Delineation of Soil Parameters to assess Ecosystem Degradation using Spectral Mixture Analysis

A. Bayer [a,*], M. Bachmann [a], A. Müller [a]

[a] German Remote Sensing Data Center, German Aerospace Center, 82234 Wessling, Germany – anita.bayer@dlr.de

Abstract – Land degradation processes in the subtropical Thicket Biome in the Eastern Cape Province, South Africa, are observed and monitored. As a result, a significant loss of soil quality on such sites has been recorded. This study focuses on the determination of fundamental soil parameters like organic carbon, iron, and clay in order to assess ecosystem degradation. The test site in South Africa is surveyed for ground truth and hyperspectral image data are obtained. We take advantage of spectral mixture analysis to approximate the 'pure' soil signal from mixing pixel signatures. For a subsequent quantification of soil parameters, spectral feature analysis is linked with multiple linear regression techniques. For organic carbon and iron, calibrations of high accuracy are used for the prediction of image data. The results highly correlate with measured contents. In contrast, the quantification of clay content is still problematic mostly due to the existence of soil structural crusts.

Keywords: Land degradation, soil organic carbon, imaging spectroscopy, spectral feature analysis, multiple endmember spectral mixture analysis.

1. INTRODUCTION

The Subtropical Thicket Biome (Eastern Cape Province, South Africa) with its dense shrub vegetation in pristine conditions stores exceptionally large amounts of carbon for a semi-arid region. The predominant endemic succulent shrub *Portulacaria afra* (colloquially called Spekboom) accounts for the fixation of large amounts of carbon in its biomass and the peripheral soils. Once this type of ecosystem is degraded, it is not able to recover naturally. During the degradation process, mainly due to overgrazing by goats, semi-arid shrub land with a closed canopy changes to an open savannah-like system (figure 1). This is accompanied by a severe loss of biodiversity and ecosystem carbon stocks (Mills and Fey, 2003). Quantitative maps outlining the amount of soil organic matter, as well as iron and clay content in the soils of degraded Thicket ecosystems, would be very valuable to steer the ongoing restoration activities in the area. Therefore, spatially continuous quantitative information derived from remote sensing data is of high interest for the area.

This research study aims at quantifying key soil parameters (organic carbon, iron, and clay content) based on airborne hyperspectral data. A combination of spectral feature analysis and multivariate statistical approaches are applied. This paper presents the methodological framework and intermediate results of the delineation of soil constituents focusing on organic carbon. The potential of using spectral information in the visible and shortwave infrared range for quantitative analysis of soil properties is well known. Many studies exist where laboratory, field, and airborne hyperspectral data in combination with different modelling approaches are used to describe constituents of the upper soil layer (e.g. Gomez et. al, 2008, Stevens et. al, 2008). A summary of key studies using imaging spectroscopy to study soil properties can be found in Ben-Dor et. al, 2009.

2. BASE DATA AND METHODS

Thicket shrub vegetation covers approximately 10 % of the land area of the Eastern Cape Province. The slightly hilly terrain is mainly used for livestock and game farming. The underlying shales and sandstones have resulted in the development of loamy and sandy soils at surface (Cambisols and Luvisols, see Mills and Cowling, 2006). Within this region a test area (75 x 3 km) was selected with the highest variance of Subtropical Thicket vegetation classes (see figure 2). This increases the transferability of methods developed to the entire Thicket Biome region.

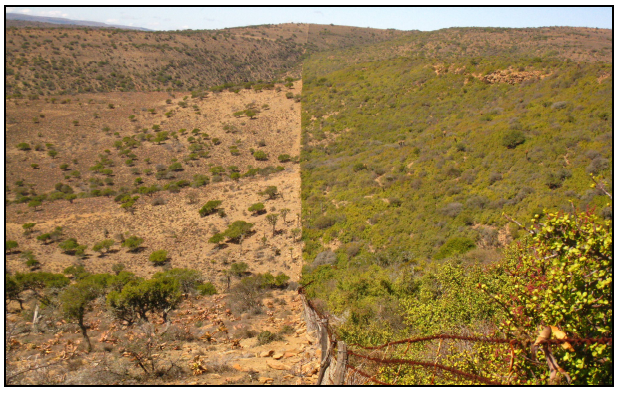

Figure 1. Highly degraded on the one side and on the other side only slightly influenced nearly pristine Thicket shrub vegetation. Those high contrasts can be found at fence lines between pasture and game farms in the Eastern Cape Province.

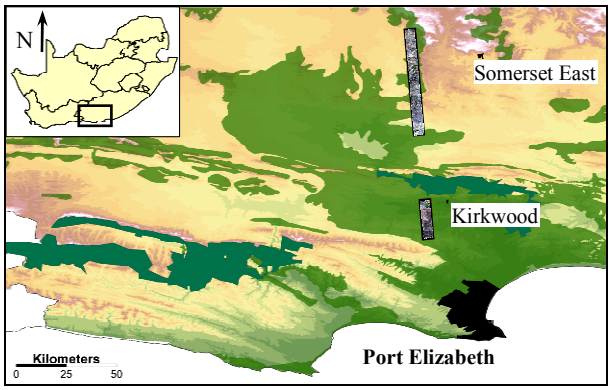

Figure 2. Distribution of Subtropical Thicket Vegetation and location of the research area in the Eastern Cape, South Africa.
Subtropical Thicket Vegetation,
Nature Reserves.

* Corresponding author

Table A. Range of values determined by chemical analysis of 164 collected soil samples (analysis conducted by University of Stellenbosch, South Africa).

	C_{org} [%]	Iron [%]	Clay [%]
Min	0.21	0.90	0.00
Max	5.85	10.62	23.80
Mean	1.21	3.06	6.46

Hyperspectral remote sensing data obtained from the airborne HyMap sensor (operated by HyVista Corporation, Australia) were recorded over the study area in October 2009. Additional ground reference data was collected in two field surveys in 2009. Within 164 plots, field spectra of exposed soils (using an ASD Field Spec Pro Spectrometer), soil samples, information about vegetation coverage density and type, and surface conditions (especially soil crusts and stone coverage) were taken. The soil samples were chemically analysed for organic carbon, dithionite extractable iron, and clay content using the Walkley-Black, CBD and pipette method, respectively (see table A). After the samples were dried, ground and sieved, the soil fraction less than 2 mm was spectrally measured in the laboratory. Hyperspectral image data are orthorectified and corrected for atmospheric effects using the ATCOR procedure including inflight calibration to enhance the accuracy of spectral signals (Richter, 2010). HyMaps bad bands (1, 95 to 100) are removed and a Savitzky-Golay filter is applied to remove small remaining artefacts. Field and laboratory data sets are also Savitzky-Golay filtered and resampled to HyMaps spectral resolution and the 119 suitable spectral bands.

2.1 Spectral Mixture Analysis

To minimize the spectral influence of vegetation on a non homogenous soil pixel, a linear mixture model is applied. For selected subscenes potential spectral endmembers are derived using the automated Sequential Maximum Angle Convex Cone approach (SMACC, see Gruninger et. al, 2004). The algorithm searches for extreme pixels in the n-dimensional feature space of hyperspectral image data. To support the detection of endmembers for bare soil and dry vegetation SMACC is also applied to a preclassified image, where pixels with a Normalized Differenced Vegetation Index (NDVI) higher than 0.2 are excluded. From this pool of endmembers we selected five to six endmembers for each of the classes: photosynthetic active vegetation (PV), non-photosynthetic active vegetation (NPV), and bare soil (BS). To perform linear spectral unmixing we utilized Multiple Endmember Spectral Mixture Analysis (µMESMA) according to Bachmann, 2007. This approach was developed for the retrieval of subpixel ground cover fractions in semi-arid regions. A high accuracy for µMESMA unmixing was detected for semi-arid regions in south-western Spain which is characterised by similar land coverage (Bachmann, 2007). The automated algorithm derives quantitative cover fractions for the two vegetation classes PV and NPV and BS according to the basic equation of the linear mixture model

$$S = EM \cdot A \quad (1)$$

The spectrum S measured in m bands can be modeled by the endmember matrix EM with n endmembers and the abundance matrix A of the dimensions m*n. µMESMA is applied using a partially constrained algorithm, with abundances of each class between 0 and 1, but the sum-to-1 constraint not fulfilled.

2.2 Approximation of soil signature

The unmixing results are used to approximate the soil spectral signal by simple residual analysis of the linear mixture equation

$$SR = S - EM_{PV} \cdot A_{PV} - EM_{NPV} \cdot A_{NPV} \quad (2)$$

with the soil residuum SR, the original pixel spectrum S and the particular endmember EM of PV and NPV used by µMESMA to determine the related abundances A for each spectrum. The residual part we refer to as soil residuum SR, which includes the spectral signal coming from the soil fraction and any unexplained signals not triggered by bare soil or the two vegetation classes. A successful approximation of the soil signal presumes in the first instance an accurate mixture model for each pixel and some additional criteria that assure the quality of the extracted spectra. The following quality requirements are set up to filter out these pixels: Sum of the three abundances not less than 0.5, minimal soil abundance of 0.4, abundance of PV or NPV not higher than 0.25, and a low root mean squared error (RMSE). In addition, the mean reflectance of the extracted soil residuum is set to be above 10 %, to filter out dark pixels where eventually unexplained signal highly influences the calculated soil residuum. A pixel fulfilling these requirements is supposed to show a valid soil spectrum, while other pixels are masked.

Figure 3 shows examples for approximated soil spectra of two pixels with different soil fractions. For the pixel spectrum of example 1 µMESMA calculated abundances of 3 % for PV and 13 % for NPV, respectively, while example 2 shows the extracted soil residuum for a comparably dark pixel with 3 % of PV and 24 % of NPV. To approximate the 100 % soil endmember of each pixel the calculated residual spectrum is scaled according to the remaining abundance of SR with

$$A_{SR} = 1 - A_{PV} - A_{NPV} \quad (3)$$

Hence, the reconstructed soil residuum of example 1 is brighter (A_{SR} = 0.84) and the soil residuum of example 2 (A_{SR} = 0.73) darker than the related original pixel spectrum. In the visible range a reduction of the red edge characteristic for vegetation is apparent.

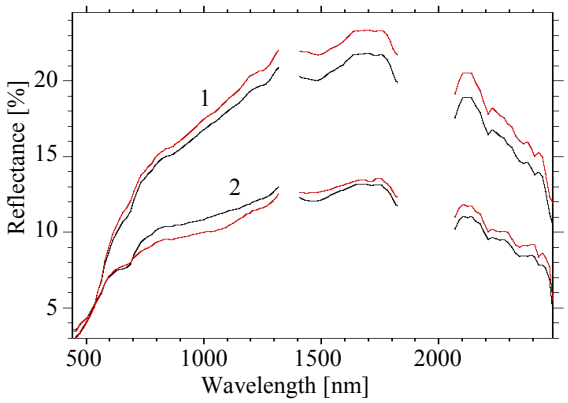

Figure 3. Examples for reconstructed 'pure' soil signatures from pixel spectra of different µMESMA abundances. ~Original pixel spectrum, ~scaled approximated soil residuum.

2.3 Determination of soil constituents

A first analysis for testing the correlation of soil spectral signatures with measured soil contents was conducted for the laboratory and field dataset by using statistical regression techniques. For this purpose the ParLes software for chemometrics which is based on partial-least-squares regression was applied (Viscarra Rossel, 2008). Correlation accuracies of R^2_{Cal} between 0.75 and 0.83 were achieved for soil organic carbon, iron, and clay (Bayer, 2010).

Table B. Number of spectral features used and total number of calculated variables for determination of key soil constituents.

	Absorption features (6 var.)	Curve features (1 var.)	Features of convex hull (2 var.)	Total number of variables
C_{org}	2	0	2	16
Iron	3	1	1	21
Clay	4	0	2	28

The approximated soil spectra are used for the quantification of the three chosen soil parameters from airborne hyperspectral image data. The developed approach parameterises characteristic spectral features and determines their relationship to measured chemical contents by subsequent regression analysis. In the first step, the input spectra are analysed for spectral characteristics in terms of specific absorption features (AF), features of the spectral curve itself (CV), and its continuum as convex hull of each spectrum (CO). Diagnostic spectral characteristics described in respective literature are included here (e.g. Viscarra Rossel and Behrens, 2010, Gomez et. al, 2008). For instance, for delineation of organic carbon two AF (around 1726 and 2340 nm) and two CO features (VIS range between 450 and 740 nm and NIR/SWIR range between 1000 and 1750 nm) are used. Special care is taken not to include spectral characteristics in wavelength ranges where typical features of PV and NPV occur. This minimises the influence of vegetation signature potentially remaining in the spectrum of the approximated soil residuum. Each AF is parameterised from continuum removed spectra by the following six variables: depth and wavelength of maximal absorption, absorption at supposed characteristic wavelength, feature width, area, and asymmetry. Each CV feature is characterised by the mean slope of the spectral curve in a defined wavelength range. For CO features the mean slope and mean reflectance of the convex hull in a defined wavelength range are calculated. Table B shows the number of spectral features analysed for the determination of soil organic carbon, iron, and clay.

With the calculated feature variables and the chemical reference values related to this parameter, a first multiple regression including leave-one-out analysis was conducted to identify and remove outlier samples as well as insignificant variables. Variables are considered insignificant if the absolute value of a regression coefficient is smaller than twice its standard deviation, calculated during the leave-one-out analysis (see Kessler, 2007). According to the partial-least squares principle a final multiple linear regression model is established for each soil parameter including the related significant feature variables and chemical reference values.

3. RESULTS

On the hyperspectral image data of the Thicket Biome, South Africa, spectral mixture analysis was carried out with overall low RMSE errors. Thus, it can be assumed that with the three classes PV, NPV, and BS and the given endmembers the mixture model is capable of modelling most of the spectral mixture for the specific land coverage and species occurrence present in the Thicket Biome. An example of image-derived abundances and calculated RMSE is given in Figure 4. Areas where green vegetation is detected correlate with vegetated spots clearly visible in the coloured-infrared image. Dry vegetation appears throughout the image, pixel having between 15 and 35 % NPV are prevalent. Bare soil is the dominant fraction with abundances between 35 and 80 %, except where green vegetation is present. The results show that the majority of pixels in this ecosystem can be explained with sufficient accuracy. A mixture of bare soil and dry vegetation is predominant.

The calibration accuracy R^2_{cal} of the regression model developed for organic carbon is 0.89 for the laboratory spectral dataset and with 0.82 slightly lower when utilizing the field dataset. The accuracy for iron is 0.79 and for clay 0.54, respectively (laboratory dataset). Thus, the model accuracies for soil organic carbon and iron, determined by a combination of spectral feature analysis combined with linear multiple regression techniques, are in the same range as statistical PLSR approaches provide. Though, the described approach is thought to be more robust as it reduces statistical adaptation. Thus it will allow a transfer of the established regression coefficients to hyperspectral airborne data collected of the study area.

Prediction results are shown with the focus on soil organic carbon (figure 5). The calibration was done by using the laboratory spectral dataset. So as to take neighbourhood effects into account, a moving window filter is applied to the maps of pixelwise calculated soil organic carbon contents. The filter with a small kernel size of 3 averages a pixel if no more than two pixels are missing in its direct surrounding. Thus, the filter interpolates singular missing pixels but does not extrapolate into borders of larger areas of excluded pixels (compare figure 5 A and B). The small filter kernel retains the manipulation of the statistics of calculated values at a low extent. Calculated C_{org}

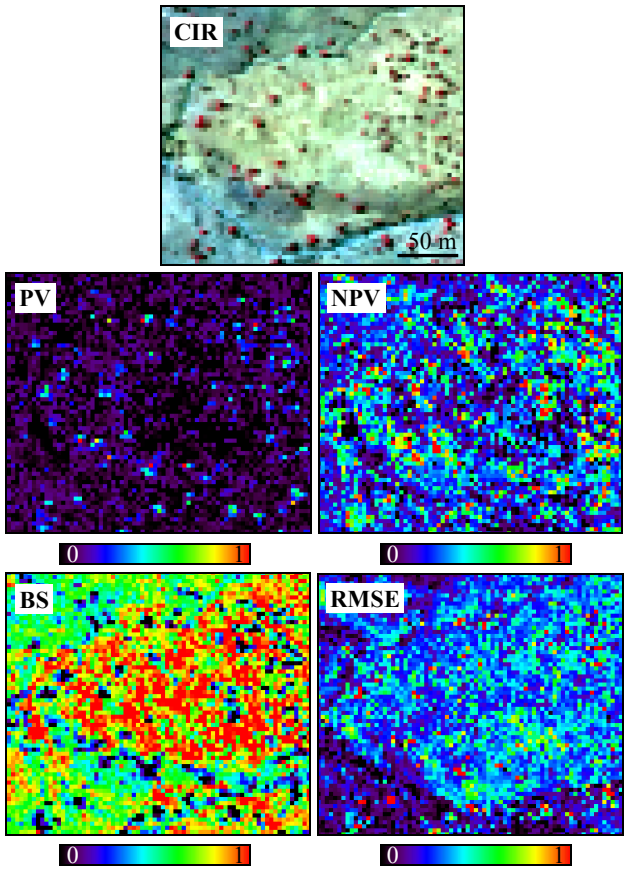

Figure 4. µMESMA derived quantitative cover fractions for a detail of the HyMap data (shown as coloured-infrared image). The abundances for classes photosynthetic active vegetation (PV), non-photosynthetic active vegetation (NPV), and bare soil (BS) are given as well as the RMSE determined for each pixel.

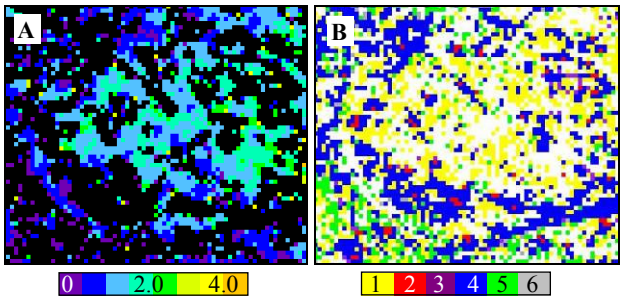

Figure 5. Quantification of soil organic carbon from HyMap image data, A: Derived contents of soil organic carbon grouped in classes ranging up to 4.0 % C_{org}, B: Reasonings for exclusion of pixels (1: Abundance of PV or NPV > 0.25, 2: High RMSE, 3: Sum of abundances < 0.5, 4: Soil abundance < 0.4, 5: Inaccurate calculation of C_{org} by feature-based model, 6: Mean reflectance of extracted soil residuum < 10 %).

ranges from 0 to 4.0 %, reaching the highest values in the surrounding of spots with green and dry vegetation, respectively (figure 5 A). This can be a result of vegetation signal remaining after the extraction procedure. As well it can be logical consequence of increased input of organic carbon around growing and recycled vegetation. Figure 5 B shows the detailed reasoning for the exclusion of pixels where no soil organic carbon could be quantified. Nevertheless, the 'valid' pixels provide a detailed impression of soil organic carbon distribution close to the surface. The comparison of calculated contents of soil organic carbon with chemical contents measured from field samples show good correlations. This leads to the assumption that for this purpose a calibration established using laboratory spectral data can be transferred to airborne hyperspectral data.

Comparable feature-based approaches are also conducted for soil iron and clay content. Especially when focussing on clay, surface structural crusts must be taken into account. This is because in crusted soils, clay content is increased directly at the surface and slightly reduced underneath.

4. CONCLUSIONS AND OUTLOOK

From the preliminary results it is concluded that a detailed quantification of soil organic carbon can be achieved with the proposed combination of methods. Spectral mixture analysis is applied as a type of pre-treatment step for the modelling of 'pure' soil spectra from mixing signatures by residual analysis. Subsequent analysis of characteristic spectral features (absorption features, characteristics of the spectral curve itself, and its continuum) are linked to measured chemical contents using multiple regression techniques. High correlation accuracies can be determined for the calibration by using laboratory and field spectral datasets (R^2_{Cal} = 0.89 for C_{org}). Results predicting soil organic carbon from airborne HyMap image data show a high correlation with contents of the upper soil layer measured from field samples. Detailed maps of soil parameters help to visualise the total ecosystem degradation and current status of the Thicket Biome region. Degradation trends can be monitored with ongoing surveys.

Further research aims to enhance the described methodology as well as to draw special attention to the influences of land coverage and surface conditions (e.g. soil structural crusts) on an accurate derivation of soil constituents. To link soil surface information with the deeper profile, another field survey with a detailed mapping of soil profiles and soil types in the working area is scheduled for summer 2011. A further objective is to test and adapt the proposed methodology on data with the designated specifications of the future EnMAP satellite in order to allow for future satellite surveys.

REFERENCES

M. Bachmann, "Automatisierte Ableitung von Bedeckungsgraden durch MESMA-Entmischung," PhD Thesis, University of Würzburg, Germany, 2007.

A. Bayer, „Quantitative derivation of Key Soil Parameters on the Basis of Hyperspectral Remote Sensing Data," In: Proc. of ESA Hyperspectral Workshop, Frascati, Italy, 2010.

E. Ben-Dor, S. Chabrillat, J.A.M. Demattê, G.R. Taylor, J. Hill, M.L. Whiting and S. Sommer, "Using Imaging Spectroscopy to study soil properties," Remote Sensing of Environment, vol. 113, p.p. 38-55, 2009.

C. Gomez, R.A. Viscarra Rossel, and A.B. McBratney, "Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy," Geoderma, vol. 146, p.p. 403-411, 2008.

J. Gruninger, A.J. Ratkowski and M.L. Hoke, "The Sequential Maximum Angle Convex Cone (SMACC) Endmember Model," in: Proceedings SPIE, Algorithms for Multispectral and Hyperspectral and Ultraspectral Imagery, vol. 5425-1.

W. Kessler, "Multivariate Datenanalyse," Wiley-VCH Verlag, Weinheim, Germany, 2007.

A.J. Mills and M.V. Fey, "Declining soil quality in South Africa: effects of land use on soil organic matter and surface crusting," South African Journal of Science, vol. 99, p.p. 429-436, 2003.

A.J. Mills and R.M. Cowling, "Rate of Carbon Sequestration at Two Thicket Restoration Sites in the Eastern Cape, South Africa," Restoration Ecology, vol. 14, p.p. 38-49, 2006.

R. Richter, "Atmospheric / Topographic Correction for Airborne Imagery," DLR report DLR-IB 565-02/10, Wessling, Germany, 2010.

A. Stevens, B. van Wesemael, H. Bartholomeus, D. Rosillon, B. Tychon, and E. Ben-Dor, "Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils," Geoderma, vol. 144, p.p. 395-404, 2008.

R.A. Viscarra Rossel, "ParLeS: Software for chemometric analysis of spectrometric data," Cemometrics and Intelligent Laboratory Systems, vol. 90, p.p. 72-83, 2008.

R.A. Viscarra Rossel and T. Behrens, "Using data mining to model and interpret soil diffuse reflectance spectra," Geoderma, vol. 158, p.p. 46-54, 2010.

ACKNOWLEDGEMENTS

This PhD research project is part of the *Helmholtz EOS Network* as collaboration of German *Helmholtz* research centers. The *PRESENCE Network* (a *Living Lands* initiative), South Africa, and its partners support the efforts on-site.