# Spectroscopy-supported digital soil mapping

V.L. (Titia) Mulder

**Thesis committee**

**Promotor**
Prof. Dr M.E. Schaepman
Professor of Remote Sensing, University of Zürich, Switzerland
Professor of Geo-information Science and Remote Sensing,
Wageningen University

**Co-promotor**
Dr S. de Bruin
Assistant professor, Laboratory of Geo-Information Science and Remote Sensing
Wageningen University

**Other members**
Prof. Dr P.C. de Ruiter, Wageningen University
Prof. Dr J. Wallinga, Wageningen University
Dr P. Lagacherie, UMR LISAH, INRA, Montpellier, France
Prof. Dr A. Stein, University of Twente, Enschede, The Netherlands

This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PE&RC)

# Spectroscopy-supported digital soil mapping

V.L. (Titia) Mulder

**Thesis**
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on 21 October 2013
at 4 p.m. in the Aula.

V.L. (Titia) Mulder
Spectroscopy-supported digital soil mapping
188 pages

PhD thesis, Wageningen University, Wageningen, NL (2013)
With references, with summaries in Dutch and English

ISBN 978-94-6173-690-1

# Table of contents

# Chapter 1

# General introduction

## 1.1 General introduction

In the epoch of the Antropocene, humans have become the main driver to global environmental change (Steffen et al., 2004). Human-driven changes such as population growth, land use change and greenhouse gas emissions have affected the functioning of the Earth system. This has resulted in pressure on food security and the necessity to adapt to climate change. As part of the Earth's spheres, the pedosphere is responding and contributing to these environmental changes (Macías and Arbestain, 2010). Observed changes in the functioning of the pedosphere renewed the recognition that soil resources provide key ecosystem services. Soils play a fundamental role for assuring food security and are essential to be considered in climate change adaptation (Global Soil Parnership, 2011; Grunwald, 2011). These recent developments also led to the establishment of the Global Soil Partnership (Global Soil Parnership, 2011). This partnership aims to enhance and apply knowledge of soil resources, improve their global governance and standardization.

The return of soils on the political and global research agenda (Hartemink, 2008) is also reflected by recent initiated research efforts which aim to respond to the need for up-to-date and high resolution soil information, thereby exploring new techniques and methodologies (Hartemink and McBratney, 2008). iSoil (van Egmond et al., 2009) and Digisoil (Grandjean and DIGISOIL Team, 2011) are examples of projects that focused on method development. Their main aim was to integrate new sensor technologies with soil science and digital soil mapping (McBratney et al., 2003). Research was conducted at small scale in order to fully exploit the data richness offered by both the sensors and soil legacy data. Other research efforts were more tailored to large-scale soil property mapping and harmonizing available soil data. Such research is crucial to the identification of threats to soil resources as identified by e.g. UN Convention to Combat Desertification (UNCCD) (United Nations, 1994) and the EU Soil Thematic Strategy (Commission of the European Communities, 2006). Currently available continental or global soil data are incomplete and only indirectly relate to problems of environmental change (Grunwald et al, 2011). For example, the information content of the digital soil map of the world (Fig. 1.1) provides the soil type along with specific profile descriptions that typifies the soil unit. Opposite to soil properties, it is generally found that this information does not provide direct information on soil resources. As a result, the GlobalSoilMap.net project (Sanchez et al., 2009) and e-SOTER (van Engelen, 2008) were initiated as the first steps to provide soil information for addressing the described current large-scale environmental issues. The GlobalSoilMap.net project aimed to make a new digital soil map of the world using state-of-the art technologies.

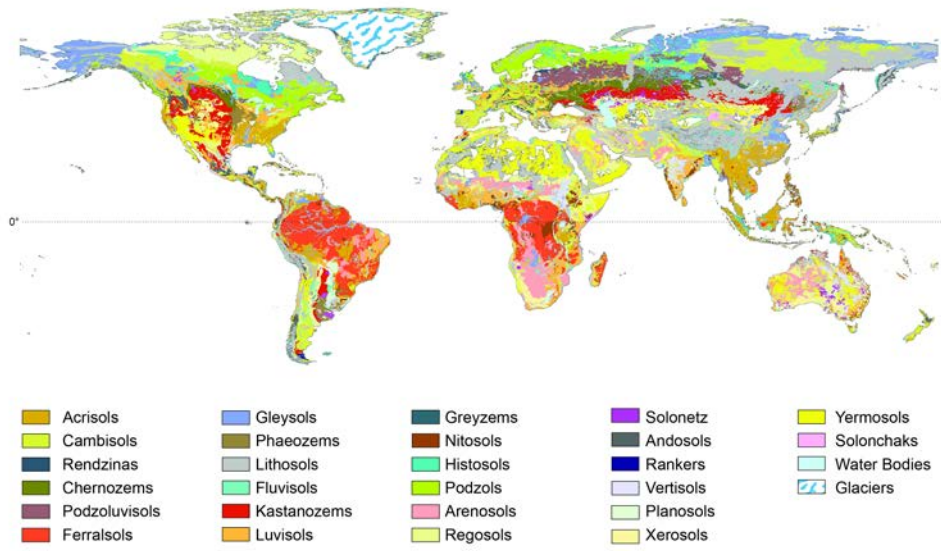


Figure 1.1: Digital soil map of the world, (FAO, 2007).

The research presented in this thesis was carried out in the context of the EU Seventh Framework Programme (FP7) project e-SOTER (van Engelen, 2008). The Group on Earth Observation (GEO) initiated a Global Earth Observation System of Systems (GEOSS). Within this framework, e-SOTER delivered a web-based regional pilot platform with soil and terrain data, methodologies and applications (Battrick, 2008). The project had two major research objectives. The first objective was to improve the current SOil and TERrain (SOTER) methodology (van Engelen and Wen, 1995) at scale 1:1 million, focusing on harmonization of existing soil legacy data and creating exhaustive soil maps using moderate-resolution optical remote sensing (RS) combined with existing soil legacy data. The second objective aimed to develop advanced RS applications within 1:250.000 scale study areas, including geomorphic landscape analysis and RS of soil attributes. This thesis is the result of the latter objective and contributed to the improvement of SOTER spatial and attribute data by developing methods to obtain semantic soil data from remote and proximal sensing (PS) data.

The remainder of this chapter describes the framework in which the work of this thesis was done. Section 1.2 introduces conventional and digital soil mapping (CSM and DSM respectively). Section 1.3 provides the basic principles of RS and PS with emphasis on the added value for DSM. In section 1.4, the objectives and research questions are defined while the final section provides the outline of this thesis.

## 1.2 Conventional and digital soil mapping

### 1.2.1 Conventional soil mapping (CSM)

One of the first soil scientists was perhaps Olivier Serries (1539 - 1619), who wrote *Théâtre d'Agriculture* (1600), a famous textbook of French agriculture. Vasily Dokuchaev (1846 - 1903) has been credited for laying the foundations of soil science at the end of the 19$^{th}$ century. Early 20$^{th}$ century, the expert on pedology Eugene Hilgard (1833 - 1916), was considered to be the father of 'modern' soil science in the USA. This modern soil science is nowadays referred to as CSM which matured in the second half of the 20$^{th}$ century (Schelling, 1970). The original methodologies were refined over time. At the end of the 20$^{th}$ century, the major principles and practices for making and using soil surveys were described in the standard from the Soil Survey Division Staff – the soil survey manual (Soil Survey Division Staff, 1993). CSM typically employs the free survey method to create soil maps. First, a mental soil-landscape model is made. Soil boundaries are defined based on landscape features from aerial photograph interpretations. Next, sample locations are selected that are likely to be most informative. The mental model is refined based on these field observations. Additional samples might be obtained and finally the map unit composition is determined. The map is a general-purpose map with soil classes and additional soil profile descriptions characterizing each map unit (Bregt, 1992). The CSM approach can result in accurate maps if the survey is well performed but there are some major limitations to CSM. The practical limitation to CSM, especially for large or inaccessible areas, is the intensive field work required to produce accurate and detailed maps (Bui et al., 1999). Apart from that, the soil surveyor's mental soil-landscape model is difficult to reproduce. Also, the discrete homogeneous polygons mapping units have a qualitative nature and lack quantified measures of accuracy (Kempen, 2011). Central to CSM is the description of the soil by a soil classification system (IUSS Working group WRB., 2006), which is a hierarchical and inflexible system. To address current environmental issues, more flexible and quantitative methods are required to study soils and their relation or function to environmental factors and threats (Bouma et al., 2012; Hartemink and McBratney, 2008).

### 1.2.2 Digital soil mapping (DSM)

DSM relies on field, laboratory and RS and PS soil observations, integrated with quantitative methods to infer spatial patterns of soils across various spatial and temporal scales (Grunwald, 2010). Using a broad range of data sources and methods, DSM aims to provide up-to-date and accurate soil maps to meet the current and future need for soil information. The DSM approach is both data and environmental-centred and so uses the data as a starting point to study the spatial distribution of

soils and soil properties. This makes DSM flexible and more suitable in providing soil information for specific applications compared to CSM.

The basis of DSM is the application of pedometric methods that predict the spatial and temporal distribution of soil types and soil properties. The conceptual framework in which the pedometric methods are applied is the *State Factor Equation of soil formation*, first introduced by Jenny (1941). This work states that soils can be described by the main environmental soil forming factors, which are: climate, organisms, relief, parent material and time (CLORPT). DSM uses this concept to develop empirical models that relate observations of soil properties with environmental variables that describe the main soil forming factors (i.e. CLORPT). Refinements of this modeling framework were made over the years, including the SCORPAN (McBratney et al., 2003) framework which is spatially explicit, and the STEP-AWBH (Grunwald, 2011) which is both spatially and temporally explicit. Typically, the environmental variables are exhaustive georeferenced data layers, including digitized geological and soil maps, satellite images and derivatives of the latter. There are no prerequisites on the type of model; regression models, regression trees and various other data mining techniques have been proven successful in establishing the statistical relations. Overall, DSM is indeed flexible, quantitative and accurate (Chapter 2 and references in there). Nevertheless, there are some critical points to consider. First, the models are typically not easy to transfer to other regions because the prediction models are based on the feature space of the study area which may not be applicable in another area. Secondly, compared to CSM products, DSM maps are developed for specific purposes rather than for general applications, which reduces its use to a limited public. Finally, DSM is not standardized and the use and interpretation of models by other users requires a clearly written report with supplementary information and instructions.

### 1.2.3 Spatial modeling of soil properties at regional scale

DSM successfully delivers up to date soil information at smaller scales (Gomez et al., 2012; Kempen, 2011). Still, difficulties remain for large-scale assessment with respect to data availability and scaling issues (Minasny, 2012; Minasny et al., 2013). In this work the following definitions of scale were used; local is smaller than $10^4$ $km^2$, regional varies between $10^4$ $km^2$ and $10^7$ $km^2$ and global is larger than $10^7$ $km^2$ (Mulder et al., 2011b). Larger spatial scales or studies refer to studies which are done at a scale that reaches beyond the local scale. Typically, large-scale assessments rely on relative sparse samples which impede the development of soil prediction models. The relation between the target and predictor variable tends to decrease with increasing scale and extent of the study area. Large-scale DSM is not straight forward and involves more than simply applying existing methods to large scale areas. Grunwald et al. (2011) emphasized that the assessment of spatial and temporal autocorrelations depend on the density and distribution of soil observations within a

landscape. Accordingly, the sample size of target variables and the sampled variability of the soil and environmental observations influence the accuracy of soil prediction models (Vasques et al., 2012). Advanced soil collection methods, suitable for large-scale studies, are needed to advance DSM predictions in space and time. It is expected that datasets, where the density and scale of soil observations more closely resemble the spatial resolution of the CLORPT factors, can elucidate the scaling behaviour of soil properties and processes across spatial and temporal scales. Apart from that, advancements to quantify soil properties and processes are expected in improved inference methods and less so on improved methodologies or models for predicting soil properties.

The use of geostatistical methods has some advantages as compared to the "CLORPT"-approach. Geostatistical methods are data driven rather than knowledge driven, and deal with geographical space rather than feature space. This allows for transferability to other areas, as a result. Furthermore, in addition to the structural part (drift or trend) that is modelled by the "CLORPT"-approach, the spatially correlated random part of variation is modelled. Here, the trend-component can be estimated using the auxiliary variables, using e.g. regression kriging (Hengl et al., 2007a). Modelling both the structural part and the random part of variation generally results in a higher prediction accuracy; kriging provides the best linear unbiased estimator of an unknown location along with the prediction uncertainty (Goovaerts, 1999). Overall, geostatistical methods allow for more in-depth analysis of prediction uncertainties and spatial processes compared to the 'CLORPT' approach. Despite the many pros for using geostatistical methods some major limitations for large-scale mapping exist. Calibration of geostatistical models, generally requires higher sampling densities and spatial dependence of the observations (Hengl et al., 2003). Furthermore, geostatistical methods involving kriging have practical limitations for large scale assessments. The most important limitation, besides data collection, is the computational feasibility for large datasets. Currently, promising research is done to ensure computation feasibility for large datasets (Katzfuss, 2011), which was especially helpful for global environmental and climate change studies (Cressie and Johannesson, 2008; De Jong et al., 2013; Furrer and Genton, 2011). They showed that the computational requirements are not necessarily the limiting factor. So far, few studies have used such approaches for soil prediction models and the soil science community could take such methods into consideration.

Concluding, to make large-scale DSM successful: (I) advanced sampling and inference methods are needed, (II) improving soil predictions requires compatibility between the variability within the sample and the predictor variables, (III) data driven geostatistical methods should be used to improve spatial predictions, and (IV) the methods must be both time and cost efficient. RS and PS methods could provide essential information to make regional-scale DSM successful, as this thesis will demonstrate.

## 1.3 Remote and proximal sensing of soils

### 1.3.1 Remote sensing

RS is about obtaining information from an object through the analysis of data acquired by a device that is not in direct contact with the object of interest. Examples of sensors used for such purpose include gravity meters measuring variations in force distributions, and sonar measuring variations in acoustic wave distributions. This thesis relied on sensors measuring variations in the emission and reflectance of electromagnetic (EM) energy by objects. EM energy sensors collect data about the emission and reflection of EM energy from objects. Sensors operating from satellite and airborne platforms can provide information on various earth surface features. This information supports inventorying, mapping and monitoring of earth resources, including soils. Note that in this thesis, the term RS refers to EM RS while other types of RS are not further discussed (Lillesand et al., 2008). The EM spectrum ranges from gamma (γ) rays at the shortest wavelengths to radio-waves at the longer wavelengths (Fig. 1.2). Most common sensing systems operate in one or several of the visible, infrared (IR) or microwave portions of the spectrum. Each part of the spectrum provides typical information for various earth resources. Sensors obtaining information from the γ-rays (Wilford et al., 1997), X-rays (Bish and Plötze, 2011) and the Mid-IR (Viscarra Rossel et al., 2006b) successfully retrieve soil information, especially soil mineralogy. However these sensors are not fully operational for spatial modelling at large scales. In this thesis, I limit myself to the use of the visible and near infrared (VNIR), shortwave infrared (SWIR) and thermal infrared (TIR) wavelength ranges. Sensor data covering these wavelengths are readily available from both satellite and airborne platforms (Lillesand et al., 2008).

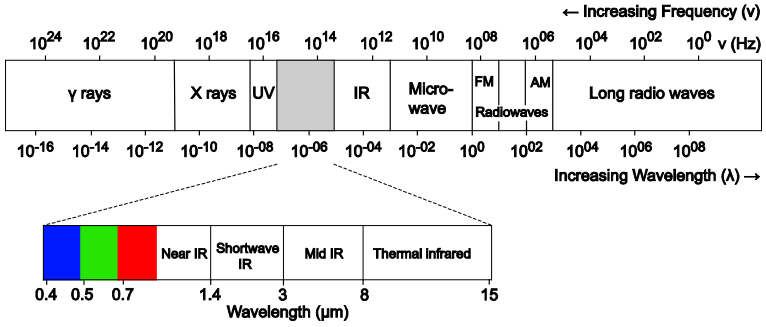


Figure 1.2: Electromagnetic spectrum (Lillesand et al., 2008).

### 1.3.2 Remote and proximal sensing methods

Deriving information from spectral data typically relies on methods such as (un)supervised classification (Friedl et al., 2002), spectral unmixing (Keshava and Mustard, 2002), band ratios (Hewson et al., 2012), absorption features analysis (Clark et al., 2003) and partial least squares regression (PLSR) (Viscarra Rossel et al., 2006b). Some spectral signatures, such as those from soil, water and vegetation, can be easily discriminated and (un)supervised classification usually results in reliable outputs. Nevertheless, satellite images take pixel-based measurements. These pixels seldom consist of a single constituent but represent a mix of multiple constituents. Fractional coverage estimates of pixel-constituents have been obtained using spectral unmixing (Keshava and Mustard, 2002). Alternatively, information from band ratios have been used for those constituents that show clear differences between spectral bands in relation to some physical property, e.g. vegetation properties (Tucker, 1979). Vegetation differentiates itself from other earth surface resources by the major difference in reflectance between the red and near-infrared wavelengths (Fig. 1.3a). Based on the difference between these spectral regions, several indices have been developed to retrieve specific information for vegetation properties (Tucker, 1979).

Quantitative information is typically retrieved from spectral data using radiative transfer modeling (Laurent et al., 2011) or absorption feature analysis (Clark et al., 2003; Sunshine and Pieters, 1998). This requires high-resolution spectral data to ensure correct detection of the specific spectral characteristics. Absorption feature analysis uses the continuum removal (CR) of the isolated feature to relate changes in depth, width or surface of these features to physical properties, e.g. clay content (Fig. 1.3b). An alternative statistical approach is PLSR, where spectral bands are used to model the relation between physical properties and reflectance (Viscarra Rossel et al., 2011). This has been useful for the retrieval of various soil properties but limitations occur when absorption features of constituents tend to overlap. In case absorption features overlap, the relation towards the physical property of interest becomes non-linear due to e.g. complex scattering behaviour of the constituents.

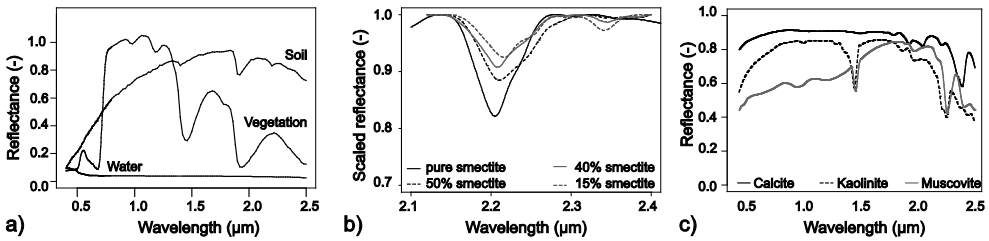


Figure 1.3: a) Spectra of soil, vegetation and water, (b) CR of SWIR reflectance of smectite spectra of varying abundances, (c) Spectra of various minerals (Clark et al., 2007).

Mineralogy is a clear example of a soil property for which the retrieval of quantitative information is limited using a strictly linear approach (Clark and Roush, 1984). Many mineral diagnostic absorption features occur in the SWIR (Fig. 1.3c); methods estimating mineralogy from a natural sample consisting of various minerals need to account for the non-linear behaviour in reflectance and overlapping absorption features. To do so, specialized modeling approaches are needed to obtain an estimate of abundance or characterization of the constituents.

### 1.3.3 Remote and proximal sensing for DSM

In this thesis, data collection by sensors on board of satellite or airborne platforms is referred to as remote sensing (RS). Those measurements that are taken at the field or laboratory level are referred to as proximal sensing (PS). An extensive review on the potential of RS and PS for soil attribute retrieval is given in Chapter 2 of this thesis. It was found that, RS provides exhaustive coverage of large areas, but with low spectral resolution. RS and derived products are therefore very suitable as predictor variables in DSM but less so to measure distinct absorption features. Satellite products such as Landsat Thematic Mapper (TM), Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER) and Moderate Resolution Imaging Spectroradiometer (MODIS) have been used as representatives for the soil forming factors soil, vegetation and parent material (CLORPT). In contrast, PS provides high spectral resolution data from individually sampled sites (Fig. 1.4). RS and PS are considered a cost and time efficient alternative compared to chemical and physical laboratory measurements to obtain soil property information from sampled sites. RS and PS are especially suitable for large-scale studies where soil legacy data is sparse.

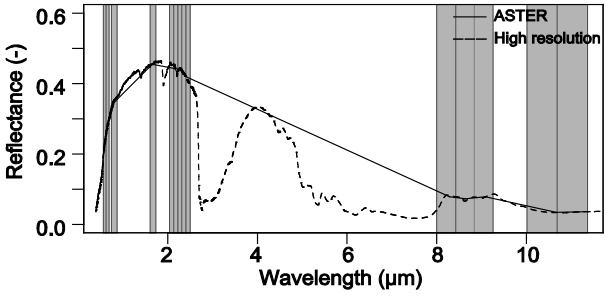


Figure 1.4: Spectral resolution of ASTER compared to laboratory spectroscopy (Clark et al., 2007).

Despite the many benefits that RS and PS methods provide for DSM, some critical issues need consideration. Optimal retrieval of detailed soil information requires operational methods and high spectral resolution imagery, yet, both criteria have not been met. Most state-of-the-art methods to retrieve soil properties are not publicly available. Suitable high spectral and spatial resolution imagery are currently not provided by satellite platforms. That is to say, no sensors are available with spectral bands covering the wavelengths required for optimal soil property retrieval. Using RS and PS is not straightforward and expert knowledge is required to maintain quality control. Incorrect measurements and pre-processing of data may easily lead to inaccurate results. Apart from these technical problems, a common criticism is that RS observations are limited to the first few centimetres of the surface. It is well known that many soil processes are best quantified using the full soil profile. Nevertheless, RS allows spatial explicit characterization of soil properties at the surface. Using pedotransfer functions, properties have been predicted for the full profile (Hong et al., 2013; McBratney et al., 2002). A combination of the latter two could result in a more refined description of soil properties within an area. Besides, agricultural activities surely benefit from estimated topsoil parameters to optimize crop growth conditions (White et al., 2012).

Following from sections 1.2 and 1.3, soil mineralogy became the central soil property of this thesis for two reasons. First, the retrieval of soil mineralogy from PS is known to be difficult due to the complex scattering behaviour within a sample (Clark and Roush, 1984). Improving methods for such properties would substantially improve the perspectives of using PS for DSM. Second, the presences of medium and long-scale dependent spatial processes were assumed necessary to model the major soil variability at regional scale using RS (section 1.2.3). Mineralogy meets the requirements to test this hypothesis. It is known that in large areas the spatial processes of mineral variability manifest themselves at shorter and longer spatial scales (Jaquet, 1989).

## 1.4 Objective and research questions

The soil science community is expected to deliver large-scale soil information maps that readily allow addressing the related environmental questions raised on the political agenda of countries, continental and global organizations. Soil scientists debate how to meet the demands from outside the community. Prominent soil scientists have provided the community with envisions of research in the near future (Bouma, 2009; Grunwald, 2011; Hartemink and McBratney, 2008). The ultimate goal would be to develop a global soil model which is sufficiently flexible to predict various soil properties at different spatial and temporal scales and in which existing soil data and environmental data are integrated in a holistic fashion. To achieve this

goal research priorities are among others, efficient data collection, spatial and temporal modeling of soil properties and solving scaling issues affecting large-scale assessments. Apart from that, harmonization of existing legacy data and improved transferability over different areas is needed. RS and PS could take a key position in providing data for the development of methods for large-scale modeling of soil properties. Especially with the advancement in computer and sensing technologies, methods for data collection and spatial modeling for large-scale areas might benefit from using spectral data. From this, the main objective of this thesis is to exploit the use of remote and proximal sensing methodologies for digital soil mapping in order to facilitate soil mapping at larger spatial scales.

To achieve this objective, an experiment at regional scale was designed. In this experiment, each step involved in the process of data collection, data retrieval from spectral data and spatial modeling of soil properties, was critically reviewed. Temporal modeling of soil properties and harmonization of legacy data was left aside. This stepwise approach allowed for the identification of the major bottlenecks which are addressed by the research questions outlined below.

1: *What is the current state-of-art in the use of remote sensing for soil and terrain mapping?*

This question is addressed by reviewing recent findings on the use of optical and microwave data for soil science applications with the emphasis on (I) spatial segmentation of the landscape, (II) measurements or prediction of soil properties by means of physically-based and empirical methods and (III) supporting spatial interpolation of sparsely sampled soil property data as a primary or secondary data source.

2: *Can major soil variability at regional scale be represented by a sparse remote sensing-based sampling approach?*

The collection of a representative soil sample for a large area, in the absence of legacy data and limited acquisition resources, demands flexibility and adaptation of sampling strategies. The challenge is to adequately plan the data collection within a sound theoretical framework using the full potential of the available RS data.

3: *Which methods allow retrieval of mineralogy from complex mixtures using proximal sensing?*

It was deemed necessary to demonstrate two recently proposed methods that allow advanced modeling of soil mineralogy using proximal sensing. The complex scattering behaviour of minerals within a soil sample may be addressed by model approaches that assume either linear or non-linear reflectance of the constituents. In this thesis, the opportunities and limitations for spectroscopic analysis of minerals

were evaluated under the assumptions of both linear and non-linear reflectance behaviour.

4: *Can scale-dependent variability be extracted from remote sensing and do model predictions improve by using scaled remote sensing data that match the variability of the sample?*

The aim is characterizing regional soil mineral composition using predictor variables derived from RS data. The sparse sample used in this work represents the major variability of soils but bears few compatibility in variability with the high-resolution RS data. The feasibility to predict soil properties using such a sample is likely to increase if the RS data represent a similar spatial scale of variability. The latter might be achieved by smoothing the RS data to medium- and long-range spatial structures, using a geostatistical approach.

## 1.5 Structure of this thesis

The key principles about soil mapping and RS and PS were described in this introductory chapter. Chapters 2 to 6 constitute the core of this thesis and address the objective and research questions presented in section 1.4.

Chapter 2 reviews previous studies related to DSM and RS and PS of soil properties. Here, the major opportunities and limitations of using spectral information for DSM were identified. This review was used as the starting point of this thesis. Chapter 3 describes a RS-based sampling approach for data collection in large areas. The sampling was constrained by lack of accurate soil legacy data and limited acquisition resources. This chapter shows how, despite these limitations, RS data allowed collecting a sample representing major soil variability at regional scale. This sample was further used for the research carried out in this thesis. Chapters 4 and 5 demonstrate two methods to retrieve soil mineralogy from PS. These chapters demonstrate advances in soil spectroscopy and also raise awareness to acquire soil information from PS. In Chapter 6, the previously described methods were integrated which allowed to characterize the regional soil mineral composition using spectroscopy and geostatistics. This chapter demonstrates the full potential of RS and PS for large-scale DSM. Chapter 7 concludes this thesis by summarizing and discussing the main findings in relation to the problem setting discussed in this chapter. Literature references have been combined in the *References* section at the end of the thesis.

# Chapter 2

# The use of remote sensing in soil and terrain mapping – A review

*This chapter is based on:*
V.L. Mulder, S. de Bruin, M.E. Schaepman, T.R. Mayr (2011). The use of remote sensing in soil and terrain mapping – A review. *Geoderma*, 162 (1-2), 1-19.
DOI: 10.1016/j.geoderma.2010.12.018

## Abstract

This article reviews the use of optical and microwave remote sensing data for soil and terrain mapping with emphasis on applications at regional and coarser scales. Remote sensing is expected to offer possibilities for improving incomplete spatial and thematic coverage of current regional and global soil databases. Traditionally, remotely sensed imagery have been used to support segmentation of the landscape into rather homogeneous soil–landscape units for which soil composition can be established by sampling. Soil properties have also been inferred from optical and microwave data using physically-based and empirical methods. Used as a secondary data source, remotely sensed imagery may support spatial interpolation of sparsely sampled soil property data. Soil properties that have been measured using remote or proximal sensing approaches include mineralogy, texture, soil iron, soil moisture, soil organic carbon, soil salinity and carbonate content. In sparsely vegetated areas, successful use of spaceborne, airborne, and in-situ measurements using optical, passive and active microwave instruments has been reported. On the other hand, in densely vegetated areas, soil data acquisition typically relied on indirect retrievals using soil indicators, such as plant functional groups, productivity changes, and Ellenberg indicator values. Several forms of kriging, classification and regression tree analyses have been used jointly with remotely sensed data to predict soil properties at unvisited locations aiming at obtaining continuous area coverage. We expect that remotely sensed data from existing platforms and planned missions can provide an important data source supporting digital soil mapping. Yet, most studies so far have been performed on local scale and only few on regional or smaller map scale. Although progress has been made, current methods and techniques still bear potential to further explore the full range of spectral, spatial and temporal properties of existing data sources. For example, spaceborne spectroscopy has been of limited use in retrieving soil data when compared to laboratory or field spectroscopy. To date, there is no coherent methodology established, where approaches of spatial segmentation, measurements of soil properties and interpolation using remotely sensed data are integrated in a holistic fashion to achieve complete area coverage. Such approaches will enhance the perspectives of using remotely sensed data for digital soil mapping.

## 2.1 Introduction

Soil and terrain information is needed for policy-making, land resource management, and for monitoring the environmental impact of development. Lack of comprehensive information about global, national or local land resources increases the risk of releasing uninformed policy decisions, avoidable continued degradation of land and water resources, and excessive carbon emission to the atmosphere and renders it finally less likely that the Millennium Development Goals will be achieved. The viability and cost of vital infrastructure is affected by this information shortage just as much as the food and water security and response to environmental change (van Engelen, 2008). Global and regional models that address climate change, land degradation and hydrological processes need soil input parameters with complete area coverage, but currently there are only few spatially exhaustive datasets available (Anderson et al., 2008; Bastiaanssen et al., 2005).

In recent decades the soil science community has made great efforts to develop regional and global soil and terrain databases. Currently, there are several georeferenced soil databases available at map scales smaller than 1: 250.000; namely the Harmonized World Soil Database at a map scale of 1:5 M (million) developed by the FAO-UNESCO (FAO et al., 2008); The European Soil Database at a map scale of 1:1 M, which is part of the European Soil Information System — EUSIS (Le Bas et al., 1998). The latter is the product of a collaborative project involving all the European Union and neighbouring countries, that has been active for the past 20 years (King et al., 1994). Further, the latest version of the European Soil Database (v2.0) includes an extended geometric component 'The Soil Geographical Database of Eurasia' (Lambert et al., 2002), which also covers the Russian Federation, Belarus, Moldova and Ukraine (Morvan et al., 2008); The Soil and Terrain Digital Database (SOTER), which incorporates quantitative information on soils and terrain at map scales 1:1 M and 1:5 M (Oldeman and van Engelen, 1993); Although partly implemented, the geo-referenced Soil Database for Europe at a map scale of 1:250.000, is an extendable database to which users can submit their local soil and terrain databases. For the latter, there is a manual aiming for consistence among soil surveyors (Finke et al., 2001). Other examples of soil databases with a continental scale are the SOTER database for different parts of Africa at a scale of 1:2 M (Dijkshoorn, 2003; van Engelen et al., 2006) and the SOTER database for Latin America and the Caribbean at a scale of 1:5 M (Dijkshoorn et al., 2005). There are many national soil databases such as the American Web Soil Survey (WSS) (Soil Survey Staff - Natural Resources Conservation Service) and the Soil Survey Geographic Data Base (SSURGO) from the Natural Resources Conservation Service (NRCS) (Soil Survey Staff - Natural Resources Conservation Service); the Australian Soil Resource Information System (ASRIS) from CSIRO Australia (Australian

Department of Agriculture Fisheries and Forestry); Available from the Agriculture and Agri-Food Canada: the Canadian Soil Information System (CANSIS) and the National Soil Database (NSDB) of Canada (Agriculture and Agri-Food Canada., 2010); and the Russian Soil map at a scale of 1: 2.5 M (Stolbovoi and McCallum, 2002). For an extended inventory of available soil databases we refer to Rossiter (2004) and Nachtergaele (1999).

The above suggests there is already much soil information available. Nevertheless, a major problem is inconsistency in data acquisition because the data have been collected nationally at various scales, using different standards and methods. Apart from that, developing and transitional countries typically lack digital and accessible soil information. Available data sets for these countries are mostly at small to medium scales and have been produced through international projects. Larger scale digital soil data are limited in availability to the USA, Canada, Australia and Europe. However, available soil databases mapped at large scale often have inconsistencies, e.g. the present geographical coverage for the European continent is uneven between and even within countries. National and regional European networks are much denser in northern and eastern regions than in southern Europe (Morvan et al., 2008).

Remote sensing (RS) may offer possibilities for extending existing soil survey data sets. The data it provides can be used in various ways. Firstly, it may help segmenting the landscape into internally more or less homogeneous soil–landscape units for which soil composition can be assessed by sampling using classical or more advanced methods. Secondly, RS data can be analysed using physically-based or empirical methods to derive soil properties. Moreover, RS imagery can be used as a data source supporting digital soil mapping (DSM) (Ben-Dor et al., 2008; Slaymaker, 2001). Finally, RS methods facilitate mapping inaccessible areas by reducing the need for extensive time-consuming and costly field surveys.

Although RS and soil spectroscopy have been recognized as a potentially effective and cost-efficient technology, they are not yet routinely used in soil surveys. Our knowledge of how to apply advances in RS to soil and terrain mapping is still incomplete (Ben-Dor et al., 2008). The ability to apply RS methods and improve coherence in soil and terrain mapping on a global scale, could contribute to the Global Soil Observing System, which is planned by the Global Earth Observation System of Systems (GEOSS) to meet the need for land resources information (Battrick, 2005). Using more coherent data sets with exhaustive coverage would also improve the identification of threats to soil quality as identified by e.g. UNCCD (United Nations, 1994), the EU Soil Thematic Strategy (Commission of the European Communities, 2006), the Canadian Soil Quality Program (Spiess, 2003), the United States Natural Resources Conservation Service (United States Department of Agriculture, 2006 ) and the Australian natural resource management (NRM) programs (Australian Government, 2010). RS has been used to identify these threats

and to support soil functional mapping such as water and nitrogen stress (Liaghat and Balasundram, 2010; Yi et al., 2008) and soil erosion (Ben-Dor et al., 2009; Metternicht and Fermont, 1998).

This paper aims to review publications from a wide range of sources and outlines a methodological framework that facilitates soil and terrain mapping from a soil survey-oriented view in combination with remote and proximal sensing (RS and PS) methodologies. The review focuses on the use of optical and microwave data for soil science applications. Airborne geophysical (e.g. magnetic, electromagnetic and radioactive) surveys have been used in geological and soil mapping (Martelet et al., 2006; Saunders et al., 1999; Wilford et al., 1997). However, the data used in these surveys are not as extensively available as optical and microwave data, which makes them less suitable for regional soil and terrain mapping.

The structure of this paper is based on the well-known State Factor Equation of soil formation, where soil is described as a function of CLimate, Organisms, Relief, Parent material and Time, referred to as CLORPT (Jenny, 1941) and its closely related SCORPAN soil spatial prediction function (SCORPAN – SSPFe). SCORPAN includes the same factors as CLORPT but also spatial (cross) correlation of soil properties and presence of spatially autocorrelated errors (McBratney et al., 2003). Several factors of soil formation can be derived from RS (Buis et al., 2009; French et al., 2005; Schmidtlein et al., 2007; Singhroy et al., 2003).

We review the use of RS and PS for (1) identifying any of the factors of soil formation to stratify the landscape, i.e. into large relatively homogeneous soil-landscape units which can be used as covariate for DSM or whose soil composition can be determined by classical sampling, (2) allowing measurement or prediction of soil properties by means of physically-based and empirical methods, and (3) supporting spatial interpolation of sparsely sampled soil property data as a primary or secondary data source. Note that in this review we use the term proximal sensing (PS) for laboratory and field measurements. The following definition of DSM is adopted: ‘the creation and population of spatial soil information by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems (Lagacherie et al., 2007; McBratney et al., 2003)’ (Carré et al., 2007).

## 2.2 Spatial stratification of the landscape

A common way of spatially segmenting the landscape is to divide it into internally more or less homogeneous and mutually contrasting landform units (Hewitt, 1993; Hudson, 1992). Soil–landform units are specialized landform units expected to be relatively homogeneous in terms of the main factors including parent material (Hengl and Reuter, 2009; McBratney et al., 2003). Soil-landform maps thus provide a tool for identifying locations where different geomorphic processes dominate. Landform

maps are typically suitable predictors of soil types because soil development often occurs in response to the underlying lithology and water movement in the landscape (Ballantine et al., 2005; McKenzie and Ryan, 1999). For classical soil mapping, such as the SOTER methodology, the landform maps can be used to draw soil boundaries. Within these units, soil samples can be taken and a soil type or soil association can be assigned to the different units (van Engelen and Wen, 1995). In DSM, soil landform maps may be used as auxiliary data source (discussed later).

### 2.2.1 Landform mapping

Traditionally, landform mapping is done by visually interpreting aerial photographs (Dent and Young, 1981). Nowadays, with access to fast computers and digital sources such as Digital Elevation Models (DEMs) – typically acquired by RS – it can be done digitally. Typically, the surface is parameterized by attributes such as elevation, slope, aspect, plan and profile curvature, and flow accumulation (Moore et al., 1993) to obtain relief or surface topography units (Fig. 2.1). These attributes quantify the role of topography in redistributing water in the landscape and in modifying the amount of solar radiation received at the surface which may affect the pedogenesis and thereby the soil characteristics (Wilson and Gallant, 2000). There are many definitions for landform mapping, as is described by Dehn et al. (2001).

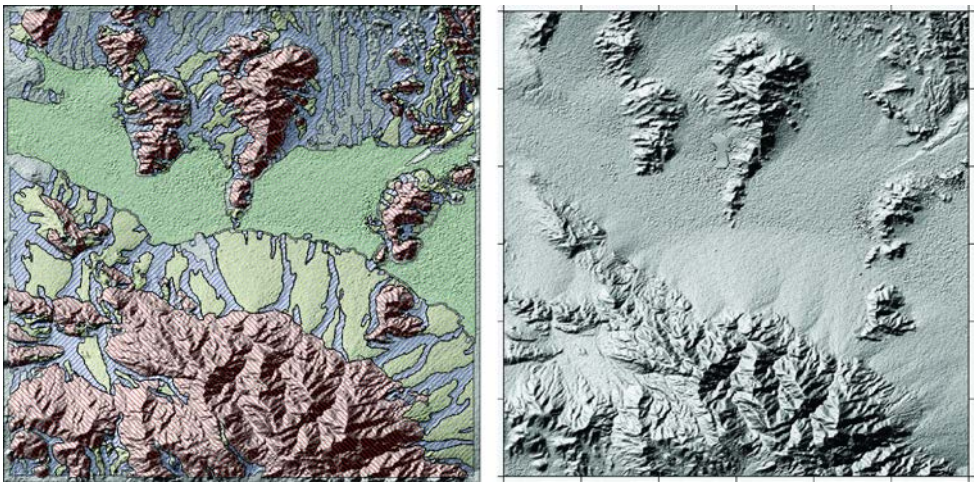

Figure 2.1: (Left) Landform delineations: red hatch – mountains, yellow hatch - alluvial fans, green hatch – valleys, blue hatch - wash drainage areas, unshaded - unclassified areas and (right) the corresponding shaded relief : 4x vertical exaggeration (right). (Reprinted with permission, Leighty (2004))

In this review the description of landform mapping given by Minár and Evans (2008 p. 1) is being adopted; Hence, the main objective of landform mapping is "The identification of the most specific geometrical geomorphic elements with maximal change of genetic, geometric and process character" (Minár and Evans, 2008 p. 1). In other words, the internal homogeneity and external contrasts of landforms in terms of their geometry should reflect their genesis and recent dynamic (Minár and Evans, 2008). Consequently, the land surface form is characterized by a complex structure of nested hierarchies of relief units (Dikau, 1989). Three types of relief units are distinguished, based on increasing complexity. Firstly, elementary forms, which represent the smallest and simplest geometric units. Secondly, the landforms that are composites of elementary forms and thirdly, the landform patterns which are associations of landforms (Minár and Evans, 2008). Most of these topographic attributes are calculated from directional derivatives of a DEM (Florinsky, 1998). Supervised or unsupervised classification for landscape segmentation can be performed using crisp or fuzzy clustering methods, ranging from local up to global studies (Dobos et al., 2005; Klingseisen et al., 2008; Moore et al., 2003; Prima et al., 2006; Schmidt and Hewitt, 2004; Tribe, 1992; van Asselen and Seijmonsbergen, 2006), as well as object-based landform mapping (Otto et al., 2010a;b; Raper and Livingstone, 1995).

There are several general problems which occur with the use of most automated landform classification methods. One of these problems concerns scale-dependence of the geomorphic elements that can be recognized. To determine the scale on which the desired elements can be retrieved, the algorithms should be applied over DEMs at various resolutions. Secondly, the definition of class boundaries and semantics may differ for different classification methods, which makes it difficult to compare them (Dehn. et al., 2001). Furthermore, the recognition of geomorphic elements is strongly dependent on the input data used, and it is influenced by DEM accuracy (outlined below). Separation of small geomorphic units and recognition of different geomorphic elements in flat areas is often hindered by the presence of vegetation; therefore research is on-going for the correction of DEM's for vegetation cover (Gallant and Read, 2009; Hofton et al., 2006; Petersen et al., 2009). Another problem is that segmentation methods developed for mountainous areas do not work well in flat areas. Therefore, when dealing with large heterogeneous landscapes, methods for different terrain types have to be combined (Dobos et al., 2005; Ehsani and Quiel, 2008; 2009; Klingseisen et al., 2008; Moore et al., 2003; Prima et al., 2006; Rasemann et al., 2004; Schmidt and Hewitt, 2004). The LandMapR program developed by MacMillan et al. (2004) has been very successful in classifying a hierarchy of landform entities over a full range of spatial scales. Iwahashi and Pike (2007) developed a global terrain map with an automated nested-means classification of topography. Their map of terrain classes for the world is freely available online [http://gisstar.gsi.go.jp/terrain/front_page.htm].

### 2.2.2 Landform mapping based on combined data sources

The combination of a DEM with spectral data can improve landform classification in complex landscapes. From a DEM the basic morphometric identifying parameters are derived, as described in section 2.1, and additional spectral segmentation has been used to refine morphometrically similar landforms (Saadat et al., 2008). Different landform models have been developed for using spectral data in combination with data from SRTM and Landsat Thematic Mapper+ (TM+). The combination of these data sources resulted in better classification of landform types which are dominated by slope processes (Ehsani and Quiel, 2009; Martin and Franklin, 2005; Taramelli and Melelli, 2009). Other satellites whose data have been used for landform recognition in combination with a DEM are e.g. ASTER (Glasser et al., 2008; Saadat et al., 2008; Schneevoigt et al., 2008), Satellite Pour l'Observation de la Terre (SPOT) (Hansen et al., 2009) and Compact High Resolution Imaging Spectrometer (CHRIS) (Ulrich et al., 2009). All studies referred to above concluded that spectral data improved classification, because of increased distinction between topographically similar landforms. Unlike these studies, which were carried out at local scale, Ballantine et al. (2005) and Iwahashi and Pike (2007) used MODIS (Moderate Resolution Imaging Spectroradiometer) – and SRTM30 data respectively as the sole data source for producing a general landform map at global or regional scale.

Applications of Synthetic Aperture Radar (SAR) data and combinations of SAR and multispectral data have also been extensively studied within the context of improved landform recognition on a local scale (Madhavan et al., 1997; Singhroy and Molch, 2004). Different wavelengths of the SAR signal enable structural analysis of elements in specific size classes, while polarization angles are particularly sensitive for directional structures. The lineament orientations or faults of geomorphological units are enhanced by different single polarized images and multipolarization combinations from SAR (Fig. 2.2). Moreover, SAR is cloud-penetrating while the strength of backscatter depends on the dielectric properties of surface materials (e.g. soil water content), and the copolarization sensitivity to surficial sediments, both improve the classification of exposed surficial sediments (Singhroy and Molch, 2004).

Singhroy et al. (2003) fused RADARSAT with TM images and employed spectral classification for distinguishing surficial deposits, moisture conditions and vegetation cover, which facilitated the interpretation and delineation of terrain units within an area of about 4800 $km^2$ (Singhroy et al., 2000; 2003). Subsequently, C-band SAR images were used in densely vegetated areas to produce an image of the surface envelope of the canopy enhanced by highlights and shadows related to surface structures and erosional features. The above examples illustrate how automated analyses can benefit from complementary information provided by radar and spectral imagery. Combinations of SAR and multispectral images have also

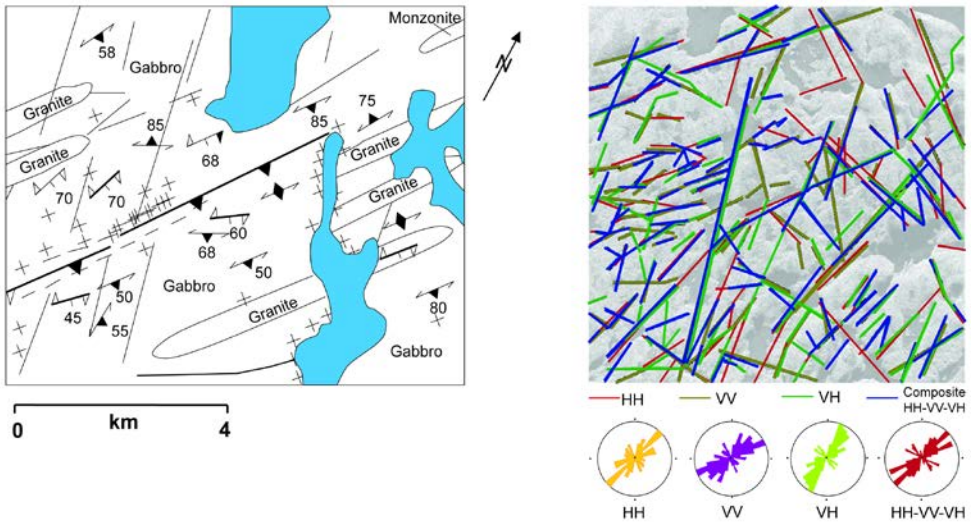


Figure 2.2: (Left) Original map of geological setting and (right) different structural orientation interpreted from each polarization. (Reprinted with permission, Singroy and Molch (2004))

been visually interpreted for identifying geological structural features. However, visual interpretation is time consuming and it is sensitive to human interpretation errors (Madhavan et al., 1997).

### 2.2.3 Digital elevation models

The most widely used sources of DEM data are Light Detection and Ranging (LIDAR) and SAR and stereo-correlation of images. Dependent on the sensor flight altitude, LIDAR allows highly accurate and very densely sampled elevation points (Woolard and Colby, 2002). Processing of LIDAR data involves filtering irregularly spaced data points to obtain terrain elevation projected onto a regular grid (Brennan and Webster, 2006; Hodgson et al., 2003). SAR data are typically processed using interferometric techniques. SAR data are either airborne or spaceborne; near-global coverage, between approximately 60° northern latitude and 56° southern latitude has been achieved with the Shuttle Radar Topography Mission (SRTM) (Farr, 2000). Compared to typical LIDAR data sets, SRTM has much poorer spatial resolution, but unlike the former, SRTM data is easily accessible and even available for free (Farr, 2000). Recently, the ASTER Global Digital Elevation Map (GDEM), created by stereo-correlation of ASTER imagery, has been made available for free to the public. The ASTER GDEM has a spatial resolution of 30m and has near global coverage (METI/ERSDAC, 2009). A fine resolution (2.5 m) DEM can be generated with the ALOS/PRISM which is a panchromatic RS instrument specially designed for stereo mapping (Earth Observation Research Center, Japan Aerospace Exploration Agency, 2010).

A main limitation for both, LIDAR and SAR based approaches is vegetation cover density. For LIDAR, too small gap fractions in the canopy prevent the laser pulse to reach the ground, for SAR decorrelation of moving foliage is the main issue. Further, in rugged terrain, the observational geometry of SAR limits assessing continuous DEMs. Vertical and horizontal accuracies vary for the different data sets. For example, SRTM is claimed to have a vertical absolute accuracy, which is a measure of how accurate elevation is at each pixels, less than 16m, and an absolute horizontal accuracy of 20m (Farr, 2000). For the ASTER GDEM the estimated absolute accuracies are 20m vertically and 30m horizontally (METI/ERSDAC, 2009). For ALOS/PRISM the estimated absolute vertical accuracy is about 6m and the absolute horizontal accuracy is 2.5m (Earth Observation Research Center, Japan Aerospace Exploration Agency, 2010). The estimated absolute accuracies for the Ice, Cloud and land Elevation Satellite (ICEsat) Geoscience Laser Altimetry System (GLAS) are 16.1m horizontally and 13.2m vertically. GLAS offers global coverage of raw altimetric data, but processed DEM's are available only regionally owing to the necessity of specific corrections for surface types (Pagnutti and Ryan, 2009). There are many different airborne LIDAR sensors on the market; for example, the Airborne Laser Terrain Mapper (ALTM) of Optech has a vertical accuracy of 8-11 cm and a horizontal accuracy of 2-3 cm on the ground (Rayburg et al., 2009).

### 2.2.4 Vegetation patterns and indices

Spatial and temporal variations in vegetation indices have been have been found to be linked to prevailing climate, ecosystem, terrain and physical soil properties (Singh et al., 2004); The Normalized Difference Vegetation Index (NDVI) is one of the most common indicators of crop growth characteristics and, indirectly, of specific site qualities (Sommer et al., 2003; Sumfleth and Duttmann, 2008). Tucker (1979) introduced NDVI and the Global Inventory Modeling and Mapping Studies (GIMMS) data set, the latter of which provides a time series of NDVI data (Julien and Sobrino, 2009; Los et al., 1994). A serious problem in partly vegetated areas is the influence of soil background reflectance on NDVI, which produces decreasing NDVI values with increasing soil brightness under otherwise identical circumstances (Huete, 1988; Tucker et al., 1985). Therefore, several variations on the NDVI have been developed, e.g. the Soil Adjusted Vegetation Index (SAVI) (Huete, 1988; Rondeaux et al., 1996), the Transformed SAVI (TSAVI) (Baret et al., 1989; Rondeaux et al., 1996), the Modified SAVI (MSAVI) and the Global Environment Monitoring Index (GEMI) (Qi et al., 1994; Rogan and Yool, 2001; Rondeaux et al., 1996). For a more extensive overview of vegetation indices the reader is referred to Dorigo (2007) and Huete (1988).

Examples of soil properties that have been related to monotemporal NDVI imagery in local scale studies are root zone soil moisture (Wang, 2008), soil colour (Singh et al., 2006), soil texture and water-holding capacity (Lozano-Garcia et al.,

1991) and soil carbon and nitrogen content (Sumfleth and Duttmann, 2008). Alternatively, NDVI time series have been used to derive soil patterns by analysing changing NDVI values during a growing season and the onset of senescence during a dry season, for example Lozano-Garcia et al. (1997). Hansen et al. (2009) found larger changes in vegetation greenness and canopy water absorbance on steeply sloping valley sides with sandy soils than in nearly flat, waterlogged valley bottoms. Dobos et al. (2000) found that the use of spectral indices such as NDVI in combination with a DEM often produced soil pattern delineations comparable to existing regional scale soil and terrain data. To our knowledge, in regional studies NDVI data have been related to soil type patterns rather than to specific soil properties.

The use of bio-geographical gradients in non-linear ordinations combines the information from vegetation ecology and RS methods (Schmidtlein et al., 2007). The approach entails analysing trends present in sets of floristic variables and establishing the nature of interrelationship between them (Armitage et al., 2004). Reflectance values have been related to ordination axes by PLSR and the resulting regression equations were then applied on the spectral image. By this method the compositional variation was mapped for an area of 25 ha ($R^2$ = 0.79), using continuous fields (Schmidtlein et al., 2007). Such bio-geographical patterns can be related to precipitation, temperature and soil conditions on a regional scale (Mahecha and Schmidtlein, 2008; Schmidt and Hewitt, 2004). Alternatively, stratification using indicator species of vegetation for specific habitats enables soil types to be allocated to specific strata, and vice versa (Mücher et al., 2009). However, the success of the latter method is limited to the availability of data for potential natural vegetation (PNV) and indicator species, and has been unsuccessful in ecoregions significantly altered by humans.

Obviously, vegetation does not allow direct measurement of pure soil spectra. Removing the spectral influence of vegetation from the signal may improve the mapping of soil attributes (Bartholomeus, 2009). Both Bierwirth (1990) and Luo et al. (2005) describe a spectral unmixing technique which is based on the assumption that pixel reflectance is a linear mix of component reflectance. With the unmixing technique the abundance of the different endmembers were determined. Then the signature of vegetation corresponding to its abundance fraction was eliminated, and other endmember signatures covered by vegetation were replaced by scaling their abundance fractions to sum the original pixel total.

## 2.3 Measurement of soil properties on bare soil

Stoner and Baumgardner (1981) identified five characteristic soil spectral reflectance curve forms which they considered representative of the diversity of soil reflectance found in wide ranges of naturally occurring surface soils. These curve forms were identified by curve shape and the presence or absence of absorption features representing distinctive organic matter and iron content, as well as texture. Several decades later it was demonstrated that many soil attributes can be measured by spectral analysis of soil samples under laboratory conditions. Examples include sand, silt and clay (Chang, 2002; Hahn and Gloaguen, 2008; Minasny and McBratney, 2008; Nanni and Demattê, 2006; Salisbury and D'Aria, 1992), $Fe_2O_3$, $SiO_2$, $Al_2O_3$ (Boardman, 1994; Genú and Demattê, 2006; Nanni and Demattê, 2006; Stoner and Baumgardner, 1981), soil organic matter (Ben-Dor et al., 2002; Chang, 2002; Gomez et al., 2008b; McCarty et al., 2002; Metternicht, 2003; Stoner and Baumgardner, 1981; Viscarra Rossel et al., 2006b), soil moisture, salt and carbonates (Ben-Dor et al., 2002; Farifteh et al., 2006). McBratney et al. (2006) introduced the use of pedotransfer functions for predicting functional soil properties such as water content and pH buffering capacity from spectrally analysed soil samples. Demattê and Garcia (1999) related spectral reflectance to soil weathering state associated to different forms of iron and the texture of soils developed from basaltic rock.

The above shows that under laboratory conditions several soil attributes can be determined by spectral analysis. However, airborne or spaceborne spectroscopy complicates the measurement owing to atmospheric influences (Gail et al., 1994; Richter and Schläpfer, 2002), structural effects, lower spectral and spatial resolution, geometric distortions and spectral mixture of features (Kriebel, 1978; Richter and Schläpfer, 2002). For soil applications, (partial) coverage of the soil with photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV) and lichens can be a limitation as well; in densely vegetated areas the cover of PV and NPV can be up to 100%. In tundra and open woodland habitats, lichens and mosses can cover as much as 70 % of the surface (Solheim et al., 2000).

In the following subsections we review the different soil attributes that can be determined by laboratory and field spectroscopy (i.e. proximal sensing) as well as airborne and spaceborne spectroscopy (i.e. remote sensing) of bare or sparsely vegetated soil. Unlike globally important soil properties such as texture, organic matter, moisture and mineralogy, other soil properties are particularly of local or regional relevance. Examples are iron content which is pertinent to significantly weathered soils, soil salinity which affects semi-arid to arid climates and carbonates that are indicative of specific parent materials. Furthermore, the spatial coverage of different RS products varies from near global coverage to local areas up to single scattered images.

Table 2.1(p.28) gives an overview of various soil and terrain attributes that have been retrieved from RS and PS data. The feasibility of the retrieval has been scaled between 'low' and 'high'. This feasibility rating is based on a multi-criteria analysis of the reviewed literature cited in this paper. The criteria taken into account are the quality of the dataset i.e. being the number of samples and the research methodology, the accuracy of obtained results, the number of studies reported and the applicability to field surveys. Each criterion was assigned separately and weighed equally to obtain the final rating. Accompanying tables (Table 2.2a, b and c, p. 29) give the specifications of the different sensors used in the reviewed work.

### 2.3.1 Mineralogy

The analysis of mineralogy with spectral PS has made great progress over the last years. Nowadays, several institutes provide spectral libraries with comprehensive collections of a wide variety of materials. For example, the ASTER spectral library version 2.0, which is a collection of contributions from the Jet Propulsion Laboratory, Johns Hopkins University and the United States Geological Survey, is a widely used spectral library which contains over 2400 spectra of a wide variety of minerals, rocks, vegetation and manmade materials covering the wavelength range 0.4-15.4 μm (Baldridge et al., 2008). Methods such as partial least square regression (PLSR) can be used to match collected spectral samples to those in the spectral libraries (Viscarra Rossel, 2008; Viscarra Rossel et al., 2009).

With RS, mineralogy can be determined from the spectral signature of rock outcrops or from the mineral composition of bare in-situ soils. In order to discriminate between different minerals, subtle differences in the spectral signature throughout the VNIR (Visible and Near Infrared) – TIR (Thermal Infrared) are used. Therefore, satellite data with a fine spectral resolution are needed, as only with a fine spectral resolution can subtle spectral differences be detected in the signal. Additionally, fine spatial resolution is beneficial, as it reduces the number of elements represented within a pixel, which enhances the unmixing results and thereby the detection of minerals. The spatial and spectral resolutions of Landsat TM and MODIS have been found to be too coarse for determining mineral composition (Dobos et al., 2000; Kettles et al., 2000; Teruiya et al., 2008). However, the combination of Landsat TM data and ASTER data has been useful because the general lithological variability is mapped with Landsat TM whereas ASTER maps the different mineral groups.

Table 2.1 Feasibility* of determining soil and terrain attributes with RS and PS instruments.

| Attributes | Radar | | Lidar | Optical | |
|---|---|---|---|---|---|
| | Passive | Active | | Multispectral | Spectroscopy |
| Terrain attributes | | | | | |
| Elevation | - | high | high | medium | - |
| Slope | - | high | high | medium | - |
| Aspect | - | high | high | medium | - |
| Dissection | - | medium-high | medium-high | low-medium | - |
| Landform unit | - | medium-high | medium-high | medium-high | low-medium |
| Digital Soil Mapping | - | high | medium-high | medium-high | medium |
| Soil type | - | - | - | medium | high |
| Soil attributes – proximal sensing | | | | | |
| Mineralogy | - | - | - | - | high |
| Soil texture | - | high | - | - | high |
| Iron content | - | - | - | - | medium-high |
| Soil organic carbon | - | - | - | - | high |
| Soil moisture | high | high | - | - | high |
| Soil salinity | - | - | - | - | medium |
| Carbonate content | - | - | - | - | medium |
| Nitrogen content | - | - | - | - | high |
| Lichen | - | - | - | - | medium-high |
| Photosynthetic vegetation | - | - | - | - | medium-high |
| Nonphotosynthetic | - | - | - | - | medium-high |
| Soil attributes – remote sensing | | | | | |
| Mineralogy | - | - | - | medium | medium-high |
| Soil texture | - | medium | - | medium | medium |
| Iron content | - | - | - | low | medium |
| Soil organic carbon | - | - | - | low | high |
| Soil moisture | medium-high | medium-high | - | medium | low-medium |
| Soil salinity | - | medium-high | - | low-medium | medium |
| Carbonate content | - | - | - | low-medium | low-medium |
| Nitrogen content | - | - | - | - | medium |
| Lichen | - | - | - | low-medium | medium |
| Photosynthetic vegetation | - | - | - | medium | medium-high |
| Nonphotosynthetic vegetation | - | - | - | medium | medium-high |
| Ellenberg indicator values | - | - | - | - | low |
| Plant functional type | - | - | - | low-medium | low |
| Vegetation Indices | - | - | - | high | medium |
| Land cover | - | low-medium | - | medium-high | high |
| Land degradation | - | low-medium | - | high | low-medium |

*Feasibility (1-5) = weighted average of scores for the number of studies reported, dataset quality, obtained result and applicability to field surveys. Low = 1, low-medium = 2, medium = 3, medium-high = 4 and high = 5.

Table 2.2a: Current systems providing optical data for soil and terrain mapping.

| Optical | | Subsystem | Spectral bands | Spectral range (μm) | Spatial resolution (m) | Spatial* coverage |
|---|---|---|---|---|---|---|
| Spaceborne | | | | | | |
| | Landsat | VNIR -TIR | 8 | 0.45-12.50 | 15-60 | Global |
| | MODIS | VNIR-TIR | 36 | 0.40-14.40 | 250-1000 | Global |
| | MERIS | VNIR | 15 | 0.39-1.040 | 300 | Global |
| | ASTER | VNIR-TIR | 15 | 0.52-11.65 | 15-90 | Global |
| | Hyperion | VNIR-SWIR | 242 | 0.40-2.500 | 30 | Regional |
| | ALOS/PRISM | VIS | 1 | 0.52-0.77 | 2.5 | Local |
| Airborne | | | | | | |
| | AVIRIS | VNIR | 224 | 0.38-2.500 | 4-20 | Local |
| | HyMap | VNIR -SWIR | 128 | 0.45-2.480 | 2-10 | Local |
| | ROSIS | VNIR | 115 | 0.42-0.873 | 2 | Local |
| | DAIS-7915 | VNIR-TIR | 79 | 0.45-12 | 3-10 | Local |

* Local: $<10^4$ km$^2$, regional: $>10^4$ km$^2$, $<10^7$km$^2$ and global = $>10^7$ km$^2$

Table 2.2b: Current systems providing radar data for soil and terrain mapping.

| Radar | | Subsystem* | Data collection | Spatial resolution (m) | Spatial** coverage |
|---|---|---|---|---|---|
| Spaceborne | | | | | |
| | SRTM | C-, X-band | active | 30 | Global |
| | RADARSAT2 | C-band | active | 3-100 | Global |
| | ASAR Envisat | C-band | active | 30-1000 | Global |
| | PALSAR scansar | L-band | active | 100 | Global |
| | PALSAR | L-band | active | 24-88 | Global |
| | Polarimetric SMOS | L-band | passive | 1000 | Global |
| Airborne | | | | | |
| | E-SAR (3-85 cm) | X-, C-, L-, P-band | active | 2-4 | Local |
| | GeoSAR | | active | 3 | Local |
| | MIRAMAP | X-, P-band X-, C-, L-band | passive | 5-50 | Local |

*X-band: 2.5-4 cm L-band: 15-30 cm
C-band: 4-8 cm P-band: 30-85 cm

** Local: $<10^4$ km$^2$, regional: $>10^4$ km$^2$, $<10^7$km$^2$ and global = $>10^7$ km$^2$

Table 2.2c: Current systems providing LIDAR data for terrain mapping.

| LIDAR | | Spectral range (nm) | Spatial resolution (m) | Spatial coverage |
|---|---|---|---|---|
| Spaceborne | | | | |
| | ICEsat GLAS | 532, 1064 | 70 | Global |
| Airborne | | | | |
| | ALTM Gemini | 1064 | 2-3.5 | Local |
| | ALTM Orion | 1064 | <1.5 | Local |

* Local: $<10^4$ km$^2$, regional: $>10^4$ km$^2$, $<10^7$km$^2$ and global = $>10^7$ km$^2$

Hyperion (Mahoney et al., 2002), airborne AVIRIS (Fig. 2.3) (Green et al., 1998; 2003; Kruse et al., 2003) and Hymap (Martini et al., 2004) spectrometers may be better suited because of their higher spatial and spectral resolution. For example, AVIRIS data has been used to analyse the variation in soil type and their mineralogical and chemical compositions. This is achieved by mapping $SiO_2$ and $Al_2O_3$ in order to estimate the Ki-index, an indicator of the degree of soil weathering (Galvão, 2008). However, the spatial extent of the latter products is smaller, and this limitation applies also to Hyperion and airborne products (Table 2.2a and b).

Several methods relying on spectral RS data have been developed for geological mapping. The spectral features of typical rocks on Earth are mostly found in the TIR region, where quartzoze, carbonate, silicate and mafic minerals can be discerned. Several indices have been developed, such as the Quartz Index, Carbonate Index and Mafic Index (Ninomiya et al., 2005). They are claimed to be suitable for regional to global mapping, but so far they have only been tested in arid and semi-arid regions with ASTER data. Results suggest robustness of the indices for detecting rock types under different climatic circumstances and elevations which applied to reported case studies (Ninomiya et al., 2005).

In local studies, advanced methods for deriving minerals from ASTER data have resulted in classification accuracies up to 86%. Examples of powerful sub-pixel unmixing analysis tools are the Successive Projection Algorithm (SPA) (Zhang et al., 2008), Spectral Angle Mapper (SAM), Constrained Energy Minimization (CEM) and spatial-spectral endmember extraction (SSEE) tool (García-Haro et al., 2005; Rogge et al., 2007; Rowan and Mars, 2003; Zhang et al., 2007). The Tetracorder tool, on the other hand, consists of a set of algorithms within an expert system decision-making framework for soil and terrain mapping. The expert system rules are implemented in a decision tree in which multiple algorithms are applied to the spectral data. The system contains a large spectral library with soil mineral properties and land cover types from all over the world. The results obtained with the Tetracorder show that many different minerals can be identified, as has been shown in figure 2.3 (Clark et al., 2003).

A comparison of the spaceborne Hyperion and airborne AVIRIS spectrometers revealed that Hyperion provides similar basic mineralogical information, but that it is unable to distinguish subtle spectral differences due to its much lower signal-to-noise ratios (Kruse et al., 2003; Kruse et al., 2002). Cudahy et al. (2001) evaluated the impact of the characteristic low signal-to-noise ratio (S/N < 40:1) of Hyperion. They minimized the influence of noise and improved diagnostic spectral signatures using radiative transfer-based atmospheric correction of Hyperion radiance to surface reflectance. Even though the derived surface composition maps generated from endmember images were noisy, they showed the same spatial patterns and correlated

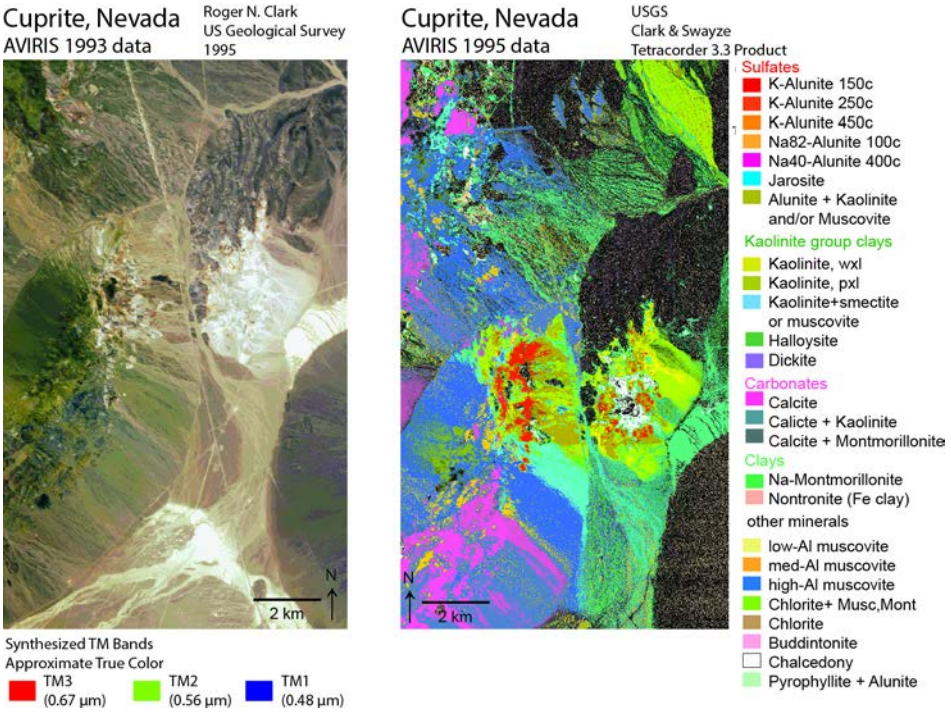


Figure 2.3: (Left) True color composite of Cuprite, Nevada and (right) the corresponding mineral map derived from AVIRIS data (Reprinted with permission, Clark et al. (2003)).

well with known geology (Cudahy et al., 2001). Also, in another study in which Hyperion data were used in combination with additional information obtained from spectral analysis of field samples, the resulting classification accurately mapped different types of rock outcrop (Mahoney et al., 2002). When Hyperion imagery is used in combination with multispectral images such as ALI (Advanced Land Imager) (Beck, 2003) or ASTER, both the calibration of the multispectral data and the mapping accuracy can be improved. The three sensors offer complementary capabilities: ALI is well suited for distinguishing iron oxides, iron hydroxides and iron sulphates, whereas ASTER enables distinctions to be made between clay and sulphate mineral species (Hubbard et al., 2003; Hubbard and Crowley, 2005).

### 2.3.2 Soil texture

In PS, soil texture is typically determined by multiple linear regression or PLSR. Calibration of these models is mostly done using data from a sample. Results show that these methods are useful tools for predicting soil texture, but calibration of the models is based on local conditions and therefore these models will typically not work outside the studied areas (Demattê et al., 2007; Minasny et al., 2008; Thomasson et al., 2001; Viscarra Rossel et al., 2006b).

Apan et al. (2002) used ASTER bands 2, 8 and the first principal component of ASTER imagery for determining broad texture classes. Differences between clay-rich and quartz-rich soils can be locally or regionally mapped based on specific absorption features. Clay minerals have typical hydroxyl absorption at 2200 nm, referred to as the SWIR Clay Index (Chabrillat, 2002). This feature can be captured with bands 5 and 6 of ASTER. The presence of quartz can be detected using thermal bands between 8000 nm and 9500 nm in which the restrahlen feature occurs, which correspond with bands 10 to 14 of ASTER. The combination of ASTER bands 5 and 6 and thermal infrared bands 10 and 14 can then be used to discriminate both dark clayey soils and bright sandy soils from nonphotosynthetic vegetation on a local scale, but results are influenced by organic matter (Breunig et al., 2008; Salisbury and D'Aria, 1992).

AVHRR (Advanced Very High Resolution Radiometer) has been used to map the spatial extent of clay content by means of multivariate prediction models (Odeh and McBratney, 2000). Landsat TM, SPOT and airborne spectroscopy have been used to determine different soil texture classes by correlation of image data with laboratory analysis (Barnes and Baker, 2000). The different soils were classified with accuracy from 50% up to 100%. Note however that this study was conducted on a plot-scale with an exhaustive soil sample dataset; due to the availability of this large dataset higher accuracies were obtained compared to other studies. Only few researchers explored Hyperion data for mapping soil texture; the main reason for this is the earlier mentioned low signal-to-noise ratio and additionally required heavy pre-processing. Even so, Chabrillat et al. (2002) successfully identified, after noise reduction, expansive clays in the Colorado Front Range Urban Corridor when vegetation cover was less than 10%.

In contrast with the use of optical imagery, there is little experience in using radar to retrieve soil texture. Singh and Kathpalia (2007) developed a modelling approach based on a genetic algorithm, which included empirical modelling to simultaneously retrieve soil moisture, roughness and texture from the dielectric constant derived from ERS-2 SAR backscatter data. Although the results were in agreement with field observations, they concluded that there are problems with the retrieval of input variables of the model.

### 2.3.3 Soil moisture

Microwave RS of soil moisture content is based on the contrast in dielectric properties between dry soil and water. Currently, the most advanced index on soil moisture is the Soil Water Index (SWI) (Wagner et al., 2007), in which the METOP ASCAT and ENVISAT ASAR GM data are combined into one layer with a resolution of 1 km. ENVISAT ASAR GM provides backscatter data on a monthly basis, at best, with a resolution of 1 km on a regional scale. The ERS 2 scatterometer and METOP ASCAT provide global backscatter data at a coarser spatial resolution (25 km) but on a finer temporal resolution. Surface soil moisture is derived from the backscatter data and then used to compute the SWI, which gives relative values of soil water content over the rooting depth. The boundaries of the backscatter data related to soil moisture are set on the basis of a long-term change detection approach. The temporal resolution of the SWI is 2 weeks. The index is particularly useful for monitoring changes in soil water content over time, and is unsuitable to quantify the soil water content (Wagner et al., 2007; Wagner and Scipal, 2000).

The recently launched passive microwave SMOS (Soil Moisture and Ocean Salinity) and future satellite SMAP (Soil Moisture Active Passive) will have a global coverage with 1 km resolution and a temporal resolution up to approximately 3 to 5 days. The algorithms devised for the retrieval of soil moisture data from SMOS are promising. The core of the SMOS L2 processor is the inversion of the L-MEB (L-band Microwave Emission of the Biosphere) model (Wigneron et al., 2007), which is used as a forward emission model to simulate the L-band emission from the soil–canopy layer. The L-MEB model uses a simplified (zero-order) radiative transfer equation to predict the surface brightness temperature. The modelled surface soil moisture (0-3 cm) is expected to be accurate to within 4.0% volumetric water content (Panciera et al., 2009; Wigneron et al., 2007).

A different approach to estimating soil moisture is the use of surface energy balance models. These studies are typically done on plot to local scale and produce spatiotemporal predictions of actual evapotranspiration (ET) which can be linked with soil water. There are several models available; the most widely used are (1) the Soil Energy BALance (SEBAL), in which soil and vegetation contributions to ET are aggregated (Bastiaanssen et al., 2005); (2) the Two-Source Energy Balance (TSEB) modelling approach, which discriminates between soil and vegetation (Anderson et al., 2008); (3) the Surface Energy Balance System (SEBS) (Su, 2002) which uses both the optical and thermal parts of the electromagnetic spectrum to estimate turbulent atmospheric fluxes and surface evaporation (Van Der Kwast, 2009). ASTER and MODIS images have been used for retrieving the surface variables required as inputs for energy balance modelling (French et al., 2005; Su et al., 2005). The main difficulties using surface energy balance models are obtaining all the necessary data at the proper spatial resolution and the calibration of the model.

Currently, the most advanced approaches used for estimating root-zone soil moisture are based on assimilation of RS observations into soil–vegetation–atmosphere transfer (SVAT) models. These models can be divided into thermal RS and water and energy balance (WEB) approaches. The WEB-SVAT (Water and Energy Balance – Soil Vegetation Atmosphere Transfer Modeling) model uses measured precipitation and predicted evapotranspiration. The model is based on forcing a prognostic root-zone water balance model with observed rainfall and predicted evapotranspiration. In RS SVAT approaches, the radiometric temperature is derived from thermal RS and combined with vegetation information obtained at the VNIR wavelengths in order to solve the surface energy balance; this method does not explicitly quantify soil moisture but uses a thermal-based proxy variable for the availability of soil water in the root zone and the onset of vegetation water stress (Crow et al., 2008).

Under laboratory conditions, spectral PS with statistical methods has been used for quantifying soil water content. Examples of such methods are the soil line which plots near infrared as a function infrared reflectance (Baret et al., 1993; Demattê et al., 2006) and multiple regression with the water absorption features centred at 1400 nm, 1900 nm and 2200 nm as the independent predictors. However, the latter method will most likely not work under field conditions owing to strong absorption of radiance by water vapour in the atmosphere.

### 2.3.4 Soil organic carbon

Most research on the determination of soil organic carbon (SOC) with RS has been performed at plot scale (< 1km$^2$). The spectral data is usually obtained from PS so vegetation does not disturb the signal. Correlation coefficients in the range of $0.87<R^2<0.98$ between spectrally measured and chemically analysed samples have been obtained using mid infrared and combined diffuse reflectance spectroscopy (Barnes et al., 2003; Chang, 2002; McCarty et al., 2002; Viscarra Rossel et al., 2006b).

One of the indicators used for soil organic carbon mapping with RS is soil colour; dark soils typically contain more soil organic matter than pale soils. This darkening of soil with higher organic carbon content is due to the effect of saturated organic matter and to variations in the composition and quantity of black humic acid and soil moisture (Viscarra Rossel et al., 2006a). This is why the visible part of the spectrum has often been used to map SOC by soil colour. However, the relationships are not sufficiently robust for practical application on a wide variety of soils (Viscarra Rossel et al., 2006a).

Used techniques employ the shape of the reflectance spectrum, for example using band depth analysis and principal component analysis (Palacios-Orueta et al., 1999; Palacios-Orueta and Ustin, 1998). Alternatively, multivariate regression modelling such as Partial Least-Square Regression (PLSR) and multiple linear regression can be

used. By means of these methods, different topsoil parameters are determined from the spectral signature contained in a single imaging spectrometer image, where the various variables are represented by different combinations of absorption features across the spectra. Applying multiple linear regression to an airborne DAIS7915 calibration dataset obtained satisfying results for soil organic matter; results could be improved by use of better quality of spectroscopy data (Ben-Dor et al., 2002). Selige et al. (2006) used both multiple linear regression and PLSR, with the latter having the highest predictive power. Spaceborne imaging spectrometer data have not been often used for predicting soil organic carbon; but Gomez et al. (2008b) applied advanced spectral unmixing methods to Hyperion data for a 16 $km^2$ study area and obtained similar SOC fractions as those in field observations.

When mapping soil organic carbon on a large mapping scale without extensive calibration by soil samples, a solution could be to use indices based on spectral reflectance for quantifying soil organic carbon. The amount of SOC is then detected with reflectance spectroscopy based on the constituents of SOC: cellulose, starch and lignin. Good relations have been found for indices based on the visible part of the spectrum ($R^2$=0.80) and for the absorption features related to cellulose (around 2100 nm) ($R^2$=0.81). The best index-based relations were compared to results for PLSR ($R^2$=0.87). PLSR proved to be much less sensitive towards extrapolation of the model beyond the mineralogy and SOC levels used during the calibration. Although, the indices seem promising, they must still be tested on spaceborne sensors which have lower signal-to-noise ratio. Application in areas that have significant vegetation cover will be a challenge as well (Bartholomeus et al., 2008).

### 2.3.5 Iron content

Soil iron can be seen as an indicator of soil fertility and the age of the sediments (Bartholomeus et al., 2007). Over the years, PS has proven to be useful for determining soil iron content in soil samples and at plot scale (Demattê, 2002; Nanni and Demattê, 2006). But also, RS has been successfully used for determining the presence of iron over areas up to 500 $km^2$. Both soil colour (Escadafal, 1993) and absorption features have been used to derive iron content (Farrand and Harsanyi, 1997; Palacios-Orueta and Ustin, 1998; Warell, 2003). Iron oxide and iron hydroxides have specific absorption features that are located in the VNIR and can be measured from multispectral or imaging spectrometer images (Abrams and Hook, 1995). However, these features are confounded if there is vegetation cover (Xu et al., 2004).

Only a few methods have been developed to *quantify* soil iron content. Though Landsat TM has been used for this purpose, the low spectral resolution means that the absorption features are not unequivocally discernible and therefore the results are not accurate (Deller, 2006). Bartholomeus et al. (2007) were among the first to quantify soil iron content on the basis of airborne optical data. They determined the

iron content in Mediterranean soils in partly vegetated areas, using ground-based spectral reflectance and airborne imaging spectroscopy. The use of two iron-related absorption features as well as a ratio-based Redness Index, which is the ratio of the reflectance in the red part of the spectrum divided by the sum of total visible reflectance, gave fairly good correlations ($R^2$=0.67 and $R^2$=0.51, respectively) on samples measured under laboratory conditions. Unfortunately, the relations were weak ($R^2$ = 0.26) when applied to airborne ROSIS (Reflective Optics System Imaging Spectrometer) data. The relations appeared to be sensitive to vegetation cover, but a combination of the Redness Index and relations based on the absorption feature made the model more robust against the influence of vegetation cover (Bartholomeus et al., 2007).

### 2.3.6 Soil salinity

In arid and semi-arid climates, precipitation is insufficient to maintain a regular percolation of rainwater through the soil, so soluble salts accumulate, with consequences for soil properties such as structure. Both radar and optical RS have been used for mapping soil salinity. Microwave RS of salinity is based on the dielectric properties of the soil, since salinity is a key element of the electric conductivity (Aly et al., 2007). The dielectric constant is a complex number consisting of a real part, which is related to soil moisture, and an imaginary part, which is related to salinity. Using inverse modelling, the imaginary part can be calculated and calibrated with soil salinity (Bell et al., 2001; Taylor et al., 1996; Yun et al., 2003). Soil salinity classes have been successfully derived on a local scale (< 500 $km^2$) with the C-, P-, and L-bands of airborne and spaceborne radar systems; best results are obtained using L-band data because long wavelengths penetrate soil and vegetation to a greater extend than higher frequencies (Bell et al., 2001; Lasne et al., 2008; Taylor et al., 1996).

The spectral response patterns of saline soils are a function of the quantity and mineralogy of the salts they contain (Mougenot et al., 1993). Using spectral absorption features, spectral PS can be used to provide information on the presence of salt minerals and it enables salt-affected soils to be quantified (Weng et al., 2008). Salinized soils have distinctive spectral features in the VNIR parts of the spectrum, related to water in hydrated evaporite minerals. They show absorption features at 505 nm, 920 nm, 1415 nm, 1915 nm and 2205 nm. However, laboratory spectral analyses revealed that salt-affected soil samples did not exhibit all of the diagnostic absorption features that were found in the spectra of the pure salt minerals. Though the regression models had accuracies up to $R^2$ = 0.8 (Farifteh et al., 2008). Salt scalds and highly salinized soil show additional absorption features at 680 nm, 1180 nm and 1780 nm. These features enable the recognition of minerals such as gypsum, bassanite, and polyhalite, which can be used as indicators. Another informative property is that at approximately 2200 nm hydroxyl features become less pronounced

when samples are more saline. The reduction of the 2200 nm absorption intensity may be a result of a loss of crystallinity in clay minerals. Yet another potentially usable characteristic is that the overall slope of the reflection curve between 800 nm and 1300 nm decreases as samples become more saline (Taylor and Dehaan, 2000).

Using RS on a local scale (< $10^4$ km), broad salinity classes can be mapped with ASTER (Melendez-Pastor et al., 2010), Hymap (Dehaan and Taylor, 2003), Landsat TM and ALI imagery – the latter two using the Salinity Index (SI) and the Normalized Salinity Index (NSI) (Bannari et al., 2008; Jabbar and Chen, 2008; Odeh and Onus, 2008). Weng et al. (2008) were able to discriminate 5 classes of saline soils with Hyperion data for an area of about 1200 $km^2$. Alternative methods for mapping saline areas are based on detecting the presence of salt scalds and halophytic vegetation. However, spectral resolution must be high in order to detect the different vegetation types (Dehaan and Taylor, 2001).

A major constraint to using PS and RS data for mapping salinity is related to the fact that there is a strong vertical, spatial and temporal variability of salinity in the soil profile. Spectral data acquisition does not allow information to be extracted from the entire soil profile, since only the soil surface is observed. This can be overcome by integrating RS with simulation models and geophysical surveys (Farifteh et al., 2006; Metternicht, 2003; Mougenot et al., 1993). Direct and precise estimation of salt quantities is difficult using satellite data with a low spectral resolution because these fail to detect specific absorption bands of some salt types and the spectra interfere with other soil chromophores (Mougenot et al., 1993).

### 2.3.7 Carbonates

RS allows distinction between common carbonate minerals on the basis of unique spectral features found in the SWIR, and especially in the TIR region. In that region the minerals have a low emissivity from 1095 up to 1165 nm and high emissivity from 8125 to 1095 nm. The Calcite Index, for example, is based on this difference in emissivity and has been used on a single ASTER image of 60*60 km (Yoshiki et al., 2002; Yoshiki et al., 2004). Alternatively, the specific absorption features of carbonate have been analysed with derivative analysis on PS data. Derivatives of second order or higher should be relatively insensitive to variations in illumination intensity whether caused by changes in sun angle, cloud cover, or topography (Hu, 2007; Plaza et al., 2008). Under laboratory conditions this method worked well ($R^2$ = 0.64), but when applied to airborne data with a pixel size of $25m^2$, the performance decreased ($R^2$= 0.46). This was attributed to radiometric and wavelength calibration uncertainties as well as possible residual atmospheric effects (Lagacherie et al., 2008).

### 2.3.8 Nonphotosynthetic vegetation

Nonphotosynthetic vegetation (NPV) such as crop residues, woody stems and forest litter influences the spectral response of bare soil and thus the accuracy of determined soil properties. For example, the presence of senescent vegetation and litter affects the relationship between reflectance and several soil attributes such as total Fe, $TiO_2$ and $Al_2O_3$ in the transition from the red to the near-infrared interval. The lignin/cellulose spectral feature of the crop residues also effects the appearance of the 2200-nm absorption band related to clay minerals (Galvão, 2008).

Imaging spectrometers and advanced multispectral sensors potentially allow discrimination between crop residues and soil. Attempts to measure crop residue cover using RS have had mixed success, however. In areas having strongly contrasting soil and residue reflectance, Landsat TM-based indices such as the simple ratio-type vegetation indices (SRTVI) and normalized difference-type vegetation indices (NDTVI) successfully identified broad crop residue cover classes for a coverage ranging from small plots up to 1500 $km^2$ (Serbin et al., 2009). However, these indices were less effective when used in regions with different soil types because of the poor contrasts between crop residues and many soils (Qi et al., 2002). This poor performance was partly ascribed to the broadness of the bands of Landsat TM, which are unable to discriminate between specific material absorptions that occur in the 1000 to 2500 nm wavelength region (Mirik et al., 2005). By contrast, ASTER has been successfully used to discriminate NPV from bare soil by means of band ratios (Breunig et al., 2008). Spectral unmixing of AVIRIS data (Asner and Heidebrecht, 2003; Galvão et al., 2001; Roberts et al., 1993) and Hyperion data (Asner and Heidebrecht, 2003; Huete et al., 2003) in areas up to 100 $km^2$ produced even better results.

The imaging spectrometer Cellulose Absorption Index (CAI) and the multispectral Lignin–Cellulose Absorption (LCA) index are reflectance band height indices that use three spectral bands between 2000 and 2400 nm to estimate crop residue cover (Daughtry et al., 2005; Serbin et al., 2009). Whereas CAI showed a clear separation of values between common soil minerals and crop residues, LCA did not. Spectral confusion may occur because common soil minerals such as amphiboles, chlorites, iron hydroxides, and – especially – carbonate have LCA values similar to those of crop residues (Daughtry and Hunt Jr, 2008; Serbin et al., 2009).

### 2.3.9 Lichens

Exposed bedrock is often partly covered with lichens varying in colour, from blackish-brown to orange. Lichens might prevent the transmission of light to the underlying rock substrate and effectively mask the mineral substrate (Bechtel et al., 2002). They may increase, decrease or not influence the spectral reflectance of a rock

surface, depending on the spectral contrast between the lichen and the bare rock (Satterwhite et al., 1985). Zhang et al. (2005) and Bechtel et al. (2002) found that, under laboratory conditions, a single lichen endmember allowed for spectral unmixing of lichens and rock, due to the spectral similarity of various crustose/foliose lichen species in the short-wave IR. For both the reflectance and normalized reflectance data, spectral mixture analysis results correlated well ($R^2$> 0.9) with endmember abundances estimated from digital photography (Zhang et al., 2005). A number of minerals (clay and carbonates, for example) also have absorption bands in the SWIR, but the shapes and positions of the mineral bands differ from those of the lichen. The absorption at 1700 nm might be most useful in determining areal lichen cover because this is not significantly affected by common mineral absorption bands (Cloutis and Edward, 1989). Moreover, the discrimination of different lichen types is possible in the visible spectral region, but there the inference of mineral absorptions may be problematic (Bechtel et al., 2002; Cloutis and Edward, 1989; Kiang et al., 2007). Kiang et al. (2007) studied the full reflectance spectrum of lichen under laboratory conditions and found that the 'red edge', which is typical for photosynthetic species, is negligible in lichen and that the overall influence of lichen on the NIR is a lowering of the overall reflectance. Therefore, we expect that it is likely that the overall influence of lichen on the retrieval of soil properties is much lower compared to other vegetation types. But, in areas with exhaustive lichen cover, the overall lower reflectance has to be accounted for. We are not aware of any RS-based lichen mapping study on a plot or local scale.

### 2.3.10 Soil proxies

The efficiency of using RS to map soil properties in densely vegetated areas depends on indirect relations between vegetation and soil attributes. As already outlined in section 2.2.4, vegetation indices and time series can be used to delineate soil patterns. To retrieve soil properties, more detailed information on the vegetation cover is needed. Two useful but prospective proxy indicators have been used to retrieve soil properties from RS: Plant Functional Types (PFT) and Ellenberg indicator values.

A central tenet in the concept of PFT is that morphological and physiological adaptations are linked in predictable ways by resource limitations, responses to disturbance, biotic factors or other aspects of the environment. The extent to which such linkages are generalized will determine the ability to detect functional types with RS (Ustin and Gamon, 2010). For example, abiotic factors that affect biodiversity are the nutrients available, such as nitrogen, and the prevailing climatic conditions. In some cases, low levels of nutrients lead to high levels of biodiversity (Forde et al., 2008). Diekmann (2003) shows that the relation between nutrient requirements of plants and nutrient availability in soils can be used to derive soil attributes. Accordingly, the concept of plant functional types (PFT) can be used to

derive the specific type or group of species that grow on typical soils. Schaepman et al. (2007) showed that PFT may be derived from high resolution imaging spectrometer data on a plot level. Sun et al. (2008) developed the current global MODIS PFT product, which is a map with the distribution and abundance of major plant functional types. Ustin and Gamon (2010) proposed the new concept of 'optical' types. They argue that functional types can be distinguished largely on the basis of optical properties detectable by RS. To fully utilize the potential of RS, data must be combined with ecological models linking structural, physiological and phenological traits based on resource constraints. See Ustin and Gamon (2010) for an overview of different sensors and their implications for assessing PFT's. Hence, PFT regulate or are regulated by ecosystem processes and have discrete different functions within the ecosystems.

Different PFT have a particular distribution in relation to geography or environment, e.g. species of ultramafic soils or acidophilous bog species (Wilson, 1999). Therefore, PFT could be explained by the DEM-derived terrain variables which describe the landscape structure. Buis et al. (2009)) found that PFT strongly depended on bedrock cover, which emphasizes the dominance of local water redistribution processes for the PFT. However, for soil and terrain mapping the method should be inverted since the PFT should indicate soil attributes.

For the same reasons as the PFT, Ellenberg indicator values can be used to retrieve soil attributes. Originally, the Ellenberg indicator values were calculated for flora mapped on the basis of intensive fieldwork (Ellenberg, 1988). However, Schmidtlein (2005) showed that imaging spectroscopy can be used as a tool for mapping Ellenberg indicator values for soil water content, soil pH and soil fertility. The Ellenberg indicator values scale the flora of a region along gradients reflecting light, temperature, moisture, soil pH, fertility and salinity. This way, the flora can be used to monitor environmental change and thereby changes in the soil (Diekmann, 2003; Hill et al., 2000).

## 2.4 The use of remote sensing in digital soil mapping

It is unlikely that the variables measured using RS will exhaustively cover the area to be mapped. Accurate estimation of soil attributes is hampered if the pixels have a vegetation cover over 20% (Bartholomeus et al., 2007). Another problem is that the spectral signatures of urban areas, roads and water surfaces do not contain information relevant for soil and terrain mapping. Therefore, these areas should be masked, but this produces incomplete coverage of the remaining area from which RS estimates of soil attributes can be obtained. Finally, RS typically provides information about only the surface layer.

However, by combining RS, DEM's and soil sample data using DSM methods, a complete coverage can be produced and the accuracy of the estimated soil properties

can be improved. Over the past decade many research papers have dealt with DSM, which has been defined as 'the creation and population of spatial soil information by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems (Lagacherie et al., 2007; McBratney et al., 2003)' (Carré et al., 2007). It allows for the prediction of soil properties or classes using soil information and environmental covariates of soil (Carré et al., 2007; Dobos et al., 2006). As this paper concerns applications of RS, the reader is referred to Boettinger et al. (2010), Grunwald (2009), McBratney et al. (2003) and Scull et al. (2003) for reviews on DSM.

RS and PS can provide data sources to be used in DSM. There are different ways in which the data can be used to obtain complete coverage. One way is to fill the gaps by directly interpolating RS-based measurements of soil properties, using these data as primary data source. This approach is suitable if legacy soil data is scarce or unavailable. Alternatively, if legacy soil data are available, soil and terrain attributes derived from RS or soil proxies can be used as secondary variables to improve the interpolation of existing soil data (McBratney et al., 2003).

### 2.4.1 Soil spatial prediction

In this section some methods which can be used for soil spatial prediction using RS and PS are briefly discussed. A distinction is made between the use of RS and PS as primary data source and the use of RS as secondary data source. The single use of RS retrievals as primary data is possible if the spatial coverage is high and well correlated in space, e.g. in sparsely vegetated areas. If the spatial coverage is sparse, e.g. in case of RS in vegetated areas, or PS with a field spectrometer, then the data can be used as a primary data source but other secondary exhaustive data sources, such as a DEM, a geological map, or RS images have to be used to obtain full area coverage and improve the interpolation results. RS retrievals have been used both as primary or secondary variables by Ben-Dor et al. (2002), Castrignanò et al. (2000), Knotters et al. (1995), McBratney et al. (1991) and Odeh et al. (1994).

#### *2.4.1.1 Remote sensing as primary data source*

When mapping spatial patterns, pixels can be treated as regionalized variables using geostatistical techniques (Heuvelink and Webster, 2001; McBratney et al., 2003). In this way, an exhaustive coverage of RS retrievals can be used as a primary data source for mapping soil attributes in areas which are, for example, covered by vegetation or masked by clouds. Using univariate kriging it is possible to map continuous soil properties and classes by accounting for spatial correlation between sampled and nearby unsampled locations. Simple kriging may not be the method of choice in heterogeneous areas because the mean is deemed constant and known throughout the area. Ordinary kriging is often regarded more appropriate, because using a restricted search neighbourhood, it adapts to local fluctuation of the mean by

limiting the domain of stationarity of the mean to the local neighbourhood (Goovaerts, 1999).

Also (generalized) linear models with independent variables such as slope, curvature, wetness-index and soil profile information have been used to derive soil attribute maps. Despite satisfactory results obtained in homogenous landscapes, multiple regression has its limits in complex heterogeneous areas (Gessler et al., 1995; Moore et al., 1993; Odeh et al., 1994). However, the same limitation applies to ordinary and simple kriging.

*2.4.1.2 Remote sensing as secondary data source*

When measurements are sparse or poorly correlated in space, the estimation of the primary attribute is generally improved by accounting for secondary information from other related categorical or continuous attributes such as a digital elevation model, RS data or land use maps. PS can be used as a primary data source and RS can be used as one of the secondary data sources to predict soil properties from PS. This way, the large spectral resolution of the PS data can be combined with the spatial coverage of the RS data. In any case, with PS, either field or laboratory measurements are taken which generates a relatively sparse sample that can be used as primary data source or as a covariable (in co-kriging) for soil spatial prediction on a dense grid. The primary attribute can be predicted with kriging within strata, or some combination of regression analysis and kriging or cokriging (Heuvelink and Webster, 2001; Knotters et al., 1995). If the secondary information is not exhaustively sampled, the estimation can be done by cokriging (Knotters et al., 1995). In contrast to the previously described methods, regression kriging does not assume a stationarity of the observations themselves but rather of the residuals between observations and a modelled trend (Cressie, 1991). The advantage of regression kriging is the ability to extend the method to a broader range of regression techniques and to allow separate interpretation of the two interpolated components (Hengl et al., 2007).

### 2.4.2 Classification and regression trees

The technology involved in collecting and analysing data has become more and more powerful in recent years. Stimulated by the need to analyse massive amounts of data efficiently, data mining was born and has recently developed rapidly. When multisource data are used in a classification, advanced non-parametric classifiers such as neural network, classification tree, evidential reasoning, or knowledge-based approaches are typically preferred over parametric classification algorithms. The underlying assumption is that the relation between soil types and the additional attributes is expected to be non-linear (Hahn and Gloaguen, 2008; Lu and Weng, 2007; Zhai et al., 2006). These methods have given good results when used to extract geographical information, such as land cover, from RS on a local scale and these

methods appeared to be more flexible and robust to non-linear relations. However, the use of too many additional attributes appeared to be limiting the performance for univariate decisions trees (Friedl and Brodley, 1997; Pal and Mather, 2003; Santos and Amaral, 2004)).

Classification and regression tree analysis is a data mining technique which splits a full data set in a sequence of binary splits on dependent (RS) variables aiming for pure nodes with respect to the target variable. The input variables are environmental variables that are believed to reflect the properties integrated by a soil surveyor when making the source map. A tree model is trained and tested by growing and pruning it, using only a proportion of the available data. The model is then used to predict the full extent of the original map (Quinlan, 1993; Salford-Systems, 2002). Classification trees have been used to estimate soil properties and to create soil maps (Lagacherie and Holmes, 1997; McBratney et al., 2000; McKenzie and Ryan, 1999; Omuto and Shrestha, 2007; Scull et al., 2005; Xu et al., 2005). Based on the Soil–Landscape paradigm, explicit links have been made between data, information and knowledge (MacMillan et al., 2007; Moran and Bui, 2002). Accordingly, Quinlan (1993) concludes that there is sufficient predictive capacity in the environmental correlation attributes representing geology, terrain, and soil/water/vegetation interactions to model a known soil map. Note that regression tree analyses may also be used within a regression kriging context (see above) (Hengl et al., 2004; McBratney et al., 2000).

## 2.5 Conclusions and outlook

This article reviewed the use of remote sensing (RS) for soil survey. To summarize: RS provides data (1) supporting the segmentation of the landscape into rather homogeneous soil-landscape units whose soil composition can be determined by sampling or that can be used as a source of secondary information, (2) allowing measurement or prediction of soil properties by means of physically-based and empirical methods, and (3) supporting spatial interpolation of sparsely sampled soil property data as a primary or secondary data source. Table 2.3 gives an overview of the various methods discussed in this paper. Spatial segmentation by automated segmentation of the landscape to support soil–landscape mapping is typically based on first- and second-order derivatives of DEMs, observed parent material and vegetation patterns. Spatial and temporal changes in vegetation indices and biogeographical gradients have been used to improve spatial segmentation.

Extended spectral libraries are available which aid the research on soil attribute retrieval with PS (Abrams and Hook, 2001; Clark et al., 2003). A wide variety of soil attributes have been derived with use of statistical and chemometric analysis of spectroscopic data (Minasny and McBratney, 2008; Viscarra Rossel and McBratney, 2008) which can be used for DSM (Minasny et al., 2009). However, as can be seen in Table 2.1, the feasibility to derive these soil attributes is on average ‘medium’ which

means that current methods are not fully developed yet. The retrieval of soil attributes with RS has made progress, particularly since the launch of advanced multispectral sensors and imaging spectrometers such as ASTER and Hyperion, which have made it possible to detect subtle differences between spectral signatures. Either indices, proxies, quantities or patterns for many soil attributes that are important for deriving soil and terrain maps have been derived from RS. However, the number and feasibility of soil attributes that have been derived from RS images is much lower compared to the use of PS (Table 2.1). Due to the heterogeneity of landscapes and the spatial resolution of the imagery (Table 2.2a, b and c) it is often difficult to find pure pixels representing soil or bare rock. Advanced unmixing tool methods, such as Tetracorder (Clark et al., 2003) are needed to extract sub-pixel soil and rock composition. Finally, the spatial extent of most reported work was restricted to local studies (Table 2.3).

RS data have been used in DSM as covariates for the prediction of soil classes or soil properties. Often, the use of spectral imagery for the spatial prediction of soil properties is based on the spatial relation between existing soil data and observed patterns in the imagery, and not on physically based retrievals, such as soil moisture (Dobos et al., 2000; Stoorvogel et al., 2009). Over the last years spectral PS showed to be useful as part of DSM (Minasny et al., 2009; Viscarra Rossel and McBratney, 2008). Dependent on spatial and spectral resolution, spatial coverage and the availability of legacy data, RS and PS data are either used as primary or secondary data source for the spatial prediction of soil properties. In vegetated areas soil proxies, such as NDVI, plant functional type or Ellenberg indicator values, have been used to derive soil properties, but with mixed success. Alternatively, data mining techniques such as classification trees – which are generated from a matrix of environmental variables – have been used to estimate soil properties and to create soil maps.

Although much progress has been made, current PS methods are not readily implemented at spaceborne level. There are, however, space-based instruments that partially support such approaches (Pieters et al., 2009) or will be available soon (Stuffler et al., 2009). The spectral band settings and improved signal-to-noise performance of upcoming spectrometers in space will certainly improve the retrievals of soil based information using advanced spectral mixing approaches. Secondly, most methods used for retrieving soil attributes have been developed using local or regional correlation approaches, and may not scale for operational use over vast areas. Considering the use of RS for regional DSM, research is needed on extending current methods beyond the plot. Indications are that perspectives exist to develop methods for large scale mapping as indicated in Iwahashi and Pike (2007), Ballantine et al. (2005), Wagner et al. (2007) and Ninomiya et al. (2005). Thirdly, although experiments retrieving soil information work well when using PS, their accuracy drops when (larger-scale) RS methods are being used. This accuracy drop is mainly

caused by sensor noise (Phillips et al., 2009), directional reflectance (Kriebel, 1978), topographic (Richter and Schläpfer, 2002), atmospheric distortions (Gail et al., 1994; Richter and Schläpfer, 2002), and increased mixture of soil properties with decreasing mineral abundances. Since advances in PS have evolved much faster than in RS, a technology gap still has to be bridged.

Future studies will therefore focus on the improved integration of PS and RS using scaling-based approaches in order to make optimal use of all data sources available. Revisit time or temporal approaches are still limited by satellite orbital constraints and/or data download capacity. Soil moisture based retrievals have become increasingly feasible with the launch of SMOS (Soil Moisture Ocean Salinity), but its spatial resolution is still too coarse for soil plot-size retrievals. Certainly, the planned availability of SMAP will further contribute to improved retrievals, including freeze/thaw status of the surface.

This review has shown that future research will aim for the integrated use of RS methods for spatial segmentation, as well as measurements and spatial prediction of soil properties to achieve complete area coverage. In-situ or PS methods are readily available and we will be seeing future instruments launched soon supporting these methods at larger spatial scales finally enhancing the perspectives of DSM.

Table 2.3: Summary of retrieval methods for soil and terrain attributes (* Laboratory = 0 m, Plot = > 0 – 1 $km^2$, Local = > 1 $km^2$ – $10^4$ $km^2$, Regional = > $10^4$ $km^2$ – $10^7$ $km^2$, Global = > $10^7$ $km^2$).

| Soil attributes | Sensor class | Method | Reference | Spatial extent* |
|---|---|---|---|---|
| Carbonate | Remote sensing | Calcite index | Yosiki et al. (2002) | Local |
| | | | Yosiki et al. (2004) | Local |
| | Proximal sensing | Continuum removal | Lagacherie (2006 | Laboratory, Plot, |
| Soil Salinity | Remote sensing | Characterization by salt scalds | Bell et al. (2001) | Local |
| | | SAR C-,P- and L-bands | Taylor et al. (1996) | Local |
| | | | Jabbar (2008) | Local |
| | | Salinity Index | Odeh (2008) | Local |
| | | Normalized Salinity index | Dehaan and Taylor (2003) | Local |
| | | Matched Filtering | Weng et al. (2008) | Local |
| | Proximal sensing | Multivariate regression and VNIR signatures | Melendez-Pastor et al. (2010) | Local |
| | | | Farifteh et al. (2008) | Laboratory |
| Soil iron content | Remote sensing | Principal component analyses | Xu et al. (2004) | Local |
| | | | | Plot |
| | | Redness Index | Bartholomeus et al. (2007) | Laboratory, Plot |
| | Proximal sensing | Soil colour | Escadafel (1993) | Local |
| | | Constrained Energy Minimization | Farrand and Harsanyi (1997) | Plot |
| | | Continuum removal | Bartholomeus et al. (2007) | |
| Soil organic carbon (SOC) | Remote sensing | SOC indices | Bartholomeus et al. (2008) | Plot |
| | Proximal sensing | Multivariate regression modeling | Ben-Dor et al. (2002) | Local |
| | | | Selige et al. (2006) | Plot |
| | | PLSR | Gomez (2008b) | Local |
| Soil moisture | Remote sensing | Soil Water Index | Wagner et al. (2007) | Regional |
| | | | Wagner and Scipal (2000) | Regional |
| | | Surface energy balance models | Su et al. (2005) | Local |
| | | | Bastiaanssen et al. (2005) | Local |
| | | | Crow et al. (2008) | Plot |
| | Proximal sensing | Soil line technique, MLR of absorption dip | Dematte et al. (2006) | Local |

Table 2.3 continued: Summary of retrieval methods for soil and terrain attributes.

| Soil attributes | Sensor class | Method | Reference | Spatial extent |
|---|---|---|---|---|
| Clay/sand fraction – texture | Remote sensing | Merged radar and spectral data | Singhroy et al. (2003) | Local |
| | | | Apan et al. (2002) | Local (60*60km) |
| | | Band ratios from ASTER data | Breunig (2008) | Local (60*60km) |
| | | | Datt et al. (2003) | Local (7.5*100km) |
| | | | Chabrillat (2002) | Local |
| | Proximal sensing | SWIR Clay Index | Dematte et al. (2007) | Local (plot/laboratory) |
| | | Matched filtering algorithm | Selige et al. (2006) | Local (88 ha) |
| | | | Barnes and Baker (2000) | Local (20*40 km) |
| | | Regression equations field/lab | Thomasson et al. (2001) | Plot |
| | | | Sullivan et al. (2005) | Plot |
| | | Multiple linear regression | Viscarra Rossel et al. (2006b) | Plot |
| | | | Minasny and McBratney (2008) | Plot |
| | | Partial Least Square Regression | | |
| Mineralogy | Remote sensing | Quartz Index, Carbonate Index, Mafic Index in TIR | Ninomiya et al. (2005) | Local |
| | | Spectral Angle Mapper | Kruse et al. (1993) | Local |
| | | | Rowan (2003) | Local |
| | | | Farrand and Harsanyi (1997) | Local |
| | | Constrained Energy Minimization technique | Chabrillat (2002) | Local |
| | | Pixel Purity Index on MNF image | Boardman (1994) | Local (plot/laboratory) |
| | | | Rogge et al. (2007) | Local |
| | | | Garcia-Haro (2005) | Local |
| | | SMA, MESMA, VMESMA, SSEE, SPA | Zhang (2008) | Local |
| | | Spectral Feature Fitting | Galvão et al. (2001) | Local |
| Soil proxies | Remote Sensing | Plant Functional Type | Mücher et al. (2009) | Regional |
| | | | Schaepman et al. (2007) | Plot |
| | | | Sun (2008) | Global |
| | | Ellenberg's indicator values | Ustin and Gamon (2010) | Local – global |
| | | | Diekmann (2003) | Local (plot) |
| | | | Schmidtlein (2005) | Local |
| | | NDVI | Wang et al. (2007) | Local |
| | | | Singh et al. (2006) | Local |
| | | | Lozano-Garica et al. (1991) | Local |
| | | | Sumfleth and Duttmann (2008) | Local |

Table 2.3 continued: Summary of retrieval methods for soil and terrain attributes.

| Terrain attributes | Sensor class | Method | Reference | Spatial extent* |
|---|---|---|---|---|
| Elevation | Remote sensing | Summary statistics | Florinsky (1998)<br>Wilson and Gallant (2000) | Local – Regional<br>Local – Global |
| Slope gradient | Remote sensing | Directional derivatives | Florinsky (1998)<br>Wilson and Gallant (2000) | Local – Regional<br>Local – Global |
| Relief intensity | Remote sensing | Directional derivatives | Florinsky (1998)<br>Wilson and Gallant (2000) | Local – Regional<br>Local – Global |
| Major landform | Remote sensing | Automated classification, DEM | Tribe (1992)<br>MacMillan et al. (2004)<br>Irvin et al. (1997) | Local<br>Local<br>Plot |
| | | Fuzzy classification, DEM | Iwahashi and Pike (2007)<br>Schmidt and Hewitt (2004)<br>Klinkseisen (2008)<br>Rasemann et al. (2004) | Global<br>Regional<br>Local<br>Plot, Local |
| | | Object-based classification | Otto et al. (2010b)<br>Raper and Livingstone (1995) | Local<br>Local |
| | | Expert driven multi-level approach, LIDAR DEM | van Asselen and Seijmonsbergen (2006)<br>Tagil and Jennes (2008) | Local<br>Local |
| | | Semi-automatic based on Topographic Position Index | Saadat et al. (2008) | Local |
| | | Combination of DEM classification and spectral data | Ehsani and Quiel (2008a)<br>Ehsani and Quiel (2008b)<br>Kettles (2000)<br>Teruiya (2008)<br>Ballantine (2005) | Local<br>Regional<br>Local<br>Local<br>Regional |
| | | DEM classification and backscatter data | Singhroy (2003)<br>Singhroy (2004) | Local<br>Local |
| Regional slope | Remote sensing | Directional derivatives DEM | Dobos et al. (2005) | Regional |
| Hypsometry | Remote sensing | Directional derivatives DEM | Dobos et al. (2005) | Regional |
| Dissection | Remote sensing | Directional derivatives DEM | Florinsky (1998) | Local – Regional |
| Permanent water surface | Remote sensing | Pixel Purity Index<br>Spatial Spectral Endmember Extraction | Boardman (1994)<br>Rogge et al. (2007 | Laboratory<br>Local |
| Texture non-consolidated material | Remote sensing / Proximal sensing | Spectral unmixing | Apan et al. (2002)<br>Breunig (2008) | Regional<br>Local |

Table 2.3 continued: Summary of retrieval methods for soil and terrain attributes.

| Vegetation | Sensor class | Method | Reference | Spatial extent |
|---|---|---|---|---|
| Vegetation patterns | Remote sensing | NDVI | Tucker (1979) | Global |
| | | | Huete (1988) | Global |
| | | Biogeographical ordination | Schmidtlein et al. (2007) | Plot |
| | | | Mahecha and Schmidtlein (2008) | Regional |
| Photosynthetic vegetation | Remote sensing | Vegetation Indices | Tucker (1979) | Global |
| | | | Huete (1988) | Global |
| | | | Qi et al. (1994) | Plot |
| | | Spectral unmixing | Luo et al. (2005) | Local |
| Nonphotosynthetic vegetation | Remote sensing / Proximal sensing | Cellulose Absorption Index (CAI) | Daughtry et al. (2005) | Local |
| | | | Serbin et al. (2009) | Local |
| | | | Daughtry et al. (2005) | Local |
| | Remote sensing / Proximal sensing | Lignin Cellulose Absorption index (LCA) | Serbin et al. (2009) | Local |
| | | | Mirik et al. (2005) | Plot |
| | Remote sensing | Simple ratio-type vegetation indices (SRTVI) | Mirik et al. (2005) | Plot |
| | | Normalized difference-type vegetation indices (NDTVI) | Asner and Heidebrecht (2003) | Local |
| | | Probalistic spectral mixture model | Heute et al. (2003) | Local |
| | | Spectral unmixing | Souza et al. (2003) | Plot |
| Lichens | Remote sensing / Proximal sensing | Spectral unmixing | Zhang et al. (2005) | Laboratory |
| | | | Bechtel et al. (2002) | Laboratory |

# Chapter 3

# Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data

*This chapter is based on:*
V.L. Mulder, S. de Bruin, M.E. Schaepman (2012). Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 21(1), 301-310. DOI: 10.1016/j.jag.2012.07.004

## Abstract

This paper presents a sparse, remote sensing-based sampling approach making use of conditioned Latin Hypercube Sampling (cLHS) to assess variability in soil properties at regional scale. The method optimizes the sampling scheme for a defined spatial population based on selected covariates, which are assumed to represent the variability of the target variables. The optimization also accounts for specific constraints and costs expressing the field sampling effort. The approach is demonstrated using a case study in Morocco, where a small but representative sample record had to be collected over a 15.000 $km^2$ area within 2 weeks. The covariate space of the Latin Hypercube consisted of the first three principal components of ASTER imagery as well as elevation. Comparison of soil properties taken from the topsoil with the existing soil map, a geological map and lithological data showed that the sampling approach was successful in representing major soil variability. The cLHS sample failed to express spatial correlation; constraining the LHS by a distance criterion favoured large spatial variability within a short distances resulting in an overestimation of the variograms nugget and short distance variability. However, the exhaustive covariate data appeared to be spatially correlated which supports our premise that once the relation between spatially explicit remote sensing data and soil properties has been modelled, the latter can be spatially predicted based on the densely sampled remotely sensed data. Therefore, the LHS approach is considered as time and cost efficient for regional-scale surveys that rely on remote sensing-based prediction of soil properties.

## 3.1 Introduction

Soil and terrain information is needed for many purposes including policy-making, land resource management, and for monitoring environmental impact of development (Battrick, 2005). Especially in developing countries, information about soils is sparse while at the same time resources for data acquisition are limited (Rossiter, 2004). Therefore, these resources should be used efficiently which means that acquisition plans have to be effectively fulfilling the goals; usually only few sites can be visited while they should represent the variability present in the area. Fortunately, often auxiliary data is available (e.g. remote sensing (RS) data) serving as a proxy representing the environmental conditions under which the soils have developed (Mulder et al., 2011b). In case of strong correlation, these data can be used in combination with a small sample of primary data to predict soil properties. Based on the soil-landscape paradigm (de Bruin et al., 1999; Hudson, 1992; Jenny, 1941), we postulate the existence of a relation between surface reflectance and soil properties and present a purposive sampling approach using RS data to select sites that are expected to represent major soil variability. The aim of sampling is to model the above-mentioned relation; however model selection and calibration fall outside the scope of the current paper. Examples of methods used for such purpose are linear regression (Gessler et al., 1995; Odeh et al., 1994) and regression trees (Lagacherie and Holmes, 1997; Scull et al., 2005). Both the selection and the calibration of regression models benefit from data covering the full range of the explanatory variables, for example to avoid extrapolation and decide upon the polynomial order (Rawlings et al., 1998). Any voids in the surface predicted by the regression model (e.g. because of partial cloud cover) will be interpolated, using a geostatistical method (Hengl et al., 2007; Heuvelink and Webster, 2001).

Over the past decades, extensive work has been published on sampling schemes for soil mapping. De Gruijter et al. (2006) distinguish two fundamentally different statistical approaches for sampling and inference, the design-based (DB) strategies and the model-based (MB) strategies. The DB strategies involve probability sampling and design-based inference and aims to estimate the population parameter. That is, a pre-specified number of sample locations are randomly selected from the area with known probabilities, and the population parameters are estimated from the sample data based on the selection probabilities. Examples include (Stratified) Simple Random Sampling. Without prior knowledge, simple random sampling is typically chosen. If there are strata delineating more or less homogeneous subgroups, these can be used to increase efficiency of the estimate (Brus, 1994; Cochran, 1977). Alternatively, regression models can be used to increase efficiency at the estimation stage (Brus, 2000). MB strategies can be used for predicting population statistics such as global means, but their most common application is in mapping some target

variable. MB strategies do not require probability sampling; the spatial distribution of the target variable in the study area is modelled as a realization of a stochastic process. For MB strategies, purposive sampling is in general more efficient than probability sampling (Brus and de Gruijter, 1997). Sampling may be guided by a geometric criterion leading to spatial coverage sampling (Walvoort et al., 2010) or may aim to minimize the variance of the prediction error or more precisely the kriging variance (van Groenigen and Stein, 1998) or universal kriging variance (Brus and Heuvelink, 2007; Zhu and Stein, 2006). For an extensive discussion of DB and MB strategies the reader is referred to Brus and de Gruijter (1997).

In our work the situation considered is one where the affordable sampling density is expected to be far too low for geostatistical interpolation (kriging). Rather, spatial prediction will be based on a modelled relation between spectral data obtained over a vast area by using both, RS data and measured target soil properties. The shape and coefficients of this relation are yet to be established using the collected data. Furthermore, given budgetary constraints, we set conditions on the sample size, the time spend on acquisition and the difficulty of reaching individual locations. This renders response-surface designs such as discussed by Meyers et al. (2011), Lesch (2005) and an equal range design as discussed by Hengl et al. (2003) infeasible, since we need to collect the sample data while minimising some cost-distance and we do not yet know the shape of the regression model; even a regression tree (Breiman et al., 1984) belongs to the possibilities. We aim to sample the feature space such that it will enable us to both select and calibrate a regression (tree) model of the relation between RS spectral data and target soil properties. A candidate sampling strategy is Latin Hypercube Sampling (LHS). If $n$ is the desired sample size, LHS stratifies the marginal distributions of the covariates into $n$ equally probably intervals and randomly samples the multivariate strata such that all marginal strata are included in the sample. While LHS is probability sampling, conditioning the LHS on any constraints and sampling costs leads to a purposive sampling strategy since the inclusion probabilities of locations are modified by the conditioning criteria. For example, remote sites may have zero inclusion probability while they do belong to the population. Conditioned Latin Hypercube Sampling (cLHS) has recently been proposed as an efficient sampling approach (Minasny and McBratney, 2006). Adamchuk et al. (2011) assessed the approach for an application in precision agriculture on a plot scale.

In this paper we apply and evaluate cLHS as proposed by Minasny and McBratney (2006), using auxiliary data from RS data to optimise a small sample (maximum of 100 points) of multiple soil properties covering a 15,000 $km^2$ area in Morocco. The optimization was conditioned by a cost function which accounts for accessibility and travel time in combination with accomplishment of an LHS. Section two entails a detailed description of the method which is next demonstrated by the case study in section three. The approach is evaluated by (1) comparing the coverage

of the covariate space and the cost of the cLHS with those of two alternative sampling strategies, (2) by comparing the thematic data obtained by the cLHS with the existing soil map, a geological map and lithological data and (3) variogram analysis of both the sparse and the exhaustive dataset. The paper concludes with a reflection on the proposed methodology.

## 3.2 Methods

### 3.2.1 Conditioned sampling

In this paper, thematic space refers to the space spanned by the target soil variables while covariate space refers to the space spanned by the covariates used for the (constrained) LHS. The aim of our sampling is to create a dataset covering the covariate space, while honouring constraints and minimizing costs related to sample size, time spent on sampling and accessibility. Assuming a relation between the two spaces, the sample would also cover thematic space. For the design of this sampling we made use of cLHS (Minasny and McBratney, 2006). Each of the marginal distributions of the covariate space is divided into equiprobable intervals that are each targeted to be sampled once. Conditioned LHS aims at allocating individual sites to each of the intervals while simultaneously honouring constraints and minimizing the costs of sampling. In doing so, costs associated with failure to cover the covariate space are weighed with the constraints on sampling. Based on Minasny and McBratney (2006) the following objective function was used (Eq. 3.1):

$$J = w1 * \sum_{i=1}^{n} \sum_{j=1}^{k} \left| \eta_{ij} - 1 \right| + w2 * \sum_{p=1}^{n} c_p \qquad \textit{Equation 3.1}$$

where $n$ denotes the sample size, $k$ is the number of variables, $\eta_{ij}$ is the number of times that an interval $i$ for variable $j$ is sampled and $c_p$ is the cost associated with sampling point $p$. The weights *w1* and *w2* specify the relative weight of the cLHS component and the sampling costs, respectively.

The method was implemented by the following steps:

(1) Define the spatial population. i.e. the spatial population comprises the exhaustive sample of locations within the study area.

(2) Select the covariates.
The main criteria for selecting covariates include: (1) representative covariates are expected to have high correlation with the target variable, (2) the number of locations where predictive covariates are available needs to be much larger than the intended sample size and (3) covariates are cheaper to

measure than the target variable itself (Davis, 2002; Webster and Oliver, 2007).

(3) Parameterize the objective function (Eq. 3.1).
Constraints are modelled by setting high costs to a site violating the constraints. Potential constraints are the maximum distance to travel off-road, maximum slope steepness, water bodies and areas which are inaccessible or forbidden to enter. Cost may for example, increase linearly with increasing distance from the road and increasing steepness of the landscape. In this study the weights *w1* and *w2* were based on a series of trials, as explained below (see par. 3.3.1.4.).

(4) Implement and decide on optimizer.
The objective function (Eq. 3.1) eventuates the following problem: Given population $N$ with covariate data ($X$), select $n$ sites ($n < N$) so that the multivariate distribution of $X$ is maximally stratified while the cost component is minimized. To solve this combinatorial problem, a heuristic optimiser such as simulated annealing is required (Kirkpatrick et al., 1983).

(5) Run the optimization and evaluate.
The suggested sampling scheme should be evaluated carefully since the weights *w1* and *w2* influence the scheme strongly, so some experimentation may be required to obtain satisfactory results (see step 3).

The final sample will be an (approximate) LHS representation of the covariate space. Implementation of the resulting sampling strategy is assumed to provide a sample representing the major soil variability within the study area.

### 3.2.2 Auxiliary data from remote sensing

Several soil forming factors (climate, organisms, relief and parent material) of the State Factor Equation of soil formation (Jenny, 1941) have been found to be expressed in spectra such as recorded by RS satellites (Buis et al., 2009; French et al., 2005; Schmidtlein et al., 2007; Singhroy et al., 2003). The use of RS data as proper covariates for soil mapping has a proven track record (Boettinger et al., 2010) and previous research on vegetation monitoring demonstrated the usefulness of RS data as covariates for cLHS (Lin et al., 2011). These findings strengthen our premise that RS imagery provides useful covariates for sampling soil variability. Also, the extent and spatial coverage of RS data as well as their relatively low cost indicate their potential usefulness as covariates (see step 2 above). The first few Principal Components (PC) of spectral image data do represent the major orthogonal axes of variability in the reflectance data but using reduced dimensionality (Wold et al., 1987). PC analysis is one of the most common methods to reduce data dimensionality arising from high resolution spectrometers (Mulder et al., 2011b and references therein). Digital elevation data can be used as a proxy for the factor ‘relief’ (Debella-Gilo and Etzelmüller, 2009; McBratney et al., 2003).

## 3.3 Case study

### 3.3.1 Implementation of the RS-based cLHS approach

In this section the procedure as described in section 3.2, is demonstrated step-by-step with a case study in Morocco.

*3.3.1.1 Study area*

The regional case study is located in Northern Morocco, centred at around 34.0º N, -4.5º W and covers an area of 15.000 km$^2$ (Fig. 3.1). While the Rif Mountains, an area of highlands, form the northern border, the Anti-Atlas mountain range is the southern border with areas of plateaus and intermountain valleys in between. Elevation ranges between +4 and + 2350 meters a.s.l. The climate is typically warm temperate with dry and hot summers (Kottek et al., 2006). The main land use is dominated by a mosaic of vegetation and croplands, bare areas, sparse vegetation and open evergreen forest (European Space Agency GlobCover Project, 2008).

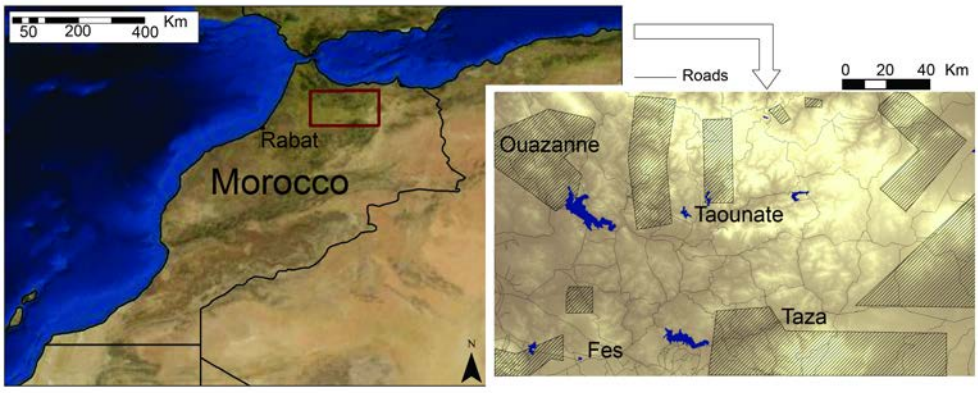


Figure 3.1: Study area, located in North Morocco, grey shades refer to elevation (Cartography by Navteq Inc., 2011). The pre-defined sub-areas are indicated by the striped areas.

*3.3.1.2 Selection of predictive covariates*

The covariate space used in cLHS consisted of the first three principal components from the Advanced Spaceborne Thermal Emission and Reflectance radiometer (ASTER) imagery and the ASTER GDEM (digital elevation model). Compared to other multispectral sensors offering a global coverage, ASTER has a relative high spatial and spectral resolution and favourable positions of the spectral bands for the retrieval of soil properties (METI/ERSDAC, 2009). ASTER has been proven to be a useful data source for the retrieval of various soil properties and NDVI (Mulder et al., 2011b; Tucker et al., 1985).

The ASTER sensor records data in 15 spectral bands of which 4 in VNIR with a resolution of 15 m, 6 in the SWIR with a resolution of 30 m and 5 in the TIR region with a resolution of 90 m (Abrams and Hook, 2001). A mosaic of these images was taken in the same season over the years 2005 and 2007. The ASTER GDEM is a digital elevation model (DEM) at 30 m resolution generated by image matching of ASTER imagery (METI/ERSDAC, 2009). The first three Principal Components (PC) of ASTER VNIR-SWIR (30 m resolution) were used because they represented the major variance (>90%) present in the reflectance data in 3 uncorrelated variables. The latter, however, is not required for LHS (McKay et al., 1979).

*3.3.1.3 Parameterization of constraints and cost*

The field campaign was to be completed within two weeks. Considering the size of the study area, sparseness of the road network and the available transport, it was estimated that 100 sites could be visited during the campaign. To accomplish a sample size of 100 sites within two weeks, the following constraints/conditions were set: (1) the cost to travel off-road was set to increase linearly (1000 units/km) with increasing distance up to a maximum of $5.5 * 10^5$ units, that is the maximum cost reached by the maximum distance from a road within the area; (2) the costs for slopes steeper than $45^0$, rivers and lakes were set to $1.0 * 10^6$, to prevent these sites from being selected during the optimization for the cLHS. Areas which are inaccessible or forbidden to sample were masked out.

*3.3.1.4 Optimization of the constrained LHS*

The weights of *w1* and *w2* were set after running a series of trials, as follows; by systematically increasing *w2* from 0 to 0.5 in steps of 0.1 and with $w1 = 1 - w2$, the spatial distribution of the points within the subareas was visually assessed and the distance to the road network was calculated. By setting *w1* and *w2* to 0.9 and 0.1, respectively, all sites were located within 5 km distance to the road network, which was deemed feasible. The subareas shown in figure 3.2 were chosen for practical reasons, i.e. computational feasibility (simulated annealing is computationally heavy) and accessibility of the terrain. The Latin hypercube computed over the full image was to be realized within these subareas. For computational reasons, histogram

computations of the marginal distribution were based on data from the central cells of a 3x3 moving-window. Within the subareas the cLHS was obtained by sampling the *n* equiprobable intervals of the covariates while accounting for the costs of sampling (Eq. 3.1). Optimization was achieved by simulated annealing using the cooling schedule described by Press et al. (1992), Equation 3.2:

$$p = e^{-\frac{\Delta E}{T_i}} \qquad \text{with } T_i = T_{i-1} * k \qquad \textit{Equation 3.2}$$

where $p$ is the probability of accepting a solution that worsens the objective function (configurations improving the objective function are always accepted), $\Delta J$ is the change in the objective function, $k$ a constant factor decreasing the temperature at each iteration and $T_i$ is the temperature in iteration $i$. The initial value for $T$ was set to 1 and k was fixed to 0.95. The maximum number of iterations was set to $1*10^5$. Small changes in $\Delta J$ were observed after 55000 iterations which provided good confidence that the optimum was found. The software for cLHS optimization was implemented in Matlab (Mathworks, 2009).

To evaluate the effectiveness of optimizing the objective function (Eq.3.1) with cLHS, both components of the function were calculated for a sample realized by simple random sampling and systematic random sampling (de Gruijter et al., 2006). These sampling strategies are, similar to the cLHS sample, realized within the defined subareas (Fig. 3.2).

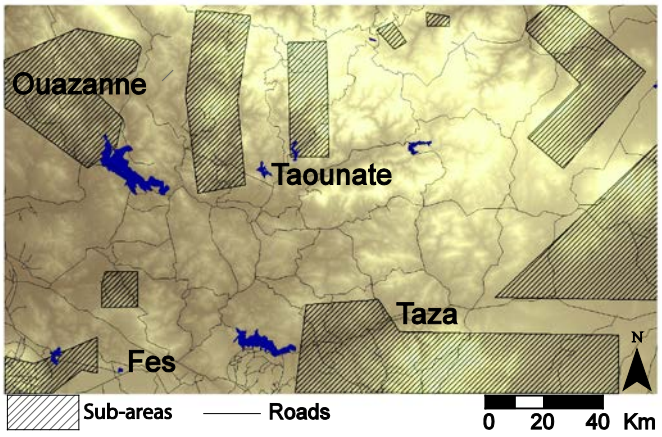


Figure 3.2: Selected sub-areas for fieldwork, as indicated by the striped areas.

### 3.3.2 In-situ data collection

At each site a 5x5 m plot was laid out. To sample the variability within the plot, a composite soil sample was taken from the top 5 cm of the soil at the corners and the centre of the plot. Sampling the top 5 cm is sufficient because ASTER data provides information of the material composition of the surface. Also, a sample of the local parent material was collected from nearby rock outcrops; this material was used for comparison with information from the Geological Map of Africa (Association of African Geological Surveys, 1963). At each location the soil profile, vegetation and landscape properties were described. Additionally, to georeference the sites, GPS-coordinates were recorded and photos were taken. The soils were classified according to the French soil classification "Classification des Sols, 1967" (Aubert et al., 1967). In the laboratory, the soil samples were dried at 70° C, sieved at 2 mm and analysed on organic matter (OM) content, pH, CEC, EC, texture and mineralogy (Baize and Jabiol, 1995; Bergmann, 1998).

### 3.3.3 Evaluation

#### *3.3.3.1 Field conditions*

During the fieldwork campaign, some sites could not be visited due to unforeseen accessibility problems and delays, leading to loss of time and consequently to a reduced sample of 73 sites. On some occasions, local experts suggested alternative locations where similar soils that could otherwise not be reached could be sampled. To assess the effects of these adaptations in the field, the realized sample was compared to the original sample by analysis of the marginal distribution of each covariate. Ideally, every interval of the marginal distributions of the covariates should have a frequency of 1.

#### *3.3.3.2 Variogram analyses*

Variograms may provide essential information for future modelling of the target variables sampled with the cLHS. We anticipate that the variograms of data obtained from the cLHS will exaggerate short-distance spatial variability since the cLHS approach favours high variability at short distances. On the other hand, variograms of exhaustively sampled covariates are expected to provide insight into the potential use of geostatistical interpolation of any voids in the surface predicted by the regression model. For calculation of the variograms, the soil property data were log transformed when deemed necessary. To assess the spatial correlation of the full spatial population, a large sample (75000) was randomly taken from the covariate space of

Table 3.1: Auxiliary data available for data comparison.

| Legacy data | Scale / resolution | Coverage |
|---|---|---|
| Local soil maps (INRA Maroc, 2010) | 1: 50.000 | Partial |
| Soil map of Morocco (Cavallar, 1950) | 1: 2M | Full |
| Geological Map of Africa (Association of African Geological Surveys, 1963) | 1: 5M | Full |
| ASTER imagery | 30m x 30m | Partial |
| ASTER GDEM | 25m x 25m | Full |

the complete research area, their variograms were computed and compared to the variograms of the covariates from the cLHS data. Calculations were performed using the *R language and environment for statistical computing* (R Development Core Team, 2011) and contributed packages *gstat* and *sp* (Didier et al., 2011; Pebesma, 2004).

*3.3.3.3 Coverage of the thematic space*

To verify whether the cLHS sample covered the thematic space, the collected soil data were compared with 'known variability' from available legacy soil data. Note that the legacy data is by no means intended as a reference data set; it is only used to verify whether reported variability is also represented by the cLHS sample. Data quality differs strongly among the data sets, especially concerning resolution and extent (Table 3.1).

Information on soil mineralogy was not directly available, however, based on the Geological Map of Africa (1:5M) and lithological data obtained from drillings, a basic map of the surface mineralogy was made. It was assumed that the geological map units relate to the lithological classes. The dominant minerals representing these lithological classes were used to determine the surface mineralogy within the geological mapping units (Table 3.2).

Table 3.2: Lithology obtained from drillings with associated minerals (*Source:* Kauffman, 2011).

| Thematic data | | | | |
|---|---|---|---|---|
| Lithological classes | Basalt | Granite | Shale | Sandstone |
| | Conglomerate | Limestone | Marlstone | Schist |
| | Dolomite | Slate | | |
| Minerals associated to the lithological classes | Calcite | Chlorite | Microcline | Feldspars |
| | Dolomite | Smectite | Kaolinite | Hematite |
| | Quartz | Muscovite | | |
| Representative mineral classes for surface mineralogy | Calcite / Clay mineral | | Quartz / Calcite / Clay minerals | |
| | Calcite / Clay mineral / Quartz | | Quartz / Chlorite / Feldspar | |
| | Feldspar / Muscovite / Clay minerals | | Quartz / Feldspar | |
| | Feldspar / Muscovite / Quartz | | | |

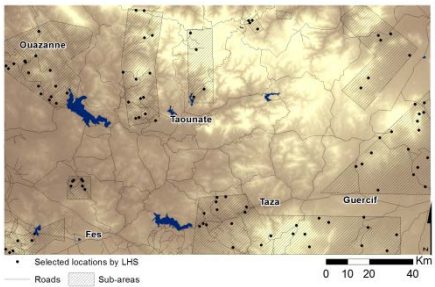


Figure 3.3: Sites selected by the LHS.

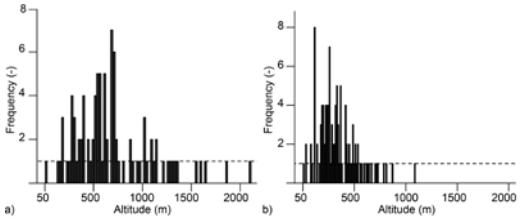


Figure 3.4: Sampled intervals for the covariate elevation for (a) simple random sampling and (b) systematic random sampling, the cLHS sample is represented by the horizontal line.

These were compared to the determined mineralogy at the sampled sites. 248 drillings with classified lithology were made available from the 'Green Water Credits'-project (Kauffman, 2011). By definition, the first unweathered layer was used as main lithology representing surface mineralogy.

## 3.4 Results and discussion

Figure 3.3 shows the sites after optimization of the sampling scheme. The 100 points were distributed all over the sub-areas in the study area. Figure 3.4 shows the coverage of the covariable elevation, where the optimal cLHS is represented by the horizontal line. With respect to the first component of the objective function (Eq. 3.1), where the cLHS is optimized to sample each interval of the covariate space, both the systematic and the simple random sample resulted in a sample not fully covering the covariate space. The second component of the objective function, the associated cost for the cLHS, was at least a factor 10 less compared to the systematic and simple random sample. These results confirm in contrary to cLHS, that these sampling strategies are suboptimal in fulfilling our sampling goals with respect to (1) coverage of the covariate space and (2) minimization of cost.

### 3.4.1 Data collection of the cLHS covariate space

Following the objective function (Eq. 3.1) and optimization as discussed in section 3.3, the results of the realized cLHS are presented in figure 3.5 a-d, where the horizontal line represents the optimized LHS. Due to missing points some intervals remained unsampled while relocation of sites caused oversampling of other intervals. The histogram of elevation shows some large gaps at higher altitudes because of the accessibility problems. Despite the relocation of sites, the histograms of the first three principal components of the ASTER imagery show less oversampling of intervals compared to the histogram for elevation. In all cases, the sample covered the full range of the marginal distributions. The gaps are expected to have only minor influence on the calibration of the prediction models described in the introduction

because the unsampled intervals are spread rather than clustered over the full range of the marginal distributions.

### 3.4.2 Coverage of thematic space

*3.4.2.1 Soil classes*

In Table 3.3 the comparison of the sampled soil classes with the Soil map of Morocco (Cavallar, 1950) and local soil maps (INRA Maroc, 2010) is shown for the single class mapping units (classes des sols). Unfortunately, eighteen field classifications could not be compared with mapped data since they were located outside the area covered by available soil maps. Twenty nine locations could be identified by a single mapped soil class and the other locations are within soil association on the Soil Map of Morocco (1:1.5M). However, these twenty nine sites do cover most of the mapped soil classes. The 'Sols hydromorphes' and 'Sols Sodiques' which are included in the legacy data set were missed by the constrained LHS sample. This might be contributed to the lack of spectral differentiation due to spectral mixing and their restricted occurrence. The two classes are reported to occupy a small proportion of the study area (1.96%). On the other hand, spectra and samples were collected of a sodic soil and a salt crust which were not included in the legacy dataset.

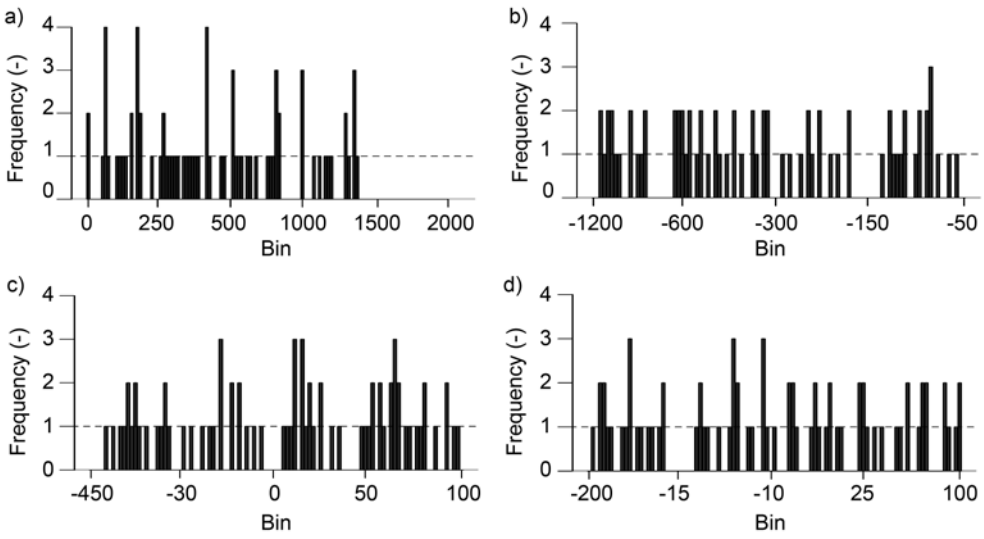


Figure 3.5: Realized samples (a) DEM, (b) PC1, (c) PC2 and (d) PC3, the optimal LHS is represented by the horizontal line.

Table 3.3: Total number of sites within each mapped soil class (Aubert et al., 1967) in the study area. Comparison based on the Soil map of Morocco (1: 1.5M) (Cavallar, 1950) and local maps (INRA Maroc, 2010).

| Classes des sols | Explanation | No. of sites |
|---|---|---|
| Classe de sols minéraux bruts | Shallow soils without well- defined horizons | 3 |
| Classe de sols peu évolués | Juvenile soils | 9 |
| Classe des vertisols | Vertic properties | 4 |
| Classe des sols calcimagnésiques | Soils enriched in $Ca^{2+}$ and $Mg^{2+}$ | 5 |
| Classe des sols isohumiques | Isohumic properties | 3 |
| Classe des sols brunifiés | Brunified soils | 1 |
| Classe des sols à sesquioxydes de fer | Soils with sesquioxidic concretions of iron | 1 |
| Classe des sols ferrallitiques | Soils with evidence of fersiallitic weathering | 3 |
| Classe des sols hyromorphes | Hydromorphic properties | 0 |
| Classe des sols sodiques | Salt affected soils | 0 |

### *3.4.2.2 Lithology*

The lithology for the rocks collected in the field was in agreement with the lithology obtained from the drillings (Table 3.2). A total of 57 sites were correctly assigned to the representative mineral classes for surface mineralogy, as described in section 3.3.3.3 (Table 3.4). The remaining 20 sites did not correctly represent the mapping unit at the specific site, however, this might be caused by the coarse resolution of the geological map, which fails to represent local variability.

### *3.4.2.3 Soil properties*

Table 3.5 shows that the sampled soil properties covered the range of soil texture, OM, pH, CEC and EC reported for the soil types (IUSS Working group WRB, 2006) present in the area according to the legacy data. Figure 3.6 shows that the sampled texture classes were in agreement with the legacy data (Aubert et al., 1967). These results show that both the spectral (section 3.4.1) and thematic space were well covered by the cLHS sample.

Table 3.4: Total number of sites within each mapped soil class (Aubert et al.,1967) in the study area. Comparison based on the Soil map of Morocco (1: 1.5M) and local maps.

| Mineral class | No. of sites |
|---|---|
| Calcite / Clay mineral | 1 |
| Calcite / Clay mineral / Quartz | 16 |
| Feldspar / Muscovite / Clay minerals | 12 |
| Feldspar / Muscovite / Quartz | 2 |
| Quartz / Calcite / Clay minerals | 1 |
| Quartz / Chlorite / Feldspar | 24 |
| Quartz / Feldspar | 1 |

Table 3.5:Range of sampled soil properties from laboratory analysis.

| Soil properties | Range |
|---|---|
| Clay (%) | 5.0 - 47.5 |
| Silt (%) | 2.5 – 53.0 |
| Sand (%) | 0.7 - 75.1 |
| $pH_{H2O}$ | 5.3 - 8.6 |
| $pH_{KCl}$ | 4.0 - 8.7 |
| Organic Matter (%) | 0.28 - 9.30 |
| CEC (meq%) | 6.50 - 77.20 |
| EC (ms/cm) | 0.23 – 248.0 |

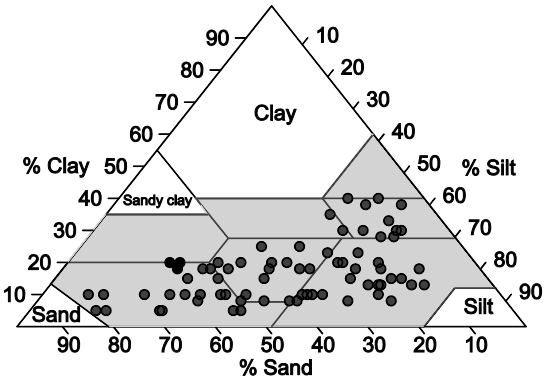


Figure 3.6: Ternary diagram of soil texture with the sampled texture classes. The shaded area represents the texture classes according to legacy data; points represent the individual samples.

### 3.4.3 Variogram analysis

*3.4.3.1 Variogram analysis of the sampled soil property data*

Figures 3.7 a-f present the variograms calculated for the soil property data. Please note the differences in the y-axes used, caused by the different measurement units and variances. None of the figures show evidence of spatial correlation, i.e. the semivariance does not systematically increase with increasing distance between the pairs of points considered. Accordingly, all variograms lack a clear structural component and they can be modelled by a pure nugget effect. These results are no surprise: the cLHS approach favoured sampling high variability within short distance from the road network.

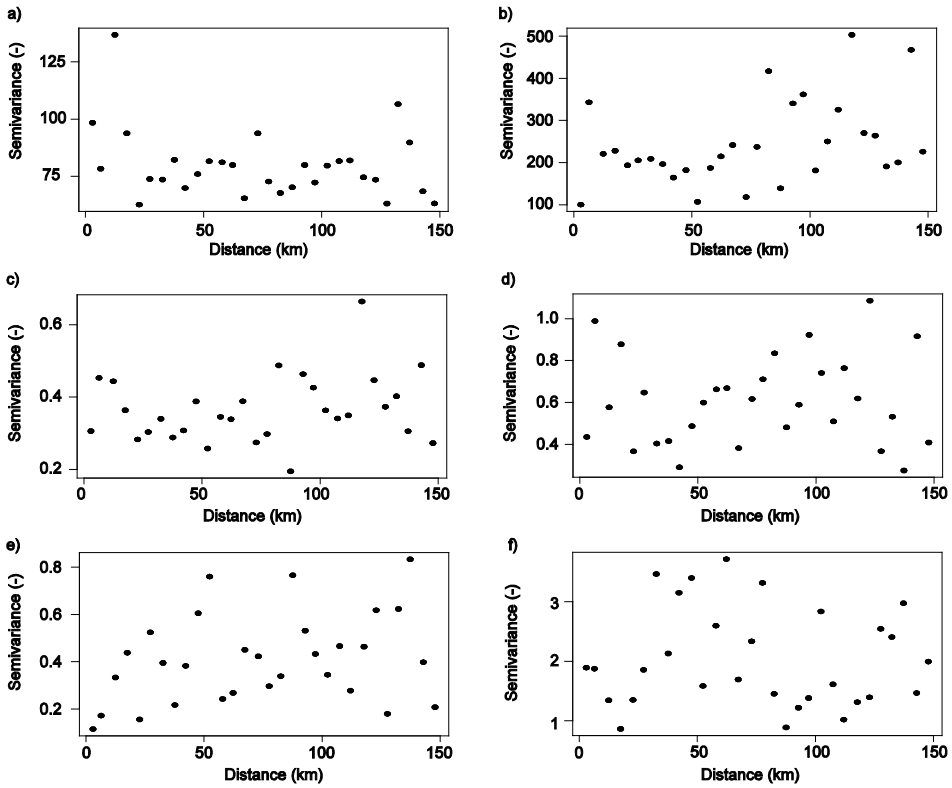


Figure 3.7: Variograms calculated for the soil property data from the LHS sample, (a) clay (%), (b) total silt (%), (c) Log (total sand (%)), (d) Log (Organic Matter (%)), (e) pH-$H_2O$ and (f) Log (Electric Conductivity).

### *3.4.3.2 Variogram analysis of the cLHS covariate space*

Figure 3.8 (left) shows the variograms for the realized cLHS covariate space. Contrary to expectations, the semivariances for PC3 attain higher values than those of PC2. This can be attributed to the sampling effect; the sample has one or a few observations in a far tail which affects the semivariance in several (notably short) distance classes. Unlike our expectations, figure 3.8a does suggest spatial structure of elevation data in the cLHS. However around a lag distance of 30 km the experimental semivariances show erratic behaviour, which did not disappear when manipulating the bin sizes (not shown here). Similar behaviour can be observed in figures 3.8c and g. Summarizing, most variograms in figure 3.8 (left) show weak spatial structure at best, particularly when compared to the variograms of the large sample (Fig. 3.8 right). Again this behaviour can be contributed to the set distance criteria which resulted in an overestimation of the variograms nugget and short distance variability.

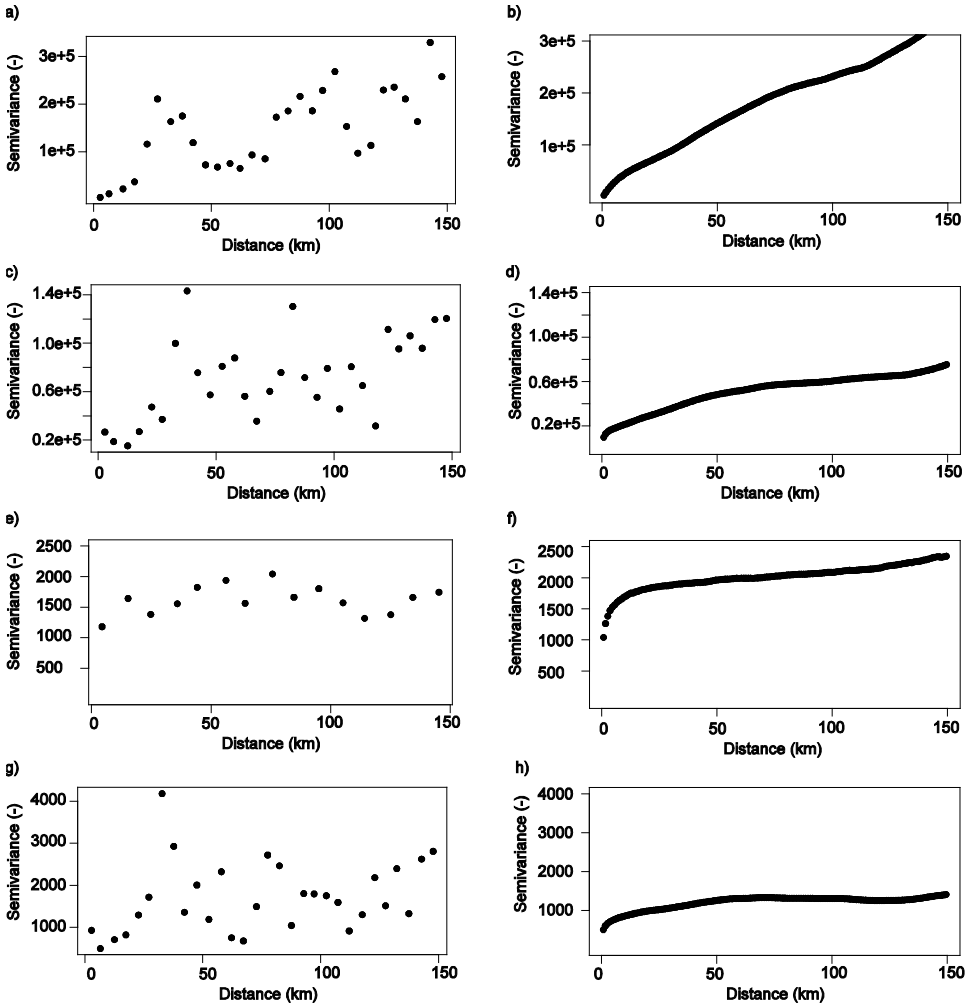


Figure 3.8: Variograms calculated for the LHS covariates : (a) DEM LHS sample, (b) DEM large sample, (c) PC1 LHS sample, (d) PC1 large sample, (e) PC2 LHS sample, (f) PC2 large sample, (g) PC3 LHS sample and (h) PC3 large sample.

The variograms of the large sample indicate that RS-data are strongly spatially correlated (Fig. 3.8 right). In contrast to the approximate cLHS sample, the variograms of the largely sampled first three principal components follow the theoretical pattern of decreasing variances (Wold et al., 1987; see the sills of the variograms). The strong correlation of the covariate data supports our premise that once the relation between RS and soil properties has been modelled, the RS data can provide the spatial basis for mapping soil properties.

## 3. 5 Conclusions

The RS-based sparse sampling approach presented and evaluated in this paper aims to sample variability in soil properties. Subsequently, the sample data will be used to characterize the relationship of the chemical and physical properties with their spectral characteristics measured with PS and RS.

The main findings of this paper are; (1) Latin Hypercube Sampling constrained on the accessibility of sites can produce a relatively small sample that represents the spectral variability within the considered covariate space, (2) the dataset obtained by the constrained LHS, represented variability in soil properties when compared to legacy data and (3) the constrained LHS sample lacked spatial correlation, because the procedure favoured large spatial variability within a constrained neighbourhood. This implies that a LHS sample constrained on accessibility is unsuitable for variogram estimation. Consequently, the data are unsuitable for spatial prediction models using kriging, which makes sense because no criteria on spatial coverage and distribution are included in the target function.

However, the acquired dataset is expected to provide information for studying soil-landscape relationships originating from the State Factor Equations of soil formation (Jenny, 1941) based on spectral data. The imagery is then used for providing spatial and spectral information on a regional scale, which makes sense because the variograms of the large sample showed that the RS data are strongly spatially correlated with ranges up to several decades of kilometres. The cLHS sample provides the data needed for fitting quantitative relationships between the target soil properties and the covariates (McBratney et al., 2003).

It is therefore expected that the RS-based sparse sampling approach is time and cost efficient for regional-scales soil survey. If data availability is scarce, this approach provides a small dataset with soil properties that can next be used for mapping soil properties using modelled relations between properties and spectra; once the relation between RS and the soil properties has been modelled the latter can be spatially predicted, based on the densely sampled RS data.

**Acknowledgements**
We thank INRA Maroc for their support with the fieldwork campaign. We acknowledge financial support from the EU FP7 program under the e-SOTER project (contract 211578). We thank the ‘Green Water Credits’-project for making the lithological data available.

# Chapter 4

# Retrieval of composite mineralogy by VNIR spectroscopy

*This chapter is based on:*
V.L. Mulder, S. de Bruin, M.E. Schaepman (2012). Retrieval of composite mineralogy by VNIR spectroscopy. *Digital Soil Assessments and Beyond – Proceedings of the Fifth Global Workshop on Digital Soil Mapping*, Sydney, Australia, pp. 373-376. DOI: 10.1201/b12728-73

## Abstract

This research analyses the complexity of retrieving composite soil mineralogy from proximal sensed data. Using measured spectra from soil mixtures, retrieval of composite mineralogy is hampered by the co-occurrence of absorption features at similar wavelengths. Complex interactions between different mineral components may further limit retrieval of separate endmembers. Accounting for the co-occurrence of absorption features but eluding the complex interaction between the components, an advanced classification routine for spectroscopic measurements was tested to assess composite mineralogy in terms of presence and absence of minerals. This was done by linearly forward modelling of the spectrum from the fraction of the constituents within the sample. The classification routine was applied both to the modelled and the measured spectra. Results show that by using the modelled spectra more minerals can be retrieved with higher accuracy because the diagnostic absorption features of the minerals were manifested. Due to the complex interaction between the minerals, features were less distinct or even absent in the measured data which hampered the classification routine to correctly retrieve composite mineralogy. Post-classification indicates that the measured samples can be grouped by dominant mineralogy due to the high spectral similarity between the modelled and measured spectra. This method is deemed particularly useful for regional-scale studies after generalization of the minerals into mineral groups.

## 4.1 Introduction

Visible and Near Infrared (VNIR) spectroscopy has proven to be an efficient method for the determination of various soil properties (Mulder et al., 2011b). From these properties, soil mineralogy is an important indicator for soil formation and parent material characterization. The standard determination of mineralogy by powder diffraction is costly and labour-intensive and VNIR spectroscopy might be an efficient alternative. Currently, individual soil minerals can often be determined by their characteristic absorption features and using spectral unmixing techniques (Mulder et al., 2011b). Nevertheless, simultaneous retrieval of multiple minerals is limited by the co-occurrence of absorption features at similar wavelengths, nonlinear mixing (or scattering) phenomena (Singer, 1981; Sunshine et al., 1990), and measurement noise (Stenberg et al., 2010). The question of whether linear or nonlinear processes dominate the spectral signatures of mixed spectra is still an unresolved issue. The linear approach has been demonstrated to be an useful technique for interpreting the variability of mineralogy in remote sensing (RS) data (Clark et al., 2003; Rezaei et al., 2011). Although these methods seem to be working to a certain quality, when using proximal sensing (PS) data the linear approach becomes invalid. Because, the linear approach is only strictly valid for the situation where the endmembers are arranged in discrete, segregated patches on the surface. This condition is never met when unmixing soil constituents since they exist in intimate association with one another (Singer, 1981; Sunshine et al., 1990).

Linear unmixing approaches, such as Multiple Endmember Spectral Unmixing Analysis (MESMA), have difficulties with deriving multiple endmembers for mineral mixtures because they assume the reflectance being a combination of linear functions as required for these analysis routines (Mulder et al., 2011a). Alternatively, the methods based on Tetracorder aim to match diagnostic absorption features with spectra from a large spectral library (Clark et al. 2003). As such, it does do not directly account for non-linearity but reduces the influence of non-linear behaviour. The results obtained with the Tetracorder show that many different minerals can be identified, however, the retrieval of mixtures with the current approach remains limited (Clark et al., 2003).

Therefore, in this paper we analyse if the retrieval of composite mineralogy with the Material Identification and Characterization Algorithm (MICA) (Kokaly, 2011) is improved by using linearly modelled spectra. By linearly modelling the spectrum from the fraction of the constituents within the sample the complex interaction between the components is avoided while the co-occurrence of absorption features is accounted for. The performance of PRISM in retrieving composite mineralogy is evaluated by comparing the results from the modelled spectra with measured spectra.

## 4.2 Material and methods

### 4.2.1 Materials

The study area was located in Northern Morocco, centred at 34,0ºN, -4,5ºW and covers an area of 15,000 $km^2$, with the Rif Mountains being the northern border and the Anti-Atlas Mountains being the southern border. At 73 sites, a mixed soil sample from the corners and the centre of a 15*15 meter plot was taken of the top 5 cm of the soil. The soil samples were dried at 70° C, sieved and crushed to a powder (<65 µm). The total sample of 77 observations was collected by a predefined sampling scheme which was designed in such a way that the variability in soil properties, including mineralogy, was fully sampled (Mulder et al., 2012a).

Mineralogy of the samples was determined on non-oriented powder specimens with X-ray diffraction analysis. A second sample preparation for clayey materials produced oriented specimens, thereby facilitating their identification. The changes in the reflex positions in the XRD pattern by intercalation of ethylenglycol were used for identification of smectite. X-ray measurements were made using a Bragg-Brentano Theta-2-Theta diffractometer Philips PW1820 using Cu Kalpha radiation. The qualitative phase composition was determined in comparison to the ICDD PDF database using the software DIFFRACplus (Bruker AXS, 2012). The quantitative mineral composition of the samples was calculated by Rietveld analysis using the Rietveld program AutoQuan (Bergmann, 1998; GE SEIFERT; Kleeberg, 2005).

Spectral measurements were taken under laboratory conditions with an ASD Fieldspec Pro FR spectroradiometer covering the 350-2500 nm wavelength region. The samples were illuminated using a stabilized quartz-tungsten light source with a condenser and Koehler illuminator, securing isotropic illumination within the field-of-view. In order to obtain a measurement field-of-view of 2 cm, a $5^0$ foreoptics was used and the sensor was placed 23 cm above the sample. The powdered samples were placed in pure crystal sample holders to reduce scattering from the sample holder, since crystal lacks distinct absorption features. Preprocessing of the spectral data was done in IDL ENVI. The determination of composite mineralogy was performed with the MICA module that is part of the 'USGS Processing Routines in IDL for Spectroscopic Measurements' (PRISM) software (Kokaly, 2011) which includes the processing routines of Tetracorder (Clark et al. 2003).

### 4.2.2 Methods

Linearly modelling the spectrum from its components eludes complex interaction between the minerals and therefore it is expected that the diagnostic features of individual constituents remain apparent. In contrast, soil samples may represent an intimate mix of mineral constituents, which implies that specific spectral properties

combine in a complex way which affects features that are diagnostic of the individual minerals (Singer, 1981). The linear model is described by:

$$R_m = \sum_{i=1}^{n} \rho_i * \mathrm{f}_i \qquad \textit{Equation 4.1}$$

Where $n$ denotes the mineral components within the sample, $\rho_i$ the reflectance curve of component $i$, $f_i$ the fraction of component $i$ and $R_m$ the modelled reflectance curve of the measured sample.

The complex interaction between constituents could influence the reflectance curve significantly. Although, the linear model does not include these effects, the final reflectance curve ($R_m$) must be representative for the measured reflectance; if the two spectra appear to be uncorrelated, further analysis concentrating on the relation between the measured and modelled spectra will not contribute to an operational method for the retrieval of composite mineralogy. Therefore, their spectral similarity was judged based on the spectral angle between the two spectra (Kruse et al., 1993). A spectral angle smaller than 0.1 radians was considered as the threshold, indicating similarity.

The classification routine is based on the current expert system decision-making framework of Tetracorder (Clark et al., 23). The central concept of the system is the detection of diagnostic absorption features; these features are unique to particular materials in shape. The features were isolated by continuum removal, then the shape of the absorption features were modelled and matched to references features from a spectral library.

The similarity between the diagnostic features of the sample with library spectra was quantified by using modified least squares calculation, the resulting coefficient of determination ($R^2$) was used as the final "fit" value, that is, the measure of agreement between the spectra. To discriminate between materials having similar diagnostic absorption features, constraints were included to rule out the co-occurrence of other materials (Clark et al. 2003) (Fig. 4.1). The following steps have been taken for mineralogical classification:

- Calculation of the continuum removal spectra for the spectra of (*a*) pure minerals, (*b*) the measured reflectance of the soil samples and (*c*) the linear modelled spectra of the minerals within the soil samples.
- Defining and selection of the diagnostic absorption features by determination of the shoulders and the depth of the absorption features from *a*, *b* and *c*.

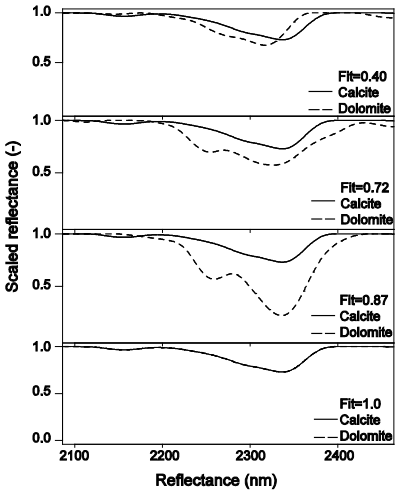


Figure 4.1. The continuum removed spectra from 4 minerals are compared to the reference spectra "Calcite" by a modified least squares calculation, the best match (Fit) to "Calcite" is calcite with a $R^2$ of 1.

- Defining and selecting the 'not features' (Clark et al., 2003) by analyzing the unique and co-occurring diagnostic features from *c*.
- Defining the weights of the diagnostic absorption features. The weights which are known to be represented poorly in the measured soil samples, as judged from the comparison of the linear mixed mineral spectra, do get a higher weight in order to be detected.
- Run the classification routine for (1) the measured spectra (*b*) and (2) the linear modelled spectra (*c*) with a spectral library containing the pure minerals (*a*). Then, match the measured samples with the modelled samples with use of the Spectral Angle Mapper (Kruse et al. 1993).

## 4.3 Results

The approach is demonstrated by sample no. 41. This sample consists of 44% quartz, 31% calcite and 7% kaolinite, 7% muscovite and 11% of residual minerals. Most of the diagnostic features occur in the SWIR and many of these features of the 4 minerals co-occur at similar wavelengths (Fig. 4.2). However, unique diagnostic features were found around 956 nm (kaolinite), 910 nm (muscovite), 1998 nm (calcite) and 1943 nm (quartz). From this figure it becomes clear that the shape and depth of the features are not a strictly linear representation of the co-occurring minerals within the measured spectra. For example, the spectra of calcite and muscovite, as well as the mixed spectra, show clear absorption

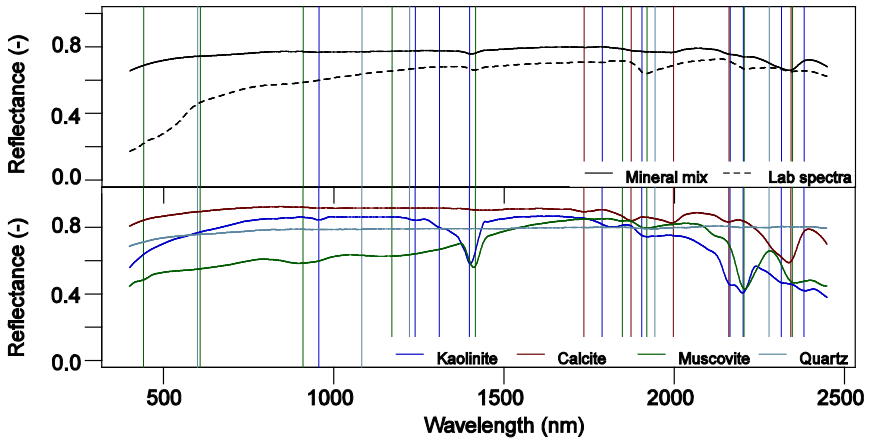


Figure 4.2. (top) Reflectance of the linear mixed and measured reflectance of sample 41, (bottom) Reflectance of the pure minerals (abundance > 5%) present in sample 41. The vertical lines indicate absorption features of the various minerals.

features at 2340 nm, however the measured spectrum shows presence of the absorption features, but less wide and deep. This can be contributed to the coating of the mineral by calcium carbonate or iron oxide thereby reducing the depth of the diagnostic absorption features of the coated mineral. Also, the spectrum of quartz does not depict distinct absorption features which will likely limit detection of its presence.

The results show the limitation of Tetracorder to retrieve the presence of minerals, with similar diagnostic absorption features within a soil mixture, by the pure minerals included in the spectral library (Table 4.1). From the measured reflectance of sample 41, quartz and chlorite were not detected in the spectrum by their absorption features, while especially quartz has a high abundance. The best match for the measured reflectance was obtained for calcite ($R^2$ = 0.84) and muscovite ($R^2$ = 0.72). Results for the modelled spectra demonstrate that more minerals were retrieved, including quartz, with higher coefficients of determination. However, the expected limitations such as coating of minerals and lack of absorption features, indeed limit the retrieval of e.g. quartz. For the complete dataset, the total and correctly observed minerals by PRISM were compared to the determined mineralogy extracted by the XRD method (Table 4.2). This table shows that the correct detection with respect to the total detected minerals was higher using the modelled spectra. In addition, the dominant mineral within the sample was correctly assigned as primary match 58 of overall 77 samples for the modelled spectra against 39 of overall 77 samples for the measured spectra.

Table 4.1: Comparison of minerals determined by X-ray diffraction with the best match ($R^2$) of minerals from the PRISM classification routine applied on the measured and modelled reflectance of sample 41.

| X-ray diffraction | | PRISM | | |
|---|---|---|---|---|
| Mineral | Abundance (%) | Mineral | Measured $R^2$ | Modelled $R^2$ |
| Quartz | 44 | Quartz | - | 0.62 |
| Calcite | 31 | Calcite | 0.84 | 0.87 |
| Kaolinite | 7 | Kaolinite | 0.58 | 0.80 |
| Muscovite | 7 | Muscovite | 0.72 | - |
| Chlorite | 4 | Chlorite | - | 0.69 |
| Feldspar | 4 | Feldspar | 0.56 | 0.63 |

Table 4.2: Comparison of mineral detection by PRISM for the measured and modelled reflectance with mineralogy as determined by X-ray diffraction (XRD) for the complete data set of 77 samples.

| Minerals | XRD | Measured | | Modelled | |
|---|---|---|---|---|---|
| | Total | Total | Correct | Total | Correct |
| Quartz | 77 | 1 | 1 | 36 | 22 |
| Montmorillonite | 35 | 10 | 4 | 52 | 35 |
| Kaolinite | 49 | 63 | 37 | 58 | 41 |
| Plagioclase | 67 | 66 | 59 | 49 | 46 |
| Muscovite | 66 | 73 | 62 | 34 | 26 |
| Goethite | 13 | 18 | 4 | 5 | 4 |
| Calcite | 62 | 61 | 51 | 61 | 52 |
| Chlorite | 50 | 8 | 6 | 30 | 24 |

Quartz does not have diagnostic absorption features in the VNIR, which results in poor detection especially for the measured data. Goethite was overestimated by the measured dataset which might be contributed to its nature to form layers around other materials instead of small grains. Plagioclase does not manifest diagnostic features related to the vibrational absorption processes in the longer wavelengths (2000-2500 nm). Instead, we used the diagnostic features in the 400-1500 nm region which arise from the electronic absorption processes of its weathering products. The underestimation of plagioclase for the modelled data can be contributed to the absence of weathering products in the pure spectra used for the model. The same arguments partly explain the difference in both data sets for muscovite. In addition, montmorillonite was overestimated while muscovite was underestimated for the modelled spectra. This can be contributed to the presence of calcite within those samples; the strong increase in reflectance of calcite at 2336 nm flattens the 2355 nm diagnostic feature of muscovite which results in a spectrum similar to montmorillonite. Concerning the measured spectra, montmorillonite did not clearly manifest the diagnostic features originating from the vibrational absorption

processes. This might well be a nonlinear mixing (or scattering) phenomena of the minerals, resulting in underestimation of the presence of montmorillonite.

Despite the high spectral similarity (average spectral angle < 0.04 rad.) of the continuum removed spectra, the modelled and measured spectra were not matched exactly with each other. However, the top five best matches of the measured spectra with the modelled spectra were in most cases samples similar in composition but deviating abundance. The best match ($R^2$=0.64) was obtained for modelled spectra with similar dominant minerals but less variability in the residual mineral composition.

## 4.4 Discussion and conclusions

This research confirmed the complexity of retrieving composite soil mineralogy from proximal sensed data. Using measured spectra from soil mixtures, retrieval of composite mineralogy was hampered by the co-occurrence of absorption features at similar wavelengths. Due to the complex interaction between the different components, retrieval of each separate endmember was limited when using PRISM. By excluding the interaction between the components while accounting for the co-occurrence of absorption features, PRISM was able to retrieve the composite mineralogy in terms of presence of minerals. By generalization of the minerals into groups this method could be particularly useful for regional scale studies where mineralogy can be used as key soil property for parent material characterization and soil formation. The expert system decision-making framework in PRISM was based on *a priori* knowledge from the data collected in the field. Application of the method elsewhere will require adjustment using data acquired from a sample covering the attribute space, e.g. such as proposed in Mulder et al. (2012a).

However these results are not optimal for local scale studies whereby more detailed information is required, such as mineral abundance, in order to predict specific soil threats (e.g. salinization or swelling and shrinking of soils). Successful methods focus on the interaction between minerals within the intimate mixture by addressing the interaction with a non-linear model. Promising methods include the photometric model of Hapke (Warell and Davidsson, 2010, and references therein) and the modified Gaussian model of absorption bands (Sunshine et al. 1990). These methods have been successfully used for extra-terrestrial studies, yet, they have not become operational for terrestrial studies focusing on soil properties. Therefore, to make optimal use of the upcoming high quality remote sensing data, we have to consider such approaches to provide the soil science community with high quality data to support their work.

**Acknowledgements**

We thank Dr. C. Mavris and Prof. Dr. M. Egli (Physical Geography, University of Zürich) and Prof. Dr. M. Plötze (IGT Institute for Geotechnical Engineering – ETH Zürich) for their support for the X-ray diffraction analysis and interpretation.

# Chapter 5

# Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1-2.4 µm) and regression tree analysis

*This chapter is based on:*
V.L. Mulder, M. Plötze, S. de Bruin, M.E. Schaepman, C. Mavris, R.F. Kokaly, M. Egli (2013). Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1-2.4 µm) and regression tree analysis. *Geoderma,* (207-208), 279-290. DOI: 10.1016/j.geoderma.2013.05.011

## Abstract

This paper presents a methodology for assessing mineral abundances of mixtures having more than two constituents using absorption features in the 2.1–2.4 µm wavelength region. In the first step, the absorption behaviour of mineral mixtures is parameterized by exponential Gaussian optimization. Next, mineral abundances are predicted by regression tree analysis using these parameters as inputs. The approach is demonstrated on a range of prepared samples with known abundances of kaolinite, dioctahedral mica, smectite, calcite and quartz and on a set of field samples from Morocco. The latter contained varying quantities of other minerals, some of which did not have diagnostic absorption features in the 2.1–2.4 µm region. Cross validation showed that the prepared samples of kaolinite, dioctahedral mica, smectite and calcite were predicted with a root mean square error (RMSE) less than 9 wt.%. For the field samples, the RMSE was less than 8 wt. % for calcite, dioctahedral mica and kaolinite abundances. Smectite could not be well predicted, which was attributed to spectral variations of the cations within the dioctahedral layered smectites. Substitution of part of the quartz by chlorite at the prediction phase hardly affected the accuracy of the predicted mineral content; this suggests that the method is robust in handling the omission of minerals during the training phase. The degree of expression of absorption components was different between the field sample and the laboratory mixtures. This demonstrates that the method should be calibrated and trained on local samples. Our method allows the simultaneous quantification of more than two minerals within a complex mixture and thereby enhances the perspectives of spectral analysis for mineral abundances.

## 5.1 Introduction

Soil mineralogy is an important indicator for soil formation and parent material characterization. Among other minerals in soils like quartz, feldspars and carbonate minerals, clay minerals are the main secondary phases formed by the weathering of the parent material. The abundance of different clay minerals and their structural features become useful indicators in defining the evolutional stage of a soil (Egli et al., 2008; Hong et al., 2013; Mavris et al., 2011; Sedov et al., 2003). In environmental and geological studies, the characterization (and quantification) of soil mineralogy is typically achieved using X-ray diffraction (XRD). XRD is broadly acknowledged as the essential tool for mineral determination of mono- or multi-mineral mixtures (Bish and Plötze, 2011; Gomez et al., 2008a; Mulder et al., 2011b; Omotoso et al., 2006). The basic limitation of XRD is that the analysis must be carried out indoors, basically due to sample preparation requirements and specific laboratory treatments necessary for some clay minerals, such as glycolation and heating after various cation saturations. Visible Near Infrared and Shortwave Infrared (VNIR/SWIR) spectroscopy has proven to be an efficient method for the determination of various soil properties since measurements can be done with little effort and in situ (Ben-Dor et al., 2009; Viscarra Rossel et al., 2006b). In this paper we propose and demonstrate its use for simultaneous quantification of mineral abundances from complex mixtures.

Some minerals such as quartz, and low iron feldspars do not show absorption features in the 0.350–2.500 µm wavelength range except for the features arising from $Fe^{2+/3+}$ related to their weathering products (Clark et al., 1990). Detection of minerals having absorption features within the 0.350–2.500 µm range have been successfully obtained using linear spectral unmixing techniques (Dennison and Roberts, 2003). However, these analyses were limited to estimating the main component within a sample having the most distinct absorption feature (Mulder et al., 2012b). Linear mixing behaviour of spectra, however, is highly unlikely in soils because the mineral constituents are typically in intimate association with one another. Influencing factors are e.g. the opaqueness of minerals and coating by other minerals. Furthermore, simultaneous retrieval of multiple mineral abundances from reflectance spectra in the 0.350–2.500 µm region is affected by the co-occurrence of absorption features at similar wavelengths arising from overtones and combinations of the fundamental absorptions of OH, $H_2O$ and $CO_3$ which occur at wavelengths greater than 2.500 µm, nonlinear mixing (or scattering) phenomena (Singer, 1981; Sunshine et al., 1990), and measurement noise (Stenberg et al., 2010). Hence, reflectance spectra of mixtures are typically a complex result from the combinations of the spectral characteristics of the constituents (Clark et al., 1990), as illustrated in figure

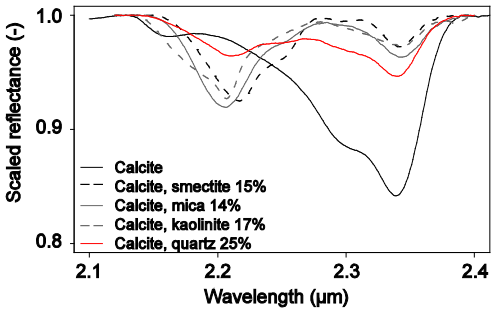


Figure 5.1: Continuum removed reflectance of calcite and mixtures containing calcite. The mixtures contain minerals of which the additional mineral with absorption features has an abundance of approximately 15% and a sample of calcite with 25% quartz (Spectra originate from field samples measured in this experiment, the calcite spectra contains a small trace of mica).

5.1. A comparison of the diagnostic features of pure calcite with the continuum removed reflectance (Clark, 1998) of samples containing a spectrally dominant mineral shows that e.g. in the presence of quartz the double absorption feature near 2.300–2.350 µm is present but much less distinct while it is absent in mineral mixtures of calcite with kaolinite or dioctahedral mica at approximately 15% abundance. Note also that the absorption near 2.150 µm is absent in the smectite and dioctahedral mica mixtures while it changes the typical absorption of kaolinite. Depending on the composition, the abundance and the spatial arrangement of the minerals, the total reflectance resulting from the scattering of the minerals within the intimate mixture produces positional shifts, changes in intensity, disappearance of absorption features or changes in their shape.

Methods aiming to match diagnostic absorption features with spectra from a large spectral library include the Tetracorder (Clark et al., 2003) and the CRISM Analysis Tool (CAT) (Flahaut et al., 2012). While the extended library enables application to unknown areas without the need of calibration on local samples, the retrieval of the mineral composition of complex mixtures remains limited because spectral mixing effects may yield diagnostic features not distinct enough to be matched to minerals in the spectral library. Theoretically, the spectra could be matched to the corresponding spectra with known abundances in the library. However the spectra to be included in the library of various minerals and the possible variation in mixtures of these would follow combinatory logic (Mulder et al., 2012b). So, the methods are commonly applied to characterizing mineral composition in terms of presence or absence but not quantifying mineral abundances (Clark et al., 2003). Non-linear models, such as the single scattering albedo model of Hapke (Hapke, 2002; Warell and Davidsson, 2010) have been successful in predicting the abundances of minerals in intimate mixtures. The main reason why such a nonlinear approach is not widely adopted is the amount of detailed information needed on the scattering properties of all endmembers to perform the calculations (Keshava and Mustard, 2002). Alternatively, the modelling

of reflectance and the inference of absorption components within complex features can be done by fitting Gaussian curves or modified Gaussian curves to the absorption features and absorption components in reflectance spectra of minerals also referred to as spectral deconvolution (Burns, 1993; Noble et al., 2006; Roush and Singer, 1986; Singer, 1981). Sunshine et al. (1990) provided the explanation for Gaussian behaviour of absorption features. The signal sensed by a spectrometer corresponds to the mean response from massive amounts of electronic and vibrational processes that cause absorption around specific wavelengths (absorption bands). Owing to the Central Limit Theorem, an absorption feature closely resembles a Gaussian distribution. Alternatively, the exponential Gaussian optimization (EGO) of Pompilio et al. (2009) has been designed to model absorption components which are not Gaussian in shape and accounts for saturation and asymmetry effects. The use of such a quantitative deconvolution method for a spectrum of a specific composition is dependent only on the spectra and absorption of the minerals themselves rather than the detailed information on the scattering properties required for the Hapke model (Shepard and Helfenstein, 2007). It provides the means to study the individual absorption components in spectra and interpretation of these can then be analysed in terms of composition (Sunshine et al., 1990).

Modified Gaussian models have been demonstrated in laboratory experiments by mixtures with two constituents of interest using either multiple linear regression techniques (Bishop et al., 2011; Singer, 1981) or the ratio between intensities of absorption components (Kanner et al., 2007; Sunshine and Pieters, 1998). It has thus been assumed that the model parameters vary as a linear function of the relative proportions of the constituents in the mixture (Pompilio et al., 2009; Sunshine and Pieters, 1998). Samples with similar mineralogy but unknown abundance can then be predicted by the calibrated mixtures models. However, such approach is insufficient for the prediction of mixtures with more than two minerals. Model parameters might vary linearly over a short range of the mixture possibilities but over the complete range of mixture possibilities, non-linearity dominates. As a result, a different type of analysis is required to relate the EGO parameters to the mineral content in order to determine abundances of three or more minerals within a mixture. We propose a recursive partitioning of the data by regression tree analysis (Breiman et al., 1984). Regression tree analysis allows dealing with nonlinearity and interactions between the EGO parameters. Regression trees can be trained by setting decision rules based on the predictive structure of the dataset with mineral mixtures (Breiman et al., 1984). This approach is an often used data mining technique in several disciplines (De'Ath and Fabricius, 2000; McBratney et al., 2003; Yang et al., 2003).

Below we give details on combining the deconvolution by EGO and the use of regression trees on the EGO parameter values for quantifying mineral abundances of mixtures having more than two constituents. The approach is demonstrated on a range of prepared samples with known abundances of kaolinite, dioctahedral mica,

smectite, calcite and quartz and on a set of field samples from Morocco, which were quantitatively analysed by XRD analysis.

## 5.2 Methods

### 5.2.1 Spectral deconvolution by Gaussian modelling of absorption components

Deconvolution of the spectra by fitting Gaussian curves needs to be concerned with partly overlapping absorptions components (Sunshine and Pieters, 1993) as well as the presence of amorphous materials and impurities that may modify absorption band shapes and contribute to saturation and asymmetry of spectral features (Burns, 1993; Pompilio et al., 2009). The Modified Gaussian Model (MGM) describes absorption components as modified Gaussian distributions that are parameterized by a band centre, band width (full width at half maximum) and band strength (amplitude intensity), for more details see Sunshine et al. (1990) and Kanner et al. (2007). In several studies MGM has been successfully used to model overlapping absorptions components (Bishop et al., 2011; Kanner et al., 2007; Lane et al., 2011; Ogawa et al., 2011; Pinet, 2007; Sunshine and Pieters, 1998). Because we have mineral mixes with many constituents, the absorption components may be modified in shape due to saturation and asymmetry effects. It is important to include a parameter on saturation. When saturation occurs the absorption band is reduced in depth and becomes flatter near the minimum. Without information on saturation the spectral deconvolution might be less accurate because it cannot discriminate between overlapping and saturated absorption bands (Pompilio et al., 2009). Therefore, the deconvolution of the spectra will be performed with the exponential Gaussian optimization (EGO) of Pompilio et al. (2009) to model those absorption components which are not Gaussian in shape and account for saturation and asymmetry effects.

The EGO algorithm of Pomilio et al. (2010; 2009) first fits a continuum linear in wavelength over the isolated absorption feature and secondly fits the feature's individual absorption components by a number of EGO profiles in log reflectance. These are then combined into the final fitted curve of an isolated feature. The continuum over the isolated feature is set according to equation 5.1:

$$C(\lambda) = c_0 + c_1\lambda^{-1} \qquad \textit{Equation 5.1}$$

where $c_0$ is the offset and $c_1$ a constant for the slope of the continuum and $\lambda$ is the wavelength (Sunshine et al., 1990). In wavelength space, the continuum becomes a flat line in infrared and a curved line at shorter wavelengths (Clénet et al., 2011). Then, for each absorption component an EGO profile is described, using the five parameters position, intensity, width, band saturation and asymmetry (Eq. 5.2):

$$EGO(\lambda) = -\frac{s}{1-e^{-0.5t}}\left[1 - e^{-0.5\left(te^{-0.5\left(\frac{\lambda-\mu}{\sigma+k(\lambda-\mu)}\right)^2}\right)}\right] \qquad \textit{Equation 5.2}$$

Where $\lambda$ represents the wavelength in micrometres [µm], $s$ the band amplitude intensity, $\mu$ the centre and $\sigma$ the full width at half maximum (FWHM) of the EGO profile. The parameter $t$ is used to model band saturation, and $k$ is the coefficient for asymmetry.

The complete model algorithm is described by the fitted continuum $C(\lambda)$ and the sum of the different EGO profiles superimposed onto the continuum (Eq. 5.3):

$$\ln(R(\lambda)) = C(\lambda) + \Sigma EGO(\lambda) \qquad \textit{Equation 5.3}$$

where $R$ is the reflectance spectrum as a function of the wavelength, $C(\lambda)$ indicates the continuum as a function of wavelength and $EGO(\lambda)$ are the individual profiles fitted to each absorption feature again as a function of wavelength. Log transformed reflectance is used since absorption is assumed having a logarithmic dependence between transmissivity of light through a substance and the product of the path length and the absorption coefficient of the substance (Beer Lambert law) (Sassaroli and Fantini, 2004). The appropriate number of absorption components to be used in the EGO analysis depends on the number of known minerals present in the material mixture and the number of unique electronic and vibrational absorptions for each mineral (Kanner et al., 2007). For a graphical demonstration of the EGO algorithm we refer to the work of Pompilio et al. (2010; 2009).

We used a similar approach as Pompilio et al. (2010; 2009) to evaluate the deconvolution of the spectra. So called *best fit* models were assessed using the root mean squared error (RMSE) of the estimated natural log of reflectance. Based on visual interpretation, additional EGO profiles were included at subsequent iterations when the current set of profiles did not achieve a proper fit in specific wavelength ranges. These EGO profiles were centred around the position (µm) of the selected absorption components, as discussed in section 5.2.3.1. A best fit of the estimated natural log reflectance results from negligible improvement of the corresponding RMSE with subsequent iterations. For a complete description of the statistics we refer to Appendix A in Pompilio et al. (2009).

The EGO routine was implemented within the R-environment by Pompilio et al. (2009) ( R Development Core Team, 2011). The approximation of the parameters is obtained by optimization of the parameters in the model aiming for the least RMSE using a Levenberg-Marquardt approach (Garbow et al., 1980; Moré, 1978; Press et al., 1992).

### 5.2.2 Determination of mineralogical composition by regression trees

Regression tree analysis (RTA) is a flexible method for specifying the conditional distribution of a variable *y*, given a vector of predictor variables *X* (Breiman et al., 1984). The goal of the regression tree is to create relatively homogeneous subsets. This is done by recursively partitioning the data in binary splits based on a single predictor variable. The partition is determined by splitting rules, evaluating the best split at each internal node with respect to homogeneity of the two subsets. Each observation is ultimately assigned to a unique terminal node based on the splitting rules set for each node in the tree (Breiman et al., 1984).

In this work, mineral abundance was predicted by coupling RTA with the EGO results whereby a separate regression tree was trained for the individual minerals. In the first instance, a maximum tree was grown to the point where additional splits could not be made due to lack of data, using the EGO profile parameters for the corresponding mineral as the predictor variables. Subsequently, the least important splits were removed by pruning the tree based upon the standard error of the estimate from cross-validation. The optimal tree was derived by using the so called *1SE* rule, introduced by Breiman (1984). By this rule we select the simplest tree where the error estimate is within one standard error of the lowest error of the estimate. Thereby we reduce the instability of the model and also the number of parameters used for the prediction model. In addition, the stopping rules used for training of the trees included (Breiman et al., 1984); the minimum number of observations that must exist in a node in order to be considered for a split were set to 3 for the laboratory experiment and to 5 for the field experiment; a split was accepted if the overall coefficient of determination ($R^2$) of leave-one-out cross validation increased at each step by at least 0.01.

An additional validation was done to test the model performance with a number of samples which were not included for training of the trees. By setting aside 16 randomly selected samples from the field experiment (total 77 samples), training of the trees was based on the remaining samples. The accuracy of the predicted mineralogy from the cross validation and the external validation was assessed by the coefficient of determination ($R^2$) from the CART analysis and the root mean square error (RMSE) of the predicted mineral abundances compared to the known mineralogy. Calculations were performed using the *R language and environment for statistical computing*, version 2.14.1 (R Development Core Team, 2011) and the contributed package *rpart* (Ripley, 2011). For more details about RTA we refer to Breiman et al. (1984).

### 5.2.3 Case studies

#### *5.2.3.1 Laboratory experiment*

Mineral samples

The laboratory experiment was conducted on spectra of physical mixtures of almost pure minerals. The minerals included in the mixtures were kaolinite, dioctahedral mica (illite), smectite, calcite and quartz, those found to be dominantly present in the study area (see section 5.2.3.2). Since quartz was ubiquitous in the study area a fixed amount of quartz was added to each sample. Quartz is known to lack absorption features in the studied wavelength range. Nevertheless, inclusion of quartz in physical mixtures is important because of potential secondary impact, from scattering of light by quartz, on the absorption features of other minerals (Clark, 1999); however, in this study the direct estimation of quartz abundance using EGO and regression tree analysis is not supported because of absence of absorption features (see also figure 5.2). The minerals were obtained either from CMS source clay repository (kaolinite KGa-2 Georgia U.S.A., Fe-chlorite CCa-1 California U.S.A.) or are industrial products (quartz and calcite from Fluka, illite Sarospatakite from Füzerradvany Hungary, and montmorillonite cloisite-Na from Southern Clay Products U.S.A.). A dataset of 35 samples was used for training the regression trees. To obtain the required opacity for the spectral measurements each sample had a gravimetric content of 15 g.

Samples 1-25 consisted of gravimetric, intimate mineral mixes of pure kaolinite, dioctahedral mica, smectite, calcite and quartz with a grain size of < 63 µm. Each of these samples had a quartz content of 25 wt. %, while the other minerals contributed to the remaining 75 wt. % with abundance as indicated in Table 5.1. A pure sample of each mineral was included as reference for the absorption characteristics of the pure minerals. With this laboratory experiment the true composition of the sample can be assumed to be known and the chance that there are unknown minerals influencing the model performance is minimized.

A sensitivity test was done to test the model performance against the presence of an unknown mineral on which the regression trees were not trained. Missing specific absorption components might reduce the statistical relationship between the EGO profiles and mineral contents included in the analysis. Therefore, in five samples chlorite was admixed, which is another common sheet silicate in soils (Dalton et al., 2004).

Table 5.1: Overview of prepared mineral samples, content is given in wt. % of the total mass of 15 g.

| Sample no. | Smectite | Kaolinite | Dioctahedral mica | Calcite | Quartz | Chlorite |
|---|---|---|---|---|---|---|
| 1 | 10 | 5 | 40 | 20 | 25 | 0 |
| 2 | 45 | 15 | 15 | 0 | 25 | 0 |
| 3 | 5 | 5 | 50 | 15 | 25 | 0 |
| 4 | 20 | 10 | 10 | 35 | 25 | 0 |
| 5 | 10 | 15 | 15 | 35 | 25 | 0 |
| 6 | 0 | 0 | 60 | 15 | 25 | 0 |
| 7 | 10 | 15 | 40 | 10 | 25 | 0 |
| 8 | 45 | 20 | 5 | 5 | 25 | 0 |
| 9 | 25 | 15 | 30 | 5 | 25 | 0 |
| 10 | 0 | 10 | 40 | 25 | 25 | 0 |
| 11 | 20 | 15 | 25 | 15 | 25 | 0 |
| 12 | 15 | 10 | 20 | 30 | 25 | 0 |
| 13 | 5 | 5 | 55 | 10 | 25 | 0 |
| 14 | 0 | 0 | 20 | 55 | 25 | 0 |
| 15 | 0 | 0 | 0 | 75 | 25 | 0 |
| 16 | 60 | 5 | 10 | 0 | 25 | 0 |
| 17 | 50 | 10 | 5 | 10 | 25 | 0 |
| 18 | 15 | 0 | 30 | 30 | 25 | 0 |
| 19 | 0 | 20 | 55 | 0 | 25 | 0 |
| 20 | 0 | 0 | 50 | 25 | 25 | 0 |
| 21 | 15 | 10 | 35 | 15 | 25 | 0 |
| 22 | 5 | 0 | 35 | 35 | 25 | 0 |
| 23 | 15 | 5 | 20 | 35 | 25 | 0 |
| 24 | 50 | 5 | 20 | 0 | 25 | 0 |
| 25 | 10 | 0 | 45 | 20 | 25 | 0 |
| Pure minerals | | | | | | |
| 26 | 100 | 0 | 0 | 0 | 0 | 0 |
| 27 | 0 | 100 | 0 | 0 | 0 | 0 |
| 28 | 0 | 0 | 100 | 0 | 0 | 0 |
| 29 | 0 | 0 | 0 | 100 | 0 | 0 |
| 30 | 0 | 0 | 0 | 0 | 100 | 0 |
| Addition of chlorite* | | | | | | |
| 31 (1) | 10 | 5 | 40 | 20 | 15 | 10 |
| 32 (2) | 30 | 10 | 15 | 0 | 20 | 25 |
| 33 (4) | 20 | 10 | 10 | 35 | 15 | 10 |
| 34 (8) | 45 | 20 | 5 | 5 | 20 | 5 |
| 35 (14) | 0 | 0 | 20 | 55 | 10 | 15 |

* Numbers within brackets refer to the corresponding number of the sample without chlorite

Spectral measurements and reflectance of minerals

Spectral measurements were performed under laboratory conditions with an Analytical Spectral Devices (ASD) Fieldspec 3 spectroradiometer. The spectroradiometer covered the 350–2500 nm wavelength region with a resolution of 3 nm at 700 nm and 10 nm at 1400/2100 nm (1.4 nm and 2nm sampling interval). The samples were measured using an ASD High Intensity Muglight (4W quartz tungsten halogen lamp) to minimize measurement errors associated with stray light and specular reflection from the minerals. The powdered samples were placed in pure quartz sample holders to avoid scattering from the sample holder, since quartz lacks distinct absorption features in the measured wavelength region. All sample holders were calibrated against a non-reflecting reference standard prior to sample measurements.

The reflectance of pure minerals — including those studied in this paper — was extensively reviewed by Clark et al. (1990) and Swayze et al. (2003). Figure 5.2 presents the continuum removed (CR) spectra of the pure minerals, the spectra show clearly the strong absorption features of water around 1.400 µm and 1.900 µm, related to the overtones and combinations of the fundamental vibrational characteristics of water as well as absorption from the OH bends. The diagnostic characteristic of kaolinite is the double absorption feature around 2.150 µm and 2.200 µm; Dioctahedral mica has two primary absorption features centred around 2.200 µm and 2.350 µm. Smectite contains water in the crystal structure and therefore exhibits strong absorption near 1.400 µm and 1.900 µm. In the SWIR, it can be recognized by the single sharp and symmetrical absorption at 2.200 µm due to the AlOH bend and the smaller absorption around 2.250 µm. For calcite, second and third overtones and combinations of the $CO_3$ fundamentals occur in the near IR. The two strongest absorptions are found at 2.500–2.550 µm and 2.300–2.350 µm. Three weaker absorption bands occur near 2.120–2.160 µm, 1.970–2.000 µm and 1.850–1.870 µm. In the SWIR region, calcite is distinguished from the other minerals by the weaker absorption near 2.120–2.160 µm and the double absorption feature near 2.300–2.350 µm (Clark et al., 1990). Based on the bends which discriminate the minerals from each other we study the wavelength range between 2.100–2.400 µm. Within this wavelength range quartz does not have any absorption feature. Direct estimation of quartz abundances using EGO and regression tree analysis is therefore not supported. As indicated earlier, it is important to include quartz into the mixtures. The presence of quartz affects the scattering of light, thereby affecting the spectral features of the other minerals within the mixture. While the shape of the CR feature of a mineral is unlikely to be affected by quartz, the depth, asymmetry and saturation of the fitted Gaussians will likely be affected (Clark et al., 1990).

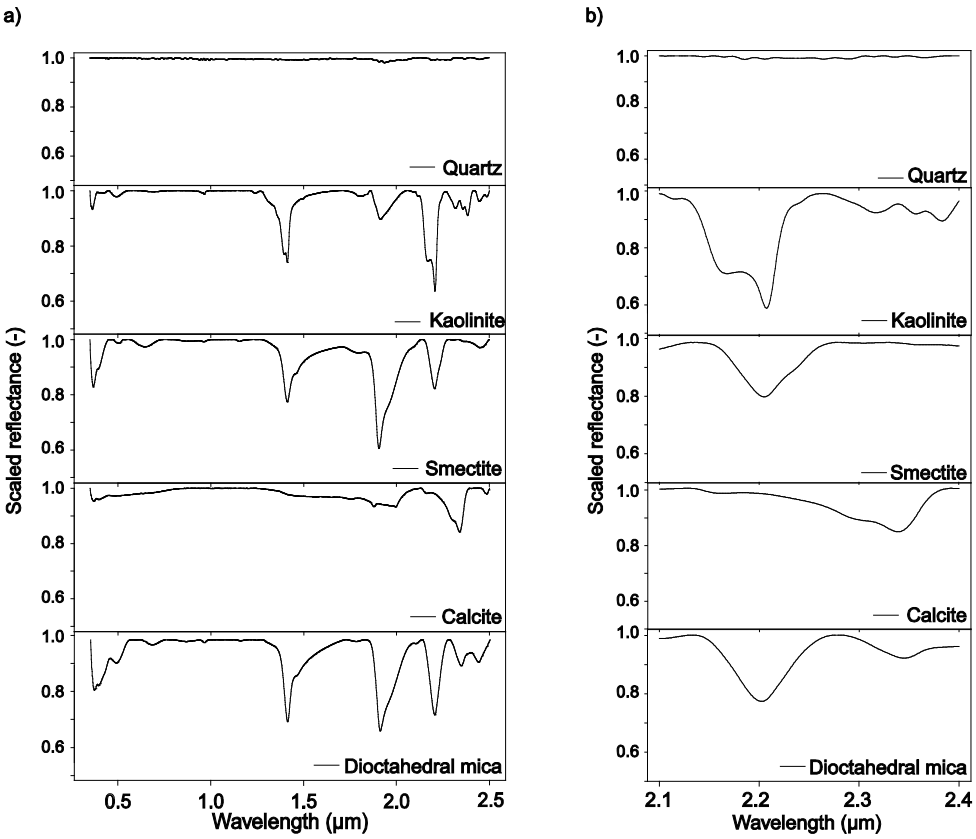


Figure 5.2: Continuum removed spectra from the pure minerals in this study, right: full VNIR - SWIR, left: between 2.1 – 2.4 µm (Spectra originate from samples measured in this experiment).

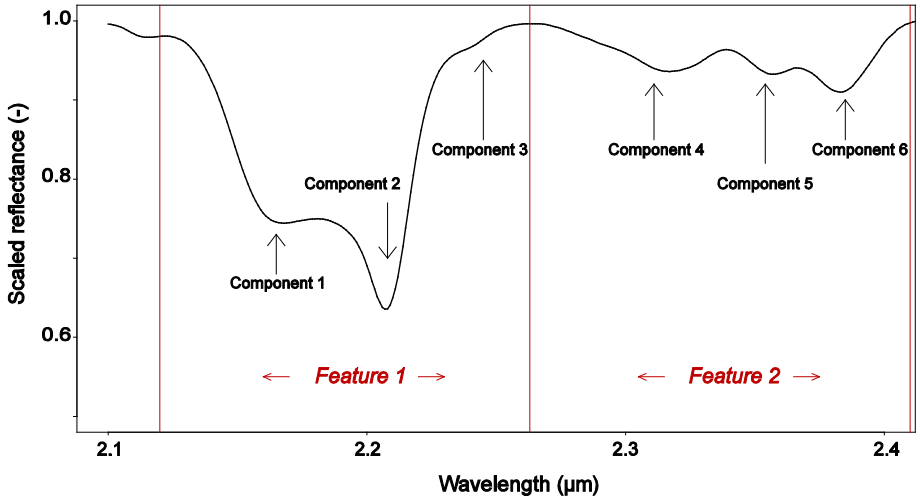


Figure 5.3: Continuum removed spectrum of kaolinite where the absorption components within the 2.100 - 2.400 µm region are indicated along with the two spectral regions within the studied wavelength range (red lines).

Selection of absorption components for curve fitting by EGO

Based on the number of minerals and their specific absorption behaviour, as discussed in the previous section, we expected that six individual absorption components centred around 2.170 µm, 2.210 µm, 2.250 µm, 2.310 µm, 2.350 µm and 2.380 µm would be required to optimally model abundances. From here on, these components are referred to as absorption components 1 to 6. For those spectra having 2 absorption features within the 2.100–2.400 µm region, the reflectance was split into the two primary absorption features around 2.120–2.270 µm and 2.280–2.400 µm, referred to as "feature 1" and "feature 2", respectively. Figure 5.3 illustrates the positioning of the individual absorption components and features using the spectrum of pure kaolinite (Fig. 5.3).

*5.2.3.2 Field experiment*

Field sampling

The regional case study was located in Northern Morocco, centred at around 34.0ºN,-4.5ºW and covers an area of 15,000 km$^2$. While the Rif Mountains, an area of highlands, form the northern border, the Anti-Atlas Mountains is the southern border with areas of plateaus and intermountain valleys in between. This area offered a diverse lithological setting including sedimentary, igneous and metamorphic rock types. For training of the regression trees, a sample was collected which covered the variability in mineralogy present in the study area. We designed a sparse, remote sensing-based sampling approach making use of conditioned Latin Hypercube

Sampling (cLHS) to assess variability in soil properties at regional scale (Mulder et al., 2012a).

At 73 sites, a mixed soil sample from a 15×15 m. plot was taken of the top 5 cm of the soil. Two additional samples originating from a soil profile and duplicate samples of 2 sites were included which resulted in a total sample size of 77. The soil samples were dried at 70 C, sieved and crushed to a powder (<20 µm) sieved and crushed to a powder (<20 µm) with a McCrone micronising mill in ethanol. To improve mineral quantification by X-ray diffraction, the organic matter was destroyed for samples having organic matter content higher than about 10%. Spectral measurements were taken with the same setup as described in section 5.2.3.1 and the absorption components used for curve fitting and training of the regression tree were identical to those described in section 5.2.3.1.

X-ray powder diffraction for analysis of soil mineralogical composition

Mineralogy of the fine powdered samples (< 0.20 µm) was determined on randomly oriented powder specimens with X-ray diffraction analysis. The second sample preparation for clayey materials produces oriented specimens, which enhanced the basal reflections from the clay minerals thereby facilitating their identification. The changes in the reflex positions in the XRD pattern by intercalation of ethylene glycol were used for the identification of smectite. X-ray measurements were made using a Bragg-Brentano Theta-2Theta diffractometer Philips PW1820 using Cu Ka alpha radiation. The instrument was equipped with an automatic divergence slit, a graphite monochromator and a NaI scintillation counter. The powder samples were step-scanned at room temperature from 2 to 70°2Theta (step width 0.03° 2Theta, counting time 4 s). The qualitative phase composition was determined in comparison to the ICDD PDF database using the software DIFFRACplus (Bruker AXS, 2012). The quantitative mineral composition of the samples was calculated by Rietveld analysis using the Rietveld program AutoQuan (GE SEIFERT) (Bergmann, 1998; Kleeberg, 2005).

The ranges and variety of minerals within the field samples, determined by XRD analysis, are presented in Table 5.2. Due to the similar structure of illite and muscovite, the minerals were not separated with the XRD analysis and are treated as the sum of dioctahedral mica in the field experiment. The interstratified illite/smectite minerals are abbreviated as I/S ML. This mineral type was not defined more in detail with respect to the ratio illite/smectite and order of layering. For the experiments those minerals were included which (1) had absorption features in the 2.100–2.400 µm region (Fig. 5.4), (2) were frequently present in the samples and (3) had an average content larger than 5 wt. %. Following these criteria, calcite, dioctahedral mica, smectite and kaolinite were considered for in this study. The other minerals were deemed either subordinate to the selected minerals or did not contain the required spectral features to estimate abundances using spectral unmixing

Table 5.2: Ranges of the mineral abundance (wt. %) present in the field samples, as determined by XRD analysis.

| Mineral | Average (wt. %) | Range (wt. %) | Error ($3\sigma$) (wt. %) | No. samples |
|---|---|---|---|---|
| Anatase | 1.0 | 0.4 – 1.9 | 0.2 - 0.3 | 41 |
| Aragonite | 7.1 | 1.9 – 11.1 | 0.7 – 0.8 | 3 |
| Calcite | 21.4 | 0.7 – 75.3 | 0.3 – 1.6 | 63 |
| Chlorite | 5.0 | 0.8 – 25.4 | 0.4 – 1.8 | 52 |
| Cristobalite | 2.8 | 1.2 – 4.5 | 0.3 – 0.5 | 3 |
| Epidote | 1.9 | 1.9 | 0.5 | 1 |
| Dolomite | 6.9 | 0.8 – 33.9 | 0.3 – 0.9 | 29 |
| Goethite | 5.2 | 1.2 – 11.6 | 0.8 – 1.3 | 14 |
| Gypsum | 1.5 | 1.5 | 0.5 | 1 |
| Hematite | 1.4 | 0.4 – 4.6 | 0.2 – 0.4 | 27 |
| Dioctahedral mica | 10.5 | 1.6 – 56.9 | 0.4 – 2.1 | 66 |
| Kaolinite | 6.7 | 1.3 – 16.7 | 1.1 – 2.7 | 49 |
| K-feldspar | 3.1 | 0.9 – 18.6 | 0.5 – 1.3 | 42 |
| Palygorskite | 6.8 | 6.8 | 0.8 | 1 |
| Plagioclase Ab | 4.9 | 1 – 23.8 | 0.4 – 1.2 | 69 |
| Quartz | 42.5 | 12.6 – 83.9 | 0.6 – 1.8 | 77 |
| Rutile | 1.1 | 0.6 – 1.8 | 0.2 – 0.4 | 10 |
| Smectite | 28.6 | 9.4 – 52.9 | 1.4 – 3.0 | 25 |
| I/S ML | 20.7 | 6.5 – 42.4 | 2.3 – 3.6 | 10 |

approaches. The use of field samples may introduce several errors which could reduce the accuracy of the approach. The uncertainties are related to additional traces of other minerals and organic matter, significant amounts of minerals without diagnostic features in the 2.100–2.400 µm region and measurement errors. Also, differences in the accuracy of the XRD analysis might influence the prediction accuracy. As can be seen from Table 5.2, the error for the determined content of smectite (1.4 – 3.0 wt. %) and I/S ML (2.3 – 2.6 wt. %) was higher compared to the errors for calcite (0.3 – 1.6 wt. %), dioctahedral mica (0.4 – 2.1 wt. %) and kaolinite (1.1 – 2.7 wt. %). Note that the regression tree was trained only on the abundances of the minerals of interest. These minerals occur within an unknown matrix, which might interfere with the measurements and the subsequent analysis. However, owing to the sampling design we assume that the samples also represent the non-analysed minerals associated with the composition of interest. Although unknown constituents may affect the spectral behaviour this does not necessarily invalidate the approach since regression tree analysis can deal with non-linear behaviour and interactions. Training of the regression trees was done by using the XRD-determined abundance of each mineral as the weight percentage of the mineral with respect to the total

weight of the sample. Therefore, the regression tree predictions are also expressed as absolute weight percentages.

## 5.3 Results

### 5.3.1 Curve fitting

Table 5.3 shows the parameters of the EGO profiles fitted for the pure minerals; calcite was fitted with three EGO profiles positioned around the typical absorption components of calcite. The spectrum of calcite was fitted with one feature ranging from 2.120 to 2.400 µm. Dioctahedral mica required two EGO profiles positioned around the two strongest typical absorption components. The feature was split into two absorption features to obtain a better fit. Smectite was fitted with two EGO profiles positioned around the typical absorption features of smectite by one feature within the 2.100–2.280 µm wavelength range. Kaolinite was fitted with four EGO profiles and was split into the wavelength ranges of 2.120–2.240 µm and 2.270–2.450 µm. The typical absorption between 2.120 and 2.240 µm was more difficult to fit by an EGO profile, this feature had the lowest fit of all the modelled features. Overall, the low values for the RMSE of the estimated log reflectance per feature show that the EGO profiles fitted on the typical absorption components of the mineral result in an accurate estimate of the measured log reflectance of the absorption features.

Table 5.3: Model parameter values of the fitted EGO profiles to the pure minerals (a-d). In addition, the RMSE of the fitted log reflectance compared to the measured log reflectance is given for each mineral. The EGO profile numbers correspond to the absorption components as identified in section 2.3.1. Each absorption component is described by a position (µm), width as the full width at half maximum (µm), intensity (µm), saturation (-) and asymmetry (-).The values listed for each continuum are the offset and slope of a straight line in wavelength and natural log reflectance.

Table 5.3a: Calcite.

| EGO Profile | Position | Intensity | Width | Saturation | Asymmetry |
|---|---|---|---|---|---|
| 1 | 2.158 | 0.012 | 0.010 | -2.890 | 0.098 |
| 2 | - | - | - | - | - |
| 3 | 2.288 | 0.082 | 0.028 | $-7.95*10^{-5}$ | -0.356 |
| 4 | - | - | - | - | - |
| 5 | 2.338 | -0.173 | 0.021 | $1.10*10^{-4}$ | -0.1000 |
| 6 | - | - | - | - | - |
| | | Feature1 | | Feature 2 | |
| Continuum | | | | | |
| | Offset | -5.165 | | - | |
| | Slope | 1.274 | | - | |
| RMSE | | 0.0008 | | - | |

Table 5.3b: Dioctahedral mica.

| EGO Profile | Position | Intensity | Width | Saturation | Asymmetry |
|---|---|---|---|---|---|
| 1 | - | - | - | - | - |
| 2 | 2.202 | 0.261 | 0.024 | $-3.68*10^{-5}$ | 0.011 |
| 3 | - | - | - | - | - |
| 4 | - | - | - | - | - |
| 5 | 2.341 | 0.050 | 0.014 | -2.501 | -0.111 |
| 6 | - | - | - | - | - |
| | | Feature1 | | Feature 2 | |
| Continuum | | | | | |
| | Offset | -1.868 | | -4.723 | |
| | Slope | 3.456 | | -9.971 | |
| RMSE | | 0.0035 | | 0.0011 | |

Table 5.3c: Kaolinite.

| EGO Profile | Position | Intensity | Width | Saturation | Asymmetry |
|---|---|---|---|---|---|
| 1 | 2.164 | - 0.360 | 0.025 | $-2.87 * 10^{-4}$ | -0.326 |
| 2 | 2.206 | 0.0446 | 0.011 | $-1.48 *10^{-5}$ | -0.147 |
| 3 | - | - | - | - | - |
| 4 | - | - | - | - | - |
| 5 | 2.357 | 0.035 | $6.20*10^{-3}$ | 2.097 | 0.055 |
| 6 | 2.384 | 0.076 | $9.82*10^{-3}$ | 0.355 | -0.012 |
| | | Feature1 | | Feature 2 | |
| Continuum | | | | | |
| | Offset | -8.329 | | -3.27 | |
| | Slope | 17.590 | | 6.27 | |
| RMSE | | 0.0166 | | 0.0014 | |

Table 5.3d: Smectite.

| EGO Profile | Position | Intensity | Width | Saturation | Asymmetry |
|---|---|---|---|---|---|
| 1 | - | - | - | - | - |
| 2 | 2.205 | 0.196 | 0.018 | 0.920 | -0.073 |
| 3 | 2.239 | 0.062 | 0.010 | 0.883 | 0.101 |
| 4 | - | - | - | - | - |
| 5 | - | - | - | - | - |
| 6 | - | - | - | - | - |
| | | Feature1 | | Feature 2 | |
| Continuum | | | | | |
| | Offset | -1.563 | | - | |
| | Slope | 3.091 | | - | |
| RMSE | | 0.0016 | | - | |

Table 5.4 provides the results of the deconvolution of the spectra from the 35 samples within the laboratory experiment. The total number of fitted profiles is given and the minimum and maximum parameter values as well. In addition, the average goodness-of-fit of all samples is provided by the average RMSE of the estimated log reflectance per feature for all 35 samples along with the standard deviation of the RMSE. The low average RMSE and standard deviation of the estimated log reflectance shows that the curve fitting approach approximates the measurements well. Splitting the reflectance into two spectral regions was necessary for most samples. The samples containing high concentrations of calcite were modelled by one feature. Samples containing relatively high concentrations of all four minerals required more EGO profiles compared to samples with specific dominant minerals to obtain the same accuracy.

The position of the EGO profiles corresponded to those of the diagnostic absorption features of the individual minerals, with minor shifts to smaller and longer wavelengths. In conclusion, the six expected absorption components of the minerals within the spectra were chosen to reflect their dimensionality. The minimum and maximum values of the saturation and asymmetry deviated strongly compared to the values of the EGO profiles fitted to the pure minerals, especially for the profiles 1, 3, 5 and 6. This indicates that the central limit theorem does not fully apply to the spectra of mineral mixtures. To accurately fit the absorption around the defined components, the EGO profile is required to be modelled with use of the parameters for asymmetry and saturation.

Table 5.4: Model parameters of the EGO profiles fitted to the 35 prepared samples. For each EGO profile the total number of fitted profiles and the minimum and maximum parameters values are given. In addition, the average RMSE and standard deviation of the fitted log reflectance for the full sample is given.

| EGO Profile | No. of samples | Position | | Intensity | | Width | | Saturation | | Asymmetry | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Min | Max | Min | Max | Min | Max | Min | Max | Min | Max |
| 1 | 24 | 2.15 | 2.193 | -0.360 | 0.13 | 0.01 | 0.02 | -4.842 | 38.88 | -0.38 | 0.098 |
| 2 | 32 | 2.20 | 2.208 | 0.005 | 0.44 | 0.01 | 0.02 | -0.501 | 0.920 | -0.26 | 0.221 |
| 3 | 24 | 2.21 | 2.253 | -0.069 | 0.10 | -0.01 | 0.08 | -3.343 | 14.57 | -0.23 | 0.199 |
| 4 | 10 | 2.28 | 2.309 | -0.057 | 0.08 | 0.01 | 0.03 | -3.987 | 3.705 | -0.43 | 0.027 |
| 5 | 27 | 2.33 | 2.357 | -0.173 | 0.07 | 0.01 | 0.02 | -44.87 | 13.61 | -0.35 | 0.055 |
| 6 | 14 | 2.36 | 2.385 | 0.006 | 0.07 | 0.01 | 0.02 | -30.50 | 6.107 | -0.45 | 0.093 |

| | Feature1 | | Feature 2 | |
|---|---|---|---|---|
| | Average | Standard deviation | Average | Standard deviation |
| RMSE | 0.00189 | 0.0028 | 0.00058 | 0.00030 |

Each absorption component is described an EGO profile which has a position (μm), width as the full width at half maximum (μm), intensity (μm), saturation (-) and asymmetry (-). The RMSE provides the measure for goodness-of-fit of the estimated log reflectance compared to the measured log reflectance of the prepared samples.

Table 5.5 provides the results of the deconvolution of the spectra from the 77 samples within the field experiment. In addition, the average goodness-of-fit of all samples is provided by the average RMSE of the estimated log reflectance per feature for all 77 samples along with the standard deviation of the RMSE. The average RMSE and its standard deviation of the estimated log reflectance show that the curve fitting resulted in small differences between the observed and predicted log reflectance. Of most interest are the differences between the obtained parameter values compared to those of the laboratory experiment. The field samples required fewer EGO profiles than the laboratory experiment which can be attributed to the fact that most of these samples consisted of at most two dominant minerals. The intensities were lower and the saturation and asymmetry effects strongly increased. The shifts in position towards higher and lower wavelengths were also stronger but they were still well modelled by the six absorption components of the minerals. Fewer EGOs were fitted around the smaller absorption components (1, 4 and 6) which could indicate that the weaker absorption components of the minerals, around 2.170 µm, 2.310 µm and 2.380 µm, became subsidiary absorptions. Overall, the appearances of individual absorption components were less intense in the field samples due to the noise introduced relating to the additional traces of other minerals, significant concentrations of minerals without diagnostic features in the 2.100–2.400 µm region (Table 5.2).

Table 5.5: Model parameters of the EGO profiles fitted to the 77 field samples. For each EGO profile the total number of fitted profiles and the minimum and maximum parameters values are given. In addition, the average RMSE and standard deviation of the fitted log reflectance for the full sample is given.

| EGO Profile | No. fitted | Position | | Intensity | | Width | | Saturation | | Asymmetry | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Min | Max | Min | Max | Min | Max | Min | Max | Min | Max |
| 1 | 37 | 2.171 | 2.198 | 0.008 | 0.08 | 0.01 | 0.03 | -6.82 | 8.253 | -0.17 | 0.11 |
| 2 | 75 | 2.202 | 2.216 | -0.077 | 0.17 | 0.01 | 0.03 | -9.11 | 1.136 | -0.22 | 0.27 |
| 3 | 72 | 2.215 | 2.333 | -0.026 | 0.06 | 0.02 | 0.02 | -4.95 | 48.96 | -0.16 | 0.30 |
| 4 | 13 | 2.281 | 2.309 | -0.017 | 0.04 | -0.01 | 0.04 | -11.81 | 14.15 | -0.45 | 0.29 |
| 5 | 65 | 2.312 | 2.369 | -0.014 | 0.06 | 0.006 | 0.19 | -21.98 | 56.68 | -1.02 | 1.69 |
| 6 | 9 | 2.376 | 2.396 | -0.013 | 0.02 | 0.005 | 0.51 | -6.863 | 11.14 | -0.31 | 3.62 |

| | Feature1 | | Feature 2 | |
|---|---|---|---|---|
| | Average | Standard deviation | Average | Standard deviation |
| RMSE | 0.0020 | 0.0006 | 0.00056 | 0.00034 |

Each absorption component is described by a position (µm), a full width at half maximum (µm), intensity (µm), saturation (-) and asymmetry (-). The RMSE provides the measure of goodness-of-fit of the predicted log reflectance compared to the observed log reflectance of the field samples.

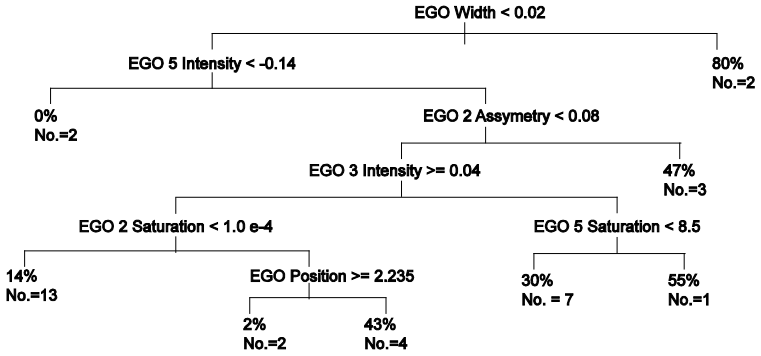


Figure 5.5: Regression tree for dioctahedral mica, the thresholds at a specific split with the assigned EGO parameter, the mineral abundance (wt.%) and the number of samples for each terminal node are given.

## 5.3.2 Prediction of mineralogy by regression trees

*5.3.2.1 Laboratory experiment*

The mineral abundances of the prepared samples were predicted by pruned regression trees which included a maximum of seven EGO variables (Table 5.6). The splits of the regression trees were mainly based on the position, intensity and width of the EGO profiles. The predicted samples fell into the numerical ranges defined by the terminal nodes of the regression tree. For example, figure 5.5 shows the pruned regression tree for dioctahedral mica where the variables with the threshold to split the dataset are given at the nodes and the total number of samples assigned to each terminal node.

Table 5.6: Variables used at splits in the pruned regression trees for the laboratory experiment.

| | EGO profile parameters | | | | |
|---|---|---|---|---|---|
| Regression tree | Position | Intensity | Width | Asymmetry | Saturation |
| Kaolinite | | EGO 2* | EGO 1 | | |
| | | EGO 3 | EGO 2 | | |
| | | | EGO 3 | | |
| Dioctahedral mica | EGO 3 | EGO 3 | EGO 2 | EGO 2 | EGO 2 |
| | | EGO 5 | | | EGO 5 |
| Smectite | EGO 3 | EGO 1 | EGO 2 | EGO 2 | EGO 2 |
| Calcite | EGO 1 | EGO 2 | EGO 2 | EGO 1 | |
| | | | EGO 5 | | |

* EGO 1: 2.170 µm, EGO 2: 2.210 µm, EGO 3: 2.250 µm and EGO 5: 2.350 µm

Cross-validation of the pruned regression trees shows that (Fig. 5.6 a-d); kaolinite resulted in six terminal nodes and abundance was predicted with an RMSE of 4 wt. % and $R^2$ of 0.92. Dioctahedral mica resulted in eight terminal nodes and was predicted by the regression tree with an RMSE of 9.3 wt. % and $R^2$ of 0.82. Calcite resulted in seven terminal nodes and was predicted by the regression tree with an RMSE of 7.3 wt. % and $R^2$ of 0.88. Smectite resulted in seven terminal nodes and was predicted by the regression tree with an RMSE of 8 wt. % and $R^2$ of 0.85. Calculation of the covariance of the predicted minus the measured mineral abundance showed that neither the over- or under estimation of a mineral were correlated to the other minerals.

For dioctahedral mica the major splits in the regression tree were based on the position and intensity of the absorption component around 2.250 µm, the width, asymmetry and saturation around 2.210 µm. Smaller splits used the intensity and saturation around 2.350 µm (Table 5.6). The major splits for the pruned regression tree for kaolinite were based on the intensity and width of the EGO profiles around 2.210 µm and 2.250 µm and smaller splits used the width around 2.170 µm (Table 5.6). Although the small subsidiary absorptions at wavelengths > 2.250 µm could be fitted, they were not significantly present within the mixture to be used as a prediction parameter.

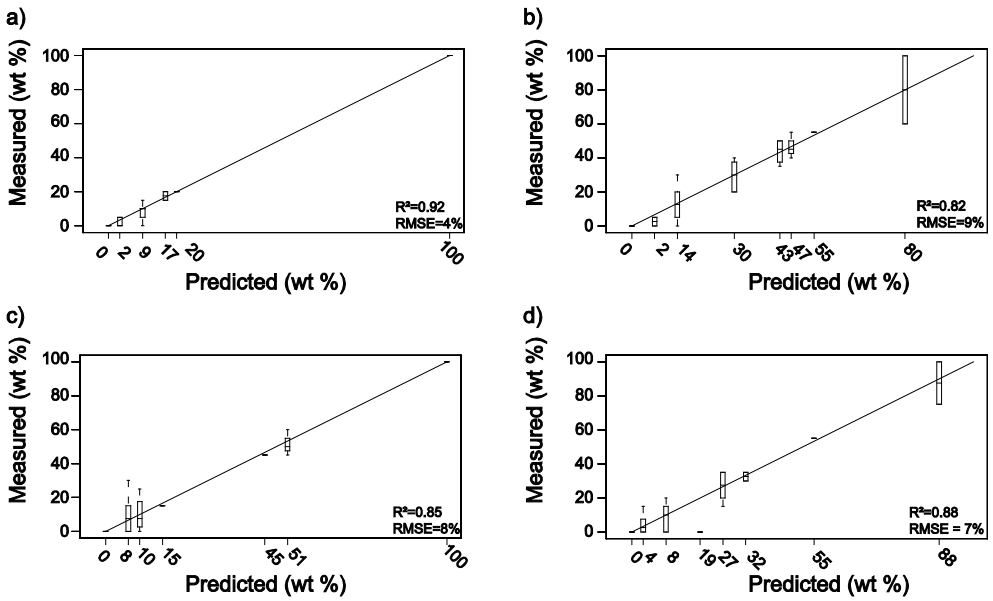


Figure 5.6 a-d: Predicted relative mineral content (wt.%) from the laboratory experiment compared to known mineral abundance of a) kaolinite, b) mica, c) smectite and d) calcite. The predicted mineralogy is presented by boxplots of the samples which were assigned to the terminal nodes of the regression tree.

The EGO profiles corresponding to the typical absorption components of smectite around 2.210 µm and 2.250 µm were used for partitioning of the dataset; the asymmetry and saturation of the EGO profile around 2.210 µm and the position around 2.250 µm were used for the major splits. The typical v-shape of the smectite absorption becomes less distinct and intense within mixtures due to saturation. With increasing saturation of the absorption component around 2.250 µm, the width of the EGO profile around 2.210 µm and the asymmetry of the EGO profile around 2.250 µm increases. The prediction of calcite was based on the width of the EGO profile around 2.350 µm, and the position around 2.170 µm and the parameters related to the EGO profile around 2.210 µm. Taking into account the strong absorption components of calcite around 2.350 µm (Fig. 5.2), it was to be expected that changes related to this feature would be important for defining the splits of the regression tree.

In table 5.7, the differences between the predicted mineralogy from the samples with added chlorite are given. Small errors occur up to 3 wt. %, with the exception of smectite in sample 2 and dioctahedral mica in sample 14, and the latter deviating just 6 wt. % difference with the measured content. Mineral concentrations in most of these "contaminated" samples were predicted correctly. This indicates that the pruned regression trees trained on the specific mineral absorption components were rather insensitive to contamination by chlorite.

Table 5.7: Difference in the predicted mineralogy of the samples containing additional chlorite.

| Sample No. | Mineral | Content (%) | Predicted content (%) | | Difference between predicted contents |
|---|---|---|---|---|---|
| | | | Original sample | Sample with chlorite | |
| 1 | Kaolinite | 5 | 8.9 | 8.9 | 0.0 |
| | Dioctahedral mica | 40 | 29.3 | 29.3 | 0.0 |
| | Smectite | 10 | 8.5 | 8.5 | 0.0 |
| | Calcite | 20 | 27.5 | 27.5 | 0.0 |
| 2 | Kaolinite | 15 | 8.9 | 8.9 | 0.0 |
| | Dioctahedral mica | 15 | 13.9 | 13.9 | 0.0 |
| | Smectite | 45 | 51.3 | 8.5 | -42.7 |
| | Calcite | 0 | 4.4 | 0.0 | -4.4 |
| 4 | Kaolinite | 10 | 8.9 | 8.9 | 0.0 |
| | Dioctahedral mica | 10 | 13.9 | 13.9 | 0.0 |
| | Smectite | 20 | 10.0 | 8.5 | -1.5 |
| | Calcite | 35 | 27.5 | 27.5 | 0.0 |
| 8 | Kaolinite | 20 | 17.5 | 17.5 | 0.0 |
| | Dioctahedral mica | 5 | 13.9 | 13.9 | 0.0 |
| | Smectite | 45 | 45.0 | 45.0 | 0.0 |
| | Calcite | 5 | 4.4 | 8.3 | 3.9 |
| 14 | Kaolinite | 0 | 2.1 | 2.1 | 0.0 |
| | Dioctahedral mica | 20 | 29.3 | 13.9 | -15.4 |
| | Smectite | 0 | 8.5 | 8.5 | 0.0 |
| | Calcite | 55 | 55.0 | 55.0 | 0.0 |

Finally, the absorption components around 2.210 µm and 2.350 µm were most diagnostic; modelled changes of these components and their relation to the other components were key to predicting mineralogy in composite mixtures. These results and the constructed regression trees show that indeed there is a statistical relationship between the model parameters and their representativeness of the complex interaction between diagnostic absorptions and mineralogy. Although intensity and widening were the most important parameters, some regression trees required the parameters related to saturation or asymmetry which stresses the added value of including these parameters for the deconvolution of the spectra. Note that our sample was small, which resulted in simple trees with few terminal nodes; a larger sample of mineral mixtures spanning the volume space may produce more complex trees with finer resolution in the terminal nodes.

*5.3.2.2 Field experiment*

The relative mineral abundance in the field samples was predicted by training and pruning the regression trees by the same EGO profiles used for the laboratory experiment. New regression trees were trained because unknown constituents may have affected the spectral behaviour and thereby changing the relation of the EGO parameters with mineral abundances. The regression trees included a maximum of five variables for their splits (Table 5.8). The coefficient of determination ($R^2$) and RMSE of predicted mineral content (wt. %) obtained from cross validation of the pruned regression trees were generally lower compared to the controlled experiments. This indicates that the fit of the trees were less accurate and the estimated error in the mineral abundance was consequently lower. Because more samples were available for training of the tree, more terminal nodes could be made and thereby the error estimate of the mineral abundance reduced. The splits of the regression trees were mainly based on intensity and asymmetry rather than the position, intensity and width of the trees in the laboratory experiment; Kaolinite resulted in seven terminal nodes and abundance was predicted with an RMSE of 3 wt.% and $R^2$ of 0.70 (Fig. 5.7a). The deviations occur in samples where absence was predicted while the actual content of the samples where within a range of 0-8 wt. % kaolinite. Dioctahedral mica resulted in eight terminal nodes and was predicted by the regression tree with an RMSE of 5 wt. % and $R^2$ of 0.63 (Fig. 5.7b). Calcite resulted in seven terminal nodes and was predicted by the regression tree with an RMSE of 8 wt. % and $R^2$ of 0.80 (Fig. 5.7d). The nodes including the absence and traces of calcite contained the samples with the greatest deviation resulting in a general underestimation of calcite. Unfortunately, with respect to smectite it was not possible to build a regression tree with a better fit and accuracy ($R^2$=0.40, RMSE=12 wt. %, Fig. 5.7c) without violating the criteria set for pruning the regression tree. Calculation of the covariance of the predicted minus the measured mineral abundance

showed that neither the over- or under estimation of a mineral were correlated to the other minerals.

The independent validation on 16 samples resulted in an RMSE of 5 wt. % for kaolinite, 9 wt. % for dioctahedral mica, wt. 18% for smectite and 14 wt. % for calcite. The overall lower RMSE (wt. %) within the field experiment can be attributed to specific mineral composition of the field samples compared to the prepared samples. The relative abundance of minerals was usually dominated by two minerals with only smaller contributions of other minerals. Also, more samples were available for training the regression tree. From figure 5.7 (a-d) it can also be observed that the model has difficulties with the prediction of the absence of minerals; the main reason can be attributed to the precision of the tree, due to the small sample (smectite, the smallest set consisted of 25 samples) few terminal nodes could be set. For both kaolinite and dioctahedral mica the model estimates a presence up to 5 wt. % in samples where these minerals are absent. For calcite and smectite the abundance estimate for absent minerals is up to a maximum of 10 wt. %. These lower thresholds for kaolinite and dioctahedral mica can be attributed to lower abundance in the samples used for training the trees.

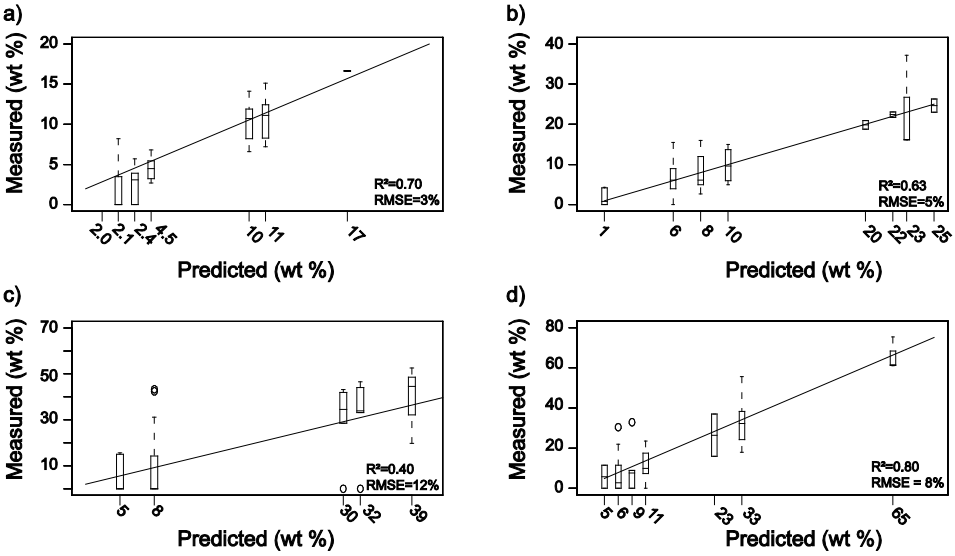


Figure 5.7a-d: Predicted relative mineral content (wt.%) from the field experiment compared to the values obtained from the XRD analysis of a) kaolinite, b) mica, c) smectite and d) calcite. The predicted mineralogy is presented by boxplots of the samples which were assigned to the terminal nodes of the regression tree.

Table 5.8: Variables used at splits in the pruned regression trees for the field experiment.

| Regression tree | EGO parameter values | | | | |
|---|---|---|---|---|---|
| | Position | Intensity | Width | Asymmetry | Saturation |
| Kaolinite | EGO 3* | | EGO 2 | EGO 2<br>EGO 3 | |
| Dioctahedral mica | | EGO 2<br>EGO 5 | | EGO 2<br>EGO 3 | EGO 2 |
| Smectite | | EGO 2<br>EGO 3 | | EGO 2<br>EGO 3 | |
| Calcite | EGO 5 | EGO 5 | | EGO 1<br>EGO 2 | |

* EGO 1: 2.170 µm, EGO 2: 2.210 µm, EGO 3: 2.250 µm and EGO 5: 2.350 µm

## 5.4 Discussion

This work demonstrated that the combination of spectral deconvolution of SWIR spectra with regression tree analysis allows simultaneous quantification of more than two minerals within a mixture. The fact that more than two minerals can be simultaneously quantified is an improvement on similar MGM-based methods. Key to this improvement was the regression tree analysis that followed the EGO analysis.

It was found that the degree of expression of absorption components was different between the field samples and the laboratory mixtures. Due to the nature of the field samples, the simple representation of the complex scattering behaviour by a few Gaussian bands required asymmetry and saturation to accurately deconvolve the spectra. Also, asymmetry of the EGO profiles proved to be an important parameter for the estimation of mineral content with field samples. For operational use, these results emphasize the importance of using (1) field samples for training of the model rather than laboratory mixes and (2) deconvolution using the EGO algorithm. For terrestrial studies it is therefore recommended that a representative surface sample is collected using a strategic sampling design, such as the one used by Mulder et al. (2012a).

The regression tree analysis was an improvement on the band ratios (Kanner et al., 2007; Noble et al., 2006) and the multiple linear regression (MLR) (Bishop et al., 2011) that were used to relate mineral abundances to the parameterized reflectance of mineral mixtures. Regression trees can deal with nonlinearity and interactions between the EGO parameters which made it possible to accurate predict mineral abundances from complex mixtures. The regression tree analysis used in this work uses a heuristic approach of making local optimal choices and finding a global optimum (cf., greedy algorithm), whereby construction of the tree is based on a local optimization, that is, at each node the data is partitioned – giving the best result for that specific node. This local optimization can result in an initial split based on some criteria which, at subsequent levels, results in suboptimal splits. Improvements are

expected by using less greedy algorithms such as boosted regression trees or random forest (Brown et al., 2006).

Alternatively, methods based on the single scattering albedo model of Hapke (2002), such as the work of Warell and Davidsson (2010) and Mustard and Pieters (1987), have proven successful in simultaneous retrieval of mineral abundances from prepared mixtures. Despite that, the accuracy of estimated mineral abundances from natural samples proved to be rather low. Our method is a strong competitor of the Hapke-based models. Especially, considering the required detailed information on the scattering properties of all endmembers in the model of Hapke (Keshava and Mustard, 2002) and the lower accuracy for estimates from natural samples (Warell and Davidsson, 2010). Nevertheless, there are limitations to our approach and possible solutions, as outlined below.

### 5.4.1 Limitations and outlook

Retrieval of smectite abundances

A likely explanation for the unsuccessful prediction of smectite is that smectite is a group consisting of monoclinic clay minerals such as e.g. montmorillonite. Within this group the clays have different ions within the octahedral layer; particularly $Fe^{3+}$ or $Al^{3+}$, but also $Mg^{2+}$, occur depending on the dioctahedral or trioctahedral character of the octahedral layer (Moore and Reynolds Jr., 1997). $Mg^{2+}$ is not commonly present in a dioctahedral layer but can be present due to charge compensation in the tetrahedral layers. Within the studied wavelength range strong absorption features of minerals are related to the overtones and combinations of the fundamentals of either the OH bend of AlOH, MgOH or $Fe^{3+}$OH. Also, within the laboratory experiment Na-rich montmorillonite was used, while the field samples originate from a calcite-rich environment resulting in Ca-rich montmorillonite. As discussed by Clark et al. (1990), variations in the Na/Ca ratio cause shifts of the 2.200 µm feature and with increasing $Ca^{2+}$ concentrations, the 2.250 µm feature becomes less pronounced (see also figures 5.2 and 5.4). The prediction model for smectite was trained using both absorptions, which subsequently resulted in poor performance. As indicated by Bishop et al. (2011); if differences in the octahedral cations occur, samples should not be treated spectrally as one group. So, by discriminating among the dioctrahedral clay minerals by their variations in their cation concentrations and by training the regression trees individually, the prediction of smectite abundance could be improved.

Outlook

A contribution that could not yet be included are abundance estimates from other minerals having their absorption features outside the studied wavelength range. Common minerals found on Earth, besides quartz and feldspars, e.g. goethite, hematite or chlorite, have their electronic absorption bands at shorter wavelengths.

Similar, dominant soil minerals such quartz, alkali feldspars and plagioclases have fundamental absorptions in the MIR and occasionally have broad electronic absorption around 1.2 microns due to structural Fe. Within the domain of reflectance spectroscopy it would be recommended to extend the studied wavelength range to the full VNIR-SWIR range (0.4–3.0 µm) to allow analyses of more minerals, which is feasible since most spectrometers usually cover this wavelength range.

Another challenge for the future includes the integration of the suggested approach with a spatial component. Spatially explicit information provides important information on parent material and soil formation (Egli et al., 2008; Mavris et al., 2011). Currently, satellites do not provide the fine spectral resolution needed to accurately deconvolve the spectra into components that can be related to mineral abundances. As a result, recent research efforts especially focus on the exploration and identification of minerals rather than abundances using RS data (Lau et al., 2012; Sgavetti et al., 2009; van der Meer et al., 2012; Viscarra Rossel, 2011). Compared to our work these studies do have a spatial context and provide insight in the spatial distribution of mineral characterization within a study area. However, combining the presented work with additional geostatistical analysis could be key to e.g. the creation of mineral abundance maps of large areas.

## 5.5 Conclusions

The retrieval of mineralogy from SWIR reflectance presented and evaluated in this paper aims to estimate the modal abundance of multiple minerals by a spectral curve fitting approach. Subsequently, the model parameters were used to predict mineral abundance by regression tree analysis. By using EGO, the spectra between 2.100 µm and 2.400 µm were accurately fitted in both the laboratory and field experiment (average RMSE of 0.004). This required the isolation of the absorption features and modelling of the individual absorption components within these absorption complexes. The position of the absorption components compared to those of the known diagnostic absorption features of the individual minerals showed shifts towards shorter and longer wavelengths. The parameters addressing the saturation and asymmetry of the absorption components appeared to be essential to obtain an accurate fit for the field samples.

The regression trees calibrated for the laboratory experiment based the splits mainly on the position, intensity and width of the fitted absorption components. The calibrated trees in the field experiment resulted in splits based on intensity and asymmetry rather than position, intensity and width. The selected parameters reflected the major changes in absorption of specific minerals and to changes in relation to the other known minerals.

The cross-validation results showed that the regression tree models were able to predict the mineral abundances well by establishing the statistical relationship with

the EGO model parameters. Within the laboratory experiment, abundances of kaolinite, dioctahedral mica, calcite and smectite were predicted with respectable RMSE values of less than 9 wt.% and a minimum detection limit up to 10 wt.%. Prediction of mineralogy with field samples showed good results for calcite, dioctahedral mica and kaolinite, with RMSE values less than 8 wt. %, similar minimum detection limits but lower coefficients of determination. Prediction of smectite abundance was less successful due to the spectral variations related to differences in the octahedral cations in the smectites of the analysed powders. Substitution of part of the quartz by chlorite at the prediction phase hardly affected the accuracy of the predicted mineral contents; this suggests that the method is robust in handling the omission of minerals during the training phase. On the other hand, the degree of expression of absorption components was different between the field samples and the laboratory mixtures. This demonstrates that the method should be calibrated and trained on a training set representing the range of local mineral compositions. With this study we demonstrated that our method allows estimation of more than two minerals within a mixture and thereby enhances the perspectives of spectral analysis for mineral abundances.

**Acknowledgements**

We acknowledge financial support from the EU FP7 program under the e-SOTER project (contract 211578). This study benefitted greatly from the review by G.A. Swayze and other anonymous reviewers. We thank H. Bartholomeus for his useful comments and L. Pompilio for making available the EGO algorithm and her insightful comments. R. Kokaly was supported by the Climate and Land Use Change Research & Development Program of the U.S. Geological Survey. We applied the SDC approach for the sequence of authors (Tscharntke et al., 2007).

Disclaimer: "Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government."

# Chapter 6

# Characterizing regional soil mineral composition using spectroscopy and geostatistics

*This chapter is based on:*
V.L. Mulder, S. de Bruin, J. Weyermann, R.F. Kokaly, M.E. Schaepman (Accepted for publication). Characterizing regional soil mineral composition using spectroscopy and geostatistics. *Remote Sensing of Environment.*

## Abstract

This work aims at mapping major mineral variability at regional scale based on scale-dependent spatial variability observed in remote sensing (RS) data. The objectives were to improve prediction models by using smoothed RS data that matched the scale of variability of the target sample and to evaluate the use of spectroscopy for soil mineral mapping. The analyses involved two datasets: (1) mineral abundances obtained from X-ray diffraction measurements, and (2) mineral categories derived from laboratory spectral analysis using the Material Identification and Characterization Algorithm (MICA). The RS data were smoothed to represent the medium and long-range spatial variability using Fixed Rank Kriging. The smoothed RS data resembled more closely the spatial variability of soil and environmental properties at regional scale. This was demonstrated by stronger correlations of mineralogy with the smoothed RS data than with the original RS data. Highest model accuracies resulted from the models using multi-scale soil-landscape relationships. Maps of predicted mineral categories and mineral abundances had similar patterns. The MICA analysis successfully predicted dominant mineral categories but was less able to model subtle differences. High correlation coefficients were obtained for the abundances of calcite ($R^2$=0.71, RMSE = 12 wt. %) and mica ($R^2$=0.70, RMSE= 6 wt. %) and moderate accuracies for smectite ($R^2$=0.57, RMSE = 8.3 wt. %) and kaolinite ($R^2$=0.45, RMSE = 3.8 wt. %). Moreover, soil mineralogy obtained from spectroscopy provided sufficiently detailed data to characterize major mineral variability (overall accuracy for mica, smectite and kaolinite were 76%, 89% and 86% respectively).

## 6.1 Introduction

As part of the ecosystem, soils are responding and contributing to global environmental changes including climate change. Soil resources play a fundamental role for assuring food security and are essential to be considered in climate change adaptation measures (Global Soil Parnership, 2011; Grunwald, 2009). Accordingly, representative and up-to-date soil information for regional to global scales is needed. This provides new challenges to digital soil mapping (DSM) in relation to scaling issues, prediction accuracies and computational requirements.

DSM has evolved to being a discipline where geo-statistics, terrain analysis and remote sensing (RS) are synergistically used to predict spatial soil variability. Over the past decade, many contributions have assessed DSM, and some extensive reviews on the topic have been written (Grunwald, 2009; McBratney et al., 2003; Scull et al., 2003). At smaller scales, DSM has been successful in delivering up-to-date soil information (Ben-Dor et al., 2002; Castrignanò et al., 2000; McBratney et al., 1991; Odeh et al., 1994). Still, difficulties exist for large-scale assessments with respect to data availability and scaling issues (Minasny, 2012).

In this work the following definitions of scale were used; local is smaller than $10^4$ $km^2$, regional varies between $10^4$ $km^2$ and $10^7$ $km^2$ and global is larger than $10^7$ $km^2$ (Mulder et al., 2011b). Larger spatial scales or studies refer to studies which are carried out at a scale beyond the local scale. Typically, large-scale assessments rely on relatively sparse samples which impede the development of soil prediction models; many predictor variables are obtained from satellite platforms and using such data requires comprehensive preprocessing. Accordingly, the sample size of target variables and the sampled variance of the soil and environmental observations influence the accuracy of soil prediction models (Vasques et al. 2011).

Within this framework, we aimed at mapping major soil variability at regional scale using predictor variables derived from RS data. The foci were (1) to improve prediction models using scaled variability in RS data that matched the scale of variability of the target sample and (2) to evaluate the use of spectroscopy for mineral mapping. We have a sparse sample that represents the major variability of soils at regional scale, collected previously (Mulder et al., 2012a). We state that for modelling soil properties using this sample, the predictor variables are required to represent the same scale. This was realized by considering medium and long range spatial variability in RS data. Smoothing RS data to the corresponding ranges allows expressing major variability rather than the noisy RS signals, including short range variability and measurement errors (Burrough, 1983). The spatial variability in the RS data can be assessed by determining the spatial structures in their variogram. The Soil-Landscape paradigm (Jenny 1941) is the basic principle underlying this study. Here, it is stated that soil properties depend on the environmental conditions and thus

soil properties can be predicted using the statistical relation between soil properties and environmental predictors. Therefore, we assume that the observed spatial variance at specific scales in RS data represents the spatial variance of soil properties at the same scale. Accordingly, prediction models (McBratney et al., 2006) using a representative sample and the smoothed RS data are deemed more representative for large-scale soil mapping. We chose soil mineralogy for the known presence of different spatial structures in large areas and the relative low variability over middle to long ranges (Jaquet, 1989).

Predicting soil properties using spectroscopy is applicable to larger-scale studies because it is a time and cost-efficient method to collect information (Viscarra Rossel et al., 2011). However, the derived information differs from that obtained from traditional chemical analysis. Typically, less refined outputs are generated from spectroscopy, especially for those materials that are difficult to estimate, e.g. mineralogy (Stenberg et al., 2010). In the presented work, we were interested in major mineral variability (i.e. areas where high concentrations of specific minerals occur) rather than quantifying soil properties at high resolution. Spectral analysis has been applied to provide such information, e.g. the work of Clark et al. (2003) and Swayze et al. (2003). Hence, for evaluating the use of spectroscopy for mineral mapping, we modelled surface mineralogy using both spectroscopy-derived and chemical-derived soil mineralogy.

Factorial kriging enables multi-scale analysis of small data sets (Goovearts, 1997). Unfortunately, the use of large datasets in conventional kriging approaches results in computationally forced limitations, mainly owing to the calculation of the inverse covariance matrix (Cressie and Johannesson, 2008). Local kriging neighbourhoods were used with moving windows for larger data sets (Harris et al., 2011). This appeared to be suboptimal because the long scale is not well represented, since the data is still too voluminous. To ensure computational feasibility, methods have been developed that approximate the covariance function by e.g. covariance tapering (Furrer et al., 2006) or approximate iterative methods (Nychka et al., 2002; Quiñonero-Candela and Rasmussen, 2005). Nevertheless, these methods still have difficulties calculating the sparse covariance matrix for massive datasets. We used the Fixed Rank Kriging (FRK) method, to overcome the difficulties with large datasets (Cressie and Johannesson, 2008). FRK introduces a dimension reduction for the calculation of the (inversed) covariance matrix. This reduction is based on the assumption that the process of interest can be expressed as a linear combination of spatial basis functions and a fine-scale-variation component. The size of the covariance matrix is thereby reduced to the number of spatial basis functions. The basis functions are typically chosen at various resolutions to capture different scales of spatial variation. FRK thus allows modelling the spatial variability at specific scales of interest. As such, FRK has not been employed yet in DSM assessments, however, it was deemed most suitable for large-scale DSM.

Here, we present a study aiming at characterizing regional soil mineral composition, based on scale-dependent spatial variability as observed in RS data, using a 15.000 km$^2$ area located in Northern Morocco. The analysis of scale-dependent variability, the subsequent smoothing of RS data with FRK and the development of the prediction models was done using ASTER satellite imagery and derived products. In the final step, the resulting mineral maps for both approaches were carefully evaluated and final recommendations were given concerning the use of remote and proximal sensing (RS and PS) data for large-scale DSM.

## 6.2 Data

### 6.2.1 Study area and site data

The 15.000 km$^2$ study area is located in Northern Morocco, centred at approximately 34.0ºN, 4.5ºW. The Rif Mountains, an area of highlands, forms the northern border of the study area. The Anti-Atlas Mountains are located on the southern border. In between, the study area consists of plateaus and intermountain valleys. This area offered a diverse lithological setting, including sedimentary, igneous and metamorphic rock types. We designed a sparse RS-based sampling approach, making use of conditioned Latin Hypercube Sampling (cLHS) to assess the major soil variability present in the study area. The sample consists of 73 sites, located within predefined subareas (Fig. 6.1), and proved to be representative for soil variability at regional scale. For more details on the sampling strategy and the mineral analysis we refer to Mulder et al. (2012a; 2013).

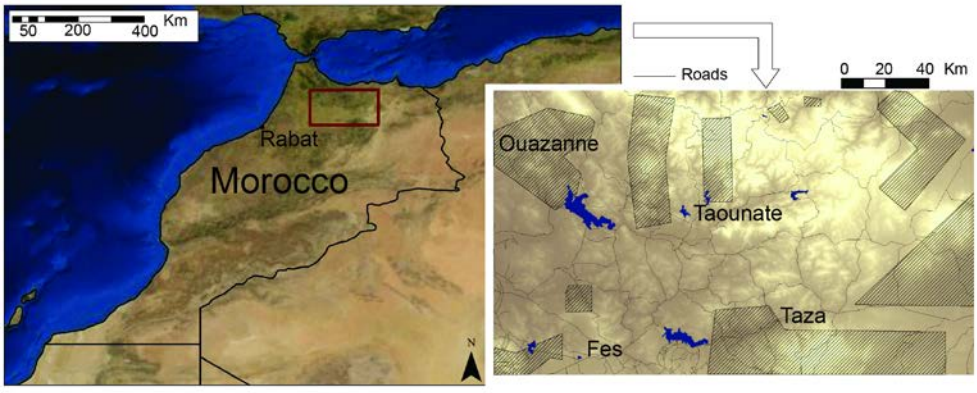


Figure 6.1: Study area with predefined subareas (Mulder et al., 2012a).

*6.2.1.3 Soil mineralogy from x-ray diffraction analysis*

The mineralogy of the samples was determined by X-ray diffraction (XRD) analysis (Mulder et al., 2013) (Table 6.1). Illite and muscovite could not be separated by the XRD analysis due to their similar structure, and in this work they were referred to as (dioctahedral) mica. Interstratified illite/smectite is abbreviated by I/S ML. The minerals used for this study were required to (1) have absorption features in the 2.100–2.400 µm region (Clark, 1999); this allows comparison at a later stage with the spectroscopy-derived maps, (2) be frequently present in the samples and (3) have an average content larger than 5 wt. %. Following these criteria, calcite, dioctahedral mica, smectite and kaolinite were used for further analysis (Mulder et al., 2013).

Table 6.1: Mineral abundances (wt. %) present in the samples as determined by XRD analysis (Mulder et al., 2013).

| Mineral | Average (wt. %) | Range (wt. %) | Error (3σ) (wt. %) | **No. samples** |
|---|---|---|---|---|
| Anatase | 1.0 | 0.4 - 1.9 | 0.2 - 0.3 | 41 |
| Aragonite | 7.1 | 1.9 - 11.1 | 0.7 – 0.8 | 3 |
| Calcite | 21.4 | 0.7 - 75.3 | 0.3 – 1.6 | 63 |
| Chlorite | 5.0 | 0.8 - 25.4 | 0.4 – 1.8 | 52 |
| Cristobalite | 2.8 | 1.2 - 4.5 | 0.3 – 0.5 | 3 |
| Epidote | 1.9 | 1.9 | 0.5 | 1 |
| Dolomite | 6.9 | 0.8 - 33.9 | 0.3 – 0.9 | 29 |
| Goethite | 5.2 | 1.2 - 11.6 | 0.8 – 1.3 | 14 |
| Gypsum | 1.5 | 1.5 | 0.5 | 1 |
| Hematite | 1.4 | 0.4 - 4.6 | 0.2 – 0.4 | 27 |
| Dioctahedral mica | 10.5 | 1.6 - 56.9 | 0.4 – 2.1 | 66 |
| Kaolinite | 6.7 | 1.3 - 16.7 | 1.1 – 2.7 | 49 |
| K-feldspar | 3.1 | 0.9 - 18.6 | 0.5 – 1.3 | 42 |
| Palygorskite | 6.8 | 6.8 | 0.8 | 1 |
| Plagioclase Ab | 4.9 | 1 - 23.8 | 0.4 – 1.2 | 69 |
| Quartz | 42.5 | 12.6 - 83.9 | 0.6 – 1.8 | 77 |
| Rutile | 1.1 | 0.6 - 1.8 | 0.2 – 0.4 | 10 |
| Smectite | 28.6 | 9.4 - 52.9 | 1.4 – 3.0 | 25 |
| I/S ML | 20.7 | 6.5 - 42.4 | 2.3 – 3.6 | 10 |

*6.2.1.4 Soil mineralogy from spectroscopy*

Studies employing spectral feature analysis and comparison methods have shown that many different materials can be identified and mapped in terrestrial imaging spectrometer data (Clark et al., 2003; Dalton et al., 2004; Kokaly et al., 2013. Van der Meer (2012) reviewed the use of multi- and hyperspectral for geologic RS. The reviewed studies were limited to characterizing key minerals at local scales. Most detailed information was obtained from Hyperion data supported by ground truth data or not. The methods included linear spectral unmixing (Chudnovsky 2009; 2011) or Mixture Tuned Matched Filtering (Kodikara et al., 2012). These methods appeared to be suboptimal to map or quantify mineral mixtures when overlapping absorption features exist. Successful examples include the Tetracorder (Clark et al. 2003) and the Material Identification and Characterization Algorithm (MICA) (Kokaly et al., 2011a). MICA is a module of the USGS Processing Routines in IDL for Spectroscopic Measurements (PRISM) software (Kokaly et al., 2011a) and has been especially helpful for the characterization of the dominant mineralogy over large areas (Kokaly, 2011b; Mulder et al., 2012b).

The central concept of the MICA classification routine is the detection of diagnostic absorption features. These features are unique to particular materials. The diagnostic features are isolated by continuum removal and matched to reference features from a spectral library. To discriminate between materials having similar diagnostic absorption features, constraints are included to rule out the co-occurrence of other materials (Clark et al., 2003; Mulder et al., 2012b). Spectral measurements were taken under laboratory conditions with an ASD Fieldspec Pro FR spectroradiometer covering the 350-2500 nm wavelength region. The samples were illuminated using a stabilized quartz-tungsten light source with a condenser and Koehler illuminator, securing isotropic illumination within the field-of-view. A measurement field-of-view of 2 cm was obtained by using a $5^0$ foreoptics while placing the sensor 23 cm above the sample. The 2 mm samples were placed in pure crystal sample holders to reduce scattering from the sample holder. Preprocessing of the spectral data and the determination of main mineralogy was done using the 'Processing Routines in IDL for Spectroscopic Measurements' (PRISM) software (Kokaly, 2011a). The diagnostic features within the Short-Wave Infrared (SWIR) were used to derive the soil mineral categories. Discriminating between a calcite-rich and -poor environment gave the most accurate output for six mineral categories (Table 6.2). This classification relied on 48 spectral endmembers from the USGS spectral library version 6 (Clark et al., 2007). The overall accuracy was 0.52, based on the confusion matrix between the MICA output categories and reclassified abundances derived from x-ray diffraction.

Table 6.2: Soil mineral categories derived with PRISM.

| Mineral category | Spectral endmembers | No. of samples MICA | No. of samples XRD |
|---|---|---|---|
| Smectite | Montmorillonite (SAz-1, STx-1)<br>Montmorillonite + Illite (CM37, CM42) | 6 | 15 |
| Smectite calcite-rich | Montmorillonite (50%) + Calcite (25%) +Dolomite (25%)<br>Montmorillonite (50%) + Calcite (50%)<br>Montmorillonite (67%) + Calcite (33%)<br>Montmorillonite (20%) + Calcite (80%) (AMX15, AMX19)<br>Montmorillonite (50%) + Dolomite (50%) (AMX10, AMX21) | 27 | 18 |
| Dioctahedral mica | Illite (CU00-5b)<br>Illite (GDS 4)<br>Illite IMt-1 (a,b)<br>Muscovite – medium low Al<br>Muscovite (CU93-1) low Al<br>Muscovite + Chlorite (CU91-253D)<br>Muscovite (GDS 113, GDS 116)<br>Chlorite + Muscovite (CU93-65A) | 40 | 32 |
| Dioctahedral mica calcite-rich | Muscovite (67%) + Calcite (33%) | 23 | 28 |
| Kaolinite | Kaolinite (CM3, CM7, CM9)<br>Kaolinite (KGa-2)<br>Kaolinite + Muscovite (intimate mix)<br>Kaolinite (50%) + Muscovite (50%)<br>Kaolinite (50%) + Smectite (50%) | 17 | 24 |
| Kaolinite calcite-rich | Kaolinite (30%) + Calcite (70%)<br>Kaolinite (20%) + Calcite (80%) | 38 | 22 |

### 6.2.2 Remote sensing data

The explanatory variables used for predicting soil mineralogy, based on research listed in Table 6.3, were:

- ASTER satellite imagery (Abrams and Hook, 2001);
- ASTER digital elevation model (Abrams and Hook, 2001);
- The first three principal components (Wold et al., 1987) of ASTER VNIR-SWIR bands;
- Normalized difference vegetation index (NDVI) (Tucker, 1979).

#### *6.2.2.1 ASTER satellite data*

The spectral data products used for this work were (1) the ASTER Visible and Near Infrared (VNIR), (2) crosstalk corrected SWIR surface reflectance and (3) surface emissivity derived from the thermal infrared (TIR) (Abrams and Hook, 2001). A total of nine images of 60 x 60 km were required. These images were recorded in June and September 2005 and early May 2007. The selection of images from similar seasonal conditions minimized the cloud and vegetation coverage but small phenological differences did happen to occur over the years. Additionally, the images do not completely cover the area (Fig. 6.2).

Preprocessing of satellite data

We applied a modified Sun-Canopy-Sensor (SCS+C) method, to reduce topographic effects in the data (Soenen et al., 2005). This method appeared to be more effective than a pure cosine correction, a reason for which is in the fraction of forested hills covering substantial areas in the nine ASTER tiles. Pixels with slopes steeper than 40º were overcorrected which resulted in overestimated reflectance values (Soenen et al., 2005) and therefore these pixels were masked out. The striping in the TIR bands was reduced by applying a low-pass filter (Cudahy, 2012). Next, due to differences in gains and offsets of the images, a stepwise approach was used for mosaicking. This resulted in a global response of reflectance values (Cudahy, 2012) for the area. Finally, the data were resampled to 90 meters to match the resolution of the TIR data. Pixels having substantial vegetation cover (NDVI > 0.35 (Serbin et al., 2009)), clouds, snow and water were masked out to retain a strong signal from the soils.

Table 6.3: Remote sensing data.

| Data | | Wavelength (µm) | Resolution (m) | Information |
|---|---|---|---|---|
| ASTER VNIR | Band 1 | 0.520 – 0.600 | 15 | Vegetation and iron bearing minerals (Cudahy et al,. 2012) |
| | Band 2 | 0.630 – 0.690 | 15 | |
| | Band 3N | 0.780 – 0.860 | 15 | |
| ASTER SWIR | Band 4 | 1.600 – 1.700 | 30 | Clay and carbonate (Hewson et al., 2005) |
| | Band 5 | 2.145 – 2.185 | 30 | |
| | Band 6 | 2.185 – 2.225 | 30 | |
| | Band 7 | 2.235 – 2.285 | 30 | |
| | Band 8 | 2.295 – 2.365 | 30 | |
| | Band 9 | 2.360 – 2.430 | 30 | |
| ASTER TIR | Band 10 | 8.125 – 8.475 | 90 | Clay, quartz and carbonate-rich minerals (Hewson et al., 2005) |
| | Band 11 | 8.475 – 8.825 | 90 | |
| | Band 12 | 8.925 – 9.275 | 90 | |
| | Band 13 | 10.25 – 10.95 | 90 | |
| | Band 14 | 10.95 – 11.65 | 90 | |
| ASTER GDEM | Band 3N,3B | - | 30 | Proxy for topography (Jenny, 1941) |
| NDVI | Band 2,3 | - | 15 | Crop growth characteristics and, indirectly, characteristic for specific site qualities (Sommer et al., 2003; Sun et al., 2008) |
| PC 1:3 | Band 1:9 | - | 30 | Major variability present in the spectral data (Wold et al., 1987) |

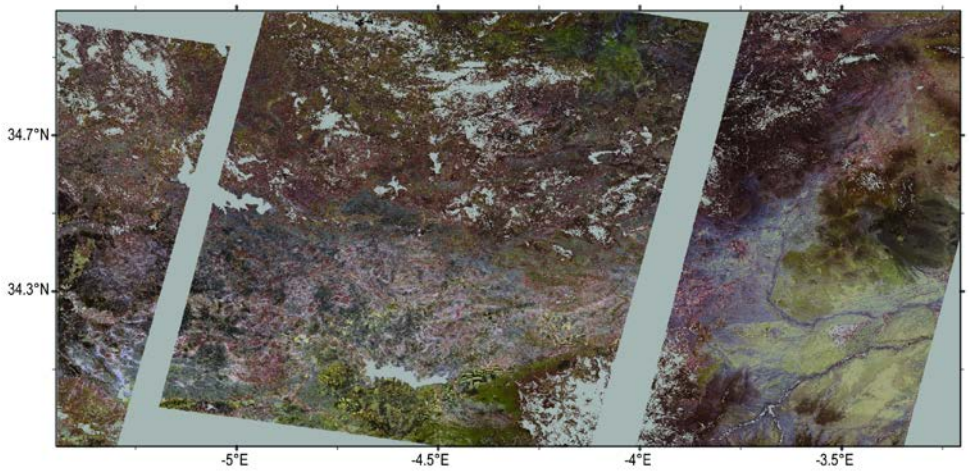


Figure 6.2: False colour composite of the topographically corrected, masked mosaic. Grey colours indicate no data.

## 6.3 Methods

Variogram analysis of the RS data was used for confirming the presence of medium and long-range structures of variability. Based on these findings, the RS data were smoothed to the corresponding ranges. Next, prediction models were made relating the mineral abundances and categories to the RS data. The considered scales for these models were (1) the pixel level, (2) the medium scale, (3) the long scale and (4) a combination of both medium and long scale. The most accurate prediction models for mineral abundances and categories were applied to the full extent of the study area and finally the maps were compared to each other.

### 6.3.1 Analysis of spatial structures within the remote sensing data

The scale-dependent spatial variability observed in RS data was determined by the range where changes in spatial structure occurred for each individual band or derived product. These structures were modelled by fitting variogram models to empirical variograms (Cressie, 1993; Pebesma, 2004). The Matérn model was applied because of its flexibility to model different variogram shapes (Matérn, 1960; Minasny and McBratney, 2005). Two structures were expected to suffice for an accurate fit (Goovearts, 1997). The accuracy of the fit was determined by calculating the RMSE of the fitted and actual semivariance of the lags of the variogram. The maximum lag-distance considered was 60 kilometres, corresponding to one third of the diagonal of the study area. More insight in the spatial processes was obtained by evaluating the relative contributions of the nugget and the partial sills to the total semivariance, the range parameter and kappa values. Ensuring that the spatial variability was captured at the scale of interest, the ranges of these structures were subsequently used as indicators for the resolution of the basis functions for FRK. Calculations were performed using the *R language and environment for statistical computing* (R Development Core Team, 2011) and contributed packages *gstat* and *sp* (Pebesma, 2004; Renard, 2011).

### 6.3.2 Data smoothing using Fixed Rank Kriging

Fixed Rank Kriging (FRK) aims to predict a spatial process by separating the signal from the noise, based on a large number of observations that contain measurement errors. To ensure computational feasibility even for massive datasets, FRK introduces a dimension reduction. This is done by expressing the process of interest as a linear combination of spatial basis functions and a fine-scale-variation component. The latter is characterized by a covariance matrix. Plugging in estimates of all parameters, we can obtain the Fixed Rank Kriging (FRK) predictor. In our case these are the smoothed RS signal and the accompanying mean squared prediction

error. This allows us to express major variability rather than the noisy RS signals, including short range variability and measurement error. FRK uses the signal that remains after removal of a deterministic trend. This signal is called the random spatial-variation component. Maximum-likelihood estimates of the covariance matrix of the vector of basis-function weights and the variance of the fine-scale-variation component can be obtained using the expectation-maximization (EM) algorithm, implemented in the Spatial Random Effect (SRE)-model (Katzfuss and Cressie, 2009). Hereby, the variance of the measurement error is assumed to be known and the spatial basis functions are representative for the scale of spatial variation. Recommendations for the spatial basis functions are that the center points (1) vary in spatial resolution in order to capture different scales of spatial variations and (2) regularly cover the spatial domain. This approach is based on Cressie and Johannesson (2008) and Katzfuss and Cressie (2011), among others.

Model settings SRE-model

Here, bisquare functions were used as basis functions, as suggested by previous work of Nychka et al. (2002). The center points of the functions were positioned on a regular grid at locations with data. The appropriate distance between the center points were obtained from variogram analysis, as was described in section 6.3.1. Two sets of spatial basis functions were selected to capture the spatial variation for both the long and medium range (Katzfuss and Cressie, 2011). The resolution and number of center points of the basis functions are given in Table 6.4. The observations used to estimate the spatial process were obtained by extracting them from the RS data, using a 7x7 window; this corresponded to one observation every 670 meter (Table 6.4). The number of observations slightly differed due to differences in spatial coverage resulting from various masks that were applied to the RS data. The variance of the measurement error and fine scale variation were estimated by fitting a variogram, using a subsample consisting of 15.000 observations (Katzfuss and Cressie, 2011). The overall mean was subtracted from the data, as is similarly done in simple kriging where the mean is assumed to be known and constant throughout the area; the residual component was modelled as a stationary random function with zero mean (Goovaerts, 1997).

Settings of the EM-algorithm

The maximum-likelihood estimates of the covariance matrix of the vector of basis-function weights and the variance of the fine-scale-variation component were obtained by iteratively solving the expectation-maximization (EM) algorithm. Convergence of the algorithm was monitored by evaluating the change in the

Table 6.4: SRE model settings.

| Model setting | VNIR-SWIR | TIR | PC | DEM | NDVI |
|---|---|---|---|---|---|
| Resolution of basis functions | | | No. of center points | | |
| 30 km (long range) | 23 | 23 | 26 | 26 | 26 |
| 20 km (long range) | 63 | 63 | 66 | 66 | 66 |
| 9 km (medium range) | 237 | 236 | 234 | 235 | 234 |
| 5 km (medium range) | 660 | 659 | 660 | 660 | 660 |
| No. of observations used for the approximation of the covariance matrix (7x7 window) | 46349 | 46237 | 51106 | 52376 | 52155 |
| No. of FRK prediction locations (450 m. resolution) | 115514 | 115514 | 115514 | 115514 | 115514 |

covariance matrix and fine-scale-variation components. All unique elements were stacked into a vector and the change was evaluated against the stopping criteria that determined the order of difference ($\varsigma$). E.g., $\varsigma = r^2 * 10^{-6}$ for the long range process, where $r^2$ is the order of magnitude of the parameters in the vector. The pre-specified values in $\varsigma$ are given in Table 6.5 and were deemed sufficiently low (Katzfuss, 2011). The choice of starting values for the parameter vector that maximizes the likelihood functions in the EM algorithm, were set to the default values to satisfy the requirements of a valid starting covariance matrix ($K^{[0]}$) and positive initial fine-scale-variation ($\sigma^{2}_{\xi}{}^{[0]}$) (Katzfuss and Cressie, 2011).

FRK predictions

Plugging in estimates of all parameters, we obtained the Fixed Rank Kriging (FRK) predictors at a spatial resolution of 450 m; this was deemed sufficiently detailed for mapping regional scale variability (Table 6.4). The FRK prediction errors were converted to relative values ((the prediction error / the range of the RS signal)*100) for comparison. Next, the effect on the relation between smoothed RS data and mineralogy was evaluated by assessing the correlation of mineralogy with the original, medium and long-scaled data. Preparation of the FRK input data and analysis of the FRK predictors were performed using the *R language and environment for statistical computing* (R Development Core Team, 2011) and contributed packages *raster* (Hijmans and van Etten, 2012) and *sp* (Pebesma, 2004). The FRK was performed in Matlab using the freely available FRK code provided by the SSES Program (http://www.stat.osu.edu/~sses/research.html).

Table 6.5: Settings of the EM algorithm.

| Model setting | Parameter | Value |
|---|---|---|
| Starting values to set the initial conditional expectation of the likelihood values (where *v*^2 is the variance of residuals and *I* the identity matrix) | Covariance matrix $K^{[0]}$ | 0.9 $v\text{^}2I$ |
| | Fine scale variation process ($\sigma^2_{\xi^{[0]}}$) | 0.1 $v\text{^}2$ |
| Pre-specified values for the stopping criteria of the iteration process | Short range process | $1*10^{-5}$ |
| | Long range process | $1*10^{-6}$ |

### 6.3.3 Digital soil mapping for mineral characterization

*6.3.3.1 DSM using soil mineralogy derived by x-ray diffraction*

Prediction of mineral abundances using the data from the x-ray analysis needs a modelled relation between the continuous target variables and continues predictor variables. Popular models include regression tree analysis (Breiman et al., 1984; Viscarra Rossel, 2011) and artificial neural networks (Malone et al., 2009). Linear regression models are most commonly used in DSM (Cambule et al., 2013; McBratney et al., 2003). We used multiple linear regression (MLR) for predicting mineral abundances based on (smoothed) RS data.

Variable selection in MLR was based on an exhaustive search for the best subsets explaining the most variation in the target variable while keeping the explanatory variables to a minimum (highest $R^2_{adj}$). Models were calibrated for the considered scales at (1) the pixel level, (2) the medium scale, (3) the long scale and (4) a combination of both medium and long scale. The models were validated using leave-one-out cross-validation and the most accurate models were applied to the full extent of the study area. The data-range of predictor variables might have been smaller at the calibration stage, potentially resulting in unrealistic estimates when applied to the full extent (outside the range of 0-100 wt. %). Abundances outside the [0, 100]-interval were set to "nodata". The calculations were performed using the *R language and environment for statistical computing* (R Development Core Team, 2011) and contributed packages *leaps* (Miller, 2002), *regr0* (Stahel, 2011) and *raster* (Hijmans and van Etten, 2012).

*6.3.3.2 DSM using soil mineralogy derived from spectroscopy*

Prediction of mineral categories requires a modelled relation between the categorical target variables and the continuous predictor variables (RS data). Such can be achieved by regression tree analysis (Breiman et al., 1984; Felicísimo et al., 2012), artificial neural networks (Behrens et al., 2005) and multinomial logistic regression (MNL) (Ardö et al., 1997; Hengl et al., 2007; Kempen et al., 2012). For this work we choose for MNL because it was deemed more suitable for our small dataset. The following mineral categories were predicted for each mineral: "absence", "presence" and "presence Ca-rich" (Table 6.2). "Presence Ca-rich" refers to the occurrence of a mineral in a calcite-rich environment.

Models were calibrated for the considered scales, being (1) the pixel level, (2) the medium scale, (3) the long scale and (4) a combination of both medium and long scale. The accuracy of the predicted mineral categories was assessed by calculating the error matrix of the observed versus the predicted mineral categories, using the training data. In addition, the change in the predicted probability of each category per sample was evaluated. These classification probabilities provide an internal measure of model performance (Stehman, 1997). The most accurate MNL models were applied to the full extent of the study area. The calculations were performed using the *R language and environment for statistical computing* (R Development Core Team, 2011) and contributed package *nnet* (Venables and Ripley, 2002) and *DAAG* (Maindonald and Braun, 2010).

*6.3.3.3 Evaluation of characterized regional soil mineral composition*

Comparison of abundance maps with the mineral sample

Independent validation of the maps was not possible because the complete sparse sample was needed to fit the MLR and MNL models. Alternatively, the distribution in mapped and modelled mineral abundances was compared to obtain insight in major differences. A random sample consisting of 10,000 points was taken from the mineral maps. The Wilcoxon test was used to test whether both samples had the same distribution and mean ($p$=0.01) (Wilcoxon, 1945) and if not, the shift in the median was defined. This shift provides insight in possible over- or underestimations in the modelled abundances. The histograms were given for visualisation of the mapped and measured mineral variability.

Comparison of mineral categories maps with the mineral sample

Comparing the distribution of mapped mineral categories with the sample required a reclassification of the measured abundances into the mineral categories (Table 6.2). Finding the threshold in abundance where minerals were successfully characterized by MICA is subjective and hampers a quantitative comparison. Initially, the threshold for a calcite-rich environment was set to the abundance of the first quartile of the distribution, being 16 wt. % calcite. The threshold for mica, kaolinite and smectite was set to an abundance larger than 0 wt. %. Next, the similarity between the abundances and the predicted categories ($p$=0.01) was determined using the Chi-squared test (Snedecor and Cochran, 1989). The initial thresholds may be unrealistic and hamper the comparison of the modelled the mineral abundance and -category maps (next section). As such, more realistic thresholds were defined by adjusting the thresholds to a level where the population percentages of reclassified abundances matched with the mineral categories. These adjusted thresholds were assumed to be the detection limit in mineral abundance for MICA.

Comparison of mineral abundances and categories maps

The abundance maps were first reclassified to the mineral categories using the thresholds resulting from the Chi-squared test. Next, to provide insight in the differences, the error matrix was calculated between both maps using all pixels (Stehman, 1997). Finally, these maps were compared, using the error matrix and observed spatial patterns.

All calculations were performed using the *R language and environment for statistical computing* (R Development Core Team, 2011).

## 6.4 Results

### 6.4.1 Analysis of spatial scales within the remote sensing data

The variogram analysis confirmed the presence of medium and long-range spatial processes within the RS data. The variograms from the VNIR, SWIR, TIR sensors and principal components resulted in similar variograms; therefore one band per sensor was presented (Fig. 6.3). The VNIR data had high nuggets (42% of the total semivariance), little variability was captured by the medium-range structure (14%) while the long-range structure captured 44% of the total semivariance. The nuggets of the SWIR data were 20% of the total semivariance, 22% and 58% was captured by the medium and long-range structure, respectively. The nuggets, medium and long-range structures of the TIR data were, 40%, 34%, and 26% respectively. The nuggets of the PC's was 26% of the total semivariance, most variability was accounted for by the medium-range structure (40%) while the long-range structure captured 33% of the total semivariance. The NDVI had a high nugget (38%), 25% and 37% of the semi-variance was captured by the medium and long-range structure, respectively. The long-range structure of the DEM was modelled by a power model with a linearity parameter of 1.01 indicating an almost linear structure. The semivariance in elevation (DEM) was 100% explained by the medium-range structure plus the nugget.

The medium-scale structures had low kappa-values (< 0.51) and range parameters of maximal 9.5 kilometres, indicating strong spatial variability over shorter distances (Minasny and McBratney, 2005). The long-scale structures had kappa-values between 1.9 and 6 and a range between 20 and 40 kilometre, indicating a smooth spatial process over long distances. The data smoothing with FRK need to capture the spatial variability corresponding to the range parameters of the medium- (9.5 km) and long-scale structures (up to 40 km). We decided using the basis function-parameterizations listed in Table 6.4 (resolution of basis functions).

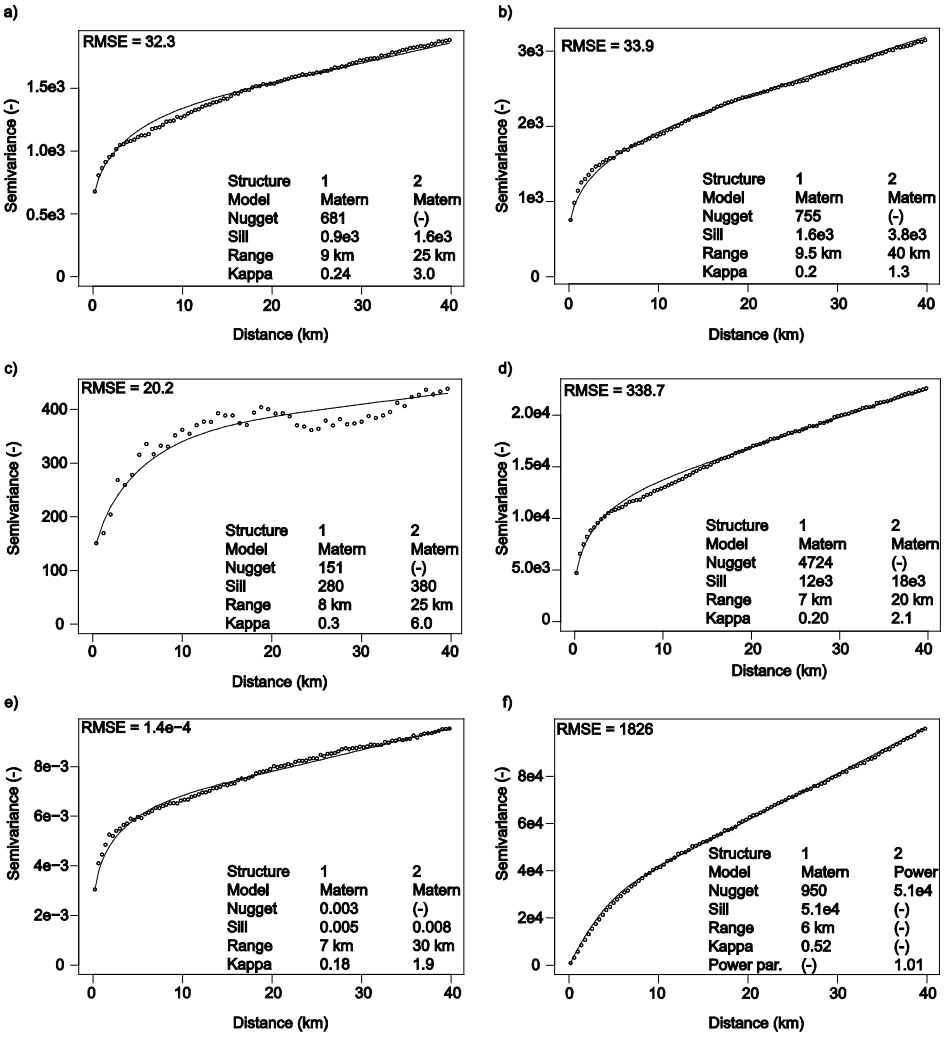


Figure 6.3: Variograms of the (a) VNIR band 1, (b) SWIR band 1, (c) TIR band 1, (d) PC 1, (e) NDVI and (f) DEM. The RMSE of the fitted model is given along with the model parameters of the medium (1) and long-range (2) spatial structures.

### 6.4.2 Data smoothing using Fixed Rank Kriging

Smoothing of RS data provided images with full coverage, representing either the medium or long- range variability within the RS data (Fig. 6.3 and 6.4). As expected, the prediction errors for the medium-range data (Fig. 6.4c and d) increased in the interpolated areas. The image representing the long-range variability (Fig. 6.4a) shows broad spatial patterns, however, also differentiated areas with higher prediction errors (Fig. 6.4b). This was expected because it is the difference between the local observations and long-scaled estimated RS value. The averaged FRK prediction errors at the medium scale did not deviate strongly, except for the TIR data (Table 6.6). Higher prediction errors were found for data having high nugget and semivariance at short distances, e.g. NDVI (Fig. 6.3e) against PC 1 (Fig. 6.3d). The opposite was found for the long-scale RS data where the prediction error decreased with increasing partial sill, e.g. SWIR (Fig. 6.3b) against DEM (Fig. 6.3f). Likely, the high errors within the TIR data originate from the striping effect.

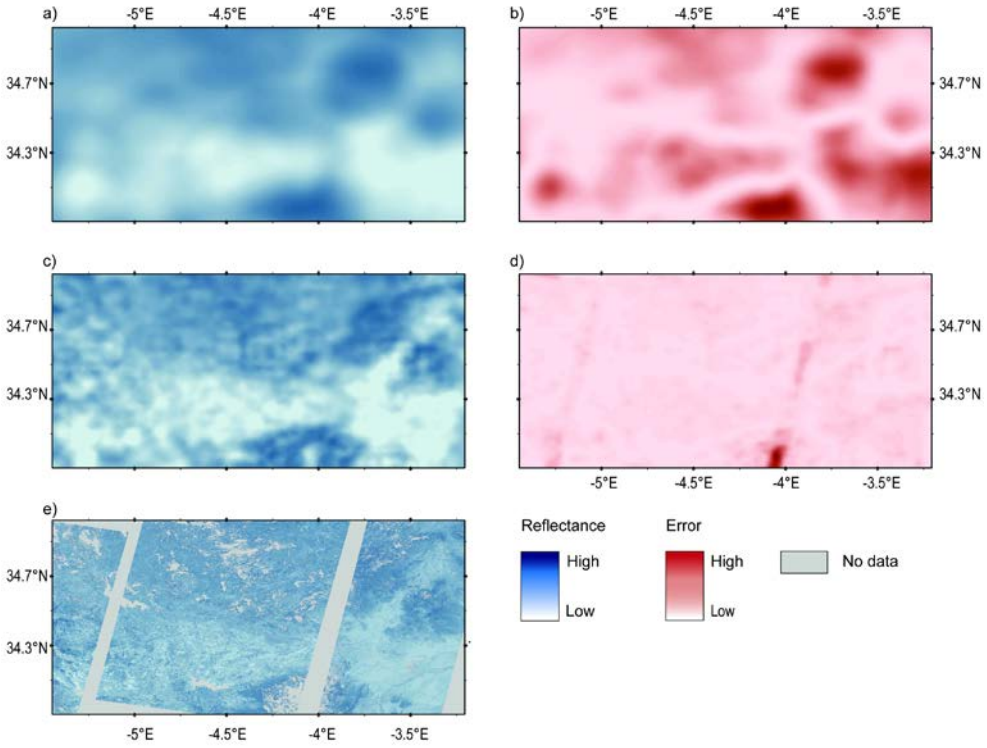


Figure 6.4: Smoothed RS data (band 1) to the (a) long-range and (c) medium-range compared to (e) the original data along with the prediction errors (b,d) introduced by FRK.

Table 6.6: Relative prediction errors (%), averaged per sensor for the medium and long-range RS data.

| Range | VNIR | SWIR | TIR | PC | DEM | NDVI |
|---|---|---|---|---|---|---|
| Medium | 6.3 | 5.4 | 11.8 | 6.5 | 3.1 | 8.0 |
| Long | 3.7 | 4.5 | 32.0 | 6.4 | 9.4 | 12.0 |

The correlation between mineralogy and smoothed RS data was found to be higher for several RS variables (Fig. 6.5). Calcite (Fig. 6.5a) showed stronger correlation for the VNIR- SWIR bands at both scales. The effect of smoothing on the TIR and PC data had mixed effect on the correlation. Most striking was the higher correlation for PC1 and lower correlations for PC3 and NDVI. The correlation of mica improved using scaled RS data (Fig. 6.5b). Compared to calcite, the effect was less in the VNIR-SWIR bands but more apparent in the TIR bands. By removing the noise from the data, the typical absorption features of kaolinite in the SWIR were expected to improve the correlation. This was indeed observed for the long-range data (Fig. 6.5c). In contrast, the correlations slightly decreased with the medium-range data, except for the NDVI and TIR band 13. Smectite had low correlation with the RS data (Fig. 6.5d) but smoothing increased the correlation with the TIR data. Smectite has strong absorption features in the SWIR bands, even so the correlation of smectite did not increase with these bands after smoothing.

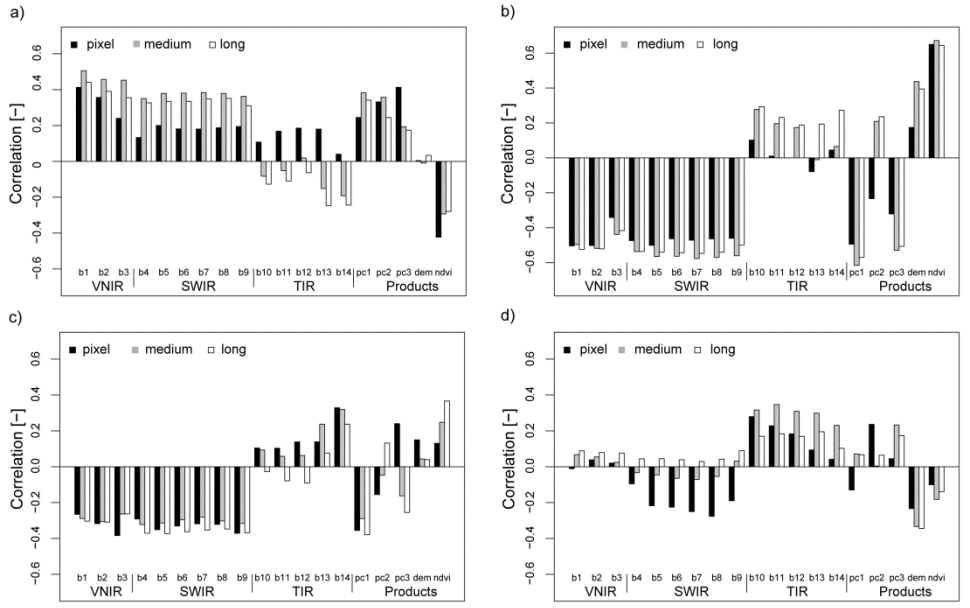


Figure 6.5: Correlation of RS data representing the pixel, medium and long-range variability with (a) calcite, (b) mica, (c) kaolinite and (d) smectite abundances (wt. %).

### 6.4.3 Prediction models for soil mineralogy

*6.4.3.1 Prediction of mineral categories using spectroscopy*

The overall accuracy of the MNL models using the original RS data were relative high (> 0.74) (Table 6.7). The smoothed RS data, regardless of scale, improved the overall accuracy little, except for mica using medium-scale RS data. Nevertheless, using a combination of medium and long-scaled RS data substantially improved the overall accuracy of the models. The probability of a sample belonging to a specific class also increased using a combination of medium and long-scaled RS data (Fig. 6.6). Many samples were positioned around the center of the diagram using the original RS data, indicating substantial confusion between the categories. This improved by using scaled RS data, because many samples moved towards the corners of the diagrams. The remaining samples located along the axes indicated confusion between two categories, which was largest for absence and presence in a calcite-poor environment. High confusion between the smectite classes remained, despite using scaled RS data (Fig. 6.6c).

Table 6.7: Overall classification accuracy of the predicted mineralogy by MNL using the sample data.

| Mineral | Pixel | Medium | Long | Medium + Long |
|---|---|---|---|---|
| Mica | 0.74 | 0.70 | 0.79 | 0.76 |
| Smectite | 0.78 | 0.79 | 0.74 | 0.89 |
| Kaolinite | 0.75 | 0.79 | 0.76 | 0.86 |

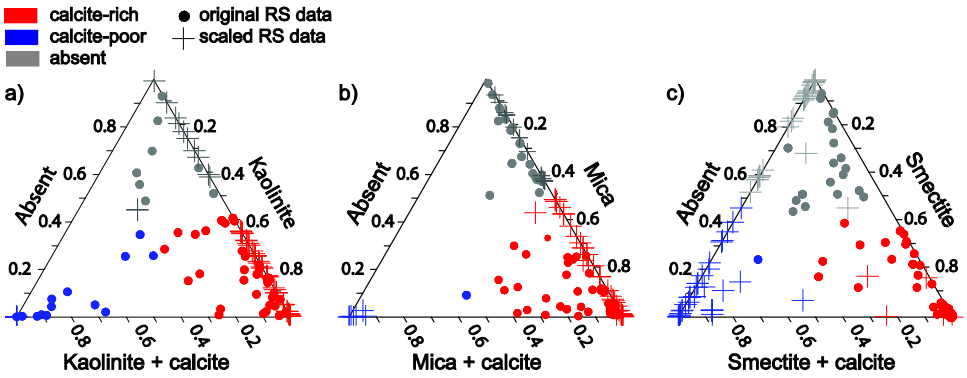


Figure 6.6: Ternary diagrams with MNL classification probabilities of kaolinite (a), mica (b) and smectite (c).

*6.4.3.2 Prediction of mineral abundances based on x-ray diffraction analysis*

Models using the original RS data resulted in low coefficients of determination and high RMSE values for all minerals and were deemed unsuitable for predicting mineral abundances (Table 6.8). The use of individual scaled RS data had mixed effect on the prediction models and most accurate models were derived for calcite and mica. We found that the prediction models resulted in highest accuracy using both the medium and long-range RS data. The mica and calcite models performed well for both the medium and the combined ranges; the model statistics showed high $R_{adj}^2$ and lowest RMSE for predictions from the calibration and cross-validation. The goodness-of-fit for smectite indicated that 57% of the variation was explained by the model. The model for kaolinite was limited to explain 45% of the variation. These results indicated that the use of smoothed data improved model performance.

### 6.4.4 Digital soil mapping for mineral characterization

*6.4.4.1 Application of prediction models to the study area*

This section presents the spatial explicit prediction models for mica because of its high predictive accuracy (Table 6.8). The effect of using medium (Fig. 6.7b) and long-range variability (Fig. 6.7c) was clearly observed. Employing the smoothed data, areas of high and low abundances were well differentiated whereas the original data produced a map that hardly showed spatial patterns. Model predictions in areas interpolated with FRK seamlessly joined most of the spatial patterns of areas where RS data were available. The mineralogy was not well characterized in the nodata-area in the East, running from North to South using medium-scaled RS data. In this area, the accumulated FRK prediction errors (Fig. 6.4) were too high and the

Table 6.8: Model accuracy for the predicted minerals by MLR

| Mineral | Scale | $R_{adj}^2$ | $RMSE_{cal}$ | $RMSE_{val}$ |
|---|---|---|---|---|
| Calcite | Pixel | 0.42 | 12.8 | 14.2 |
| | Medium | 0.54 | 10.6 | 17.3 |
| | Long | 0.34 | 12.7 | 22.9 |
| | Medium + long | 0.71 | 8.9 | 12.0 |
| Mica | Pixel | 0.46 | 3.4 | 3.7 |
| | Medium | 0.61 | 4.9 | 9.0 |
| | Long | 0.55 | 5.2 | 8.9 |
| | Medium + long | 0.70 | 4.6 | 6 |
| Smectite | Pixel | 0.53 | 7.4 | 10.2 |
| | Medium | 0.29 | 9.2 | 25.8 |
| | Long | 0.36 | 7.8 | 19.5 |
| | Medium + long | 0.57 | 6.0 | 8.3 |
| Kaolinite | Pixel | 0.26 | 2.9 | 3.2 |
| | Medium | 0.10 | 3.6 | 7.9 |
| | Long | 0.24 | 2.9 | 4.5 |
| | Medium + long | 0.45 | 2.5 | 3.8 |

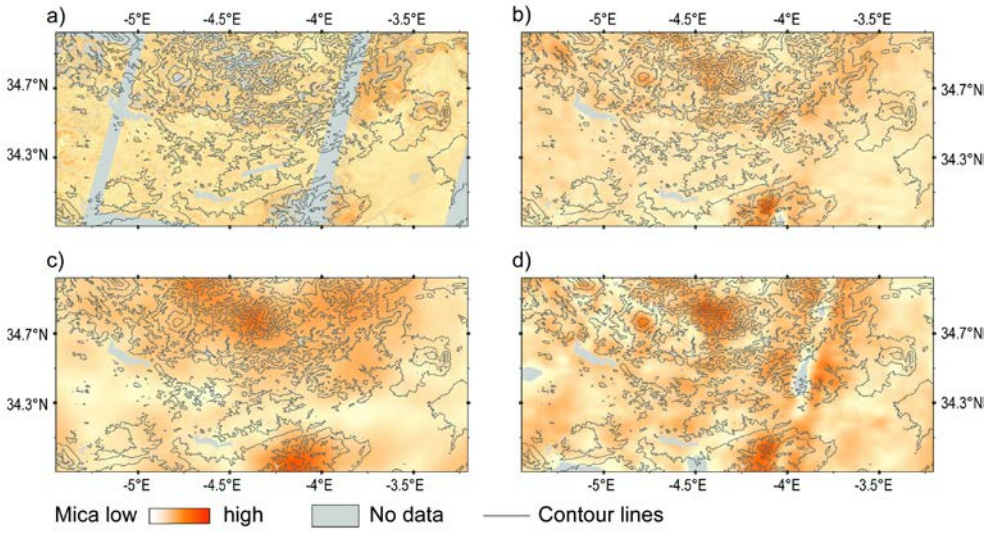


Figure 6.7: Predicted mica based on XRD data and using the (a) original RS data, (b) RS data representing medium-range variability, (c) RS data representing long-range variability and (d) combination of scaled RS data.

predicted abundances disrupted the smooth spatial patterns observed in the map (Fig. 6.7d). The mapped calcite abundances were overestimated in the 30 wt. % and 50 wt. %-interval (Fig. 6.8a) causing a shift in the median of 4.5 wt. % ($p$=0.02). Similarly, mica was overestimated in the 15 wt. %-interval (Fig. 6.8b) causing a shift in the median of 1.5 wt. % ($p$=0.02). The distribution of modelled and measured kaolinite abundances were similar ($p$=0.05) (Fig. 6.8c). Smectite abundances were predicted outside the range of measured abundances (Fig. 6.6d), yet, they originated from the same distribution ($p$=0.53). The modelled and measured mineral categories did not have a similar distribution ($p < 0.01$) using the initial thresholds (Table 6.9). Increasing the threshold of mica to 5 wt. % resulted in a similar distribution ($p = 0.51$). 22 wt. % of the mapped kaolinite falsely represents the presence of kaolinite in either a calcite-rich or -poor environment ($p < 0.01$). Smectite required a lower calcite threshold (3 wt. %, $p = 0.63$).

Table 6.9: Proportional distribution of the sampled and modelled mineral categories.

| Mineral class | Mica | | | Kaolinite | | Smectite | | |
|---|---|---|---|---|---|---|---|---|
| | Sample | Sample* (mica 5 wt. %) | Map | Sample | Map | Sample | Sample* (calcite 3 wt. %) | Map |
| Presence Ca-rich | 0.35 | 0.24 | 0.20 | 0.23 | 0.29 | 0.21 | 0.34 | 0.40 |
| Presence | 0.49 | 0.40 | 0.45 | 0.40 | 0.56 | 0.27 | 0.14 | 0.13 |
| Absence | 0.15 | 0.36 | 0.35 | 0.37 | 0.15 | 0.52 | 0.52 | 0.47 |

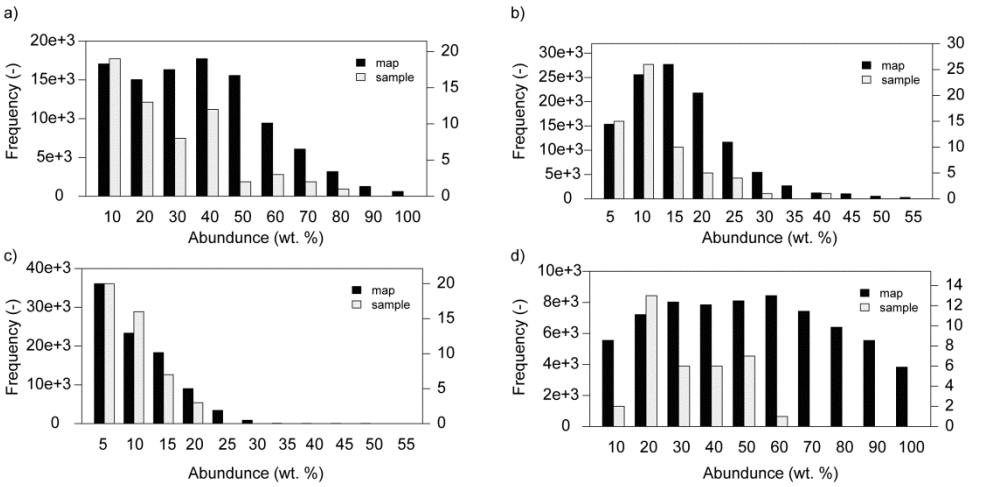


Figure 6.8: Histograms of the sampled (right axis) and mapped (left axis) abundances (wt. %) of (a) calcite, (b) mica, (c) kaolinite and (d) smectite.

### *6.4.4.2 Evaluation of predicted soil mineral composition using spectroscopy or x-ray diffraction*

The overall similarity between the maps with reclassified abundances and mineral categories were relative low (between 0.34 and 0.44, Table 6.10). Nevertheless, substantial agreement between to two approaches was found using a more qualitative comparison using the error matrices and observed spatial patterns. Areas with high abundances (Fig. 6.9) coincide with the minerals classified as present either in a calcite- rich or poor environment (Fig. 6.10 and Table 6.10). High concentrations of calcite (Fig. 6.9d) showed similar patterns as mica (Fig. 6.10a) and kaolinite (Fig. 6.10b) in a calcite-rich environment, although, the error matrices indicated relative high misclassifications. The absence of smectite had high similarity (Table 6.10c) but the other two classes deviated strongly, which agrees with the findings in section 6.3.2.1. High user accuracy was found for the presence of kaolinite, but substantial spatial differences were observed, especially in the North (Fig. 6.9b and Fig. 6.10b). As a result, the error matrix shows low agreement between the mapped absence of kaolinite. Finally, the effect of the FRK prediction errors were less pronounced in the mapped mineral categories.

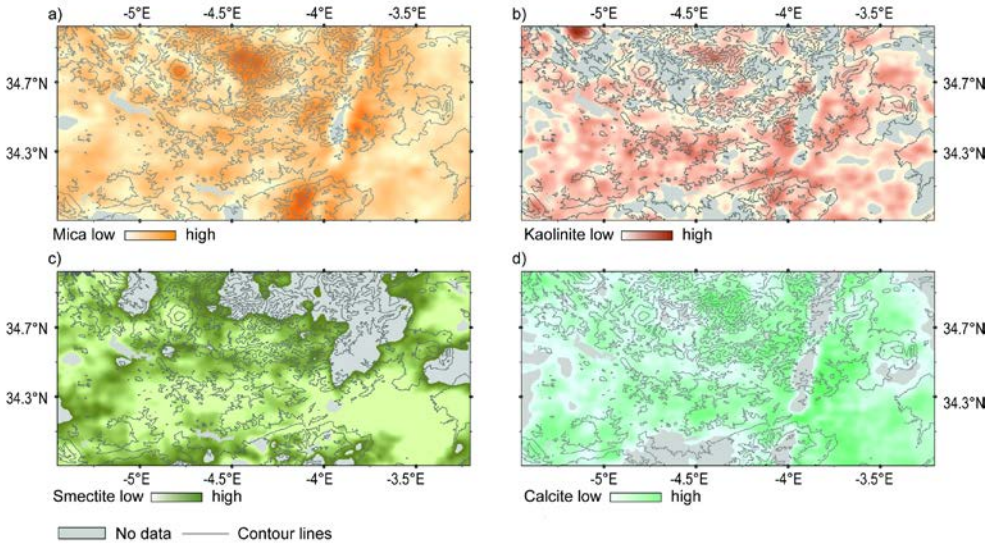


Figure 6.9: Predicted soil mineral abundances, based on XRD data and scaled RS data.

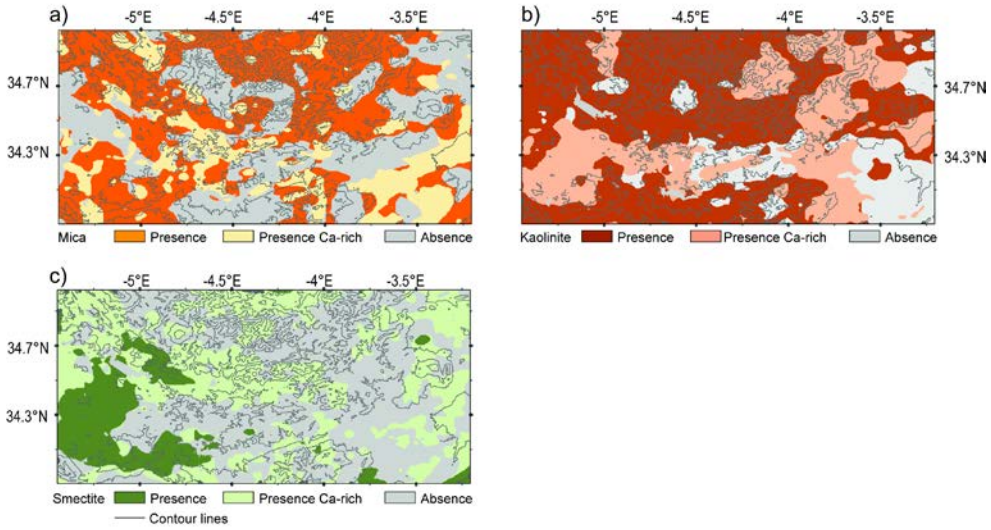


Figure 6.10: Predicted soil mineral categories based on spectral point measurements and scaled RS data.

Table 6.10a: Error matrix (%) mica characterization.

| Reclassified abundances | Presence Ca-rich | Presence | Absence | User accuracy |
|---|---|---|---|---|
| Presence Ca-rich | 0.13 | 0.28 | 0.14 | 0.23 |
| Presence | 0.04 | 0.13 | 0.12 | 0.43 |
| Absence | 0.03 | 0.04 | 0.07 | 0.50 |
| Producer's Accuracy | 0.64 | 0.28 | 0.22 | 0.34* |

*Overall accuracy

Table 6.10b: Error matrix (%) kaolinite characterization.

| Reclassified abundances | Presence Ca-rich | Presence | Absence | User accuracy |
|---|---|---|---|---|
| Presence Ca-rich | 0.16 | 0.24 | 0.08 | 0.34 |
| Presence | 0.04 | 0.12 | 0.02 | 0.66 |
| Absence | 0.08 | 0.19 | 0.05 | 0.16 |
| Producer's Accuracy | 0.57 | 0.22 | 0.34 | 0.34* |

*Overall accuracy

Table 6.10c: Error matrix (%) smectite characterization.

| Reclassified abundances | Presence Ca-rich | Presence | Absence | User accuracy |
|---|---|---|---|---|
| Presence Ca-rich | 0.16 | 0.06 | 0.17 | 0.40 |
| Presence | 0.10 | 0.04 | 0.07 | 0.20 |
| Absence | 0.13 | 0.02 | 0.22 | 0.60 |
| Producer's Accuracy | 0.40 | 0.34 | 0.47 | 0.43* |

*Overall accuracy

## 6.5 Discussion

### 6.5.1 Main findings

#### *6.5.1.1 Data collection*

The MICA analysis applied to the samples had difficulties in matching some minerals present in the sample with the spectral library (Mulder et al., 2012b). The MICA analysis used in this study was adapted from an approach developed at the remote sensing level of spectroscopy (Kokaly et al 2011a) rather than the laboratory level. To improve the analysis, the expert-decision rules need to be further refined to match the level of detail of high-resolution spectral data. More specifically, this involves specific adjustments of the characteristics describing the diagnostic absorption features. Despite that, MICA delivers insightful information for large-scale studies, consisting of different mineral categories. This provides important information on parent material and soil formation. The highest interest for regional

scale studies, are the coarse differences and general spatial patterns of minerals. Therefore, spectroscopy is considered to be an important and cost efficient way to obtain mineral information. In addition, mineral compositions derived using the MICA algorithm are based on absorption features in reflectance signatures, thus the basis of mineral identification is similar to absorption processes causing variations in the satellite remote sensing data. By contrast, XRD results are based on different types of processes arising in the interaction of electromagnetic energy with the sample.

In this work, x-ray analysis was used to obtain mineral abundances. This involved the use of costly equipment while the analysis was labour-intensive. Alternatively, mineral abundances could be estimated using other spectral modelling approaches, such as proposed by Mulder et al. (2013).

*6.5.1.2 Variogram analysis and Fixed Rank Kriging*

The different scales of spatial variability within the area were determined by variogram analysis (Fig. 6.3). The high nugget values (20% to 43% of the total semivariance) suggested the presence of a substantially amount of noise in the original RS data. Also, for some RS data more variance was captured by the long-range structure than the medium-scale structure. These findings were further used in the FRK. Here, we found lower prediction errors for RS data having lower sill values at the medium-scale. This indicated that the basis functions had more difficulty in capturing the spatial process for strongly spatially correlated RS variables. The prediction errors might be further reduced by increasing the number of basis functions. At the long-scale the opposite was found, this indicated that the increased nugget reduced the prediction accuracies for those RS data having lower spatial correlation. Overall, the resulting FRK prediction errors and the relation with the variogram were confirm the expectations (Isaaks and Srivastava, 1989).

FRK generated a full coverage by interpolating missing areas with low FRK prediction errors up to a range of 6 kilometres. The missing data in the Eastern part of the area, running from North to South (120 km) having a width of 10 kilometre, appeared to be too large. The FRK positioned the medium-range basis functions at a resolution of 5 and 9.5 kilometre. As a result, there were too few observations available to accurately interpolate. The resulting prediction errors accumulated in the mineral prediction models, increased the average prediction errors (Table 6.6) and resulted in unrealistic predictions. As such, the accuracy of interpolation by FRK is mainly defined by the extent of the missing data and the desired map scale. Discrepancies between the latter substantially increase the prediction errors. The MNL was more robust for this phenomenon because the RS data was related to three categories rather than a continuous range of abundances.

*6.5.1.3 Prediction models for soil mineralogy*

Initially, the effect of smoothing RS data seemed relatively small in relation to mineralogy (Fig. 6.5). However, the prediction models clearly indicated that using original RS data resulted in lower model performance (Table 6.6 and 6.8). Noise in the original data (i.e. short scale variability) was removed by the FRK, leading to improved relations of smoothed RS data with mineralogy. In addition, the model accuracy and sample sizes of kaolinite were smaller than for calcite and mica. This may indicate that the MLR might have been hampered by sample size rather than by the soil-landscape relationships. Alternative to MLR, different analysis might improve the prediction of mineralogy. MLR assumes a linear relationship of mineralogy towards the predictor variables but this relation might be more complex. Recursive partitioning of the data, for example, can deal with nonlinearity and interactions between predictor variables (Breiman et al., 1984).

The most accurate models utilized a combination of scaled RS data. This indicates that variability of environmental factors explain soil variability better at a specific scale. In addition, the resulting spatial patterns of the soil system's variability at regional scale were more representative. Vasques et al. (2012) also found that the use of medium and long-range variability observed in spatial data were needed to model major soil variability at large scales. They found that soil carbon was controlled by ecological processes that evolved and interacted over a range of spatial scales across the landscape. They suggested there were two appropriate scales to observe soil carbon and their results supported our use of multi-scale soil-landscape relationships.

*6.5.1.4 Evaluation of maps*

The soil mineral maps showed similarity between both approaches (Fig. 6.9 and 6.10). Unfortunately, the overall accuracy was rather low (between 0.34 and 0.44, Table 6.10). The modelled calcite and mica abundances were overestimated. As a result, the thresholds which were based on the measured abundances, were too low (Fig. 6.8). This resulted in higher misclassification of the presence of mica or kaolinite in a calcite-rich or -poor environment (Table 6.10a, b). Besides, comparison of kaolinite abundances (Fig. 6.9b) and categories (Fig. 6.10b) may not be that informative. Typically, kaolinite can be well discriminated by spectral analysis and the MNL model resulted in accurate models, while the MLR model was weak ($R^2$ = 0.45) (Table 6.7 and Fig. 6.6). This may imply that the MICA approach was more reliable for kaolinite. Reflecting on the methods and findings of this paper several sources of errors could be identified that contributed to the disagreement, (1) the reclassification of abundances into mineral categories was subjective due to the unknown spectral detection thresholds and (2) soil mineral abundances were less accurately modelled due to the accumulation of the FRK predictions errors. As such, major differences resulted from the classification of absence in areas where low or unrealistic abundances were mapped. Overall, the spectroscopy-based maps

characterized regional soil mineral composition and defined areas with a high likelihood of occurrence of minerals.

The evaluation in section 6.4.4 was based on the distribution of the collected sample (Mulder et al., 2012a). Here, two assumptions were made, first, the spatial distribution of mineralogy was assumed to be fully sampled. Second, the overrepresented and missing intervals of the LHS were not considered to influence the distribution of mineralogy. The first assumption is inherent to the choice of using cLHS (Minasny and McBratney, 2006). The RS data was assumed to be representative of the landscape and its environmental variables (CLORPT, Jenny, 1941). By sampling the first three principal components of ASTER data and elevation, the sampling aimed to optimally capture the environmental variability which was thus the proxy for soil variability. As such, reflectances are a result of the environmental variables but also an expression of interactions. The precise interactions are unknown but they might influence soil variability, e.g. soil-vegetation feedbacks (Milcu et al., 2012) and, therefore, the sample may not be an optimal representation of the actual soil variability. The second assumption should be further investigated because the realized sample did not fully cover the LH covariate space. Therefore, careful interpretation of the statistical tests employed in section 6.4.4 is in place. Additional field data may be acquired to complete the coverage of the LH, and independent data is deemed necessary to validate the final prediction models and maps.

### 6.5.2 Imaging spectroscopy for geologic applications

Studies reviewed by Van der Meer (2012) demonstrate that, at the moment, RS data does not provide the high spectral resolution needed to quantitatively map mineralogy. This was also supported by Cudahy et al. (2012) who emphasize that the ASTER data does not provide sufficient spectral resolution to determine the abundance or chemistry of specific minerals. The physical nature of minerals is complex and difficult to model in a quantitative way using spectroscopy at the remote sensing level. Nevertheless, using an advanced spectral feature comparison method, spectroscopy at the remote sensing level can deliver informative data supporting regional-scale DSM; MICA allowed characterizing more minerals with a higher accuracy than other spectral feature comparison methods. Besides, a geostatistical approach using a small sample having the required quantitative data and spectral resolution (laboratory level) significantly improves the feasibility to quantitatively map mineralogy. Finally, this approach can also be applied to model other environmental properties on various spatial scales, because of the underlying basic principle, the Soil-Landscape paradigm (Jenny, 1941). DSM uses this paradigm to develop empirical models that relate observations of various soil properties, e.g. soil carbon, with environmental variables that describe the main soil-forming factors (McBratney et al., 2003; Vasques et al., 2012).

## 6.6 Conclusion and outlook

In this work we aimed at characterizing regional soil mineral composition, based on scale-dependent spatial variability observed in RS data. The focus of this work was to improve prediction models using scaled variability in RS data that matched the scale of variability of the target sample. The results indicated that the use of RS data smoothed to target scales generally resulted in better prediction accuracies than when using the original RS data. Moreover, soil information obtained from spectroscopy provided sufficiently detailed data to characterize the spatial distribution of the soil mineral categories, present in the study area.

The best models utilized a combination of medium and long-range variability. This indicates that the variability of environmental factors improve the explained soil variability when represented at specific scales. As such, large-scale DSM approaches, based on the Soil-Landscape paradigm (Jenny, 1941), may improve their empirical models by using these multi-scale soil-landscape relationships. In addition, PS provided important primary information while RS provided all secondary data for the prediction models. Obtaining spectroscopic data is time and cost efficient, therefore, its use should be seriously considered for large-scale assessments. This work contributed substantially to DSM by advancing methods for large-scale assessments. The prediction models were improved by (1) representing explanatory variables more closely to the spatial variability of soil and environmental properties at regional scale and (2) using existing multi-scale soil-landscape relationships to predict soil properties. The approach was cost-efficient because it relies on globally available data and advanced soil collection methods while computational demands are relatively low.

### Outlook

This work showed the potential of using scale-dependent RS data for large-scale DSM. However, this study was one of the first steps exploring the use of large datasets, geostatistical approaches and RS and PS methods simultaneously. Future research could focus on using additional environmental data to account for all factors of soil formation (Jenny, 1941). Another interest would be studying the information originating from micro-scale or local scale variability, which was not accounted for in this study (Katzfuss, 2011; Katzfuss, 2013; Sang and Huang, 2012). This allows mapping those regions that were now considered to be outliers due to their local occurrence, e.g. hydrothermal alteration zones having specific mineral occurrences. These areas are important to improve our understanding of local soil formation. Fixed Rank Kriging (FRK) is also an option that could be employed for geologic remote sensing in vegetated areas. FRK allows to generate a full coverage by interpolating areas where vegetation hampers extracting mineral information (Cressie and Johannesson, 2008). Yet, the feasible depends on the coverage and spatial

distribution of vegetation and the desired map scale of the modelled properties. For this, future research is needed to define its use in vegetated areas.

**Acknowledgements**

We acknowledge financial support from the EU FP7 program under the e-SOTER project (contract 211578). R. Kokaly was supported by the Climate and Land Use Change Research & Development Program of the U.S. Geological Survey. We thank M. Katzfuss for insightful comments on Fixed Rank Kriging. We thank M. Plötze and C. Mavris for their assistance is the mineral diffraction analysis. We acknowledge INRA Maroc for their support with sampling field data.

Disclaimer: "Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government."

# Chapter 7

# Synthesis

## 7.1 Main findings

Global environmental changes have resulted in moderated key ecosystem services that soils provide. It is necessary to have up to date soil information on regional and global scales to ensure that these services continue to be provided. As a result, digital soil mapping (DSM) research priorities are among others, advancing methods for data collection and analyses tailored towards large-scale mapping of soil properties. Within this context, the objective of this thesis was to exploit the use of remote and proximal sensing methodologies for digital soil mapping in order to facilitate soil mapping at regional scale. Different methods were assessed and developed for operational DSM at regional scale, to reach this objective. Each of these methods focussed on one of the research questions presented in section 1.4. Below, the answers to the main research questions are revisited and the results are discussed in relation to the problem setting discussed in Chapter 1.

1: *What is the current state-of-art in the use of remote sensing for soil and terrain mapping?*
The use of existing remote and proximal sensing (RS and PS) methods for soil survey support three main components in DSM (Chapter 2). First, RS data support the segmentation of the landscape into homogeneous soil-landscape units whose soil composition can be determined by sampling. These soil-landscape units have also been used as a source of secondary information in DSM. Secondly, RS and PS methods allow for inference of soil properties using physically-based and empirical methods. Thirdly, RS data supports spatial interpolation of sparsely sampled soil property data as a primary or secondary data source. The progress made in this field is discussed below, as well as remaining limitations.

PS has been successfully used for deriving quantitative and qualitative soil information (Viscarra Rossel et al., 2006). Most reported studies were done at the laboratory and local scale focusing on estimating soil properties having clear absorption features (Ben-Dor et al., 2008). These studies showed the high potential of PS but for large-scale mapping of soil properties, methods need to be extended beyond the plot. Various soil properties are difficult to characterize using spectroscopy due to the lack of diagnostic absorption features and complex scattering behaviour within the soil mixture (Clark and Roush, 1984). Quantification and qualification of such soil properties require methods that model the complex scattering behaviour of soils (Chapter 4). For that, the sample preparation, spectral measurements, data analysis and model parameterization require special expertise (Pompilio et al., 2010). Besides, these methods have not been fully developed.

Important qualitative and, to a lesser extent, quantitative soil information has been derived from RS data (Chapter 2). It was found that RS-derived information has a lower accuracy and feasibility to obtain information compared to PS. RS provides

qualitative information on those soil properties that have clear diagnostic absorption features. The main limiting factors to derive the same soil properties as with PS are (1) the coarse spatial and spectral resolution, (2) the low signal-to-noise ratio of high-resolution RS data and (3) the bands of multi-band RS sensors have not been positioned at diagnostic wavelengths.

Reported DSM-studies made limited use of the various methods that are available for spectroscopy and geostatistics (Ben-Dor et al., 2009; Dewitte et al., 2012). It was found that current research using RS data typically produces qualitative outputs. Also, the overall model accuracy reduced with increasing scale of the study area. This was contributed to incompatibility between the RS data and the available sample data. From the viewpoint of the soil scientists, the major gap is the lack of readily available RS-based soil products. Currently, soil scientists generate their own input data for their models. However, they may be limited in their knowledge of RS and PS tools and methods. Here, communication is the limiting factor which is needed to initiate a multidisciplinary approach for soil mapping.

Chapter 2 of this thesis indicated a large potential of using RS and PS methods for DSM, yet, advances are deemed necessary to fully develop large-scale methodologies. Advances may be expected in developing more quantitative methods and enhanced geostatistical analysis using RS and PS data by making use of recent developments in DSM-related disciplines.

2: *Can major soil variability at regional scale be represented by a sparse remote sensing-based sampling approach?*

Budgetary constraints commonly limit the sample size of soil mapping efforts at larger scales. Furthermore, areas might not be accessible due to either the rough terrain or political reasons. Also, often there is limited soil legacy data available supporting the sampling strategy (de Gruijter et al., 2006). Collecting an exhaustive sample that honours a geometric criterion or represents the full variability requires extensive fieldwork and analysis, or may not be feasible at all. Considering the involved cost and time, current methods for data collection are not adequate (Grunwald, 2011). The consensus is to collect a sparse sample that captures the major soil variability while minimizing the acquisition efforts.

RS data can serve as proxy for soil variability alternative to the required soil legacy data that supports the sampling strategy. The Soil-Landscape paradigm (Jenny, 1941) is the basic principle underlying the use of RS data. This paradigm indicates that the degree of sampled variability from exhaustive RS data is representative for the same degree of soil variability within the landscape. A sampling strategy aiming to sample the feature space of the RS data with minimal acquisition efforts is the most efficient way to collect data. Latin Hypercube Sampling (cLHS) allows optimizing such a sampling strategy (Minasny and McBratney, 2006).

It was found that the RS-based sampling strategy successfully represented major soil variability (Chapter 3). Apart from that, the cLHS sample did not express spatial correlation; constraining the LHS by a distance criterion favoured large spatial variability within short distances. The absence of spatial correlation in the sampled soil variability precludes the use of additional geostatistical analyses to spatially predict soil properties. Predicting soil properties using the cLHS sample is thus restricted to a modelled relation between the sample and exhaustive predictor variables. The RS data were deemed suitable predictors. This data provided the necessary spatial information because of the strong spatial correlation while the spectral information provided the variability of the environment (Chapter 3 and 6).

The LHS approach is considered a time and cost efficient method for acquiring information on soil resources over extended areas. This information is crucial to the identification and monitoring of those resources (Chapter 1). If data availability and resources are scarce, this approach provides a small representative dataset with soil properties. These can be used for mapping soil properties using modelled relations between the sampled properties and RS data (Chapter 6).

3: *Which methods allow retrieval of mineralogy from complex mixtures using proximal sensing?*

PS has been recognized as potential data source for retrieval of mineralogy and other soil properties (Chapter 1). Various methods are available for estimating soil properties from PS data. Most of these methods provide qualitative soil information and are restricted to derive information of soil properties that have distinct spectral diagnostic features (Chapter 2). For these reasons, PS-based methods are needed for providing quantitative information on a wider set of soil properties. Soil mineralogy can be used for developing such methods (Clark and Roush, 1984).

Estimating mineralogy using PS and a linear modelling approach is hampered by the strong influence of complex scattering within the mixture and overlapping absorption features (Chapter 4). It was demonstrated that a state-of-the-art linear spectral feature comparison method, the Material Identification and Characterization Algorithm (MICA) (Kokaly, 2011a), can be used to characterize the dominant mineralogy (Chapter 4). Mineral categories that can be characterized include the presence of mica, kaolinite and smectite either in a calcite rich or poor environment, among others. Moreover, this method appeared sufficiently detailed to map regional mineral variability (Chapter 6). Retrieval of refined information from natural samples, such as mineral abundances is more complex, since mineral absorption features may be less distinct or even absent. Estimating mineral abundances requires a method that accounts for the interaction between minerals within the intimate mixture. This can be done by addressing the interaction with a non-linear model.

Quantifying mineral abundances of mixtures having more than two constituents can be estimated using PS (Chapter 5). Spectral deconvolution of SWIR spectra (2.1–

2.4 µm) coupled with regression tree analysis allows simultaneous quantification of more than two minerals within a mixture. Accurate deconvolution of natural samples using Gaussian bands requires the use of Exponential Gaussian Optimization (EGO) (Pompilio et al., 2009) rather than the Modified Gaussian Model (MGM) (Sunshine and Pieters, 1993). It was found that the modelled asymmetry and saturation were key to deconvolve the reflectance, when using the simple representation of the complex scattering behaviour by a few Gaussian bands. Also, asymmetry of the EGO profiles shows to be an important predictor variable for the prediction of mineral content of field samples. Compared to MGM and the single scattering albedo model of Hapke (Hapke, 2002), the presented model is easier to parameterize and model accuracies for estimated mineral abundances from natural samples are higher.

Based on the findings of Chapter 4 and 5, it can be concluded that more advanced unmixing algorithms do improve the estimation soil mineralogy. Firstly, a linear model within an expert-based decision framework allows characterizing dominant mineralogy. Secondly, EGO, in combination with regression trees, allows quantification of mineral abundances within complex mixtures. Spectral deconvolution has not been widely used to retrieve soil information using spectroscopy. Employing spectral deconvolution coupled with regression tree analysis does not require additional information compared to linear spectroscopic analysis methods. This approach advances existing PS methods and has the potential to quantify other soil properties, e.g. soil moisture (Whiting et al., 2004). As such, the soil science community was provided an improved inference method to derive and quantify soil properties.

4: *Can scale-dependent variability be extracted from remote sensing and do model predictions improve by using scaled remote sensing data that match the variability of the sample?*

Using a sparse sample for regional-scale DSM is a challenging task. Especially for soil prediction models that rely on the statistical relation between the sample and predictor variables. These models may be hampered in determining the relation of a sparse sample with predictor variables for properties with high spatial variability over short distances. The relation tends to weaken with increasing scale and extent of the study area, which results in lower prediction accuracies. This can be contributed to a lack of compatibility between the predictor variables and the sample. Model accuracies might be improved by using predictor variables that represent the same scale.

Major mineral variability can be characterized using a sparse sample, spectroscopy and geostatistics. First, the RS-based predictor variables can be smoothed to improve the compatibility of the data with the sample that is representable for major variability at regional scale. This can be achieved by considering the medium and long-range spatial variability in the RS data (Goovearts,

1997). Smoothing the data to the corresponding ranges can be done using Fixed Rank Kriging (FRK) (Cressie and Johannesson, 2008). This method allows smoothing the massive RS datasets to the observed ranges and removes the short scale variability and measurement noise. The resulting RS data resemble more closely the spatial variability of soil and environmental properties at regional scale. Besides, FRK provides a full data coverage for a study area because missing data is interpolated. Next, with this scaled RS data, regression models can be made for predicting the major mineral variability. Here, the combined use of the medium and long-scaled data allows to model multi-scale soil-landscape relationships (Vasques et al., 2012). The original RS data was found to be unsuitable for developing regression models predicting calcite, mica, kaolinite and smectite (Chapter 6). On the other hand, the models using scaled RS data substantially increased the prediction accuracies. This is mainly contributed to the improved compatible of the data and the use of multi-scale soil-landscape relationships (Chapter 6). Following this approach, more accurate predictions and representative spatial patterns of major soil variability can be obtained.

Alternatively to the DSM approach, imaging spectroscopy has been used to map soil mineralogy. Recent studies (van der Meer et al., 2012) demonstrated that, at the moment, RS data does not provide the high spectral resolution that is needed to quantitatively map soil mineralogy. The physical nature of minerals is too complex (Clark, 1999) to be modelled in a quantitative way using imaging spectroscopy alone. The use of a geostatistical approach in combination with a small representative sample substantially improves the feasibility to quantitatively map mineralogy.

In the context of spatial modelling at regional scale, the aim was to model areas where high concentrations of specific soil properties occur, rather than quantifying soil properties at high resolution. This consensus resulted in improved model accuracies and thus higher quality of delivered soil information. In addition, this approach provides (soil) information over vast areas at a spatial and thematic resolution that corresponds to other fundamental (global) data sets such as those for weather, climate, land cover and geology (Hengl, 2009). Environmental models will improve with a proper integration of soil data, which can be achieved when the data have matching scales.

In conclusion, scale-dependent variability in RS data can be extracted and does improve soil prediction models that rely on a sparse sample. This method has potential to model various natural resources at large spatial scales and thereby enhances the perspective of a global system for inventorying and monitoring the earth’s soil resources.

## 7.2 Reflection and outlook

This research was motivated by the need to improve our knowledge about remote sensing (RS) and proximal sensing (PS) methods for digital soil mapping (DSM) at regional scale. Scientifically, this thesis contributed to the development of methodologies, which aim to optimally use RS and PS for DSM to facilitate soil mapping at regional scale. The main contributions of this thesis with respect to the latter are (I) the critical evaluation of recent research achievements and identification of knowledge gaps for large scale DSM using RS and PS data, (II) the development of a sparse RS based sampling approach to represent major soil variability at regional scale, (III) the evaluation and development of different state-of-the-art methods to retrieve soil mineral information from PS, (IV) the improvement of spatially explicit soil prediction models and (V) the integration of RS and PS methods with geostatistical and DSM methods. Following these five contributions, it can be concluded that: *Improvements in regional-scale DSM result from the integrated use of RS and PS with geostatistical methods.* In every step of the soil mapping process, spectroscopy can play a key role and can deliver data in a time and cost efficient manner. Although existing methods have demonstrated the value of spectral data in DSM, this thesis stressed that methods need the support of geostatistics and ground truth data in order to advance models for regional-scale DSM. This thesis improved the integration for large-scale DSM, despite the limited available soil legacy data and tight constraints on budget and time to obtain data for DSM. Actually, the constraints forced us to find sophisticated solutions and to combine the most recent advances made in PS, RS and geostatistics. Reflecting on the findings in this thesis, some issues have not been discussed yet. These are focussed on (I) quantification of soil mineralogy using PS and (II) mapping soil properties at regional scale using PS, RS and geostatistics.

### 7.2.1 Quantification of soil mineralogy using proximal sensing

This thesis has shown the feasibility to estimate up to four minerals from a soil mixture (Chapter 5). This is a substantial improvement compared to methods based on the MGM (Sunshine and Pieters, 1993), since these were limited to estimate two minerals within prepared mixtures. Methods that were based on Hapke's (Hapke, 1981) model, proved to be evenly successful in estimating mineral abundances, although their analysis on natural samples was less accurate. The strongest point of our method is the need of few input parameters to quantify up to four minerals, namely the measured reflectance and known mineral variability within the sampled area. Like with other methods, the weakest point of our approach is that training of the model requires laboratory spectral measurements of natural samples. Prerequisite is thus a small sample that is representative for the study area. For this purpose, a

RS-based sampling approach was developed (Chapter 3) where data can be collected in a cost and time efficient manner.

The X-ray analysis was the most time and cost demanding analysis of the method. An alternative, unexplored option to obtain a set of present mineralogy is running MICA on high spectral resolution RS data. This set of minerals can be used to train the model for the quantification of mineral abundances from the PS data. In this thesis, the feasibility for mineral characterization was already demonstrated in Chapter 4 to 6. Future research using these methods will prove its efficiency in collecting soil information for large areas.

Currently, the method presented in Chapter 5 allows quantifying mineral abundances up to four minerals while it is known that soils contain a high variety of minerals (Hochella, 2002). The value of knowing the main minerals, however, provides essential information about soils (Egli et al., 2003; Mavris et al., 2011). Overall, the natural samples used in this study consisted of two main minerals with an additional two minerals having minor abundances (secondary minerals), but the remaining minerals were subordinate in abundances. The quantification of the main minerals without knowledge of subordinate minerals provides information on e.g. soil formation or parent material. If this information would be limited to two minerals, without supporting secondary minerals, the information content may be too limited to draw conclusions. It remains a scientific challenge to quantify the full set of minerals using PS. Also, modelling of subordinate minerals is not solved, yet this is of importance for mining of rare earth minerals (Hochella, 2002).

The soil spectroscopy community has not yet explored spectral deconvolution for assessing soil properties using PS other than mineralogy and soil moisture (Whiting et al., 2004). Various methods for estimating properties of the topsoil using PS were found to be sufficient accurate compared to chemical soil analysis. The remaining inaccuracies in estimated soil properties of the topsoil, such as soil organic matter, have often been contributed to other constituents in soil samples (Bartholomeus et al., 2008). This implies actually, that the inaccuracies are a consequence of overlapping absorption features, which were not accounted for by the employed inference methods. Although it has not yet been investigated, the prediction of various soil properties may be further improved by using the method presented in Chapter 5.

### 7.2.2 Mapping soil properties at regional scale using PS, RS and geostatistics

This thesis has shown the added value of advanced PS, RS and geostatistical methods to obtain soil information. At the same time, Chapter 2 listed various studies that were successful in mapping soil properties at regional scale using RS data only. One of the best examples includes the ASTER Geoscience product (Cudahy, 2012) and the mineral maps that were made for Australia (Lau et al., 2012). The maps presented in this thesis differ in some ways compared to these. The main differences are the

higher spatial resolution of the ASTER Geoscience products and the more qualitative nature of the mineral estimates. The higher spatial resolution is favoured by soil scientists. However, as was discussed in Chapter 6, a high spatial resolution does not necessarily improve the accuracy of predicted soil properties when modelled at regional scale and the added value of high-resolution data should be reconsidered for regional-scale models.

One strong point of our method is the flexibility to model other soil properties because of the underlying basic principle, the Soil-Landscape paradigm (Jenny, 1941). DSM uses this paradigm to develop empirical models that relate observations of various soil properties with environmental variables that describe the main soil forming factors (Chapter 1 and 2). Another strong point of our method is the ability to interpolate gaps in the spatial coverage of employed auxiliary data. Using RS in areas having a climate other than arid or semi-arid will always require interpolation of masked out areas either due to clouds, high vegetation cover or other non-climatic related features. Apart from that, many soil properties are difficult to quantify using RS data due to the spectral resolution and the position of the sensors' bands. As mentioned in Chapter 1, modelling soil properties in a regional or global-scale context benefits more from quantification of soil properties, compared to qualitative soil information. Indeed, this thesis has shown the need to use a representative sample and laboratory measurements for PS analysis. Nevertheless, using this data in combination with RS and geostatistics allows to spatially explicit quantify soil properties at regional scale providing data at a resolution fitting that scale.

A good example that could benefit from the presented methods in this thesis is global soil carbon (SOC) modelling (Minasny et al., 2013). Envisioning a method to achieve this, a combination of PS, RS and geostatistics is deemed necessary. Mapping of carbon pools is complex and mapping the spatial and temporal processes using RS-based ratios, such as used for the ASTER geoscience products, does not suffice. Key to global SOC mapping is quantifying both the slow and fast carbon pools (Knorr et al., 2005). The slow carbon pools are below ground and are difficult to model using solely RS data, if not impossible. The slow and fast carbon pools could be quantified based on legacy, in-situ, and observational (RS) data. That is, the below ground carbon information requires on-site data-collection where existing soil legacy data may be exploited for sampling or spatial modelling at a later stage. The collected data may be analysed using PS, chemical analysis or a combination of the two techniques. Spatially explicit SOC maps can be derived using this set of information, RS data and derived products as exhaustive predictor variables and spatial and temporal statistics. The methods presented in this thesis can contribute to global mapping of SOC pools (Croft et al., 2012) with respect to soil data collection and spatial modelling of SOC. Nevertheless, more research is needed on the temporal component (Croft et al., 2012).

### 7.2.3 Modelling global environmental changes

The recognition that RS, PS and geostatistics are key to improvement of regional-scale modelling of the environment increases the feasibility to provide soil and other environmental information at large scales. Delivering up-to-date accurate spatial soil information is a major advancement with respect to the information that is currently being used in environmental-change models. These models need this information of the pedosphere for model parameterization and calibration, in addition to the information of the other Earth's spheres. Also, at the moment they do not incorporate supporting soil data for the prediction of future changes in e.g. vegetation shifts or changes in soil resources (Grunwald, 2011; Varella et al., 2010). This introduces large uncertainties for modelling e.g. future food security.

Two major issues are at the core of this problem. First, there is a lack of available soil information at regional and global scale (Chapter 1). Second, modelling future changes requires an understanding of changes in the past. At the moment, analysis methods have not been fully developed to advance this understanding with respect to changes in soil resources. The soil science community is aware of these shortcomings and current efforts are on data harmonization (Panagos et al., 2011; Sulaeman et al., 2012) while research efforts are initiated for temporal modelling of soil properties (Banwart, 2011). Despite these initiatives, it is expected that the existing soil data have insufficient coverage and thematic variability for regional and global models. The time and cost associated with collecting sufficient data are comprehensive. Therefore, it is important to develop new methods, for the benefit of various research disciplines focussing on modelling environmental changes, climate change adaptation, food security and soil services. This thesis contributed to these essential developments.

### 7.2.4 Outlook

As previously discussed, an interesting and contemporary research topic, is global soil carbon mapping. Many papers have described the importance of this research; soil carbon is one of the most important soil services and is needed in climate change adaptation and food security research (Bond-Lamberty and Thomson, 2010; Lal, 2004). However, for future research in soil science, it is more important to develop methods that allow modelling a wider set of soil properties. Regional and global-scale environmental models do have need for various soil parameters and therefore methods are required to be flexible in delivering soil information. Considering the need and the content of such soil information, spatiotemporal modelling is the future of soil mapping (Heuvelink and Webster, 2001; Katzfuss and Cressie, 2012). However, to actually develop a spatiotemporal model, many advances have to be made, especially in the temporal domain. Again, RS and PS with geostatistics will play the key role. Spatiotemporal modelling requires the presence of a monitoring

system (Jonckheere, 2012); the integration of different sensor and measurement techniques would allow for a robust monitoring system. This is an efficient alternative, because spectral measurements are less expensive to collect and analyse compared to traditional techniques (Croft et al., 2012). Integration of laboratory, field and airborne or satellite platform sensors will allow to monitor changes in soil, vegetation and their feedbacks (Milcu et al., 2012) over various spatial and temporal scales. Such data, in combination with soil profile data, is prerequisite for accurate spatiotemporal modelling of soil resources. Other priorities that deserve attention in future research are listed below.

Spectroscopy for quantifying soil properties

RS and PS methods bear potential to be further exploited. Major contributions include the development of more quantitative approaches, such as proposed in this thesis, and the integration of various RS and PS sensors. Remote sensors which were not included in this thesis form a valuable input as well. These include, gamma-ray spectroscopy (Wilford et al., 1997), radar (Merlin et al., 2013) or electric conductivity (Lambot et al., 2004). These techniques provide subsurface information and are less sensitive to vegetation cover, while imaging spectroscopy is limited to the surface. These remote sensors do not yet deliver data on the global scale but it may contribute substantially to a system for inventorying and monitoring the earth's soil resources over various spatial and temporal scales.

Geostatistics for large datasets

Another remaining challenge for large-scale DSM, is the use of voluminous datasets (Katzfuss, 2011; Katzfuss, 2013). Up to today, little research has been done by the DSM community to advance methods with geostatistical approaches that can deal with large datasets. Accurate spatial and spatiotemporal modelling approaches that allow transferability to other areas rely on geostatistical approaches and exhaustive datasets. Chapter 6 suggested some alternatives that enhance the perspectives of large-scale DSM.

Finally, in our quest to develop such methods, we need to combine legacy, in-situ, and observational data and modelling to allow better prediction of soil properties. This will enable us in the near future to deliver more accurate and comprehensive information about soils, soil resources and ecosystem services provided by soils at regional and, ultimately, global scale.

# References


Abrams, H., Hook, S., 2001. ASTER User Handboek. Jet Propulsion Laboratory Pasadena, CA, USA.

Abrams, M., Hook, S.J., 1995. Simulated ASTER data for geologic studies. IEEE Transactions on Geoscience and Remote Sensing 33(3), 692-699.

Adamchuk, V.I., Viscarra Rossel, R.A., Marx, D.B., Samal, A.K., 2011. Using targeted sampling to process multivariate soil sensing data. Geoderma 163(1-2), 63-73.

Agriculture and Agri-Food Canada., 2010. The Canadian Soil Informayion System (CANSIS) and the National Soil Databse (NSDB).

Aly, Z., Bonn, F.J., Magagi, R., 2007. Analysis of the Backscattering Coefficient of Salt-Affected Soils Using Modeling and RADARSAT-1 SAR Data. Geoscience and Remote Sensing, IEEE Transactions on 45(2), 332-341.

Anderson, M.C., Norman, J.M., Kustas, W.P., Houborg, R., Starks, P.J., Agam, N., 2008. A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. Remote Sensing of Environment 112(12), 4227-4241.

Apan, A., Kelly, R., Jensen, T., Butler, D., Strong, W., Basnet, B., 2002. Spectral discrimination and separability analysis of agricultural crops and soil attributes using ASTER imagery., 11th ARSPC, Brisbane, Australia.

Ardö, J., Pilesjö, P., Skidmore, A., 1997. Neural networks, multitemporal landsat thematic mapper data and topographic data to classify forest damages in the Czech republic. Canadian Journal of Remote Sensing 23(3), 217-229.

Armitage, R.P., Kent, M., Weaver, R.E., 2004. Identification of the spectral characteristics of British semi-natural upland vegetation using direct ordination: a case study from Dartmoor, UK. International Journal of Remote Sensing 25(17), 3369 - 3388.

Asner, G.P., Heidebrecht, K.B., 2003. Imaging spectroscopy for desertification studies: comparing AVIRIS and EO-1 Hyperion in Argentina drylands. IEEE Transactions on Geoscience and Remote Sensing 41(6), 1283-1296.

Association of African Geological Surveys, 1963. Geological map of Africa (1:5M). ASGA-UNESCO, Paris.

Aubert, A., Betremieux, R., Bonfils, P., Bonneua, M., Boulaine, J., Dejou, J., Delmas, J., Drouineau, G., Duchaufour, P., Dupuis, J., Dutil, P., Flon, H., Fournier, F., Gelpe, J., Geze, B., Hevert, J., Henin, S., Horemans, M., Jacquin, F., Jamagne, M., Maignien, R., Meriaux, S., Pedro, G., Portier, J., Segalen, R., Servat, E., Vigneron, J., 1967. Classification des Sols. INRA.

Australian Department of Agriculture Fisheries and Forestry, Australian Soil resource Information System (ASRIS).

Australian Government, 2010. Caring for our country - Monitoring, evaluation, reoring and improvement strategy. Commonwealth of Australia.

Baize, D., Jabiol, B., 1995. Guide pour la descirption des sols, INRA edition, INRA, Paris.

Baldridge, A.M., Hook, S.J., Grove, C.I., Rivera, G., 2008. The ASTER Spectral Library Version 2.0, Jet Propulsion Laboratory.

Ballantine, J.A.C., Okin, G.S., Prentiss, D.E., Roberts, D.A., 2005. Mapping North African landforms using continental scale unmixing of MODIS imagery. Remote Sensing of Environment 97(4), 470-483.

Bannari, A., Guedon, A.M., El-Harti, A., Cherkaoui, F.Z., El-Ghmari, A., 2008. Characterization of Slightly and Moderately Saline and Sodic Soils in Irrigated Agricultural Land using Simulated Data of

Advanced Land Imaging (EO-1) Sensor. Communications in Soil Science and Plant Analysis 39(19), 2795-2811.

Banwart, S., 2011. Save our soils. Nature. 474(7350), 151-152.

Baret, F., Guyot, G., Major, D.J., 1989. TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation, Geoscience and Remote Sensing Symposium, 1989. IGARSS'89/ 12th International Canadian Symposium on Remote Sensing., 1989. IEEE, New York, Vancouver, pp. 1355-1358.

Baret, F., Jacquemoud, S., Hanocq, J.F., 1993. About the soil line concept in remote sensing. Advances in Space Research 13(5), 281-284.

Barnes, E.M., Baker, M.G., 2000. Multispectral data for mapping soil texture: Possibilities and limitations. Applied Engineering in Agriculture 16(6), 731-741.

Barnes, E.M., Sudduth, K.A., Hummel, J.W., Lesch, S.M., Corwin, D.L., Yang, C., Daughtry, C.S.T., Bausch, W.C., 2003. Remote- and ground-based sensor techniques to map soil properties. Photogrammetric Engineering and Remote Sensing 69(6), 619-630.

Bartholomeus, H., 2009. The influence of vegetation on the spectroscopic estimation of soil properties. Ph.D. Thesis, Wageningen UR, Wageningen, The Netherlands, 1-143 pp.

Bartholomeus, H., Epema, G., Schaepman, M.E., 2007. Determining iron content in Mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy. International Journal of Applied Earth Observation and Geoinformation 9(2), 194-203.

Bartholomeus, H.M., Schaepman, M.E., Kooistra, L., Stevens, A., Hoogmoed, W.B., Spaargaren, O.S.P., 2008. Spectral reflectance based indices for soil organic carbon quanitification. Geoderma 145(9).

Bastiaanssen, W.G.M., Noordman, E.J.M., Pelgrum, H., Davids, G., Thoreson, B.P., Allen, R.G., 2005. SEBAL Model with Remotely Sensed Data to Improve Water-Resources Management under Actual Field Conditions. Journal of Irrigation and Drainage Engineering 131(1), 85-93.

Battrick, B., 2005. GEOSS - 10-Year Implementation Plan, ESA Publications Division.

Bechtel, R., Rivard, B., Sánchez-Azofeifa, A., 2002. Spectral properties of foliose and crustose lichens based on laboratory experiments. Remote Sensing of Environment 82(2-3), 389-396.

Beck, R., 2003. EO-1 User Guide. USGS Earth Resources Observation Systems Data Center, Cincinatti, Ohio.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. Journal of Plant Nutrition and Soil Science 168(1), 21-33.

Bell, D., Menges, C., Ahmad, W., van Zyl, J.J., 2001. The Application of Dielectric Retrieval Algorithms for Mapping Soil Salinity in a Tropical Coastal Environment Using Airborne Polarimetric SAR. Remote Sensing of Environment 75(3), 375-384.

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using Imaging Spectroscopy to study soil properties. Remote sensing of environment 113(Supplement 1), S38-S55.

Ben-Dor, E., Patkin. K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data - a case study over clayey soil in Israel. International Journal of Remote Sensing 23(6), 20.

Ben-Dor, E., Taylor, R.G., Hill, J., Demattê, J.A.M., Whiting, M.L., Chabrillat, S., Sommer, S., Sparks, D.L., 2008. Imaging spectrometry for soil applications. Advances in Agronomy 97, 321-392.

Bergmann, J., Friedel, P., Kleeberg, R.,, 1998. BGMN - a new fundamental parameters based Rietveld program for laboratory X-ray

sources, it's use in quantitative analysis and structure investigations, CPD Newsletter. Commission of Powder Diffraction, International Union of Crystallography, pp. 5-8.

Bierwirth, P.N., 1990. Mineral mapping and vegetation removal via data-calibrated pixel unmixing, using multispectral images. International Journal of Remote Sensing 11(11), 1999-2017.

Bish, D.L., Plötze, M., 2011. X-ray Powder Diffraction with Emphasis on Qualitative and Quantitative Analysis in Industrial Mineralogy. In: G. Christidis (Ed.), Industrial Mineralogy, EMU Notes in Mineralogy, pp. 35-76.

Bishop, J.L., Gates, W.P., Makarewicz, H.D., McKeown, N.K., Hiroi, T., 2011. Reflectance spectroscopy of beidellites and their importance for Mars. Clays and Clay Minerals 59(4), 378-399.

Boardman, J.W., 1994. Geometric mixture analysis of imaging spectrometry data, Geoscience and Remote Sensing Symposium, Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation (2364),. 2369-2371.

Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S., 2010. Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Progress in Soil Science; Volume 2. Springer Science+Business Media B.V., Dordrecht.

Bond-Lamberty, B., Thomson, A., 2010. Temperature-associated increases in the global soil respiration record. Nature 464(7288), 579-582.

Bouma, J., 2009. Soils are back on the global agenda: Now what? Geoderma 150(1–2), 224-225.

Bouma, J., Broll, G., Crane, T.A., Dewitte, O., Gardi, C., Schulte, R.P., Towers, W., 2012. Soil information in support of policy making and awareness raising. Current Opinion in Environmental Sustainability.

Bregt, A., 1992. Processing of soil survey data. Ph.D. Thesis, Wageningen University, Wageningen, The Netherlands.

Breiman, L., Friedman, J.H., Olshen, R.A., Sonte, C.J., 1984. Classification and regression trees. Wadsworth International Group, Belmont, California.

Brennan, R., Webster, T.L., 2006. Object-oriented land cover classification of lidar-derived surfaces. Canadian journal of remote sensing 32(2), 162-172.

Breunig, F.M., Galvão, L.S., Formaggio, A.R., 2008. Detection of sandy soil surfaces using ASTER-derived reflectance, emissivity and elevation data: potential for the identification of land degradation. International Journal of Remote Sensing 29(6), 1833 - 1840.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132(3-4), 273-290.

Bruker AXS, 2012. DIFFRACplus Sotware Solutions for X-ray Powder Diffraction.

Brus, D.J., 1994. Improving design-based estimation of spatial means by soil map stratification. A case study of phosphate saturation. Geoderma 62(1-3), 233-246.

Brus, D.J., 2000. Using regression models in design-based estimation of spatial means of soil properties. European Journal of Soil Science 51(1), 159-172.

Brus, D.J., de Gruijter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80(1-2), 1-44.

Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138(1-2), 86-95.

Bui, E.N., Loughhead, A., Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. Australian Journal of Soil Research 37(3), 495-508.

Buis, E., Veldkamp, A., Boeken, B., van Breemen, N., 2009. Controls on plant functional surface cover types along a precipitation gradient in the Negev Desert of Israel. Journal of Arid Environments 73(1), 82-90.

Burns, R.G., 1993. Mineralogical Applications of Crystal Field Theory 2ed. Camebridge University Press, Cambridge UK.

Burrough, P.A., 1983. Multiscale sources of spatial variation in soil. I. The application of fractal concepts to nested levels of soil variation. Journal of Soil Science 34(3), 577-597.

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., 2013. A methodology for digital soil mapping in poorly-accessible areas. Geoderma 192(0), 341-353.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond DSM. Geoderma 142(1-2), 69-79.

Cartography by Navteq Inc., 2011. Map of Morocco - Bing Maps Web Mapping Service. Microsoft.

Castrignanò, A., Giugliarini, L., Risaliti, R., Martinelli, N., 2000. Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics. Geoderma 97(1-2), 39-60.

Cavallar, W., 1950. Soils du Maroc (1: 1.5M). l'Annexe du Maroc de l'Institut Geographique National, Rabat, Morocco.

Chabrillat, S., Goetz, A.F.H., Krosley, L., Olsen, H.W., 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. Remote Sensing of Environment 82(2-3), 431-445.

Chang, C.W., Laird,D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science 167(2), 110-116.

Chudnovsky, A., Ben-Dor, E., Kostinski, A.B., Koren, I., 2009. Mineral content analysis of atmospheric dust using hyperspectral information from space. Geophysical Research Letters 36(15), L15811.

Chudnovsky, A., Kostinski, A., Herrmann, L., Koren, I., Nutesku, G., Ben-Dor, E., 2011. Hyperspectral spaceborne imaging of dust-laden flows: Anatomy of Saharan dust storm from the Bodélé Depression. Remote Sensing of Environment 115(4), 1013-1024.

Clark, R.N., 1998. Material absorption band depth mapping of imaging spectrometer data using a complete band shape least-squares fit with library reference spectra. U.S. Geological Survey, Denver, pp. 11.

Clark, R.N., 1999. Chapter 1: Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: A.N. Rencz (Ed.), Manual of Remote Sensing. John Wiley and Sons, New York, pp. 3-58.

Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High Spectral Resolution Reflectance Spectroscopy of Minerals. Journal of Geophysical Research. 95(B8), 12653-12680.

Clark, R.N., Roush, T.L., 1984. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. Journal of Geophysical Research 89(B7), 6329-6340.

Clark, R.N., Swayze, G.A., Livo, K.E., Kokaly, R.F., Sutley, S.J., Dalton, J.B., McDougal, R.R., Gent, C.A., 2003. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. Journal of Geophysical Research 108(5), 44.

Clark, R.N., Swayze, G.A., Wise, R., Livo, E., Hoefen, T., Kokaly, R., Sutley, S.J., 2007. USGS digital spectral library splib06a. U.S. Geological Survey.

Clénet, H., Pinet, P., Daydou, Y., Heuripeau, F., Rosemberg, C., Baratoux, D., Chevrel, S., 2011. A new systematic approach using the Modified Gaussian Model: Insight for the characterization of chemical composition of olivines, pyroxenes and olivine–pyroxene mixtures. Icarus 213(1), 404-422.

Cloutis, E.A., Edward, A., 1989. Spectral reflectance properties of lichens: remote sensing discrimination of lichen and rock, International Geoscience and Remote Sensing Symposium. Digest, pp. 2506.

Cochran, W.G., 1977. Sampling techniques. 3th ed. John Wiley & Sons, New York.

Commission of the European Communities, 2006. Communication from the commission to the council, the European Parliament, the European economic and social committee and the committee of the regions - Thematic Strategy for Soil Protection. COM(2006)231 final, Brussels.

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. Journal of the Royal Statistical Society. Series B: Statistical Methodology 70(1), 209-226.

Cressie, N.A.C., 1991. Statistics for spatial data. John Wiley & Sons, Inc., Iowa.

Croft, H., Kuhn, N.J., Anderson, K., 2012. On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. Catena 94, 64-74.

Crow, W.T., Kustas, W.P., Prueger, J.H., 2008. Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model. Remote Sensing of Environment 112(4), 1268-1281.

Cudahy, T., 2012. Satellite ASTER Geoscience Product Notes for Australia, Version 1. Report no. EP-30-07-12-44, CSIRO, Australia..

Cudahy, T.J., Hewson, R., Huntington, J.F., Quigley, M.A., Barry, P.S., 2001. The performance of the satellite-borne Hyperion hyperspectral VNIR-SWIR imaging system for mineral mapping at Mount Fitton, South Australia., International Geoscience and Remote Sensing Symposium IGARSS '01. Proceedings. 2001 IEEE International Sydney, Australia, pp. 314-316.

Dalton, J.B., Bove, D.J., Mladinich, C.S., Rockwell, B.W., 2004. Identification of spectrally similar materials using the USGS Tetracorder algorithm: the calcite–epidote–chlorite problem. Remote Sensing of Environment 89(4), 455-466.

Daughtry, C.S.T., Hunt, E.R., Jr., Doraiswamy, P.C., McMurtrey III, J.E., 2005. Remote sensing the spatial distribution of crop residues. Agronomy Journal 97(3), 864-871.

Daughtry, C.S.T., Hunt Jr, E.R., 2008. Mitigating the effects of soil and residue water contents on remotely sensed estimates of crop residue cover. Remote Sensing of Environment 112(4), 1647-1657.

Davis, J.C., 2002. Statistics and data analysis in geology. Wiley, New York.

De'Ath, G., Fabricius, K.E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81(11), 3178-3192.

de Bruin, S., Wielemaker, W.G., Molenaar, M., 1999. Formalisation of soil-landscape knowledge through interactive hierarchical disaggregation. Geoderma 91(1-2), 151-172.

de Gruijter, J., Bierkens, M., Brus, D.J., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Earth and Environmental Science. Springer, Berlin.

De Jong, R., Schaepman, M.E., Furrer, R., De Bruin, S., Verburg, P.H., 2013. Spatial relationship between climatologies and changes in global vegetation activity. Global Change Biology 19(6), 1953-1965.

Debella-Gilo, M., Etzelmüller, B., 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. Catena 77(1), 8-18.

Dehaan, R., Taylor, G.R., 2003. Image-derived spectral endmembers as indicators of salinisation. International Journal of Remote Sensing 24(4), 775 - 794.

Dehaan, R.L., Taylor, G.R., 2001. Mapping irrigation-induced salinity with hyperspectral imagery, International Geoscience and Remote Sensing Symposium IGARSS '01.

Proceedings. 2001 IEEE International pp. 293-295

Dehn., M., Gartner, H., Dikau, R., 2001. Principles of semantic modeling of landform structures. Computer & Geosciences 27(8), 1005-1010.

Deller, M.E., 2006. Facies discrimination in laterites using Landsat Thematic Mapper, ASTER and ALI data- examples from Eritrea and Arabia. International Journal of Remote Sensing 27(12), 22.

Demattê, J.A.M., 2002. Characterization and discrimination of soils by their reflected electromagnetic energy. Balanço social, pesquisa agropecuária brasileira 37(10), 1445-1458.

Demattê, J.A.M., Galdos, M.V., Guimarães, R.V., Genú, A.M., Nanni, M.R., Zullo, J., 2007. Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. International Journal of Remote Sensing 28(17), 3813-3829.

Demattê, J.A.M., Garcia, G.J., 1999. Alteration of soil properties through a weathering sequence as evaluated by spectral reflectance. Soil Science Society of America Journal 63(2), 327-342.

Demattê, J.A.M., Sousa, A.A., Alves, M.C., Nanni, M.R., Fiorio, P.R., Campos, R.C., 2006. Determining soil water status and other soil characteristics by spectral proximal sensing. Geoderma 135, 179-195.

Dennison, P.E., Roberts, D.A., 2003. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sensing of Environment 87(2–3), 123-135.

Dent, D., Young, A., 1981. Soil Survey and Land Evaluation. George Allen & Unwin, London.

Dewitte, O., Jones, A., Elbelrhiti, H., Horion, S., Montanarella, L., 2012. Satellite remote sensing for soil mapping in Africa: An overview. Progress in physical geography 36(4), 514-538.

Didier, R., Bivand, R.S., Pebesma, E.J., Virgilio Gomez-Rubio., V., 2011. Applied Spatial Data Analysis with R. Mathematical Geosciences 43(5), 607-609.

Diekmann, M., 2003. Species indicator values as an important tool in applied plant ecology - a review. Basic and Applied Ecology 4(6), 493-506.

Dijkshoorn, J.A., 2003. SOTER database for Southern Africa (SOTERSAF), ISRIC - World Soil Information, Wageningen.

Dijkshoorn, J.A., Huting, J.R.M., Tempel, P., 2005. Update of the 1:5 million Soil and Terrain Database for Latin America and the Caribbean (SOTERLAC; version 2.0), ISRIC - World Soil Information, Wageningen.

Dikau, R., 1989. The Application of a Digital Relief Model to Landform Analysis in Geomorphology. In: J.F. Raper (Ed.), Three Dimensional Applications in Geographical Information Systems. Taylor & Francis, London, pp. 51-77.

Dobos, E., Daroussin, J., Montanarella, L., 2005. An SRTM-based procedure to delineate SOTER Terrain Units on 1:1 and 1:5 million scales, EUR 21571 EN. Office for Official Publications of the European Communities, Luxembourg, pp. 55.

Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L., Helt, T., 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma 97(3-4), 367-391.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International journal of applied earth observation and geoinformation 9(2), 165-193.

Egli, M., Mirabella, A., Fitze, P., 2003. Formation rates of smectites derived from two Holocene chronosequences in the Swiss Alps. Geoderma 117(1–2), 81-98.

Egli, M., Nater, M., Mirabella, A., Raimondi, S., Plötze, M., Alioth, L., 2008. Clay minerals, oxyhydroxide formation, element leaching

and humus development in volcanic soils. Geoderma 143(1-2), 101-114.

Ehsani, A.H., Quiel, F., 2008. Geomorphometric feature analysis using morphometric parameterization and artificial neural networks. Geomorphology 99(1-4), 1-12.

Ehsani, A.H., Quiel, F., 2009. A semi-automatic method for analysis of landscape elements using Shuttle Radar Topography Mission and Landsat ETM+ data. Computers & Geosciences 35(2), 373-389.

Ellenberg, H., 1988. Vegetation ecology of Central Europe. Cambridge University Press, Cambridge.

Escadafal, R., 1993. Remote sensing of soil color: principles and applications. Remote Sensing Reviews 7(3-4), 261-279.

European Space Agency GlobCover Project, 2008. GlobCover Land Cover v2 2008 database. In: MEDIAS-France (Ed.).

FAO, 2007. Digital Soil Map of the World. In: F. Nachtergaele (Ed.). Food and Agriculture Organization of the United Nations, Rome, Italy.

FAO, IIASA, ISRIC, ISS-CAS, JRC, 2008. Harmonized World Soil Database (version 1.0). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Farifteh, J., Farshad, A., George, R.J., 2006. Assessing salt-affected soils using remote sensing, solute modelling, and geophysics. Geoderma 130(3-4), 191-206.

Farifteh, J., van der Meer, F., van der Meijde, M., Atzberger, C., 2008. Spectral characteristics of salt-affected soils: A laboratory experiment. Geoderma 145(3-4), 196-206.

Farr, T.G., 2000. The shuttle radar topography mission. IEEE Aerospace Conference Proceedings (1), pp. 63.

Farrand, W.H., Harsanyi, J.C., 1997. Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique. Remote Sensing of Environment 59(1), 64-76.

Felicísimo, A.M., Cuartero, A., Remondo, J., Quirós, E., 2012. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides, 1-15.

Finke, P., Hartwich, R., Dudal, R., Ibanez, J., Jamagne, M., King, D., Montanarella, L., Yassoglou, N., 2001. Georeferenced soil database for europe: manual of procedures (version 1.1). EUR 18092 en, Office for Official Publications of the European Communities, Italy.

Flahaut, J., Quantin, C., Clenet, H., Allemand, P., Mustard, J.F., Thomas, P., 2012. Pristine Noachian crust and key geologic transitions in the lower walls of Valles Marineris: Insights into early igneous processes on Mars. Icarus 221(1), 420-435.

Florinsky, I.V., 1998. Combined analysis of digital terrain models and remotely sensed data in landscape investigations. Progress in Physical Geography 22(1), 28.

Forde, S.E., Beardmore, R.E., Gudelj, I., Arkin, S.S., Thompson, J.N., Laurence D. Hurst, L.D., 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 11(455), 220-224.

French, A.N., Jacob, F., Anderson, M.C., Kustas, W.P., Timmermans, W., Gieske, A., Su, Z., Su, H., McCabe, M.F., Li, F., Prueger, J., Brunsell, N., 2005. Surface energy fluxes with the Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER) at the Iowa 2002 SMACEX site (USA). Remote Sensing of Environment 99(1-2), 55-65.

Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment 61(3), 399-409.

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C., 2002. Global land cover mapping from MODIS: Algorithms and early results. Remote Sensing of Environment 83(1-2), 287-302.

Furrer, R., Genton, M.G., 2011. Aggregation-cokriging for highly multivariate spatial data. Biometrika 98(3), 615-631.

Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance Tapering for Interpolation of Large Spatial Datasets. Journal of Computational and Graphical Statistics 15(3), 502-523.

Gail, P.A., Jinxue, W., Michael, L.H., Kneizys, F.X., James H. Chetwynd, Jr., Laurence, S.R., Kimball, L.M., Robert, A.M., Eric, P.S., Shepard, William, O.G., Leonard, W.A., John, E.A.S., 1994. History of one family of atmospheric radiative transfer codes. In: K.L. David (Ed.). SPIE, pp. 170-183.

Gallant, J.C., Read, A., 2009. Enhancing the SRTM Dataa for Australia. In: R. Purves, S. Gruber, R. Straumann (Eds.), Geomorphometry 2009 Conference Proceedings, Zurich, Switzerland.

Galvão, L.S., Formaggio, A.R., Couto, E.D., Roberts, D.A., 2008. Relationships between the mineralogical and chemical composition of tropical soils and topography from hyperspectral remote sensing data. ISPRS Journal of Photogrammetry and Remote Sensing 63(2), 259-271.

Galvão, L.S., Pizarro, M.A., Epiphanio, J.C.N., 2001. Variations in reflectance of tropical soils: Spectral-chemical composition relationships from AVIRIS data. Remote Sensing of Environment 75(2), 245-255.

Garbow, B.S., Hillstrom, K.E., More, J.J., 1980. MINPACK project. http://www.netlib.org/minpack/.

Garcia-Haro, F.J., 1997. Modelizacio´n y estimacio´n de para´metros relacionados con lacubierta vegetal en teledeteccio´n., University of Valencia, Valencia.

García-Haro, F.J., Sommer, S., Kemper, T., 2005. A new tool for variable multiple endmember spectral mixture analysis (VMESMA). International Journal of Remote Sensing 26(10), 2135-2162.

Genú, A.M., Demattê, J.A.M., 2006. Determination of soil attribute contents by means of reflected electromagnetic energy. International Journal of Remote Sensing 27(21), 4807-4818.

Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modelling and spatial prediction of soil attributes. International Journal of Geographical Information Science 9(4), 421 - 432.

Glasser, N.F., Jansson, K.N., Harrison, S., Kleman, J., 2008. The glacial geomorphology and Pleistocene history of South America between 38°S and 56°S. Quaternary Science Reviews 27(3-4), 365-390.

Global Soil Parnership, 2011. FAO, Rome, Italy.

Gomez, C., Lagacherie, P., Coulouma, G., 2008a. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. Geoderma 148(2), 141-148.

Gomez, C., Lagacherie, P., Coulouma, G., 2012. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. Geoderma 189-190, 176-185.

Gomez, C., Viscarra Rossel, R.A., McBratney, A.B., 2008b. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma 146(3-4), 403-411.

Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89(1-2), 1-45.

Goovearts, P., 1997. Geostatistics for Natural Resource Evaluation. Oxford University Press, New York.

Grandjean, G., DIGISOIL Team, 2011. Scientific synthesis of the DIGISOIL project. Report N° FP7-DIGISOIL-D5.1 pp. 54.

Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M., Olah, M.R., Williams, O., 1998. Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Remote sensing of environment 65(3), 227-248.

Green, R.O., Pavri, B.E., Chrien, T.G., 2003. On-orbit radiometric and spectral calibration characteristics of EO-1 Hyperion derived with an underflight of AVIRIS and in situ measurements at Salar de Arizaro, Argentina. Geoscience and Remote Sensing, IEEE Transactions on 41(6), 1194-1203.

Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152(3-4), 195-207.

Grunwald, S., 2010. The current state of digital soil mapping and what is next. In: J. Boetinger, D.W. Howell, A.C. Moore, A.E. Hartemink, S. Kienst-Brown (Eds.), Digital soil mapping: Bridging research, production and environmental applications. Springer, Heidelberg, pp. 3-12.

Grunwald, S., 2011. Digital soil mapping and modeling at continental scales: Finding solutions for global issues. Soil Science Society of America Journal 75(4), 1201-1213.

Hahn, C., Gloaguen, R., 2008. Estimation of soil types by non linear analysis of remote sensing data. Nonlinear Processes in Geophysics 15(1), 115-126.

Hansen, M.K., Brown, D.J., Dennison, P.E., Graves, S.A., Bricklemyer, R.S., 2009. Inductively mapping expert-derived soil-landscape units within dambo wetland catenae using multispectral and topographic data. Geoderma 150(1-2), 72-84.

Hapke, B., 1981. Bidirectional Reflectance Spectroscopy 1. Theory. Journal of Geophysical Research 86(B4), 3039-3054.

Hapke, B., 2002. Bidirectional reflectance spectroscopy. 5. The coherent backscatter opposition effect and anisotropic scattering. Icarus 157(2), 523-534.

Harris, P., Brunsdon, C., Fotheringham, A.S., 2011. Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. Stochastic Environmental Research and Risk Assessment 25(2), 123-138.

Hartemink, A.E., 2008. Soils are back on the global agenda. Soil Use and Management 24(4), 327-330.

Hartemink, A.E., McBratney, A., 2008. A soil science renaissance. Geoderma 148(2), 123-129.

Hengl, T., 2009. Chapter 3: Hands-on software, A practical guide to geostatistical mapping. Luxembourg: Office for Official Publications of the European Communit, Ispra, Italy, pp. 165.

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007a. About regression-kriging: From equations to case studies. Computers & Geosciences 33(10), 1301-1315.

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120(1-2), 75-93.

Hengl, T., Reuter, H.I., 2009. Geomorphometry: concepts, software, applications. Developments in soil science (33). Elsevier, Amsterdam, The Netherlands.

Hengl, T., Rossiter, D., Stein, A., 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. Australian Journal of Soil Research 41(8), 1403-1422.

Hengl, T., Toomanian, N., Reuter, H.I., Malakouti, M.J., 2007b. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. Geoderma 140(4), 417-427.

Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present, and future. Geoderma 100(3-4), 269-301.

Hewitt, A.E., 1993. Predictive modelling in soil survey. Soils and Fertilizers 56(3), 10.

Hewson, R.D., Cudahy, T.J., Jones, M., Thomas, M., 2012. Investigations into soil composition and texture using infrared spectroscopy (2-14 m). Applied and Environmental Soil Science 2012.

Hewson, R.D., Cudahy, T.J., Mizuhiko, S., Ueda, K., Mauger, A.J., 2005. Seamless geological map generation using ASTER in the Broken

Hill-Curnamona province of Australia. Remote Sensing of Environment 99(1-2), 159-172.

Hijmans, R.J., van Etten, J., 2012. Geographic analysis and modeling with raster data. R Package version 2.1-25.

Hill, M.O., Roy, D.B., Mountford, J.O., Bunce, R.G.H., 2000. Extending Ellenberg's Indicator Values to a New Area: An Algorithmic Approach. Journal of Applied Ecology 37(1), 3-15.

Hochella, M.F., 2002. Sustaining Earth: Thoughts on the present and future roles of mineralogy in environmental science. Mineralogical Magazine 66(5), 627-652.

Hodgson, M.E., Jensen, J.R., Schmidt, L., Schill, S., Davis, B., 2003. An evaluation of LIDAR- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs. Remote sensing of environment 84(2), 295-308.

Hofton, M., Dubayah, R.C., Blair, J.B., Rabine, D., 2006. Validation of SRTM elevations over vegetated and non-vegetated terrain using medium footprint lidar. Photogrammetric Engineering and Remote Sensing 72(3), 7.

Hong, S.Y., Minasny, B., Han, K.H., Kim, Y., Lee, K., 2013. Predicting and mapping soil available water capacity in Korea. PeerJ1, e71.

Hu, T., 2007. Nonlinear control design for linear differential inclusions via convex hull of quadratics. Automatica 43(4), 685-692.

Hubbard, B.E., Crowle, J.K., Zimbelman, D.R., 2003. Comparative alteration mineral mapping using visible to shortwave infrared (0.4-2.4 μm) Hyperion, ALI, and ASTER imagery. IEEE Transactions on geoscience and remote sensing 41(6 PART I), 1401-1410.

Hubbard, B.E., Crowley, J.K., 2005. Mineral mapping on the Chilean-Bolivian Altiplano using co-orbital ALI, ASTER and Hyperion imagery: Data dimensionality issues and solutions. Remote Sensing of Environment 99(1-2), 173-186.

Hudson, B.D., 1992. The Soil Survey as Paradigm-based Science. Soil Science Society of America Journal 56(3), 836-841.

Huete, A., Miura, T., Gao, X., 2003. Land cover conversion and degradation analyses through coupled soil-plant biophysical parameters derived from Hyperspectral EO-1 Hyperion. IEEE Transactions on Geoscience and Remote Sensing 41(6 PART I), 1268-1276.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25(3), 295-309.

INRA Maroc, 2010. Local soil maps of North Morocco. INRA Maroc, Rabat, Morocco.

Isaaks, E.H., Srivastava, R.H., 1989. Ordinary Kriging, Applied geostatistics Oxford University Press, New York, pp. 278-322.

IUSS Working group WRB, 2006. World reference base for soil resources World Soil Resources Reports No. 103. FAO, Rome.

Iwahashi, J., Pike, R.J., 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86(3-4), 409-440.

Jabbar, M.T., Chen, X., 2008. Land degradation due to salinization in arid and semi-arid regions with the aid of geo-information techniques. Geo-spatial Information Science 11(2), 112-120.

Jaquet, O., 1989. Factorial kriging analysis applied to geological data from petroleum exploration. Mathematical Geology 21(7), 683-691.

Jenny, H., 1941. Factors of soil formation, A system of quantitative pedology. McGraw-Hill, New-York.

Jonckheere, I., 2012. The role of satellite data for the national forest monitoring systems in the context of REDD+, European Space Agency, (Special Publication) ESA SP.

Julien, Y., Sobrino, J.A., 2009. The Yearly Land Cover Dynamics (YLCD) method: An analysis of global vegetation from NDVI and

LST parameters. Remote Sensing of Environment 113(2), 329-334.

Kanner, L.C., Mustard, J.F., Gendrin, A., 2007. Assessing the limits of the Modified Gaussian Model for remote spectroscopic studies of pyroxenes on Mars. Icarus 187(2), 442-456.

Katzfuss, M., 2011. Hierarchical spatial and spatiao-temporal modeling of massice datasets, with application to global mapping of CO2, PhD thesis, The Ohio State University.

Katzfuss, M., 2013. Bayesian nonstationary spatial modeling for very large datasets. Environmetrics, Cornell University Library, arXiv:1204.2098v5

Katzfuss, M., Cressie, N., 2011. Tutorial on Fixed Rank Kriging (FRK) of CO2 data, Department of Statistics, The Ohio State University, Columbus.

Katzfuss, M., Cressie, N., 2012. Bayesian hierarchical spatio-temporal smoothing for very large datasets. Environmetrics 23(1), 94-107.

Kauffman, S., 2011. Green Water Credits - Policy Brief. ISRIC.

Kempen, B., 2011. Updating soil information with digital soil mapping, PhD thesis, Wageningen University, Wageningen, 213 pp.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., 2012. Soil type mapping using the generalised linear geostatistical model: A case study in a Dutch cultivated peatland. Geoderma 189–190(0), 540-553.

Keshava, N., Mustard, J.F., 2002. Spectral unmixing. IEEE Signal Processing Magazine 19(1), 44-57.

Kettles, I.M., Rencz, A.N., Bauke, S.D., 2000. Integrating Landsat, geologic, and airborne gamma ray data as an aid to surficial geology mapping and mineral exploration in the Manitouwadge area, Ontario. Photogrammetric Engineering and Remote Sensing 66(4), 437-445.

Kiang, N.Y., Siefert, J., Govindjee, Blankenship, E., R., 2007. Spectral signatures of photosynthesis. I. Review of earth organisms. Astrobiology 7(1), 222-251.

King, D., Daroussin, J., Tavernier, R., 1994. Development of a soil geographic database from the Soil Map of the European Communities. Catena 21(1), 37-56.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by Simulated Annealing. Science 220(4598), 671-680.

Kleeberg, R., 2005. Results of the second Reynolds Cup contest in quantitative mineral analysis. CPD Newsletter 30, 3.

Klingseisen, B., Metternicht, G., Paulus, G., 2008. Geomorphometric landscape analysis using a semi-automated GIS-approach. Environmental Modelling & Software 23(1), 109-121.

Knorr, W., Prentice, I.C., House, J.I., Holland, E.A., 2005. Long-term sensitivity of soil carbon turnover to warming. Nature 433(7023), 298-301.

Knotters, M., Brus, D.J., Oude Voshaar, J.H., 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67(3-4), 227-246.

Kodikara, G.R.L., Woldai, T., van Ruitenbeek, F.J.A., Kuria, Z., van der Meer, F., Shepherd, K.D., van Hummel, G.J., 2012. Hyperspectral remote sensing of evaporate minerals and associated sediments in Lake Magadi area, Kenya. International Journal of Applied Earth Observation and Geoinformation 14(1), 22-32.

Kokaly, R.F., 2011a. PRISM: Processing routines in IDL for spectroscopic measurements (installation manual and user's guide, version 1.0), U.S. Geological Survey Open-File Report pp. 432.

Kokaly, R.F., Couvillion, B.R., Holloway, J.M., Roberts, D.A., Ustin, S.L., Peterson, S.H., Khanna, S., Piazza, S.C., 2013. Spectroscopic remote sensing of the distribution and persistence of oil from the Deepwater Horizon spill in Barataria Bay marshes. Remote Sensing of Environment 129, 210-230.

Kokaly, R.F., King, T.V.V., Hoefen, T.M., Dudek, K.B., and Livo, K.E.,, 2011b. Surface materials map of Afghanistan: carbonates, phyllosilicates, sulfates, altered minerals, and other materials:U.S. Geological Survey Scientific Investigations Map 3152–A, one sheet, scale 1:1,100,000.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift 15(3), 5.

Kriebel, K.T., 1978. Average variability of the radiation reflected by vegetated surfaces due to differing irradiations. Remote Sensing of Environment 7(1), 81-83.

Kruse, F.A., Boardman, J.W., Huntington, J.F., 2003. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. IEEE Transactions on Geoscience and Remote Sensing 41(6 PART I), 1388-1400.

Kruse, F.A., Boardman, J.W., Huntington, J.F., Mason, P., Quigley, M.A., 2002. Evaluation and validation of EO-1 Hyperion for Geologic Mapping, International Geoscience and Remote Sensing Symposium (IGARSS) '02. Proceedings. 2002 IEEE International pp. 593-595.

Kruse, F.A., Lefkoff, A.B., Boardman, J.B., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment 44(2-3), 145-163.

Lagacherie, P., Baret, F., Feret, J., Netto, J.M., Robbez-Masson. J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sensing of Environment 112(3), 825-835.

Lagacherie, P., Holmes, S., 1997. Addressing geographical data errors in a classification tree for soil unit prediction. International Journal of Geographical Information Science 11(2), 183 - 198.

Lagacherie, P., McBratney, A.B., Voltz, M., 2007. Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science, 31. 1 ed. Elsevier, Amsterdam.

Lal, R., 2004. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. Science 304(5677), 1623-1627.

Lambert, J.J., Daroussin, J., Eimberck, M., Le Bas, C., Jamagne, M., King, D., Montanarella, L., 2002. Soil geogrpahical database for Eurasia & Mediterranean: Instructions guide for elaboration at scale 1;1,000,000 (version 4). Office for Official Publications of the European Communities, Luxembourg.

Lambot, S., Slob, E.C., Van Bosch, I.D., Stockbroeckx, B., Vanclooster, M., 2004. Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties. IEEE transactions on geoscience and remote sensing 42(11), 2555-2568.

Lane, M.D., Glotch, T.D., Dyar, M.D., Pieters, C.M., Klima, R., Hiroi, T., Bishop, J.L., Sunshine, J., 2011. Midinfrared spectroscopy of synthetic olivines: Thermal emission, specular and diffuse reflectance, and attenuated total reflectance studies of forsterite to fayalite. Journal of Geophysical Research 116(E8), E08010.

Lasne, Y., Paillou, P.H., Ruffié, G., Serradilla, C., Demontoux, F., Freeman, A., Farr, T., McDonald, K., ChapmanB., 2008. Effect of salinity on the dielectric properties of geological materials: Implication for soil moisture detection by means of radar remote sensing. IEEE Transactions on Geoscience and Remote Sensing 46(6), 1674-1688.

Lau, I.C., Cudahy, T.J., Caccetta, M.C., Kobayashi, C., Kashimura, O., Kato, M., Wheaton, G.A., Carter, D.J., 2012. Mapping surface soil mineralogy using hyperspectral and ASTER imagery: An example from Mullewa, Western Australia, Digital Soil Assessments and Beyond - Proceedings of the Fifth Global Workshop on Digital Soil Mapping, pp. 365-371.

Laurent, V.C.E., Verhoef, W., Clevers, J.G.P.W., Schaepman, M.E., 2011. Inversion of a coupled canopy–atmosphere model using multi-angular top-of-atmosphere radiance

data: A forest case study. Remote Sensing of Environment 115(10), 2603-2612.

Le Bas, C., King, D., Jamagne, M., Daroussin, J., 1998. The European Soil Information System, European Soil Bureau Research, Luxembourg.

Lesch, S.M., 2005. Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. Computers and Electronics in Agriculture 46(1–3), 153-179.

Liaghat, S., Balasundram, S.K., 2010. A review: The role of remote sensing in precision agriculture. American journal of agricultural and biological sciences 5(5), 6.

Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2008. Remote sensing and image interpretation, 6ed. Wiley & Sons, New York.

Lin, Y.-P., Chu, H.-J., Huang, Y.-L., Tang, C.-H., Rouhani, S., 2011. Monitoring and identification of spatiotemporal landscape changes in multiple remote sensing images by using a stratified conditional Latin hypercube sampling approach and geostatistical simulation. Environmental Monitoring and Assessment 177(1), 353-373.

Los, S.O., Justice, C.O., Tucker, C.J., 1994. A global 1° by 1° NDVI data set for climate studies derived from the GIMMS continental NDVI data. International Journal of Remote Sensing 15(17), 3493-3518.

Lozano-Garcia, D.F., Fernandez, R.N., Johannsen, C.J., 1991. Assessment of regional biomass-soil relationships using vegetation indexes. IEEE Transactions on Geoscience and Remote Sensing 29(2), 331-339.

Lu, D., Weng, G., 2007. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing 28(5), 823-870.

Luo, H., Ye, H., Ke, Y., Pan, J., Gong, J., Chen, X., 2005. Removing vegetation using unsupervised fully constrained least squares linear spectral mixture analysis method in soils surveying by remote sensing Proceedings of SPIE. Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, INTERNATIONAL pp. 90-98.

Macías, F., Arbestain, M.C., 2010. Soil carbon sequestration in a changing global environment. Mitigation and Adaptation Strategies for Global Change 15(6), 511-529.

MacMillan, R.A., Jones, R.K., McNabb, D.H., 2004. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). Computers, Environment and Urban Systems 28(3), 175-200.

MacMillan, R.A., Moon, D.E., Coupé, R.A., 2007. Automated predictive ecological mapping in a Forest Region of B.C., Canada, 2001-2005. Geoderma 140(4), 353-373.

Madhavan, B.B., Venkataraman, G., Shah, S.D., Mohan, B.K., 1997. Revealing the geology of the Great Nicobar Island, Indian Ocean, by the interpretation of airborne synthetic aperture radar images. International Journal of Remote Sensing 18(13), 2723-2742.

Mahecha, M.D., Schmidtlein, S., 2008. Revealing biogeographical patterns by nonlinear ordinations and derived anisotropic spatial filters. Global Ecology and Biogeography 17(2), 284-296.

Mahoney, S., James, P., Mauger, A., Heinson, G., 2002. Geologic and regolith mapping for mineral exploration in the Gawler Craton of South Australia using Hyperion and other remote sensing techniques, International Geoscience and Remote Sensing Symposium (IGARSS) '02. Proceedings. 2002 IEEE International pp. 1779-1781.

Maindonald, J., Braun, W., 2010. Data Analysis and Graphics Using R. Camebridge Series in Statistical and Probabilitic Mathematics. 3 ed. Camebridge University Press, Camebridge.

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154(1-2), 138-152.

Martelet, G., Truffert, C., Tourlière, B., Ledru, P., Perrin, J., 2006. Classifying airborne radiometry data with Agglomerative Hierarchical Clustering: A tool for geological mapping in context of rainforest (French Guiana). International Journal of Applied Earth Observation and Geoinformation 8(3), 208-223.

Martin, Y.E., Franklin, S.E., 2005. Classification of soil- and bedrock-dominated landslides in British Columbia using segmentation of satellite imagery and DEM data. International journal of remote sensing 26(7), 1505-1509.

Martini, B.A., Cocks, T.D., Cocks, P.A., Pickles, W.L., 2004. Operational airborne hyperspectral remote sensing for global geothermal exploration. Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International, pp. 626.

Matérn, B., 1960. Spatial variation. Meddelanden från Statens Skogsforskningsinstitut 49(5).

Mathworks, 2009. Matlab 7.8.0 The Mathworks Inc.

Mavris, C., Plötze, M., Mirabella, A., Giaccai, D., Valboa, G., Egli, M., 2011. Clay mineral evolution along a soil chronosequence in an Alpine proglacial area. Geoderma 165(1), 106-117.

McBratney, A.B., Hart, G.A., McGarry, D., 1991. The use of region partitioning to improve the representation of geostatistically mapped soil attributes. European journal of soil science 42(3), 513-532.

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma 109(1–2), 41-73.

McBratney, A.B., Minasny, B., Viscarra Rossel, R.A., 2006. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. Geoderma 136(1-2), 272-278.

McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97(3-4), 293-327.

McBratney, A.B., Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117(1-2), 3-52.

McCarty, G.W., Reeves, 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 Science Society of America Journal 66(2), 640-646.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21(2), 239-245.

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89(1-2), 67-94.

Melendez-Pastor, I., Navarro-Pedreño, J., Koch, M., Gómez, I., 2010. Applying imaging spectroscopy techniques to map saline soils with ASTER images. Geoderma 158(1-2), 55-65.

Merlin, O., Escorihuela, M.J., Mayoral, M.A., Hagolle, O., Al Bitar, A., Kerr, Y., 2013. Self-calibrated evaporation-based disaggregation of SMOS soil moisture: An evaluation study at 3km and 100m resolution in Catalunya, Spain. Remote sensing of environment 130, 25-38.

METI/ERSDAC, N.L., USGS/EROS, 2009. ASTER Global DEM Validation summary report.

Metternicht, G.I., 2003. Categorical fuzziness: A comparison between crisp and fuzzy class boundary modelling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios. Ecological Modelling 168(3), 371-389.

Metternicht, G.I., Fermont, A., 1998. Estimating Erosion Surface Features by Linear Mixture Modeling. Remote sensing of environment 64(3), 254-265.

Meyers, R., Montgommery, D.C., Anderson - Cook, C.M., 2011. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 2 ed. John Wiley & Sons.

Milcu, A., Lukac, M., Subke, J.-A., Manning, P., Heinemeyer, A., Wildman, D., Anderson, R., Ineson, P., 2012. Biotic carbon feedbacks in a materially closed soil-vegetation-atmosphere system. 2(4), 281-284.

Miller, A.J., 2002. Subset Selection in Regression, Second Editon. In: Monographs on statistics and applied probability Taylor & Francis Group.

Minár, J., Evans, I.S., 2008. Elementary forms for land surface segmentation: The theoretical basis of terrain analysis and geomorphological mapping. Geomorphology 95(3-4), 236-259.

Minasny, B., 2012. Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012, Sydney, Australiea. CRC Press, Sydney.

Minasny, B., McBratney, A.B., 2005. The Matérn function as a general model for soil variograms. Geoderma 128(3-4 special issue), 192-207.

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences 32(9), 1378-1388.

Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemometrics and Intelligent Laboratory Systems 94(1), 72-79.

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital Mapping of Soil Carbon, Advances in Agronomy, pp. 1-47.

Minasny, B., McBratney, A.B., Pichon, L., Sun, W., Short, M.G., 2009. Evaluating near infrared spectroscopy for field prediction of soil properties. Australian Journal of Soil Research 47(7), 664-673.

Minasny, B., McBratney, A.B., Tranter, G., Murphy, B.W., 2008. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. European Journal of Soil Science 59(5), 960-971.

Mirik, M., Norland, J.E., Crabtree, R.L., Biondini, M.E., 2005. Hyperspectral one-meter-resolution remote sensing in Yellow Stone Park, Wyoming: II. Biomass. Rangeland Ecology Management 58, 7.

Moore, A.B., Morris, K.P., Blackwell, G.K., Jones, A.R., Sims, P.C., 2003. Using geomorphological rules to classify photogrammetrically-derived digital elevation models. International Journal of Remote Sensing 24(13), 2613-2626.

Moore, D.M., Reynolds Jr., R.C., 1997. Diffraction and the Identification and Analysis of Clay Minerals. 2 ed. Oxford University Press, New York.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57(2), 443-452.

Moran, C.J., Bui, E.N., 2002. Spatial data mining for enhanced soil map modelling. International Journal of Geographical Information Science 16(6), 533-549.

Moré, J.J., 1978. The Levenberg-Marquardt Algorithm: Implementation and Theory. Lecture Notes in Mathematics 630. Springer-Verlag, New-York.

Morvan, X., Saby, N.P.A., Arrouays, D., Le Bas, C., Jones, R.J.A., Verheijen, F.G.A., Bellamy, P.H., Stephens, M., Kibblewhite, M.G., 2008. Soil monitoring in Europe: A review of existing systems and requirements for harmonisation. Science of The Total Environment 391(1), 1-12.

Mougenot, B., Pouget, M., Epema, G.F., 1993. Remote sensing of salt affected soils. Remote Sensing Reviews 7(3-4), 241-259.

Mücher, C.A., Hennekens, S.M., Bunce, R.G.H., Schaminée, J.H.J., Schaepman, M.E., 2009. Modelling the spatial distribution of Natura 2000 habitats across Europe. Landscape and Urban Planning 92(2), 148-159.

V.L. Mulder, S. de Bruin, J. Weyermann, R.F. Kokaly, M.E. Schaepman (Accepted for publication). Characterizing regional soil mineral composition using spectroscopy and geostatistics. *Remote Sensing of Environment.*

Mulder, V.L., de Bruin, S., Schaepman, M.E., 2011a. Towards spectroscopic modelling of composite mineralogy, 9th Swiss Geoscience Meeting, Zürich, Switzerland.

Mulder, V.L., de Bruin, S., Schaepman, M.E., 2012a. Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. International Journal of Applied Earth Observation and Geoinformation 21(1), 301-310.

Mulder, V.L., de Bruin, S., Schaepman, M.E., 2012b. Retrieval of composite mineralogy by VNIR spectroscopy, 5th Global Workshop on Digital Soil Mapping 2012, Sydney, Australia.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011b. The use of remote sensing in soil and terrain mapping - A review. Geoderma 162(1-2), 1-19.

Mulder, V.L., Plotze, M., de Bruin, S., Schaepman, M.E., Mavris, C., Kokaly, R., Egli, M., 2013. Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1 µm - 2.4 µm) and regression tree analysis. Geoderma (207-208), 279-290.

Mustard, J.F., Pieters, C.M., 1987. Quantitative abundance estimates from bidirectional reflectance measurements. Journal of Geophysical Research: Solid Earth 92(B4), E617-E626.

Nachtergaele, F.O., 1999. From the soil map of the world to the digital global soil and terrain database: 1960-2002. In: M.E. Sumner (Ed.), Handbook of soil science. CRC Press, Boca Raton

Nanni, M.R., Demattê, J.A.M., 2006. Spectral reflectance methodology in comparison to traditional soil analysis. Soil Science Society of America Journal 70(2), 393-407.

Ninomiya, Y., Fu, B., Cudahy, T.J., 2005. Detecting lithology with Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) multispectral thermal infrared 'radiance-at-sensor' data. Remote Sensing of Environment 99(1-2), 127-139.

Noble, S.K., Pieters, C.M., Hiroi, T., Taylor, L.A., 2006. Using the modified Gaussian model to extract quantitative data from lunar soils. Journal of Geophysical Research E: Planets 111(11).

Nychka, D., Wikle, C., Royle, J.A., 2002. Multiresolution models for nonstationary spatial covariance functions. Statistical Modelling 2(4), 315-331.

Odeh, I.O.A., McBratney, A.B., 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi Valley of eastern Australia. Geoderma 97(3-4), 237-254.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63(3-4), 197-214.

Odeh, I.O.A., Onus, A., 2008. Spatial analysis of soil salinity and soil structural stability in a semi-arid region of New South Wales, Australia. Environmental Management 42(2), 265-278.

Ogawa, Y., Matsunaga, T., Nakamura, R., Saiki, K., Ohtake, M., Hiroi, T., Takeda, H., Arai, T., Yokota, Y., Yamamoto, S., Hirata, N., Sugihara, T., Sasaki, S., Haruyama, J., Morota, T., Honda, C., Demura, H., Kitazato, K., Terazono, J., Asada, N., 2011. The widespread occurrence of high-calcium pyroxene in bright-ray craters on the Moon and implications for lunar-crust composition. Geophysical Research Letters 38(17), L17202.

Oldeman, L.R., van Engelen, V.W.P., 1993. A world soils and terrain digital database (SOTER) - An improved assessment of land resources. Geoderma 60(1-4), 309-325.

Omotoso, O., McCarty, D.K., Hillier, S., Kleeberg, R., 2006. Some successful approaches to quantitative mineral analysis as

revealed by the 3rd Reynolds Cup contest. Clays and Clay Minerals 54(6), 748-760.

Omuto, C.T., Shrestha, D.P., 2007. Remote sensing techniques for rapid detection of soil physical degradation. International Journal of Remote Sensing 28(21), 4785 - 4805.

Otto, J.-C., Dikau, R., Löwner, M.-O., 2010a. New GML-Based Application Schema for Landforms, Processes and Their Interaction, Landform - Structure, Evolution, Process Control. Lecture Notes in Earth Sciences. Springer Berlin / Heidelberg, pp. 21-36.

Otto, J.-C., Dikau, R., Schneider, M., Klein, R., 2010b. Semi-Automatic Digital Landform Mapping, Landform - Structure, Evolution, Process Control. Lecture Notes in Earth Sciences. Springer Berlin / Heidelberg, pp. 37-51.

Pagnutti, M., Ryan, R., E, 2009. Automated DEM Validation using ICESAT GLASS DATA, ASPRS/MAPPS 2009 Fall Conference, San Antonio, Texas.

Pal, M., Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment 86(4), 554-565.

Palacios-Orueta, A., Pinzon, J.E., Ustin, L.S., Roberts, D.A., 1999. Remote sensing of soils in the Santa Monica Mountains: II. Hierarchical foreground and background analysis. Remote Sensing of Environment 68(2), 138-151.

Palacios-Orueta, A., Ustin, L.S., 1998. Remote sensing of soil properties in the Santa Monica Mountains I. Spectral analysis. Remote Sensing of Environment 65(2), 170-183.

Panagos, P., van Liedekerke, M., Montanarella, L., 2011. Multi-scale European Soil Information System (MEUSIS): A multi-scale method to derive soil indicators. Computational Geosciences 15(3), 463-475.

Panciera, R., Walker, J.P., Kalma, J.D., Kim, E.J., Saleh, K., Wigneron, J.-P., 2009. Evaluation of the SMOS L-MEB passive microwave soil moisture retrieval algorithm. Remote Sensing of Environment 113(2), 435-444.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & geosciences 30(7), 683-691.

Petersen, G., Lebed, L., Fohrer, N., 2009. SRTM DEM levels over papyrus swamp vegetation - a correction approach. Advances in GeoSciences 21, 4.

Phillips, R.D., Blinn, C.E., Watson, L.T., Wynne, R.H., 2009. An Adaptive Noise-Filtering Algorithm for AVIRIS Data With Implications for Classification Accuracy. IEEE Transactions on Geoscience and Remote Sensing 47(9), 3168-3179.

Pieters, C.M., Boardman, J., Buratti, B., Chatterjee, A., Clark, R., Glavich, T., Green, R., Head, J., Isaacson, P., Malaret, E., McCord, T., Mustard, J., Petro, N., Runyon, C., Staid, M., Sunshine, J., Taylor, L., Tompkins, S., Varanasi, P., White, M., 2009. The Moon mineralogy mapper (M3) on Chandrayaan-1. Current Science 96(4), 500-505.

Pinet, P.C.H., F.; Clenet, H.; Chevrel, S.; Daydou, Y.; Baratoux, D.; Rosemberg, C.; Bibring, J. P.; Poulet, F.; Gondet, B.; Mustard, J.; Le Mouélic, S.; Bellucci, G.; OMEGA Team, 2007. Mafic Mineralogy Variations Across Syrtis Major Shield and Surroundings as Inferred from Visible-Near-Infrared Spectroscopy by OMEGA/Mars Express, Seventh International Conference on Mars, Pasadena, California, pp. 3146.

Plaza, A., Huber, S., Schaepman, M., Clevers, J., Malenovsky, Z., Carrère, V., Gamba, P., Zagajewski, B., 2008. D6.1 – Algorithm Theoretical Basis Document. University of Extremadura and Wageningen University, pp. 117.

Pompilio, L., Pedrazzi, G., Cloutis, E.A., Craig, M.A., Roush, T.L., 2010. Exponential Gaussian approach for spectral modelling: The EGO algorithm II. Band asymmetry. Icarus 208(2), 811-823.

Pompilio, L., Pedrazzi, G., Sgavetti, M., Cloutis, E.A., Craig, M.A., Roush, T.L., 2009. Exponential Gaussian approach for spectral modeling: The EGO algorithm I. Band saturation. Icarus 201(2), 781-794.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1992. Numerical recipes in FORTRAN: the art of scientific computing. 2 ed. Cambridge University Press, Cambridge.

Prima, O.D.A., Echigo, A., Yokoyama, R., Yoshida, T., 2006. Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. Geomorphology 78(3-4), 373-386.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sensing of Environment 48(2), 119-126.

Qi, J., Marsett, R., Heilman, P., Biedenbender, S., Moran, S., Goodrich, D., Weltz, M., 2002. RANGES improves satellite-based information and land cover assessments in southwest United States. EOS Transitions 83(51).

Quinlan, J.R., 1993. C4.5: Programs for machine learning. Morgan Kaufmann Publishers, San Mateo, California.

Quiñonero-Candela, J., Rasmussen, C.E., 2005. A unifying view of sparse approximate Gaussian process regression. Journal of Machine Learning Research 6, 1939-1959.

R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raper, J., Livingstone, D., 1995. Development of a geomorphological spatial model using object-oriented design. International Journal of Geographical information systems 9(4), 359 - 383.

Rasemann, S., Schmidt, J., Schrott, L., Dikau, R., 2004. Geomorphometry in mountain terrain. In: M.P. Bishop, J.F. Shroder (Eds.), GIS & Mountain Geomorphology. Praxis Publishing Ltd, Omaha.

Rawlings, J.O., Sastry, G.P., Dickey, D.A., 1998. Applied Regression analysis - A Research Tool. Mathematics and Statistics. Springer New York.

Rayburg, S., Thoms, M., Neave, M., 2009. A comparison of digital elevation models generated from different data sources. Geomorphology 106(3-4), 261-270.

Renard, D., 2011. Roger S. Bivand, Edzer J. Pebesma, Virgilio Gomez-Rubio: Applied Spatial Data Analysis with R. Mathematical Geosciences 43(5), 607-609.

Rezaei, Y., Mobasheri, M.R., Zoej, M.J.V., Schaepman, M.E., 2011. Endmember Extraction Using a Combination of Orthogonal Projection and Genetic Algorithm. Geoscience and Remote Sensing Letters, PP(99), 1-5.

Richter, R., Schläpfer, D., 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction. International Journal of Remote Sensing 23(13), 2631 - 2649.

Ripley, B., 2011. Recursive partitioning, RPackage version 3.1-50.

Roberts, D.A., Smith, M.O., Adams, J.B., 1993. Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. Remote Sensing of Environment 44(2-3), 255-269.

Rogan, J., Yool, S.R., 2001. Mapping fire-induced vegetation depletion in the Peloncillo Mountains, Arizona and New Mexico. International Journal of Remote Sensing 22(16), 3101 - 3121.

Rogge, D.M., Rivard, B., Zhang, J., Sanchez, A., Harris, J., Feng, J., 2007. Integration of spatial-spectral information for the improved extraction of endmembers. Remote Sensing of Environment 110(3), 287-303.

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sensing of Environment 55(2), 95-107.

Rossiter, D.G., 2004. Digital soil resource inventories: Status and prospects. Soil use and management 20(3), 296-301.

Roush, T.L., Singer, R.B., 1986. Gaussian Analysis of Temperature Effects on the Reflectance Spectra of Mafic Minerals in the 1-µm Region. Journal of Geophysical Research 91(B10), 10301-10308.

Rowan, L.C., Mars, J.C., 2003. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. Remote Sensing of Environment 84(3), 350-366.

Saadat, H., Bonnell, R., Sharifi, F., Mehuys, G., Namdar, M., Ale-Ebrahim, S., 2008. Landform classification from a digital elevation model and satellite imagery. Geomorphology 100(3-4), 453-464.

Salford-Systems, 2002. CART User's Guide: An Implementation of the Original CART Methodology. Salford Sytems, Sandiago.

Salisbury, J.W., D'Aria, D.M., 1992. Infrared (8-14 μm) remote sensing of soil particle size. Remote Sensing of Environment 42(2), 157-165.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., De Lourdes Mendonça-Santos, M., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T.G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.L., 2009. Digital soil map of the world. Science 325(5941), 680-681.

Sang, H., Huang, J.Z., 2012. A full scale approximation of covariance functions for large spatial data sets. Journal of the Royal Statistical Society. Series B: Statistical Methodology 74(1), 111-132.

Santos, M.Y., Amaral, L.A., 2004. Mining geo-referenced data with qualitative spatial reasoning strategies. Computers & Graphics 28(3), 371-379.

Sassaroli, A., Fantini, S., 2004. Comment on the modified Beer-Lamert law for scattering media. Physics in Medicine and Biology 49(14), 3.

Satterwhite, M.B., Ponder Henley, J., Carney, J.M., 1985. Effects of lichens on the reflectance spectra of granitic rock surfaces. Remote Sensing of Environment 18(2), 105-112.

Saunders, D.F., Burson, K.R., Thompson, C.K., 1999. Model for hydrocarbon microseepage and related near-surface alterations. American Association of Petroleum Geologists Bulletin 83(1), 170-185.

Schaepman, M.E., Wamelink, G.W.W., Van Dobben, H.F., Gloor, M., Schaepman-Strub, G., Kooistra, L., Clevers, F.G.P.W., Schmidt, A., Berendse, F., 2007. River floodplain vegetation scenario development using imaging spectroscopy derived products as input variables in a dynamic vegetation model. Photogrammetric Engineering and Remote Sensing 73(10), 10.

Schelling, J., 1970. Soil genesis, soil classification and soil survey. Geoderma 4(3), 165-193.

Schmidt, J., Hewitt, A., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. Geoderma 121(3-4), 243-256.

Schmidtlein, S., 2005. Imaging spectroscopy as a tool for mapping Ellenberg indicator values. Journal of Applied Ecology 42(5), 966-974.

Schmidtlein, S., Zimmermann, P., Schüpferling, R., Weiβ, C., Austin, M., 2007. Mapping the floristic continuum: Ordination space position estimated from imaging spectroscopy. Journal of Vegetation Science 18(1), 131-140.

Schneevoigt, N.J., van der Linden, S., Thamm, H.-P., Schrott, L., 2008. Detecting Alpine landforms from remotely sensed imagery. A pilot study in the Bavarian Alps. Geomorphology 93(1-2), 104-119.

Scull, P., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: A review. Progress in Physical Geography 27(2), 7.

Scull, P., Franklin, J., Chadwick, O.A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. Ecological Modelling 181(1), 1-15.

Sedov, S., Solleiro-Rebolledo, E., Morales-Puente, P., Arias-Herreìa, A., Vallejo-Gèomez, E., Jasso-Castaneda, C., 2003. Mineral and organic components of the buried paleosols of the Nevado de Toluca, Central

Mexico as indicators of paleoenvironments and soil evolution. Quaternary International 106-107, 169-184.

Selige, T., Bohner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136(1-2), 235-244.

Serbin, G., Daughtry, C.S.T., Hunt Jr, E.R., Reeves Iii, J.B., Brown, D.J., 2009. Effects of soil composition and mineralogy on remote sensing of crop residue cover. Remote Sensing of Environment 113(1), 224-238.

Sgavetti, M., Pompilio, L., Roveri, M., Manzi, V., Valentino, G.M., Lugli, S., Carli, C., Amici, S., Marchese, F., Lacava, T., 2009. Two geologic systems providing terrestrial analogues for the exploration of sulfate deposits on Mars: Initial spectral characterization. Planetary and Space Science 57(5-6), 614-627.

Shepard, M.K., Helfenstein, P., 2007. A test of the Hapke photometric model. Journal of Geophysical Research 112(E3), E03001.

Singer, R.B., 1981. Near-Infrared spectral reflectance of mineral mixtures: Systematic combinations of pyroxenes, olives, and iron oxides. Journal of Geophysical Research 86(B9), 7967-7982.

Singh, D., Herlin, I., Berroir, J.P., Silva, E.F., Simoes Meirelles, M., 2004. An approach to correlate NDVI with soil colour for erosion process using NOAA/AVHRR data. Advances in Space Research 33(3), 328-332.

Singh, D., Kathpalia, A., 2007. An efficient modeling with GA approach to retrieve soil texture, moisture and roughness from ERS-2 SAR data. Progress in Electromagnetics Research 77, 121-136.

Singh, D., Meirelles, M.S.P., Costa, G.A., Herlin, I., Berroir, J.P., Silva, E.F., 2006. Environmental degradation analysis using NOAA/AVHRR data. Advances in Space Research 37(4), 720-727.

Singhroy, V., Barnett, P.J., Assouad, P., Molch, K., 2003. Terrain interpretation from SAR Techniques, International Geoscience and Remote Sensing Symposium (IGARSS) '03. Proceedings. 2003 IEEE International (1), 106-108.

Singhroy, V., Molch, K., 2004. Geological case studies related to RADARSAT-2. Canadian Journal of Remote Sensing 30(6), 10.

Singhroy, V.H., 2000. Satellite based engineering-terrain mapping of Canada's boreal forest region, 14th International Conference on Applied Geological Remote Sensing. Springer Netherlands, pp. 131-140.

Slaymaker, O., 2001. The role of remote sensing in geomorphology and terrain analysis in the Canadian Cordillera. International Journal of Applied Earth Observation and Geoinformation 3(1), 7.

Snedecor, G.W., Cochran, W.G., 1989. Statistical Methods. 8 ed. Iowa State University Press.

Soenen, S.A., Peddle, D.R., Coburn, C.A., 2005. SCS+C: A modified sun-canopy-sensor topographic correction in forested terrain. IEEE Transactions on Geoscience and Remote Sensing 43(9), 2148-2159.

Soil Survey Division Staff, 1993. Soil Survey Manual, U.S. Department of Agriculture Handbook 18. Soil Conversation Service.

Soil Survey Staff - Natural Resources Conservation Service, American Web Soil Survey (2012). In: U.S. department of. Agriculture (Ed.), Natural Resources Conservation Service, United States Department of Agriculture.

Soil Survey Staff - Natural Resources Conservation Service, Soil Survey Geographic (SSURGO) (2012). In: U.S. department of. Agriculture (Ed.), Natural Resources Conservation Service, United States Department of Agriculture.

Solheim, I., Engelsen, O., Hosgood, B., Andreoli, G., 2000. Measurement and Modeling of the Spectral and Directional Reflection Properties of Lichen and Moss Canopies. Remote sensing of environment 72(1), 78-94.

Sommer, M., Wehrhan, M., Zipprich, M., Weller, U., zu Castell, W., Ehrich, S., Tandler, B.,

Selige, T., 2003. Hierarchical data fusion for mapping soil units at field scale. Geoderma 112(3-4), 179-196.

Spiess, D., 2003. Canada - Alberta Environmentally Sustainable Agriculture (CAESA): Soil Inventory Project Procedures Manual - The Soil Inventory Process, Government of Alberta - Agriculture and Rural development, Alberta.

Stahel, W.A., 2011. The R-Function regr for an Augmented Regression Analysis. ETH Zurich, Zurich, Switzerland, pp. 11.

Steffen, W., Sanderson, R.A., Tyson, P.D., Jager, J., Matson, P.A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.J., Turner, B.L., Wasson, R.J., 2004. Global Change and the Earth System: a Planet under Pressure. Global Change - The IGBP book series. Springer Berlin Heidelberg, Berlin, Germany.

Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 62(1), 77-89.

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and Near Infrared Spectroscopy in Soil Science, Advances in Agronomy, pp. 163-215.

Stolbovoi, V., McCallum, I., 2002. Land Resources of Russia. International Institute for Applied Systems Analysis, Austria.

Stoner, E.R., Baumgardner, M.F., 1981. Characteristic variations in reflectance of surface soils. Soil Science Society of America Journal 45(6), 5.

Stoorvogel, J.J., Kempen, B., Heuvelink, G.B.M., de Bruin, S., 2009. Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. Geoderma 149(1-2), 161-170.

Stuffler, T., Förster, K., Hofer, S., Leipold, M., Sang, B., Kaufmann, H., Penné, B., Mueller, A., Chlebek, C., 2009. Hyperspectral imaging - An advanced instrument concept for the EnMAP mission (Environmental Mapping and Analysis Programme). Acta Astronautica 65(7-8), 1107-1112.

Su, H., McCabe, M.F., Wood, E.F., Su, Z., Prueger, J.H., 2005. Modeling Evapotranspiration during SMACEX: Comparing Two Approaches for Local- and Regional-Scale Prediction. Journal of Hydrometeorology 6(6), 910-922.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrological Earth System Sciences 6(1), 85-100.

Sulaeman, Y., Minasny, B., McBratney, A.B., Sarwani, M., Sutandi, A., 2012. Harmonizing legacy soil data for digital soil mapping in Indonesia. Geoderma 192, 77-85.

Sumfleth, K., Duttmann, R., 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. Ecological Indicators 8(5), 485-501.

Sun, W., Liang, S., Xu, G., Fang, H., Dickinson, R., 2008. Mapping plant functional types from MODIS data using multisource evidential reasoning. Remote Sensing of Environment 112(3), 1010-1024.

Sunshine, J.M., Pieters, C.M., 1993. Estimating Modal Abundances From the Spectra of Natural and Laboratory Pyroxene Mixtures Using the Modified Gaussian Model. Journal of Geophysical Research 98(E5), 9075-9087.

Sunshine, J.M., Pieters, C.M., 1998. Determining the composition of olivine from reflectance spectroscopy. Journal of Geophysical Research 103(E6), 13675-13688.

Sunshine, J.M., Pieters, C.M., Pratt, S.F., 1990. Deconvolution of Mineral Absorption Bands: An Improved Approach. Journal of Geophysical Research 95(B5), 6955-6966.

Swayze, G.A., Clark, R.N., Goetz, A.F.H., Chrien, T.G., Gorelick, N.S., 2003. Effects of spectrometer band pass, sampling, and signal-to-noise ratio on spectral identification using the Tetracorder algorithm. Journal of Geophysical Research 108(E9).

Taramelli, A., Melelli, L., 2009. Map of deep seated gravitational slope deformations susceptibility in central Italy derived from

SRTM DEM and spectral mixing analysis of the Landsat ETM+ data. International Journal of Remote Sensing 30(2), 357 - 387.

Taylor, G., Dehaan, R., 2000. Salinity mapping with hyperspectral imagery. In: G. Taylor (Ed.), 14th International Conference, Applied Geologic Remote Sensing. Springer Netherlands, Las Vegas, pp. 512-519.

Taylor, G.R., Mah, A.H., Kruse, F.A., Kierein-Young, K.S., Hewson, R.D., Bennett, B.A., 1996. Characterization of saline soils using airborne radar imagery. Remote Sensing of Environment 57(3), 127-142.

Teruiya, R.K., Paradella, W.R., Dos Santos, A.R., Dall'Agnol, R., Veneziani, P., 2008. Integrating airborne SAR, Landsat TM and airborne geophysics data for improving geological mapping in the Amazon region: The Cigano Granite, Caraja's Province, Brazil. International Journal of Remote Sensing 29(13), 3957-3974.

Thomasson, J.A., Sui, R., Cox, M.S., Al-Rajehy, A., 2001. Soil reflectance sensing for determining soil properties in precision agriculture. Transactions of the ASABE 44(6), 1445-1453.

Tribe, A., 1992. Automated recognition of valley lines and drainage networks from grid digital elevation models: a review and a new method. Journal of Hydrology 139(1-4), 263-293.

Tscharntke, T., Hochberg, M.E., Rand, T.A., Resh, V.H., Krauss, J., 2007. Author Sequence and Credit for Contributions in Multiauthored Publications. PLOS Biology 5(1), e18.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8(2), 127-150.

Tucker, C.J., Vanpraet, C.L., Sharman, M.J., van Ittersum, G., 1985. Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel: 1980-1984. Remote Sensing of Environment 17(3), 233-249.

Ulrich, M., Grosse, G., Chabrillat, S., Schirrmeister, L., 2009. Spectral characterization of periglacial surfaces and geomorphological units in the Arctic Lena Delta using field spectrometry and remote sensing. Remote Sensing of Environment 113(6), 1220-1235.

United Nations, 1994. Elaboration of an internation convetion to combat desertification in countries experiencing serious drought and/or desertification, particularly in Africa.: United Nations (Ed.), pp. 58.

United States Department of Agriculture, N.R.C.S., 2006 Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. U.S. Department of Agriculture Handbook 296. U.S. department of. Agriculture (Ed.).

Ustin, S.L., Gamon, J.A., 2010. Remote sensing of plant functional types. New Phytologist 186(4), 795-816.

van Asselen, S., Seijmonsbergen, A.C., 2006. Expert-driven semi-automated geomorphological mapping for a mountainous area using a laser DTM. Geomorphology 78(3-4), 309-320.

Van Der Kwast, J., 2009. Quantification of top soil moisture patterns Evaluation of field methods, process-based modelling, remote sensing and an integrated approach, PhD Thesis, Utrecht University, Utrecht, The Netherlands 1-317 pp.

van der Meer, F.D., van der Werff, H.M., van Ruitenbeek, F.J., Hecker, C.A., Bakker, W.H., Noomen, M.F., van der Meijde, M., Carranza, E.J.M., de Smeth, J.B., Woldai, T., 2012. Multi- and hyperspectral geologic remote sensing: A review. International Journal of Applied Earth Observation and Geoinformation 14(1), 112-128.

van Egmond, F.M., Dietrich, P., Werban, U., Sauer, U., 2009. iSOIL: exploring the soil as the basis forquality crop production and food security. Quality Assurance and Safety of Crops & Foods, 4.

van Engelen, V., 2008. Annex 1 - Description of Work. EU 211578, ISRIC, Wageningen, The Netherlands.

van Engelen, V.W.P., Verdoodt, A., Dijkshoorn, J.A., Van Ranst, E., 2006. Soil and Terrain Database of Central Africa (DR of Congo, Burundi and Rwanda). ISRIC - World Soil Information, Wageningen.

van Engelen, V.W.P., Wen, T.T., 1995. Global and national soils and terrain digital databases (SOTER): procedures amnual (revised edition). ISRIC – World Soil Information, Wageningen, The Netherlands.

van Groenigen, J.W., Stein, A., 1998. Constrained Optimization of Spatial Sampling using Continuous Simulated Annealing. Journal of Environmental Quality 27(5), 1078-1086.

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. Environmental Modelling & Software 25(3), 310-319.

Vasques, G., Grunwald, S., Myers, D.B., 2012. Associations between soil carbon and ecological landscape variables at escalating spatial scales in Florida, USA. 27(3), 355-367.

Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. Statistics and Computing. Springer, Berlin.

Viscarra Rossel, R.A., 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemometrics and Intelligent Laboratory Systems 90(1), 72-83.

Viscarra Rossel, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. Journal of Geophysical Research 116(F4), F04023.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal Soil Sensing. An Effective Approach for Soil Measurements in Space and Time, Advances in Agronomy, pp. 237-282.

Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. Geoderma 150(3-4), 253-266.

Viscarra Rossel, R.A., McBratney, A.B., 2008. Diffuse Reflectance Spectrscopy as a Tool for Digital Soil Mapping. In: A.E. Hartemink, A.B. Mcbratney, M. De Lourdes Mendonca-Santos (Eds.), Digital Soil Mapping with Limited Data. Springer Science + Business Media B.V., pp. 165-172.

Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B., 2006a. Colour space models for soil science. Geoderma 133(3-4), 320-337.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006b. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131(1-2), 59-75.

Wagner, W., Pathe, C., Sabel, D., Bartsch, A., Kunzer, C., Scipal, K., 2007. Experimental 1 km soil moisture products from ENVISTAT ASAR for Southern Africal, SHARE project and the MISAR project. European Space Agency.

Wagner, W., Scipal, K., 2000. Large-scale soil moisture mapping in Western Africa using the ERS Scatterometer. IEEE transactions on Geoscience and Remote Sensing 38(4), 6.

Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Computers & geosciences 36(10), 1261-1267.

Wang, Y.Y., 2008. Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data. International Journal of Remote Sensing 29(10), 2993-3010.

Warell, J., 2003. Properties of the Hermean regolith: III. Disk-resolved vis-NIR reflectance spectra and implications for the abundance of iron. ICARUS 161(2), 24.

Warell, J., Davidsson, B.J.R., 2010. A Hapke model implementation for compositional analysis of VNIR spectra of Mercury. Icarus 209(1), 164-178.

Webster, R., Oliver, M.A., 2007. Geostatistics for environmental scientists. Statistics in practice. Wiley, Chichester.

Weng, Y., Gong, P., Zhu, Z., 2008. Reflectance spectroscopy for the assessment of soil salt content in soils of the Yellow River Delta of China. International Journal of Remote Sensing 29(19), 5511 - 5531.

White, J.W., Andrade-Sanchez, P., Gore, M.A., Bronson, K.F., Coffelt, T.A., Conley, M.M., Feldmann, K.A., French, A.N., Heun, J.T., Hunsaker, D.J., Jenks, M.A., Kimball, B.A., Roth, R.L., Strand, R.J., Thorp, K.R., Wall, G.W., Wang, G., 2012. Field-based phenomics for plant genetics research. Field Crops Research 133, 101-112.

Whiting, M.L., Li, L., & Ustin, S.L. (2004). Predicting water content using Gaussian model on soil spectra. Remote Sensing of Environment 89, 535-552

Wigneron, J.P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.J., Richaume, P., Ferrazzoli, P., de Rosnay, P., Gurney, R., Calvet, J.C., Grant, J.P., Guglielmetti, M., Hornbuckle, B., Matzler, C., Pellarin, T., Schwank, M., 2007. L-band Microwave Emission of the Biosphere (L-MEB) Model: Description and calibration against experimental data sets over crop fields. Remote Sensing of Environment 107(4), 639-655.

Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1(6), 80-83.

Wilford, J.R., Bierwirth, P.N., Craig, M.A., 1997. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. AGSO Journal of Australian Geology and Geophysics 17(2), 201-216.

Wilson, J.B., 1999. Guilds, Functional Types and Ecological Groups. Oikos 86(3), 507-522.

Wilson, J.P., Gallant, J.C., 2000. Digital Terrain Analysis. In: J.P. Wilson, J.C. Gallant (Eds.), Terrain analysis: Principles and Applications. John Wiley & Sons, Inc, New York, pp. 1-27.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems 2(1-3), 37-52.

Woolard, J.W., Colby, J.D., 2002. Spatial characterization, resolution, and volumetric change of coastal dunes using airborne LIDAR: Cape Hatteras, North Carolina. Geomorphology 48(1-3), 269-287.

Xu, M., Watanachaturaporn, P., Varshney, P.K., Arora, M.K., 2005. Decision tree regression for soft classification of remote sensing data. Remote Sensing of Environment 97(3), 322-336.

Xu, Y., Lin. Q., Shao, Y., Wang, L., 2004. Extraction mechanism of Alteration Zones using ASTER Imagery, International Geoscience and Remote Sensing Symposium (IGARSS) '04. Proceedings. 2004 IEEE International pp. 2.

Yang, L., Xian, G., Klaver, J.M., Deal, B., 2003. Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data. Photogrammetric Engineering and Remote Sensing 69(9), 1003-1010.

Yi, Q.-x., Huang, J.-f., Wang, F.-m., Wang, X.-z., 2008. Quantifying biochemical variables of corn by hyperspectral reflectance at leaf scale. Journal of Zhejiang University - Science B 9(5), 378-384.

Yoshiki, N., Maldague, X.P., Rozlosnik, A.E., 2002. Mapping quartz, carbonate minerals, and mafic-ultramafic rocks using remotely sensed multispectral thermal infrared ASTER data, SPIE, pp. 191-202.

Yoshiki, N., Meynart, R., Neeck, S.P., Shimoda, H., Lurie, J.B., Aten, M.L., 2004. Lithologic mapping with multispectral ASTER TIR and SWIR data, SPIE, pp. 180-190.

Yun, S., Huadong, G., Qingrong, H., Yuan, L., Qing, D., Chunming, H., 2003. Effect of dielectric properties of moist salinized soils on backscattering coefficients extracted from RADARSAT image, Geoscience and Remote Sensing Symposium, 2003. IGARSS '03.

Proceedings. 2003 IEEE International, pp. 2789-2791 vol.2784.

Zhai, Y., Thomasson, J.A., Boggess, J.E., Sui, R., 2006. Soil texture classification with artificial neural networks operating on remote sensing data. Computers and Electronics in Agriculture 54(2), 16.

Zhang, J., Rivard, B., Rogge, D.M., 2008. The Successive Projection Algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data. Sensors 8(2), 1321-1342.

Zhang, J., Rivard, B., Sánchez-Azofeifa, A., 2005. Spectral unmixing of normalized reflectance data for the deconvolution of lichen and rock mixtures. Remote Sensing of Environment 95(1), 57-66.

Zhang, X., Pazner, M., Duke, N., 2007. Lithologic and mineral information extraction for gold exploration using ASTER data in the south Chocolate Mountains (California). ISPRS Journal of Photogrammetry and Remote Sensing 62(4), 12.

Zhu, Z., Stein, M.L., 2006. Spatial sampling design for prediction with estimated parameters. Journal of Agricultural, Biological, and Environmental Statistics 11(1), 24-44.

# Summary

Global environmental changes have resulted in changes in key ecosystem services that soils provide. It is necessary to have up to date soil information on regional and global scales to ensure that these services continue to be provided. As a result, Digital Soil Mapping (DSM) research priorities are among others, advancing methods for data collection and analyses tailored towards large-scale mapping of soil properties. Scientifically, this thesis contributed to the development of methodologies, which aim to optimally use remote and proximal sensing (RS and PS) for DSM to facilitate regional soil mapping. The main contributions of this work with respect to the latter are (I) the critical evaluation of recent research achievements and identification of knowledge gaps for large-scale DSM using RS and PS data, (II) the development of a sparse RS-based sampling approach to represent major soil variability at regional scale, (III) the evaluation and development of different state-of-the-art methods to retrieve soil mineral information from PS, (IV) the improvement of spatially explicit soil prediction models and (V) the integration of RS and PS methods with geostatistical and DSM methods.

A review on existing literature about the use of RS and PS for soil and terrain mapping was presented in Chapter 2. Recent work indicated the large potential of using RS and PS methods for DSM. However, for large-scale mapping, current methods will need to be extended beyond the plot. Improvements may be expected in the fields of developing more quantitative methods, enhanced geostatistical analysis and improved transferability to other areas. From these findings, three major research interests were selected: (I) soil sampling strategies, (II) retrieval of soil information from PS and (III) spatially continuous mapping of soil properties at larger scales using RS.

Budgetary constraints, limited time and available soil legacy data restricted the soil data acquisition, presented in Chapter 3. A 15.000 $km^2$ area located in Northern Morocco served as test case. Here, a sample was collected using constrained Latin Hypercube Sampling (cLHS) of RS and elevation data. The RS data served as proxy for soil variability, as alternative for the required soil legacy data supporting the sampling strategy. The sampling aim was to optimally sample the variability in the RS data while minimizing the acquisition efforts. This sample resulted in a dataset representing major soil variability. The cLHS sample failed to express spatial correlation; constraining the LHS by a distance criterion favoured large spatial variability over short distances. The absence of spatial correlation in the sampled soil variability precludes the use of additional geostatistical analyses to spatially predict soil properties. Predicting soil properties using the cLHS sample is thus restricted to a modelled statistical relation between the sample and exhaustive predictor variables.

For this, the RS data provided the necessary spatial information because of the strong spatial correlation while the spectral information provided the variability of the environment (Chapter 3 and 6). Concluding, the RS-based cLHS approach is considered a time and cost efficient method for acquiring information on soil resources over extended areas.

This sample was further used for developing methods to derive soil mineral information from PS, and to characterize regional soil mineralogy using RS. In Chapter 4, the influences of complex scattering within the mixture and overlapping absorption features were investigated. This was done by comparing the success of PRISM's MICA in determining mineralogy of natural samples and modelled spectra. The modelled spectra were developed by a linearly forward model of reflectance spectra, using the fraction of known constituents within the sample. The modelled spectra accounted for the co-occurrence of absorption features but eluded the complex interaction between the components. It was found that more minerals could be determined with higher accuracy using modelled reflectance. The absorption features in the natural samples were less distinct or even absent, which hampered the classification routine. Nevertheless, grouping the individual minerals into mineral categories significantly improved the classification accuracy. These mineral categories are particularly useful for regional scale studies, as key soil property for parent material characterization and soil formation. Characterizing regional soil mineralogy by mineral categories was further described in Chapter 6. Retrieval of refined information from natural samples, such as mineral abundances, is more complex; estimating abundances requires a method that accounts for the interaction between minerals within the intimate mixture. This can be done by addressing the interaction with a non-linear model (Chapter 5).

Chapter 5 showed that mineral abundances in complex mixtures could be estimated using absorption features in the 2.1–2.4 µm wavelength region. First, the absorption behaviour of mineral mixtures was parameterized by exponential Gaussian optimization (EGO). Next, mineral abundances were successfully predicted by regression tree analysis, using these parameters as inputs. Estimating mineral abundances using prepared mixes of calcite, kaolinite, montmorillonite and dioctahedral mica or field samples proved the validity of the proposed method. Estimating mineral abundances of field samples showed the necessity to deconvolve spectra by EGO. Due to the nature of the field samples, the simple representation of the complex scattering behaviour by a few Gaussian bands required the parameters asymmetry and saturation to accurately deconvolve the spectra. Also, asymmetry of the EGO profiles showed to be an important parameter for estimating the abundances of the field samples. The robustness of the method in handling the omission of minerals during the training phase was tested by replacing part of the quartz with chlorite. It was found that the accuracy of the predicted mineral content was hardly affected. Concluding, the proposed method allowed for estimating more than two

minerals within a mixture. This approach advances existing PS methods and has the potential to quantify a wider set of soil properties. With this method the soil science community was provided an improved inference method to derive and quantify soil properties

The final challenge of this thesis was to spatially explicit model regional soil mineralogy using the sparse sample from Chapter 3. Prediction models have especially difficulties relating predictor variables to sampled properties having high spatial correlation. Chapter 6 presented a methodology that improved prediction models by using scale-dependent spatial variability observed in RS data. Mineral predictions were made using the abundances from X-ray diffraction analysis and mineral categories determined by PRISM. The models indicated that using the original RS data resulted in lower model performance than those models using scaled RS data. Key to the improved predictions was representing the variability of the RS data at the same scale as the sampled soil variability. This was realized by considering the medium and long-range spatial variability in the RS data. Using Fixed Rank Kriging allowed smoothing the massive RS datasets to these ranges. The resulting images resembled more closely the regional spatial variability of soil and environmental properties. Further improvements resulted from using multi-scale soil-landscape relationships to predict mineralogy. The maps of predicted mineralogy showed agreement between the mineral categories and abundances. Using a geostatistical approach in combination with a small sample, substantially improves the feasibility to quantitatively map regional mineralogy. Moreover, the spectroscopic method appeared sufficiently detailed to map major mineral variability. Finally, this approach has the potential for modelling various natural resources and thereby enhances the perspective of a global system for inventorying and monitoring the earth's soil resources.

With this thesis it is demonstrated that RS and PS methods are an important but also an essential source for regional-scale DSM. Following the main findings from this thesis, it can be concluded that: *Improvements in regional-scale DSM result from the integrated use of RS and PS with geostatistical methods.* In every step of the soil mapping process, spectroscopy can play a key role and can deliver data in a time and cost efficient manner. Nevertheless, there are issues that need to be resolved in the near future. Research priorities involve the development of operational tools to quantify soil properties, sensor integration, spatiotemporal modelling and the use of geostatistical methods that allow working with massive RS datasets. This will allow us in the near future to deliver more accurate and comprehensive information about soils, soil resources and ecosystem services provided by soils at regional and, ultimately, global scale.

# Samenvatting

Mondiale veranderingen in het functioneren van het systeem aarde en het milieu hebben veranderingen gebracht in de ecosysteemdiensten die bodems leveren. Om te garanderen dat deze diensten ook in de toekomst geleverd worden, is inzicht nodig in de huidige status van bodems en hun ecosysteemdiensten. Onderzoekslijnen, binnen de wetenschappelijke bodemgemeenschap, richten zich daarom op het ontwikkelen van methoden om gegevens te verzamelen en te analyseren die specifiek gericht zijn op regionale en globale studies. De wetenschappelijke bijdrage van dit werk omvat de ontwikkeling van methoden die optimaal gebruik maken van aardobservatie (AO) en spectroscopie voor grootschalige, digitale bodemkartering. De belangrijkste resultaten van dit werk richten zich op (I) het kritisch evalueren van recente onderzoeken en het identificeren van kennishiaten op het gebied van grootschalige bodemkartering die gebruik maken van AO en spectroscopie, (II) het ontwikkelen van een bemonsteringsmethode die enkel gebruik maakt van AO, (III) de evaluatie en ontwikkeling van geavanceerde methoden voor het bepalen van bodemmineralogie met spectroscopie, (IV) het verbeteren van regionale ruimtelijke modellen voor het voorspellen van bodemeigenschappen en (V) de integratie van AO en spectroscopie met geostatistische en digitale karteringsmethoden.

Hoofdstuk 2 is een literatuurstudie naar het gebruik van AO en spectroscopie voor bodem- en terreinkartering. Recente studies wezen erop dat AO en spectroscopie veel potentie hebben om positief bij te dragen aan het verbeteren van bestaande methodes voor grootschalige kartering. Toch moeten de huidige methoden nog verder ontwikkeld worden om op grotere schaal toegepast te kunnen worden. Verbeteringen kunnen verwacht worden op het gebied van de ontwikkeling van kwantitatieve methoden, verbeterde geostatistische analyses en toepasbaarheid van methoden naar andere gebieden. Gebaseerd op deze bevinden zijn drie onderzoekslijnen vastgesteld: (I) strategieën voor bodembemonstering, (II) bepalen van bodemeigenschappen met gebruik van spectroscopie en (III) ruimtelijk karteren van bodemeigenschappen op grote schaal.

Budgettaire beperkingen, gelimiteerde tijd en beschikbare bodemgegevens waren bepalend voor de data acquisitie in het veld (Hoofdstuk 3). Een gebied ter grote van 15.000 $km^2$ in het noorden van Marokko diende als studiegebied. In dit gebied was een monster verzameld gebaseerd op constrained Latin Hypercube Sampling (cLHS) van AO en hoogtegegevens. De AO gegevens werden als alternatief gebruikt voor de bodemgegevens die nodig zijn ter ondersteuning van de methode. Het bemonsteringsschema richtte zich op het volledig bemonsteren van de variabiliteit in deze data terwijl, tegelijkertijd, de kosten voor acquisitie beperkt werden. Dit monster bleek representatief voor de algemene bodemvariabiliteit in het gebied. Het

cLHS-monster was niet in staat om de ruimtelijke correlatie van bodemeigenschappen te beschrijven. Dit was te verwachten aangezien de LHS gelimiteerd was met een criterium dat de voorkeur had om veel ruimtelijke variabiliteit te bemonsteren over korte afstanden. Dit leidde ertoe dat het monster niet geschikt was om bodem mineralogie te voorspellen met geostatistische methoden. Als alternatief moest een empirische methode gebruikt worden die het monster relateert aan de AO gegevens om zo mineralogie te voorspellen. Hiervoor zijn de AO gegevens uitermate geschikt, mede omdat deze sterke ruimtelijke correlatie vertoond en spectraal representatief zijn voor de bodem variabiliteit (Hoofdstuk 3 en 6). De cLHS, gebruikmakende van AO gegevens, word beschouwd als een tijd- en kosteneffectieve bemonsteringsmethode om bodemgegevens in te winnen voor grote gebieden.

In Hoofdstuk 4, werden de invloed van complexe spectrale interacties van en tussen mineralen op het bepalen van mineralogie onderzocht. Dit was gerealiseerd door mineralogie te bepalen met PRISM's MICA met de originele spectra en lineair gemodelleerde spectra van de minerale fracties aanwezig in het monster. De gemodelleerde spectra bevatten de specifieke minerale absorptie kenmerken welke voorkomen op dezelfde golflengtes maar de complexe spectrale interactie tussen de mineralen werden zo uitgesloten. Voor de gemodelleerde spectra bleek dat meerdere mineralen bepaald konden worden met een hogere nauwkeurigheid. In de originele spectra waren de specifieke minerale absorptie kenmerken minder duidelijk te onderscheiden of zelfs afwezig. Hierdoor is MICA minder goed in staat om mineralen te classificeren. Het samenvoegen van individuele mineralen in categorieën resulteerde in een accuratere classificatie. Deze categorieën zijn bijzonder geschikt voor regionale studies en kunnen dienen als belangrijke bodemeigenschap om moedermateriaal of bodemvorming te karakteriseren. Het karteren van deze minerale categorieën is verder beschreven in Hoofdstuk 6. Voor het bepalen van gedetailleerdere informatie zijn methoden nodig die zich richten op de wisselwerking tussen aanwezige mineralen (Hoofdstuk 5).

Hoofdstuk 5 heeft laten zien dat hoeveelheden aan mineralen in samengestelde mengsels geschat kunnen worden met behulp van absorptiekenmerken die voorkomen in de golflengtes tussen 2.1-2.4 µm. De karakteristieke absorptiekenmerken waren eerst geparameteriseerd met behulp van exponentiële Gaussische optimalisatie (EGO). De mineraalhoeveelheden werden geschat door middel van regressiebomen en de resulterende EGO parameters. Het schatten van mineraalhoeveelheden was getest op samengestelde monsters die calciet, kaoliniet, dioctahedral mica en montmorillonite bevatten en natuurlijke monsters met soortgelijke samenstelling. Deze experimenten toonden aan dat de voorgestelde methode goed in staat is om hoeveelheden te schatten. Ook bleek de noodzaak om EGO te gebruiken; het nauwkeurig modelleren van de complexe spectrale interactie in natuurlijke monsters met enkele Gaussische modellen vereiste het gebruik van de parameters asymmetrie

en verzadiging van het absorptiekenmerk. Daarnaast had asymmetrie een sterke voorspellende waarde om mineraalhoeveelheden in complexe mengsels te schatten. In een derde experiment werd de robuustheid van het model getest door een deel van de quartz in de monsters te vervangen met chloriet. De schatting van hoeveelheden bleek hier nauwelijks door beïnvloed. Tot slot, de methode kan hoeveelheden van meer dan 2 mineralen in een bodemmonster schatten en heeft de potentie om verschillende bodemeigenschappen te kwantificeren; hiermee heeft de bodemkundige gemeenschap een verbeterde methode voor het verkrijgen van kwantitatieve bodeminformatie met behulp van spectroscopie.

De laatste uitdaging in dit proefschrift was het karakteriseren van regionale mineralogie gebruikmakende van het marginale monster. Regionale modellen zijn minder goed in staat om de relatie te leggen tussen voorspellende variabelen en een bodemeigenschap welke grote ruimtelijke variabiliteit heeft. Hoofdstuk 6 beschreef een methode om modellen te verbeteren door de schaal-afhankelijke ruimtelijke variabiliteit die aanwezig is in AO gegevens te gebruiken. Fixed Rank Kriging maakte het mogelijk om de middellange en lange schaal te modelleren ondanks de enorme omvang van de dataset. Het resultaat was een verbeterde weergave van de regionale ruimtelijke variabiliteit van zowel bodem- als omgevingsfactoren. De originele AO data bleek ongeschikt om hoeveelheden te voorspellen en minder accuraat voor de klassen. Verdere verbeteringen waren behaald door bestaande bodem-landschapsrelaties op verschillende schaalniveaus te gebruiken. De kaarten van voorspelde mineraalklassen en -hoeveelheden voor het studiegebied in Noord Marokko, hadden grote overeenkomsten. MICA was in staat categorieën te voorspellen van mica, kaoliniet en smectiet, in een kalkrijk of –arm milieu. Het gebruik van een geostatistische methode in combinatie met een klein maar representatief monster verbetert de kwantificatie van regionale mineralogie. Tot slot, deze methode heeft de potentie om meerdere natuurlijke hulpbronnen te modelleren. Hiermee vergroten we het perspectief voor een mondiaal systeem voor het inventariseren en monitoren van bodems en andere natuurlijke hulpbronnen.

Dit proefschrift toont aan dat AO en spectroscopie een belangrijke bron zijn voor regionale bodemkartering. Op basis van deze thesis kan geconcludeerd worden dat: *Regionale bodemkartering verbeterd door geïntegreerd gebruik te maken van spectroscopie en aardobservatie met geostatistische methoden*. In elke stap van bodemkartering kan spectroscopie een sleutelrol spelen en efficiënt gegevens leveren. Echter, duidelijk is dat er problemen moeten worden opgelost in de nabije toekomst. Onderzoeksprioriteiten moeten zich richten op het ontwikkelen van (I) operationele methoden voor kwantitatieve bepalingen van bodemeigenschappen, (II) de integratie van sensoren, (III) ruimtelijk-temporeel modelleren en (IV) het gebruik van geschikte geostatistische methoden voor omvangrijke AO data. Deze ontwikkelingen zullen het mogelijk maken om, in de nabije toekomst, informatie te leveren over bodems en de geleverde ecosysteemdiensten.

# Acknowledgements

Looking back and reflecting on the past five years, I realize that the successful completion of my PhD thesis would not have been possible without the support and contributions I received during the past years. I am grateful to many people and therefore I take the opportunity here to thank all of you. First of all, I would like to deeply thank my supervisors Michael Schaepman and Sytze de Bruin. Thank you Michael, you gave me good opportunities to specialize myself in remote sensing of soils. You have always given me enough food for thoughts which allowed me to go beyond set standards. Thanks, for giving me the time and support to fulfill my quest. Sytze, as my daily supervisor you were always there for me, and I am grateful for all your time and patience during the many discussions. Also, thank you for keeping my feet to the ground while giving me the freedom to find my way around in science. Harm Bartholomeus, being the only person involved in proximal soil sensing at CGI, I would like to thank you for the many discussions we had during the coffee breaks and the constructive feedback to my work.

It fits a soil scientist to get the hands dirty by doing fieldwork. The fieldwork campaign, which led to the results in this thesis, could not have been possible without the support of various people. I know that the campaign was ambitious and that circumstances where challenging, but we managed. So many thanks go to my driver Bouamer Rached, who drove us around for at least twelve hours a day. Thank you Elouadi Mbarek, for your local soil-knowledge and our nice French – English teaching moments. Being out and about in the field, collecting data was a tough job therefore I want to thank Karim Andich and Jeroen van Bruggen who literally lifted my burden in the field.

During the course of my PhD, my work got tailored towards mineralogy. I would like to thank my co-authors Michael Plötze, Christian Mavris and Raymond Kokaly, who gave me the mineralogical lectures that were needed to perform my work. Also, I am grateful to Michael and Christian, for helping me to perform the necessary mineralogical analysis at the University of Zürich and the ETH. Raymond, I am grateful for the extensive email-conversations on mineral spectroscopy and the constructive feedback on my work.

I would like to thank all my colleagues from the Centre for Geo-Information, Wageningen and Remote Sensing Laboratories, Zürich. A few people I want to mention in particular; my amazing office mates in Wageningen: Lucia, Rogier, Daniëla, Nandika, Brice, Kim and Loïc. My other amazing office mates in Zurich: Valérie, again Lucia, Lucie, Petra, Parvis and Hossein. Hans Roelofsen, Marc Vis and Christian Mavris, both of you thanks for all the good times and interesting discussions about how science “rocks”! Rogier, *intelligenti pauca*: thanks. Thank you

all, for the great time I had with you, including all the fun we had with science-comics, elongated coffee breaks, and the many bbq's and beers after work.

And then there are some more friends who were able to give me the necessary distractions and mental support during my PhD. Thanks to Aimee, Caroline, Christel, Corinne, Nanette and Rineke for all the nice dinners, trips and 'wine'-evenings. Rineke, thanks for your amazing 'hollandse nuchterheid' and for being my paranymph! Els, you have been such a great housemate. Thanks, for providing me with proper nutrition after some long office-days and taking care of Spock while I was away on one of my many trips.

Last but not least, I would like to thank my family. Pap en mam, Tobias en Nathalie, bedankt voor alle steun en inspirerende woorden. Deze hebben mij keer op keer de kracht gegeven om, op de moeilijke momenten, toch door te zetten. Nathalie, bedankt voor je hulp als paranimf en alle steun tijdens de afgelopen jaren.

# List of publications

Peer reviewed journals

**Mulder, V.L.**, de Bruin, S., Weyermann, J., Kokaly,R., Schaepman, M.E., (Accepted for publication). Characterizing regional soil mineral composition using spectroscopy and geostatistics. *Remote Sensing of Environment*.

**Mulder, V.L.**, Plötze, M., de Bruin, S., Schaepman, M.E., Mavris, C., Kokaly, R., Egli, M., (2013). Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1-2.4 µm) and regression tree analysis. *Geoderma*, (107-108), 279-290.

**Mulder, V.L.**, de Bruin, S., Schaepman, M.E., (2012). Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. *International Journal of Applied Earth Observation and Geoinformation* 21,(1), 301-310.

**Mulder, V.L.**, de Bruin, S., Schaepman, M.E.; Mayr, T. (2011). The use of remote sensing in soil and terrain mapping - A review. *Geoderma* 162, (1-2), 1-19.

Other scientific publications

Bartholomeus, H.M., Roosjen, P., Clevers, J., Suomalainen, J., **Mulder, V.L.**, Kooistra, L. (2013). Estimation of soil clay content using multidirectional laboratory spectroscopy measurements. In: *In: Proceedings of the 8th EARSeL Imaging Spectroscopy workshop*, 8-10 April 2013. - Nantes, France.

Bartholomeus, H.M., Roosjen, P., Clevers, J., Suomalainen, J., **Mulder, V.L.**, Kooistra, L. (2013). Estimation of soil organic matter using multidirectional reflectance measurements. In: *Proceedings of the 16$^{th}$ International conference on Near Infrared Spectroscopy*, 2-7 June 2013. – Montpellier, France.

Bakker, A., **Mulder, V.L.**, (2012). Digital Soil Mapping: Predicting soil texture at a regional scale using remote sensing. *Geophysical Union (EGU) Conference*, 22-27 April 2012. - Vienna, Austria.

Bock, M., Köthe, R., Schuler, U., Günther, A., Pickert, E., Willer, J., Baritz, R., Zawadzka, J., Mayr, T., **Mulder, V.L.**, Moussadek, R., van Engelen, E., (2012). Soil and terrain modelling to develop medium to small scale conceptual soil maps. *Geophysical Union (EGU) Conference*, 22-27 April 2012. - Vienna, Austria.

**Mulder, V.L.**, de Bruin, S., Schaepman, M.E., (2012). Soil mapping at regional scale using Remote Sensing –integrating multiple research methods. *Geophysical Union (EGU) Conference*, 22-27 April 2012. - Vienna, Austria.

**Mulder, V.L.**, de Bruin, S., Schaepman, M.E., (2012). Retrieval of composite mineralogy by VNIR spectroscopy. In: *Proceedings 5th Global Workshop on Digital Soil Mapping 2012*, 10-13 April 2012. - Sydney, Australia.

**Mulder, V.L.**, de Bruin, S., Schaepman, M.E., (2011). Towards spectroscopic modelling of composite mineralogy. *9th Swiss Geoscience Meeting*, 11-13 November 2011. - Zürich, Switzerland.

**Mulder, V.L.**, de Bruin, S., Schaepman, M.E., (2011). The use of remote sensing for soil mapping at regional scale. *Soil Science in a Changing World*, 18-22 September 2011. - Wageningen, The Netherlands.

**Mulder, V.L.**, de Bruin, S., Schaepman, M.E. (2011). Soil mapping at regional scale using ASTER and VNIR spectroscopy. *In: Proceedings of the Second Global Workshop on Proximal Soil Sensing*, 15-18 May 2011. - Montreal, Canada.

**Mulder, V.L.**, Bartholomeus, H.M. (2009). The influence of slope in the quantification of soil iron content with spectral reflectance based iron indices. *In: Proceedings of the 6th EARSeL SIG IS workshop Imaging Spectroscopy: Innovative tool for scientific and commercial environmental applications*, 16-19 March 2009. - Tel Aviv, Israel.

**Mulder, V.L.** (2008). The influence of palaeo-climatic changes on fluvial landscape evolution during the quaternary: a case study of the Upper Thames basin, UK, with the LAPSUS model. *MSc thesis, Wageningen University*, 127p.

**Mulder, V.L.** (2007). The influence of slope in the quantification of soil iron content with spectral reflectance based iron indices. *MSc thesis, Wageningen University*, 66p.

# Short biography

Titia Mulder was born in Delft (The Netherlands) on the 26th of January, 1984. When she was one year old, the family moved to Barneveld. There, she attended primary school at the Spreng, and in 1996 she started secondary school at Scholengemeenschap Johannes Fontanus College. In 2003, she obtained her VWO diploma, after which she decided to move to Wageningen to study Soil, Water & Atmosphere, at Wageningen University.

In 2006, she obtained her BSc degree with a minor degree in Geo-Information Science and Remote Sensing and she continued her studies at Wageningen University. Not knowing whether she should study Soil Science or Geo-information Science and Remote Sensing she followed both curricula. Her major MSc thesis was about the influence of palaeo-climatic changes on fluvial landscape evolution during the quaternary applied to a case study of the Upper Thames basin, UK.

In the course of her studies she got interested into the combination of soil science with remote sensing. Therefore, she performed a second MSc thesis about the influence of viewing geometry in the quantification of soil iron content using imaging and laboratory spectroscopy. She finalized her studies with an internship at the European Space Research Institute of the European Space Agency (ESA/ESRIN), located in Frascati, Italy. She contributed to the project DesertWatch by evaluating the use of satellite retrieved soil moisture as indicator for desertification. In 2008, she obtained her MSc degree in Soil Science, specializing in Land Dynamics and a minor MSc degree in Geo-Information Science and Remote Sensing.

After MSc graduation she got the opportunity to work on this thesis where she could continue to work on the topics of soil science and remote sensing. In the course of five years, she worked on the topics presented in this thesis; she actively participated in various scientific conferences and workshops and was involved with MSc-level teaching.

Titia's research interest are related to spatial modeling of soil resources at large scales, with emphasis on the integration of remote and proximal sensing methods with geostatistics. This includes the design of sampling strategies, retrieval of soil information from laboratory spectroscopy, kriging approaches for large datasets and the development of soil prediction models.

# PE&RC PhD Training Certificate


With the educational activities listed below the PhD candidate has complied with the educational requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

**Review of literature (6 ECTS)**
The use of remote sensing in soil and terrain mapping - A review

**Writing of project proposal (4.5 ECTS)**
Retrieval of soil properties at various spatial, spectral and temporal scales using remote sensing (2008)

**Post-graduate courses (3 ECTS)**
- Hyper-I-net Summer School; GIRS (2008)
- Geostatistics; PE&RC (2010)

**Laboratory training and working visits (4.5 ECTS)**
- Exchange program with University of Zürich; RSL, University of Zürich, Switzerland (2009-2012)
- XRD analysis of soil samples; department of Geography, University of Zürich, Zürich, Switzerland and ETH – IGT – Institute for Geotechnical Engineering, Zürich, Switzerland (2011-2012)

**Invited review of (unpublished) journal manuscript (2 ECTS)**
- Geoderma: Regional predictions of soil properties using hyperspectral data (2011)
- Geoderma: Topsoil organic carbon prediction using VNIR-SWIR reflectance spectra at lab, field and laboratory level (2011)
- Applied and Environmental Soil Science: The effect of pre-treatment on chemometry using imaging spectroscopy (2012)
- Geoderma: Identification and mapping of Vertisols using SPOT 5 Images (2012)

**Competence strengthening / skills courses (4.5 ECTS)**
- PhD Competence assessment; WGS (2009)
- The art of writing; WGS (2009)
- Scientific writing; WGS (2009)
- Afstudeervak organiseren en begeleiden; EDU (2009)

**PE&RC Annual meetings, seminars and the PE&RC weekend (1.5 ECTS)**
- PE&RC Annual meetings: expect the unexpected (2008), intelligent nature: on the origin of communication (2009), selling science, why and how scientist sell science (2010), extreme life: exploring life in the extremes and the extremes in life (2012)
- Other PE&RC Symposia: 3rd Dutch remote sensing symposium (2011)

**Discussion groups / local seminars / other scientific meetings (7.5 ECTS)**
- Spatial Methods discussion group (2008-2011)
- Remote sensing thematic group (2008-2011)
- e-SOTER project meetings (2008-2011)
- PhD Defence mini-seminars (2009, 2011, 2012)
- Invited lecture Agroscope, Zürich: the use of remote sensing for soil terrain mapping at regional scale (2011)

**International symposia, workshops and conferences (9 ECTS)**

- 6th EARSel SIG IS Workshop; poster presentation; Tel Aviv, Israël (2008)
- 2nd Global Workshop on Proximal Soil Sensing: oral presentation; Montreal, Canada (2011)
- Soil Science in a Changing World; poster presentation; Wageningen, the Netherlands (2011)
- 9th Swiss Geoscience Meeting (SGM); poster presentation; Zürich, Switzerland (2011)
- 5th Global Workshop on Digital Soil Mapping; oral presentation + reporter for session; Sydney, Australia (2012)
- European Geoscience Union, General Assembly 2012; organizing session: digital soil mapping: novel approaches and sensing techniques to the prediction of key soil properties; Vienna, Austria (2012)

**Lecturing / supervision of practical 's / tutorials (3 ECTS)**

- Remote sensing and GIS integration; 3 days (2009)
- Introduction to Geo-information science; 15 days (2010)

**Supervision of 2 MSc students (6 ECTS)**

- Digital soil mapping of soil texture at a regional scale with the use of remote sensing; 7 days
- Mangrove's structural complexity and disturbances; a case study of the Berau Delta, East Kalimantan, Indonesia; 3 days