

# Visualizing small scale variability of clay content on soils at Mt. Kilimanjaro by VIS-NIR spectroscopy

Anna Kühnel<sup>1\*</sup>, Christina Bogner<sup>2</sup>, Holger Pabst<sup>3</sup>, and Bernd Huwe<sup>1</sup>

<sup>1</sup>Soil Physics Group, BayCEER, University of Bayreuth, Germany

<sup>2</sup>Ecological Modelling, BayCEER, University of Bayreuth, Germany

<sup>3</sup>Department of Soil Science of Temperate Ecosystems, University of Göttingen, Germany

## Abstract

We investigated the feasibility of VIS-NIR reflectance spectra to predict clay content for different land-use forms in situ. We used partial least square regression on an independent validation dataset and root mean squared error and the Akaike information criterion to evaluate our model. With this model we predicted clay content in four soil profiles on a 3 x 3 cm scale. Models performed well for spectra taken in the laboratory (RPD > 2; R<sup>2</sup> > 0.76). The accuracy for in situ predictions however varies between the land-use forms and predictions are preliminary.

**Keywords:** agricultural soils; diffuse reflectance spectroscopy; spatial variation; clay

## Introduction

The conversion of natural or semi-natural ecosystems to anthropogenic land-use forms often results in degradation of soil quality and altered ecosystem functions like water and carbon storage or erosion control. To infer the implications of these changes, fast and accurate predictions are required. This is especially important for the sub-Saharan ecosystems where information on soil properties is still rather scarce. Visible (VIS) and near-infrared (NIR) spectroscopy is a fast method to predict various soil properties simultaneously at comparatively low costs and has been widely used under laboratory conditions (Chang et al., 2001; Viscarra Rossel et al. 2006; Awiti et al., 2008). Using VIS-NIR spectroscopy directly in the field is not yet as reliable (Morgan et al., 2009; Nocita et al., 2011). However, it provides a direct and non-destructive method, if stable models can be developed. The goal of this study is to visualize the small scale variability of clay content, in situ. We use VIS-NIR reflectance spectra of soil to build a model based on partial least square regression (PLSR) to predict clay content for different land-use forms.

## Materials and methods

### Study site

The study was conducted on the southern slopes of Mt. Kilimanjaro, Tanzania (3°4'33"S, 37°21'12"E). The natural ecosystem of the lowlands around Mt. Kilimanjaro (up to 1100m a.s.l.) is savannah that developed on superficial deposits from the volcano. Mean annual rainfall fluctuates between 400–900mm (Soini, 2005), the main soil type is Vertisol and the vegetation is dominated by *Balanitis aegyptiaca* and different *Acacia* species. The savannah ecosystem is threatened by the transformation into fields, as the increasing population needs arable land, where maize and sunflowers are grown.

The submontane zone, an area between 1100–1800m a.s.l. on the southern slopes of Mt. Kilimanjaro is mainly covered by homegarden ecosystems, a traditional agroforestry

system, where banana (*Musa* spp.) and coffee (*Coffea arabica*) trees are grown together with a variety of smaller crops (Fernandes et al., 1985) and coffee plantations. Mean annual rainfall is between 1200 and 2000mm (Soini, 2005), the main soil types of the higher elevations are Andosols, more weathered soils develop into Vertisols. We have selected four of the typical ecosystems, namely natural savannah, maize field, homegarden and coffee plantation. Soil under coffee plantation was described as Haplic Vertisol, soil under homegarden and savannah as Sodic Vertisol and soil under maize field as Thephric Cambisol (FAO, IUSS Working Group, WRB, 2007).

#### Soil sampling and laboratory analysis

In each of the four selected ecosystems a soil pit was dug to a depth of at least 100cm or until continuous rock was reached. One profile wall was carefully cleaned of roots and debris and a frame of 0.5 x 1m with 3 x 3cm segments was put on it. Each segment was then scanned with the contact probe attached to an Agrispec portable spectrometer (ASD, Boulder Colorado) in the spectral range of 350–2500nm in 1nm intervals. Small soil core samples (diameter 2.5cm) were taken for validation. For the model calibration, soil samples were collected from 25 different sites with a soil auger and different soil horizons were separated, resulting in 146 samples. All samples were oven-dried at 45° for 24h and sieved <2 mm. Clay content was measured using a Master Sizer S particle size analyzer.

#### Spectral measurements and model calibration

For spectral measurements a well-mixed aliquot of the dried sample was scanned with the same device as used in the field. The instrument was calibrated with a Spectralon® white tile prior to measurements. For each sample as well as for the calibration with the white reference 30 reflectance spectra were averaged to reduce the noise.

Each spectrum was corrected for the ASD offset between the detectors with the additive method (Becvar et al., (2006 - 2008)). Then, a wavelet transformation was performed and the spectra were transformed into absorbance values. Afterwards, noisy portions of the spectra were removed and only the range from 500 nm to 2400 nm was kept. The dataset was split into a calibration and a validation dataset by randomly choosing 3/4 for calibration. The number of components for the optimal PLSR model was chosen based on the leave-one-out cross validation. The root mean squared error of prediction

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (1)$$

the Akaike Information Criterion,

$$AIC = N \log(RMSE) + 2m \quad (2)$$

the coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

and the ratio of percent deviation

$$RPD = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{RMSE} \quad (4)$$

were calculated, where N is the number of samples, m is the number of model parameters,  $x_i$  is the predicted value,  $y_i$  is the observed value and  $\hat{y}_i$  is the mean of the observed values. The model with the lowest AIC was chosen, as AIC helps to select a model that represents the variability in the data without causing it to overfit (Viscarra Rossel, 2008). All analyses were performed in R (R Development Core Team, 2011).

## Results and Discussion

### Model calibration

We have chosen a model with 10 components. The means and the range of clay content in calibration and validation datasets were similar, with the validation dataset covering the whole range of measurements (Table 1). Other studies predicting the clay content showed slightly better  $R^2$  and RMSE values. Stenberg (2010), for example, analysed the effect of different pre-treatments of the samples on clay content and found  $R^2 > 0.86$ . Considering the classification of RPD values by Viscarra Rossel et al., 2006, our models performed well and quantitative predictions are very good (RPD > 2.0).

Table 1. Parameter of the PLSR model

Clay	Mean	Range	$R^2$	RMSE	RPD
Calibration	51.2	20.0 - 85.0	0.84	8.0	2.04
Validation	50.8	21.0 - 76.0	0.80	7.1	2.26

### Model validation

Correlations between a) the air-dried spectra taken in the laboratory and b) the spectra taken in the field and the measured values were calculated (Figure 1). When looking at all plots, the PLSR model for the prediction of clay content with the air dried spectra performed quite well ( $R^2 = 0.75$ ; RPD = 2.02), whereas predicting clay with field spectra resulted only in  $R^2$  of 0.27 and a RPD value of 1.18 (Table 2).

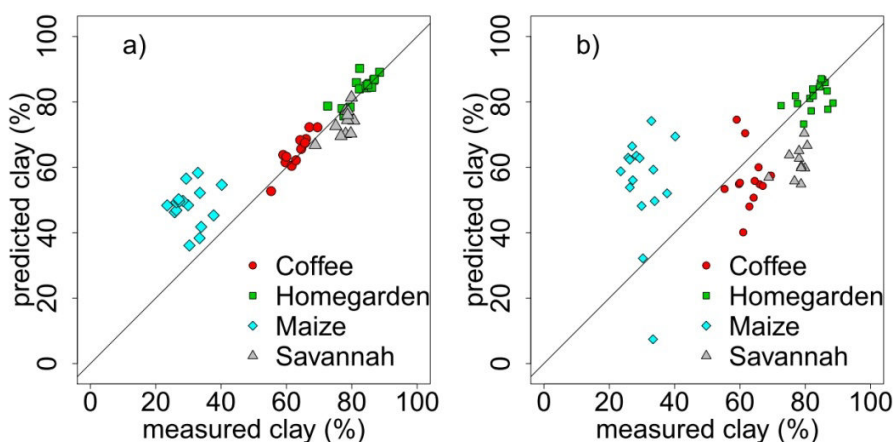


Figure 1. Predicted versus measured clay content for a) spectra taken under laboratory conditions and b) field spectra

Table 2. Validation parameters for laboratory and field predictions of clay content

Plot	Air-dried spectra			Field spectra		
	RMSE	RPD	$R^2$	RMSE	RPD	$R^2$
Homegarden	3.0	1.48	0.51	4.4	0.96	0.15
Coffee plantation	3.0	1.33	0.38	11.6	0.33	-8.8
Savannah	5.1	0.65	-1.64	16.7	0.2	-27.6
Maize field	19.3	0.24	-17.8	29.7	0.15	-43.6
All Plots	11.0	2.02	0.75	18.5	1.18	0.27

There are, however, large differences between the individual plots. The clay content of the homegarden, coffee and savannah profiles could be predicted quite accurately from air dried spectra. In contrast, the clay content in the maize field was poorly predicted. This could be due to a high amount of volcanic material in the soil. In our study we had only few samples of this material, so that it is probably under-represented in our calibration dataset. Due to the fact that scanning was conducted during the rainy season, field spectra were probably affected by the moisture content of the soil. As soil moisture has a strong influence on the reflectance spectra (Lobell et al., 2002), spectra taken under field conditions are often not reliable. Other factors influencing the spectra are the size and shape of the particles and the distribution of voids (Chang et al., 2001), thus smearing of clay during surface preparation or differences in bulk density could have an effect.

#### Small scale variability in the field

For each 3 x 3cm segment of the profiles clay content was predicted from field spectra with the respective model (Figure 2). Differences between the ecosystems are clearly visible. Soil in the homegarden ecosystem showed high clay content throughout the profile. In the maize profile the starting of the Cv-horizon at about -30cm was clearly visible. The accuracy of these predictions however is not yet satisfying and interpretations are preliminary.

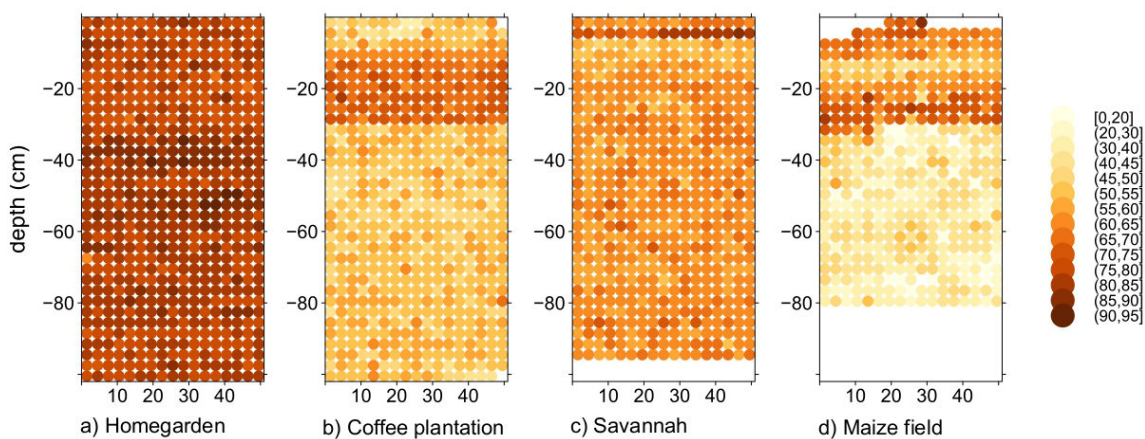


Figure 2. Small scale variability of clay content (%) in the different ecosystems

#### **Conclusions**

VIS-NIR spectroscopy is a fast and promising tool, but not yet applicable for detecting small scale differences in the field. Moisture content in the field and the different structure of the soil in situ compared to sieved samples needs to be taken into account. Possible solutions are preprocessing the calibration dataset with external parameter orthogonalisation, as suggested by Minasny et al., 2011 or to consider the difference between field and laboratory spectra. Whenever predictions are more accurate however, VIS-NIR spectroscopy can be used to assess the spatial organisation of soils rapidly and helps to understand the functioning of the soil within the ecosystem.

## Acknowledgements

This study was funded by the German Research Foundation (DFG) within the Research-Unit 1246 (KiLi) and supported by the Tanzanian Commission for Science and Technology (COSTECH), the Tanzania Wildlife Research Institute (TAWIRI) and the Mount Kilimanjaro National Park (KINAPA). Additionally we want to thank Johannes Hepp for assistance in the field and laboratory.

## References

- Awiti, A. O., M. G. Walsh, K. D. Shepherd, and J. Kinyamario (2008). "Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence". In: *Geoderma* 143.1-2, pp. 73–84
- Becvar M. and Hirner, A. and U. Heiden ((2006 - 2008)). DLR Spectral Archive. url: [http://cocoon.caf.dlr.de/intro\\_en.html](http://cocoon.caf.dlr.de/intro_en.html)
- Chang, C.-W., D. A. Laird, M. J. Mausbach, and C. R. Hurburgh (2001). "Near-Infra- red Reflectance Spectroscopy – Principal Components Regression Analyses of Soil Properties". In: *Soil Sci. Soc. Am. J.* 65.2, pp. 480–490
- FAO, IUSS Working Group, WRB (2007). World Reference Base for Soil Resources, first update 2007. World Soil Resources Report 103. Rome, Italy.
- Fernandes, E. C. M., A. Oktingati, and J. Maghembe (1985). "The Chagga homegardens: a multistoried agroforestry cropping system on Mt. Kilimanjaro (Northern Tanzania)". In: *Agroforestry Systems* 2, pp. 73–86
- Lobell, D. B. and G. P. Asner (2002). Moisture Effects on Soil Reflectance.
- Minasny, B., A. B. McBratney, V. Bellon-Maurel, J.-M. Roger, A. Gobrecht, L. Ferrand, and S. Joalland (2011). "Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon". In: *Geoderma* 167-168, pp. 118-124
- Morgan, C. L., T. H. Waiser, D. J. Brown, and C. T. Hallmark (2009). "Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy". In: *Geoderma* 151.3-4, pp. 249 –256
- Nocita, M., L. Kooistra, M. Bachmann, A. Müller, M. Powell, and S. Weel (2011). "Predictions of soil surface and topsoil organic carbon content through the use of laboratory and field spectroscopy in the Albany Thicket Biome of Eastern Cape Province of South Africa". In: *Geoderma* 167-168, pp. 295–302
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria.
- Soini, E. (2005). "Changing livelihoods on the slopes of Mt. Kilimanjaro, Tanzania: Challenges and opportunities in the Chagga homegarden system". In: *Agroforestry Systems* 64.2, pp. 157–167
- Stenberg, B. (2010). "Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon". In: *Geoderma* 158.1-2. Diffuse reflectance spectroscopy in soil science and land resource assessment, pp. 15 –22
- Viscarra Rossel, R., R. McGlynn, and A. McBratney (2006). "Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy". In: *Geoderma* 137.1-2, pp. 70–82
- Viscarra Rossel, R. (2008). "ParLeS: Software for chemometric analysis of spectroscopic data". In: *Chemometrics and Intelligent Laboratory Systems* 90.1, pp. 72–83