spectral library

In September 2010, field spectrometry measurements of the dominant terrain features (i.e. endmembers) were collected in the burned areas three years after the fire. Measurements were obtained within one hour of local solar noon on clear-sky days with a single channel spectroradiometer (UniSpec-SC) covering the 300-1100 nm spectral domain with a 3.7 nm resolution (PP Systems, 2006). 59 top-of-canopy (TOC) measurements of regenerating vegetation were recorded. Canopy height ranged between 0.5 and 2 m which made it possible to collect TOC signatures without the need of a measurement platform. In addition, 39 spectra of non-photosynthetic (i.e. brown) vegetation and 29 spectra of shallow and gravelly soils of both flysch and limestone sediments were also obtained: 15 above flysch substrate and 14
above limestone substrate. The spectra of each class were collected from various locations throughout the study area. All measurements were obtained while holding the sensor 0.3 to 0.5 m above the sample. The shadow endmember was assumed to be a uniformly dark component and was modeled as a flat 1% reflectance (Lelong et al., 1998; Somers et al., 2009a, 2010ab). The spectra were resampled to the TM wavebands to facilitate further analysis. Figure 2A shows the mean spectral signatures of each constituting endmember. The TM visual and near-infrared (VNIR) band passes are indicated in the figure. In the area two main substrate classes appear: limestone and flysch sediments. Corresponding spectral signatures are plotted with dashed lines in figure 2A, whereas figure 2B shows the occurrence of these two substrate classes in the 2007 burned areas. This classification was obtained after interpreting and digitizing the geological map of the area (Institute for Geology and Mineral Exploitation 1983). The difference in the substrates' optical properties is clear from the figure. The limestone substrate is relatively bright compared to the darker flysch substrate.

FIGURE 2 HERE

2.2.2 Line transect data

78 line transect plots were sampled to estimate the abundance of regenerating vegetation in the 2007 burned areas three years post-fire, in September 2010. 46 plots were measured in flysch areas whereas the remaining 32 samples were obtained on a limestone substrate. The sample scheme was designed for the 30 m Landsat resolution. The plots were selected during several one-day hikes based on a stratified sampling approach taking into account the constraints on mainly accessibility and time, encompassing the range of variability in recovery rates in the study area. The plot's centre coordinates were recorded with a handheld Garmin eTrex Visa Global Positioning System (GPS, 15 m error in x and y: Garmin, 2005). To minimize the influence of spatial autocorrelation plots were located at least 500 m apart. They consist of two perpendicular 60 m line transects, of which the first was directed north-
south. Using the point-intercept method (Bonham, 1989; Clemente et al., 2009; van Leeuwen et al., 2010; Vila and Barbosa, 2010) at a 1 m interval along the line transect, vegetation abundance was estimated. The fraction of vegetation cover equals the total number of vegetation interception points divided by the total number of interception points (Bonham, 1989) (Figure 3). 60 m linear transects were preferred to 30 m transects to anticipate potential satellite misregistration. Moreover, samples were located in relatively homogeneous areas of regrowth. Figure 4 shows example plot photographs of shrub land at different recovery rates.

FIGURE 3 HERE

FIGURE 4 HERE

2.3 Satellite data and preprocessing

One 30 m resolution Landsat TM image (path/row 184/34, acquired on July 18, 2010) was used in this study. We have tried to minimize the difference in phenology between the image and field data acquisition, however, small differences in phenology and resulting reflectance might influence the analysis. The image of July 18, 2010 was the acquisition that most closely resembled the ecosystem status as measured during the field campaign in September 2010. Analysis was restricted to wavebands between 400 and 1100 nm because of the consistency with the field spectral library. In addition, these wavebands show a higher signal-to-noise ratio than other spectral regions (Chen and Vierling, 2006). For TM imagery this results in four spectral bands: blue (B, 450-520 nm), green (G, 520-600 nm), red (R, 630-690 nm) and near infrared (NIR, 760-900).

The TM image was geometrically corrected using a set of homologous points of a previously georeferenced TM image of the study area (Veraverbeke et al., 2010ab, 2011). The resulting Root Mean Squared Error (RMSE) was lower than 0.5 pixels. The image was registered in Universal Transverse Mercator (UTM, zone 34S), with ED 50 (European Datum 1950) as geodetic datum.
Raw digital numbers (DNs) were scaled to at-sensor radiance values \( L_s \) (Chander et al., 2007). The radiance to reflectance conversion was performed using the COST method (Chavez 1996):

\[
r_a = \frac{\pi(L_s - L_d)}{(E_o / d^2)(\cos \theta_z)^2}
\]

where \( r_a \) is the atmospherically corrected reflectance at the surface; \( L_s \) is the at-sensor radiance (\( \text{Wm}^{-2}\text{sr}^{-1} \)); \( L_d \) is the path radiance (\( \text{Wm}^{-2}\text{sr}^{-1} \)); \( E_o \) is the solar spectral irradiance (\( \text{Wm}^{-2} \)); \( d \) is the earth-sun distance (astronomical units); and \( \theta_z \) is the solar zenith angle. The COST method is a dark object subtraction (DOS) approach that assumes 1% surface reflectance for dark objects (e.g. deep water).

Additionally, it was necessary to correct for differing illumination effects due to topography. This was done based on the modified c-correction method (Veraverbeke et al., 2010c), a modification of the original c-correction approach (Teillet et al., 1982), using a digital elevation model (DEM) and knowledge of the solar zenith and azimuth angle at the moment of image acquisition. Topographical slope and aspect data were derived from a 30 m DEM (Hellenic Military Geographical Service, HMGS) resampled and co-registered with the TM images. The illumination is modeled as:

\[
\cos \gamma_i = \cos \theta_p \cos \theta_z + \sin \theta_p \sin \theta_z \cos (\varphi_a - \phi_o)
\]

where \( \gamma_i \) is the incident angle (angle between the normal to the ground and the sun rays); \( \theta_p \) is the slope angle; \( \theta_z \) is the solar zenith angle; \( \varphi_a \) is the solar azimuth angle; and \( \phi_o \) is the aspect angle. Then terrain corrected reflectance \( r_t \) is defined as:

\[
r_t = r_a \left( \frac{1 + c_k}{\cos \gamma_i + c_k} \right)
\]
where $c_k$ is a band specific parameter $c_k = b_k / m_k$ where $b_k$ and $m_k$ are the respective
intercept and slope of the regression equation $r_a = b_k + m_k \cos \gamma_i$. Since topographic
normalization works better when applied separately for specific land cover types (Bishop and
Colby, 2002) specific $c$-values for the recovering 2007 scars were calculated by masking the
unburned areas (Veraverbeke et al., 2010c).

2.4 SMA

SMA is a commonly used image analysis technique to derive abundance estimates of
dominant ground components (e.g. green vegetation, substrates, etc.). Although some authors
recognize the occurrence of multiple photon scattering (Ray and Murray, 1996; Somers et al.
2009b), most vegetation monitoring studies consider a mixed pixel spectrum ($r$) as a linear
combination of pure spectral signals of its constituent components or endmembers, weighted
by their corresponding sub-pixel fractional covers (Adams et al., 1986):

$$r = Mf + \epsilon$$  \hspace{1cm} (4)

where $M$ is a matrix in which each column corresponds with the pure spectral signal of a
specific endmember, $f$ is a column vector $[f_1,...,f_m]^T$ denoting the cover fractions occupied by
each of the endmembers in the pixel. In this study, green vegetation, brown vegetation,
substrate and shadow are the endmembers of interest. $\epsilon$ represents the residual error.

Equation 4 is often solved by estimating abundance fractions using least squares error
estimates. Once the pure spectral signals of the endmembers are known, the fraction vector $f$ is
calculated by minimizing the following equation:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left( \sum_{j=1}^{m} (M_{i,j} \times f_j) - r_i \right)^2$$  \hspace{1cm} (5)

where $n$ is the number of spectral bands (Barducci and Mecocci, 2005). Generally, physically
meaningful abundance estimates are obtained by constraining the cover fraction to sum to
unity and to be positive (Roberts et al., 1993).
Endmembers may be derived from spectral libraries built from field or laboratory measurements (Roberts et al., 1998). Yet, endmember reference spectra can also be derived directly from the image data themselves (Bateson et al., 2000). Even in quickly recovering ecosystems, the diameter of woody individuals seldom exceeds 2 m in a medium-term perspective (3 years post-fire) (Keeley and Keeley, 1981; Malanson and Trabaud, 1988; Clemente et al., 1996). As a result, the occurrence of pure image pixels in the post-fire recovery areas is very rare at the Landsat 30 m resolution. As a consequence we acquired pure field spectra as described in section 2.2.1. To account for endmember variability, several authors suggest to evaluate multiple endmember combinations from the spectral library instead of using a fixed mean signature per endmember (Roberts et al., 1998; Asner and Lobell, 2000). Then, pixels are iteratively decomposed using different sets of endmember combinations and ultimately these fractional covers corresponding with the iteration that revealed the lowest least squares error are selected. This method is widely known as MESMA (Roberts et al., 1998). MESMA, however, does not always select the most appropriate endmember spectrum (Rogge et al., 2006). A prior segmentation of the imagery in zones that reveal a high similarity in the spectral properties of a certain endmember has been presented as a sound and computationally efficient solution for this issue (Rogge et al., 2006).

We executed three linear unmixing models. Each model used the mean spectrum as endmember for green and brown vegetation. The difference between the different models, however, is the definition of the substrate endmember:

- The first model using the mean substrate spectrum as soil endmember and is referred to as simple SMA.
- The second model is a simple MESMA in which two different soil spectra are incorporated (the mean flysch and limestone spectra).
The third model forces the choice between the mean limestone or flysch spectrum based on ancillary data. We used a generalized lithological map (Figure 2B) to ensure the proper substrate endmember selection. This technique is referred to a segmented SMA.

Preliminary experiments indicated that it was impossible to discriminate between the brown vegetation and substrate endmembers. This is explained by their high spectral similarity (Figure 2A) and corroborates with previous findings of Goodwin et al. (2005), Gill and Phinn (2009) and Somers et al. (2010b). As such, the best characterization of image variance was achieved with a three-endmember (green vegetation, substrate and shadow) model. To obtain ecologically meaningful estimates, the shadow cover fraction cover was distributed over the green vegetation and substrate components, proportionally to the estimated fractional cover of these components (Roder et al., 2008).

2.5 Analysis method

2.5.1 Simulated data

The analysis is twofold. Firstly, we used the spectral library with pure substrate (29) and vegetation signals (59) to create simulated mixed pixels. According to equation 4, a total of 1000 mixed vegetation-substrate spectra were calculated. 500 of them were constituted with a limestone spectrum while for the other half a flysch endmember was used. Pure pixel spectra combinations and fractional covers were randomly assigned to each pixel. To account for ambient and instrumental error, normally distributed noise was added to the signal (with a mean of zero and standard deviation ranging from 0% to 15% of the mixed signal, Asner and Lobell, 2000). Subsequently, each mixed spectrum was unmixed using the three different models. The first model, traditional simple SMA, uses one spectrum for each endmember. The second model, MESMA, chooses the substrate endmember (flysch or limestone) corresponding with the lowest residual error. Finally, the segmented SMA model forces the
choice between the limestone or flysch endmember based on ancillary knowledge. Simulated
data supply a reliable means to evaluate the performance of the various models as it inherently
provides correct validation data (Rogge et al., 2006). The performance of each model was
expressed in the coefficient of determination ($R^2$) of the linear regression with the estimated
vegetation fractions as independent variable and the modeled fractional vegetation covers a
dependent variable. Separate regression models were performed for the limestone mixtures,
the flysch mixtures and the pooled dataset combining limestone and flysch mixtures. In
addition, the selection of the proper substrate endmember by the MESMA model was
evaluated using the knowledge of the set-up of the simulation experiment as reference data.

2.5.2 Landsat imagery

The second part of the analysis focused on the Landsat TM data. The same three unmixing
models were applied and vegetation fractional covers of the line transect locations were
extracted by calculating the mean index value of a 3-by-3 pixels matrix. It is widely accepted
that using the mean of a pixel matrix minimizes the effect of potential misregistration (Ahern
et al. 1991). Linear regressions were performed to correlate the TM fractional covers
(independent variables) and line transect field data of vegetation recovery (dependent
variables). Regression model results were compared using the $R^2$ statistic. Again, separate
regression models were performed for the 32 limestone plots, for the 46 flysch samples and
for the 78 field ratings together. The ancillary knowledge of the constituting substrate
endmember was also used to assess the performance of the MESMA model's endmember
spectrum selection. The best method was used to map the vegetation abundance three years
after the large 2007 Peloponnese wildfires.

3 Results

3.1 Simulated data
Figure 5 displays the scatter plots and regression lines of the simulation experiments. In figure 5A the results of the traditional SMA model are visualized, while figure 5C and 5E respectively depict the outcomes of the MESMA and segmented SMA models. A comparison between the simple SMA and MESMA model learns that the $R^2$ between modeled and estimated fraction covers was higher for MESMA compared to simple SMA for the flysch subset, limestone subset and the whole dataset (respectively 0.75, 0.75 and 0.68 for simple SMA and 0.79, 0.79 and 0.77 for the three datasets for MESMA). However, the goodness-of-fit of the segmented SMA for the pooled dataset was yet higher ($R^2 = 0.79$), whereas $R^2$ values of the substrate subsets were equal to the MESMA model. Moreover, for the segmented SMA the regression parameters of the flysch subset, limestone subset and pooled dataset closely resembled each other (slope respectively 0.77, 0.78 and 0.78 and intercept 0.12 for three datasets) whereas with simple unmixing regression slope (0.90 for the flysch subset, 0.70 for the limestone subset and 0.74 overall) and intercept (-0.04 for the flysch subset, 0.21 for the limestone subset and 0.12 pooled) significantly diverged. Also for MESMA a similar divergence was present in the data: the regression slope equalled respectively 1, 0.79 and 0.86 for the flysch, limestone and pooled data, whereas the intercept was respectively -0.08, 0.12 and 0.04. The divergence of the different regression lines as observed with simple SMA and MESMA was especially obvious for low vegetation cover estimates. Figures 5B, 5D and 5F respectively show the same model as presented in figures 5A, 5C and 5E, however, in these models randomly distributed noise was added. This did not impact the trends described above, however, $R^2$ values revealed a small drop compared to their noise-free counterpart. The only exception against this drop was the traditional SMA model of the pooled dataset which retained its $R^2 = 0.68$. 

FIGURE 5 HERE
The error matrix of the selection of the substrate spectrum by MESMA based on simulated data is tabulated in Table 1. The overall accuracy equalled 61% and a relatively low Kappa coefficient of 0.21 was obtained. The MESMA model's substrate spectrum selection revealed a high omission error for the flysch class (producer's accuracy of 29%) and relatively high commission error for the limestone class (user's accuracy of 57%).

**TABLE 1 HERE**

### 3.2 Landsat imagery

Figure 6 presents scatter plots and regression line between the line transect field ratings and the vegetation fractional covers retrieved from the Landsat imagery. In corroboration with the results from the simulations (Figure 5), the regression parameters of the segmented SMA model were very similar for the flysch subset, limestone subset and pooled dataset (slope respectively 1.02, 0.99 and 1.03 and intercept respectively -0.06, -0.08 and -0.08). This contrasts with the more differing regression slope and intercept of the simple SMA (slope respectively 1.07, 0.93 and 0.83 for the flysch subset, limestone subset and pooled dataset and intercept respectively -0.15, -0.01 and 0) and MESMA models (slope respectively 0.71, 0.89 and 0.86 for the flysch subset, limestone subset and pooled dataset and intercept respectively 0.10, -0.01 and 0.06). For the simple SMA, this did not result in less optimal regression models for the substrate subsets, however, the overall $R^2$ was clearly higher for the model that forced the flysch-limestone endmember choice ($R^2=0.70$ versus $R^2=0.65$ for simple SMA). For MESMA, the goodness-of-fit was lower for both subset and pooled data (e.g. $R^2=0.63$ for the pooled dataset). In addition, the regression lines of the segmented SMA more closely resembled the expected one-one line compared to the other models.

**FIGURE 6 HERE**

The error matrix of the selection of the substrate spectrum by MESMA based on field data is listed in Table 2. Similar to the results of table 1, the overall accuracy equalled 62% and a
relatively low Kappa coefficient of 0.18 was obtained. The MESMA model's substrate spectrum selection revealed a high omission and commission error for the limestone class which resulted in a relatively low producer's accuracy (41 %) and user's accuracy (54 %) for this class. Producer's and user's accuracy for the flysch category were slightly higher (respectively 76 % and 65 %).

The ancillary information of figure 2B was used to differentiate between relatively bright (limestone) and dark (flysch) substrates when mapping the post-fire vegetation cover while accounting for background variability using the segmented SMA model (Figure 7).

4 Discussion

Post-fire recovery landscapes essentially are mixed vegetation-substrate environments. A plethora of studies made use of this feature to map post-fire vegetation cover with the NDVI (a.o. Viedma et al., 1997; Díaz-Delgado et al., 2003; McMichael et al., 2004; Malak and Pausas, 2006; Clemente et al., 2009). To obtain qualitative fractional cover maps, these index values require a prior calibration with field estimates of vegetation cover (Clemente et al., 2009). In this study, SMA demonstrated to be a strong alternative for the spectral indices approach, as SMA outputs fraction images without an initial regression fit between remotely sensed data and field ratings.

The regression fit between the line transect field estimates of recovery and the most optimal SMA resulted in moderate-high $R^2 = 0.70$. The residual variation can be explained by the fact that both field and remotely sensed estimates are imperfect proxies for vegetation cover. The line transect method is a relatively rough approach to estimate fractional vegetation cover while several noise factors hamper satellite image analysis. Inaccurate atmospheric correction (Gong et al., 2008), suboptimal illumination correction (Veraverbeke et al. 2010c), sensor noise (Plaza et al., 2004), slight differences in acquisition timing between field and image data or the unmixing model structure itself (e.g. non-linear mixing due to multiple photon
scattering among different ground components, Borel and Gerstl, 1994; Somers et al., 2009b) are all known to create noise in image analyses. The influence of soil brightness variation, however, was a very important factor impacting model performance. Both the simulation experiment and Landsat application demonstrated that accounting for soil brightness variations by the segmented approach significantly improved the SMA model. The simple SMA with one single spectrum for each endmember provided reasonable regression models for each substrate class separately, however, model performance of the pooled dataset was considerably weaker. This is explained by the fact that traditional SMA resulted in clearly different regression lines depending on substrate class (Figure 5A, 5B and 6A). In other words, the relationship between the observed (field or modeled) vegetative fraction and the estimated fraction from the simple SMA model was determined by the brightness of the background. Thus, neglecting this background brightness difference produced a weaker overall fit. Moreover, the simple SMA model underestimates the vegetative fraction in limestone areas while in flysch areas the opposite is true. As shown in figure 2A the optical properties of these two substrate types are clearly different. They represent a relatively bright (limestone) and dark (flysch) background. MESMA is the most widely used technique to include endmember variability in a SMA model (Roberts et al., 1998). Table 1 and 2, however, clearly indicated that MESMA did not manage to select the appropriate substrate spectrum in this case study. The Kappa coefficients of 0.18 an 0.21 for respectively the simulation experiment and the Landsat application revealed that the substrate spectrum selection was only slightly better than an agreement by chance. As a consequence, MESMA did not solve the substrate variability issue in this application. This can be explained by the fact that the spectral signatures of limestone and flysch are almost linear translations of each other (Figure 2A). Due to the lack of shape differences between these two substrate spectra, MESMA did not demonstrate a strong tendency to select the appropriate soil endmember. In
contrary, the ultimate selection of the substrate endmember appeared to be rather arbitrarily. It is recognized that when different substrate endmember spectra reveal clear shape differences, MESMA can be a very straightforward solution to find the proper substrate spectrum based on an iterative process (Roberts et al., 1998).

Because of the failure of the MESMA model in this case study, we applied a segmented approach in which the substrate endmember choice was based on ancillary knowledge (i.e. the simulation set up in the case of the simulations and a generalized lithological map for the Landsat application). For this model, regression slope and intercept did not depend on substrate class (Figures 5E, 5F and 6C). So irrespective which substrate type, the regression lines were similar. As a consequence, potential over- or underestimation of vegetative cover was eliminated and the performance of the pooled regression model was equally high. The SMA model that accounts for soil brightness variations also produced regression fits very close to the expected one-one line, which proves its consistency. These beneficial results of the segmentation approach corroborate with Rogge et al. (2006) who demonstrated the effectiveness of prior segmentation to overcome poor endmember spectrum selection by MESMA. In addition, limiting the number of the potential endmember spectra favors the computational efficiency compared to MESMA models (Rogge et al., 2006).

In post-fire recovery studies using SMA, Riaño et al. (2002), Roder et al. (2008) and Vila and Barbosa (2010) all employed one single substrate endmember. Disregarding soil brightness variations potentially adds an explanation to the observed suboptimality of the SMA outcomes observed by Roder et al. (2008) and Vila and Barbosa (2010). We want to remark that in the simulation model vegetation cover was slightly overestimated for very low vegetative covers, while the model slightly underestimated the vegetative fraction for mixtures in which the vegetation component dominates (Figure 5). For extreme fractional vegetation covers (close to zero and one) the SMA simulation models showed a tendency to estimate vegetative
fractional cover as respectively zero and one. This explains the slight over- and underestimation observed in the simulation experiment. Due to the fact that most field ratings range between 20-70% vegetative coverage, this behavior is not present in the regression fit between Landsat and line transect data. In contrast, the overall regression intercept of the modified SMA regression model is slightly negative (-0.08). However, the general SMA constraint that fraction estimates have to be positive (Roberts et al., 1998), prevents the occurrence of negative fractional covers without biophysical meaning. In the field, the presence of extreme fractional covers (close to zero or one) was extremely rare, so these cases do not nullify the performance of the model. In this respect, a totally different scenario would emerge when one would aim to estimate the post-fire vegetation regrowth very shortly after the fire, e.g. one year after the fire. Then, it would be wise to additionally evaluate the model performance for very low vegetation covers. However, in contrast with our study, a one year post-fire assessment would also need to include a char endmember in the model (Lewis et al., 2007; Robichaud et al., 2007).

A drawback of the proposed method is the need of ancillary data. With a combination of field knowledge and lithological maps it is relatively easy to construct spectrally similar lithological units, however, this possibility depends on the availability of such data layers. Besides among substrates, endmember variability is also present among vegetation species. In our case study, however, the variability in the spectral response of different vegetation species was very small compared to large spectral differences between the substrate classes. For this reason and because of the small sensitivity of broadband sensors to discriminate between different vegetation types (Somers et al. 2010a), we disregarded vegetation variability in our analyses. Other pathways to improve the accuracy of the recovery assessment are multiple. A possible amelioration could be the inclusion of the short-wave infrared (SWIR: 1300-1700 nm) and mid infrared (MIR: 1700-2400 nm) spectral regions in the unmixing process. These
spectral regions have proven to be very effective in discriminating soil and vegetation (Drake et al., 1999; Asner and Lobell, 2000). Moreover, the SWIR-MIR spectrum is very sensitive to moisture content (Hunt and Rock, 1989; Zarco-Tejada et al., 2003) and are consequently strongly related to plant water content. Carreiras et al. (2006) demonstrated that adding the SWIR-MIR Landsat bands resulted in better estimates of tree canopy cover in Mediterranean shrublands. To retain consistency with the field spectral library these wavebands were not included in our study (Somers et al., 2010a). Additionally, enhancing the spectral resolution by employing hyperspectral data would increase the amount of spectral detail which would benefit the differentiation between spectra. By including more and other spectral wavebands the unmixing model could gain discriminative power. Potentially, this would make it even possible to distinguish between non-photosynthetic vegetation and substrate (Asner and Lobell, 2000; Somers et al; 2010a), which appeared to be impossible based on the Landsat VNIR bands.

5 Conclusions

Using a combination of field and simulation techniques, the importance of accounting for background brightness variability in estimating fractional vegetation cover using SMA was highlighted. Although the traditional SMA model in which the substrate endmember was defined as the arithmetic mean of two flysch and limestone substrates subclasses resulted in reasonable regression fits for the flysch and limestone datasets separately, the regression fit performed on the pooled dataset was considerable weaker. The regression lines of the different datasets (only limestone, only flysch and pooled) significantly diverged and as such vegetative cover estimations depended on substrate type. The use of a single spectrum substrate endmember thus resulted in an over- or underestimation of the vegetative cover fraction related to background brightness differences. Traditionally, MESMA is applied to address the endmember variability issue, however, in this case study MESMA did not manage
to select the appropriate substrate endmember due to the lack of shape difference between the flysch and limestone spectra. Therefore, a prior segmentation based on ancillary information (lithological map) was executed to incorporate soil color variation in a segmented SMA model. This model forces the proper substrate endmember spectrum choice. The overall regression fit of the segmented approach significantly improved and the discrepancy between the regression of the different subsets significantly reduced. Moreover, the resulting regression line very closely resembled the expected one-one line between observed and estimated fractional vegetation covers.

This paper demonstrated the utility of SMA for monitoring post-fire vegetation regeneration three year after the 2007 Peloponnese wildfires. Although a segmented approach to account for soil brightness variations significantly improved the model, further research is required to evaluate the model’s performance for other soil types, with other image data and at different post-fire timings.

Acknowledgements

The study was financed by the Ghent University special research funds (BOF: Bijzonder Onderzoeksfonds). Part of the work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The contribution of Dr. Ben Somers is funded by the Belgian Science Policy Office in the frame of the STEREO II programme – project VEGEMIX (SR/67/146).

References


European Commission, 2005. Soil atlas of Europe, Office for Official Publications of the European Communities, Luxembourg


PP Systems, 2006. Unispec-SC (single channel) spectral analysis system operation manual version 2.02, Hertfordshire


Figure 1. Location of the study area (the areas encircled with black represent the 2007 burned areas) and distribution of the field plots (marked with white crosses) (Landsat Thematic Mapper image July 18, 2010 RGB-432).

Figure 2. Mean spectral signatures of green vegetation, brown vegetation, and substrate acquired in the field with a Unispec single channel field spectroradiometer (A). The shadow endmember is modeled as a flat 1% reflectance (Lelong et al., 1998). Specific spectra for limestone and flysch substrate are indicated by the dashed lines. The Thematic Mapper (TM) visual and near infrared bandpasses are also shown. B shows the presence of flysch and limestone substrates in the 2007 burned areas (based on Institute for Geology and Mineral Exploration, 1983).

Figure 3. Line transect plot design (Bonham 1989)

Figure 4. Example plot photographs of shrubland with a high (A), moderate (B) and low (C) recovery rate.

Figure 5. Scatter plots and regression lines of modeled versus estimated fractional vegetation cover of the simulation experiments for the noise-free simple Spectral Mixture Analysis (SMA) (A), the noise-free multiple endmember SMA (MESMA) (C) and the noise-free segmented SMA (E) and the equivalent models with noise (Asner and Lobell, 2000) (respectively B, D and F). Separate scatter plots and regression lines are displayed for the flysch subset (n = 500) and limestone subset (n = 500). Regression lines of the pooled dataset (n = 1000) are also indicated.

Figure 6. Scatter plots and regression lines of line transect ratings versus fractional vegetation cover derived from Landsat imagery for the simple Spectral Mixture Analysis (SMA) (A), the multiple endmember SMA (MESMA) (B) and segmented SMA (C). Separate scatter plots and regression lines are displayed for the flysch subset (n = 46) and limestone subset (n = 32). Regression lines of the pooled dataset (n = 78) are also indicated.

Figure 7. Fractional vegetation cover map three years after the fires based on the segmented SMA model.

Table 1. Error matrix of the substrate spectrum selection by the multiple endmember Spectral Mixture Analysis (MESMA) model for the simulation experiment. The reference data were retrieved from the experimental set-up.

Table 2. Error matrix of the substrate spectrum selection by the multiple endmember Spectral Mixture Analysis (MESMA) model for the Landsat application. The reference data are the line transect field plots.