

# SOIL SPECTROSCOPY AS A TOOL TO ASSESS ORGANIC CARBON, IRON OXIDES, AND CLAY CONTENT IN THE SUBTROPICAL THICKET BIOME OF EASTERN CAPE PROVINCE OF SOUTH AFRICA

Marco Nocita<sup>1,2</sup>, Martin Bachmann<sup>1</sup>, Andreas Müller<sup>1</sup>, Lammert Kooistra<sup>2</sup>, Mike Powell<sup>3</sup>, Silvia Weel<sup>3</sup>

<sup>1</sup> DLR-DFD German Aerospace Center, German Remote Sensing Data Center, 82234 Wessling, Germany

<sup>2</sup> Wageningen University, Centre for Geo-Information, Droevendaalsesteeg 3 6708 PB Wageningen, The Netherlands

<sup>3</sup> Presence-Earthcollective, PO Box, 237 Patensie 6335, Eastern Cape, South Africa

[marco.nocita@dlr.de](mailto:marco.nocita@dlr.de)

**Keywords:** soil spectroscopy, land degradation, organic carbon, iron oxides, clay content, thicket biome

## Abstract

In the subtropical thicket biome of the Eastern Cape province of South Africa, heavy browsing by goats, which, remove shrub biomass more rapidly than it is replaced, transforms the dense closed-canopy shrubland into an open savanna-like system. This transformation causes a lot of changes, among which, soil fertility depletion.

This document presents a project dealing with organic carbon, iron oxides, and clay content assessment, in the degraded thicket biome, through the combination of soil spectroscopy and partial least square regression (PLSR) techniques. The study area is a transect crossing in direction south east-north west the Eastern Cape province of South Africa, from latitude -33.57 to -32.59 and longitude 25.38 (eastern extreme) to 25.26 (western extreme). The study area has been selected based on a GIS analyses, realized overlayng Vegetation Type, Rainfall and Topografy data sets. 113 points have been visited over a distance of 130 Km. At every point field spectroscopy measurements have been realized and soil samples of the first cm (topsoil) and of the 0-20 cm have been collected. The soil samples have been chemically and spectrally analyzed. The present study models the relationships between soil spectral reflectance's, measured in situ and in the laboratory, and the soil parameters taken in consideration. Results indicated as soil stoniness is an important variables to consider for the soil properties prediction models creation. The PLSR models developed with laboratory and field spectra offered very good results for the prediction of OC (Calibration and validation  $R^2 > 0.7$ , RMSEV < 0.6), and mediocre results for the [Fe] prediction (RMSEV always > 0.55). The clay content prediction models production is at a preliminary stage. Based on this research, future studies, based on the up-scaling process of the obtained regression models to air-borne and space-borne HyperSpectral device, will be realized.

## 1. Introduction

The estimation of soil properties, such as organic carbon (OC), iron oxides, and clay contents, are important not only to quantify soil fertility, but also basic factors to understand the soil capability to resist to phenomenon of land degradation. Shepherd and Walsh (2002) pointed out the tremendous need for new techniques to

measure soil properties that are faster and cheaper than the time consuming and expensive conventional soil laboratory methods.

In the last years, it has been shown that spectroscopy in the visible-near infrared (VNIR) regions offers a rapid, inexpensive and non destructive technique to quantify soil properties, with comparable quality; moreover VNIR spectroscopy, requiring

less sample preparation, with less or no chemical reagents, is highly adaptable to automated and in situ measurements, and has the potential to analyze various soil properties simultaneously (McCarty et al., 2002; Viscarra Rossel et al., 2006).

For example, under laboratory conditions, coupling visible-near-infrared (VNIR) spectroscopy with multivariate calibration, can accurately determine the organic matter (Reeves et al., 2002; Salgo et al., 1998) and clay content (Ben-Dor et al., 1995; Kooistra et al., 2001) of the soil.

Most VNIR studies are conducted under controlled laboratory conditions, but investigations done in situ (Daniel et al., 2003; Kooistra et al., 2003) have produced promising results, allowing the development of models that can be extended to data from airborne and space-born hyperspectral imaging spectrometers, favouring the estimation of soil parameters for large areas instead of just point measurements.

The objective of this study is to calibrate and validate multiple regression models for the prediction of soil OC, iron oxides and clay content, through the application of soil spectroscopy techniques, along a transect crossed in the subtropical thicket biome (STB) of the Eastern Cape Province of South Africa

The STB of South Africa is centred in the south-western part of the Eastern Cape Province (33°S, 25°E).

The vegetation is characterized by a matrix of succulent shrubs (e.g. *Portulacaria afra*), spinescent shrubs (e.g. *Azima tetracantha*, *Gymnosporia polycantha*, *Putterlickia pyracantha*, *Rhus longispina*) and low-growing trees (<5 m) (e.g. *Pappea capensis*, *Euclea undulata*, *Schotia afra*) (Skead 1987; Kerley et al. 1999).

Despite a long association with indigenous large herbivores (Midgley 1991; Kerley et al. 1995), thicket is surprisingly sensitive to injudicious pastoralism (Stuart-Hill 1992). During the 20<sup>th</sup> century, heavy browsing by goats transformed the dense closed-canopy shrubland into an open

savannah-like system with a cover of ephemeral grasses and forbs.

This transformation caused a lot of changes, as the reduction of available water due to the diminishing below-ground biomass, the extreme carbon storage decrease due to the frequent fires (avoided by thicket vegetation) and the permanent plant cover degradation (causing soil temperature boost), and the crushing of the soil structure (Mills and Fey, 2004).

In order to arrest these degradation processes, the subtropical thicket restoration project (STRP), initiated by the Working for Woodland program of the South African Government, points to re-establish the vegetation mosaic and the soil fertility level, developing cost effective methods to check, for example, soil components content monitoring.

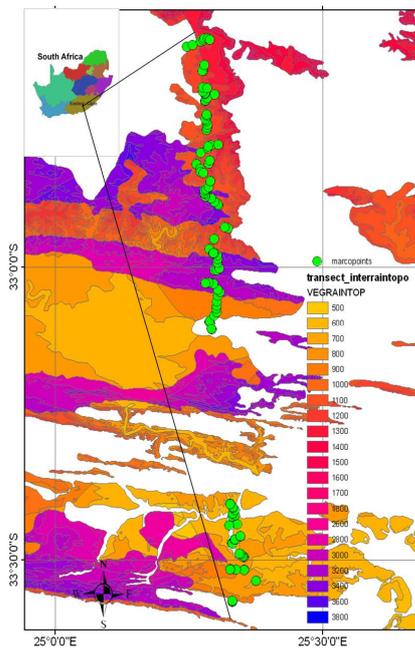
This paper presents the results of a study developing soil properties small scale prediction model, as the base research for future up-scaling processes, to all the subtropical thicket biome, will be realized with air-borne and space-borne hyperspectral data collection campaigns.

## 2. Methodology

### 2.1 Study area

The study area is a transect crossing in direction south east-north west the Eastern Cape province of South Africa from latitude -33.57 to -32.59 and longitude 25.38 (eastern extreme) to 25.26 (western extreme). The selection of the study area was based on a GIS analyses which took in consideration three available datasets: Vegetation Type, Rainfall and Topography. Vegetation type and topography datasets refers to the STB, giving 21 vegetation types, and an altitude range from 100 to 1100 m a.s.l. (11 classes), respectively. Rainfall dataset covered the full Eastern Cape province giving a rainfall range between 200 to 1000 mm/yr (8 classes). The three datasets have been overlaid, obtaining a stratified map, with 31 biome classes. The transect covered 21 out of 31 biome classes. 113 points have been visited over a distance of 130 km (figure 1).

The choice to select a transect as study area, along the STB, was based on i) the possibility to cover the biggest possible number of classes originated with the GIS analyses, and ii) on the future flight campaign, for the collection of hyperspectral data, which will follow the same line traced by the studied transect; in fact, the next step will be to use the soil spectral library created for this project, to build soil properties prediction models, which will be up-scaled to airborne hyperspectral data.



**Figure1:** map of the study area

## 2.2 Data collection

The data, used for the prediction models, have been collected during the soil sampling and the field spectroscopy campaigns.

### 2.2.1 Soil sampling campaign (SSC)

During the SSC, 113 soil samples have been collected according to the following scheme:

a 20x20x20 cm hole has been dugged, the soil collected in one bag, well homogenized and divided in two bags, one with soil destined to soil chemical analyses, and one for soil laboratory spectral analyses.

As stated before, the soil collected was referred to a depth up to 20 cm, but in some cases, due to shallow stones, it was not possible to reach the 20 cm depth.

Land use, vegetation type, soil properties, and root content have been estimated, for very visited point

### 2.2.2 Field spectroscopy campaign (FSC)

The FSC characterized the second part of the fieldwork.

One month after the SSC, 111 plots, out of 113 sampled during the SSC, have been visited to collect with an ASD Fieldspec-Pro radiometer, in 1 nm steps in the 350-2500nm wavelength range, soils spectral reflectance, using a contact probe device in order to eliminate the effect of vegetation on the soil spectra collected.

For every plot, a 50x50 cm plot has been defined, and ten spectral measurements have been collected, 5 with stones layer and 5 without. The five measurements, to repeat twice, followed the scheme 1 in the centre and four at the corner of the plot.

Afterwards, from the same spot, a sample has been taken from the upper surface (till 1cm), homogenized, and divided in two bags, one for chemical analyses and one for laboratory spectral analyses.

## 2.3 Soil samples analyses

All the soil samples collected during SSC and FSC have been chemically analyzed in order to establish OC, iron oxides and clay contents.

OC was determined using the Walkley and Black method (Walkley and Black, 1934). The mechanical clay content has been identified with the segmentation procedure (grain size < 2 mm) (Baize, and Jabiol, 1995), while the iron content has been assessed measuring the concentration of the dithionite-extractable iron oxides (Agbnenin, 2003).

The same samples collected during SSC and FSC have been spectrally measured under laboratory controlled conditions, with the same spectrometer used in the field. The topsoil samples have been analyzed with and without stones, in order to determine the influence of the stone layer on the soil spectral reflectance, as assessed with the field spectrometer, while the 0-20 cm soil samples have been analyzed without stones. The samples without stones have been analyzed after creating sub-sampling portions of the ground (< 2 mm soil, ~20 g) (Viscarra-Rossell, 2005). All the samples have been illuminated with two-quartz halogen lamps (1000 W each), mounted on a tripod of zenith angle of 30°. The reflected light has been assessed in nadir position. 4 measurements have been taken, rotating clockwise the sample of 90°.

#### 2.4 Models construction

All the spectra collected in the field and in the laboratory have been corrected for the ASD “jump” at 1000 nm (additive correction method), for spectralon reflectance, and averaged for subsequent proceedings.

Based on chemical analyses results, the dataset has been divided in training (2/3) and test set (1/3).

Prior to performing the statistical analyses, spectra from 350 to 399 nm, from 796 to 814 nm, and from 2401 to 2500 nm have been excluded, as insensitive or influenced by artefacts produced by spectrometer (Viscarra-Rossell et al., 2006).

Several pre-processing techniques, commonly used in soil spectroscopy, have been applied for the enhancement of spectral features. Calibration models for OC, iron oxides, and clay content have been developed both without pre-processing and applying several data manipulation: transformation of Reflectance (R) spectra in log (1/R); spectral normalization performed using multiplicative spectral correction (MSC) (Geladi et al., 1985); random noise reduction and signal to noise ratio (SNR) improvement realized using the Savitzky-Golay filter (Savitzky and Golay, 1964), with a second order polynomial fit and a variable window size of either 3, 6, and 10; spectral resolution enhancement and background effect elimination with first derivative application; data pre-treatment using Mean-Centre function.

Once data pre-processing has been completed, the models have been developed using Partial Least Square Regression (PLSR) techniques (Cozzolino and Moron, 2003). The number of factors to take in consideration for the PLSR analyses has been decided based on a leave one-out cross-validation (CV) approach to the training set (Reeves et al., 2002). For the selection of an optimal, parsimonious PLSR model different factors have been taken into account: the root mean squared error of the cross validation (RMSECV), for the accuracy of CV; coefficient of determination of the cross validation (R<sup>2</sup>CV) (Wold and Sjöström, 2001); Akaike Information Criterion (AIC), representing the variability in the data without causing it to overfit (Li et al., 2002); the smallest possible number of factors.

The test set has been used for the models validation and; R<sup>2</sup> between measured and predicted values for the soil parameters, and root mean square error of the validation (RMSEV) have been used to evaluate the established model (Kooistra et al., 2003).

For topsoil samples, superficial stoniness effect on spectral reflectance, has been

tested for both laboratory and field spectral data.

Models accuracies have been tested both including and excluding water absorption bands (WAB) in all the spectral datasets. ParLes 3.1 is the software used to develop the models (Viscarra-Rossel, 2008).

### 3. Results and discussion

#### 3.1 Chemical analyses

The chemical analysis results for topsoil (table 1) and 0-20cm (table 2) samples

show wide ranges of

concentration. This explains the choice to create 3 different training and test set for both topsoil and 0-20cm soil samples and to build independent

prediction models for the three analyzed properties.

Quite good correlations exist between OC and Iron

oxides concentration ([OC], [Fe]) in topsoil and 0-20cm

layer, indicating that topsoil could be representative for the deeper layer for these two properties (table 3). The weak correlation between Topsoil and 0-20cm clay concentration [Clay] could be caused by several factors as land-use, characterized in this area by either goat farming or game reserve activities. Soil is not undergone to intense ploughing activities, which could cause the stability of [OC] and [Fe] from Topsoil to 0-20cm, but significant differences of [Clay], favoured by alluviation deposit phenomena, and constant transport of small particles from the topsoil till the under layer.

#### 3.2 Interpretation of soil spectral reflectance.

In order to resume the differences of spectral reflectance between i) laboratory and field spectra, ii) topsoil and 0-20 cm soil samples, and iii) samples with and without stones, mean spectra have been calculated (fig. 2).

**Table 1: topsoil chemical analyses results**      **Table 2:0-20cm chemical analyses results**

0-20 cm					
OC%	Plot	Clay%	Plot	Fe%	Plot
0.2	65	3.94	85	1.40	15
0.21	17	4.02	46	1.48	21
0.22	37	4.93	10	1.50	31
0.23	39	4.95	17	1.53	86
0.24	18	5.04	20	1.73	92
0.29	10	5.21	88	1.80	23
0.32	24	6.14	41	1.80	30
0.32	85	7.01	110	1.80	32
...	...	...	...	...	...
...	...	...	...	...	...
3.29	6	26.80	86	4.45	35
3.55	34	27.58	55	4.63	4
3.56	31	33.37	82	4.63	8
3.6	28	36.95	78	5.03	3
4.82	32	39.07	103	5.05	36
5.05	33	44.60	7	7.28	1

Topsoil					
OC%	Plot	Clay%	Plot	Fe%	Plot
0.18	39	0.90	18	0.73	18
0.21	35	1.10	85	1.35	17
0.21	85	1.90	10	1.48	10
0.22	18	1.92	106	1.48	106
0.24	65	1.94	56	1.53	24
0.24	91	2.93	17	1.53	47
0.25	36	2.94	54	1.53	64
0.25	106	2.94	110	1.63	31
...	...	...	...	...	...
...	...	...	...	...	...
4.21	34	23.49	82	4.10	35
4.32	6	26.25	98	4.18	3
4.32	32	26.92	97	4.18	37
5.92	33	31.45	68	4.25	33
5.92	47	32.25	60	6.43	1
6.03	78	39.15	26	6.90	26



		Long				factors						
		(1/R)										
0-20	76	37	x	SG (2, 6)	1st deriv	Mean center	0.86	0.359	5	0.8	0.8	0.356
0-20 (NWAB)	76	37	x	SG (2, 6)	1st deriv	Mean center	0.86	0.359	5	0.8	0.8	0.354
TLNS	75	36	x	SG (2, 6)	1st deriv	Mean center	0.87	0.458	5	0.8	0.8	0.41
TLNS (NWAB)	75	36	x	SG (2, 6)	1st deriv	Mean center	0.86	0.464	4	0.8	0.8	0.405
TFNS	75	36	x	SG (2, 6)			0.81	0.557	8	0.8	0.8	0.432
TFNS (NWAB)	75	36	x	SG (2, 6)		Mean center	0.81	0.547	9	0.7	0.7	0.469
TFS	75	36	x	SG (2, 6)	1st deriv	Mean center	0.71	0.685	5	0.7	0.6	0.578
TFS (NWAB)	75	36	x	SG (2, 3)	1st deriv	Mean center	0.71	0.681	7	0.7	0.7	0.588

**0-20: 0-20cm lab spectra model; (NWAB): NO water absorption bands; TLNS: topsoil lab spectra model without stones; TFNS: topsoil field spectra model without stones; TFS: topsoil field spectra model with stone**

**Table 5: calibration and validation of iron oxides prediction models**

spectral data			data manipulation				Calibration			Validation	
	training	test	R to Log (1/R)	LS and BL corr	De-noising	Pretreatment	R <sup>2</sup> CV	RMSECV	PLSR factors	R <sup>2</sup>	R <sup>2</sup> adj
0-20	76	37	x	MSC		MC	0,151	0,871	4	0,285	0,265
0-20 (NWAB)	76	37	x	MSC	SG (2,10)	MC	0,119	0,876	4	0,246	0,224
TLNS	75	36	x	MSC	SG(2;10)	MC	0,188	0,882	4	0,211	0,188
TLNS (NWAB)	75	36	x	MSC		MC	0,176	0,898	5	0,261	0,240
TFNS	75	36	x	MSC		MC	0,205	0,913	6	0,288	0,267
TFNS (NWAB)	75	36	x	MSC		MC	0,185	0,901	6	0,358	0,339
TFS	75	36	x			MC	0,138	0,912	6	0,283	0,262
TFS (NWAB)	75	36	x			MC	0,129	0,918	6	0,221	0,198

**0-20: 0-20cm lab spectra model; (NWAB): NO water absorption bands; TLNS: topsoil lab spectra model without stones; TFNS: topsoil field spectra model without stones; TFS: topsoil field spectra model with stones**

Topsoil field no stones (TFNS) model without WAB was built including 9 factors. It gave less error than the TFNS with WAB, but the latter one was built with 8 factors and, once validated, produced less error and higher  $R^2$  than the former one.

TFNS without WAB received mean centre pre-treatment while TFNS with WAB was not pre-treated.

Topsoil field stone (TFS) and TFS without WAB spectra, after receiving the same data manipulation, produced similar results both for calibration and for validation. TFS without WAB was built with 7 factors while TFS only with 5 factors (table 4).

All the OC models developed produce quite good results. Models developed with spectra collected in the laboratories seem to be more accurate (lower RMSE) than the ones developed with field data. However, results obtained with models developed with field spectra assure a quite good level of prediction capacity, with  $R^2$  values both for calibration and for validation always  $> 0.7$ .

### 3.1.1 Iron Oxides

The models developed to predict Iron Content (table 5) didn't produce as good results as OC prediction models.

Calibration models produced similar accuracies for laboratories and field spectral data, with a RMSECV always lower than 1.

Models developed with laboratory data were built with fewer factors than the one produced with field data. There are not significant differences between laboratories and field models produced considering or excluding WAB. Once validated, the models which produced the higher accuracies (lower RMSEV) were the ones produced in the field (the best Iron oxides prediction model is tfnos without WAB with a  $R^2$  of 0.358), indicating no differences when stoniness is a factor included in the data analyses (table 5). In order to improve the prediction models accuracies, more analyses,

selecting, for example, specific explaining wavelengths, need to be done

### 3.1.2 Clay content

The clay content prediction models production is at a preliminary stage. At the moment, we are not able to present the definitive results, due to difficulties encountered during the data analyses.

## 4. Conclusions

This study examines the possibility to predict topsoil and 0-20cm layer OC, Iron Oxides and Clay content combining field spectroscopy and PLSR techniques.

The results of the chemical analyses indicate that topsoil and 0-20cm [OC] and [Fe] are correlated (Paerson's correlation coefficient= 0.77 and 0.68, respectively) while for [clay] there is very low correlation value (table 3).

The mean soil spectra calculated demonstrated there are differences of reflectance values between soils measured with and without stones, indicating as stoniness should be considered as an important factor, especially in relation to the up-scaling process to airborne and space-borne hyperspectral data (figure 2).

The PLSR models developed with laboratory and field spectra offered very good results for the prediction of OC (Calibration and validation  $R^2 > 0.7$ , RMSEV $<0.6$ ), and mediocre results for the [Fe] prediction (RMSEV always  $>0.55$ ) (table 4 and 5). These results indicate that, for the STB, combining soil spectroscopy and PLSR, does favour an accurate prediction of OC, while further investigation need to be realized to improve Iron oxides prediction models.

## 5. Acknowledgements

The authors would like to thank PRESENCE and Earthcollective for the assistance and dedication during the South African time. Moreover special thanks go to the German Aero-Space Agency to

believe in the project and to strongly support it technically and economically,

## 6. References

- Agbnenin, A., 2003. Extractable Iron and Aluminum Effects on Phosphate Sorption in a Savanna Alfisol. *Soil Science Society of America Journal* 67:589-595
- Baize, D., Jabiol, B., 1995. Guide pour la description des sols. Paris, INRA edition
- Ben-Dor, E., Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal* 59, 364– 372
- Cozzolino, D., Moron, A., 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *Journal of Agricultural Science* Volume 140, pages 65-71
- Geladi P., MacDougall D., Martens H., 1985. *Applied Spectroscopy* 39 (1985) 491.
- Kerley, G.I.H., M.H. Knight, and M. De Kock. 1995. Desertification of subtropical thicket in the Eastern Cape, South Africa: are there alternatives? *Environmental Monitoring and Assessment* 37:211-230.
- Kerley, G.I.K., A.F. Boshoff, and M.H. Knight. 1999. Ecosystem integrity and sustainable landuse in the Thicket Biome, South Africa. *Ecosystem Health* 5:104-109.
- Kooistra, L., Wehrens, R., Leuven, R.S.E.W., Buydens, L.M.C., 2001. Possibilities of visible near infrared spectroscopy for the assessment of soil contamination in river floodplains. *Anal. Chim. Acta* 446, 97–105.
- Kooistra, L., Wanders, J., Epema, G.F., Leuven, R.S.E.W., Wehrens, R., Buydens, L.M.C. 2003. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Anal. Chim. Acta* 484, 189–200.
- Li, B., Morris, J., Martin, E.B., 2002. Chemometrics and intelligent laboratory systems. 24 (2002) 79
- McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* 66, 640–646
- Midgley, J.J. 1991. Valley Bushveld dynamics and tree euphorbias. In *Proceedings of the first Valley Bushveld/Subtropical Thicket Symposium*, ed. P.J.K Zacharias, G.C. Stuart-Hill, and J.J. Midgley, 8-9. Special Publication, Grassland Society of Southern Africa.
- Mills, A.J., and M.V. Fey. 2004. Soil carbon and nitrogen in five contrasting biomes of South Africa exposed to different land uses. *South African Journal of Plant and Soil* 21:81-90.
- Reeves III, G. McCarty, T. Mimmo., 2002. The potential of diffuse spectroscopy for the determination of carbon inventories in the soil. *Environ. Pollut.* 116. S277-284
- Salgo, J. Nagy, J. Tarnoy, P. Marth, O. Palmai, G. Szabo- Kele, J. Near Infrared Spectrosc. 6, 1998, 199.
- Savitzky A., Golay M.J.E., 1964. *Analytical Chemistry* 36 (1964) 1627.
- Skead, C.J. 1987. Historical mammal incidence in the Cape Province. Volume 2: The eastern half of the Cape Province, including the Ciskei, Transkei and Griqualand. The Chief Directorate Nature and Environmental Conservation of the Cape of Good Hope, Cape Town, South Africa.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.* 66, 988–998.
- Stuart-Hill, G.C. 1992. Effects of elephants and goats on the Kaffrarian succulent thicket of the Eastern Cape, South Africa. *Journal of Applied Ecology* 29:699-710.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131, 59–75.

Viscarra Rossel, R.A. , 2008. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and*

*Intelligent Laboratory Systems*, 90: 72-83. doi:10.1016/j.chemolab.2007.06.006

Walkley, A., Black, I.A., 1934. An estimation of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science* 37, 29–37.

Wold, S., Sjöström, M., 2001. *Chemometrics and intelligent laboratory systems*. 58 (2001) 109