Faculty of Agricultural Science

Institute of Crop Science

University of Hohenheim

Fertilisation and Soil Matter Dynamics

Prof. Dr. Torsten Müller

Development of MidDRIFTS methodologies to support mapping of physico-chemical soil properties at the regional scale

Dissertation

Submitted in Fulfillment of the requirement for the degree

“Doktor der Agrarwissenschaften”

(Dr.sc.agr./Ph.D. in Agricultural Science)

to the

Faculty of Agricultural Sciences
Presented by

Reza Mirzaeitalarposhti

Ghaemshahr, Iran

2014

This thesis was accepted as a doctoral dissertation in fulfillment of the requirements for the degree "Doktor der Agrarwissenschaften" (Dr.sc.agr. / Ph.D. in Agricultural Sciences) by the Faculty Agricultural Sciences of the University of Hohenheim on 05.08.2014.

Date of oral examination: 30.09.2014

Examination Committee

Supervisor and Reviewer	Prof. Dr. Torsten Müller
Co-Reviewer	Prof. Dr. Georg Cadisch
Additional Examiner	Prof. Dr. Ellen Kandeler
Head of the oral examination	Prof. Dr. Markus Rodehutscord

Acknowledgments

The writing of this dissertation has been hard work, ridiculously encountered difficulties at times, but immensely satisfying in the end. During this doctoral journey I have met so many people who helped and encouraged me to fulfill my PhD work. So, it is a great pleasure to appreciate all these nice people who have contributed in some way to accomplish this dissertation directly or indirectly.

Foremost, it is with immense gratitude that I gratefully acknowledge my direct supervisor, Prof. Dr. Torsten Müller, for accepting me to join in his research team and for his confidence, initiating the research, and support along the way. I am extremely grateful and indebted to him for his expert, sincere, and valuable guidance and encouragement extended to me.

My sincere thanks also goes to Prof. Dr. Georg Cadisch, as my second supervisor, for the continuous support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I also wish to express my sincere thanks to Prof. Dr. Ellen Kandeler for her readiness to be part of my thesis examiner and defense.

It gives me great pleasure in acknowledging the support and help of Dr. Frank Rasche during the thesis writing. His advice and comments on paper writing was always constructive and informative and encouraged me to look for ways to find a better solution.

I take this opportunity to thank to my close collegue, Dr. Michael Scott Demyan for his kind cooperation and guidance during field and lab work and thesis writing. It has been great to have always quick response from him to my questions and postings. We spent lots of time together in field and laboratory with lots of lasting memories.

I would also like to thank Dr. Homan Liaghati for his constant support and help before starting my PhD in Germany. He intiated contacted with my supervisor to recommend me for studying at Hohenheim.

I am gratefull to my senior colleques Prof. Dr. Ismaeil Shahbazi & Dr. Hossein Kermanian who supported me all the time and encouraged me to continue studying abroad.

I also thank the Science, Research and Technology Ministry of Iran for awarding the scholarship for studying in Germany.

Many thanks are also given to the technicians at Institute of Fertilization and Soil Matter Dynamics (340i) for providing me all support and material preparing and sample measurements during my laboratory work, Mrs. Ruckwied, Mr. Bucher, Mr. Bremer, Mrs. Haake, Mrs. Zimmermann and Mrs. Ochott.

Very special thanks go to Mrs. Schöllhammer and Mrs. Berghammer two nice and kind personel at Institute of Fertilization and Soil Matter Dynamics (340i) for their continuous help and support in terms of all official and social matters I was dealing with.

My appreciation also goes to my all colleagues and friends at my institute and the University of Hohenheim.

Lastly I am very grateful to my family and especially my parents, for their support, encouragement and patience. Last but not the least; I would like to thank Mehrangiz, my wife, for her love and support, helping me get through this process and for providing emotional support and care and for being a constant source of inspiration. Without her, I would have been lost a long time back. I would also like to acknowledge my lovely daughter, Aynaz, the absolute joy of my life.

Table of Contents

Abbreviations

C: carbon

C:N ratio: The ratio between total organic carbon and total nitrogen

CFE: Chloroform fumigation extraction

C_{HWE}: hot water extractable C

C_{mic}: Microbial biomass C

CSM: Conventional soil mapping

CV: Coefficient of variation

DSM: Digital soil mapping

FTIR-ATR: Fourier Transform Infrared-Attenuated Total Reflectance

GCV model: Generic cross-validated models based on PLSR

GIC model: Generic independent calibrated models based on PLSR

IRS: Infrared spectroscopy

K: Kraichgau study region

K_CV: Region specific cross-validated model for Kraichgau region

MD: Mahalanobis distance

MDT: MD threshold

MidDRIFTS: Mid-infrared diffuse reflectance infrared Fourier transform spectroscopy

MIRS: Mid-infrared spectroscopy

MLR: Multiple linear regression

MSC: Multiplicative scatter correction

MSE: Mean square error

N: Nitrogen

N_{HWE}: hot water extractable N

NIRS: Near-infrared spectroscopy

N_{mic}: Microbial biomass N

N_{min}: Mineral N

OK: Ordinary Kriging

PCR: Principal component regression

PLSR: Partial least square regression

R^2: Coefficient of determination

$RMSE_{CV}$: Root mean square error of cross validation

$RMSE_P$: Root mean square error of prediction

RMSSE: Root mean square standardized error

RPD: Residual prediction deviation

RSS: Residual sum of square

SA: Swabian Alb study region

SA_CV: Region specific cross-validated model for Swabian Alb region

SLF: Subtraction of a linear function

SOC: Soil organic carbon

SOM: Soil organic matter

Std.dev: Standard deviation

TC: Total carbon

TIC: Total inorganic carbon

TN: Total nitrogen

TOC: Total organic carbon

V/NIRS: Visible/Near infrared spectroscopy

VN: Vector normalization

WHC_{max}: Maximum water holding capacity

Chapter 1

General Introduction

1. General Introduction

1.1. Background

The soil or pedosphere, which is the upper layer of terrestrial ecosystems, plays an important role as the basis for terrestrial life on earth. Soil is a complex medium which consists mainly of organic compounds namely soil organic matter (SOM), inorganic materials (clay minerals, carbonate) and living soil organisms like microorganism, plant and animal species. Soils have a fundamental role regarding ecosystem services and sustainability, water balance and carbon cycling. Besides, as a biological regulator, protect the environment and buffer extreme natural environmental pressures imposed upon it (Jones, et al. 2012). Soil is a biologically active and dynamic system where the matter and energy are interchanged between the soil and surrounding atmosphere, biosphere and hydrosphere continuously. These matter flows are important for soil profile development and play a crucial role in influencing soil quality. Soil is seen as a fundamental natural resource which plays an important role in food, fiber and energy production to meet the needs for human life. Increasing world population accompanied with rising global demand for food and energy, has placed tremendous pressure on natural resources, especially agricultural land (Brown, 2006; Zaehle, et al. 2007). Consequences of global climate changes i.e., elevated atmospheric CO_2 and increased air temperature (IPCC, 2007) affect agricultural ecosystems by influencing overall plant biomass production, and allocation of carbon to the soil pool (Davidson and Janssens, 2006). Additionally, climate change alters the balance of matter and energy fluxes to or from soil to the surrounding environment and may change the soil from a sink to a source of atmospheric carbon. Soils of terrestrial ecosystems are an important reservoir of both organic and inorganic carbon (2500 Pg of soil organic and inorganic carbon (SOC and SIC, respectively) pools to 1 m depth), which has marked influence on the global C cycle (Batjes, 2002; Grace, 2004). In addition, the ability of the terrestrial sink (temperate latitudes) in taking up the atmospheric C has been estimated 2-3.4 Gt C yr^{-1} depending on the source availability and environmental condition (Tans, et al. 1990). Since the Kyoto Protocol (UN, 1998) highlighted the crucial role of SOC on environmental processes, soil's potential sink or source for atmospheric CO_2 has received a great amount of interest as methods to mitigate climate change have been proposed (Metz, et al. 2007; Davidson and Janssens, 2006). Although both agricultural and forestry activities have greatly contributed to CO_2 release into the atmosphere (Lal, 2008), agricultural soils can also act as a net C sink to partially mitigate the increasing CO_2 emissions.

Due to the importance of agricultural soils and to protect important functions related to the soils, regular monitoring of essential soil properties such as SOC, nutrient status and soil pH are fundamental to detect changes in their status at various scales. The knowledge is needed to implement appropriate soil management in agricultural production and environmental protection (Du and Zhou, 2009). However, soil properties often vary greatly in space and time (Minasny, et al. 2006) which necessitates intensive sampling campaigns to explain spatial/temporal heterogeneity (Lal, et al. 2004). Characterizing soil properties relies on costly and time intensive conventional analytical approaches which are inappropriate for fast or in situ evaluation of soil properties (Du and Zhou, 2011). Therefore advanced quantification methods are needed to enhance our ability to assess spatial variability and mapping of soil properties using geo-spatial techniques such as geostatistical analysis (Shepperd and Walsh, 2002; Viscarra-Rossel, et al. 2006). Mid-infrared spectroscopy (MIRS) in combination with multivariate calibration has shown predictive potential of a variety of soil properties such as different SOM fractions and texture at various scales (Janik, et al. 1998; Reeves, et al. 2001; Viscarra-Rossel, et al. 2006). The technique has potential for rapid assessment of comprehensive soil properties to assess heterogeneity of individual soil properties from farm up to landscape level.

This dissertation is concerned with the application of new methodological approaches in quantifying soil physical and biochemical properties which provide baseline information for better understanding of soil fertility and quality and spatial variability at regional scale in south-west Germany considering two different agro-ecological regions. Therefore, promising outputs would offer an improved, high spatial resolution database which could be used for recommendations on land management and provide an advanced strategy for regular soil monitoring and assessment in future.

1.2. Importance of SOM in arable lands

Soil organic matter (SOM) is one of the most important factors in agricultural soils directly related to soil fertility for plant production. SOM is mainly derived from plant and animal residues which decomposed to pools with substantially different decomposition rates via microbial activity (Rodeghiero, et al. 2009). SOM has beneficial effects on soil structure and stability (physical properties), soil bio-chemical properties, biodiversity, water retention and it also acts as a main nutrient source for plants and microorganisms. SOM in nature is heterogeneous in terms of quality, quantity and availability. It also has a wide degree of stability ranging from most easily decomposable fractions up to compounds which are physically resistant to decomposition with turnover times ranging from days to thousands of years (Amundson, 2001; McLauchlan and Hobbie, 2004). Therefore, considering the heterogeneity of SOM, understanding the processes controlling SOM turnover to different changing factors (e.g. temperature, moisture, and management) is complex. Therefore, multiple SOM conceptual pools have been developed to assess the C dynamics in ecosystems (Jones, et al. 2005).

Among the gradients present in soil that influence the soil quality, soil organic carbon (SOC) is probably one of the most important which is directly or indirectly related to almost all physical, chemical and biological properties (Udelhoven, et al. 2003). Approximately 10% of the land surface is arable land (1.4 billion ha) containing 10 % of the terrestrial SOC stocks in surface soils (WRB, 2007; Jenkinson, et al. 1991). The biologically active carbon pool contributes mainly to soil fertility and provides nutrients for plant production (Wander, et al. 1994). Roughly 45% of mineral soils across the European Union are low or very low (0-2%) in organic carbon content (Rusco, et al. 2001). It has been reported that SOM and its consequent fertility is diminishing in farmlands due to inadequate farming practices (Gobeille, et al. 2006), inadequate fertilization and soil erosion. This has resulted in accelerated SOC mineralization and C release into the atmosphere. On the other hand SOC is highly related to soil texture and structure. The capability of soil to store carbon is highly affected by the quantity of fine mineral particles such as clay minerals (Hassink, 1997). Therefore, soil texture has significant effects on the residence time of SOC. Currently the factors affecting SOC are not sufficiently understood to derive management strategies to maintain SOC levels in the long term. Besides, the current knowledge on essential soil properties such as SOC and texture is either outdated or available in relatively low resolution only. Therefore, it is necessary to increase our knowledge about SOC and textural classes and interactions with environment and human practices to better understand matter cycling (i.e. C and N) in ecosystems. Furthermore, separating SOM into different fractions with different stability can provide appropriate data to simulate soil organic C dynamics using multi-pool models. Application of MIRS techniques may provide appropriate predicted results as well as spectral-based indicators to quantify SOC of different qualities and turnover rates in arable lands.

1.3. Soil spatial heterogeneity

Spatial heterogeneity is common in soils at different scales (Gallardo, et al. 2000) as a result of interactions between the five soil forming factors at different spatial scales i.e. climate, topography, parent material, organisms and land use management (Jenny, 1941; Gallardo, 2003; Liu, et al. 2010). These together have formed both random and systematic variations in the soil landscape. While random variability would be due to possible variability at the small scale as a result of management, the systematic variation can be attributed to significant effect of soil forming processes (Jenny, 1941). Variability in soil properties can take place from small scale (few millimeters) up to very large scale at the regional level (Garten Jr, et al. 2007). Anthropogenic influences such as farm management strongly effect soil variability especially at farm level which linked to field micro heterogeneity and differences in soil types. Up to present, the spatial variability of soil properties in diverse agro-ecological areas still remains poorly understood due to a general lack of high quality data. If the difficulties in measuring spatial variability can be overcome, the knowledge will provide a valuable insight on soil properties which can be further used for better ecosystem management to improve the system quality and performances (Tittonell, et al. 2007). Knowledge of spatial variability may be useful to predict soil properties in un-sampled locations via interpolation resulting in geostatistical soil maps at different scales (Wei, et al. 2008). Interaction between different soil properties and environmental factors can be understood and decisions can be made for improving natural resources' utilization efficiency and for the protection of the environment against unwise land use managements (Liu, et al. 2004; Huang, et al. 2006; Wei, et al. 2008). Furthermore, recognizing the spatial variability is important for designing efficient sampling strategies in future surveys (Rossi, et al. 2009).

1.4. Soil spatial modeling and mapping at regional scale

1.4.1. Conventional and digital soil mapping

Conventionally, soil maps were created based on the free survey method termed conventional soil mapping (CSM). First, mental soil-landscape models were constructed based on pre-existing knowledge. After determining the soil boundaries, the possible soil field observations were made. Thereafter, based on limited field observations, mental soil-landscape models were reproduced. Generated maps were mainly general-purpose with different soil classes and representing soil surveyors' knowledge about soil profiles and objects at each map unit (Burrough, et al. 1971). As it was based on limited numbers of sampled data in the field, inconvenient for large and inaccessible areas and qualitative-based map units or unknown quantitative accuracy were present in the maps. Then, to produce accurate and fine-resolution soil maps, intensive sample collection and data was needed which was difficult to acquire for inaccessible and large-scale areas (Bui, et al. 1999). Besides, generated soil-landscape models are implicit and not well-documented and intuitive. Therefore reproduction of these models which is a central part of mapping in the conventional approach is very difficult (Burrough, et al. 1997; Zhu, 1997). Paper-based soil maps are static and present a quality nature of variable without any quantified measures of accuracy (Hewitt, 1993; McBratney, et al. 2003). Therefore, there is a need for more flexible and quantitative methods to generate accurate and detailed soil spatial information to update currently existing maps obtained by conventional approaches (Hartemink and McBratney, 2008).

Digital soil mapping (DSM) is the creation of spatial information systems using field and laboratory measurements coupled with quantitative statistical methods (models) at various spatial and temporal scales (McBratney, et al. 2003; Grunwald, 2010). These statistical methods or models are referred to as pedometrics relate measured values of soil properties to continuous environmental information for prediction of soil spatial distributions (Kempen, et al. 2011). The pedometrics used in DSM are mainly based on empirical models, Fuzzy-set theory, geostatistical, or hybrid techniques (McBratney, et al. 2000). Empirical models are based on the main soil forming factors i.e. climate, organisms, relief, parent material and time (CLORPT) and legacy soil data. In this way, appropriate empirical models such as regression trees, simple and multiple linear regressions, generalized adaptive models and neural network models can be developed to correlate soil observations with environmental variables describing the main soil forming factors. The environmental variables can be taken from geo-referenced data layers, including geo-referenced digitized maps, aerial photos, and satellite images (McBratney, et al. 2003; Grunwald, 2011). Fuzzy-set theory provides a quantitative approach for predicting soil types and properties (Zhu, et al. 2001). While geostatistical methods include mainly the kriging estimator (simple and ordinary), hybrid techniques are combinations of geostatistics and CLORPT methods including universal kriging, cokriging, regression kriging, and factorial kriging (Goovaerts, 2001). The environmental data used in DSM for model development should comprise the main soil forming factors which are most relevant in assessing spatial variability of soil properties under evaluation. A well fitted model can be used later to estimate soil properties in un-sampled locations for interpolation (Kempen, et al. 2011). DSM is quantitative, more accurate, and flexible as compared to CSM. However, empirical models are mainly based on the feature space of the study area (boundaries) which may not apply for areas outside of boundary (Grunwald, 2009). Secondly, DSM maps can be developed for specific purposes as compared to the more general CSM (Kempen, et al. 2011).

1.4.2. Spatial modeling and digital soil mapping using geostatistics

Geostatistics is one of the most common used tools to assess soil spatial variability and to generate DSMs (Goovaerts, 1997; Webster and Oliver, 2001; McBratney and Pringle, 1997). Soil properties strongly vary in space and time with a huge amount of fluctuation occurring even over short distances (Lal, et al. 2004). Spatial variability is affected by several physical, chemical and biological processes at various spatial scales. Assessing spatial variability is fundamental for better understanding of the complex relationships between soil properties and environmental factors. The computed spatial model can be used later to estimate the soil properties in un-sampled locations to produce kriged maps. Interpolation is performed by using a kriging estimator based on a model of spatial dependence called a semivariogram. An example of an experimental semivariogram is presented in Figure 1.1. Three parameters are used to describe and compare spatial variability of soil properties. The range is the distance where the model reaches a horizontal asymptote. Sample locations separated by distances < range are spatially autocorrelated, whereas locations separated by distances > range are not. The value that the semivariogram model attains at the range distance (the value on the y-axis) is the sill. The partial sill is the sill minus the nugget. The nugget variance is the semivariogram value at zero separation distance (lag = 0) or the y-intercept (Goovaerts, 1997; Heuvelink and Webster, 2001). It is an indicator of short distance variation which includes measurement error, sampling error and unexplained and inherent variability (Marchetti, et al. 2012). Intensity of the spatial dependency is determined by the nugget-to-sill ratio (Cambardella, et al. 1994) where a ratio close to 0 indicates very strong spatial dependency, while close to 1 represents no spatial dependency. The application of geostatistics in generating soil maps has advantages as compared to empirical methods in DSM. Geostatistics is mainly based on intensive field and lab geo-referenced observations which allow its transferability to other areas while the environmental knowledge is considered as central part of mapping based on empirical methods. But, additional environmental information can be used as auxiliary data to improve the maps. Geostatistics is able to model both structured and random variation of properties in space (Hengl, et al. 2007). Although, DSM can generate accurate spatial soil information at smaller scales, successful application of empirical models at large scales is challenging because of increasing data variability and spatial domain. By increasing the scale and extent of the study area, the accuracy of relation between target variable and predictor variable decrease (Minasny, et al. 2013). If there is an appropriate sample size, DSM using geostatistical methods may provide accurate results for large-scale evaluation. Nevertheless, the main limitation involved in geostatistical methods is that the method requires large number of samples. Construction of reliable semivariograms captures both structured and random variation properly requires at least 100 observations (Webster and Oliver, 1992). Although this samples size may provide accurate estimation at small scales such a single field, larger scales require greater sample size with appropriate spatial distribution. Thus finding possible alternatives and strategies to cope with its major limitations for large scale soil mapping such as data acquisition via higher sampling density for DSM are crucial. According to Lagacherie (2008), these limitations might be overcome via i) optimal sampling strategy, ii) utilization of legacy soil data as much as possible and iii) developing new rapid and low-cost methodological approaches in data collection to reduce time and expense involved in conventional analytical approaches. Appropriate methods to overcome the limitations of data collection could be remote sensing (e.g., digital elevation models), proximal sensors (e.g., MIRS, NIRS) and digital elevation models. By using these data sources, more high quality data can be produced.

Therefore, to assess a successful large-scale spatial variability and to create up to date soil maps based on geostatistical methods, three main issues were taken into account in this dissertation: i) an efficient sampling strategy in terms of size and distribution needed to be conceived, an ii) advanced data acquisition approach was needed to create the large dataset required in geostatistics, and iii) employing the most accurate estimator in geostatistical (i.e., simple kriging and/or ordinary kriging to improve the quality of final kriged maps.

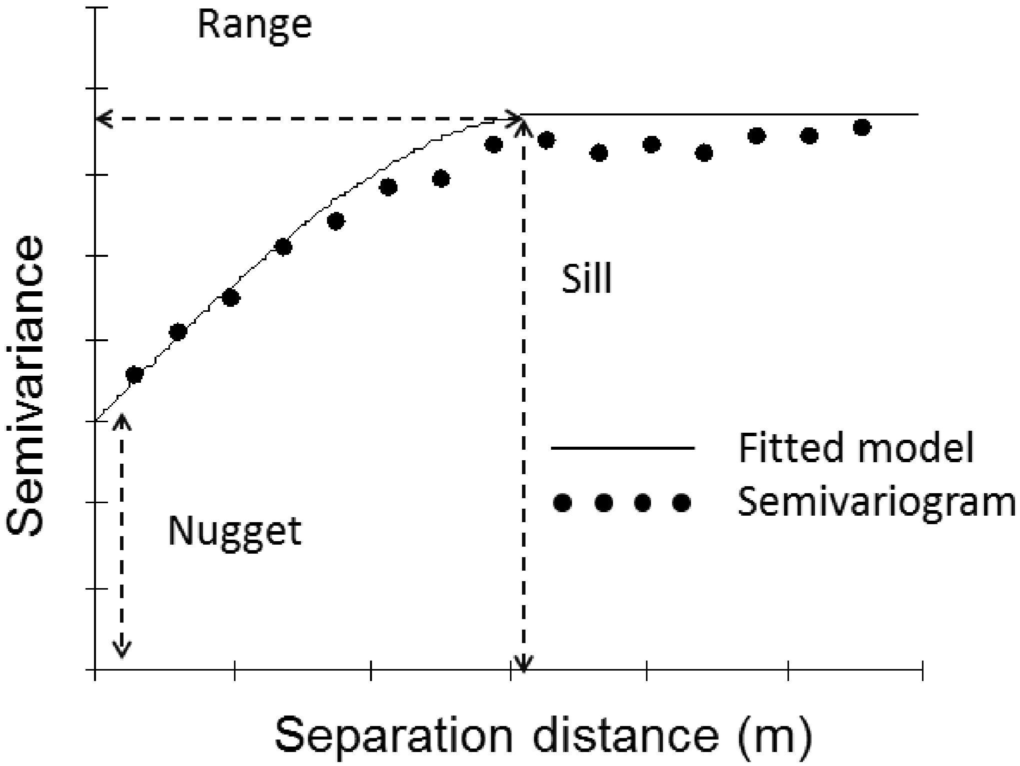

Figure 1.1. Theoretical example of an experimental semivariogram showing spatial dependency. The points are the mean values of all pairs separated by a certain lag distance.

1.5. Soil monitoring and quantification of soil carbon pools at regional scale

1.5.1. Limitations of classical analytical procedures

Estimating changes to and variability of SOC stocks at the regional scale requires appropriate amounts of geo-spatial data and sufficient environmental information of the ecosystem to which it is applied (Paustian, et al. 1997). Therefore, better understanding of SOC pools in agricultural ecosystems to update our information using a regular monitoring system to track changes is urgently needed to implement appropriate and sustainable soil management. Further, the correlation of SOC pools and influencing environmental factors are less understood at large scales (Lal, 2004b; Chaplot, et al. 2010). Therefore, development of quantitative methods to characterize essential soil properties such as C and N pools and texture at regional scale is receiving great interest. Trading of carbon credits and C-sequestration incentive programs will be reliant on methods that can quantify C in a relatively in-expensive way for large areas (Gehl and Rice, 2007). Monitoring soil carbon, nutrient status, soil pH and salinity are examples which greatly rely on a large amount of spatially dataset for monitoring and variability analysis.

Classification of various SOM fractions using conventional laboratory analysis have already been well-established to understand SOC dynamics as these methods provide a very useful indicator to assess soil quality. Several methods have been developed to quantify soil C and N in different pools including physical fractionation (Zimmerman, et al. 2007), chemical extractions (oxidation, acid hydrolysis) (Haberhauer and Gerzabek 1999), water extraction (Schulz and Körschens, 1998), chloroform fumigation (Joergensen and Mueller, 1996) and dry combustion (ISO 10694, 1995). In spite of reliable estimation of relevant soil properties, these approaches are costly and labor-intensive, sometimes using hazardous chemicals and producing chemical wastes limiting their application where large number of dataset are required (Malley, et al. 1999; Reeves, et al. 2001). For instance, measuring TOC using dry combustion needs separate determination of organic and inorganic carbon (Kimble, et al. 2001). Measuring soil carbonate using Scheibler's volumetric method (ISO 10693, 1995) is suspected to underestimate carbonate content in dolomite containing soils, as dolomite reacts slowly with 10% HCL. Using Loss-on-ignition method overestimates total SOC content as the methods depends on weight loss, which comes from not only organic matter, but also clays, other minerals and water (Cambardella, et al. 2001). In addition, deviation from the "true" value can increase when working with destructive soil chemical analyses (Huang, et al. 2005). It can be concluded that soil monitoring programs at regional scale remain still challenging and rather limited if applying costly and labor intensive conventional analytical procedures (Du and Zhou, 2009). Therefore, the need for rapid and inexpensive techniques to quantify the different pools of SOM and soil physical properties is needed to improve our understanding of the dynamics and spatio-temporal variability of SOM in terrestrial ecosystems. Recently, studies have utilized techniques which require minimum sample preparation, are relatively inexpensive and rapid and are able to characterize SOM predict other soil properties like texture, inorganic carbon soil pH etc. (Shepperd and Walsh, 2002; Viscarra-Rossel, et al. 2006).

1.5.2. Application of infrared spectroscopy in soil science

Research in infrared spectroscopy (IRS) (region from 10000-625 cm-1) goes back to early part of the 20th century, when Coblentz (1905) obtained spectra on hundreds of organic and inorganic compounds. Since that time, many developments have been made concerning commercial and research applications. The technique is seen as a reliable analytical tool for structural elucidation and compound identification over the last decades with a noticeable advancement in statistics, computing power and software (Shepherd and Walsh, 2004). The first application of IRS in soil research was in the 1960s by measurement of soil spectra for qualitative analysis of SOM, moisture and texture using the technique in the near-IR region (Bowers and Hanks, 1965). Later Hunt and Salisbury (1970), established a spectral database in laboratory to investigate mineralogical composition in soils using visible-near infrared reflectance spectroscopy. After that, the first spectral database was made in the early 1980s to investigate soils in U.S.A. established by Stoner and Baumgardner (1981). In the mid-1980s, the application of IRS entered a new phase with the development of new analytical methods to quantify different soil properties as a rapid and low-cost approach. Currently, prediction of different soil properties at various scales has been receiving greater interest together with progress made in multivariate statistics and chemometrics, and new promising aspects of this technique emerged (Chabrillat, et al. 2013).

The infrared region refers to the part of the electromagnetic spectrum between the visible and microwave regions (0.7-1000 μm or 17000-10 cm^{-1}, 780-2500 nm) (Gremlich, 2000; McBratney, et al. 2003) (Figure 1.2). Generally, the spectrum is the vibrational modes of molecular bonds which have absorption at certain wavenumbers. Therefore, each single spectrum can be used to identify the variety of chemical compounds including both organic and inorganic substances using qualitative and quantitative spectral data analysis. The IR region of the electromagnetic spectrum is subdivided into the four regions, near IR (0.8-2.5 μm or 14000-4000 cm^{-1}), MIR (2.5-25 μm or 4000-400 cm^{-1}), thermal and far IR (25 μm^{-1} mm or 400-10 cm^{-1}). In soil science both near-Infrared (NIR) and mid-Infrared (MIR) spectroscopy have been extensively utilized as a low-cost and relatively simple alternatives to quantify soil properties such as carbon contents, soil texture, nitrogen, pH, cation exchange capacity, carbonates, electric conductivity, soil microbial biomass (Viscarra-Rossel, et al. 2006).

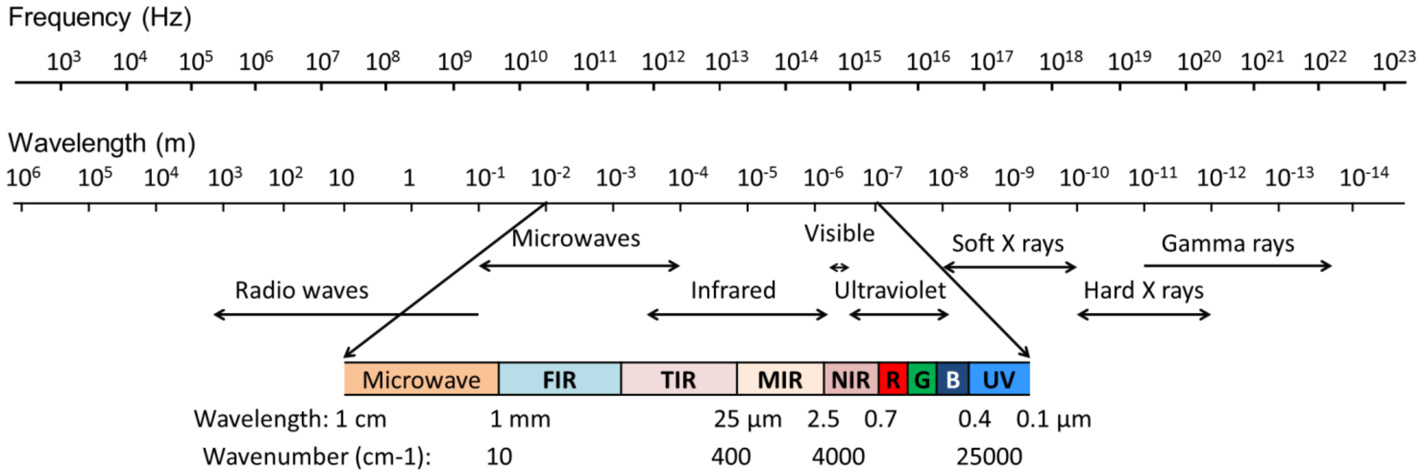

Figure 1.2. The electromagnetic spectrum highlighting boundaries for different infrared spectrum (modified from McBratney, et al. 2003). FIR: far infrared, TIR: thermal IR, MIR: mid infrared, NIR: near infrared, R: red, G: green, B; Blue, UV: Ultraviolet.

The technique minimizes time for sample preparation and is considered to be very rapid (high throughput). Each sample can be analyzed by IRS in a few seconds to obtain a single spectrum. It needs very little or no sample preparation depending on the spectroscopic instrument and sample condition. The sample only needs to be dried and ground (not needed for NIRS) before analyzing in the lab, although MIRS has been also applied on moist samples as well. The soil spectra can be achieved by using very small amount of soil sample approximately 100-200 mg and 25-100 g for MIRS and NIRS analysis respectively. It is also non-hazardous and non-destructive and the sample can be analyzed several times. It has wide-range applicability and a single spectrum can be used to predict the quantity of multiple soil properties (Barthes, et al. 2006). Although NIRS has been used in many studies to analyze multiple soil properties, the application of MIRS became more widely used in quantifying SOC and other soil properties (McCarty and Reeves, 2006; McCarty, et al. 2002; Viscarra-Rossel, et al. 2006). While the fundamental vibrations of organic and inorganic compounds absorb in the MIRS range, the peaks in NIRS range are mainly due to overtones and combinations of fundamental vibrations. Besides, there are some additional compounds which have vibrations only present in the MIRS range, such as soil carbonate (Pirie, et al. 2005) as compared to NIRS. Therefore, MIRS offers more information in its spectral range and the prediction is more precise than those obtained by NIRS, providing few and broad peaks which makes it very difficult for interpretation (Shepherd and Walsh, 2007). Restrictions of MIRS are mainly practical issues, as it requires more sample preparation and it can usually only be used in the laboratory (Janik, et al. 1998; Viscarra-Rossel, et al. 2006), although recently sensors have been developed for the field as well (Reeves, 2009).

1.5.3. Application of MIRS to quantify different soil properties

The MIR spectrum (4000 to 400 cm^{-1} wavenumbers) can be generally divided into fundamental vibrations and finger print regions. The fundamental vibration region extends from 4000-1500 cm^{-1} with several distinct bands representing vibrational modes of different functional groups. The finger print region extends from about 1450 cm^{-1} to 400 cm^{-1}, containing a complex series of peaks. The finger print region is assigned mainly to bending vibrational modes while the fundamental group region represents mainly the stretching vibrations. The finger print region is strongly influenced by the general molecular structure of the compounds and provides valuable information in investigating structural entities of compounds (Stuart, 2004).

Within the MIRS region, there are several spectral peaks (absorption bands) in MIRS soil spectra which can be assigned to different either organic or inorganic functional groups (Figure 1.3). The band from 3400-3300 cm^{-1} is mainly due to N-H stretching vibration (Baes and Bloom, 1989). The aliphatic C-H stretch is present at 3000-2850 cm^{-1} corresponding to labile organic carbon pools into the soil (Stevenson, 1982). The band at 1620 cm^{-1} (1660–1580 cm^{-1}) are assigned to aromatic C=C stretching (Baes and Bloom, 1989) and COO^- symmetric stretching (Stevenson, 1982). Aromatic C=C stretching vibrations and NH (amide II) bending vibrations are assigned to the band at 1520 cm^{-1} (1540–1503 cm^{-1}) (Stevenson, 1982), while the band at 1160 cm^{-1} (1172–1140 cm^{-1}) corresponds to the C–OH stretching of aliphatic alcoholic groups (Senesi, et al. 2003). In addition, soil minerals such as texture and carbonate also absorb in MIRS-range showing several peaks. The band of 3700-3500 cm^{-1} with two distinct peaks at 3695 cm^{-1} and 3622 cm^{-1} is assigned to the O-H vibration of clay minerals (Nguyen, et al. 1991). Additionally, there are several peaks from 1150-850 cm^{-1} which are assigned to clay minerals (Du, et al. 2007). The broad peaks between 2000 cm^{-1} and 1750 cm^{-1} are considered to represent soil non-clay minerals, mainly quartz in sand and silt relatively free from interference and overlapping (Nguyen, et al. 1991). The peak at 2686-2460 cm^{-1}, 1850-1784 cm^{-1}, 1567^{-} -295 cm^{-1}, 889-867 cm^{-1}, 734-719 cm^{-}

^{1}, and 719–708 cm^{-1} are assigned to carbonate vibrations (Tatzber, et al. 2010; Bruckman and Wriessnig, 2013).

Several FTIR-based techniques have been developed during the past decades. Transmittance was the previously preferred method which was based on the intensity of IR radiation transmitted through the sample. To analyze the soil by this method, it is necessary to prepare a pressed potassium bromide (KBr) disk using roughly an appropriate ratio of soil:KBr. In spite of clear and information-rich spectral features, this approach needs more time to prepare the disk and potentially reactions between the KBr and sample could occur in addition to water absorption. FTIR based on the Attenuated Total Reflectance (FTIR-ATR) method is another spectroscopic technique which obtains the spectrum on both solid and liquid samples and combats the most challenging aspects of infrared analyses, namely sample preparation and spectral reproducibility (Linker, 2011). One of the first applications of this technique in soil science was by Ehsani et al. (2001) to quantify nitrate concentrations in dry soil samples and was further developed by several researchers (Linker, et al. 2005; Borenstein, et al. 2006) to quantify soil nitrate using soil pastes samples. Because of strong spectral absorption in MIRS range the analysis of the spectra must be generated on KBr-diluted sample to avoid distorting the material under study (McCarty, et al. 2002). Nonetheless, KBr complicates sample preparation and has the potential to interact chemically with the sample i.e., ion exchange between sample and KBr and/or water absorption by KBr (Bertrand, et al. 2002).

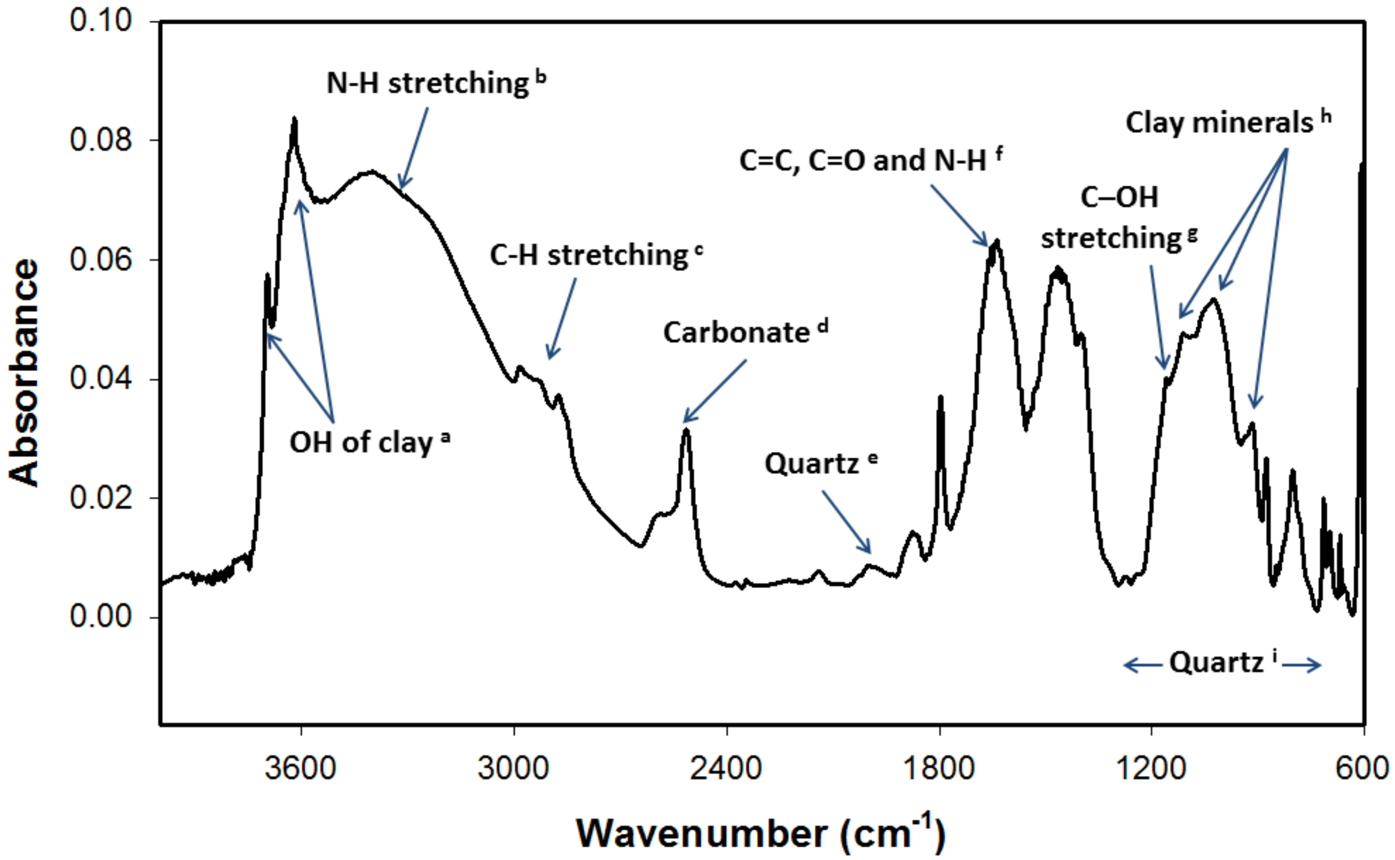

Figure 1.3. A typical electromagnetic spectrum in mid-infrared-range of a clay soil (65% clay, 2.7 % TOC, 3.5 % TIC) (chapter 2) representing main spectral features assigned to different soil organic and mineral compounds.

[a] Nguyen, et al. 1991

[b] Calderón, et al. 2011a

[c] Stevenson, 1982

[d] Tatzber, et al. 2007

[e] Nguyen, et al. 1991

[f] Stevenson, 1982, Baes and Bloom, 1989

[g] Senesi, et al 2003 [h] Madejová, 2003, Du, et al. 2007

[i] Madejová, 2003

Nevertheless, many scientists could apply MIRS on pure soil to identify variety of soil properties such as TC, OC, particle size, total sulfur, extractable Mn and exchangeable cations (Janik, et al. 1998; McCarty, et al. 2002). Diffuse reflectance infrared Fourier transform spectroscopy in mid-range (midDRIFTS) is one prospective spectroscopic technique requiring minimal soil preparation (Nguyen et al. 1991). In this technique water absorption and interference are reduced and resolution of spectrum improved as compared to transmission spectroscopy. In this technique, radiation first is emanated onto the sample surface and after absorption, refraction, reflection and scattering from the rough surface of the sample. The appearance of midDRIFTS bands in this technique are based on nonlinear scaling of intensity. Hence, magnitude of those strongly spectral bands will reduce when compared to low intensity bands and resolution of weaker bands will improve in midDRIFTS spectrum (Nguyen, et al. 1991). Although to avoid spectral distortions in midDRIFTS the sample may need to be diluted with KBr, however, quantitative analysis using midDRIFTS can be performed using pure soil samples (Janik, et al. 1998; Reeves, et al. 2001). Regardless of the method applied to analyse soil via spectroscopic techniques, quantitative mathematical analyses are required to predict target soil properties.

1.5.4. Quantitative analysis of MIRS

Spectral assessment of soil provides complex spectral data which can be related to a variety of soil properties over the entire infrared range. Therefore, quantitative and qualitative analysis of spectral data requires developing calibration models using appropriate statistical approaches. There are several sophisticated statistical techniques including multiple linear-regression (MLR) or other multivariate calibration methods such as principal component regression (PCR), and partial least squares (PLS) regression which can be used to create predictions from the spectral data for laboratory measured values (Workman, 1992).

The above approaches vary in the way a model is constructed to predict a soil property. In MLR a set of specific wavenumbers from the entire frequency range is selected and calibration is performed based on spectra data at these ranges for each single soil property (Bayer, et al. 2012). The principal component regression (PCR) is set up to utilize all or large portions of the entire frequency range. This approach first reduces the number of independent components required to describe the overall variability across the entire spectrum and between spectra and separate the spectral data into a set of scores. The analysis results in a few principal components (PCs) describing a majority of the sample set variability. Thereafter, regression analysis is made between each PC scores and measured values of desired properties separately for model calibration and validation. The accurate calibration and validation models developed independently are further used to predict soil property from new spectral data (Brereton, 1990).

Partial least squares regression (PLSR) is a multivariate calibration procedure reducing the spectral data into a lower dimensional subspace formed by a set of orthogonal latent variables to construct predictive regression models of measured soil properties (Wold et al, 1989). Like PCR, PLSR utilizes all or large portions of the entire frequency range in combination with concentration values of reference samples during model calibration and validation. But it is based on a single regression analysis between spectral data and soil property data. Increasing the irrelevant descriptors induces noise during PLSR and as a consequent the prediction power decreases. Thus to reduce noise and to increase model accuracy, a two-step descriptor selection procedure is applied. The first step is the elimination of the spectral data to a small number of latent variables (orthogonal factors), and in a second step, this subset of orthogonal latent variables is optimized to construct the model. This approach reduces noise and identifies outliers which do not fall in the calibration model range. The developed PLSR-based calibrated model can be applied to the spectra from new samples to predict the property of interest (Wold, et al. 1989). The PLSR-based calibration is considered to be more robust as compared to other calibration approaches such as PCR and MLR. It reduces noise from the data and handles multicollinearity better. There is no need to isolate specific spectral peak before calibration and the calibrated model will be performed even by small spectral variations to relate to the investigated soil properties (Haaland and Thomas, 1988).

Development of midDRIFTS-PLSR prediction models need an appropriate calibration approach i.e. independent validation and cross-validation. For independent calibration/validation the spectral data are divided into two separate subsets, one for model calibration and second is the validation dataset to independently validate the calibrated model. In the cross-validation approach called also leave-one-out cross-validation, the same dataset is used for model calibration and validation. Having a total number of N spectra, the calibration is performed using the N-1 samples and the single excluded sample is then used for model validation. The procedure is repeated N times where every single spectrum is used for model validation when it is left outside of the calibration. Generally cross-validation gives over-optimistic results and seems to be less accurate as compared to independent calibration/validation approach (Dardenne, et al. 2000; Islam, et al. 2003). Application of MIRS was considered as an advanced, high-throughput data collection strategy in the current study to provide an accurate dataset for successful large-scale mapping of different soil properties. The workflow in Figure 1.4 shows how midDRIFTS-PLSR can facilitate data collection to be integrated with geostatistics to improve digital soil mapping at regional scale.

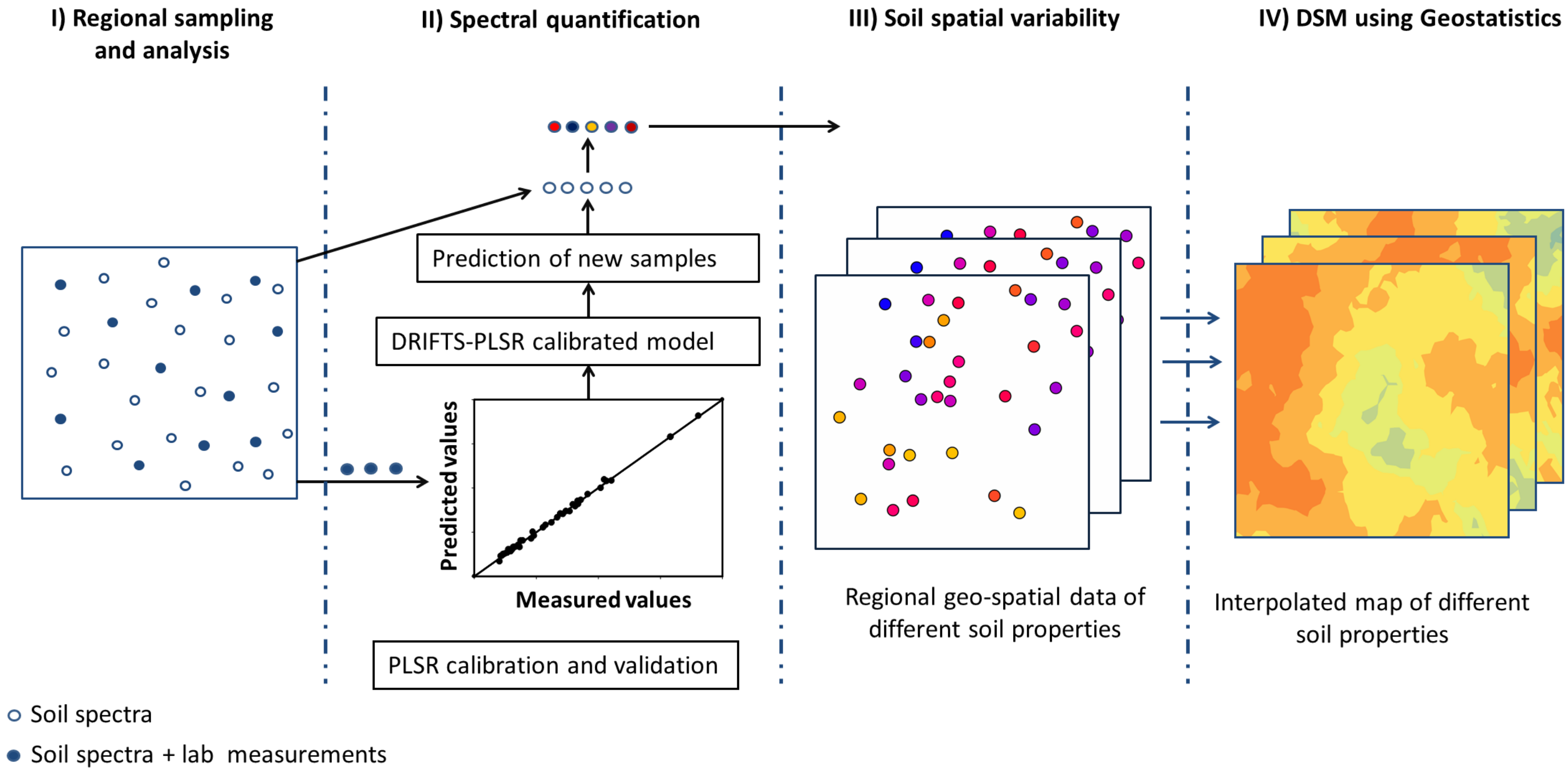

Figure 1.4. Digital soil mapping (DSM) workflow based on diffuse reflectance infrared Fourier transform spectroscopy in mid-range (midDRIFTS) coupled with geostatistical analysis. Colored dots referred as spatial soil heterogeneity.

1.6. Research hypothesis and objectives

As described above, characterizing soil properties over large spatial areas is challenging and needs intensive field and laboratory measurements because of the high variability of soil properties in space and time. Therefore, development of a new methodological strategy to acquire soil information via a more efficient, rapid and low-cost approach is of the upmost importance. Application of midDRIFTS could play an important role in collecting appropriate soil data for spatial modeling and creating soil property maps at the regional scale. To achieve this, the main objective of this survey was to assess the feasibility of midDRIFTS as a new analytical method in quantifying soil properties to better understand soil quality, to model spatial patterns and create soil property maps at the regional scale focusing on two different agro-ecological areas (40 by 40 km) in south-west Germany.

Regular soil monitoring at the regional scale requires high resolution physico-chemical soil spatial information. Hence, a multi-faceted approach was needed to address this issue. It was hypothesized that it is possible to quantify via generic midDRIFTS-PLSR calibrated models, a wide range of biochemical and physical soil properties from two regions differing in pedological conditions. Furthermore, contrary to the common perception it was further hypothesized that due to the extended data range by including both regions for PLSR-based model building, a generic independent validation approach would outperform region specific cross-validations due to the greater data range being included in the generic approach. As geostatistics requires a large number of georeferenced data, it was hypothesized that midDRIFTS-PLSR-based predicted data has the potential to be integrated with geostatisitcs for creating high resolution soil maps at regional scale. Finally, specific peaks in the mid-infrared range representing different functional SOM groups and soil inorganic compounds may provide appropriate indexes for indicating SOM quality/stability and composition at the regional scale without any PLSR calibration.

The experimental work based on the above formulated hypotheses is the subject of the following three chapters. In Chapter 2 the potential of midDRIFTS-PLSR is evaluated to predict SOM fractions and soil texture at the regional scale with emphasis on two calibration approaches i.e., independent validation, and cross-validation. In addition the efficacy of region specific cross-validation models in prediction of soil properties was compared with those of generic independent validation models.

The general objectives in chapter 3 were i.) to evaluate the large scale applicability of the most robust midDRIFTS-PLSR models (developed in chapter 2) to predict soil properties (TC, TIC, TOC, clay and silt contents) based solely on soil spectra at the regional scale and to ii.) create high quality maps of SOC and texture by integrating midDRIFTS-PLSR-based predicted data with geostatistics as an improvement over existing maps.

Chapter 4 introduces a further application of midDRIFTS spectroscopy, the specific peak area integration method, with the main objectives to generate midDRIFTS-based spectral indexes related to SOM quality and composition at the regional scale and to find the most relevant spectral index related to soil carbonate content as measured by Scheibler's method.

In chapter 5, the main findings of the study are synthesized and discussed in relation to the currently available knowledge and possible difficulties related to the research. Furthermore, an outlook and recommendations on future work are proposed.

This study was done in the framework of projects PAK 346 “Structure and Functions of Agricultural Landscapes under Global Climate Change-Processes and Projections on a Regional Scale” and FOR 1695 “Agricultural Landscapes under Global Climate Change –Processes and Feedbacks on a Regional Scale” within subproject P3. The overall objective of the projects was an integrated modeling of land-atmosphere and human-environment interactions at high spatial and temporal resolution to better assess the agricultural landscapes under future projected climate change, with the Kraichgau and Swabian Alb serving as two contrasting agricultural landscapes. The very complex land-atmospheric-socio-economic system requires a better understanding of the processes and mechanisms involved, which as a basis involves improving our information on soils. The work undertaken in this dissertation aimed to better understand the spatial nature of soil properties important SOM turnover modeling at the regional scale, which serves as an important link in modeling both plant growth and matter fluxes to the atmosphere under climate change.

Chapter 2

Quantifying physical and biochemical properties of soils from contrasting regions using midDRIFTS-PLSR - Generic independent validation outperforms region specific cross-validation

2. Quantifying physical and biochemical properties of soils from contrasting regions using midDRIFTS-PLSR - Generic independent validation outperforms region specific cross-validation

2.1. Abstract

Current knowledge on development of mid-infrared spectroscopic prediction models via partial least squares regression (PLSR) suggests that performance of models is highly affected by means of calibration and depends on the nature of the sample set. In this study we used diffuse reflectance infrared Fourier transform spectroscopy in the mid-range (midDRIFTS) to develop the most accurate PLSR models to quantify physical and biochemical properties of soils across two contrasting regions, Kraichgau (K) and Swabian Alb (SA), in Southwest Germany. We evaluated the capability of midDRIFTS-PLSR models in prediction of total carbon (TC), organic carbon (TOC), inorganic carbon (TIC), nitrogen (TN), mineral N (N_{min}), C:N ratio, hot water extractable C and N (C_{HWE}, N_{HWE}), microbial biomass C and N (C_{mic}, N_{mic}), pH, bulk density, clay, silt and sand content on a total of 126 samples using two different PLSR calibration procedures, i.e. cross-validation and independent validation. Generic calibrated models using either calibration approach were obtained from a set of 84 samples, while the remaining 42 samples were used for model validation. Cross-validation was further used to develop region specific PLSR models from a set of 42 and 21 samples for model calibration and validation respectively. Based on calibrated models, most soil properties were predicted successfully using either calibration approach with residual prediction deviation (RPD) >3 and coefficient of determination (R^2) >0.9, except for N_{min}, C/N ratio, pH, bulk density and sand. However, predictive performance of generic cross-validation models (GCV) in prediction of validation samples were considerably lower and unsuccessful compared with the generic independent validation derived models (GIC). Successful predictions were obtained only for TC, TIC, TOC, TN, silt and clay. In contrast, validation using GIC models) gave relatively the same predictive performance with those obtained in calibration except for N_{min} with RPD=1.4. Similarly, applicability of region specific cross-validated models in prediction of independent validation samples revealed considerably lower accuracies in both SA and K regions when compared to the GIC models. Successful predictions were only obtained for TC, TIC, TOC and TN in SA and TC and TIC in K. It was concluded that the GIC model provided the most accurate predictions on

independent validation sample sets as compared to both generic and region specific cross-validated models. Additionally, to calibrate robust region specific models, a higher number of samples with sufficient heterogeneity are required to allow an independent validation approach. Our results, thus, suggest a significant potential for the independent validation approach as a robust tool for predicting soil properties for soil fertility assessment, spatial analysis and modeling purposes without any further laboratory measurements. However, PLSR model can be developed further especially for those properties with low accuracies using an optimal selection of spectral peaks to integrate only the most informative peaks for model calibration.

2.2. Introduction

Recently, there has been an increasing interest of studying soil organic matter (SOM) in agricultural areas not only because of the substantial influence of SOM on nutrient availability and soil structure, but also because SOM plays an important role as a sink and source for atmospheric carbon dioxide (CO_2) (IPCC, 2007) Furthermore, climate change in addition to both land use and soil management changes is considered to alter SOM dynamics, consequently affecting role of soil as sink or source for atmospheric CO_2 (Degryze, et al. 2004; Lal, 2004a). Depletion of SOM across large scale has been proposed as a main consequence of climate change in soils (Bellamy, et al. 2005). However, different physically and chemically stabilized fractions of SOM react differently to these influences (Six, et al. 2002), which requires the measurement of not only total organic carbon (TOC) and nitrogen (TN), but of different fractions and stabilities of SOM. Nonetheless, characterising SOM pools remains challenging by common laboratory methods because of high heterogeneity of SOM both in space and time, usually requiring intensive field sampling (Malhi, et al. 2001; Baldock, et al. 2013).

As an alternative to the time and labour consuming classical laboratory analyses, several studies have utilized infrared spectroscopic approaches as a rapid-throughput and cost-effective technique to predict physical and biochemical soil properties (Viscarra-Rossel, et al. 2006; Shepherd and Walsh, 2007; Cobo, et al. 2010). Diffuse reflectance infrared Fourier transform spectroscopy in the mid-range (midDRIFTS) together with partial least squares regression (PLSR) is one prospective spectroscopic technique requiring minimal soil preparation (only drying and grinding) with lower time and cost requirements as compared to classical laboratory analyses. MidDRIFTS spectra of soil samples are related to the composition of both organic and mineral components in soils in the range from 4000 to 400 cm^{-1} (Baes and Bloom, 1989; Reeves III, et al. 2001). PLSR is a multivariate calibration procedure which reduces the spectral data into a lower dimensional subspace space formed by a set of orthogonal latent variables to construct predictive regression models of measured soil properties (Viscarra-Rossel and Behrens, 2010).

MidDRIFTS and PLSR has been applied on diverse soil data sets from farm and local scale up to continental scale. Reviews by Viscarra-Rossel et al. (2006) and Soriano-Disla et al. (2014) showed a wide range of soil properties predicted successfully by midDRIFTS-PLSR utilizing either cross-validation or independent validation approaches. While it has been shown that both cross-validation and independent calibration/validation approaches can be equally accurate (Morón and Cozzolino, 2002) and even cross-validation being preferred over independent calibration/validation (Shenk and Westerhaus, 1993), some others concluded that cross-validation results are over-optimistic when applied to unknown samples (Brunet, et al. 2007; Butkute´ and lepetiene´, 2006). Nevertheless, many studies have shown the usefulness of cross-validation in predicting a variety of soil properties using different sample sizes from 32 (Michel, et al. 2009) up to 1000 soil samples (Soriano-Disla, et al. 2013). Most recently, Vohland et al. (2014) predicted soil properties (TOC, TN, C_{mic}, C_{HWE} and soil pH) of 60 samples from different geological parent materials and agro-pedological zones using cross-validation. Moreover, independent validation has been used frequently as an appropriate approach linked to midDRIFTS-PLSR for many studies in prediction of various soil properties (Zimmermann, et al. 2007; Cobo, et al. 2010). Hence, the applicability of either calibrated method is not immediately evident. None of the above studies using midDRIFTS-PLSR have confirmed the absolute certainty of either approach. While the independent validation is widely used in data-rich situations (Varmuza and Filmoser, 2009), cross-validation is appropriate where limited data are available (Dardenne et al 2000; Bellon-Maurel and McBratney, 2011). Although development of accurate models require a certain minimum amount of soil heterogeneity (Cécillon, et al. 2009), increased soil heterogeneity (e.g. samples coming from different fields and study regions) may affect predictive performance of both cross-validation or independent validation (Bellon-Maurel and McBratney, 2011). Therefore, while a large amount of soil heterogeneity may lead to a bias in estimation via independent validation approach, dividing a dataset and using region or area specific cross-validation may result in overly optimistic assessments of the models showing a significant difference in accuracy (Nduwamungu, et al. 2009). It is important to assess cross-validation model performance using an independent sample set for verifying this model, which is lacking in the basic cross-validation approach and rarely carried out in previous studies. Due to these factors, it is not immediately evident if the best models would be obtained from cross-validation, independent validation, or a site specific model on a regionally distributed data set.

In the current study we hypothesized that is possible to quantify a wide range of biochemical and physical properties of soils derived from contrasting regions differing in pedological conditions using a generic midDRIFTS-PLSR model via either calibration approach. Furthermore, contrary to common perception we further hypothesized that due to the extended data range by including both regions for PLSR-based model building, a generic independent validation approach will outperform region specific cross-validations due to the greater data range being included in the generic approach. The main objective of our study was to test different approaches for midDRIFTS-PLSR models to quantify soil properties across two contrasting regions. Thus, to achieve this goal the specific objectives were i) to evaluate the efficacy of both generic cross-validation and independent validation approaches in quantifying midDRIFTS derived soil C and N pools, soil texture, bulk density, and pH for regionally distributed sample sets; and ii) to test the predictive performance of region specific cross-validation models with those of generic independent validation models.

2.3. Materials and methods

2.3.1. Study areas

The study areas were the Kraichgau (K) and Swabian Alb (SA) agro-ecological regions, in the Federal State of Baden-Württemberg, Southwest Germany. The distance between the centers of the two regions is roughly 120 km. The landscape of Kraichgau is situated in the north western part of Baden-Württemberg surrounded by the low mountain ranges of Odenwald, Strom, and Heuchelberg, the Black Forest and Rhine valley. The Kraichgau is a fertile and intensively cropped hilly region which is situated in a geomorphological basin between 100 to 400 m a.s.l. and mean annual temperature (MAT) between 9-10°C. The soil types in this region are mainly Leptosols and Luvisols (WRB, 2007). The mean annual precipitation (MAP) ranges from 720 to

830 mm. The fertile loess-derived soils and mild climate favour intensive agricultural activities. Approximately 83% of the total area is used for cropland mainly dominated by winter wheat, spring barley, winter rape, sugar beet as well as grain and silage maize (Ingwersen, et al. 2011). The Swabian Alb is a low mountain plateau of South-western Germany with elevations between 500-850 m a.s.l. The region is generally characterized by a cooler climate MAT of 6-7°C, MAP between 800 and 1000 mm with a high amount of summer rainfalls. The soils of the Swabian Alb developed mostly on Jurassic limestone, and are very rich in clay due to strong decalcification and weathering, and due the cooler climate also higher in TOC content (2-3.3%) compared to the Kraichgau soils (≈1%). Depending on relief position and weathering status of parent material, mainly shallow rendzic/calcaric Leptosols and chromic Cambisols or chromic Luvisols have developed in the region (WRB, 2007). Croplands make up to 52% of the agriculturally used land, while the remaining land area does not support more intensive types of land use because of dominance of shallow soils as well as steep slopes. Main crops are spring barley, winter barley, winter wheat and winter rape (Fischer, et al. 2010).

2.3.2. Soil sampling and processing

Soil samples were collected from farmers' fields at 42 locations in the two study regions (21 in each region). These fields were simultaneously used for a topsoil moisture monitoring system using time domain transmission sensors as a part of two succeeding integrated research projects. A nested sampling design was implemented at both regions, representing the full range of arable soils in the two regions. Three sampling points were selected in each field with a same crop resulting in a total set of 126 soil samples (63 samples for each region). Roughly one kg of soil (a composite of 5 sub-samples) was sampled to a depth of 30 cm at each sampling point using a stainless steel soil core sampler with 2.5 cm inside diameter. Soil samples were kept cool for transport to the laboratory. Samples were moist sieved to <4 mm and incubated at 50% WHC_{max} (maximum water holding capacity) at 20°C. A 20 g sub-sample was incubated in a Respicond IV^{TM} respirometer (Nordgren, 1988) under the same conditions to monitor CO_2 evolution. As soon as CO_2 evolution decreased to a stable basal respiration (latest reached after 72 hours), the bulk soil was sub-divided into two portions. A moist sub-sample was used for immediate biochemical analyses. The remaining sample was dried at 32°C, crushed and sieved (<2 mm) for further analysis.

2.3.3. Conventional analysis of soil samples

Hot-water extractable organic carbon (C_{HWE}) and nitrogen (N_{HWE}) were extracted from dry soil samples according to Schulz and Körschens (1998). Briefly, 20 g dried <2 mm soil (n = 2) was boiled under reflux for one hour with distilled water (1:5, w:v). After rapidly cooling to room temperature, the mixture was centrifuged for 10 minutes at 1936 rpm and the supernatant filtered through a 0.45-µm membrane filter (Minisart RC 25 syringe membrane filters, Sartorius, Göttingen, Germany). Hot water extracts were analysed for organic carbon (C_{HWE}) and total nitrogen (N_{HWE}) with a Multi N/C analyser (Analytik Jena, Jena, Germany).

Microbial biomass carbon (Cmic) and nitrogen (Nmic) were determined by chloroform-fumigation-extraction (CFE) (Joergensen and Mueller, 1996) on 20 g samples (n = 2) of <4 mm, incubated, moist soil. Briefly, two 20 g soil subsamples were taken. One portion was fumigated for 24 h at 25°C with ethanol-free chloroform ($CHCl_3$). Following CHCl3 fumigant removal, the soil was extracted with 80 ml of 0.5 M K_2SO_4 for 30 min with horizontal shaking at 200 rpm and filtered (No. 595 1/2 filter paper, Schleicher & Schuell GmbH, Dassel, Germany). The non-fumigated subsample was extracted similarly when fumigation commenced. The extract C and N concentrations were measured using a Multi N/C analyser (Analytik Jena, Jena, Germany). Cmic and Nmic were calculated as: Cmic (mg Cmic kg–1 soil) = EC / kEC, where EC = (OC extracted from fumigated soils) – (OC extracted from non-fumigated soils) and kEC = 0.45 (Joergensen and Mueller, 1996). Nmic (mg Nmic kg–1 soil) = EN / kEN, where EN = (TN extracted from fumigated soils) – (TN extracted from non-fumigated soils) and kEN = 0.54 (Joergensen and Mueller, 1996). Mineral nitrogen content (N_{min}) was taken as the sum of the ammonium and nitrate in the 0.5 M K_2SO_4 extracts of the non-fumigated soils, as measured by segmented continuous flow analysis (ALLIANCE Instruments, Evolution 2, Salzburg, Austria).

Bulk soil was analysed for TC by dry combustion according to ISO 10694 (1995) with a Vario-EL III elemental analyser (Elementar, Hanau, Germany). Scheibler's volumetric method was used to measure carbonate or total inorganic carbon (TIC) content (ISO 10693, 1995). In brief, about 0.5–1 g of ball milled and dried soil samples were treated with 10 ml of 10% HCl and connected to the gas tight Scheibler apparatus. Ambient released CO_2 derived air pressure and temperature were recorded for each sample in order to calculate the pT coefficient [pT=Air pressure in mm Hg/ (273+room temperature in °C)], which corrects the volume of CO_2 for current temperature and air pressure during the reaction. Total organic carbon (TOC) was calculated as the difference between soil total C and carbonate C (TIC) measured from Scheibler's volumetric method. Soil bulk density was determined using cylindrical core method in the field (Blake and Hartage, 1986). Soil texture was determined by the Bouyucos method and soil pH was measured in 0.01 M $CaCl_2$ with a soil:solution ratio of 1:2.5 (m:v) (Anderson and Ingram, 1993).

2.3.4. MidDRIFTS analysis of soil samples

Dried and sieved (< 2 mm) bulk soil samples were ball milled and maintained at 32°C overnight prior to midDRIFTS analysis according to Cobo et al. (2010) and Demyan et al. (2012). Briefly, a ball-milled subsample (roughly 200 mg) was scanned on a Tensor-27 mid-infrared spectrometer (Bruker Optik GmbH, Ettlingen, Germany) equipped with a potassium bromide (KBr) beam splitter and a liquid nitrogen cooled mid-band mercury-cadmium-telluride detector. A Praying Mantis diffuse reflectance chamber (Harrick Scientific Products Inc., New York, USA) was installed in the spectrometer and purged with dry air from a compressor to reduce interference from moisture and water vapour (Jun-Air International, Nørresundby, Denmark) with a flow rate of 200 litres hr^{-1}. Spectra were obtained in triplicates from each ball-milled sub-sample in the mid-infrared range (4000-600 cm^{-1}) by combining 16 individual scans at a resolution of 4 cm^{-1} recorded in absorbance units [log(1 /Reflectance)]. The spectra were first reduced to 3900 to 700 cm^{-1} to eliminate noise at both edges of each spectrum. Additionally, the CO_2 region (2400–2300 cm^{-1}) was excluded as it does not contribute to the actual soil spectrum. All replicate scans per sample were averaged for prediction purposes.

2.3.5. Spectral preprocessing treatments and midDRIFTS-PLSR model calibration

2.3.5.1. Preprocessing and sub-division of samples

All soil spectra were subjected to the different spectral pre-processing algorithms which included no data treatment, subtraction of a constant offset, subtraction of a linear function (SLF), vector normalization (VN), minimum-maximum scaling, multiplicative scatter correction (MSC), internal standard, 1st derivative, 2nd derivative, 1st derivative + SLF, 1st derivative + VN, 1st derivative + MSC (Michel, et al. 2009). Usually, pre-processing transformations of spectral data improve the accuracy of PLSR models as those tend to reduce noise, offset and bias in raw spectra (Yang and Mouazen, 2012). The mathematical pre-processing method was selected in the optimization step for each calibrated model of individual property. Seventeen data points were selected as gaps over either 1st or 2nd derivatives. In the current study, the optimal pre-processing transformation associated with the PLSR varied depending on the predicted soil properties. Dunn, et al. (2002) reported improvements of PLSR models by using 1st and 2nd derivatives while normalization and scatter corrections were used by (McCarrty, et al. 2002) and 1st derivatives by (Cobo, et al. 2010). However, little improvements of applying pre-processing to midDRIFTS-PLSR were obtained by Minasny and McBratney (2008) and Grinand et al. (2012). It seems as though a single pre-processing operation cannot generically be recommended and a variety of options must be tried in order to find the best result. Construction of PLSR models to predict a wide range of soil properties was done both for a generic model, including data from all sampling points, and for region specific models separately. Although the number of samples was sufficiently large (*n*=126 from 42 fields) to facilitate a generic calibration and independent validation, this procedure might be limited when sub-dividing the samples into the two regions (*n*=63 for each region) by which the sample number is then too small for region specific calibration and independent validation Alternatively, a cross validation procedure is particularly suited when only limited numbers of samples are available (Ludwig, et al. 2008). A flow diagram of all calibrations and validations is shown in Figure 2.1. Both strategies of independent validation (approach I) and cross-validation (approach II) were used in the current study. Therefore, for each soil property two approaches, independent validation and cross-validation, were applied to calibrate generic models (i.e., covering both study regions), while the cross-validation was further used to develop region specific models (i.e., separately for the two study regions) using the PLSR procedure. The use of term ‘generic’ throughout this paper only refers to the PLSR models built upon data set across both regions. The developments of PLSR models as well as all spectral pre-processing were performed in OPUS version 6.5 software packages (Bruker Optik GmbH). Through a random selection, 28 fields with a set of 84 soil samples (14 fields with 42 samples per study region) were selected for model calibration, while the remaining 14 fields with a set of 42 soil samples (7 fields with 21 samples per region) were used for later independent model validation via either approach. Later in model validation step, all 42 samples was considered as A, B the subset from Swabian Alb and C as a subset from Kraichgau region consists of 7 fields with 21 samples per region. No samples were excluded as outliers.

2.3.5.2. Independent validation approach

The above mentioned set of 84 soil samples was subjected to PLSR model calibration resulting in a generic calibrated model (GIC model). Thereafter, the GIC model was applied separately to the data sets A, B and C for model validation yielding prediction results GIC model-A, GIC model-B, GIC model-C (see Figure 2.1).

2.3.5.3. Cross-validation approach

To develop a generic cross-validated model (GCV model), midDRIFTS spectra from soil samples of the above mentioned set of 84 soil samples were used to develop the models by a leave-one-out cross validation approach. In addition, the soil sample sets were sub-divided into the two study regions (*n*=42). Region specific cross-validation was then performed using the midDRIFTS spectra of each study region (K and SA) individually, resulting in two separately region specific cross-validated models (SA_CV, K_CV models, respectively). To avoid optimistic assessment of predictive performance of cross-validation derived models, an independent validation varied out additionally to check the predictive performance of the models (Brown et al 2005). Hence, the GCV model was applied to the three data sets resulting in GCV model-A, GCV model-B, GCV model-C predictions. Likewise, the region specific cross-validated models SA_CV and K_CV were applied to the region subsample sets B and C, respectively. This yielded predictions SA_CV model-B and K_CV model-C.

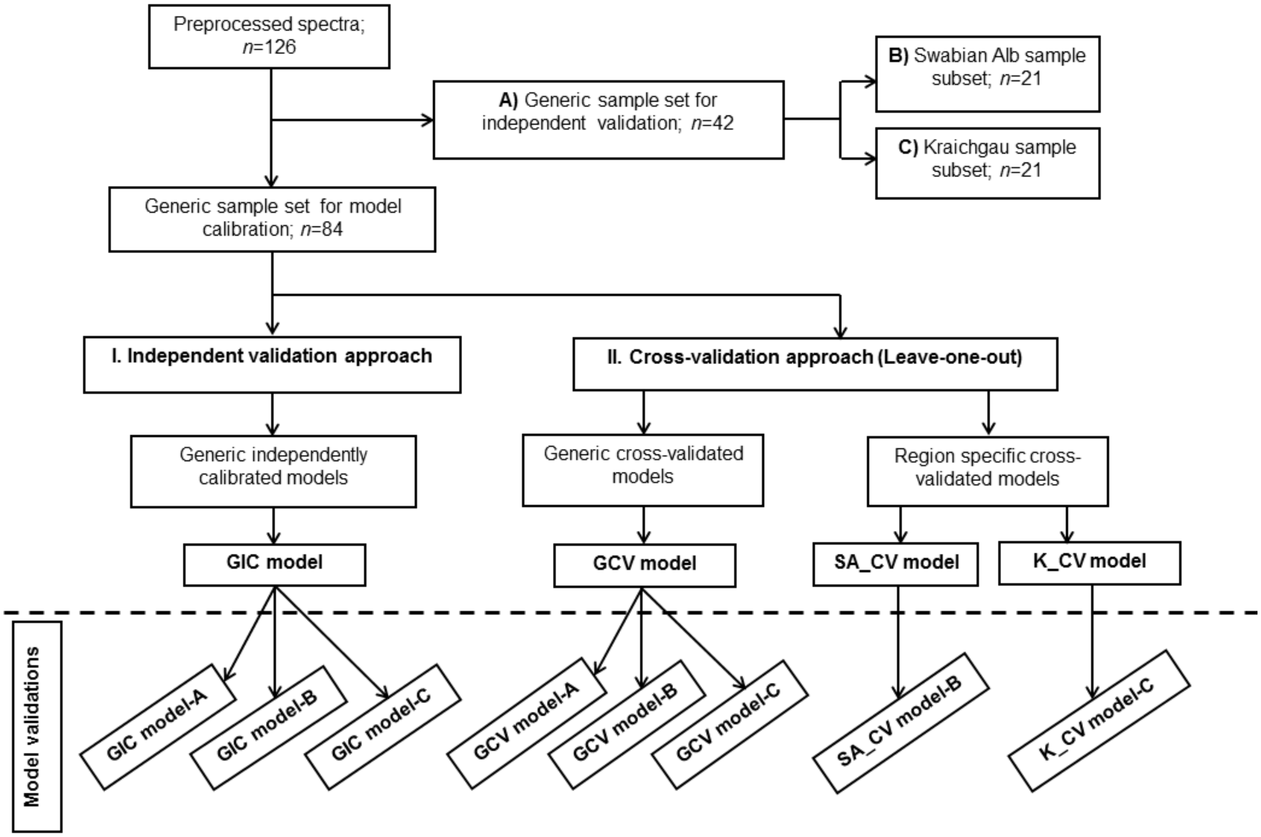

Figure 2.1. Diagram showing the PLSR analysis using the midDRIFTS spectra of soil samples associated with their laboratory measured values. GIC model: Generic independent calibrated model across both regions; GCV model: Generic cross-validated model across both regions; K_CV and SA_CV models; Region specific cross-validated models for Kraichgau and Swabian Alb respectively. Each above mentioned calibrated model was further applied to independent sample set for model validation. For example, GIC model-A is the model validation based on GIC model applied to the validation set of the two regions (*n*=42; A).

2.3.6. Statistical analysis

Descriptive statistics in addition to the coefficient of variation (CV) for assessing overall variability (Gallardo and Paramá, 2007) and significant differences of soil properties between two study regions were tested by a non-parametric method (Mann-Whitney U test) (Bekele and Hudnall, 2006). Spearman linear correlations were used to assess the relationships between hot-water extractable organic matter and other labile fractions of SOM and soil physio-chemical properties. Statistical analyses were conducted with the SPSS 11.5 software package (IBM Corporation, New York, USA).

Prediction accuracy of selected PLSR models was evaluated by the residual prediction deviation (RPD) which is the ratio of the standard deviation of the reference values and the root mean square error of predictions, the coefficient of determination (R^2) and the root mean square error of cross-validation ($RMSE_{CV}$) or of prediction ($RMSE_P$). Predictions with a RPD > 5 were considered as “excellent”; 5> RPD >3 as “acceptable”, while 3> RPD >1.4 as “moderately successful” prediction models (Pirie et al., 2005). RPDs <1.4 were considered as unsuccessful predictions (Chang, et al. 2001). The number of PLSR factors used in the calibration was determined by the prediction residual error sum of squares (PRESS) *F-statistics* from the both calibration procedures. Then the optimal number of PLS factors was used to develop a final calibration for each soil properties (McCarty, et al. 2002).

2.4. Results

2.4.1. Descriptive statistics of soil properties and differences among regions

Comparison of mean values indicated that values of measured soil properties in SA were higher ($p<0.01$) than in K except pH, TIC, and sand content (Table 2.1). This is particularly true for C_{mic} (675 vs. 262 mg C_{mic} kg^{-1} soil respectively) and TOC (3 vs. 1.1% respectively). In terms of soil texture, higher mean clay contents were obtained in SA (52.5%) as compared to K (22.4%). The variability as measured by the coefficient of variation (CV) (Table 2.1) revealed that 10 out of 15 soil properties in SA had higher variability than those in K, with the exception of C/N ratio and sand content. Among all soil chemical properties, TIC varied most, with a CV of approximately 185% for both regions.

2.4.2. Relationships between soil properties

There was a significant correlation between all examined soil properties with some exceptions (Table 2.2). High concentrations of soil C pools were paired with high concentrations of soil N pools. There was a positive correlation between TC, TIC, TOC, C_{mic} and C_{HWE} with all of N pools (i.e. TN, N_{mic}, N_{min}, and N_{HWE}; $p < 0.001$). Taking TOC as an important soil quality indicator, it was correlated positively not only with other C pools like TC, TIC, C_{mic} and C_{HWE} (R^2=0.9, R^2=0.4, R^2=0.9 and R^2=0.9 respectively; $p < 0.001$), but also with N pools such as TN, N_{min}, N_{mic} and N_{HWE} (R^2=0.9, R^2=0.8, R^2=0.7 and R^2=0.9). Soil silt and clay fractions exhibited the opposite trends. Silt content was negatively correlated with all soil properties ($p < 0.001$) with the exception of pH and bulk density, while positive correlations were found between clay and the remaining soil properties ($p < 0.001$).

Table 2.1. Descriptive statistics of soil properties (0-30 cm depth) from Kraichgau (K) and Swabian Alb (SA) sites (n=63 per each region).

Soil property[a]	Study region	mean	Std.dev	CV %	Min.	Max.	Skewness	Kurtosis	Normality (K-S[b] test)[c]	Mann-Whitney U test[d]
TC %	K	1.4	0.6	42	0.7	3.7	1.8	3.2	0.00	0.000
	SA	3.5	1.4	41	1.5	7.3	0.9	0.2	0.01	
TIC %	K	0.3	0.5	184	0	2.3	2.2	4.9	0.00	0.11
	SA	0.6	1	186	0	4.1	2.2	4	0.00	
TOC %	K	1.1	0.2	16	0.7	1.5	0.01	-0.5	0.2	0.000
	SA	3	0.8	27	1.5	4.9	0.3	-0.3	0.2	
TN %	K	0.1	0.02	14	0.1	0.2	0.6	0.09	0.00	0.000
	SA	0.3	0.08	26	0.2	0.5	0.3	-0.2	0.2	
C/N	K	9.4	1.4	14	6.7	14.9	1.6	4.1	0.00	0.000
	SA	9.8	0.5	6	8.9	11.7	1.5	3	0.00	
N_{min}	K	6.2	2.7	44	2.6	18	2.2	7.8	0.00	0.000
mg $kg^{-1}{}_{soil}$	SA	22.4	28.2	126	6.3	163	4	18	0.00	
C_{mic}	K	262	86.9	33	98	470	0.4	-0.5	0.17	0.000
mg $kg^{-1}{}_{soil}$	SA	675	246	36	276	1157	0.2	-1	0.2	
N_{mic}	K	51.3	16.7	33	15	82	-0.3	-0.8	0.03	0.000
mg $kg^{-1}{}_{soil}$	SA	98	42	43	26	196	0.3	-0.7	0.2	
C_{HWE}	K	440	128	29	207	911	1.2	2.8	0.00	0.000
mg $kg^{-1}{}_{soil}$	SA	1004	359	36	449	1838	0.2	-0.9	0.2	
N_{HWE}	K	43	13	31	14	78	0.4	0.5	0.2	0.000
mg $kg^{-1}{}_{soil}$	SA	99	43	44	34	213	0.5	-0.4	0.08	
pH	K	6.7	0.8	11	5.1	7.5	-0.6	-1.3	0.00	0.58
	SA	6.7	0.7	11	4.6	7.5	-1.6	1.6	0.00	
Bulk density	K	1.5	0.1	9	1.1	1.7	-0.7	0.3	0.2	0.000
g $cm^{-3}{}_{soil}$	SA	1.2	0.2	15	0.8	1.6	-0.2	-0.4	0.04	
Sand %	K	4.2	2.8	66	1.7	12.1	1.9	3.1	0.00	0.81
	SA	4.4	2.8	62	1.5	11.8	1.5	1.1	0.00	
Silt %	K	73.8	5.9	8	59.4	82.4	-0.4	-0.4	0.2	0.000
	SA	47	12	25	27.9	77.6	0.9	0.5	0.2	
Clay %	K	22.4	4.9	22	13.5	29.9	-0.3	-1	0.2	0.000
	SA	52.5	13	24	25	76.9	-0.4	-0.3	0.00	

[a] TC: Total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon,TN: Total nitrogen, C/N: TOC to TN ratio, C_{mic}: Microbial biomass carbon, N_{mic}: Microbial biomass nitrogen, N_{min}: Mineral nitrogen, C_{HWE}: Hot water extractable carbon, N_{HWE}: Hot water extractable nitrogen,

[b] Kolmogorov–Smirnov,

[C & D] p<0.01 significant at 1% probability; p<0.05 significant at 5% probability: p>0.05 non-significant

Table 2.2. Spearman correlation coefficients between investigated topsoil properties in the whole dataset (n=126).

Soil property[a]	C_{mic}	N_{mic}	N_{min}	TC	TIC	TOC	TN	C_{HWE}	N_{HWE}	pH	Bulk density	sand	silt
N_{mic}	0.8**[b]												
N_{min}	0.8**	0.5**											
TC	0.8**	0.8**	0.7**										
TIC	0.3**	0.6**	0.1	0.7**									
TOC	0.9**	0.8**	0.7**	0.9**	0.4**								
TN	0.9**	0.7**	0.8**	0.8**	0.3**	0.9**							
C_{HWE}	0.8**	0.8**	0.7**	0.9**	0.5**	0.9**	0.8**						
N_{HWE}	0.8**	0.8**	0.6**	0.9**	0.6**	0.9**	0.8**	0.9**					
pH	0.1	-0.3**	-0.1	0.3**	0.7**	0.2	0.1	0.2*	0.3**				
Bulk density	-0.7**	-0.6**	-0.7**	-0.7**	-0.3**	-0.8**	-0.8**	-0.8**	-0.7**	-0.1			
sand	-0.1	-0.1	0.1	0.2	0.2**	0.1	-0.1	0.1	0.1	0.2*	0.1		
silt	-0.7**	-0.6**	-0.7**	-0.7**	-0.3**	-0.8**	-0.8**	-0.7**	-0.7**	-0.2	0.7**	-0.3**	
clay	0.8**	0.6**	0.8**	0.7**	0.3**	0.8**	0.8**	0.8**	0.7**	0.2*	-0.7**	0.1	-0.9**

[a] TC: Total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon, TN: Total nitrogen, C/N: TOC to TN ratio, C_{mic}: Microbial biomass carbon, N_{mic}: Microbial biomass nitrogen, N_{min}: Mineral nitrogen, C_{HWE}: Hot water extractable carbon, N_{HWE}: Hot water extractable nitrogen, [b] *Significant at 0.05 level; ** Significant at 0.01 level.

2.4.3. Spectral assessment of investigated study regions

Averaged baseline corrected midDRIFTS spectra of bulk soils from two investigated regions are shown in Figure 2.2. The midDRIFTS spectra for both regions were generally similar in shape, but with different absorbance reflected by peak heights. This was expected as the soil from the two regions had different texture and mineralogy. The spectra showed several peaks which were assigned to organic and mineral functional groups (Janik, et al. 2007). There were a great number of peaks at the 2100 to 600 cm^{-1} region which were assigned to fundamental molecular vibrations such as quartz O-Si-O stretching peaks and several peaks representing organic matter (Nguyen, et al. 1991; Janik, et al. 2007; Spaccini and Piccolo, 2007). Maximum absorbance occurred in the 3700 to 3600 cm^{-1} region representing hydroxyl stretching vibrations associated with clay minerals. Generally, O–H (3690 to 3620 cm^{-1}) and Si–O stretching (near 1100 to 1000 cm^{-1}) may be related to clay minerals (Dixon and Weed, 1989). In the O–H stretching region, between 3800 and 3400 cm^{-1} a number of bands were observed which were related to clay minerals (Busca and Resini, 2000). Other clay minerals, such as gibbsite were found between 3610 and 3395 cm^{-1} and below 1200 cm^{-1} (White and Roth, 1986). SOM also had characteristic peaks in the mid-infrared range. Small peaks between 2960 and 2860 cm^{-1} indicated the presence of aliphatic C–H stretching according to Stevenson (1994). This band was more evident in the spectra of SA samples, which had higher concentrations of OC. Carbonates had peaks in the region of 2600 to 2500 cm^{-1} (Nguyen, et al. 1991). Peaks at 1980 cm^{-1} and 1870 cm^{-1} were assigned to quartz vibrations (Nguyen, et al. 1991) which produced higher absorbance in K. Peak near 1790 cm^{-1} was assigned to quartz vibrations (Nguyen, et al. 1991) and also C carbonyl groups in alkyl and alkyl-aryl polyester bonds (Spaccini and Piccolo, 2007). There was a prominent peak from 2000-1500 cm^{-1} appearing in both regions representing double bonds of C=C and C=O groups (Baes and Bloom, 1989). Peak at 1290 cm^{-1} was assigned to phenol vibrations according to Socrates (1980). Peak near 1160 cm^{-1}, which was more dominant in SA, was assigned to C-O bonds in both poly-alcoholic and other functional groups (Spaccini and Piccolo, 2007), while the spectral bands in the region near 1020 cm^{-1}, 920 cm^{-1}, 810 cm^{-1}, 700 cm^{-1}, and 670 cm^{-1} were associated with alumino-silicate vibrations of clay and quartz-derived fundamental O-Si-O stretching (Russel, 1987; Nguyen, et al. 1991).

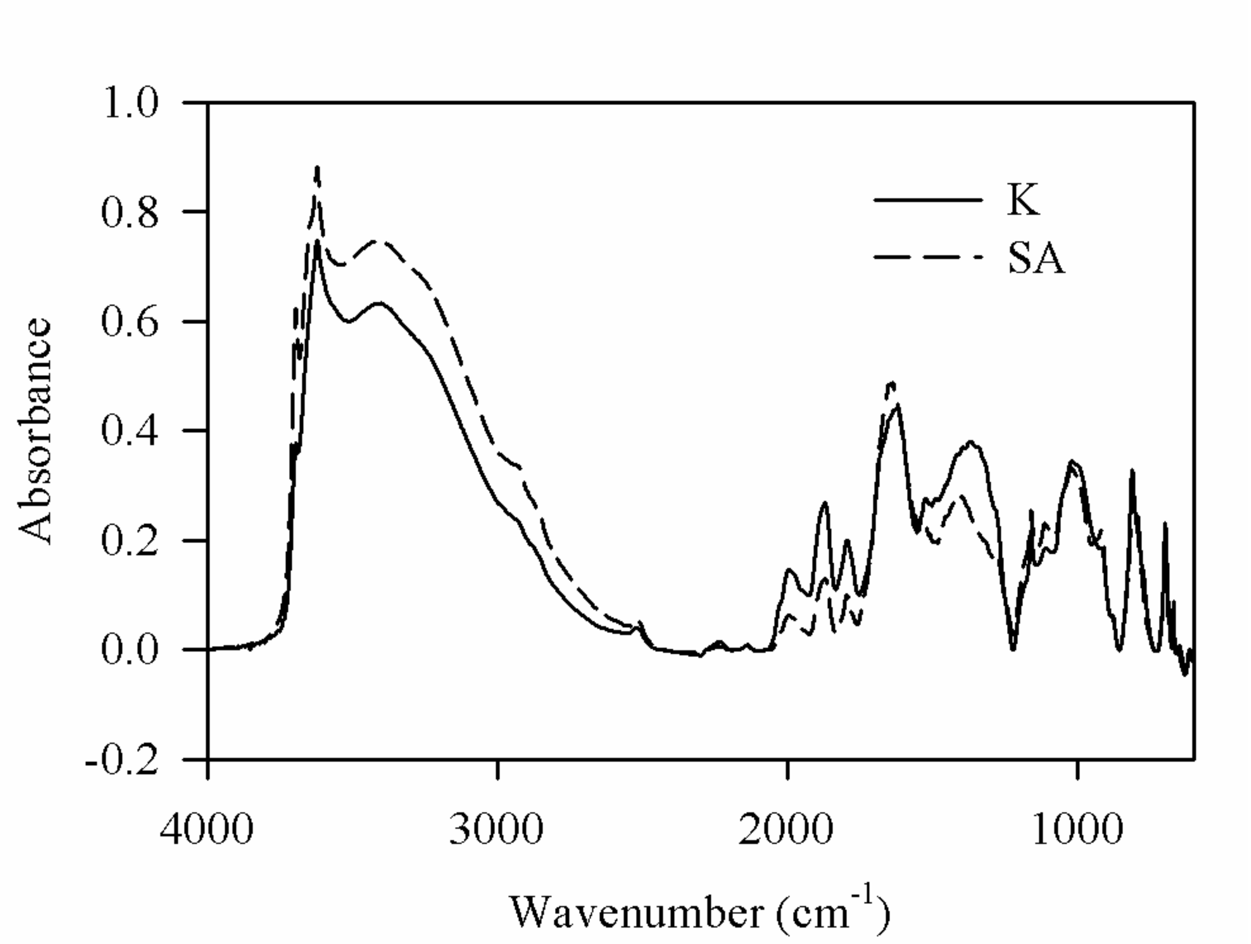

Figure 2.2. The average and baseline-corrected midDRIFTS spectra of the soils derived from the Kraichgau (K) and Swabian Alb (SA) study regions (*n*=63 per study region)

2.4.4. PLSR calibrations

2.4.4.1. PLSR calibration based on cross-validation approach across both study regions (GCV model)

Cross-validation derived models of investigated soil properties are shown in Table 2.3. Calibrated models for TC, TIC, TOC and TN were excellent with R^2= 0.99 for all and RPDs > 5. Acceptable calibrated models were also found for C_{mic}, N_{mic}, C_{HWE}, N_{HWE} and C/N ratio with R^2 and RPD values ranging between 0.88-0.95 and 3-4.3, respectively. Moderately successful calibrated models were found for N_{min}, pH, bulk density and sand contents with RPD <3. Excellent and acceptable calibrated models were obtained in particular for clay (R^2=0.97; RPD=5.7) and silt (R^2=0.95; RPD=4.5), while moderately successful for sand content (R^2=0.83; RPD=2.4) (Table 2.3).

During model optimization, different numbers of PLSR factors, data processing transformation and frequency range were used for calibration. Number of factors varied from 7 for bulk density up to 20 for TN. Factors used during calibration corresponded with the lowest $RMSE_{CV}$ and highest R^2. Taking TC, for instance, with R^2=0.99 and $RMSE_{CV}$=0.1 %, the number of factors used for model calibration was 15, while TN produced its best model performance with 20 factors. A certain number of frequency ranges with specific peaks corresponded to each soil properties. It varied between soil properties where more frequency ranges as compared to all other measured parameters were used for soil C fractions such as TC, TIC, TOC and C_{HWE} (Table 2.3).

2.4.4.2. Region specific cross-validated PLSR models (SA_CV and K_CV models)

Model performance (R^2 and RPD) for SA_CV model indicated only a slight decrease compared to the GCV model (data not shown). Exceptions were pH, sand and clay content where model accuracies improved. Excellent calibrated models were obtained for TC, TIC, TOC, and TN with R^2 values between 0.94 and 0.99 and RPD values between 4 and 18 for the SA_CV models. SA_CV models for C_{mic} and N_{HWE} were acceptable, as indicated by R^2 values of 0.88 and 0.90, and RPD values of 3 and 3.2, respectively. Remaining soil properties such as bulk density, C_{HWE}, sand and C/N ratio showed moderately successful model performances for SA study region. The K_CV models showed roughly the same calibration performance for TC, TIC and TOC as the GCV and SA_CV models. However, clearly lower K_CV models accuracies were visible for remaining soil properties except for C/N ratio where an improvement was found compared to that one for both GCV and SA_CV models. SA_CV models for soil texture yielded more accurate results, especially for silt and clay compared to the results obtained for K_CV models (data not shown).

2.4.4.3. PLSR calibration based on an independent validation approach across both study regions (GIC model)

MidDRIFTS-PLSR calibrated models using spectra of the generic calibration sample set (*n*=84) showed for C fractions excellent results, e.g., TC (RPD=17.5; R^2=0.99), TOC (RPD=11, R^2=0.99) and TIC (RPD=17; R^2=0.99) (Table 2.4). GIC model for C_{mic} (RPD=4.4 R^2=0.95) represents an acceptable model, while moderately successful predictions were found for C_{HWE}

with RPD and R^2 values of 2.8 and 0.88, respectively. Calibrations for N pools produced an excellent model for TN with R^2=0.99 and RPD=12.2; and acceptable predictions for N_{HWE} (RPD=3.2; R^2=0.9), N_{mic} (RPD=3.6; R^2=0.92), and N_{min} (RPD=3.4; R^2=0.91). In terms of soil texture, acceptable models were produced for silt and clay fractions with RPD >3, with a moderately successful calibrated obtained for sand. Calibrations with RPD <3 judged as moderately successful were C/N ratio, pH and bulk density.

In contrast to both generic and site specific cross-validation mentioned above, a lower number of factors were used for model calibration, ranging from 2 for sand up to 14 for TIC and N_{mic}. Soil spectra were subjected to different spectral pre-processing techniques. Generally, the 1st derivative in combination with some other, such as vector normalization and multiplicative scattering correction, was the most-used method for optimal calibration. As for the cross-validation approach, the model was built upon different frequency ranges for each soil property. However, there were some peaks throughout the spectral frequencies which were contributed in model calibration for several soil properties (Table 2.4).

Table 2.3. Calibration results using 84 midDRIFTS spectra of bulk soils across both regions based on cross-validation approach (GCV model).

Property	Number of factor	Preprocessing method[b]	Wavenumbers [cm^{-1}]	Model accuracy indicators [a]		
				R^2	RPD	$RMSE_{CV}$
TC	15	FD+MSC	3901-3317, 2980-2407, 2301-1618, 1281-939	0.99	15.2	0.1
TIC	18	SNV	2638-2407, 2301-1957, 1620-1279, 641-698	0.99	12.2	0.07
TOC	17	FD+MSC	3901-3317, 2980-2638, 1959-1279, 941-698	0.99	8.6	0.13
TN	20	SD	3901-3317, 2301-1279	0.99	11.3	0.01
C/N	9	MMN	3319-2397, 1959-1618	0.90	3.2	0.4
N_{min}	15	SD	3658-2638, 2301-1957	0.87	2.7	4
C_{mic}	18	SD	3658-2978, 2638-2397, 2301-939	0.95	4.3	66
N_{mic}	16	FD	3318-2977, 2300-1618, 1280-939	0.88	2.9	14
C_{HWE}	16	SD	3901-2978, 2638-2397, 2301-1618, 1281-939	0.92	3.6	112
N_{HWE}	16	SD	3901-2638, 2301-1957, 1620-1279	0.90	3.2	14
pH	15	FD+MSC	3901-3656, 2980-2638	0.71	1.9	0.4
Bulk density	7	FD+SNV	3901-3656, 2980-2397, 1959-698	0.75	2	0.11
sand	13	SD	3901-3317, 2301-1957, 1620-939	0.83	2.4	1.2
silt	9	SD	3901-3656, 3319-2638, 2301-1618, 1281-698	0.95	4.5	3.5
clay	17	MMN	3658-3317, 2638-2397, 2301-1957, 941-698	0.97	5.7	3.1

[a] TC: Total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon,TN: Total nitrogen, C/N: TOC to TN ratio, N_{min}: Mineral nitrogen, C_{mic}: Microbial biomass carbon, N_{mic}: Microbial biomass nitrogen, C_{HWE}: Hot water extractable carbon, N_{HWE}: Hot water extractable nitrogen,

[b] FD: First derivative; SD: Second derivative; MSC: Multiplicative scattering correction; SNV: Vector normalization; MMN: Min-Max normalization;

[a] R^2: coefficient of determination; RPD: residual prediction deviation; $RMSE_{CV}$: root mean square error of cross-validation

Table 2.4. Calibration results using 84 midDRIFTS spectra of bulk soils across both regions based independent validation approach (GIC model).

Property[a]	Number of factors	Preprocessing method[b]	Wavenumbers (cm^{-1})	Model accuracy indicators [c]		
				R^2	RPD	$RMSE_P$
TC	10	FD+MSC	2980-2638, 2301-1618, 1281-698	0.99	17.5	0.1
TIC	14	SLS	2640-2399, 2301-1618, 941-698	0.99	17	0.05
TOC	9	MSC	2980-2399, 2301-1618, 1281-939	0.99	11	0.1
TN	11	FD+MSC	3658-2977, 1959-939	0.99	12.2	0.01
C/N	4	FD+MSC	3658-2399, 2301-1957, 1620-1277, 941-698	0.79	2.2	0.6
N_{min}	8	FD	1620-1279	91	3.4	8.6
C_{mic}	9	FD+SNV	3319-2399, 2301-1957, 1620-1279, 941-698	0.95	4.4	70
N_{mic}	14	FD+MSC	3319-2978, 2640-2399, 2301-1618	0.92	3.6	12
C_{HWE}	5	SLS	3658-2399, 2301-1957, 1281-939	0.88	2.8	153
N_{HWE}	9	MMN	3658-2399, 2301-1618, 1281-939	0.9	3.2	15
pH	3	SLS	1959-1618, 941-698	0.56	1.5	0.5
Bulk density	3	MSC	3319-2978, 2640-2399, 1620-1279	0.56	1.5	0.15
sand	2	FD+MSC	2980-2638, 2301-1957, 941-698	0.32	1.2	2.7
silt	4	FD+SNV	3320-2978, 2301-1957, 1620-1279	0.92	3.5	4.6
clay	9	FD+MSC	3658-3317, 2980-2399, 2301-1957, 1281-698	0.97	6	3.1

[a] TC: Total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon,TN: Total nitrogen, C/N: TOC to TN ratio, N_{min}: Mineral nitrogen, C_{mic}: Microbial biomass carbon, N_{mic}: Microbial biomass nitrogen, C_{HWE}: Hot water extractable carbon, N_{HWE}: Hot water extractable nitrogen,

[b] FD: First derivative; SD: Second derivative; MSC: Multiplicative scattering correction; SNV: Vector normalization; MMN: Min-Max normalization; SLS: Straight line subtraction,

[c] Calculated for model validation; R^2: coefficient of determination; RPD: residual prediction deviation; $RMSE_P$: root mean square error of prediction

2.4.5. Independent validation of PLSR calibrated models (model validation)

2.4.5.1. Independent validation of GCV model (GCV model-A, GCV model-B, GCV model-C)

Validation of the GCV model was performed across both study regions and for each individual region. Prediction accuracies across both regions (GCV model-A) were characterized as excellent for TC, TIC, TOC, TN (RPDs >5), while there were lower prediction accuracies for remaining soil properties. Less accurate predictions were obtained for N_{mic}, N_{mic}, C_{HWE} and N_{HWE}, with RPDs of only 1.9-2.8 representing inadequate prediction accuracies. Regarding silt and clay contents, similar results as the GCV model were obtained, with only slightly lower R^2 and RPD values, still indicating acceptable validated models (Table 2.5).

Table 2.5. Prediction accuracies of cross-validation derived models (GCV-model, *n*=84; Table 5) applied to an independent test set for validation (*n*=42) across both study regions and for each individual study region (SA and K) separately (*n*=21)

Property	GCV model-A			GCV model-B			GCV model-C		
	R^2	RPD	$RMSE_P$	R^2	RPD	$RMSE_P$	R^2	RPD	$RMSE_P$
TC	0.99	11	0.13	0.98	10	0.18	0.95	4.3	0.08
TIC	0.97	5.5	0.1	0.98	6	0.13	0.94	4	0.05
TOC	0.98	7.2	0.15	0.96	5.1	0.2	0.86	2.7	0.07
TN	0.96	5.4	0.02	0.92	4.5	0.03	0.92	2.5	0.01
C/N	0.42	1.4	0.44	0.001	0.8	0.44	0.7	1.8	0.43
N_{min}	0.58	0.8	7.1	0.38	0.7	9.4	0.26	0.4	3.6
C_{mic}	0.81	2.7	110	0.77	2.2	127	0.6	1.4	90
N_{mic}	0.86	2.8	14.1	0.86	2.9	17	0.59	1.6	10
C_{HWE}	0.79	2.2	149	0.85	2.4	147	0.1	1.2	151
N_{HWE}	0.71	1.9	18	0.8	2.4	18	0.3	1.1	18.6
pH	0.21	1	0.74	0.57	1.5	0.5	0.005	1	0.9
Bulk density	0.27	1.1	0.2	0.001	1	0.24	0.16	1.1	0.15
sand	0.12	0.6	2	0.1	0.6	2.1	0.17	0.6	1.8
silt	0.88	3	5.8	0.87	2.7	6	0.05	1	5.6
clay	0.93	4	4.8	0.87	2.7	6	0.58	1.4	3

[a] TC: Total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon,TN: Total nitrogen, C/N: TOC to TN ratio, N_{min}: Mineral nitrogen, C_{mic}: Microbial biomass carbon, N_{mic}: Microbial biomass nitrogen, C_{HWE}: Hot water extractable carbon, N_{HWE}: Hot water extractable nitrogen,

[b] Calculated for model validation; R^2: coefficient of determination; RPD: residual prediction deviation; $RMSE_P$: root mean square error of prediction

Prediction results deteriorated clearly when GCV model was applied to each regional sample subsets individually for model validation. Prediction of the SA sample subsets (GCV model-B) resulted in at least acceptable predictions for TC, TIC, TOC and TN, while remaining properties were predicted with lower accuracies. Model validation with lowest prediction accuracy was obtained for K sample subsets (GCV model-C). Acceptable models were found only for TC and TIC, while the prediction of the remaining properties ranged from unsuccessful up to moderately successful predictions.

Table 2.6. Prediction accuracies of region specific cross-validation derived models (SA_CV and K_CV models applied to independent validation datasets for each individual study region separately (SA and K, respectively, *n*=21)

Property	SA_CV model-B			K_CV model-C		
	R^2	RPD	$RMSE_P$	R^2	RPD	$RMSE_P$
TC	0.96	5.2	0.9	0.96	5.1	0.06
TIC	0.98	6	0.12	0.98	8	0.03
TOC	0.91	3	0.4	0.86	2.8	0.07
TN	0.92	4.8	0.02	0.92	2.5	0.01
C/N	0.001	0.7	0.6	0.2	1	0.8
N_{min}	0.31	0.8	8.3	0.26	0.4	3.6
C_{mic}	0.80	2.2	123	0.2	1	85
N_{mic}	0.64	1.6	30	0.6	1.5	11
C_{HWE}	0.65	0.7	564	0.2	1.1	192
N_{HWE}	0.69	1.5	27	0.3	1.2	15.6
pH	0.51	1.4	0.5	0.02	0.8	1
Bulk density	0.06	0.8	0.28	0.04	0.7	0.21
sand	0.001	0.4	3.2	0.17	0.6	1.8
silt	0.77	2	10	0.01	0.8	7.2
clay	0.79	2.7	9	0.17	0.8	4.8

[a] TC: Total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon,TN: Total nitrogen, C/N: TOC to TN ratio, Nmin: Mineral nitrogen, Cmic: Microbial biomass carbon, Nmic: Microbial biomass nitrogen, C_{HWE}: Hot water extractable carbon, N_{HWE}: Hot water extractable nitrogen,

[b] Calculated for model validation; R^2: coefficient of determination; RPD: residual prediction deviation; $RMSE_P$: root mean square error of prediction

2.4.5.2. Independent validation of SA_CV and K_CV model (SA_CV model-B, K_CV model-C)

Just as with the GCV model, the region specific cross-validation models were assessed with the independent data sets. Predictions for either regional sample sub-sets were generally of lower accuracies as compared to GCV and GIC models. The only at least acceptable prediction results were obtained with SA_CV for TC, TIC and TN, and K_CV for TC and TIC. While prediction accuracies obtained on region specific sample subsets were excellent, with R^2 values of 0.97 and 0.99 for TC and TIC respectively, the K_CV model-C for the remaining soil properties were less accurate, ranging from unacceptable up to moderately acceptable models indicating lack of accuracy for prediction purposes (Table 2.6).

2.4.5.3. Independent validation of GIC model (GIC model-A, GIC model-B, GIC model-C)

Model validation results based on GIC model (Table 2.4) for all soil properties are summarized in Table 2.7 and Figure 2.3. Our calibrated models were acceptable (RPD>3) in predicting all soil C fractions of the combined sample set. Excellent predictions were obtained for TC (RPD=14.5; R^2=0.99), TOC (RPD=12; R^2=0.99), TIC (RPD=14; R^2=0.99), and C_{mic} (RPD=5; R^2=0.95). An acceptable model was obtained for C_{HWE} with measured values of RPD=3 and R^2=0.88. For N fractions, excellent predictions with R^2=0.99 and RPD=15.5 for TN and acceptable predictions for N_{mic} (RPD=3.8; R^2=0.93) and N_{HWE} (RPD=3.2; R^2=0.9) were obtained, while model performance indicators yielded weak prediction results for N_{min} (RPD=1.4; R^2=0.4). Among all variables, least accurate predictions were obtained for C/N ratio, soil pH and bulk density with the RPD <2. In terms of soil texture, excellent predictions were found for silt (R^2=0.93; RPD=3.9) and clay (R^2=0.98; RPD=7.7), whereas a relatively poor performance for sand content was found. Although predictions on the combined sample set (GIC model-A) indicated satisfactory applicability of calibrated models, it was not clear whether that calibration could be applied to a region specific sample set successfully. We illustrated the prediction results generated for each region, GIC model-B and GIC model-C separately, in Table 2.7. In general, accuracy of model predictions on region specific sample sets decreased slightly compared to the combined sample set. Although, there was a lower accuracies for GIC model-B, model efficiency did not change compared to the GIC model-A. For GIC model-C, predictions for C_{mic}, C_{HWE}, N_{HWE} and silt and clay contents produced unacceptable predictions with RPDs <3. However, model predictions on validation sample subsets generated for TC, TIC, TOC, TN, and N_{mic} were as accurate as GIC model-A and GIC model-B which still ranked as excellent models (Table 2.7).

Table 2.7. Prediction accuracies of a PLSR calibrated model (GIC model, *n*=84; Table 3) applied to an independent validation dataset (*n*=42) across both study regions and for each individual study region separately (*n*=21).

Property[a]	GIC model-A			GIC model-B			GIC model-C		
	R^2	RPD	$RMSE_P$[b]	R^2	RPD	$RMSE_P$	R^2	RPD	$RMSE_P$
TC	0.99	14.5	0.1	0.99	12	0.1	0.99	12	0.02
TIC	0.99	12	0.05	0.99	17	0.04	0.99	12	0.02
TOC	0.99	10	0.08	0.98	8	0.14	0.97	5.4	0.04
TN	0.99	15.5	0.007	0.99	14	0.007	0.96	5.2	0.004
C/N	0.67	1.8	0.34	0.43	1.3	0.27	0.78	2.1	0.3
N_{min}	0.4	1.4	4.3	0.24	1.2	4.7	0.16	1.2	1.4
C_{mic}	0.96	5	51	0.95	4.6	57	0.83	2.5	34
N_{mic}	0.93	3.8	10	0.95	4.4	11	0.92	3.5	4.5
C_{HWE}	0.88	3	113	0.88	3.1	121	0.56	1.8	104
N_{HWE}	0.9	3.2	11	0.90	3.2	12	0.83	2.4	6
pH	0.53	1.5	0.52	0.90	3.2	0.2	0.25	1.2	0.68
Bulk density	0.69	1.9	0.12	0.75	2	0.11	0.36	1.3	0.09
sand	0.15	1	1.2	0.38	1.3	1	0.48	1.5	0.8
silt	0.93	3.9	4.4	0.94	4.2	4	0.37	1.3	3.7
clay	0.98	7.7	2.4	0.98	7	2.4	0.79	2.2	1.8

[a] TC: Total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon,TN: Total nitrogen, C/N: TOC to TN ratio, N_{min}: Mineral nitrogen, C_{mic}: Microbial biomass carbon, N_{mic}: Microbial biomass nitrogen, C_{HWE}: Hot water extractable carbon, N_{HWE}: Hot water extractable nitrogen,

[b] Calculated for model validation; R^2: coefficient of determination; RPD: residual prediction deviation; $RMSE_P$: root mean square error of prediction

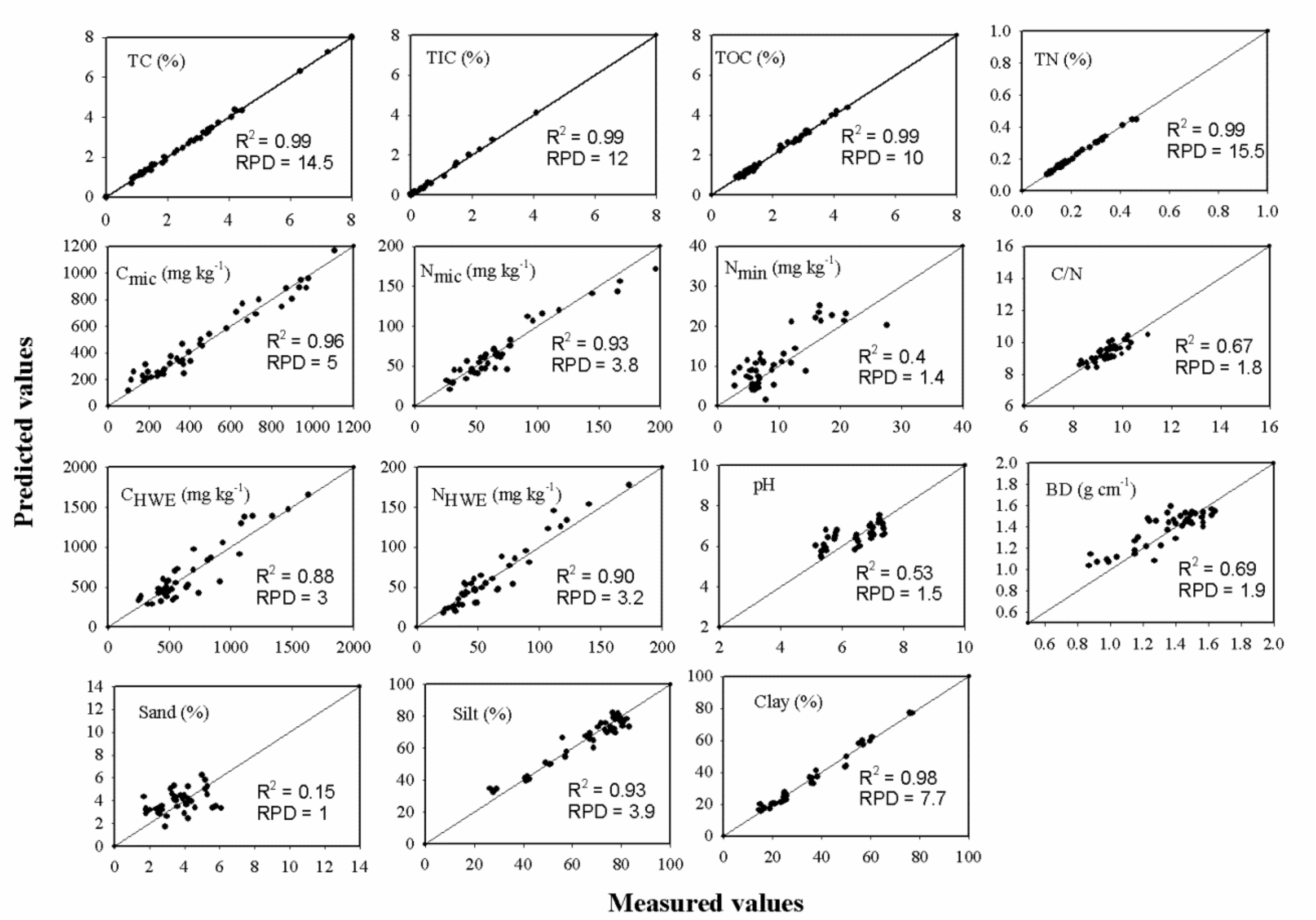

Figure 2.3. Scatter plots of PLSR-predicted vs measured values for all soil properties across both study regions (n=42) predicted independently by GIC model for investigated soil properties (referred to Table 7, GIC model-A).

2.5. Discussion

2.5.1. General considerations

It has been already showed that soil properties can be predicted using midDRIFTS with different accuracies taking soil samples in the frame of multiple spatial scales, ranging from field up to the landscape level (McCarty, et al. 2002; Madari, et al. 2006; Grinand, et al. 2012). Meanwhile, midDRIFTS-PLSR calibrated models, by any means of calibration strategies, can be strongly affected by the number of samples, similarity of calibration and validation datasets, target variables in terms of concentration, data range as well as soil texture (Bellon-Maurel, et al. 2010; Bellon-Maurel and McBratney, 2011;Yang and Mouazen, 2012). A midDRIFTS spectrum contains complex information on soil organic and mineral compounds that have mid-infrared active vibrations. These active vibrations are actually the molecular bond of biochemical constituents are viewed to overlap together entire whole spectrum. Hence, MIRS is not compound specific and its prediction is based on calibrated models which relate spectral data to concentration values of soil properties. Although, midDRIFTS-PLSR models use the most informative spectral bands to reduce the complexity and increase the model accuracy, however, irrelevant frequencies may hamper the successful calibrated models. The band between 2960 and 2860 cm^{-1} is due to C-H stretching of aliphatic compounds (Stevenson, 1994) contributing in almost all C fractions in model calibrations in current study. It is viewed that soil carbonate can overlap with this band and may affect the prediction results especially in high carbonate containing soils. However, PLSR model using whole spectrum is not limited necessarily to only one specific band. Spectral frequencies from 2200 to 1000 cm^{-1} were the most commonly used during model calibration in current study. This region assigned for several soil OM functional groups such as aromatic compounds, carbohydrates, organic acids and carbohydrate–COH stretching (Janik, et al. 1998; Viscarra-Rossel, et al. 2006) have been proposed to overlap with several mineral like clay, carbonate and silica. The midDRIFTS-PLSR calibrated model for TOC and C_{mic} for instance used several bands in MIRS spectrum are assigned for both organic matter and carbonate as well as related to other soil constituents such as clay minerals and iron oxide. Therefore, it might be built upon the organic compounds bands or as a correlation with bands representing soil minerals like clay and iron oxides (Brown, et al. 2005). Hence, the efficacy of midDRIFTS-PLSR in prediction of soil properties remains variable and complex because of the extreme spatial variability of soil. Below, possible causes of complex and variable prediction as well as differences between types of calibration are discussed.

2.5.2. Are generic calibrated models (GIC and GCV) equally valid in prediction of unknown samples?

In the current study, taking only the calibrated model results into account, i.e., GIC and GCV models, both evaluated approaches were approximately equally valid in prediction of soil properties. However, the detailed efficacies of calibrated models in predicting independent soil samples for model validation revealed some contradictions which are presented below. Despite reliable GCV models, it was not solely evident from cross-validation statistics whether they were sufficiently robust to be used on unknown samples as an independent validation is lacking in the cross-validation procedure. Overall, independent validation approach was more accurate as compared to cross-validation. According to D'Acqui et al. (2010), cross-validation derived models were shown to perform precisely in predicting properties of independent samples. On the other hand, while both cross-validation and independent validation approaches led to midDRIFTS-PLSR calibrated models with same accuracy, cross-validation derived models in predicting new independent samples resulted in considerably less accuracy as compared to independently calibrated models derived from the independent validation approach (Morón and Cozzolino, 2002; Minasny, et al. 2009). Hence, it is not totally realistic to relay on GCV models before applying on an independent test set to validate the model. Validation using GCV models gave acceptable predictions for TC, TIC, TOC, TN, silt and clay across both regions, while prediction accuracy for the remaining soil properties decreased markedly in the validation sub-set when compared to prediction results obtained by GIC models. The prediction accuracies were worse when GCV models appled to each region separately. It was attributed to the tendency of cross-validated derived models to be over-fitted during calibration (Minasny, et al. 2009; Bellon-Maurel and McBratney, 2011).

Hence, in the current study, less accurate GCV model performance on unknown samples (model validation) as compared to the GIC models was likely the result of the higher number of PLSR factors used in calibration resulting overoptimistic results in model calibration. Calibration based on fewer PLSR factors is more appropriate for verifying the models when it is applied to new samples (Madari, et al. 2006). Generally in our study, the numbers of PLSR factors used in GCV model (Table 2.3) were considerably higher than those used in GIC model (Table 2.4) especially for C_{mic}, N_{mic}, C_{HWE} and N_{HWE}. The results revealed that midDRIFTS-PLSR calibrated models by using independent validation approach outperform cross-validation derived models in prediction of soil properties on new unknown samples. Hence, sub-division of the total sample set into sub-sets for independent calibration and validation was a perquisite to avoid over-fitted calibrations which performed poorly on validation samples when cross-validation was used.

In spite of accurate prediction for GIC models in most properties of interest, however prediction accuracies were unsuccessful for pH, bulk density, C:N ratio and sand content. Prediction accuracies for pH and bulk density with indirect spectral response were moderately successful which was in line with observations by Kuang and Mouazen (2011). Unsuccessful prediction models for pH in spite of accurate prediction for TIC of soil carbonate might have been the result of recent lime application in some of the fields studied, while the lack of time has not allowed the lime to dissolve and raise the pH. Less accurate predictions were obtained for C/N, while TOC and TN were predicted well. Ratios of properties may be predicted less accurately as compared to the single quantity because of the propagation of uncertainty or errors from each single property in the ratio (Ludwig, et al. 2008). In spite of accurate prediction for silt and clay the unacceptable predictions for sand differed from that of Cobo et al. (2010). They found an RPD value of 6.8 for sand working on soils with a range in sand content between 25-75% across three investigated villages in Zimbabwe under differently soil fertility and textural classes. It might be attributed to the overlapping of the Si-O peaks with those of SOM peaks (Janik, et al. 1998) and heterogeneity of sample set as a certain amount of heterogeneity is needed for an accurate calibration (Cécillon, et al. 2009). Low sand content (2-12%) in current study were expected to be influenced more intensively by clay content. Generally, pedogenic factors such as parental material, vegetation and climate as well as soil particle size distributions are destined to alter model accuracy in the MIRS range. Thus, those conflicting results compared to findings obtained by others mentioned above (sand as an example) might be attributed mainly to the distributions of particle size and different texture classes. Higher accuracies obtained for soil biochemical characteristics (C_{mic}, N_{mic}, C_{HWE} and N_{HWE}) with compared to reported findings (Rasche et al. 2013). Soil microbiological properties like C_{mic} and N_{mic} are supposed to relate poorly with spectral measurements because of their relatively low contents in soil, less than 5% of TOC (Cohen, et al. 2005; Ludwig, et al. 2008) and it would be hindered by soil minerals and other organic peaks (Janik, et al. 1998). However, they may be predicted well due to their close relationships with SOM (Cohen, et al. 2005) and due to the strong correlation with TOC (Chodak, 2011). Our correlation results corroborated this relationship, as a significant positive correlation was found between C_{mic} and N_{mic} with TC and TOC (Table 2.2).

2.5.3. Performance of region specific cross-validation derived models (K_CV and SA_CV) in prediction of unknown samples

Prediction accuracies of the independent sample set for model validation based on site specific cross-validation derived models (K_CV model-B and SA_CV model-C) were lower compared to those of generic models, where only acceptable predictions were obtained for TC, TIC, TOC and TN in both regions (Tables 2.5 and 2.6). Different model performances for soil properties were attributed to the range of variability, soil type, site and number of samples (McBratney, et al. 2006; Kuang and Mouazen, 2011). Thus, less predictive performance of region specific cross-validated models in our study compared to the generic models was attributed to both smaller sample size and lower range of variability of the measured variables. Development of an optimum PLSR prediction model by any means of calibration approach requires a certain amount of heterogeneity (Cécillon, et al. 2009). Sub-division of our soil samples into two regional sub-sets led to a considerable decrease in model accuracy for all investigated soil properties, with model accuracy decreasing much stronger for the K_CV model-C than for the SA_CV model-B as compared to the generic models (GIC and GCV). Therefore, sufficient heterogeneity in a dataset is further possible cause affecting predictive performance of PLSR prediction model in addition to optimistic feature associated to cross-validation. The better performance of the SA_CV model-B as compared to K_CV model-C was related to both the greater variability and higher values of the measured variables found in SA than in the K region (Table 2.1). It seems as though a certain amount of variability is beneficial for model construction, but at some point becomes detrimental for accurate model construction. Based on the contrasting results reported in literature and from our own findings, it was concluded that a regional sub-division of samples is only appropriate if fundamental differences in soil properties (e.g., parent material, land use and climate) can be expected and large enough sample sets exists to employ independent validation approach for model calibration. Otherwise, it may limit the performance of PLSR procedures to find optimal model calibrations.

2.6. Conclusions

Although in previous efforts the efficacy of midDRIFTS-PLSR calibrated models has been demonstrated in successful prediction of various soil properties, there remains, several uncertainties on its potential in the frame of region dependent soil heterogeneity. Besides, effectiveness of cross-validation and independent validation derived models in predicting soil properties of unknown samples independently requires more appropriate information in terms of soil texture, SOM content, nutrient status and carbonate content. Although generic cross-validation (GCV model) models led to comparable PLSR prediction models, the most accurate PLSR predictions of “unknown” samples (model validation) were obtained using the independent validation approach. The present work demonstrated that midDRIFTS combined with PLSR was capable of predicting a wide range of biochemical and physical soil properties at regional scale from two different agro-ecological regions (40 by 40 km for each region) characterized by significantly different mean values for measured soil properties. Therefore, in data-rich situation like in current study which allows developing generic midDRIFTS-PLSR model via independent validation approach outperforms cross-validation approach. The results revealed that the generic model based on independent validation (GIC model) was the best evaluated approach to robustly and accurately predict several soil properties of new, unknown samples from the respective study regions by analyzing their midDRIFTS spectra without any further laboratory measurements. Nevertheless, creation of a truly generic midDRIFTS-PLSR calibrated model to predict any unknown samples from very different new locations requires a large calibration database from those locations. Predictions of unknown samples from new locations, which show contrasting spectral characteristics from the calibration dataset, increase the source of error resulting in declining prediction accuracies. Thus, a new calibration is needed if new, unknown samples are outside the range of the original calibration dataset and are spectrally different. In the current study, the unsuitability of site specific SA_CV and K_CV models in predicting soil properties of “unknown” samples suggested that initial cross-validation results were overoptimistic and not realistic indicators for future model performance. This shows the need, when using cross-validation to still test the model on “unknown” samples for an accuracy assessment of the predictive capability. Variability within the dataset used for calibration has a large influence on calibration and prediction accuracy. Here, the K region showing low heterogeneity seemed to be more sensitive to these criteria where prediction accuracies were lower compared to the SA region using the region specific cross-validation derived models. For promising midDRIFTS-PLSR calibrated models for soil properties in individual study regions, it is necessary to develop accurate prediction models with a large enough sample size with a sufficient heterogeneity to be based on the independent validation approach.

As final conclusion, our results proved the potential for the particular application of independent model calibration and validation to predict accurately 10 out of 15 soil properties at regional scale. MidDRIFTS-PLSR calibrated models gained provided a rapid-throughput approach for future spectral-based prediction at regions under evaluation for soil fertility assessment, modeling spatial patterns and mapping of soil properties on a regional scale, where large numbers of soil samples are required due to the unknown nature of soil variability. However, to overcome the limitations associated to cross-validation and/or accurate prediction of those properties with unsuccessful prediction a selecting an optimal of spectral peaks to reduce spectral complexity and improve the robustness of a calibrated model.

Chapter 3

Coupling midDRIFTS spectroscopy to geostatistics for regional scale mapping of soil carbon and texture in croplands in South-West Germany

3. Coupling midDRIFTS spectroscopy to geostatistics for regional scale mapping of soil carbon and texture in croplands in South-West Germany

3.1. Abstract

Assessment of spatial variability and mapping of soil properties using traditional laboratory analyses still remains challenging as large amounts of geo-referenced data is required. The aim of this study was to investigate the use of diffuse reflectance infrared Fourier transform spectroscopy in mid-range (midDRIFTS) with partial least square regression (PLSR) as a rapid-throughput method for data inputs for geostatistics. MidDRIFTS-PLSR calibrated models were used to predict soil total carbon (TC), total inorganic carbon (TIC), total organic carbon (TOC) and soil texture contents (clay, silt and sand) contents of 1170 soil samples from two regions, Kraichgau (K) and Swabian Alb (SA), in Southwest Germany. MidDRIFTS-PLSR-based predicted values were further used to be integrated with geostatisitcs using ordinary kriging (OK). The results showed that midDRIFTS-PLSR calibrated models accurately predicted 93% of the samples for all soil properties. Spatial analysis in the K region revealed moderately weak spatial dependency (Nugget-to-sill ratios > 50%) for soil TC, TIC and TOC with ranges of 30, 18 and 22 km, respectively, whereas very and moderately strong spatial dependencies (ratio ≤ 50%) were obtained for sand , silt and clay with ranges of 21, 20 and 28 km respectively. Soil properties in the SA region showed moderately strong spatial dependencies with a range of spatial autocorrelation of 8.7 km for sand up to 24 km for TIC. Ordinary kriging resulted in accurate maps of soil properties, with a root mean square standardized error (RMSSE) close to 1, except for sand in K and TIC in SA region. While the SOM map in SA and the texture map in K corresponded very well with those of reference maps, SOM map in K and texture map in SA showed dissimilarity as compared to the reference maps. This dissimilarity was attributed to soil assessment based on pedotransfer function relaying on very low resolution of the existing maps. We concluded that midDRIFTS-PLSR provided quality data which allowed integration with geostatistics to create high resolution maps of soil carbon and texture at the regional scale. It opens a new potential use as a cost-effective and promising approach to speed up and refine soil property mapping and obtain high resolution soil spatial information.

3.2. Introduction

Accurate maps of essential soil properties such soil carbon (SC) and texture are basic sources of information for decision making on lime and fertilizer application, land resource management and landscape monitoring and analysis in the frames of climate change and food security (Banwart, 2011; Lal, 2004a). Moreover, soils as a medium for plant growth play an important role in matter and energy fluxes providing soil fertility required for crop production (Lal, 2004a). Currently, lack of accurate high resolution soil data from local to global scale has been cited as the main limitation in assessing land degradation and sustainable landscape monitoring analysis (FAO, 2009). Therefore, improving the spatial resolution of georeferenced soil data especially for large scales has recently been receiving priority in both agriculturally managed and natural ecosystems (McBratney, et al. 2003; Hartemink, 2008).

The main source of soil databases in Germany is the national level map at a scale of 1:1,000,000 (Düwel, et al. 2007 a&b; FISBo BGR, 2006) and more recently a soil map at a scale of 1:200,000 (Krug and Hartwich, 2011) on the basis of previously published soil maps of the former German Democratic Republic and the Federal Republic of Germany (Hartwich, et al. 1995). These available sources lack exhaustive information with enough accuracy to be used for decision making for environmental modeling, protection and at different scales (Behrens and Scholten, 2006; Gomez, et al. 2012). Hence, the need to promote a new and powerful pedometric approach to speed-up production of high-resolution soil property mapping in Germany should receive high priority (Behrens and Scholten, 2006). On the other hand, evaluation of spatial variability to produce soil property maps using geostatistical analysis requires large numbers of samples which relies on labor intensive and time consuming traditional analytical methods (Liu, et al. 2006). To cope with such limitations on acquiring large amounts of accurate data , infrared spectroscopic (IRS) techniques linked to partial least square regression (PLSR) can complement conventional soil analyses to quantify a broad range of soil properties (Rossel and Behrens, 2010; McDowell et al. 2012; Cécillon, et al. 2009). With accurate prediction of soil properties (McBratney, et al. 2003; Viscarra-Rossel et al. 2006), spectroscopic techniques could be well suited for assessing agro-ecosystems via integration with geostatistical analysis (Shepherd and Walsh, 2007).

Studies integrating IRS with geostatistics for soil property mapping have been limited to visible and/or near-infrared spectroscopy (V/NIRS). Wetterlind et al. (2008), Odlare et al. (2005), and

Brodsky et al. (2013) applied V/NIR-PLSR models to assess soil spatial variability and create SOM and clay content maps. However, all these studies were carried out over small spatial scales (field to farm scale) mainly investigating a single soil property. Especially at large spatial scales, diffuse reflectance infrared Fourier transform spectroscopy in mid-infrared range (midDRIFTS) generally has better predictive capabilities in regards to soil carbon fractions and texture than NIR and V/NIR (McCarty, et al. 2002; McCarty and Reeves, 2006; Soriano-Disla, et al. 2014). While midDRIFTS spectra contain fundamental molecular vibrations of major soil components, NIRS spectra are overtones and combinations of fundamental vibrations with less band specificity resulting from overlapping of different vibrations (Viscarra-Rossel and Lark, 2009). Hence, working on a regional scale with large heterogeneity, NIR spectra are much more influenced and lead to a bias in predicting of soil properties derived from PLSR calibrated models (Bellon-Maurel and McBratney, 2011). Cobo et al. (2010) was first to integrate midDRIFTS-PLSR predicted soil data to compute spatial patterns of several soil properties such as C and N fractions and soil texture across three villages in Zimbabwe. They strongly recommended coupling midDRIFTS to geostatistics as a promising way to assess soil fertility and create soil map at regional and global scales for future application. With regard to the above mentioned priorities as well as the need for improving the spatial resolution of soil databases in Germany, midDRIFTS linked to geostatistical analysis could be applied to better understanding soil spatial variability and mapping of soil properties over large scales.

In the current study, previously developed midDRIFTS-PLSR calibrated and validated models at the regional scale in South-west Germany, (see chapter 2), were used to predict total carbon (TC), total organic carbon(TOC), total inorganic carbon (TIC) and soil texture (clay, silt and sand) on regional-based soil midDRIFTS spectra. It was hypothesized that midDRIFTS-PLSR-based prediction obtained from regionally distributed soil spectra provide a complete data coverage across two regions to be integrated with geostatistics for accurate assessing the spatial variability and mapping of soil properties. The main objectives of this study were: i) to evaluate the ability of midDRIFTS-PLSR calibrated models to predict soil properties (TC, TIC, TOC, clay, silt and sand contents) at regional scale using only the soil spectra, ii) to assess the spatial variability of soil carbon and texture by coupled midDRIFTS-PLSR predicted data with geostatistics, iii) to map the spatial distribution of soil carbon and texture by using ordinary Kriging estimator and

iv) to compare the accuracy of the newly created maps to the currently available lower resolution maps.

3.3. Materials and methods

3.3.1. Study sites

The study areas were the regions of Kraichgau (K) and Swabian Alb (SA), in the Federal State of Baden-Württemberg, Southwest Germany (Figure 3.1). Each study region had an area of approximately 1600 km^2. The Kraichgau region is a fertile and intensively cropped hilly region which is situated in a geomorphological basin between 100 to 400 m a.s.l. It is characterized by a mild climate with a mean annual temperature between 9-10°C and mean annual rainfall ranges from 720 to 830 mm (Clemens and Stahr, 1994; Ingwersen, et al. 2011). The Swabian Alb is a low mountain plateau of South-western Germany with elevations between 500-850 m a.s.l. The region is generally characterized by a cooler climate with an annual average temperature of 6-7°C, mean annual precipitation between 800 and 1000 mm. Soils developed mostly on Jurassic limestone, and are rich in clay due to strong decalcification and weathering, and due the cooler climate also a higher in TOC content (2-3.3 %) compared to the Kraichgau soils (≈ 1 %) (Fischer, et al. 2010).

3.3.2. Sample collection and preparation

A probability-based sampling design (e.g. simple randomized and regular grid sampling (Peterson, et al. 1999) was used across both study regions to obtain a spatially dispersed, unbiased sample of soil variables which accounted for both small and large spatial scale variation. To select the sampling points, a land use map of the state of Baden-Württemberg (Ad-hoc-AG Boden, 2005) was overlaid with the study areas' boundaries in ArcView version 10.1 (www.esri.com). For the random selection a grid of 50×50 m was overlaid on the crop land vector layer of each study region separately. Thereafter, 300 sample points were randomly selected through the entire grid for each region. Additionally, 300 points were selected based on a regular 1.5×1.5 km grid sampling scheme for each region. This approach enabled us to gain a greater heterogeneity across each region (Luck and Wu, 2002). In total 1170 potential sampling points were selected across two regions: 580 in K and 588 in SA study region (Figure 3.1).

Finally, the map of the point vector layer for each study region was projected as a KML file in Google Earth (www.earth.google.com). However, as some points were found in unsuitable places for sampling (e.g. road, water way, household, pasture or forest) they were re-located (if possible) in adjacent locations within cropped fields. Soil samples from the topsoil (0 to 30 cm) were taken during 2012 and geo-referenced using a portable GPS (GPSMAP 62st) with a horizontal accuracy of 1 to 3 m. At each sampling point five sub-samples within a 15 meter radius were bulked to form a composite sample: one central and four others located at a distance of 15 m in the four cardinal directions. Soil samples were dried at 32 °C, sieved (<2 mm) and a sub-sample used for further analysis.

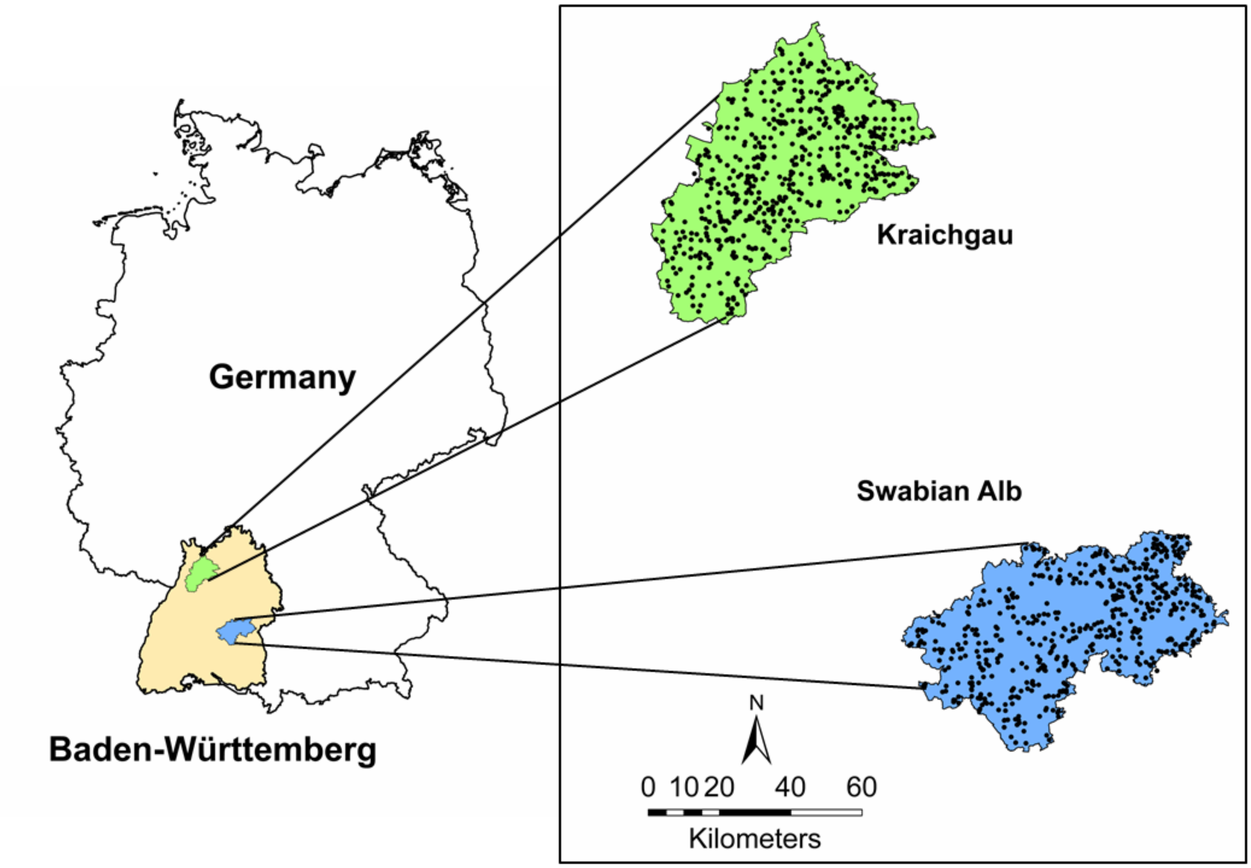

Figure 3.1. Study regions (Kraichgau and Swabian Alb) and location of individual sample points

3.3.3. MidDRIFTS analysis of soil samples and prediction

All 1170 bulk soil samples were ball milled and dried overnight at 32°C for midDRIFTS according to Demyan et al. (2012). Briefly, a ball-milled subsample was scanned on a Tensor-27 Fourier transform spectrometer (Bruker Optik GmbH, Ettlingen, Germany) equipped with a potassium bromide (KBr) beam splitter and a liquid nitrogen cooled mid-band mercury-cadmium-telluride detector. The spectra were obtained at least in triplicate in the mid-infrared range (4000-600cm^{-1}) by combining 16 individual scans at a resolution of 4 cm^{-1} recorded in absorbance units [log(1 /Reflectance)]. All spectra were averaged and baseline corrected. Additionally, the background CO_2 region (2400–2300 cm^{-1}) as well as both edges of each spectrum (>3900 and <700 cm^{-1}) were excluded.

Successful midDRIFTS-PLSR calibrated and validated models previously developed for both regions (see chapter 2) were used to predict total carbon (TC), total inorganic carbon (TIC), total organic carbon (TOC), clay (< 2 µm) and silt (2-63 µm) contents using soil spectra of samples taken during the current study. Since the model developed for sand content was not accurate, sand content was determined by subtracting the sum of silt and clay from 100. The predictions were done in OPUS software using the "Quant 2 Analyses / File list" function (Bruker Optik GmbH, Germany). The midDRIFTS-PLSR calibrated models were applied to unknown soil spectra to predict soil properties across both regions (n = 1170). According to Mark and Tunnell (1985) the predictive performance of midDRIFTS-PLSR calibrated models were judged using the principle of the Mahalanobis distance (MD) and then possible spectral outliers were identified. Therefore, each midDRIFTS-PLSR calibrated model is associated with MD threshold (MDT) calculated by the mean MD from all calibration spectra. In the current study outliers were identified by spectra for prediction which had a MD greater than three times the MDT according to the Gomez et al (2012). All spectral preprocessing and data prediction were carried out in OPUS, version 6.5 (Bruker Optik GmbH, Germany).

3.3.4. Conventional data analysis

Conventional statistics including mainly mean, range, standard deviation (SD) and coefficient of variation (CV) were performed to assess overall variability of properties before geostatitical analysis. Normal distributions of investigated variables are required to construct stable semivariograms (Webster and Oliver, 2001). Data first were tested for normality by the Kolmogorov-Smirnov (K-S) test and then non normal distributed variables were transformed to obtain approximate normality. Conventional statistics and test of normality were performed using SigmaStat 3.5 (Systat Software Inc., San José, CA, USA).

3.3.5. Geostatistical analysis of midDRIFTS-PLSR predicted soil properties

Geostatistical analyses were used to assess spatial variability and for spatial interpolation. In this study, the spatial variability of soil properties was evaluated via semivariograms which measured the average dissimilarity of data as a function of distance (Goovaerts, 1999) as illustrated in equation 1.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(i) - z(i+h)]^2 \qquad (1)$$

Where γ is the semivariance value for N data pairs (sample points) separated by a distance lag h in meters; and z the variable under consideration at positions i and $i+h$. Isotropic (omnidirectional) semivariograms were calculated using a lag distance of 1900 m as anisotropy was not detected in the variogram surface by small anisotropy ratios that were <2.5 (Wang, et al. 2010). Best fitting theoretical semivariogram models (Gaussian, linear, spherical, or exponential models) were chosen by the lowest residual sum of square (RSS) and highest coefficient of determinations (R^2) (Goovaerts, 1999; Wei, et al. 2008). Construction of semivariograms and model fitting were carried out with GS+ version 9 (Robertson, 2008). Semivariogram parameters of range, nugget (C_0), structural sill (C) and sill variance (C_0+C) were derived and used for further analysis. The range was the distance where the model reaches a horizontal asymptote indicating no spatial dependency beyond this distance. The y-axis value of the semivariogram model at the range was the sill. The nugget variance was the semivariogram value at zero separation distance indicating the variance at scales shorter than the sampling interval or random error (Heuvelink and Webster, 2001). The spatial dependency was determined using the nugget-to-sill ratio with Cambardella's classification (Cambardella, et al. 1994) which was modified by Cobo et al. (2010) with the following intensity of spatial dependency: very strong (nugget-to-sill ratio < 25%), moderately strong (nugget-to-sill ratio = 25 to 50%), moderately weak (nugget-to-sill ratio = 50 to 75%), very weak(nugget-to-sill ratio > 75%), and no spatial dependency (nugget-to-sill ratio = 100%).

Ordinary kriging (OK) as most used type of kriging was chosen in current study to map the spatial distribution of each soil properties as illustrated in equation 2.

$$Z_o = \sum_{i=1}^{N} \lambda(i)\, z(i) \qquad (2)$$

Where Z_0 is the estimated value of the soil property at the location of 0; z(i) are the known values at the sampling site i, and λ(i) is weighting factors estimated based on distance from the unsampled location using the semivariograms, N is the number of the nearest-neighbors used for estimation. The Kriged map of given soil properties was stratified into a grid with a cell size of 200 by 200 m using 12 samples as the neighborhood size. The map of SOM and texture was obtained for accuracy assessment described in section 2.4.4. To construct the SOM map, TOC values were first converted to SOM by multiplying by a factor of 1.724 according to Jackson (1958). To generate the soil texture map, silt and sand were converted from the German particle size classification to the World Reference Base particle size classification system (WRB, 2007) using R package "soiltexture". Interpolation was performed with ArcGIS version 10.1 using geostatistical analyst tools (ESRI, 2013). Soil texture maps were generated in QGIS version 2.2 (QGIS Development Team, 2014).

3.3.6. Accuracy assessment of kriged maps

The accuracy of kriged maps was assessed by mean standardized error (MSE), root mean square error (RMSE), and root mean square standardized error (RMSSE). The MSE measures the bias of estimation and values close to 0 indicate predictions are unbiased. Values close to 0 for RMSE indicate that predictions do not deviate much from measured values. Finally, RMSSE near to 1 indicates the standard errors are accurate and the correct semivariogram model was chosen (Marchetti, et al. 2012). The kriged maps of SOM and texture was further compared to the currently existing maps of SOM and texture at the scale of 1:1000000 (BGR_Legend_SOM and BGR_legend_texture) as reference according to the revised BGR Legend (Ad-hoc-AG Boden, 2005). Overall similarity between the two maps was estimated with the two-map comparison approach (Pontius, et al. 2008). The overall similarity (OS) is the ratio of intersection for each separate class (single GIS layer for each class) from currently available and newly kriged maps to the union of currently available and newly kriged maps. This ratio varies from 0% representing no overlap, to 100% showing a perfect overlap between two maps.

$$\mathrm{OS} = \frac{\sum_{i=1}^{N} \mathrm{B(i)}}{\mathrm{A}} \tag{3}$$

Where OS is the overall accuracy, A is the total area of either reference or kriged map, B is the overlapped area of each interest class, *N* is the number of class for property of interest.

3.4. Results

3.4.1. Soil carbon and texture predictions based on midDRIFTS-PLSR calibrated models

The midDRIFTS-PLSR calibrated models were used to predict the regionally distributed soil samples. A total of 582 and 588 midDRIFTS spectra of soils were used in prediction of soil properties in K and SA, respectively. Based on MD values obtained by midDRIFTS-PLSR calibrated models, most of soil spectra used in prediction resulted in MD values smaller than the MDT of the calibrated models. After outlier detection, there were for K TC, TIC and TOC 536, 510 and 558 spectra and for S TC, TIC and TOC 552, 514 and 572 spectra, respectively. In regards to soil texture, midDRIFTS-PLSR calibrated models resulted in accurate prediction of 551, 546 and 528 midDRIFTS spectra for clay, silt and sand in K, respectively. While prediction results for SA region were 562, 540 and 526 for clay, silt, and sand, respectively. Additionally, the predictive accuracy was double-checked for TC content by random sample selection (93 samples from total of 1170) across two regions for traditional laboratory measurement. The midDRIFTS-PLSR-based predicted values were plotted with those of measured values and resulted in coefficient of determination (R^2) 0.95 and 0.97 for K and SA regions respectively (Figure 3.2). With regards to the predicted values, generally higher carbon contents with a wider range were found in SA compared to the K region (Table 3.1). TC, TIC and TOC contents ranged from 0.33-6.5%, 0-2.8%, and 0.23-3.6% in K respectively while the values for SA were 0.9-10.5%, 0-2.7 and 0.83-5.5%. In terms of texture the average for clay and silt for K were 24.9% and 69.6% while SA region had 40% and 51% clay and silt, respectively. None of soil properties were normally distributed with silt and clay as exceptions. After log-transformation of TC and TOC, they were normal distributed ($p<0.05$) while TIC and sand content could not be converted to a normal distribution by any means of transformation. Hence, caution must be taken into consideration for interpreting the spatial variability and distribution for TIC and sand.

Table 3.1. Descriptive statistics of predicted soil properties from two regions: Kraichgau (K) and Swabian Alb (SA)

	TC %		TIC %		TOC %		Clay %		Silt %		Sand %	
Region	K	SA	K	SA	K	SA	K	SA	K	SA	K	SA
Min	0.33	0.9	0	0	0.23	0.83	6.6	14	11	26	0	0
Median	1.29	2.4	0.17	0.06	0.98	2.3	25	39	71.5	51	3.5	7.1
Max	6.5	10.5	2.8	2.7	3.6	5.5	51	82	82.9	81	82.2	26
Average	1.59	2.7	0.49	0.19	1.03	2.5	24.9	40	69.6	51	5.61	7.3
Range	6.18	9.6	2.8	2.7	3.9	4.7	44.4	68	72	55.4	82.2	26
SD	0.92	1.22	0.63	0.43	0.37	0.74	4.8	12	9.2	12	10.4	4.4
CV	58	45	129	230	36	30	19	30	13	24	185	060
N[a]	536	552	510	514	558	572	551	562	546	540	528	526

SD=Standard deviation, CV= Coefficient of variation, TC: total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon

[a] the total samples applied for prediction were 582 and 588 for K and SA regions respectively

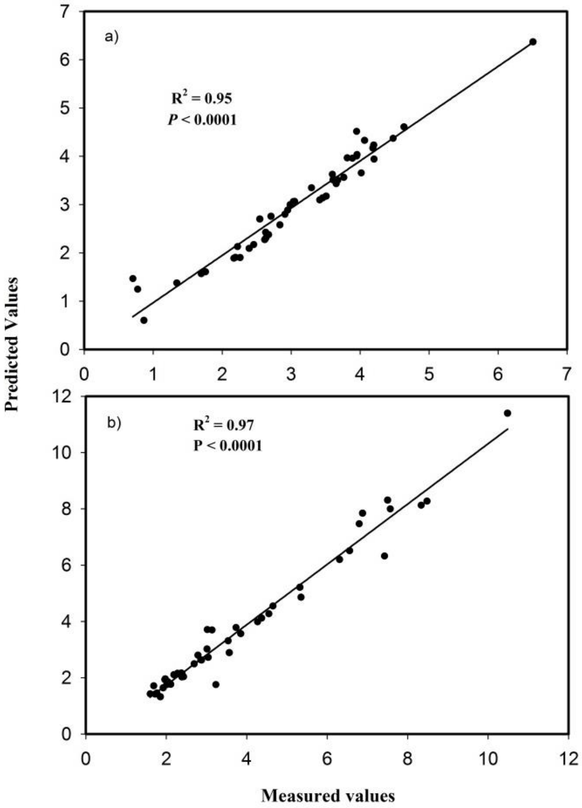

Figure 3.2. Measured total carbon (TC) and midDRIFTS predicted values for Kraichgau (a) and Swabian Alb (b) regions.

3.4.2. Analysis of spatial dependency of soil carbon and texture using geostatistics

The experimental semivariogram models of soil properties and corresponding parameters are presented in Table 3.2 and Figure 3.3 and 3.4. The experimental semivariograms indicated a clear increase of spatial dependency or dissimilarity (points on figure) with increasing distance between sampling pairs. The best fit geostatistical models were spherical for properties in K except for TIC and TOC which were fit with exponential model. In SA an exponential model was the best fit for TOC and sand, remaining soil properties were described better with the spherical model. The fitted models showed the spatial structure for properties according to the model parameters nugget, sill and range. There was a large nugget effect (random variance) for all soil properties in both regions except for silt and sand in K region (Figure 3.3 and 3.4). There were smaller nugget values for carbon fractions in SA region as compared to K region. Nevertheless, the soil texture content in K had smaller nugget values than SA region.

Table 3.2. Model parameters of theoretical semivariograms of evaluated soil properties in two study region.

Property[a]	Model	R^2	RSS	Nugget (C_0)	Sill (C_0+C)	Range (km)	$C_0/(C_0+C)$[b] (%)
				K			
TC	Sph.	0.95	0.001	0.22	0.34	30.5	65
TIC	Exp.	0.92	0.005	0.22	0.44	18.5	0.50
TOC	Exp.	0.95	0.0002	0.045	0.105	22	0.51
Clay	Sph.	0.93	24.7	12.87	27.1	28.5	0.47
Silt	Sph.	0.93	1104	8.68	80.36	19.8	0.13
Sand	Sph.	0.95	1120	19.7	122.1	21	0.16
				SA			
TC	Sph.	0.98	0.0004	0.076	0.158	15.6	0.43
TIC	Sph.	0.94	0.001	0.085	0.22	24.5	0.34
TOC	Exp.	0.91	0.0002	0.04	0.09	14.3	0.42
Clay	Sph.	0.98	320	72	157	18.8	0.46
Silt	Sph.	0.99	165	67.6	151	16.5	0.45
Sand	Exp.	0.90	15.7	10	21.2	8.7	0.47

Sph= spherical, Exp= exponential, TC: total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon, K: Kraichgau, SA: Swabian Alb, RSS= residual sum of square

[a] Model parameters for TC and TOC are based on log-transformed data

[b] Degree of spatial dependency

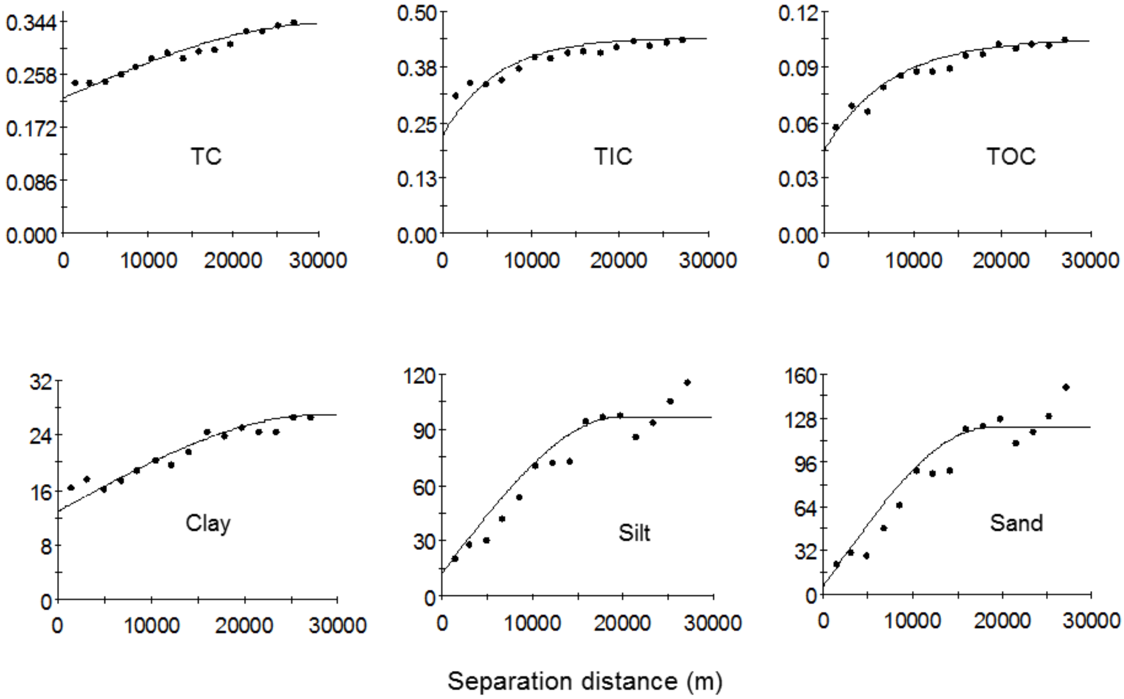

Figure 3.3. Experimental isotropic semivariograms (squares) and theoretical fitted models (line) for evaluated soil properties in Kraichgau study region. TC: total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon

The degree of spatial dependency, the nugget-to-sill ratio, varied between soil properties across the two regions ranging from 13 to 64%. Spatial dependency (nugget-to-sill ratio) of soil carbon (TC, TIC and TOC) were classified as moderately weak in K (ratio 50-64%) while moderately strong spatial dependency (25% ≤ ratio < 50%) were obtained in SA for these properties. The ratio for soil texture revealed moderately strong dependency (47%) for clay and very strong spatial dependency for silt (13%) and sand (16%) in K, while moderately strong spatial dependency was found for texture in SA with a ratio ranging from 45 to 47% (Table 3.2). The coefficient of variation (CV) was also used as an indicator of variability. However, there was no correlation between CV and nugget-to-sill ratio (Figure 3.5). A variable with a high CV did not necessarily have a high spatial dependency. For instance, TIC in both regions had extremely high CV's (129 and 230 %) while geostatistical analysis revealed moderately weak and moderately strong spatial variability in K and SA, respectively. The CV did not identify the spatial variability of measured soil variables at the un-sampled locations estimated via spatial analysis.

Spatial ranges varied from 18.5 km (TIC) to 30.5 km (clay) in K region and from 8.7 km (sand) to 24.5 km (TIC) in SA region. Beyond these ranges, there was no spatial dependence. Generally, estimated ranges for properties were shorter in SA than in K region except for TIC. Sand content in SA showed the shortest range (8.7 km) while the longest was obtained for TOC (30.5 km) in K region.

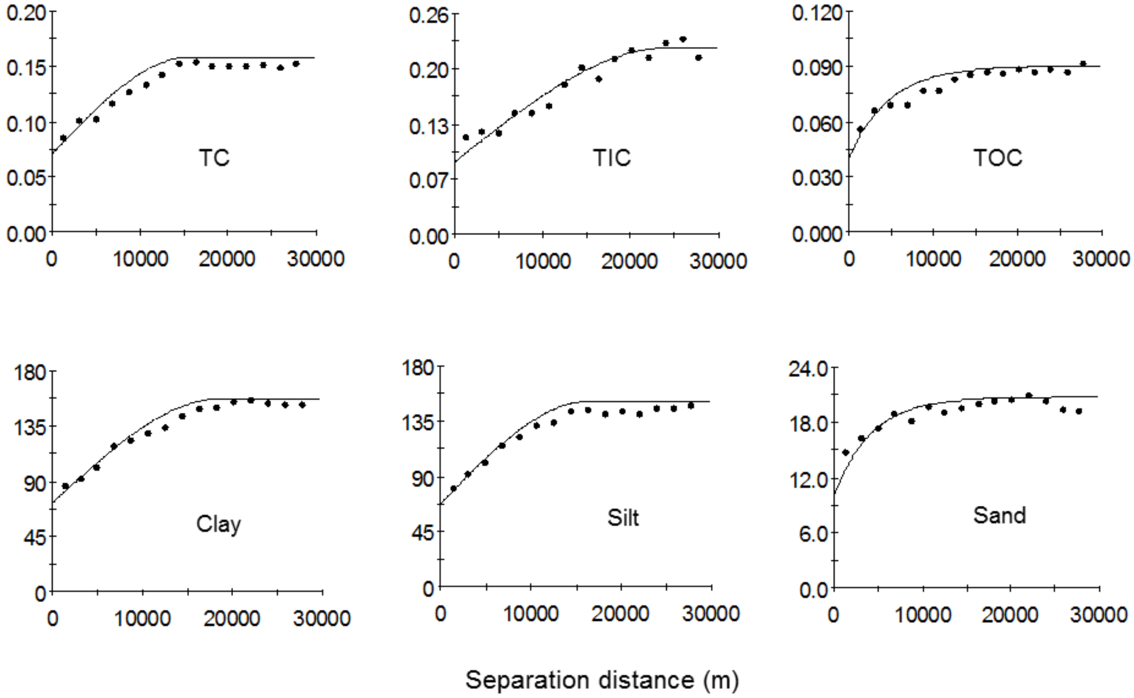

Figure 3.4. Experimental isotropic semivariograms (squares) and theoretical fitted models (line) for evaluated soil properties in Swabian Alb study region. TC: total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon

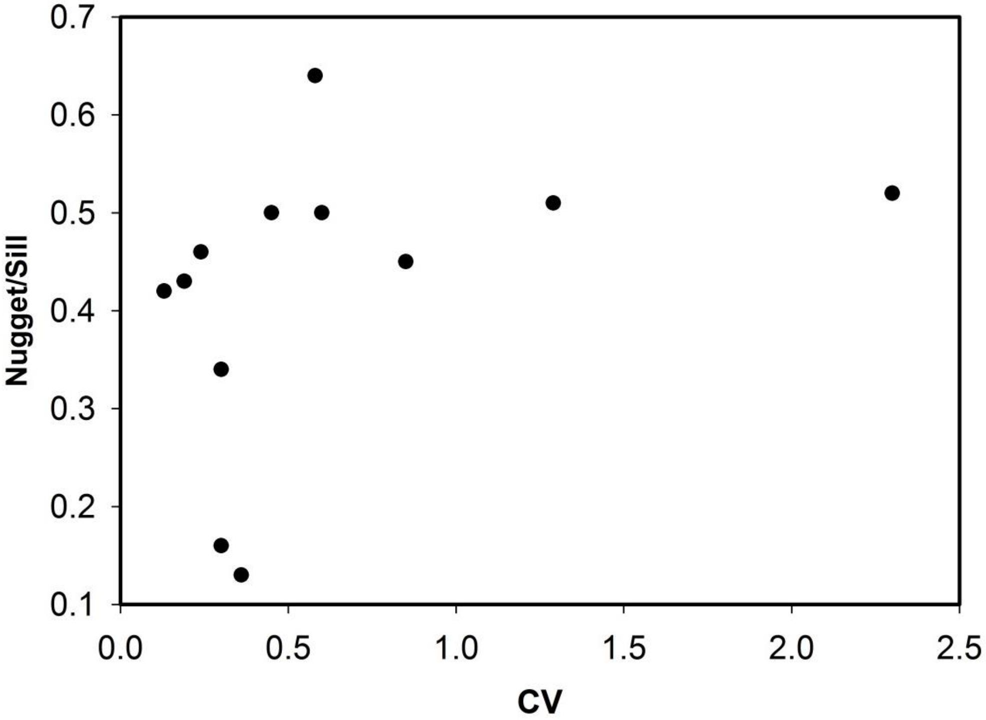

Figure 3.5. Correlation between spatial dependency indicators (Nugget-to-sill ratio) and traditional statistics-based overall variability index (CV).

3.4.3. Spatial distribution of soil carbon and texture

Figures 3.6 and 3.7 show the kriged maps of soil properties in K and SA using the fitted semivariograms parameters in ordinary Kriging (OK). The kriged maps revealed several distinct spatial patterns in both regions. Generally, TC contents decrease from west to east except in some isolated areas in between. TC content showed had highest values (2.59-6.51%) in the west which latter extended toward center part of area. Lowest TC contents (< 1%) were observed in east and north-east of K region. Except for a small area in west and south with TOC > 1%, the rest of the area showed very low TOC contents <1%. However, spatial similarity between TC and TOC was limited to the areas which contained low TIC contents. It also verified the positive significant correlation between these two variables which has been shown previously (see chapter 2). The spatial distribution of TIC differed from that of TC and TOC. There was a higher concentration band from south-west to north surrounded by low TIC areas decreasing towards the edges of the region. While the spatial distribution of sand content showed relatively less heterogeneity except in a small area in west, silt and clay contents showed strong spatial dissimilarity. Likewise, a similar spatial distribution between TC and TOC was found in SA region. High concentrations for TC in entire SA region were associated with high TOC concentrations representing a positive correlation between the two variables. Low concentration areas were limited only to the eastern part and a small area in the south-west. The distribution pattern of TIC had a north to south increasing trend at the regional scale in SA region. The kriged map of sand in SA region appears more inconsistent than those of other soil properties. It could have been also due to the low spatial dependency as shown by semivariograms. Likewise to K, the distribution map of silt content was completely reversed to that of clay content. The highest clay contents occurred in a large area in the center which continued to the south as a narrow strip.

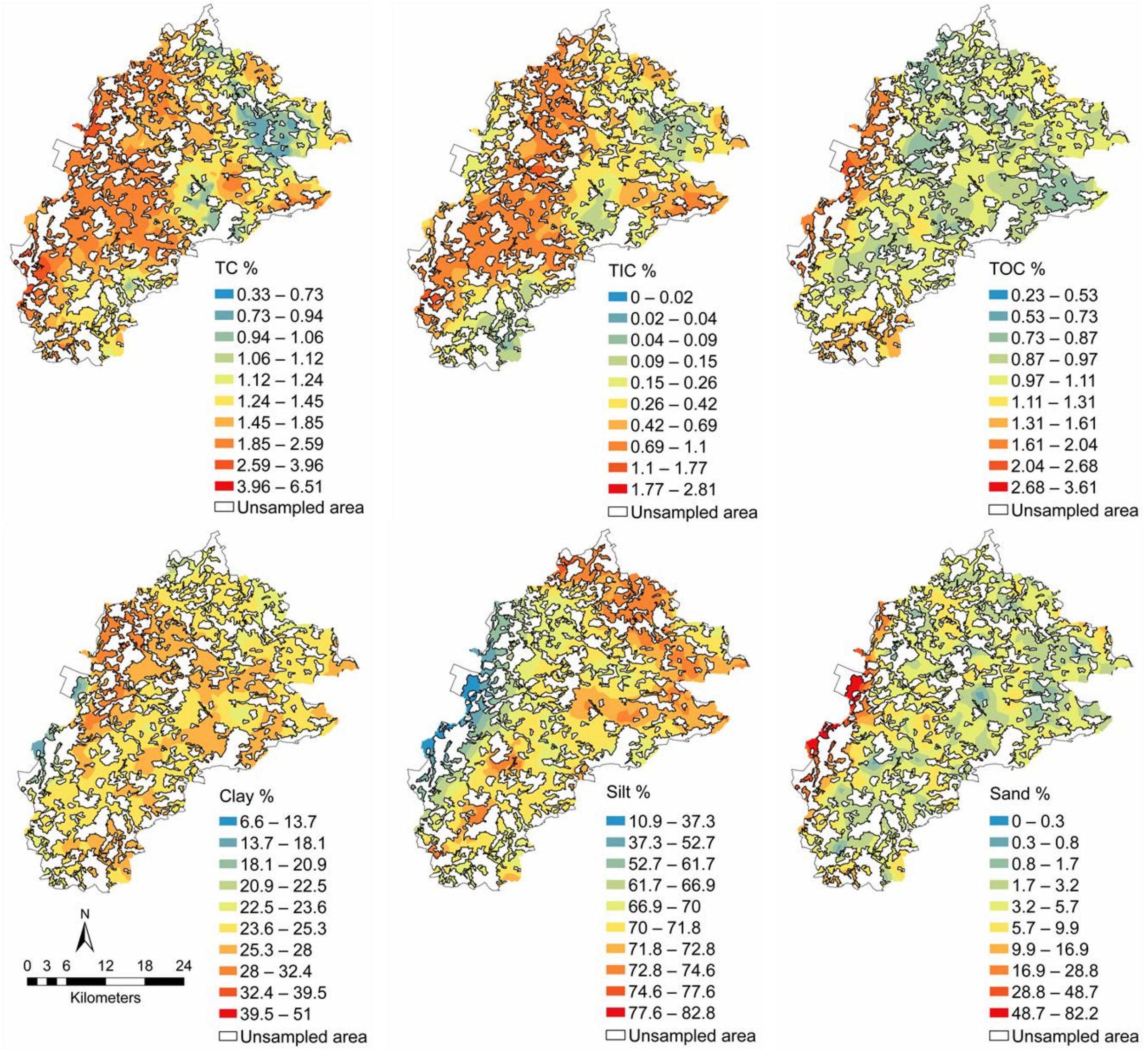

Figure 3.6. Spatial distribution maps of investigated soil properties kriged by ordinary kriging (OK) for Kraichgau region. TC: total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon

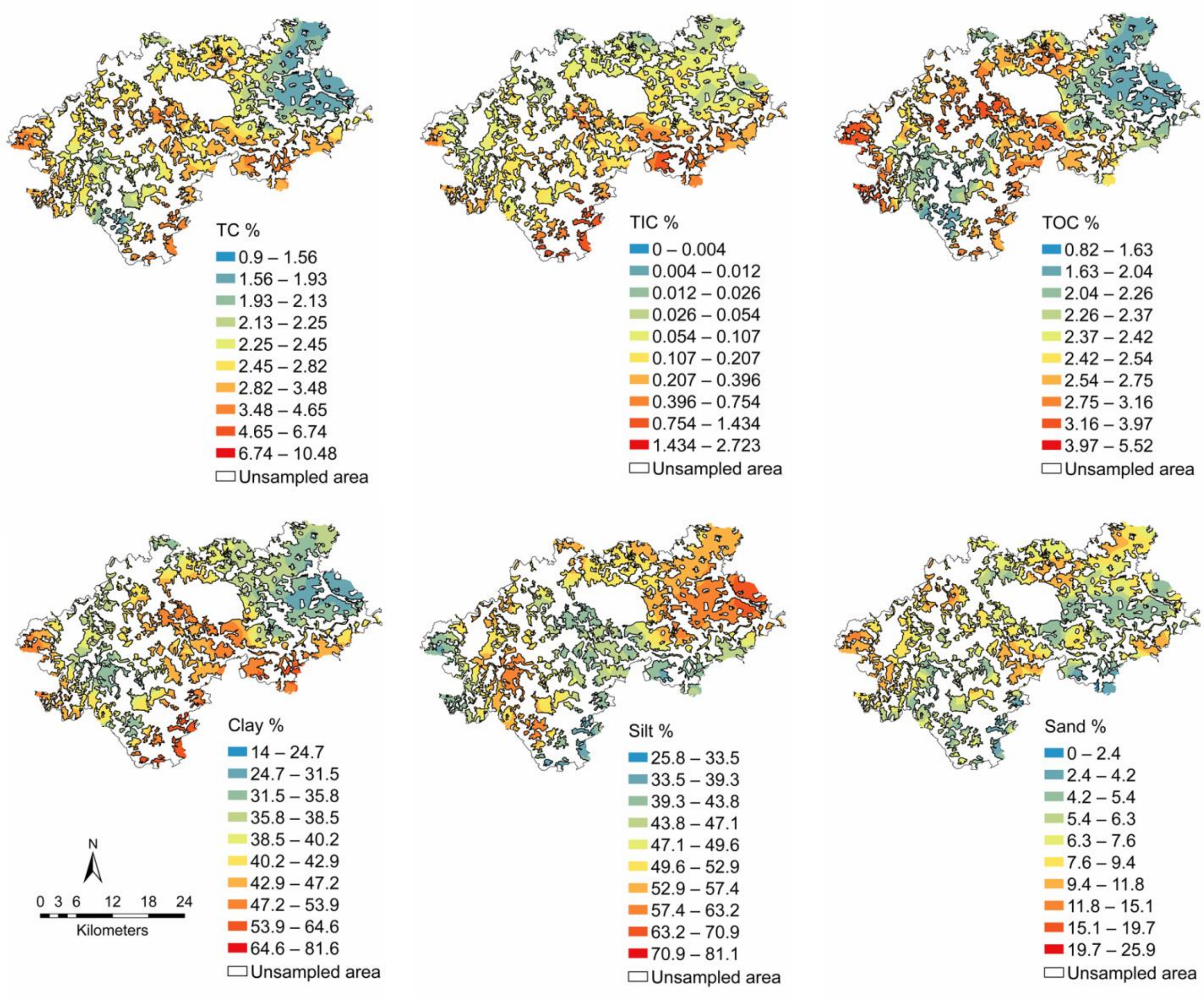

Figure 3.7. Spatial distribution maps of investigated soil properties interpolated by ordinary kriging (OK) for Swabian Alb region. TC: total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon

3.4.4. Accuracy of kriged maps

The prediction accuracy of estimated values using OK was evaluated using RMSE, MSE and RMSSE for different soil properties in both regions (Table 3.3). All soil properties in both regions were estimated well with MSE and RMSE values close to 0 and 1 respectively. The values of MSE were close to 0 for all soil properties in both regions ranging from -0.007 to 0.03 and -0.0001 to 0.002 for K and SA respectively. Values of RMSSE close to 1 indicate proper model fitting of semivariograms then used for interpolation during OK. Therefore, in K region, except for sand with a RMSSE value of 1.4, the remaining soil properties had values close to 1 indicating high kriging accuracy. The RMSSE values were also close to 1 for investigated soil properties in SA with the exception of TIC with 1.2. The estimation for sand content in K with a slightly higher RMSSE value may be attributed first to the low average contents and existence of extremely high values in the dataset.

Table 3.3. Accuracy assessment of Kriged maps of investigated soil properties by ordinary kriging

properties	RMSE		MSE		RMSSE	
	K	SA	K	SA	K	SA
TC	0.88	1.06	-0.002	-0.03	1.1	1.1
TIC	0.6	0.4	-0.002	0.002	1.1	1.2
TOC	0.29	0.64	-0.007	-0.02	1.09	1.08
Sand	6.5	3.9	0.03	0.008	1.4	0.99
Silt	5	9	-0.02	-0.001	1.1	1.04
Clay	4.6	9.4	0.01	-0.0001	1.1	1.04

RMSE: Root mean square error, MSE: Mean square standardized error, RMSSE: Root mean square standardized error,

TC: total carbon, TIC: Total inorganic carbon, TOC: Total organic carbon; K: Kraichgau; SA: Swabian Alb

The similarity assessment between existing maps and newly kriged maps was assessed both visually and statistically using two-map comparison procedure (Figure 3.8 and 3.9). A visual inspection of the SOM kriged map and BGR_Legend_SOM map in K region (Figure 3.8) indicated considerably less dissimilarity in SOM contents across the region. The overall similarity between two maps was 14%. The main difference observed was the SOM content class of 1 to <2% which was not present in reference map, while approximately 80% of total surveyed area was kriged with SOM contents 1 to <2%. Conversely, the SOM kriged map for SA matched well with the BGR_Legend_SOM map with an overlap of 65%. Likewise, the SOM class 3 to <4% was not present in the reference map while roughly 33% of whole area was class 3 to <4%

in the kriged map. Both maps of SA were dominated by SOM contents 4 to <6% which also had the highest overlap between two maps (Figure 3.9).

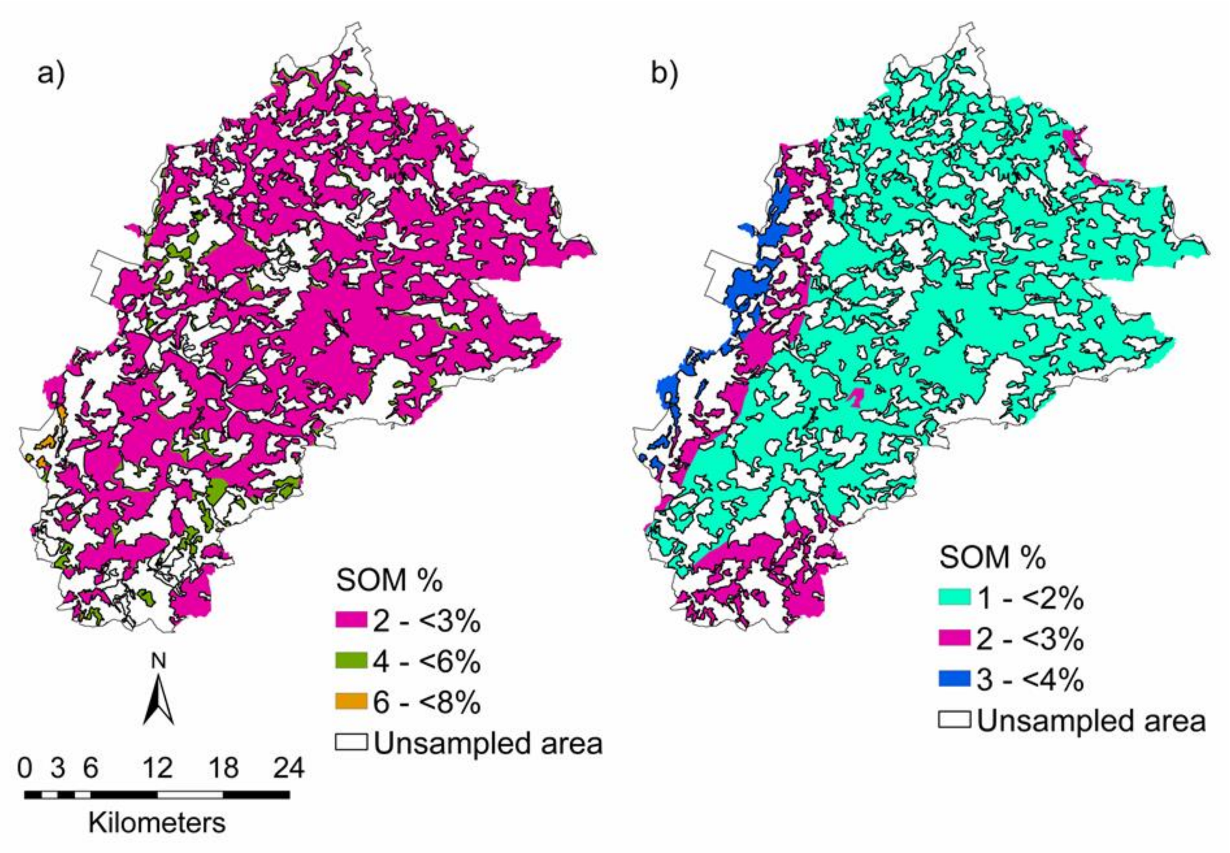

Figure 3.8. Comparison between the reference map of soil organic matter (SOM) (a) and map of SOM interpolated by ordinary Kriging (b) for Kraichgau region.

The texture map of the K region was segmented into six soil texture classes according to the World Reference Base particle size classification system. Comparison of the similarity assessment for the K study region showed that the soil texture classification had an overall similarity of 92%. In both maps, silt loam was the dominant, representing roughly 90% area (Figure 3.10). There was a large difference between soil texture maps in SA region. As it is shown in Fig. 11, the kriged soil map was classified into four texture class dominated by silty clay and silty clay loam class while clay texture was the dominant class in the

BGR_Legend_texture map, covering roughly 94% of total area. The overlap of the two maps was only 11% (Figure 3.11).

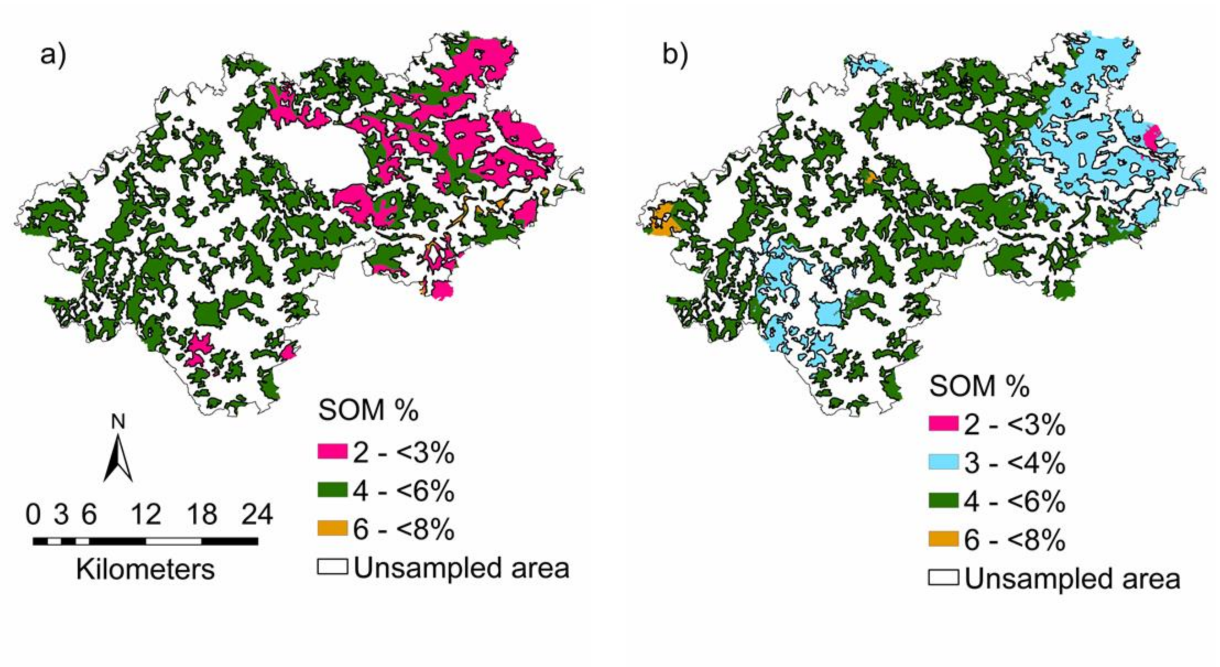

Figure 3.9. Comparison between the reference map of soil organic matter (a) and map of SOM interpolated by ordinary Kriging (b) for Swabian Alb region.

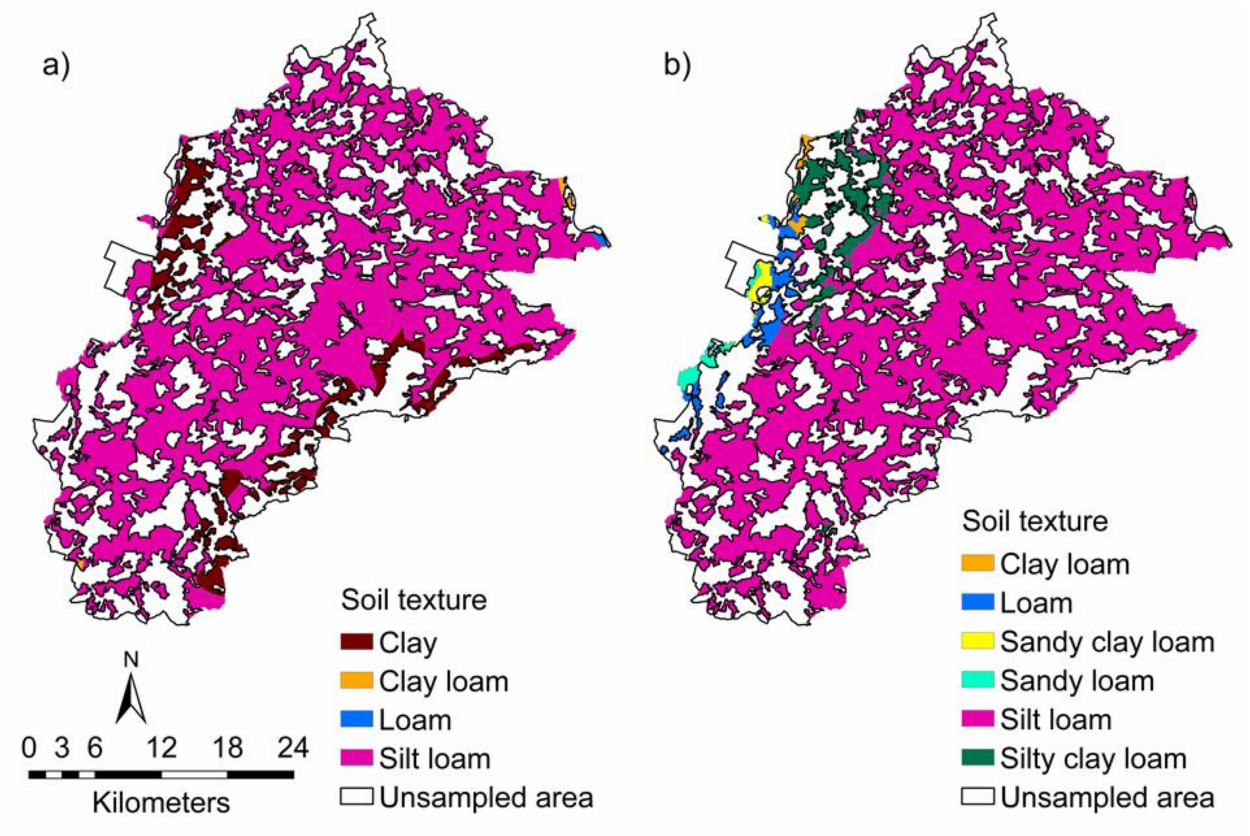

Figure 3.10. Comparison between the reference map of soil texture (a) and soil texture map interpolated (b) for Kraichgau region.

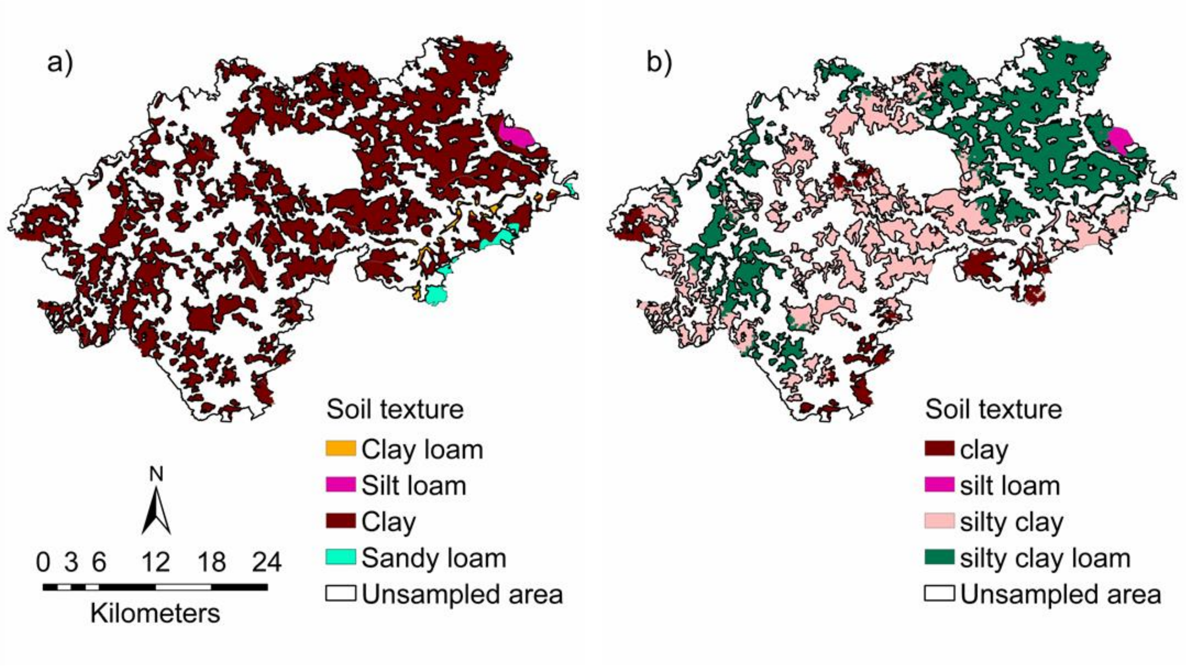

Figure 3.11. The reference map of soil texture (a) and kriged map of soil texture (b) for Swabian Alb region.

3.5. Discussion

3.5.1. Feasibility of midDRIFTS-PLSR models in the prediction of soil carbon and texture: input data acquisition for geostatistics

Up to present, application of midDRIFTS-PLSR model has been used for accurate predictions of a wide range of physical, chemical and biological soil properties (Shepherd and Walsh, 2007; Linker, 2011). However, feasibility of midDRIFTS-PLSR models for predicting soil properties using a regionally distributed sampling for geostatistical analysis in not known. In the current study, soil carbon and texture contents were predicted accurately based on previously developed midDRIFTS-PLSR calibrated and validated models (see chapter 2) applied at regional scale. In the current study, from a total of 1170 soil spectra across both regions, an average of 93% was precisely predicted. Prediction accuracies for data obtained for approximately 1090 midDRIFTS spectra were precisely evaluated via MD values. The predictions were first based on the most accurate midDRIFTS-PLSR calibrated models at regional scale with R^2 values 0.99 for soil carbon (TC, TIC and TOC) and 0.97 and 0.92 for clay and silt (see chapter 2) which all ranked as excellent models (Cobo, et al. 2010). Secondly, in the current study, prediction of these properties based on midDRIFTS spectra was further evaluated by using MD, resulted in accurate predictions except for roughly 7% of total midDRIFTS spectra which were outliers. These outliers might be attributed to the midDRIFTS spectra with extreme values which do not fall in calibration range (Soriano-Disla, et al. 2014). Prediction performance of midDRIFTS-PLSR calibrated model of TC was further checked when the predicted TC of 93 unknown midDRIFTS spectra were plotted against laboratory measured values of those samples. The results in the current study showed the high potential of midDRIFTS-PLSR calibrated models as a rapid-throughput method to predict soil properties of future unknown midDRIFTS spectra which are required as primary data input for assessing soil spatial variability and mapping.

The midDRIFTS-PLSR-based predicted dataset provided the foundation for geostatistical analysis to update our spatial knowledge of soil C and texture. Hence, how the properties are spatially correlated and distributed depend on pedogenic factors and land management that will be discussed in section 3.5.2. The regionally distributed midDRIFTS-PLSR-based predicted dataset was able to describe the spatial dependency and range of investigated soil properties clearly across both regions. Although application of midDRIFTS-PLSR is a well-known technique in the prediction of variety of soil properties, the use of spectral data in assessing spatial analysis and mapping purposes is less explored than NIR or Vis-NIR spectroscopy mainly at field or farm scale (Odlare, et al. 2005; Huang, et al. 2007; Gomez, et al. 2012). Scale issues and predicative performance of midDRIFTS-PLSR models especially at regional scale with large heterogeneity has shown its superiority compared to other studies applied V/NIRS technique. Wetterlind et al. (2008) found a $R^2 = 0.8$ and 0.89 for clay and SOM respectively with NIR-PLSR models, while we had more accurate models with $R^2 = 0.99$ and 0.97 for TOC and clay contents, respectively. Similarly, Brodsky et al. (2014) created a SOC map based on a V/NIR-PLSR model with $R^2 = 0.85$ with 128 calibration samples, which was less accurate compared to models in current study. Integration of midDRIFTS-PLSR with geostatistics was first applied by Cobo et al. (2010) in assessing spatial variability of C and N fractions at the landscape scale with comparable model accuracies to the current study, although they did not apply it for creating soil property maps. Therefore working at regional scale with large heterogeneity, NIR spectra are much more influenced by soil heterogeneity and lead to biases in predicting soil properties derived from midDRIFTS-PLSR calibrated models (Bellon-Maurel and McBratney, 2011). The prediction results of the current study showed that midDRIFTS spectroscopy can be a source of information for assessing soil spatial variability and mapping of soil properties at the regional scale. This is in line with the conclusion already drawn by Shepherd and Walsh (2007) for the broad application of geo-referenced spectroscopic data for environmental assessment and monitoring.

3.5.2. Implications of spatial pattern analysis using midDRIFTS-PLSR integrated with geostatistics

The experimental semivariograms for assessing spatial variation resulted in isotropic (omnidirectional) semivariograms with a lag distance 1900 m as anisotropy was not detected. Thus, the semivariograms depended only on the distance between sampling points (Burgos, et al. 2006). In both regions, depending on the property, spherical or exponential models explained the experimental semivariograms and indicated the presence of spatial variation. The fitted models indicated the presence of any spatial pattern and its scale. Apart from region and property, the presence of a nugget effect was attributed to the variability at short range, measurement and sampling error and random and inherent variability (Tesfahungen, et al. 2011). The presence of a nugget effect in the current study reflected the complexity of the landscape, with regards to parent material, topography, soil type and agricultural management, which can affect spatial variability within the distances smaller than sampling distance. Therefore, a proportion of the total variance was the nugget effect which cannot be explained by geostatistics in current study and as a consequence, the spatial dependency decrease by increasing in nugget effect (Gallardo and Paramá, 2007). In the current study, the proportion of total variance, which was correlated spatially (structural variation), revealed strong and moderately strong spatial dependencies for soil texture in both regions indicating that intrinsic factors such as parent material, climate, mineralogy and soil forming processes were the main causes (Wang, et al. 2010). However, moderately weak spatial dependency might be controlled by extrinsic variations such as fertilizer application, tillage, and land use changes and other management practices (Cambardella, et al. 1994; Wang, et al. 2010). Therefore, moderate spatial dependence for TC, TIC and TOC in K region might be influenced by management practices rather than intrinsic factors. Hence, soil C fractions in SA with moderately strong spatial dependency seem to be influenced much less than in K region.

The range of the semivariograms represents the length of spatial dependency and beyond this range the variable is no longer spatially correlated. Hence, with increasing range, the scale of the spatial dependency in which a variable is correlated increases (Zhang, et al. 2014). It is considered an indicator of sampling efficiency and used for a selection of an efficient sampling strategy in mapping of soil properties (Fu, et al. 2010). In the frame of landscape investigations the ranges of different soil properties vary widely due to factors affecting soil erosion and deposition such as land cover, topography, parent material and anthropogenic activities within the study region. The results from the current study were consistent with several studies that reported large differences in ranges of different soil properties over multiple spatial scales (Odlare, et al. 2005; Cobo, et al. 2010; Wang, et al. 2010). In the current study, ranges of spatial dependency differed between properties, ranging from 8.7 km to a maximum of 30.5 km. Ranges for K region were larger than those obtained for SA region. Larger ranges in K coincided with relatively large homogeneous areas in the spatial distribution maps, while these were lacking in the SA region which had shorter ranges. Indeed, soil properties with a larger range might be influenced by natural and anthropogenic factors over greater distances than those of properties with smaller ranges (Lopez-Granados, et al. 2002). In the current study, relatively greater ranges were obtained for soil properties in K than in SA except TIC. This indicates that soils in K were related to neighboring soils over greater distances as compared to the soil parameters with smaller ranges in SA. Therefore, spatial variability of different soil properties described by varied ranges is a function spatial domain (region) and the spatial scale. In this regard, the sampling distance is viewed as the most crucial factor which may affect the spatial dependency and interpolation (Mulla and McBratney, 2000). When the length of the spatial dependency (range) is much longer than the sampling interval, this is an indication that the sampling strategy was successful in capturing the spatial variability of desired variables (Fu, et al. 2010; Wang, et al. 2010). As the ranges in current study were considerably longer than sampling distance, it indicates the usefulness of sampling in terms of size and distribution to explain spatial variability of investigated soil properties. Thus, intensifying sampling would not be needed for a future regional sampling investigation, although the proportion of spatial variability appeared within distances smaller than the sampling distance which is still unexplained (micro-scale variability). This is known as nugget effect which is also attributed to measurement noise.

3.5.3. Improvement over existing soil property maps using midDRIFTS-PLSR

In the current study integration of midDRIFTS-PLSR with geostatistics produced kriged maps comparable to maps having previous pedological knowledge of the study areas. On the other hand, the spatial patterns of soil properties corroborated previous correlation results between soil properties based on soil samples used in model calibration (see chapter 2). First the quality of kriged maps based on OK was assessed with high accuracy and resulted in RMSSE's close to 1. Although sand in K and TIC in SA showed relatively lesser accuracy as compared to the others, this might be due to the contribution of extremely high and/or low values in the dataset. In K region, while most of the area is dominated by silt loam texture with sand content <20%, there was a small area on the western edge of the study area in Rhine Valley which was very sandy (sandy clay loam and sandy loam) surrounded by soils with very low sand content. The concentrations of TIC mainly derived from parent material and lime applications for adjusting soil pH, were most variable as indicated by its CV. In such a case, the smoothing effect of the kriging method might the possible cause of higher RMSE (Ren, et al. 2008). Secondly, both non-normal distributions were not corrected by any data transformation method. Hence, lower accuracy Kriged maps for TIC and sand were expected as compared to other variables with normal distributions, which are required to construct stable semivariograms (Webster and Oliver, 2001).

Comparing the maps of SOM and texture generated in current study with those of currently existing maps revealed some dissimilarity which can be discussed and addressed due to the data resolution and the method for map creation which are based on. The currently existing maps for SOM and texture maps were generated mainly via pedotransfer functions, which estimate the variable contents based on knowledge of parent material, climate and land use as well as relying on a low number of sampling points (Düwel, et al. 2007a,b). The map scale of 1:1,000,000 is very coarse even at the regional scale. However, generated maps in the current study were based on 126 measured values based on conventional laboratory analysis and 1170 midDRIFTS-PLSR-based predicted values for an area of approximately 1367 km^2 of arable land. To note, the reference maps of SOM were based on values from other land uses such as pasture and forest with greater SOM contents in addition to arable land and it may have influenced the estimate for arable lands. Besides, the dissimilarity between reference and kriged maps of SOM especially for

K regions might be attributed to depletion of SOM especially in arable lands caused by erosion, soil management practices. The soils of K region were mainly formed from loess which is very susceptible to erosion especially in crop lands which have been under intensive crop management (Clemens and Stahr, 1994). Concerning the differences observed between the newly generated soil texture maps and reference maps in SA region, it should be noted that a large part of the area (roughly 95%) was classified as clay texture in the reference map which was the largest dissimilarity. The concentration values for clay, silt and sand content for clay texture in reference map are 47%, 25% and 28% (Düwel, et al. 2007b) which seems to be very unlikely based on measurements in current study. The main differences may be caused by the low silt and high sand contents in the reference map. In the current study, the texture map was mainly silty clay and silty clay loam textures with sand contents <20% and silt contents >40%, while in the reference map relatively higher sand and less silt contents are estimated for the entire study area (clay texture). Furthermore, soils in SA are formed mainly from carbonate limestone with very rocky soils in sloping areas which were not covered in our sampling as only croplands were included in the current study, which had low sand contents. Newly created maps of SOM and texture in the current study provided better spatial information as compared to currently available maps. Therefore, soil property maps currently lacking at regional scale can be created with high resolution by the approach used in current study. The map for soil inorganic carbon, for instance, which is mainly limited only to single fields, provides an accurate source for lime requirements and pH adjustment strategies. All these results confirm midDRIFTS potential as a rapid-throughput alternative to assemble large datasets required in geostatistics for generating high resolution soil maps.

3.5.4. Relevance for future agro-ecological assessments

Effective sample size for future studies: Spatial variability of soil properties and their mapping at large-scales depends on the density and distribution of soil observations (Grunwald, 2011). On the other hand, the number of samples required to describe the soil spatial variability depends on several factors such as spatial latent autocorrelation, locations of points, and extent (Griffith, 2005). Thus, to plan an effective sampling strategy for spatial statistics in areas of similar conditions, two main issues should be taken into account, effective sample size (ESS) and sample spatial distribution. Although an approximate minimum number of samples can be calculated at different probabilities (Cobo, et al, 2010), this does not indicate how the samples should be distributed in space. Based on semivariograms ranges obtained in the current study, effectiveness of sample size and its distribution was explained. When the length of the spatial dependency (range) is much longer than the sampling interval, this is an indication that the sampling strategy was successful in capturing the spatial variability of desired variables (Fu, et al. 2010; Wang, et al. 2010). Thus, the sampling design used in current study, based on both random and stratified design, showed its feasibility in capturing regional spatial variability in the selected regions. Thus, intensifying sampling would not be needed in current study for investigated soil properties for future sampling design for regional investigations, although the proportion of spatial variability appears within the distances less than sampling distance which is still unexplained or micro-scale variability. This is known as nugget effect which is also attributed to the measurement noise. Nevertheless, spatial variability assessment and kriged maps for sand and TIC with lower accuracies might be improved by re-distribution of sample points across regions. Uncertainty propagation analysis (UPA) would be a possibility to optimize the sampling distribution. It refers to the uncertainty associated with the kriging estimator (interpolation processor) for mapping which is unavoidable and associated with geospatial data (Mowrer and Congalton, 2000). Nevertheless, once the uncertainty propagation on kriged maps is recognized, the spatial dependency might be captured effectively by a restructured sampling strategy; intensifying sampling points in areas with very high uncertainty (Brodsky, et al. 2013). It may be a possible way to explain local variation and decrease the uncertainty on kriged maps as well. Besides, lesser sampling densities (thinning of sample points) might be applied within the areas with low spatial variability with relatively smooth spatial distribution.

Utilization of Auxiliary information: Application of auxiliary information which is correlated spatially to soil properties would improve the spatial analysis and mapping of other soil properties (McBratney, et al. 2003; Zhang, et al. 2012). These can be terrain attributes such as topography (Mueller and Pierce 2003), categorical variables such as soil type and land use (Zhang, et al. 2012) and remote-sensed data such as Landsat imagery measurements (Huang et al 2007). This would provide a simple and inexpensive source of secondary information to improve the soil maps under investigation. For example, the SOM map developed in the current study does not provide information concerning soil erosion for instance. However, integration of SOM data to a digital elevation model (DEM) as secondary information may indicate soil erodibility. Thus, integration of midDRIFTS-PLSR-based predictions and remote sensed data as auxiliary environmental information would be a feasible to be linked to geostatistical analysis for high-resolution digital soil mapping.

3.6. Conclusions

In the current study, midDRIFTS-PLSR was utilized as a rapid-throughput method for providing high quality predictions of a large regionally distributed soil dataset required to investigate spatial variability of soil carbon fractions (TC, TIC and TOC) and soil texture contents (clay, silt and sand). The study confirmed that midDRIFTS-PLSR was an effective tool for providing data coverage to model spatial variability and high resolution regional scale mapping. The experimental semivariograms provided a suitable evaluation of modeling spatial variability in both regions which confirmed the suitability of the sampling design and density applied to the regions. It also revealed the degree of variability which can be caused by either extrinsic or intrinsic factors. Based on the spatial dependencies and separation distances (range) obtained for the two regions, soil carbon fractions in Kraichgau are mainly influenced by land use management strategies (fertilizer application, tillage, and land use changes and other management practices) while intrinsic factors such as soil type, mineralogy and natural soil forming processes may be influencing soil properties in Swabian Alb. The effort further resulted in the development of high resolution kriged maps for soil properties with greater accuracy as compared with existing maps. The differences were due to the different scales and numbers of observations the maps are based on. These results highlight the potential of midDRIFTS-PLSR to be used for spatial analysis and digital soil property mapping in combination with geostatistics as a reliable approach to reduce the number of conventional laboratory analyses. Due to the high accuracy, kriged maps for soil properties can be used not only to update soil property maps for the Kraichgau and Swabian Alb study regions but also to facilitate appropriate and efficient regional land use planning and as inputs for modeling SOM dynamics on the regional scale.

Chapter 4

Application of midDRIFTS-based spectral indices to investigate the composition of soil organic matter (SOM): how carbonate interferes with peaks assigned to SOM

4. Application of midDRIFTS-based spectral indices to investigate the composition of soil organic matter (SOM): how carbonate interferes with peaks assigned to SOM

4.1. Abstract

The knowledge on spectral features provides appropriate information on soil organic matter (SOM) quality and composition and a basis for multivariate quantitative measurements. This study evaluated feasibility of specific peak areas obtained by diffuse reflectance infrared Fourier transform spectroscopy in mid-range (midDRIFTS) to be used as indicators of SOM quality. Spectral peak area at specific peak obtained by integration with a locale baseline and correlated to biochemical data of bulk soil across two contrasting agroecological regions (Kraichgau (K) and Swabian Alb (SA)). A total of 126 soil samples was analyzed for different SOM pools (C_{mic}, N_{mic}, C_{HWE}, N_{HWE}, TOC and TN), C:N ratio, TIC and by midDRIFTS as well. MidDRIFTS peak areas corresponding to SOM functional groups (2930, 1620, 1520 and 1159 cm^{-1}) were assessed to study the composition of SOM. Results examining relative distribution of peaks between two regions showed that, sum of the relative peak area at 2930 cm^{-1} was larger in SA with compared to K indicated differences regarding soil type and organic C contents. In contrast, larger peak areas at 1620, 1520 and 1159 cm^{-1} obtained for soils from K region. The peak assigned for aliphatic C-H bond (2930 cm^{-1}) could provide appropriate index to investigate SOM fractions if interference signal from carbonate was taken into consideration. While weak regression performance obtained between the peak at 2930 cm^{-1} and SOM fractions (e.g., $R^2 = 0.31$ for C_{mic}), the accuracy increased to $R^2 = 0.65$ when high content carbonate samples (total inorganic carbon > 1 %) were excluded. A significant nonlinear relationship was found between the ratio of aliphatic C-H to aromatic C=C peaks (1620:2930) and C_{mic}, C_{HWE} and C/N ratio (R^2 = 0.4, 0.56 and 0.5, respectively; $P < 0.05$), but, not with the ratio of more stable C to labile fraction ($TOC-C_{HWE})/C_{HWE}$. This might be attributed to the interference of silica and other organic compounds occurred at peak 1620 cm^{-1}. The most accurate spectral index for carbonate was the peak area at 713 cm^{-1} when related to TIC obtained by Scheibler's method ($R^2 = 0.98$). Hence, it is suitable to generate a spectral index to predict TIC in various soil types without need to multivariate calibration. The results showed the feasibility of the midDRIFTS-based spectral indices to be used as an indicator of SOM quality and composition as well as to quantify TIC contents.

However, routine use of spectral indices needs to avoid mineral interferences occurred in peaks assigned for organic compounds.

4.2. Introduction

Soil organic matter (SOM) as the largest terrestrial C reservoirs has been proposed to be a fundamental sink for sequestering CO_2 into the soil by enhanced soil organic carbon (SOC) and consequently improving soil quality for plant production (Lal, 2004b; Goidts and van Wesemael, 2007). Depletion of soil organic matter (SOM) has received a great attention as malicious threat to the arable land soil resources (Van-Camp, et al. 2004). In addition, in carbonate containing soils, carbonate (e.g. total inorganic carbon) has a crucial influence on both chemical and physical soil properties linked to the soil fertility and productivity. It has also direct effect on soil pH and regulates soil`s buffer capacity (Bruckmann and Wriessnig, 2013). Between several developed methods in quantifying soil carbonate, Scheibler's method is widely used despite its weakness (Kamogawa, et al. 2001). However, characterizing different SOM fractions and their quality as well as inorganic compounds across wide-scale is still challenging because of soil heterogeneity and needs for intensive field and laboratory measurements even at the field scale (Vasques et al., 2009; Stevens et al 2008), which is not always feasible due to the time and cost involved in conventional analytical methods (Mouazen, et al. 2007; Idowu, et al. 2008; Malhi, et al. 2001). Thus efficient, more rapid and low-cost alternatives are needed to characterize important soil properties across large areas (Vasques, et al. 2009).

Mid-infrared (MIR) radiation (between 4000-400 cm^{-1}) is absorbed by different functional groups of organic and inorganic compounds and is manifested as various stretching or bending vibrations. Thus, different types of bonds, and thus different functional groups, absorb infrared radiation at different wavelengths (Viscarra-Rossel, et al. 2006**;** Cobo, et al. 2010). Application of this technique is mainly based partial least square regression (PLSR) for model calibration and validation. Thus, PLSR models predict soil properties using the entire spectrum in relation to the soil reference data during optimization (Cobo, et al. 2010; Cécillon, et al. 2009; Soriano-Disla, et al. 2014). While the soil spectrum contains many SOM functional groups representing chemical constituents of soil quality (Calderón, et al. 2013) however, PLSR calibrations does not tell us anything about soil quality. Moreover, direct use of spectral data can be used not only to identify

SOM functional groups which cannot be characterized by traditional methods but also to measure the SOM quality/stability and composition (Viscarra-Rossel, et al. 2006; Janik, et al. 2007; Calderón, et al. 2011a). Hence, direct use of spectral data at specific spectral peak area could provide generic spectral indices to understand the SOM quality and its biochemical composition when compared to chemical data (Calderón, et al. 2011a; Demyan, et al. 2012). The approach is free of PLSR model construction and can be performed on fewer sample size than those used in PLSR calibration.

Previous work has been carried out in which peaks of mid-infrared peaks were used to identify SOM and its composition. Giacometti et al. (2013) found an enrichment of the relative contribution of aliphatic peaks at 2930 cm^{-1} in relationship to the peaks at 1620 cm^{-1} and 1520 cm^{-1} assigned for aromatic compounds in manure treatments as compared to those of control and mineral fertilizer treatments. Demyan et al. (2012) applied various specific midDRIFTS peaks to reveal alteration of SOM composition under a long-term application of farmyard manure, mineral fertilizer, and a combination of both. They found positive significant correlations between labile SOM and the peak area at 2930 cm^{-1} and negative correlations with that of 1620 cm^{-1}. In addition, the ratio of 1620 to 2930 could describe SOM stability and its decomposable rate. Bernier et al. (2013) applied the same ratio to follow the changing SOM quality with different irrigation water qualities in different soil types. They also used the peaks at 2968–2838 cm^{-1} (aliphatic C-H groups) and 1645–1626 cm^{-1} (aromatic C=C and C=O groups) with a peak at 800-795 cm^{-1} of quartz as an internal reference to normalize the aliphatic and aromatic peaks. However, most of them were carried out to identify spectral signatures related to organic and inorganic compounds and their biochemical constituents based on spectral interpretation. Hence, application of specific spectral peak areas to create quantitative spectral indices with a side-by-side comparison with chemical data is not well established at present and limited to a recently publication by Demyan et al. (2012) focusing on SOM, but, not inorganic C which may interfere with bands representing SOM functional groups. It should be noted that, the soil in nature is composed of mainly mineral compounds rather than organic. The midDRIFTS spectrum of bulk soil is dominated by minerals through the entire spectral range. None of peaks assigned to organic are free of interferences from mineral side. It has been proposed that mineral such as clay, silica and carbonate interfere to organic compounds at several peaks (Stevenson, 1994; Calderón, et al. 2011b; Reeves III, 2012). The presence of soil carbonate has been demonstrated using pure carbonate in several peaks in mid-range including bands 2686-2460 cm^{-1}, 1850-1784 cm^{-1}, 1567-1295 cm^{-1}, 889-867 cm^{-1}, 734-719 cm^{-1}, and 719–708 cm^{-1} (Tatzber, et al. 2010; Ji, et al. 2009) as well as at 2910 and 2850 cm^{-1} (Legodi, et al. 2001). However, none of the studies applied the spectral data from relatively diverse and disperse sample set at regional scale. Therefore, interference and overlap between organic and mineral compounds may obscure the ultimate use of the technique (Calderón, et al. 2011b) and may be more problematic at the regional scale with higher heterogeneity of soils.

This study investigated the midDRIFTS specific peak area with chemical data of soil biochemical properties from two contrasting geographic locations representing relatively diverse SOC, carbonate content and different soil texture. We hypothesized that certain peaks (e.g. 2930 and 1620 cm^{-1}) representing different functional groups may provide an appropriate index to measure SOM quality/stability and composition at the regional scale if mineral interference (i.e., carbonate) is avoided in soil spectra. Our objectives were to i) compare spectral data of the bulk soil from two different regions to evaluate the proportion of peaks representing SOM functional groups in each region, ii) to examine the applicability of spectral indices representing both labile and resistant SOM in combination with chemical data to explore a general soil SOM stability index at the regional scale, iii) to find out most relevant spectral index to soil carbonate content measured by Scheibler's method. If succeeding, relative midDRIFTS-based indices could be calculated which do not need a specific calibration on conventionally measured soil samples.

4.3. Material and methods

4.3.1. Study area and soil sampling

The study areas were the regions of Kraichgau (K) and Swabian Alb (SA), in the Federal State of Baden-Württemberg, Southwest Germany. Each study region was approximately 1600 m^2. The Kraichgau area is a fertile and intensively cropped hilly region which is situated in a basin between 100 to 400 m a.s.l. characterized by a mild climate with a mean annual temperature between 9-10°C, thus, one of the warmest regions in Germany. The soil types in this region are mainly Luvisols and Regosols. The mean annual rainfall ranges from 720 to 830 mm. Thus, fertile loess-derived soils and mild climate predispose the region to intensive agricultural activities (Ingwersen, et al. 2011). The Swabian Alb is a low mountain plateau of South-western Germany with elevations between 500-850 m a.s.l. The region is generally characterized by a cooler climate with an annual average temperature of 6-7°C, mean annual precipitation between 800 and 1000 mm with a high amount of summer rainfalls. On the Swabian Alb, soils developed mostly on Jurassic limestone, and are very rich in clay (~ 30 %) due to strong decalcification and weathering, and due the cooler climate also a higher in TOC content (2-3.3 %) compared to the Kraichgau soils (ca. 1 %). The crop rotation is mainly dominated by spring barley, winter barley; winter wheat and winter rape (Fischer, et al. 2010).

Soil samples were collected from farmers' fields at 42 locations in the two study regions (21 in each region). These fields were simultaneously used for a topsoil moisture monitoring system using time domain transmission sensors as a part of two succeeding integrated research projects (Ingwersen, et al. 2011). A nested sampling design was implemented at both regions, representing the full range of arable soils in the two regions. Three sampling points were selected in each field with the same crop resulting in a total set of 126 soil samples (63 samples for each region). Roughly one kg of soil (a composite of 5 sub-samples) was sampled to a depth of 30 cm at each sampling point using a steel soil core sampler with 2.5 cm inside diameter. Soil samples were kept cool for transport to the lab. Samples were moist sieved to <4 mm and incubated at 50% WHCmax (maximum water holding capacity) at 20°C. A 20 g subsample was incubated in a Respicond VITM respirometer (Nordgren, 1988) under the same conditions to monitor CO_2 evolution. As soon as CO_2 evolution decreased to a stable basal respiration (latest reached after 72 hours), the bulk soil was subdivided into two portions. A moist subsample was used for immediate biochemical analyses. The remaining sample was dried at 32°C, crushed and sieved (<2 mm) for further analyses.

4.3.2. Soil analysis

Hot-water extractable organic carbon (C_{HWE}) and nitrogen (N_{HWE}) were extracted from dry soil samples according to (Schulz and Körschens, 1998). Hot water extracts were analysed for organic carbon (C_{HWE}) and total nitrogen (N_{HWE}) with a Multi N/C analyser (Analytik Jena, Jena, Germany). Microbial biomass carbon (C_{mic}) and nitrogen (N_{mic}) were determined by chloroform-fumigation-extraction (CFE) (Joergensen and Mueller, 1996) on 20 g samples (n = 2) of <4 mm, moist soil. The extract C and N concentrations were measured using a Multi N/C analyser (Analytik Jena, Jena, Germany). Bulk soil was analysed for total carbon (TC) and nitrogen (TN) by dry combustion according to (ISO 10694, 1995) with a Vario-EL III elemental analyser (Elementar, Hanau, Germany). Scheibler's volumetric method was used to measure carbonate content (ISO 10A693, 1995). The carbonate C was calculated as referred to total inorganic carbon (TIC). Total organic carbon (TOC) was calculated as the difference between soil TC and TIC measured from Scheibler's volumetric method. Soil texture was determined by the Bouyucos method (m:v) (Anderson and Ingram, 1993).

4.3.3. MidDRIFTS analysis of soil samples

Dried bulk soil samples were ball milled and maintained at $32^{\circ C}$ overnight prior to midDRIFTS analysis according to (Cobo, et al. 2010) and (Demyan, et al. 2012). Briefly, a ball-milled subsample was scanned on a Tensor-27 mid-infrared spectrometer (Bruker Optik GmbH, Ettlingen, Germany) equipped with a potassium bromide (KBr) beam splitter and a liquid nitrogen cooled mid-band mercury-cadmium-telluride detector. A Praying Mantis diffuse reflectance chamber (Harrick Scientific Products Inc., New York, US) was installed in the spectrometer and purged with dry air from a compressor (Jun-Air International, Nørresundby, Denmark) with a flow rate of 200 liters hr^{-1}. The spectra were obtained at least in triplicate in the mid-infrared range (4000-600 cm^{-1}) by combining 16 individual scans at a resolution of 4 cm^{-1} recorded in absorbance units [log(1 /Reflectance)]. The spectra were first reduced to 3900 to 700 cm^{-1} to eliminate noise at both edges of each spectrum. Additionally, the CO_2 region (2400–2300 cm^{-1}) was excluded as it does not contribute to the actual soil spectrum. The raw spectral data were pre-processed with an atmospheric correction for carbon dioxide (CO_2) and water, baseline correction, and vector normalization. The corrected spectra of the three replicates of each sample were averaged for further spectral analysis. Spectral data analysis was carried out using the spectral processing software OPUS version 6.5 (Bruker Optik GmbH, Ettlingen, Germany). Each unique single peak of midDRIFTS spectra may act as the representative for more than one functional group of both mineral and organic compounds or represent overlapping bands of compounds (Nguyen, et al. 1991). Therefore, caution must be used for band interpretations.

4.3.4. Spectral bands used for peak area integration method

Table 4.1 shows the all peaks (absorption bands) corresponding to SOM functional groups and carbonates used in current study. To develop spectral indices the peak area integration approach was used according to Demyan et al. (2012). By this, the areas of all highlighted spectral bands were calculated and then were related to different SOM fractions across two regions. At each peak, first upper and lower integration limits were determined and then a corrected peak area was calculated on the basis of a respective localized baseline (Figure 4.1). Concerning SOM functional groups peaks i.e., peaks 2930 cm^{-1}, 1620 cm^{-1}, 1520 cm^{-1}, 1159 cm^{-1}, relative peak areas (referred as rA) were calculated to quantify relative changes in the spectra (Niemeyer, et al. 1992) by dividing the values of the relative intensity of each peak by the sum of the relative intensity of all four of the investigated peaks multiplied by 100. The ratio of the rA at 1620 cm^{-1} representing aromatic C to rA at 2930 cm^{-1} assigned to aliphatic C (1620:2930 ratio) was calculated as an indicator of relatively C stability or degree of organic C decomposition (Haberhauer et al 1998). Sum of the rA of all spectra for each band (i.e., peaks 2930 cm^{-1}, 1620 cm^{-1}, 1520 cm^{-1}, 1159 cm^{-1}) was calculated for each region separately and comparison was made between two regions. In addition, to devise appropriate quantitative spectral indices, linear regression analysis was further applied between rA of peak at 2930 cm^{-1} and corresponding analytical concentration values of SOM fractions such as TOC, TN, C_{mic}, N_{mic}, C_{HWE}, N_{HWE}, C/N, (TOC-C_{mic})/C_{mic} and (TOC-C_{HWE})/C_{HWE}. Likewise, the 1620:2930 ratio was correlated to various fraction of soil C. The corrected peak areas assigned to soil carbonate i.e., peaks 2515 cm^{-1}, 875 cm^{-1}, 730 cm^{-1}, 713 cm^{-1} were also used to investigate soil TIC measured by Scheibler's method. While peaks 2515 cm^{-1}, 875 cm^{-1} and 713 cm^{-1} were used to identify and quantify soil carbonate, the peak 730 cm^{-1} was used only to show the presence of carbonate type (calcite and/or dolomite) by spectral interpretation. Regressions analysis was performed using Sigma Plot version 10.0 (Systat Software Inc., Chicago, IL, USA).

4.3.5. Acid fumigation of carbonate containing soils

Acid fumigation was used to destroy the carbonate of soil containing TIC content > 0.5 % (total of 27 samples). Additionally 6 carbonate free samples were selected to detect possible acid influence on SOC. According to Harris et al. (2001) a 200 mg ball-milled sub-sample was weighed into a ceramic crucible. A sufficient amount of distilled water (~110 µl) was added to moisten the samples to approximately field capcity before placing into the desiccator. At the same time, a 100 ml of 37% HCl was placed inside the desiccator and evacuated with a vacuum pump until the HCl started boiling. After 24 hours under fumigation the samples were then removed from the desiccator and dried at 40°C. The dried sample was then crashed and ball-milled. All samples were scanned by midDRIFTS and further used for peak area integration as mentioned above.

Table 4.1. MidDRIFTS specific peaks representing soil organic matter (SOM) functional groups and carbonate investigated in the current study.

Spectral peak	Integration limits[a] / cm^{-1}	Spectral assignment	Stability	Reference
2930	3010-2800	Aliphatic C-H stretching	Labile	Stevenson, 1982
1620	1650-1553	Aromatic C=C and C=O stretching	Stable	Baes and Bloom, 1989
1520	1540-1500	Aromatic C=C stretching	Stable	Stevenson, 1982
1159	1180-1140	C-OH stretching of aliphatic , alcoholic O-H	labile	Senesi, et al. 2003
2515	2650-2460	Carbonate	-	Tatzber, et al. 2007
875	890-860	Carbonate	-	
730	734-719	Carbonate (dolomite)	-	
713	719-708	Carbonate (calcite)	-	

[a] adjusted after original reference based on visual spectral inspection

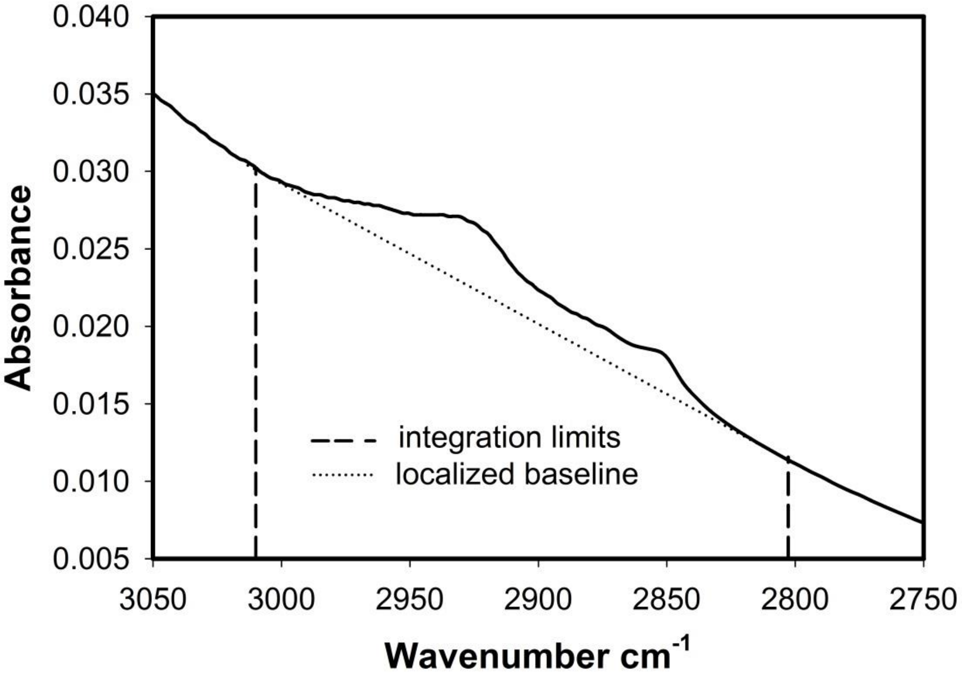

Figure 4.1. The peak area integration using localized baseline calculated to gain the corrected peak area.

4.4. Results

4.4.1. Soil organic matter contents and textural classification of investigated soils

Larger C and N contents were found in both bulk soil and SOM fractions in SA as compared to K ($P < 0.05$) (Table 4.2). Average concentration of TOC in soils of K (1.1 %) was significantly less than that of SA (3 %) ($P < 0.05$). Microbial biomass carbon and nitrogen contents (C_{mic} and N_{mic}) both were significantly higher ($P < 0.05$) in SA than K e.g. 262 versus 675 mg kg^{-1} for C_{mic}. The concentrations of hot-water extractable carbon (C_{HWE}) was also significantly different by location, where the average contents of 440 and 1004 mg kg^{-1} was obtained for K and SA, respectively. TIC was most varied within two regions with a CV value of 184 and 186 % for K and SA regions. Soil texture in K was dominated by silt, with silt loam and silty clay loam textures, while soils in SA had clay, silty clay and in a few cases silty clay loam textures (WRB, 2007). It explains also somewhat larger contents for organic matter content obtained for soils in SA.

4.4.2. Spectral features of bulk soil representing SOM functional groups

A representative baseline corrected and vector normalized midDRIFTS spectra selected from each study region is presented in Figure 4.2 a. In general, the soil spectra used in this study (n = 126) indicated a relatively diverse and wide range in SOM fractions, silt and clay and carbonate. Although, there is relatively qualitative similarities between two spectra, but, the differences can be indicated by some intensities in several distinct peaks. Therefore, differences between the two regions with different soil types and mineralogy were reflected by the midDRIFTS spectra in several peaks corresponding to the absorption of various organic and inorganic substances. The upper spectrum represents a silt-rich soil (70%) from K region with low TOC (1%) while the lower spectrum is a clay-rich soil (65%) of SA region containing 3.2% TOC and 2.2% inorganic carbon. However, once compare two spectra; differences in SOM, carbonate contents and soil clay content are easily visible in several peaks around 3700-3600 cm^{-1}, 3110-2800 cm^{-1}, 2600-2500 cm^{-1} and several peaks below 2000 cm^{-1}. Note that, in some regions differentiation between different compounds is nearly impossible as the soil mainly dominated by mineral compounds and have several peaks thought the spectrum. For instance, carbonate overlaps and have several distinct peak entire spectrum. The peaks between 3700-3400 and below 1500 are also strongly affected by O-H stretching of clay minerals (Reeves III, 2012).

Table 4.2. Summary of measured soil properties of Kraichgau (K) and Swabian Alb (SA) regions (presented in chapter 2).

		Descriptive statistics					
Properties	Study region	Minimum	Maximum	Average	CV %	Std.dev	Range
TOC / %	K	0.7	1.5	1.1	16	0.2	0.8
	SA	1.5	4.9	3	27	0.8	1.4
TN / %	K	0.1	0.2	0.1	14	0.02	0.1
	SA	0.2	0.5	0.3	26	0.08	0.3
C_{mic} / mg kg^{-1}	K	98	470	262	33	86.9	372
	SA	276	1157	675	36	246	881
N_{mic} / mg kg^{-1}	K	15	82	51.3	33	16.7	67
	SA	26	196	98	43	42	170
C_{HWE} / mg kg^{-1}	K	207	911	440	29	128	704
	SA	449	1838	1004	36	359	1389
N_{HWE} / mg kg^{-1}	K	14	78	43	31	13	64
	SA	34	213	99	44	43	179
C:N ratio	K	6.7	14.8	9.4	14	1.3	8.1
	SA	8.9	11.7	9.8	6	0.54	2.8
TIC / %	K	0	2.3	0.3	184	0.5	2.3
	SA	0	4.1	0.6	186	1	4.1
Sand / %	K	1.7	13.3	4.2	66	2.8	1.2
	SA	1.5	11.8	4.4	62	2.8	1
Silt / %	K	59.4	83.2	74	8	5.9	24
	SA	26.5	77.6	47.2	25	12	51.1
Clay / %	K	13.1	30	22.4	22	4.9	17
	SA	25	77	53	24	13	52

TOC = total organic carbon; TN = total nitrogen; C_{mic} = microbial biomass carbon; N_{mic} = microbial biomass nitrogen; C_{HWE} = hot water extractable carbon; N_{HWE} = hot water extractable nitrogen; C/N = TOC to TN ratio; TIC = total inorganic carbon.

Four main peaks of different organic functional groups representing SOM compounds are shown by a detailed illustration in Figure 4.2 b and c. Spectral differences of the peak at 2930 cm^{-1} representing mainly aliphatic C-H stretching bonds (Baes and Bloom, 1989; Janik, et al. 2007) between K and SA is consistent with the SOM content obtained by measured values. Soils from SA with larger SOM contents have higher absorbance values at 2930 cm^{-1} than K region containing less SOM. The peak at 1620 cm^{-1} mainly corresponds to resistant organic compounds with C=C and C=O bonds (Baes and Bloom, 1989; Stevenson, 1982) and form a relatively broad

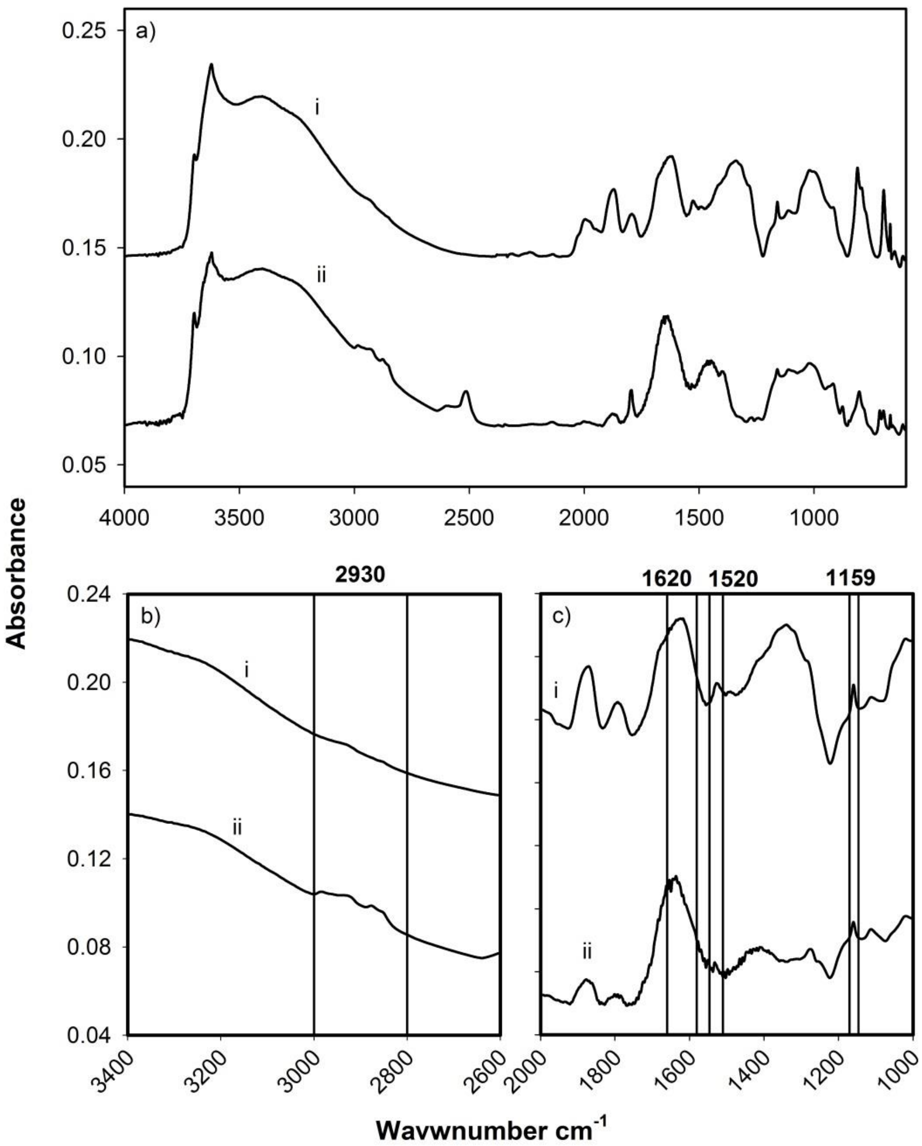

Figure 4.2. a) Baseline corrected and vector normalized midDRIFTS spectrum of bulk soil from Kraichgau (i) (20% clay and 0.9 % TOC) and Swabian Alb (ii) (65% clay and 2.7 % TOC) regions. b&c) Magnified illustration of spectral peaks SOM functional groups. 2930= C-H stretching of aliphatic compounds (CH_3 and CH_2 groups); 1620= aromatic C=C and C=O stretching; 1520= aromatic C=C stretching; 1159= C-O stretching of polysaccharide or polysaccharide-like substances.

peak which could be seen in both regions with relatively same intensity. However, the higher TOC contents of SA soil did not show the higher absorbance at peaks 1620 cm^{-1} as compared to soil from K with lesser TOC content. This peak overlaps with minerals such as silica and some other organic compounds which are not necessarily aromatic nature. The peaks at 1520 cm^{-1} and 1159 cm^{-1} representing resistant and labile organic compounds respectively were easily distinguishable. While the peaks 1620, 1520 and 1159 cm^{-1} are influenced strongly by minerals, the peak at 2930 cm^{-1} seems to be free of interference and most relevant to SOM and composition. However, carbonate may interfere with peak 2930 cm^{-1} especially in high carbonate containing soils.

4.4.3. Spectral features of bulk soils assigned for soil carbonate

Spectral differences between soils containing different carbonate content and composition can easily be distinguished in figure 4.3 a and b. According to Bruckman and Wriessnig (2013) and Tatzber et al. (2007) working on pure carbonate, four main peaks (i.e. 2515, 890, 730 and 713 cm^{-1}) representing carbonate were identified in current study. These bands are completely free of any interference and provide valuable information for carbonate quantification. Besides, the type of carbonate also can be determined by easily visual inspection. Both calcite and dolomite absorb at peak 2515 and 875 cm^{-1}. The peak 730 cm^{-1} is only assigned for dolomite, while the peak at 713 cm^{-1} represents only calcite. The uppermost spectrum was from a carbonate free soil with clearly no spectral features at all peaks. The middle spectrum represented a soil with a combination of both calcite and dolomite. Carbonate absorbs at all peaks. Carbonate does not absorb at peak 730 in lowermost spectrum, although it is from a soil containing 2.2 % TIC. Hence, the presence of spectral feature at peak 713 indicates the soil contains only calcite if the peak 730 does not exist anymore. Besides, it has been seen that the carbonate can also absorb in some other region in interference with peaks representing organic compounds.

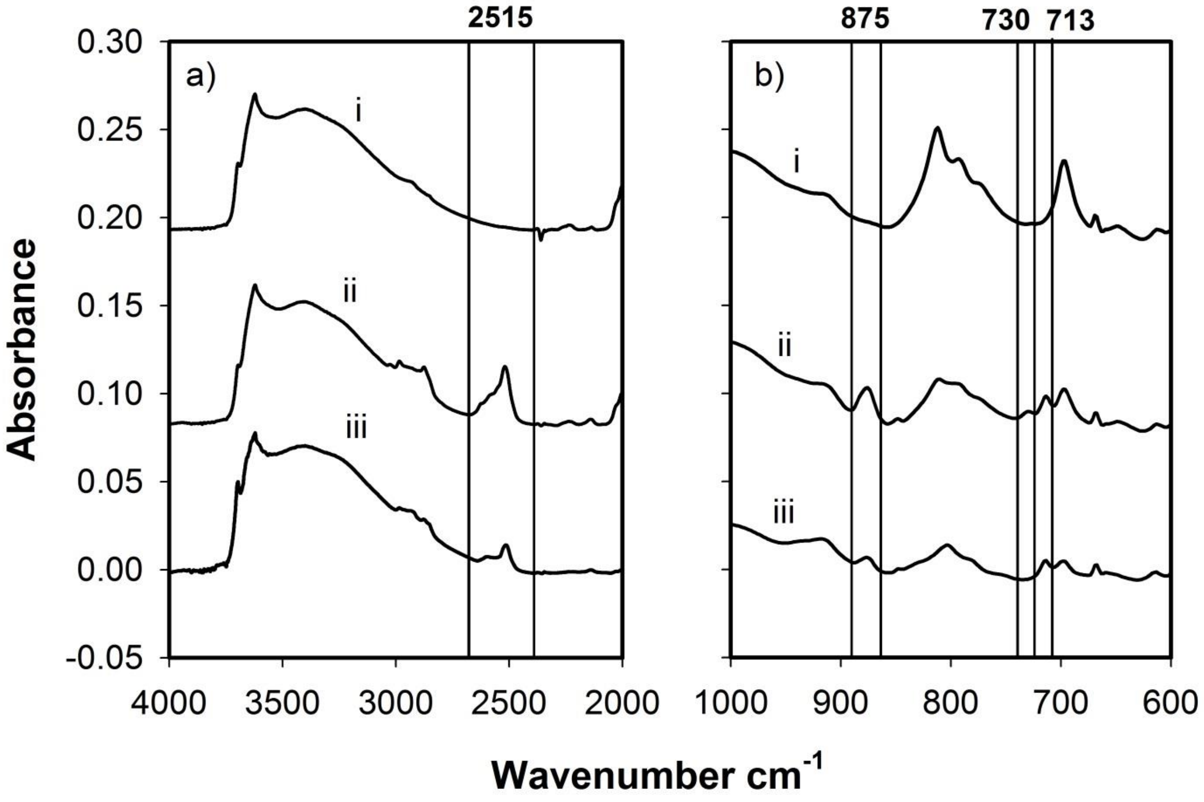

Figure 4.3. Baseline corrected of midDRIFTS spectra of a; i) carbonate free, ii) carbonate containing soil (total inorganic carbon (TIC) = 2.3 %) with a combination of calcite and dolomite, iii) calcite containing sample (TIC = 2.2). 2515 and 875 = peaks representing both calcite and dolomite; 730 = peak assigned to dolomite; 713 = peak assigned to calcite. Spectrum i is from Kraichgau and ii and iii are from Swabian Alb region.

4.4.4. Relative peak areas (rA) of SOM functional groups: comparison between two regions

Figure 4.4 shows the sum of rA for each SOM functional group peaks obtained from all soil spectra for each region separately. Significant differences were found between four peaks area as well as between regions ($P < 0.05$). The relative contribution of aromatic C=C and C=O stretching i.e. peak at 1620 cm^{-1}, was the largest proportion of the total area of the four functional group peaks in both regions (56 and 48 % for K and SA respectively). This was followed by peaks at 2930 cm^{-1}, 1159 and 1520 cm^{-1}. The comparison between the two regions indicated a significant larger relative contribution of peak at 2930 cm^{-1} in SA (38%) as compared to K (21%) ($P < 0.05$). Conversely, the relative contribution of peaks at 1620, 1520 and 1159 cm^{-1} from K soils was higher as compared to those obtained for SA soils. By increasing SOM in SA the rA of peak at 2930 cm^{-1} increased compared with K region, while rA of remaining peaks decreased in SA.

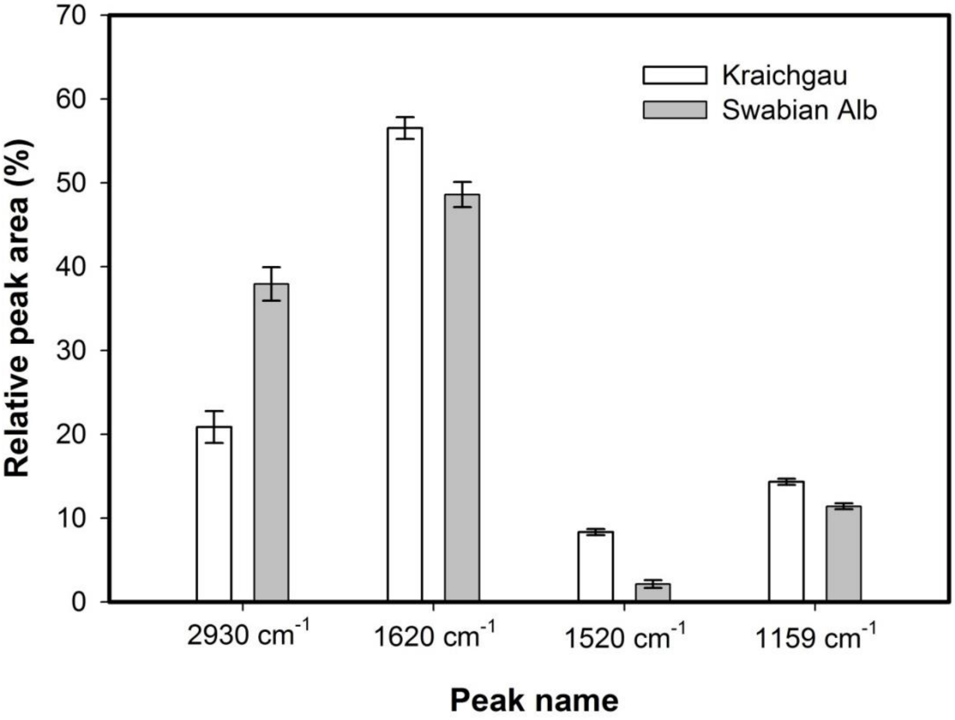

Figure 4.4. Relative midDRIFTS peak areas of soil organic matter (SOM) functional groups of bulk soils from two study regions.

4.4.5. Spectral index to characterise SOM fractions: how carbonate interferes with peak at 2930 cm^{-1}

The corrected peak area at 2930 cm^{-1} was applied to the measured values of SOM fractions for regression analysis. There was a positive correlation derived between C_{mic} and rA at 2930 cm^{-1} with a coefficient of determination (R^2) value of 0.31 ($P < 0.0001$) (Figure 4.5a). It should be noted that, this result included samples containing various TIC contents from 0 to 4.1 %. Likewise, relatively no coherent correlations were derived for other properties such as C_{HWE} and N_{mic} and N_{HWEC}, TOC, TN and C/N ratio (R^2 = 0.37-0.54; $P < 0.0001$; regression plot was not shown). Thus, the use of peak 2930 cm^{-1} as indicator of soil quality remains challenging with a relatively weak to moderate R^2 values.

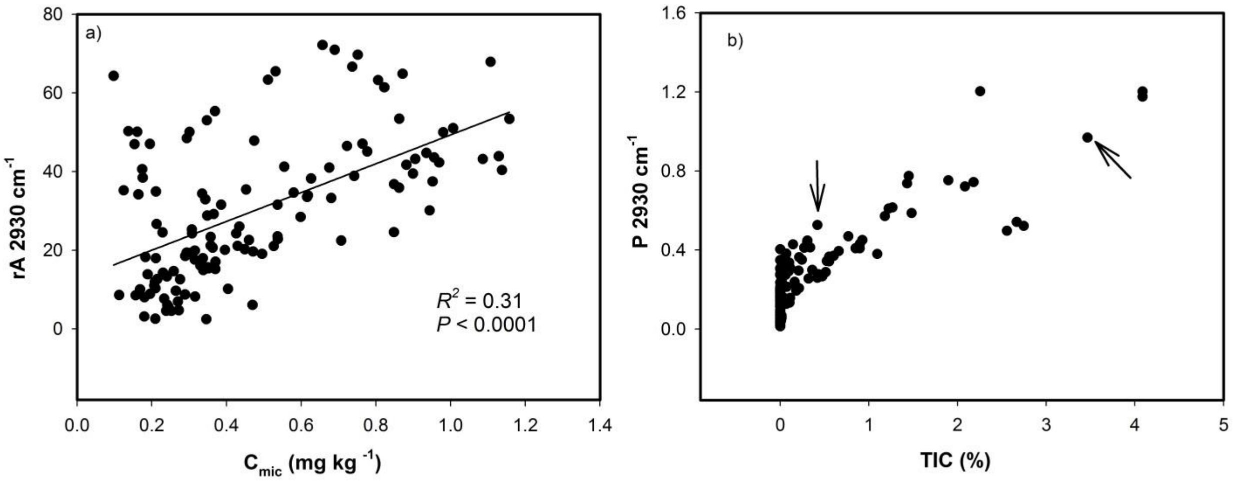

Figure 4.5. a) Relationship between relative peak area (rA) at 2930 cm^{-1} and measured microbial biomass carbon (C_{mic}) content of bulk soils; b) Relationship between corrected peak area at 2930 cm^{-1} and total inorganic carbon (TIC) contents across the two regions Kraichgau and Swabian Alb. Arrow highlighted two samples containing the same microbial biomass carbon (c.a. 750 mg kg^{-1})

Fig. 4.6 shows how acid fumigation affects spectral peaks at different frequency ranges including soil carbonate and organic compounds. It indicates the absence of major carbonate peaks especially at 2515 cm^{-1} in the spectrum after acid fumigation as compared to non-fumigated soil samples, indicating complete destruction of carbonates. But there was a noticeable decrease in peak area at 2930 cm^{-1} where the application of this peak led to any coherent correlation with none of SOM fractions. While the average peak area at 2930 cm^{-1} in non-fumigated samples was 0.62, the peak area decreased by 93% under acid fumigation to 0.04 in carbonate containing soils. Fig. 4.7 shows the carbonate free soil before and after acid fumigation. Similarly, the peak area at 2930 cm^{-1} decreased by 75% in acid fumigated as compared to that of non-fumigated samples. The spectral comparison indicates not only that carbonate but also some SOM components represented by this peak are also destroyed by acid fumigation. There were clear changes in the absobtion band of water at several peaks typically around 3400 cm^{-1} in both carbonate-free and carbonate containing samples after acid fumigation. In both soils, carbonate free and carbonate containing, the application of acid fumigation has caused a noticeable changes in midDRIFTS spectrum such as below 1500 cm^{-1}. Nearly all the peaks decrease in intensity at bands around 1450, 1000, 815 cm^{-1} except to a peak around 1200 cm^{-1}.

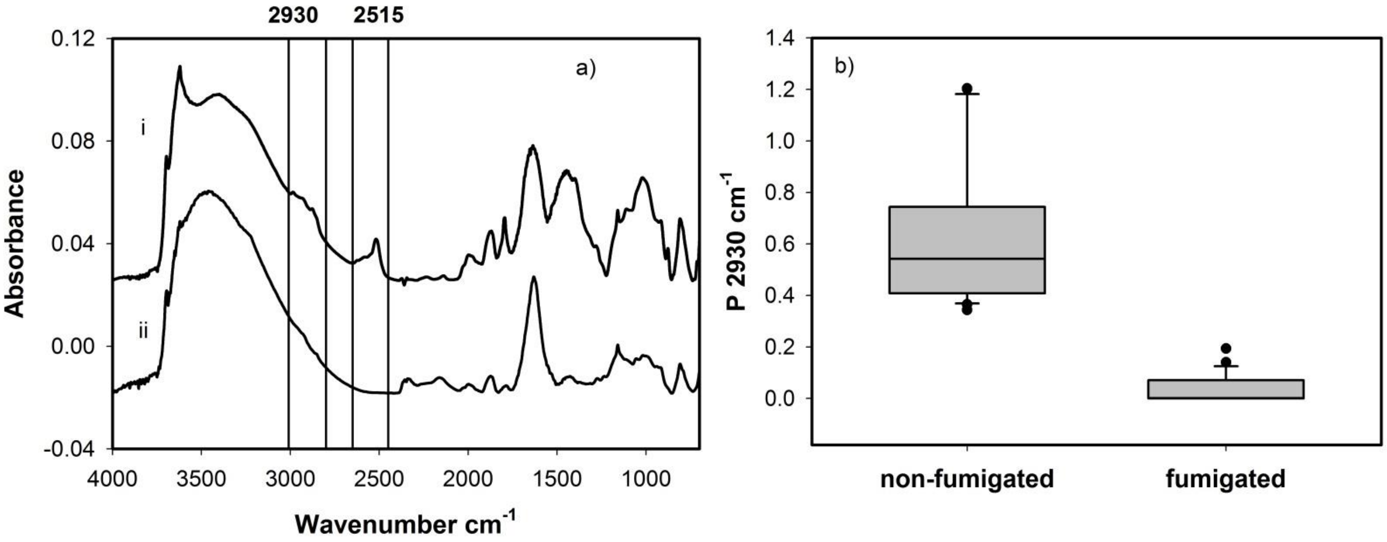

Figure 4.6. a) Averaged baseline corrected and vector normalized midDRIFTS spectrum of carbonate containing samples obtained from non-fumigation (i) and acid fumigation treatment(ii); b) Peak area at 2930 cm^{-1} of given carbonate containing soils affected by acid fumigation as compared to non-fumigated one. The middle band indicates the median (second quartile) and the bottom and top part of the box are first and third quartiles.

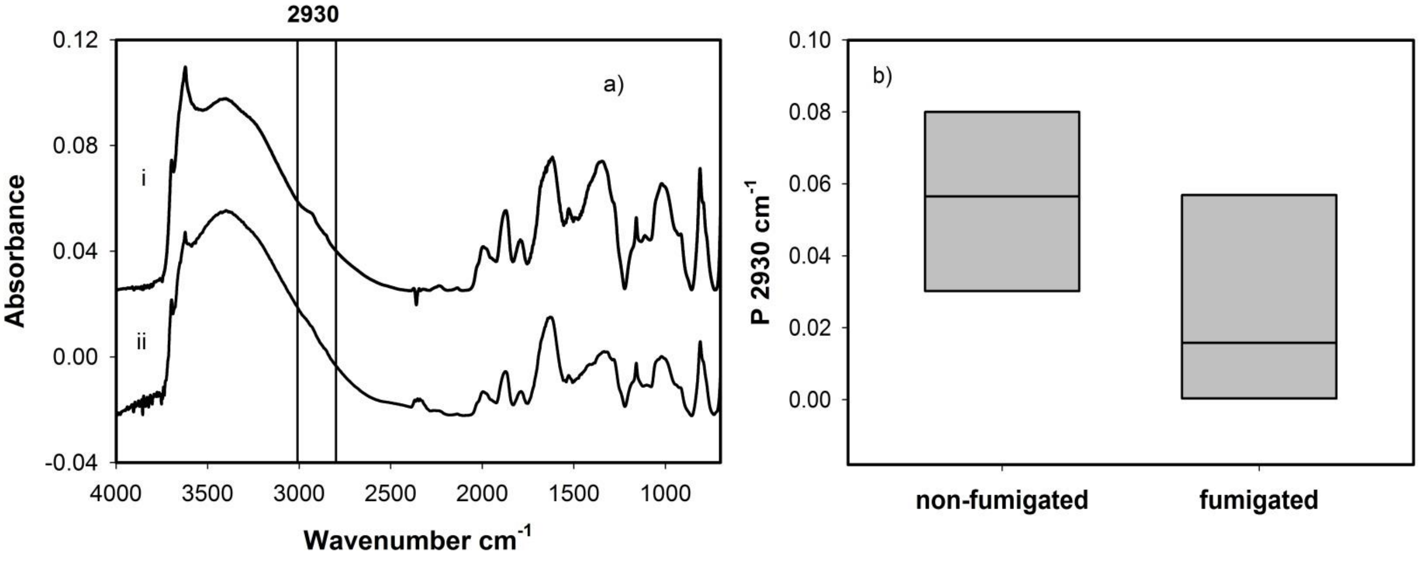

Figure 4.7. a) Averaged baseline corrected and vector normalized midDRIFTS spectrum of selected carbonate free samples obtained from non-fumigation (i) and acid fumigation treatment(ii); b) Peak area at 2930 cm^{-1} of given carbonate free soils affected by acid fumigation as compared to non-fumigated one. The middle band indicates the median (second quartile) and the bottom and top part of the box are first and third quartiles.

As mentioned previously, carbonate can overlap with the peak at 2930 cm^{-1}. Soil carbonate represents peaks at 2985 and 2880 cm^{-1} which were not present in carbonate free soil causing a big difference in peak area. The presence of carbonate was shown with a significant positive correlation between peak area at 2930 cm^{-1} and TIC contents across two regions (Figure 4.5b). It shows an increasing peak area at 2930 cm^{-1} together with TIC contents of bulk soil confirming the overlapping of carbonate with those of aliphatic compounds. For instance, two samples shown by arrows contain the same C_{mic} contents (c.a. 750 mg kg^{-1}), however, the samples with high TIC has the higher peak area at 2930 cm^{-1} roughly three times larger. We considered 1 % TIC of bulk soil as threshold which may affect the peak area at 2930 cm^{-1}. Therefore, attempt was further undertaken to develop spectral index by excluding samples containing > 1 % TIC. Noticeably, regression analysis between rA at 2930 cm^{-1} and C_{mic} contents resulted in a high correlation of determination (R^2) value of 0.65 (Figure 4.8). This considerable improvement was also seen for TOC (R^2=0.74), TN (R^2=0.71) N_{mic} (R^2=0.69), C_{HWE} (R^2=0.71), N_{HWE} (R^2=0.75), and C/N ratio (R^2 = 0.47). Moreover, a kind of textural tendency or site orientation was seen typically for TOC and TN when correlated to relative peak area at 2930 cm^{-1}. The rA at 2930 was further correlated to (TOC-C_{mic})/C_{mic} and (TOC-C_{HWE})/C_{HWE} as a ratios of stable C to C in labile fraction, but both did not lead to any relation resulted in very low R^2 value less than 0.1. Besides, the ratio of rA at peak at 1620 cm^{-1} (aromatic C) to rA at peak 2930 cm^{-1} (aliphatic C) (1620:2930 ratio) was used as an index to investigate SOM quality and decomposability of organic matter. Thus, a regression analysis between this ratio and C_{mic}, C_{HWE}, C/N and (TOC-C_{HWE})/C_{HWE} were made (Figure 4.9). Both C_{mic} and C_{HWE} were related with an exponential decay function with moderate R^2 values (0.4 and 0.56 respectively). While C/N ratio was correlated with a R^2 value of 0.5, (TOC-C_{HWE})/C_{HWE} showed hardly any relation with 1620:2930 ratio.

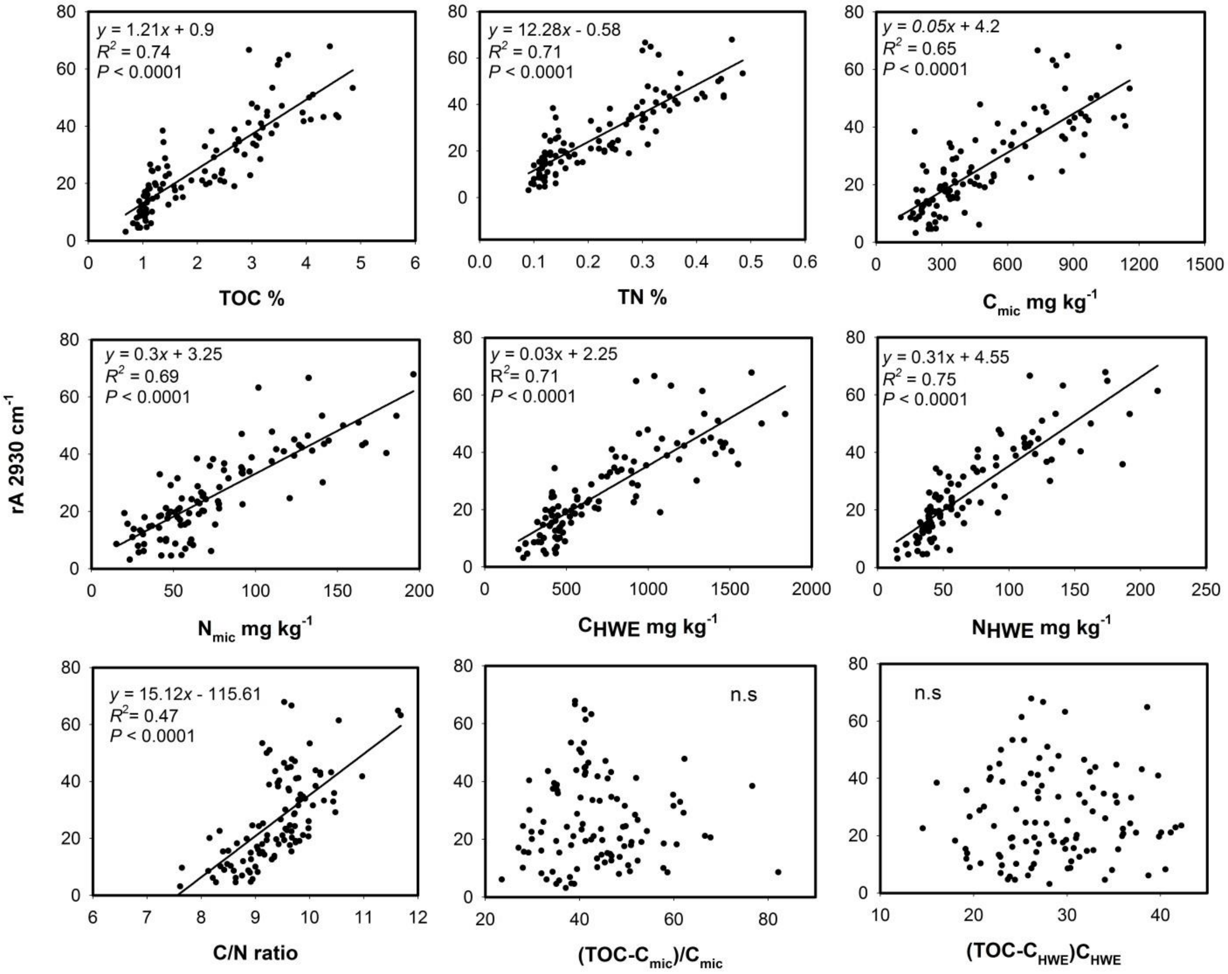

Figure 4.8. Relationship between relative peak area (rA) at peak 2930 cm^{-1} and soil properties of bulk soils across two regions.

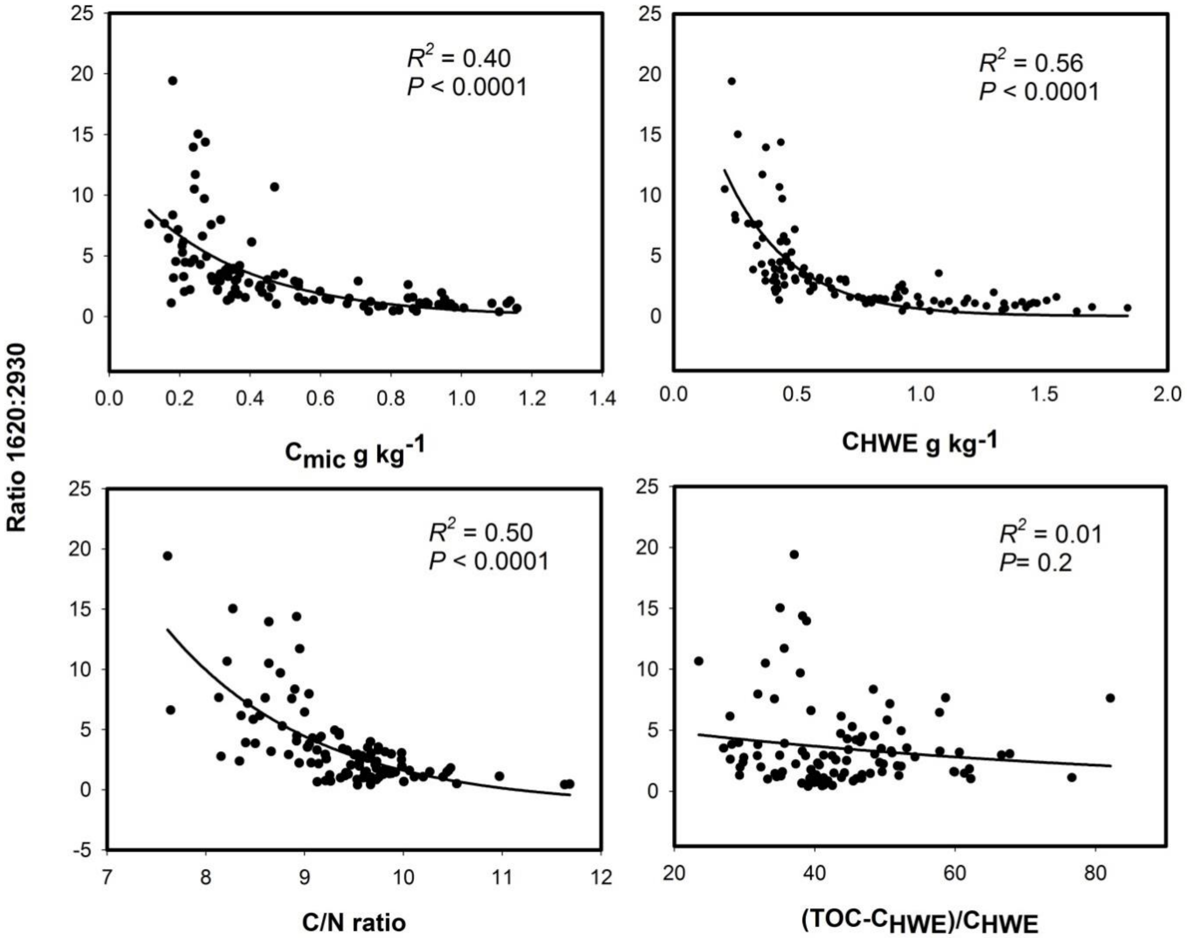

Figure 4.9. Relationship between of ratio of relative peak area at 1620 cn^{-1} to peak at 2930 cm^{-1} (1620:2930 ratio) and selected soil properties of bulk soils across two regions.

4.4.6. Spectral indices to quantify soil carbonate

Four midDRIFTS peaks assigned for soil carbonates were used to characterise soil carbonate contents and composition. There was a strong significant positive relationship between soil TIC and the peak areas at 2515 cm^{-1} and 871 cm^{-1} representing both calcite and dolomite (Figure 4.10 a and b). However, a divergent trend for each region was obtained. Soils in K present larger peak area at 2515 and 871 cm^{-1} as compare to SA at the same point of TIC content. It seems that dolomite contributes mainly in K soil in addition to calcite which is not contributed in quantifying TIC by Scheibler's method. A visual inspection of soil spectra from two regions confirmed the presence of dolomite in K region with a prominent peak area at 732 cm^{-1} which is only assigned for dolomite (Figure 4.11). Carbonate containing soils in K shows peaks at both peaks 730 ans 713 cm^{-1}, while in SA carbonate absorb only at peak 713 cm^{-1} representing only calcite. Thus, peak 713 cm^{-1} assigned for only calcite may provide a good indicator for carbonate in current study. The relationship between peak areas at 713 cm^{-1} with measured TIC contents of all soil samples is presented in Figure 4.8c. Noticeably, TIC content across both regions were correlated very well with peak area at 713 cm^{-1} with a R^2 value of 0.98 (P < 0.0001).

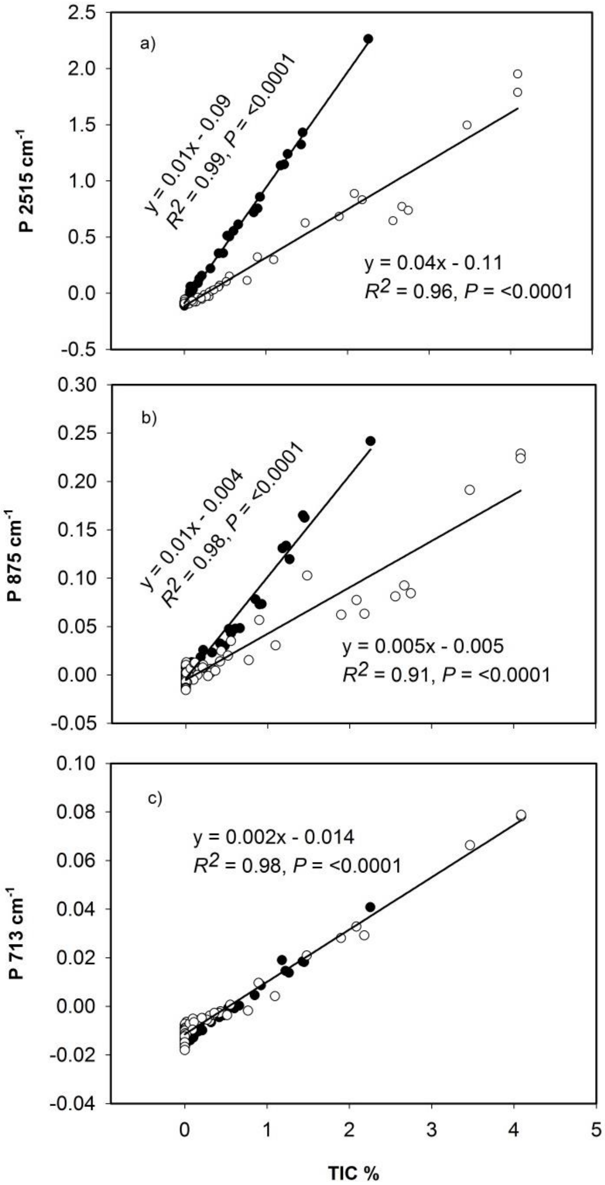

Figure 4.10. Relationship between the total inorganic carbon (TIC) as measured by the Scheibler's method and corrected peak area at: (b and c) 2515 and 871 cm^{-1} representing a combination of both calcite and dolomite; (c) Peak at 713 cm^{-1} representing only for calcite.

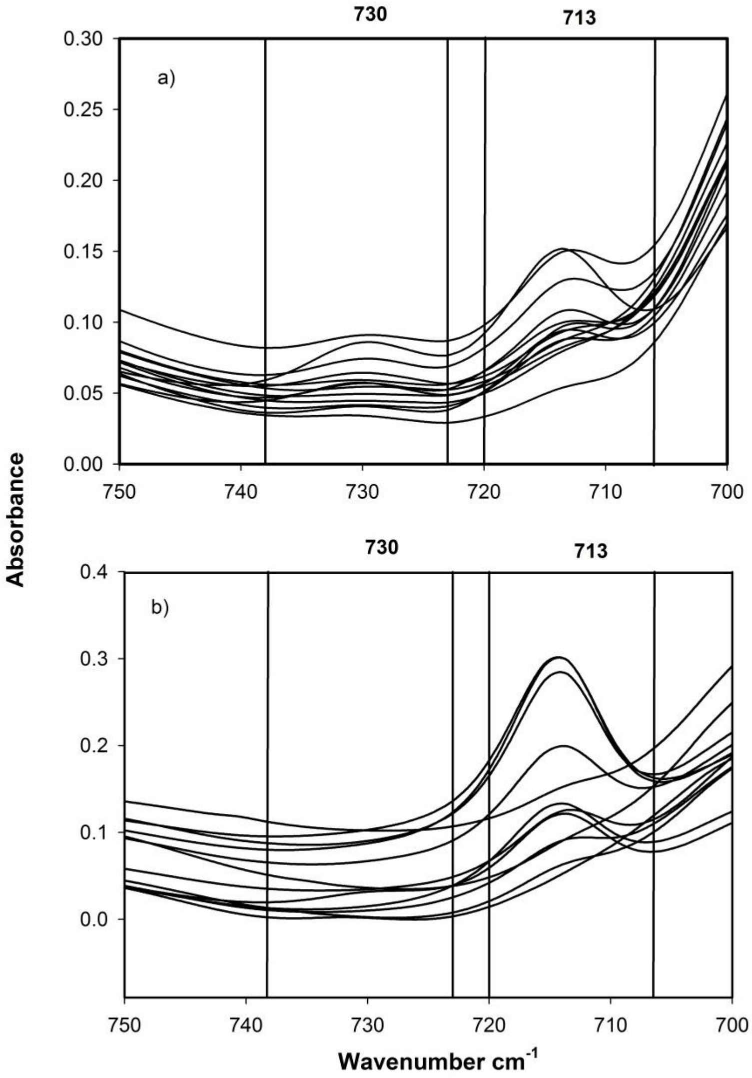

Figure 4.11. Baseline corrected of midDRIFTS spectra of carbonate containing soils for Kraichgau (a) and Swabian Alb (b) region. The peaks at 730 and 713 cm^{-1} are standing for dolomite and calcite respectively.

4.5. Discussion

4.5.1. Distribution of rA of SOM functional groups between two regions: Effect of climate and soil texture

Using the relative peak area of SOM functional groups revealed spectral differences between the two regions representing different SOM fractions and texture. Soils in SA region representing higher levels of TOC, C_{mic} and C_{HWE} lead to the higher peak area at 2930 cm^{-1}. It highlights the higher stabilization capacity of clay-rich soils in SA where adding organic substrate such as crop residue leads to larger TOC and labile SOM fractions like C_{mic} and consequently a higher peak at 2930 cm^{-1}. A similar result was found by Demyan et al. (2012) working on long term static fertilization experiment. The relative peak areas at 2930 cm^{-1} increased in farmyard manure treated soils with the highest TOC and C_{HWE} content as compared to the control containing fewer amounts of TOC and C_{HWE} (Demyan, et al. 2012). Conversely, relatively higher relative peak areas at 1620, 1520 obtained for K soils as compared to the SA soils indicated the accumulation of relatively more stable organic compounds. It might be attributed to the smaller amount of TOC, C_{mic} and C_{HWE} in K soils compared with the SA soils. Our results on relative peak area correspond with the results found by others which has already found. Enrichment at 1620 cm^{-1} peak has been found in a Halpic Chernozem soils under control and NPK treatment with relatively lower TOC (Demyan, et al. 2012). The same midDRIFTS results reported by Gerzabek, et al. (2006) in soils under follow without adding organic inputs. Similarly, in our study based on relative peak area obtained for two regions, the SOM in K contains more aromatic compounds (higher peak at 1620 and 1520 cm^{-1}) compared with the results for SA region (smaller peak at 1620 and 1520 cm^{-1}). It might be an indication of more advanced humification in K which degrades relatively easily decomposable aliphatic compounds into more stable organic constituents. Both climatic condition and textural composition might be a reason for that. Additionally, soil from two region offer different biochemical constituents increased the aromaticity nature of peaks at 1620 and 1520 cm^{-1} in K region. If taking the two peaks representing aromatic compounds, the ratio K to SA of the corrected peak area at 1620 cm^{-1} was 1.2 while this ratio at 1520 cm^{-1} was 4 indicating that different compositions of aromatic compounds were represented in the two regions. However, interference from non-aromatic organic compounds may have contributed to this difference, as band at 1620 cm^{-1} has also been assigned to amide I and II absorptions (Janik, et al. 2007; Sarkhot, et al. 2007). Although, the peak at 1159 cm^{-1} is related to organic compounds with higher decomposition rate such as polysaccharides and carbohydrates (Sarkhot, et al. 2007), but unexpectedly the SA soil with relatively larger labile fraction did not lead to higher rA at this band. One possible reason may be due to both spectral interferences and overlapping with Si–O stretching and bending of soil clay mineral such as Kaolinite in the 1120-1000 cm^{-1} range producing a shoulder in the spectrum especially in clay-rich soils (Madejová, 2003). Thus, by increasing the soil labile C fraction such as C_{mic} and C_{HWE} with together to peak at 2930 cm^{-1} may further prove labile nature of compounds related to this band. Generally, clay-rich soils contain greater SOM also tend to have greater peak areas at 2930 cm^{-1} as an indicator for labile OM fractions. The higher clay content in SA associated to the lower temperature compared with the soil in K, induce a slower decomposable rate and accumulate SOM mainly in early stage.

4.5.2. MidDRIFTS-based spectral indices as an indicator of SOM quality

Relatively weak correlation of the peak at 2930 cm^{-1} with SOM fractions have revealed the presence of overlapping signals from carbonate side at peak 2930 cm^{-1} mainly representing aliphatic compounds. Therefore, the peak area at 2930 cm^{-1} of carbonate containing soil seems to be higher as compared to those of carbonate free soils with the same TOC content. Therefore, deviation in linear regression is expected in samples with large variation in carbonate contents. The carbonate absorption at peak 2930 has been reported by several authors. According to Bruckman and Wriessnig (2013) working on spectra of pure carbonate reported aliphatic C-H stretching has some overlap with carbonate vibrations in the range from 3000-2800 cm^{-1}. Similar results also were shown by Tatzber et al. (2007) working on pure calcite and soil, where spectral signatures of carbonates were present in bands assigned to aliphatic compounds (3010-2800 cm^{-1}). Therefore, working on carbonate containing soils, it is difficult to develop spectral indices based on spectral peak at 2930 cm^{-1} to relate with labile fraction of SOM. On one hand, the application of acid fumigation to destroy soil carbonates seems to affect the SOM peak at 2930 cm^{-1} among others. Large decreases in the peak area at 2930 cm^{-1} in acid fumigated carbonate free soils as compared to non-fumigated implies that this approach to destroy soil carbonates also destroyes or alters the SOM, making direct application not useful for quantifying the peak at 2930 cm^{-1}. Additonally, reactions between soil carbonates and the acid results in salt formation, which also affect spectral peaks in the mid-IR range. Therefore, to avoid the carbonate interference effect in current study, the high carbonate containing (TIC > 1 %; *n* = 18) samples from both regions were excluded and analysis was continued with the remaining samples (*n* = 108). By this, the peak area at 2930 cm^{-1} of carbonate free samples was well correlated to the concentration of SOM fractions such as C_{mic}, N_{mic}, C_{HWE}, N_{HWE}, TC, TN and C/N. Marked improvement in correlation based on carbonate free soils (TIC < 1 %) suggests that the spectral peaks of un-treated bulk soil from a different area can be used to generate an appropriate spectral index to quantify SOM fractions. But the differences between carbonate containing and carbonate free soils must be identified and taken into consideration. However, pre-treatment of carbonate containing soils via acid fumigation (Smith et al. 1975) or ash subtraction (Reeves III, 2012) may provide a way to improve the presentation of SOM at peak at 2930 cm^{-1} by avoiding the carbonate derived interference signals. Apart from correlation accuracy, the overall trend between the relative peak areas of soil across both regions and concentration values of SOM fractions have revealed decomposability of compounds related to these peaks. C_{mic} and C_{HWE} as relatively labile SOC, has been used in several studies with a low turnover rate and serve as active pools (Prechtel, et al. 2009; Wang and Wang, 2011). Thus, positive correlations of the peak at 2930 cm^{-1} with those labile C fractions as easily decomposable compounds indicates the labile nature of compounds represented by this peak. These findings confirm the easily decomposable nature of the C-H stretching at 2930 cm^{-1} which are assigned to labile aliphatic compounds reported in several previous studies (Stevenson, 1982; Calderón, et al. 2006).

The 1620:2930 cm^{-1} ratio referring the ratio of aromatic to aliphatic peak areas varies with the intensity of decomposition or humification of SOM. Theoretically, high levels of C_{mic} and C_{HWE} as labile SOM indicate the presence of OC which is potentially readily degradable and *vice versa*. Until now, labile nature of peak at 2930 cm^{-1} and presence of relatively stable OC at peak 1620 cm^{-1} proved by correlation analysis described in previous section. The decreasing 1620:2930 cm^{-1} ratio with increasing C_{mic}, C_{HWE} and C/N ratio also confirmed the correlation results of these two peaks with chemical data. Therefore, the 1620:2930 cm^{-1} ratio indicates more easily degradable OC into the soil and will provide an appropriate spectral based index to identify SOM distribution and determine C dynamics into the soil. However, none of spectral indices i.e., the 1620:2930 ratio and peak at 2930 correlated with $(TOC-C_{mic})/C_{mic}$ and $(TOC-C_{HWE})/C_{HWE}$ as a ratio of stable C to labile C fraction. The result contradicted findings of Demyan et al. (2012). However, they used ratio of C in clay and > 1.8 g cm-3 fraction to C in < 1.8 g cm-3 fraction of soils from a fertilization experiment. Theoretically, the proportion of stable OC should increase with 1620:2930 cm^{-1} ratio by taking our correlation results into consideration. Nevertheless some other factor like soil texture and annual temperature may affect the OC stability and decomposition. The opposite results with that found by Demyan et al. (2012) may be due to the diversity of soils in current study differing in SOM content and texture. Furthermore, the interference signals from some other organic compounds and clay-bound water might be further reason. Peak at 1588 cm^{-1} due to amide II linked to amide I absorption at 1670 cm^{-1} (Janik, et al. 2007) can form a shoulder at both side of peak at 1620 cm^{-1}. Humic acids which contain nitrogen as well as clay-bound water especially in high clay content soil can also partially absorb in this region (Calderón, et al. 2011b). Therefore, application of this ratio to develop spectral based indices to quantify OC stability is remaining complicating and needs more investigation.

4.5.3. MidDRIFTS-based spectral indices as an alternative to quantify soil carbonate

Spectral data mining for carbonate resulted in strong correlation between two main peaks (2512 and 871 cm^{-1}) representing both calcite and dolomite and TIC but with a divergent trend obtained for each region. This divergent comes firstly from the type of mineral constituents contributing in soil carbonate which may vary between two regions and secondly, the method was used to quantify TIC. Soil carbonates in our study regions may consist of both calcite and dolomite. Generally, the occurrence of carbonate either as calcite and/or dolomite depends on the parent material and soil formation processes (Bruckman and Wriessnig, 2013). Additionally, in our study regions, the lime is applied by the farmers for pH adjustment which may contain a combination of both calcite and dolomite. The correlation results for peaks 2512 and 871 cm^{-1} with TIC indicates the presence of a dolomite only in K soil in addition to calcite which is not determined during laboratory measurement using Scheibler's volumetric method. The presence of dolomite in K soil was further confirmed by spectral interpretation at peak 732 cm^{-1} (the peak only assigned for dolomite in midDRIFTS spectrum) in current study. Therefore, TIC content for K region obtained by Scheibler's method is underestimated as the dolomite is not measured in this approach. Reaction time based on the type of carbonate is different in this method. Dolomite is less soluble in 10 % HCL resulting in longer dissolution times as compared to calcite.

Therefore, the values for total carbonate could be underestimated (Schlichting, et al. 1995). Therefore, while several mid-infrared peaks represent soil carbonate content, working with peaks which are representative of a combination of calcite and dolomite i.e., bands E and F remain challenging based on Scheibler-measured values. Nevertheless, the divergent trend was solved when peak at 713 cm^{-1} representing only calcite was taken into account. Consequently, a strong positive correlation obtained across both regions indicates the feasibility of bands 713 cm^{-1} to quantify only calcite concentrations in soil. Hence, the peak at 713 is the most relevant to TIC values measured by Scheibler's method in soils contain a combination of calcite and dolomite. Nevertheless, it may also provide appropriate index for only calcite containing soils (dolomite free soils). In such soils, two other peaks, i.e., 2512 and 871 also would also provide as the same as 713 which all are corroborated to the measured values obtained by Scheibler's method. However, developing a generic spectral index to quantify soil carbonate across both regions in current study or soils from contrasting geographic areas containing a mixture of calcite and dolomite, it is necessary to use the data from an analytical approach which quantifies both calcite and dolomite precisely. Taking this into account, the peak area at 2512 and 871 cm^{-1} as a most indicative peak assigned to the combination of both calcite and dolomite would provide the best spectral quantitative index to quantify soil carbonate content in soils.

4.6. Conclusions

This study used the midDRIFTS specific peak area integration method at the regional scale with soils of a relatively large range of TOC, TIC and texture contents to develop spectral indices to study the composition of SOM and to quantify soil carbonate contents. The differences between two contrasting study regions representing different soil texture classes and various carbon contents were reflected in midDRIFTS spectra typically in some predominant peaks at 2930, 1620, 1520 and 1159 cm^{-1} assigned to SOM as well as some other peaks representing soil minerals i.e., 2515 cm^{-1} for carbonate and 2000-1700 cm^{-1} for quartz. For individual peak of SOM functional groups, the regression model accuracy based on spectral index at band 2930 cm^{-1} (aliphatic C-H stretching) is promising for characterisation of SOM fractions such as C_{mic}, N_{mic}, C_{HWE}, N_{HWE}, TC, TN and C/N across two regions. Nevertheless, by increasing the soil carbonate into the samples (samples with TIC > 1 %) the peak at 2930 cm^{-1} would not lead to any appropriate index related to SOM fractions. Therefore, as far as useless of acid fumigation are concerned, the induced effect from carbonate side in that peak must be taken into consideration using either ash subtracting or developing a constant factor to neutralize carbonate inductions. The ratio of peak at 1620 and 2930 cm^{-1} (1620:2930, aromatic to aliphatic ratio), showed its potential to investigate SOM stability ratio of bulk soils linked to C:N but not when correlated to stable C/labile C ratio. This ratio was not suitable to explain the changes in stable C to labile C ($(TOC-C_{HWE})/C_{HWE}$) obtained in current study. However, it may cause by various stabilisation mechanisms and soil heterogeneity involved in individual regions bulk soil samples as well as the interference of mineral like silica and protein compounds especially at peak at 1620 cm^{-1} in high-N containing soils. Hence, midDRIFTS-based spectral indices could be a feasible indicator of SOM quality and composition if mineral interference especially at SOM fundamental bands avoided. The results showed that spectral indices to quantify soil carbonate are deeply depending to the data accuracy and method used for laboratory measurements as well as the carbonate compositional characteristic. The results showed that spectral indices can be used to identify carbonate composition and explore error measurements associated to the Scheibler's method. In addition, the spectral indices related to carbonate can be used to develop a spectral based model to quantify TIC into the soil. This is free of PLSR model calibration and validation and can be applied to either fewer sample size and/or extremely heterogeneous samples which are not suitable for PLSR calibration.

Hence, it is summarised that the application of midDRIFTS and corresponding specific peak area integration method has shown to be a potentially useful and quantitative approach in applying spectral indices at the regional scale to study SOM composition and structure as well as to predict soil carbonate which may be a generic way forward for other soils. Nevertheless, mineralogical interferences and overlapping vibrations at peaks corresponding to organic matter functional groups demands strong caution in interpreting spectral data. This approach is useful when the samples are limited and lack extreme heterogeneity which make unsuitable for midDRIFTS-PLSR calibration and validation and would be easily transferable from one sensor to another one. The spectral based derived relation in current study gives its feasibility to be used for up scaling from ex-situ to in-situ measurement up to imaging spectroscopy (mid-infrared hyper-spectral images) for measuring the spatial distribution of soil variables. However, the approach must be tested over extended spatial areas to capture larger variability in soil type and mineralogical composition.

Chapter 5

General discussion

5. General discussion

5.1. Spectral interpretation for pre-screening prior to quantitative analysis: How visual aspects of a spectrum can be related to the structure and chemistry of sample.

While quantification of soil properties via multivariate calibration is often the goal of spectroscopic applications, qualitative aspects of spectroscopy can also be used for sample discrimination prior to quantitative analysis (Coates, 2000; Linker, et al. 2006). The spectral interpretation in the current study demonstrated that visual aspects of soil spectra provided a further prospective to obtain general knowledge of unknown soils where no laboratory information is available. The concept is that soil spectra contain useful information which can be used prior to quantitative analysis (Shepherd and Walsh, 2007; Janik, et al. 2007). In the current study, midDRIFTS spectra of soils collected across two contrasting regions representing a range of heterogeneity provided extensive spectral information which gave general indications about soil quality and textural composition. Several peaks were found to be highly useful for spectral based soil quality assessment. Peaks at 2930, 1620, 1520 and 1159 cm^{-1} in the current study were assigned to SOM functional groups which corroborated findings of Bornemann and Amelung (2010) and Demyan et al. (2012). These studies showed the potential of midDRIFTS for rapid soil quality assessment with the emphasis on fundamental SOM functional group peaks. In regards to SOM, the intensity of the 2930 cm^{-1} aliphatic C-H peak was the most consistent peak, which differed between the two regions. Higher peaks in SA region were due to higher SOC, possibly as the result of lower mineralization due to cooler temperatures and higher clay content than K region. This peak was also present in PLSR models for SOC and Cmic prediction (Ludwig, et al. 2008, Rasche, et al. 2013). In addition, several distinct peaks representing soil carbonate and clay content were observed in soil spectra. The use of different peaks corresponding to molecular vibrations of organic and inorganic substances enabled us to differentiate between the two regions with different soil types and mineralogies. Since peaks and underlying functional groups are not compound specific, linkages to specific soil components may be hampered by interference and overlapping of signals from other inorganic and organic compounds (Reeves III, 2012).

A useful application of spectral screening is to discriminate soils from different origins or to classify samples into several groups for further quantitative analysis. In particular, working with a regional-based soil spectral library for model calibration, it is necessary to choose appropriate sample sub-sets to represent the variation within the whole sample set to analyze via traditional laboratory methods and PLSR model development. For this, peak area at specific peaks can be calculated to rank or to categorize spectra into several groups. Then, appropriate sample set can be selected randomly from each class for both model building and laboratory analysis. A lower accuracy PLSR model is more likely when the sample set includes extreme heterogeneity in terms of soil types, texture, and mineralogy. Therefore, separate PLSR models are necessary for each single classified spectral group. Spectral pre-screening can be used to categorize spectra into several groups which then can be modeled using each spectral group, which performs better than random selection. To achieve this, principal component analysis (PCA) has been proposed to recover variation in soil spectra and discriminate between samples from different origins by peaks associated with calcium carbonate and soil minerals (Linker, et al. 2006; Linker and Shaviv, 2008). Du et al. (2008b) pointed out that spectral information obtained by Fourier transform infrared-Photoacoustic Spectroscopy (FTIR-PAS) can be used to classify different types of clay minerals. As a result, this spectral screening would lead to noticeably improvement in model prediction as done for soil nutrients such as nitrate (Linker et al. 2006). It has been found that midDRIFTS-PLSR calibration of SOC was limited by the presence of inorganic carbon in soil spectra (Reeves, et al. 2005). Hence, differentiation between soils in terms of carbonate may improve the quality of midDRIFTS-PLSR calibrations. Spectral qualitative assessment in either visual interpretation or PCA-based discrimination provides information on soils under evaluation which can be used as a screening step prior to midDRIFTS-PLSR model construction.

5.2. Application of midDRIFTS-PLSR calibration to predict soil physical and biochemical properties at regional scale: independent validation versus cross-validation

The feasibility of studies to investigate soil heterogeneity at increasing scales up to regional level is restricted because of the considerable time and cost involved in conventional laboratory measurements. Despite widespread application of midDRIFTS-PLSR in prediction of soil properties from field up to the landscape level (McCarty, et al. 2002; Madari, et al. 2006; Grinand, et al. 2012), the models can be strongly affected by the number of samples, difference between calibration and validation datasets, target variables in terms of concentration and data range, sample preparation and soil texture (Yang and Mouazen, 2012; Bellon-Maurel, et al. 2010; Bellon-Maurel and McBratney, 2011). Hence, the efficacy of midDRIFTS in prediction of soil properties remains variable and complex because of high soil variability. The results presented in the current study (chapter 2) showed the potential of midDRIFTS-PLSR to quantify essential soil properties (e.g., soil C and N fractions, soil texture) across two contrasting regions representing high spatial heterogeneity. In the current study midDRIFTS-PLSR calibration resulted in accurate prediction for 10 out of the 15 investigated properties (R2 > 0.95 and RPD >3) were found. The results were comparable with those obtained by others (Cobo, et al. 2010; Madari, et al. 2006) and even more accurate predictionresults were achieved as compared to the few previous studies for Cmic and Nmic (Cohen et al., 2005, Ludwig et al 2008). In general, both cross-validation and independent validation approaches resulted in calibrated models of approximately the same accuracy. In the second step of midDRIFTS-PLSR application, model validation revealed inaccurate for cross-validation derived models but not for independent validation models, showing that independent validation outperformed cross-validation when applied to new unknown spectra. This was attributed to the tendency of cross-validated derived models to be over-fitted (Minasny, et al. 2009, Bellon-Maurel and McBratney, 2011) because of the higher number of factors used in calibration. This explains why an additional validation is needed to test predictive performance of cross-validated models which is lacking in this approach. The cross-validation approach was tested for each region separately as the sample size in each region (n= 63 per region) was not sufficient for independent validation. Even being region specific, lower accuracies were obtained for each region using cross-validation than with generic PLSR models. In addition to being over optimistic for predicting “unknown” samples, the lower accuracies in region specific cross-validation models were further attributed to the small sample size and insufficient heterogeneity in the dataset. Hence, the development of region specific calibrations in the current study required large enough number of samples with appropriate data coverage in terms of range and variability to allow the independent validation approach. In the current study, to develop a region specific PLSR model at least 100 samples are required to allow an independent validation modeling approach. Lastly, the generic independent validation approach outperformed region specific cross-validations due to the greater sample-set size and variability which avoids the over optimistic tendency of cross-validation calibrations when predicting “unknown” samples.

Even with the independent validation approach, calibrations were not successful (R^2<0.9 and RPD<3) for Nmin, C:N ratio, pH, bulk density and sand content. Unsuccessful calibrations for pH and bulk density which were also found by Kuang and Mouazen (2011), might be due to indirect spectral response in the Mid-IR range. Less accurate predictions obtained for C/N, were attributed to the propagation of uncertainty or errors from each single property in the ratio (Ludwig et al., 2008). Unsuccessful calibration for sand was attributed to the overlapping of the Si-O peaks with those of SOM peaks in the spectral frequencies between 1400-500 cm^{-1} (Janik, et al. 1998) and relatively low in sand content (2-12%) in data set influenced more intensively by clay content. However, to overcome the limitations associated to cross-validation and/or prediction of those properties with unsuccessful PLSR model, an optimal selection of spectral peaks is required e.g., competitive adaptive reweighted sampling to reduce spectral complexity and to improve the robustness of a calibrated model.

5.3. Practical use of midDRIFTS spectral-based index to investigate SOM quality

The results of spectral interpretation discussed previously can be used semi-quantitatively to devise spectral indexes to identify and quantify the presence of important organic functional groups in soil midDRIFTS spectra by observing fluctuations in specific spectral bands (Calderón et al., 2011b; Demyan, et al. 2012). The current study showed that peak areas at specific peaks can be used directly for assessing SOM quality in connection to measured values of SOM fractions. The approach was already applied to assess SOM quality changes under different irrigation regimes and soil types (Bernier et al., 2013) and under long term fertilizer application (Kaiser and Ellerbrock, 2005). Demyan et al. (2012) also applied the ratio of aromatic to aliphatic peaks to evaluate SOM quality and stability in fertilization experiment. However, it is less understood how spectral peak areas can be related to different SOM fractions at regional scale. The current results showed feasibility of the peak at 2930 cm^{-1} (aliphatic C-H stretching) and of the ratio 1620:2930 cm^{-1} as appropriate spectral-base indexes to characterize SOM quality. Nonetheless, interference signals from soil carbonates at peak 2930 remains challenging (Tatzber, et al. 2007). In the current study peak at 2930 cm^{-1} performed best as a soil quality index when high carbonate containing soils were excluded (i.e., TIC> 1%). Hence, mineral interference is an issue which also hampers spectral based calibration to predict SOM fractions. The 1620:2930 cm^{-1} ratio referring to the ratio of aromatic to aliphatic compounds varies with the intensity of decomposition or humification of SOM. The decreasing 1620:2930 cm^{-1} ratio with increasing Cmic, CHWE and C/N ratio also confirmed the correlation results of these two peaks with chemical data. Therefore, the 1620:2930 cm^{-1} ratio indicates more easily degradable OC into the soil and will provide an appropriate spectral based index to identify SOM distribution and determine C dynamics in the soil. Two specific interferences are carbonate interferences at peak 2930 cm^{-1} and presence of silica at peak 1620 cm^{-1} particularly in clay-rich soil. Nonetheless, the approach is applicable since it does not require a calibration and can be carried out on fewer samples like field experiment.

5.4. MidDRIFTS-based spectral index as an alternative to PLSR in quantifying soil carbonate

MidDRIFTS spectra have several carbonate peaks which are relatively free of mineral and organic interferences. As mentioned in the previous section, the peak area can be used directly to estimate soil total inorganic carbon (TIC) using the wavenumbers assigned to soil carbonate. A number of studies have used midDRIFTS spectroscopy in quantifying soil carbonate or IC based on PLSR models with variable success (Bertrand, et al. 2002; Reeves and Smith, 2009; Bellon-Maurel and McBratney, 2011). Strong positive correlation was found between spectral index at bands assigned to carbonate and TIC (R2=0.98) in the current study which was comparable with midDRIFTS-PLSR calibrated model developed for TIC (R2=0.99 see chapter 2). Although the presence of carbonate peaks in the Mid-IR has been identified in several peaks in soil spectra in comparison to pure carbonate (Tatzber, et al. 2007; Bruckman and Wriessnig, 2013), direct use of single peaks to predict IC was not previously done. In practical issue, spectral based prediction may outperform PLSR model in some situations. PLSR uses the additional wavenumbers which are not related to IC and adds noise to the model. The use of the peak area as spectral index as done in the current study are viewed free of interference which presents a unique representative signals for IC in soil for a wide range of carbonate. Hence, the approach provided a further quick assessment of carbonate or inorganic carbon as an alternative to midDRIFTS-PLSR calibrated model. However, it needs to be tested on other type of soils differeing in texture and mineralogical composition. In the current study using carbonate values measured by Scheibler's method (ISO 10693, 1995), only the peak at 713 cm^{-1} assigned to calcite resulted in the best index in soil containing both calcite and dolomite. Dolomite is not determined in Scheibler's method as it has a low solubility in 10% HCl used for this method (Schlichting, et al. 1995). In addition to accurate TIC quantification as compared to conventional laboratory analysis, the type of carbonate can be easily distinguished and specific index can be devised for each carbonate type (i.e., calcite and/or dolomite). But, for this concentration values for dolomite should be known. However, the peak ratio of dolomite/calcite might be a suitable way to correct results obtained by Scheibler's method which is mainly calcite. Nonetheless, working on soils containing only calcite all three peaks, 2515, 871 and 713 cm^{-1} can be used to develop a spectral based index in correlation to concentration values obtained by Scheibler's method. Generally the approach can be used with any soil type with various amounts of calcite and dolomite, the spectral index at 2515 and 871 cm^{-1} (assigned for both calcite and dolomite) should be tested in correlation with carbonate values obtained by a method which measures both calcite and dolomite.

5.5. MidDRIFTS-PLSR integrated with geostatistics to create high resolution soil property maps

Calibrated models for TC, TIC, TOC clay and silt were used to predict soil properties from regionally distributed soil midDRIFTS spectra across both investigated regions (n=1170). The results presented in chapter 5 demonstrated that midDRIFTS-PLSR can be successfully integrated to geostatistics to generate a large dataset in a time and cost efficient manner to assess soil spatial variability and to create high resolution soil maps. Nevertheless, it should be noted that not all samples were predicted accurately as the spectra of those samples with large dissimilarity were excluded as outliers (about 7%). It may be due to the function of different soil textures and spatial domains (McBratney, et al. 2006) which resulted in high variation of TIC, SOC carbonate and clay and silt contents that was out of the range of the calibration set. Experimental semivariograms fitted to the predicted data explained the spatial dependency precisely when appropriate models were used. Differences in spatial model parameters from fitted semivariograms (i.e. range, nugget effect and degree of spatial dependency between properties of each region) were mainly due to the intrinsic factors such as parent material, climate, mineralogy and soil forming processes factors and farming practices as well (Wang et al., 2010). The results demonstrated that soil C in the K region was highly affected by farming practices rather than intrinsic factors as indicated by greater nugget effect and lower spatial variability. Nonetheless, soil texture in both regions is controlled mainly by intrinsic factors.

Mulla and McBratney (2000) wrote that sampling density can affect spatial dependency and interpolation. The length of the spatial dependency (semivariogram range) in the current study was much longer than the sampling interval. Hence, the sampling strategy was successful in capturing the spatial variability (Fu, et al. 2010; Wang, et al. 2010). Thus, intensifying sampling would not be needed in current study regions for investigation of soil properties for future sampling, although the proportion of spatial variability within the distances appears smaller than the sampling distance which was unexplained (short-scale variability). This unexplained variation in both regions (nugget effect) was attributed to the measurement error and random variability of the property of interest.

MidDRIFTS-PLSR predicted data used in conjunction with geostatistics demonstrated its applicability to create more accurate, high resolution soil property maps at regional scale as compared to currently existing maps. The approach allowed mapping of soil properties with high accuracies. Nonetheless, the kriged maps for sand in K and TIC in SA region had relatively low accuracies which may have been due to the extremely high variability in dataset and non-normal distribution which resulted in lower accuracies as a consequence of Kriging smoothing effect (Ren, et al. 2008). The kriged map for SOM in SA region was comparable to that of reference map indicating large difference between the two maps in K region. It makes sense as reference maps were developed upon pedotransfer function with very low data resolution or coarse spatial information resulting in low-resolution spatial soil information. In addition, soil in K is loess and is very susceptible to erosion (Clemens and Stahr, 1994) which is not considered in the reference map and continued soil erosion since the map has been created. The previous findings of the application of spectroscopy and geostatisitcs were based NIR or Vis-NIR spectroscopy mainly at field or farm scale rather than MIRS for regional scale. Working on regional scale with large heterogeneity (e.g. samples from two contrasting regions and many fields within each region), midDRIFTS provides more accurate predictions as compared to NIR spectroscopy (Bellon-Maurel and McBratney, 2011). The integration of midDRIFTS and geostatistics provided a very promising approach to support digital soil property mapping at regional scale and speeds-up and refines existing soil property maps even for large geographical areas. This is particularly relevant for rapid, refined soil mapping in Germany as an alternative approach to traditional soil surveys which are very expensive and time-consuming.

5.6. Future work

MidDRIFTS can provide accurate prediction for most soil properties, with considerably less time and costs compared to conventional laboratory analyses. However, it still needs sample collection and preparation for scanning which makes it more time consuming as compared to portable NIRS (Bellon-Maurel and McBratney, 2011). Hence development of portable MIRS sensors for field application could be used to create high resolution soil property maps required in precision farming at the field scale. It has been stated that price and on-site use difficulties limit MIRS application in field. Merry and Janik (2001) successfully predicted carbonate and organic carbon, total nitrogen, cation exchange capacity and some exchangeable cations, electrical conductivity, pH, soil texture with a portable midFTIR sensor. The main issue in field applications is soil moisture and particle size which may affect both spectra and later calibration. How the spectra and/or calibration for different soil properties is affected by these factors needs to be tested by an appropriate comparison between laboratory and on-site midDRIFTS sensors.

Although the application of midDRIFTS spectroscopy in the current study was limited to arable lands, the developed models might further be used for land uses which were not included in calibration such as forest and pasture or for other geographical regions. This can be done by i) prediction of properties of new unknown samples from new location using already developed calibrated models. It would have to be tested and with regard to the prediction results observed the final decision can be made ii) re-calibration of existing PLSER models by taking samples from new locations where target soil properties already are known. Basically, midDRIFTS-PLSR model are adapted to local conditions and might not be applicable for areas outside of the area representing dissimilar spectral information as compared to those used in calibration (Bellon-Maurel and McBratney, 2011; Viscarra Rossel, et al. 2006). Therefore, using soil spectra from outside the study area could provide accurate predictions with low bias only if the soil from new locations has similar spectra as those used in calibration. Otherwise, the models must be extended for new locations by adding new soil spectra and associated laboratory measured values for different soil properties.

As discussed above, carbonates hamper the quantification of the 2930 cm^{-1} peak as related to SOM. Althought pre-treatment of carbonate containing soils via acid fumigation (Smith, et al. 1975) is seen as less descrutive of non-carbonate substances and organic matter, it still destroyes a portion of the SOM as seen by changes in midDRIFTS spectra. Additionally, the acid treatment leads to hygroscopic salt formation, which creates the problem of rapid water absorption of the sample and further interference of the mid-IR spectrum. Subtraction of ashed sample spectra from untreated ssample spectra which in theory should result in only peaks of SOM (Reeves III, 2012) has been recommended to avoid mineral interference signals such as clay and silica. Nonetheless, soil minerals undergo extreme changes (e.g. water loss, structural change)during ashing and as a resultthe spectra are also altered. Nonetheless, it might be possible to calculate a correction factor from the spectral data obtained via acid fumigation of carbonate free samples for SOM lost during fumigation.. Therefore, additional acid fumigation tests across a range of carbonate free soils are required including a diverse range of TOC contents and soil types to see if it is possible to identify an appropriate correction factor for general application of spectral indices in carbonate containing soils. All approaches are viewed as an alternative to improve the presentation of SOM functional group peaks in midDRIFTS spectra.

Spatial variability of soil properties and their mapping in large-scales depends on the density and distribution of observations (Grunwald, 2011). Thus, to plan an effective sampling strategy for spatial statistics in areas of similar conditions, two main issues should be taken into account: (1) effective sample size (ESS) and (2) sample spatial distribution. Although an approximate minimum number of samples can be calculated at different probabilities (Cobo, et al. 2010), this cannot indicate how the samples should be distributed spatially. The sampling design used in the current study showed its feasibility in capturing regional spatial variability in the selected regions, except for sand and TIC. More accurate results might be received by a re-distribution of sample points across the regions by sub-regional stratification. Uncertainty propagation analysis (UPA) would be a possibility to optimize the sampling distribution. It evaluates the uncertainty associated with the kriging estimator within the mapping procedure by calculating a standard error or deviation (Mowrer and Congalton, 2000). Hence, once the uncertainty propagation on kriged maps is recognized, the spatial dependency might be captured effectively by a restructured sampling strategy; intensifying sampling points in the areas with very high uncertainty (Brodsky, et al. 2013). It may be a possible way to explain local variation and decrease the uncertainty on kriged maps as well. Besides, lower sampling densities (thinning of sample points) might be applied within the areas showing low spatial variability with relatively smooth spatial distribution for a future sampling campaign and monitoring purposes.

Application of auxiliary information which is correlated spatially to soil properties would improve the spatial analysis and mapping of other soil properties (McBratney, et al. 2003; Zhang, et al. 2012). Remote-sensed data obtained from air and space-borne platforms such as Landsat imagery measurements combined with topographical data may provide useful information for assessing soil variability (Huang, et al. 2007). This would provide an inexpensive source of secondary information to improve the soil maps under investigation. For example, the SOM map developed in current study does not provide information concerning soil erosion. Integration of SOM data with a digital elevation model (DEM) and derived characteristics (e.g. slope length, wetness index, slope convexity, concavity) as secondary information may indicate soil erodibility status. Thus, integration of midDRIFTS-PLSR-based prediction and remote sensing-based data as an auxiliary environmental information source would be a feasible tool to be linked in geostatistical analysis for providing high-resolution digital soil property maps. Nonetheless, the routine use of remote-sensed spectral data still remains challenging because of the need for proper atmospheric and geometric corrections, low signal-to-noise ratios, spatial-temporal variability of soil surface conditions, vegetation and residue cover (Cécillon, et al. 2009).

5.7. Concluding remarks and general outlook

The results from this dissertation demonstrated that midDRIFTS-PLSR was an effective tool for characterizing soil properties, assessing soil quality, and describing soil spatial variability and finally generating high resolution digital soil property maps. In spite of the applicability of midDRIFTS in successful prediction of various soil properties, there remain several uncertainties on its potential in the frame of region dependent soil heterogeneity. Moreover, depending on data availability in terms of size and variability an appropriate PLSR calibration approach for proper prediction is required. The present work demonstrated that midDRIFTS combined with PLSR was capable of predicting a wide range of biochemical and physical soil properties at regional scale from two different agro-ecological regions characterized by significantly different mean values for measured soil properties. Although generic cross-validation (GCV model) models led to comparable PLSR prediction models, the most accurate PLSR predictions of “unknown” samples (model validation) were obtained using the independent validation approach. This shows the need, when using cross-validation to still test the model on “unknown” samples for an accuracy assessment of the predictive capability. The initial accuracies observed in cross validated models, both generic and region specific were too optimistic and not realistic for future model performance on unknown samples. Therefore, the generic models based on independent validation (GIC model) was the best evaluated approach to robustly and accurately predict several soil properties of new, unknown samples from the respective study regions by analyzing their midDRIFTS spectra without any further laboratory measurements. .

This study additionally showed that midDRIFTS could be a feasible indicator of SOM quality and composition without need to PLSR calibration. The midDRIFTS spectral-based index obtained by peak area at 2930 cm^{-1} was correlated well to SOM fractions if mineral interference especially at SOM fundamental bands avoided. Hence, mineralogical interferences and overlapping vibrations at peaks corresponding to SOM functional groups demands strong caution in interpreting spectral data. Moreover, the approach could be used to develop an index to quantify TIC undertaking bands assigned to carbonate. This is free of PLSR model calibration and validation and can be applied to either fewer sample size and/or extremely heterogeneous samples which are not suitable for PLSR calibration. Hence, the spectral based derived relation in current study gives its feasibility to be used for up scaling from ex-situ to in-situ measurement up to imaging spectroscopy (Mid-IR hyper-spectral images) for measuring the spatial distribution of soil variables. However, the approach must be tested over extended spatial areas to capture larger variability in soil type and mineralogical composition.

The database gained by midDRIFTS-PLSR calibrated models could successfully use to allow modelling spatial variability and fine resolution regional scale mapping with emphasis on geostatistics. Coupling midDRIFTS-PLSR to geostatistics provided a suitable evaluation of modelling spatial variability in both regions which confirmed the suitability of the sampling design and density applied to the whole regions. Based on spatial dependency and separation distance (range) obtained for two regions, soil carbon fractions in Kraichgau are mainly influenced by land use management strategies (fertilizer application, tillage, and land use changes and other management practices) while intrinsic factors such as soil type, mineralogy and natural soil forming processes may be controlling soil properties in Swabian Alb. The effort further resulted in the development of high resolution kriged maps for soil properties with greater accuracy as compared with existing maps.

As final conclusion, our results proved the potential for midDRIFTS-PLSR spectroscopy as a rapid-throughput method for providing high quality predictions of a large regionally distributed soil dataset. These results highlight the potential of midDRIFTS-PLSR to be used for spatial analysis and digital soil property mapping in combination with geostatistics as a reliable approach to reduce the number of conventional laboratory analyses. Due to the high accuracy, kriged maps for soil properties can be used not only to update soil property maps for the Kraichgau and Swabian Alb study regions but also to facilitate appropriate and efficient regional land use planning and as inputs for modeling SOM dynamics on the regional scale.

Outlook

The use of midDRIFTS-PLSR for broad application outside regions in this study such as national level remains challenging. Hence, establishment of a national-level spectral library for Germany would be a possible option to obtain spectral data coverage for future applications. However, the question is whether developed models in current study could predict soil properties from new location. Until now, midDRIFTS research is mainly limited to specific geographical regions for model development. Hence, differences in soil origin, traditional laboratory measures of reference samples, spectroscopic measures in terms of sample preparation, type of sensor and scanning complicate the transfer of calibrated model from one area to another one. This issue of model transferability is a main drawback linked to spectroscopy. This is also why prediction performance for a specific property varies greatly from one study to another. Although a spectral library can provide comprehensive spectral data, it needs to be established by considering standardization in both analytical reference analysis and midDRIFT spectroscopy as suggested by Viscarra-Rossel et al. (2008). Therefore, before setting up a national-based spectral library it is necessary first to establish a standardized framework for all tasks involved in midDRIFTS analysis such as sample collection, preparation, storing and scanning as well as standard analytical procedures for reference samples. Nonetheless, using standardized spectral data based will not always lead to accurate prediction models (Reeves and Smith, 2009). This contradiction comes from differences in soil type, mineralogical base and land use which may affect the calibrated models. In this regard, PLSR calibration can be developed for different area or soil types which takes differences is soils into consideration. This points out the midDRIFTS technique complexities and need for attention in the details in creating a useful soil spectral library.

Soil conceptual pools of SOM which are needed for accurate modeling of SOC dynamics, differ by soil type, location and ecosystem. Hence, it is necessary initialize SOC pool sizes to model SOC turnover to predict SOM behavior under changing climate conditions. Regional SOC modeling often considers multi-compartment models using conceptual pools representing different SOC fractions (e.g., DAISY, Century, and Roth-C). The high resolution midDRIFTS-based soil property maps from the current study can be further used for initialization of the distribution of organic matter between the different pools of the DAISY model (DAISY soil organic matter turnover model) at the regional scale. DAISY is a deterministic model which models C and N by using two main pools of soil organic matter (SOM) (native (SOM) and added (AOM) organic matter) and one main soil microbial biomass (SMB) pool (Hansen, et al. 1991; Mueller, et al. 1996). Although the added organic matter pool (AOM) is not considered in current study, however, regional-basis full data coverage on SOM initial pool sizes can be used to determine soil microbial biomass C (SMB C) and native SOM C pools (SOM C=TOC−SMB C). Spatial soil texture data is also valuable which can be used in combination with SOM pools for model initialization. The turnover rates of SOM and slow SMB pools (higher for sandy soils) as well as its stabilizing efficiency (higher for clay soils) are modified by soil texture (sand, silt, clay) in the model (Hansen, et al. 1991). Therefore, the approach opens a new perspective to gain accurate spatial data coverage of soil C and N pools which is relevant for the application of SOM simulation models on a regional scale.

Each of these three C pools in DAISY is further sub-divided into two pools with fast and slow turnover rates. Hence, future work should focus on the partitioning of the total C pool (e.g. SOM) into the respective subpools C pools (SOM1 and SOM2) via midDRIFTS spectral indexes or more advanced Pyro-MIRS indexes which may overcome some of the interferences of the midDRIFTS specific peak area approach for quality indicators.

Chapter 6

References

6. References

Ad-hoc-AG Boden. 2005. Bodenkundliche Kartieranleitung, 5. Aufl., 438 S., Hannover.

Amundson, R. 2001. The Carbon Budget in Soils. Annual Review of Earth and Planetary Sciences. 29, 535-562.

Anderson, J.M. and Ingram, J.S.I. 1993. Tropical Soil Biology and Fertility: A Handbook of Methods. CAB International, Wallingford, UK.

Baes, A.U. and Bloom, P.R. 1989. Diffuse reflectance and transmission Fourier transform infrared (DRIFT) spectroscopy of humic and fulvic acids. Soil Science Society of America Journal, 53, 695-700.

Baldock, J.A., Hawke, B., Sanderman, J. and Macdonald, L. M. 2013. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Research, 51, 577-595.

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

Barthès, B.G., Brunet, D., Ferrer, H., Chotte, J.L. and Feller, C. 2006. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. Journal of Near Infrared Spectroscopy, 14, 341-348.

Batjes, N.H. 2002. Carbon and nitrogen stocks in the soils of Central and Eastern Europe. Soil use and Management, 18, 324-329.

Bayer, A., Bachmann, M., Müller, A. and Kaufmann, H. 2012. A Comparison of Feature-Based MLR and PLS Regression Techniques for the Prediction of Three Soil Constituents in a Degraded South African Ecosystem. Applied and Environmental Soil Science, Article ID 971252, 20 pages

Behrens, T. and Scholten, T. 2006. Digital soil mapping in Germany-a review. Journal of Plant Nutrition and Soil Science, 169, 434-443.

Bekele, A. and Hudnall W. H. 2006. Spatial variability of soil chemical properties of a prairie-forest transition in Louisiana. Plant and Soil, 280,7-21.

Bellamy, P.H., Loveland, P.J., Bradley, R.I., Lark, R.M. and Kirk, G.J.D. 2005. Carbon losses from all soils across England and Wales 1978–2003. Nature, 437,245-248.

Bellon-Maurel, V., Fernandez, E., Palagos, B., Roger, J.M. and McBratney, A. 2010. Prediction of soil attributes by NIR/MIR spectroscopy. Coming back to statistics fundamentals to improving prediction accuracy. TrAC-Trends in Analytical Chemistry, 29, 1073-1081.

Bellon-Maurel, V. and McBratney, A. 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils - Critical review and research perspectives. Soil Biology and Biochemistry, 43, 1398-1410.

Bernier, M.H., Levy, G. J. Fine, P. and Borisover, M. 2013. Organic matter composition in soils irrigated with treated wastewater: FT-IR spectroscopic analysis of bulk soil samples. Geoderma, 209-210, 233-240.

Bertrand, I., Janik, L.J., Holloway, R.E., Armstrong, R.D. and McLaughlin, M.J. 2002. The rapid assessment of concentrations and solid phase associations of macro- and micronutrients in alkaline soils by mid-infrared diffuse reflectance spectroscopy. Australian Journal of Soil Research, 40, 1339-1356.

Blake, G.R. and Hartage, K.H. 1986. Bulk density, In: Klute, A. (Eds.), Methods of Soil Analysis Part 1: Physical and Mineralogical Methods. Soil Science Society of America, Madison, pp. 363-376.

Borenstein, A., Linker, R., Shmulevich, I. and Shaviv, A. 2006. Determination of soil nitrate and water content using attenuated total reflectance spectroscopy. Applied Spectroscopy, 60, 1267-1272.

Bornemann, L.G.W. and Amelung, W. 2010. Particulate organic matter at the field scale: Rapid acquisition using mid-infrared spectroscopy. Soil Science Society of American Journal, 74, 1147-1156.

Bowers, S. A. and Hanks, R. J. 1965. Reflection of radiant energy from soils. Soil Science, 100, 130-138.

Brereton, R.G. 1990. Chemometrics: applications of mathematics and statistics to laboratory systems. Ellis Horwood, London, UK.

Brodský, L., Vašát, R., Klement, A., Zádorová, T. and Jakšík, O. 2013. Uncertainty propagation in VNIR reflectance spectroscopy soil organic carbon mapping. Geoderma, 199, 54-63.

Brown, D.J., Bricklemyer, R.S. and Miller, P.R. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma, 129, 251-267.

Brown, L.R. 2006. Plan B: Rescuing a planet under stress and a civilisation in trouble. W. W. Norton & Company, New York. pp. 365.

Bruckman, V.J. and Wriessnig, K. 2013. Improved soil carbonate determination by FT-IR and X-ray analysis. Environmental Chemistry Letters, 11, 65-70.

Brunet D, Barthe´s, B.G., Chotte, J.L. and Feller, C. 2007. Determination of Carbon and Nitrogen Contents in Alfisols, Oxisols and Ultisols from Africa and Brazil Using NIRS Analysis: Effects of Sample Grinding and Set Heterogeneity. Geoderma, 139, 106-117.

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

Burgos, P., Madej´on, E., P´erez-de-Mora, A. and Cabrera, F. 2006. Spatial variability of the chemical characteristics of a trace-element-contaminated soil before and after remediation. Geoderma, 130, 157–175.

Burrough, P.A., Beckett, P.H.T. and Jarvis, M.G. 1971. The relation between cost and utility in soil survey (I–III). Journal of Soil Science, 22,359-394.

Burrough, P.A., Van Gaans, P.F.M. and Hootsmans, R. 1997. Continuous classification in soil survey: Spatial correlation, confusion and boundaries. Geoderma, 77,115-135.

Busca, G. and Resini, C. 2000. Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials. In: Meyers, R.A., (Eds.). Anonymous Encyclopedia of Analytical Chemistry. John Wiley & Sons, Ltd., Chichester, UK, pp. 10954-11008.

Butkute´, B. and lepetiene´, A.S. 2006. Application of Near Infrared Reflectance Spectroscopy for the Assessment of Soil Quality in a Long-Term Pasture. Communications in Soil Science and Plant Analysis, 37, 2389-2409.

Calderon, F., McCarty, G. and Revees, J. 2006. Pyrolisis-MS and FT-IR analysis of fresh and decomposed dairy manure. Journal of Analytical and Applied Pyrolysis, 76, 14-23.

Calderon F.J., Mikha, M.M., VIGIL, M.F., Nielsen, D.C., Benjamin, J.B. and Reeves III, J.B. 2011a. Diffuse-Reflectance Mid-infrared Spectral Properties of Soils under Alternative Crop Rotations in a Semi-arid Climate. Communications in Soil Science and Plant Analysis, 42, 2143-2159.

Calderon, F. J., Reeves III, J. B., Collins, H. P. and Paul, E. A. 2011b. Chemical differences in soil organic matter fractions determined by diffuse reflectance mid-infrared spectroscopy. Soil Science Society of America Journal, 75, 568-579.

Calderon, F., Haddix, M., Conant, R., Magrini-Bair, K. and Paul, E. 2013. Diffuse-Reflectance Fourier-Transform Mid-Infrared Spectroscopy as a Method of Characterizing Changes in Soil Organic Matter. Soil Science Society of America Journal, 77, 1591-1600.

Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F. and Konopka, A.E. 1994. Field-scale variability of soil properties in Central Iowa soils. Soil Science Society of America Journal, 58, 1501-1511.

Cambardella, C.A., Gajda, A.M., Doran, J.W., Wienhold, B.J. and Kettler, T.A. 2001 Kettler. Estimation of particulate and total organic matter by weight loss-on-ignition. In: Lal R et al. (ed) Assessment methods for soil carbon. Lewis Publ., Boca Raton, FL, pp 349-359.

Cécillon, L., Barthès, B.G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A. and Brun, J.J. 2009. Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). European Journal of Soil Science, 60, 770-784.

Chabrillat, S. E., and Ben-Dor, Viscarra Rossel, R.A. and Demattê, J.A.M. 2013. Quantitative Soil Spectroscopy. Applied and Environmental Soil Science, Article ID 616578, 3 pages

Chang, C.W., Laird, D.A., Mausbach, M.J. and Hurburgh Jr, C.R., 2001. Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. Soil Science Society of America Journal, 65, 480-490.

Chaplot V., Bouahom, B. and Valentin, C. 2010. Soil organic carbon stocks in South-east Asia: spatial variations and controlling actors. Global Change Biology, 16, 1380-1393.

Chodak, M. 2011. Near-infrared spectroscopy for rapid estimation of microbial properties in reclaimed mine soils. Journal of Plant Nutrition and Soil Science, 174, 702-709.

Clemens, G. and Stahr, K. 1994. Present and past soil erosion rates in catchments of the Kraichgau area (SW-Germany). Catena, 22, 153-168.

Coates, J. 2000. Interpretation of Infrared Spectra, A Practical Approach. Encyclopedia of Analytical Chemistry. John Wiley & Sons Ltd, Chichester

Coblentz, W.W. 1905. Investigations of Infrared Spectra, In Carnegie Institute of Washington, The Plimpton Press, Norwood, Mass., USA.

Cobo, J.G., Dercon, G., Yekeye, T., Chapungu, L., Kadzere, C., Murwira, A., Delve, R. and Cadisch, G. 2010. Integration of mid-infrared spectroscopy and geostatistics in the assessment of soil spatial variability at landscape level. Geoderma, 158, 398-411.

Cohen, M.J., Prenger, J.P. and DeBusk, W.F. 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. Journal of Environmental Quality, 34, 1422-1434.

D'Acqui, L.P., Pucci, A. and Janik, L.J. 2010. Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. European Journal of Soil Science, 61, 865-876.

Dardenne, P., Sinnaeve, G. and Baeten, V. 2000. Multivariate calibration and chemometrics for near infrared spectroscopy: Which method? Journal of Near Infrared Spectroscopy, 8, 229-237.

Davidson, E.A. and Janssens, I.A. 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature, 440, 165-173.

Degryze, S., Six, J., Paustian, K., Morris, S.J., Paul, E.A. and Merckx, R. 2004. Soil organic carbon pool changes following land-use conversions. Global Change Biology, 10, 1120-1132.

Demyan, M.S., Rasche, F., Schulz, E., Breulmann, M., Müller, T. and Cadisch, G. 2012. Use of specific peaks obtained by diffuse reflectance Fourier transform mid-infrared spectroscopy to study the composition of organic matter in a Haplic Chernozem. Europiean Journal of Soil Science, 63, 189-199.

Dixon, J.B. and Weed, S.B. 1989. Minerals in Soil Environments. Soil Science Society of America, Madison, WI, USA.

Du, C.W., Raphael, L. and Shaviv, A. 2007. Characterization of soils using photoacoustic mid-infrared spectroscopy. Applied Spectroscopy, 61, 1063-1067.

Du, C. and Zhou, J. 2009. Evaluation of soil fertility using infrared spectroscopy: A review. Environmental Chemistry Letters, 7, 97-113.

Du, C.W., Zhou, J.M., Wang, H.Y., Mang, J.B. and Zhu, A.N. 2008b. Study on the soil mid-infrared photoacoustic spectroscopy. Spectroscopy, 28, 1242-1245.

Du, C.W. and Zhou, J.M. 2011. Application of infrared photoacoustic spectroscopy in soil analysis. Applied Spectroscopy Reviews, 46, 405-422

Dunn, B.W., Beecher, H.G., Batten, G.D. and Ciavarella, S. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis - A case study from the Riverine Plain of south-eastern Australia. Australian Journal of Experimental Agriculture, 42, 607-614.

Düwel, O., Siebner, C.S. Utermann, J. and Krone, F. 2007a. Bodenarten der Böden Deutschlands. Bericht über länderübergreifende Auswertungen von Punktinformationen im FISBo BGR, Archiv der BGR, Hannover. http://www.bgr.bund.de/DE/Themen/Boden/Produkte/Schriften/Downloads/Bodenarten_Bericht.pdf?__blob=publicationFile&v=2

Düwel, O., Siebner, C.S., Utermann, J. and Krone, F. 2007b. Gehalte an organischer Substanz in Oberböden Deutschlands- Bericht über länderübergreifende Auswertungen von Punktinformationen im FISBo BGR, Archiv der BGR, Hannover. Http://www.bgr.bund.de/DE/Themen/Boden/Produkte/Schriften/Downloads/Humusgehalte_Bericht.pdf?__blob=publicationFile&v=2 ESRI 2013. ArcGIS Desktop: Release 10.2. Redlands, California, Environmental System Research Institute.

Ehsani, M. R., Upadhyaya, S. K., Fawcett, W. R., Protsailo, L. V. and Slaughter, D. 2001. Feasibility of detecting soil nitrate content using a mid-infrared technique. Transactions of the ASAE 446, 1931-1940.

FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

FISBo BGR, 2006. Bodenübersichtskarte der Bundesrepublik Deutschland 1:1000000 (Digitales Archiv FISBo BGR; BÜK 1000 N, Vers. 2.3.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D., Renner, S., Schöning, I., Schumacher, U., Wells, K., Buscot, F., Kalko, E.K.V., Linsenmair, K.E., Schulze, E.D. and Weisser, W.W. 2010. Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic and Applied Ecology, 11, 473-485.

Fu, W. J., Tunney, H. and Zhang, C. S. 2010. Spatial variation of soil test phosphorus in a long-term grazed experimental grassland field. Journal of Plant Nutrition and Soil Science, 173, 323-331.

Gallardo, A. 2003. Spatial variability of soil properties in a floodplain forest in northwest Spain. Ecosystems, 6, 564–576, 2003.

Gallardo, A., Rodríguez-Saucedo, J.J., Covelo, F. and Fernández-Alés, R. 2000. Soil Nitrogen heterogeneity in a Dehesa ecosystem. Plant and Soil, 222, 71-82.

Gallardo, A. and Paramá, R. 2007. Spatial variability of soil elements in two plant communities of NW Spain. Geoderma, 139, 199-208.

Garten Jr., C.T., Kanga, S., Bricea, D.J., Schadta, C.W. and Zho, J. 2007. Variability in soil properties at different spatial scales (1 m–1 km) in a deciduous forest ecosystem. Soil Biology and Biochemistry, 39, 2621-2627.

Gehl, R.J. and Rice, C.W. 2007. Emerging technologies for in situ measurement of soil carbon. Climate Change, 80, 43-54.

Gerzabek, M. H., Antil R. S., Kogel-Knabner I., Knicker H., Kirchmann H. and Haberhauer G. 2006. How are soil use and management reflected by soil organic matter characteristics: A spectroscopic approach. European Journal of Soil Science, 57, 485-494.

Giacometti, C., Demyan, M.S., Cavani, L., Marzadori, C., Ciavatta, C. and Kandeler, E. 2013. Chemical and microbiological soil quality indicators and their potential to differentiate fertilization regimes in temperate agroecosystems. Applied Soil Ecology, 64, 32-48.

Gobeille, A., Yavitt, J., Stalcup, P. and Valenzuela, A. 2006. Effects of soil management practices on soil fertility measurements on Agave tequilana plantations in Western Central Mexico. Soil and Tillage Research, 871, 80-88.

Goidts, E. and van Wesemael, B. 2007. Regional assessment of soil organic carbon changes under agriculture in Southern Belgium (1955–2005). Geoderma, 141, 341-354.

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

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford University Press, P. 75-116.

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

Goovaerts, P. 2001. Geostatistical modelling of uncertainty in soil science. Geoderma, 103, 3-26.

Grace, J. 2004. Understanding and managing the global carbon cycle. Journal of Ecology, 92, 189-202.

Gremlich, H. U. 2000. Infrared and Raman Spectroscopy. In: Ullmann's Encyclopedia of Industrial Chemistry. Wiley-VCH Verlag GmbH & Co. KGaA.

Griffith, D. 2005. E_ective geographic sample size in the presence of spatial autocorrelation. Annals of the Association of American Geographers, 95, 740-760.

Grinand, C., Barthes, BG., Brunet, D., Kouakoua, E., Arrouays, D., Jolivet, C., Caria, G. and Bernoux, M. 2012. Prediction of soil organic and inorganic carbon contents at a national scale (France) using mid-infrared reflectance spectroscopy (MIRS). European Journal of Soil Science, 63, 141-151.

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

Grunwald, S. 2010. The current state of digital soil mapping and what is next. In: Boetinger, J., Howell, D.W., Moore, A.C., Hartemink, A.E. and Kienst-Brown, S. (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, 1201-1213.

Haaland, D.M. and Thomas, E.V. 1988. Partial least-squares methods for spectral analyses .2. Application to simulated and glass spectral data. Analytical Chemistry, 60, 1202-1208.

Haberhauer, G., Rafferty, B., Strebl, F. and Gerzabek, M. H. 1998. Comparison of the composition of forest soil litter derived from three different sites at various decompositional stages using FTIR spectroscopy. Geoderma, 83, 331-342.

Haberhauer, G. and Gerzabek, M.H. 1999. Drift and transmission FT-IR spectroscopy of forest soils: an approach to determine decomposition processes of forest litter. Vibrational Spectroscopy, 19, 413-417.

Hansen, S., Jensen, H.E., Nielsen, N.E. and Svendsen, H. 1991. Simulation of nitrogen dynamics in the soil‑plant system using the Danish simulation model DAISY. Fertilizer Research, 27, 245‑259

Harris, D., Horwath, W.R. and van Kessel, C. 2001. Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. Soil Science Society of America Journal, 65, 1853-1856.

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

Hartemink, A.E. and McBratney, A. 2008. A soil science renaissance. Geoderma, 148, 123-129.

Hartwich, R., Behrens, J., Eckelmann, W., Haase, G., Richter, A., Roeschmann, G. and Schmidt, R. 1995. Bodenübersichtskarte der Bundesrepublik Deutschland 1:1,000,000 (BÜK 1000). Erläuterungen und Textlegende (Beiheft zur Karte). Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Hannover, Germany.

Hassink, J. 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil, 191, 77-87.

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

Heuvelink, G.B.M. and Webster, R. 2001. Modeling soil variation: past, present, and future. Geoderma, 100, 269-301.

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

Huang, P.M., Wang, M.K. and Chiu, C.Y. 2005. Soil mineral–organic matter–microbe interactions:Impact on biogeochemical processes and biodiversity in soils. Pedobiologia, 49, 609-635.

Huang, X., Senthilkumar, S., Kravchenko, A., Thelen, K. and Qi, J. 2007a. Total carbon mapping in glacial till soils using near-infrared spectroscopy, Landsat imagery and topographical information. Geoderma, 141, 34–42.

Huang, P.M., Wang, M.K. and Chiu, C.Y. 2005. Soil mineral–organic matter–microbe interactions:Impact on biogeochemical processes and biodiversity in soils. Pedobiologia, 49, 609-635.

Huang, S.W., Jin, J.Y., Yang, L.P. and Bai, W.L. 2006. Spatial variability of soil nutrients and influencing factors in a vegetable production area of Hebei Province in China. Nutrient Cycling in Agroecosystems, 75, 201-212.

Hunt, G.R. and Salisbury, J.W. 1970. Visible and near infrared spectra of minerals and rocks. I. Silicate minerals. Modern Geology, 1, 283-300.

Idowu, O.J., van Es, H.M., Abawi, G.S., Wolfe, Ball, J.I., Gugino, B.K., Moebius, B.N., Schindelbeck, R.R. and Bilgili, A.V. 2008. Farmer-oriented assessment of soil quality using field, laboratory, and VNIR spectroscopy methods. Plant and Soil, 307, 243-253.

Ingwersen ,J., Steffensa, K., Högy, P., Warrach-Sagic, K., Zhunusbayevab, D., Poltoradneva, M., Gäblera, R., Wizemannc, H.D., Fangmeierb, A., Wulfmeyerc, V. and Strecka, T. 2011. Comparison of Noah simulations with eddy covariance and soil water measurements at a winter wheat stand. Agricultural and Forest Meteorology, 151, 345-355.

IPCC, 2007. Climate change 2007. Climate change impacts, adaptation and vulnerability. Working group II. Geneva, Switzerland

Islam, K., Singh, B. and McBratney, A. 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian Journal of Soil Research, 41, 1101-1114.

ISO 10693, 1995. Soil Quality-Determination of Carbonate Content–volumetric Method. International Organization for Standardization, Geneva.

ISO 10694, 1995. Soil Quality. Determination of Organic and Total Carbon After Dry Combustion (Elementary Analysis). International Organization for Standardization, Geneva.

Jackson, M. L. 1958. Soil Chemical Analysis. Prentice-Hall., Inc., Englewood Cliffs, N. J., USA.

Janik, L.J., Merry, R.H., Forrester, S.T., Lanyon, D.M. and Rawson, A. 2007. Rapid prediction of soil water retention using mid infrared spectroscopy. Soil Science Society of America Journal, 71, 507-514.

Janik, L.J., Merry, R.H. and Skjemstad, J.O. 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? Australian Journal of Experimental Agriculture, 38, 681-696.

Janik, L.J., Merry, R.H., Forrester, S.T., Lanyon, D.M and Rawson, A. 2007. Rapid prediction of soil water retention using mid infrared spectroscopy. Soil Science Society of American Journal, 71, 507-514.

Jenkinson, D.S., Adams, D.E. and Wild, A. 1991. Model Estimates of CO_2 Emissions from Soil in Response to Global Warming. Nature, 351, 304-306.

Jenny, H. 1941. Factors of Soil Formation. McGraw-Hill, New York, 281 pp

Ji, J.F., Ge, Y., Balsam, W., Damuth, J.E. and Chen, J. 2009. Rapid identification of dolomite using a fourier transform infrared spectrophotometer (FTIR): a fast method for identifying Heinrich events in IODP Site U1308. Marine Geology, 258, 60-68.

Joergensen, R.G. and Mueller, T. 1996. The fumigation-extraction method to estimate soil microbial biomass: Calibration of the kEN value. Soil Biology and Biochemistry, 28, 33-37.

Jones, C., McConnel, C., Coleman, K., Cox, P., Falloon, F., Jenkinson, D. and Powlson, D. 2005. Global climate change and soil carbon stocks; predictions from two contrasting models for the turnover of organic carbon in soil. Global Change Biology, 11, 154-166

Kaiser, M. and Ellerbrock, R.H. 2005. Functional characterization of soil organic matter fractions different in solubility originating from a long-term field experiment. Geoderma, 127, 196-206.

Kamogawa, M,Y., Nogueira, A.R.A., Miyazawa, M., Artigas, J. and Alonso, J. 2001. Determination of soil calcareous efficiency using flow system with pervaporative separation. Analytica Chimica Acta, 438, 273-279.

Kempen B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G.B.M and de Vries, F. 2011. Efficiency Comparison of Conventional and Digital Soil Mapping for Updating Soil Maps. Soil Science Society of American Journal, 76, 2097-2115.

Kimble, J.M., Lal, R. and Follet, R.F. 2001. Methods for assessing soil C pools. In: Lal R et al. (ed) Assessment methods for soil carbon. Lewis Publ., Boca Raton, FL, pp 3-12.

Krug, D. and Hartwich, R. 2001. Die Flächendatenbank der Bodenübersichtskarte 1:200.000 (BÜK 200): Basisdaten für den länderübergreifenden Bodenschutz. In: Z. f. angew. Geol. 47/2, S. 114-120, Hannover.

Kuang, B. and Mouazen, A.M. 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. European Journal of Soil Science, 62, 629-636.

Lagacherie, P. 2008. Digital Soil Mapping: A State of the Art. In: Hartemink, A.E., McBratney, A. and Mendonça-Santos M.L. (Eds.), Digital Soil Mapping with Limited Data. pp. 3-14, Dordrecht, Springer.

Lal R., Griffin, M., Apt, J., Lave, L. and Morgan, M.G. 2004. Managing soil carbon. Science, 304), p. 393.

Lal, R. 2004a. Soil carbon sequestration impacts on global climate change and food security. Science, 304, 1623-1627.

Lal, R. 2004b. Soil carbon sequestration to mitigate climate change. Geoderma, 123, 1-22.

Lal, R. 2008. Soil carbon stocks under present and future climate with specific reference to European ecoregions. Nutrient Cycling in Agroecosystems, 81, 113-127.

Legodi, M.A, de Waal, D., Potgiete,r J.H. and Potgieter, S.S. 2001. Technical note rapid determination of $CaCO_3$ in mixtures utilising FT-IR spectroscopy. Minerals Engineering, 14, 1107-1111.

Linker, R. 2011. Application of FTIR Spectroscopy to Agricultural Soils Analysis, Fourier Transforms -New Analytical Approaches and FTIR Strategies, Prof. Goran Nikolic (Ed.), ISBN: 978-953-307-232-6, InTech, Available from:

http://www.intechopen.com/books/fourier-transforms-new-analytical-approaches-and-ftirstrategies/application-of-ftir-spectroscopy-to-agricultural-soils-analysis.

Linker, R., Shmulevich, I., Kenny, A. and Shaviv, A. 2005. Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR midinfrared spectroscopy. Chemosphere, 61, 652-658.

Linker, R., Weiner, M., Shmulevich, I. and Shaviv, A. 2006. Nitrate determination in soil pastes using attenuated total reflectance mid-infrared spectroscopy: Improved accuracy via soil identification. Biosystems Engineering, 94, 111-118.

Liu, D., Wang, Z., Zhang, B., Song, K., Li, X., Li, J., Li, F. and Duan, H. 2006. Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. Agriculture, Ecosystems and Environment, 113, 73-81.

Liu, X., Xu, J., Zhang, M. and Zhou, B. 2004. Effects of Land Management Change on Spatial Variability of Organic Matter and Nutrients in Paddy Field: A Case Study of Pinghu, China. Environmental Management, 34, 691-700.

Liu,L., Zeng, F.P., Song, T.Q., Peng, W.X., Wang, K.L., Qin, W.G. and Tan, W.N. 2010. Spatial heterogeneity of soil nutrients in Karst area' s Mulun National Nature Reserve," Chinese Journal of Applied Ecology, 21, 1667-1673.

Lopez-Granados, F., Jurado-Exposito, M., Atenciano, S., Garcia-Ferrer, A., De la Orden, M.S. and Garcia-Torres, L. 2002. Spatial variability of agricultural soil parameters in southern Spain. Plant Soil, 246, 97-105.

Ludwig, B., Nitschke, R., Terhoeven-Urselmans, T., Michel, K. and Flessa, H. 2008. Use of mid-infrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter. Journal of Plant Nutrition and Soil Science, 171, 384-391.

Madari, B.E., Reeves III, J.B., Machado, P.L.O.A., Guimarães, C.M., Torres, E. and McCarty, G.W. 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. Geoderma, 136, 245-259.

Madejová, J. 2003. FTIR techniques in clay mineral studies. Vibrational Spectroscopy, 31, 1-10.

Malhi, S.S., Grant, C.A., Johnston, A.M. and Gill, K.S. 2001. Nitrogen fertilization management for no-till cereal production in the Canadian Great Plains: A review. Soil Tillage Research, 60, 101-122.

Malley, D.F., Yesmin, L., Wray, D. and Edwards, S. 1999. Application of near-infrared spectroscopy in analysis of soil mineral nutrients. Communications in Soil Science and Plant Analysis, 30, 999-1012.

Marchetti, A., Piccni, C., Francaviglia, R. and Mabit, L. 2012. Spatial Distribution of Soil Organic Matter Using Geostatistics: A Key Indicator to Assess Soil Degradation Status in Central Italy. Pedosphere, 22, 230-242.

Mark, H.L. and Tunnell, D. 1985. Qualitative near-infrared reflectance analysis using Mahalanobis distances. Analytical Chemistry, 57, 1449-1456.

McBratney, A. B., Mendonça Santos, M. L. and Minasny, B. 2003. On digital soil mapping. Geoderma 117, 3-52.

McBratney, A. and Pringle, M. 1997. Spatial variability in soil implications for precision agriculture. In: Precision Agriculture 1997, edited by J.V. Stafford, BIOS, Oxford, UK, pp. 3-31.

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

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

McCarty, G.W. and Reeves, J.B. 2006. Comparison of NFAR infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. Soil Science, 171, 94-102.

McCarty, G.W., Reeves, J.B., Reeves, V.B., Follett, R.F. and Kimble, J.M. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Science Society of America Journal, 66, 640-646.

McDowell, M. L., Bruland, J.L., Deenik, J.L., Grunwald, S. and Knox, N.M. 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma, 189-190, 312-320.

McLauchlan, K.K. and Hobbie, S.E. 2004. Comparison of labile soil organic matter fractionation techniques. Soil Science Society of America Journal, 68, 1616-1625.

Merry, R.H. and L.J. Janik. Mid Infrared Spectroscopy for Rapid and Cheap Analysis of Soils. Australian Agronomy Conference, 3, 1130-1300.

Metz, B., Davidson, O.R., Bosch, P.R., Dave, R. and Meyer, L.A. (Eds.). 2007. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Michel, K., Terhoeven-Urselmans, T., Nitschke, R., Steffan, P. and Ludwig, B. 2009. Use of near- And mid-infrared spectroscopy to distinguish carbon and nitrogen originating from char and forest-floor material in soils. Journal of Plant Nutrition and Soil Science, 172, 63-70.

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

Minasny, B., McBratney, A.B., Malone, B.P. and Wheeler, I. 2013. Digital Mapping of Soil Carbon, Advances in Agronomy, 118, 1-47.

Minasny, B., Tranter, G., McBratney, A.B., Brough, D.M. and Murphy, B.W. 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. Geoderma, 153, 155-162.

Minasny, B.M., McBratney, A.B., Mendonca-Santos, M.L., Odeh, I.O.A. and Guyon, B. 2006. Prediction and digital mapping of soil carbon storage in the lower Namoi valley. Australian Journal of Soil Research, 44, 223-244.

Morón, A. and Cozzolino, D. 2002. Application of near infrared reflectance spectroscopy for the analysis of organic C, total N and pH in soils of Uruguay. Journal of Near Infrared Spectroscopy, 10, 215-221.

Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J. and Ramon, H. 2007. On-line measurement of some selected soil properties using a VIS-NIR sensor. Soil and Tillage Research, 93, 13-27.

Mowrer, H.T. and Congalton, R.G. 2000. Quantifying spatial uncertainty in natural resources: Theory and applications for GIS and remote sensing. Ann Arbor Press, Chelsea, 244 p.

Mueller, T., Jensen, L.S., Hansen, S. and Nielsen, N.E. 1996. Simulating soil carbon and nitrogen dynamics with the soil‑plant‑atmosphere system model DAISY. In: D.S. Powlson, D.S., Smith, P. and Smith, J.U. (Eds.), Evaluation of Soil Organic Matter Models, Using Existing Long-term Datasets, Springer, Berlin, pp. 275‑280.

Mueller, T.G. and Pierce, F.J. 2003. Soil carbon maps: enhancing spatial estimates with simple terrain attributes at multiple scales. Soil Science Society of America Journal, 67, 258–267.

Mulla, D.J and McBratney, A.B. 2000. Soil spatial variability. In Handbook of Soil Science, M. E. Sumner, Ed., pp. 321–352, CRC Press, Boca Raton, Fla, USA.

Nduwamungu, C., Ziadi, N., Tremblay, G.F., Parent, L. ´ E., Tremblay, G.F. and Thuri`es, B. 2009. Near-infrared reflectance spectroscopy prediction of soil properties: Effects of sample cups and preparation. Soil Science Society of America Journal, 73, 1896-1903.

Nguyen, T., Janik, L.J. and Raupach, M. 1991. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. Australian Journal of Soil Research, 29, 49-67.

Niemeyer, J., Chen, Y. and Bollag, J.M. 1992. Characterization of humic acids, composts, and peat by diffuse reflectance Fourier-transform infrared spectroscopy. Soil Science Society of America Journal, 56, 135-140.

Nordgren, A. 1988. Apparatus for the continuous, long-term monitoring of soil respiration rate in large numbers of samples. Soil Biology and Biochemistry, 20, 955-957.

Odlarea, M., Svenssonb, K. and Pell, M. 2005. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. Geoderma, 126, 193-202.

Paustian, K., Collins, H.P. and Paul, E.A. 1997. Management controls on soil carbon. In Soil Organic Matter in Temperate Agroecosys tems. (Eds.) Paul, E.A., Paustian, K., Elliott, E.T and Cole, C.V. CRC Press, Boca Raton, FL. pp. 15-49.

Pirie, A., Singh, B. and Islam, K. 2005. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. Australian Journal of Soil Research, 43, 713-721.

Pontius, R.G., Boersma, W., Castella, J., Clarke, K.C., de Nijs, T., Dietzel, C. et al. 2008. Comparing the input, output, and validation maps for several models of land change. Annals of Regional Science, 42, 11-47.

Prechtel, A., von Lützow, M., Schneider, B.U., Bens, O., Bannick, C.G., Kögel-Knabner, I. and Hüttl, R.F. 2009. Organic carbon in soils of Germany: Status quo and the need for new data to evaluate potentials and trends of soil carbon sequestration. Journal of Plant nutrition and Soil Science, 172, 601-614.

QGIS Development Team, 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

Rasche, F., Marhan, S., Berner, D., Keil, D., Kandeler, E. and Cadisch, G. 2013. MidDRIFTS-based partial least square regression analysis allows predicting microbial biomass, enzyme activities and 16S rRNA gene abundance in soils of temperate grasslands. Soil Biology and Biochemistry, 57, 504-512.

Reeves III, J. B. 2012. Mid-infrared spectral interpretation of soils: Is it practical or accurate? Geoderma, 189-190, 508-513.

Reeves III, J.B. 2009. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done? Geoderma, 158, 3-14.

Reeves III, J.B., McCarty, G.W. and Reeves, V.B. 2001. Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. Journal of Agricultural and Food Chemistry, 49, 766-772.

Reeves, J. B. and Smith, D. B. 2009. The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. Applied Geochemistry, 24, 1472-1481.

Reeves, J. B. 2009. Does the Spectral Format Matter in Diffuse Reflection Spectroscopy? Applied Spectroscopy, 63, 669-677.

Reeves, J. B., Francis, B. A. & Hamilton, S. K. 2005. Specular reflection and diffuse reflectance spectroscopy of soils. Applied Spectroscopy, 59, 39-46.

Ren, C.Y., Zhang, B., Wang, Z.M., Song, K.S., Liu, D.W. and Yang, G. 2008. Geostatistical analysis of soil organic carbon in the farming-pastoral ecotone of Northeast China. Geoinformatics and Joint Conference on GIS and Built Environment: Advanced Spatial Data Models and Analyses. Guangzhou, China.

Robertson, G. P. 2008. GS+: Geostatistics for the Environmental Sciences. Gamma Design Software. Plainwell, Michigan, USA.

Rodeghiero, M., Heinemeyer, A., Schrumpf, M. and Bellamy, P. 2009. Determination of soil carbon stocks and changes. In: Kutsch, W.L., Bahn, M. and Heinemeyer, A. (Eds.) Soil carbon dynamics: An integrated methodology. Cambridge University Press, Cambridge, pp. 49-75.

Rossel, R.A.V. and Behrens, T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma, 158, 46–54.

Rossi, J., Govaerts, A., Vos, B.D., Verbist, B., Vervoort, A., Poesen, J., Muys, B. and Deckers, J. 2009. Spatial structures of soil organic carbon in tropical forests - A case study of Southeastern Tanzania. Catena, 77, 19-27.

Rusco, E., Jones, R. a nd Bidoglio, G. 2001. Organic matter in the soils of Europe: Present status and future trends. European Soil Bureau, JRC Ispra, Italy, pp. 14.

Russel, J.D. 1987. Infrared methods, In: Wilson, M.J. (Eds.), A Handbook of Determinative Methods in Clay Minerals. Blackie, Glasgow, Glasgow, pp. 133-173.

Sarkhot, D.V., Comerford, N.B., Jokela, E.J., Reeves III, J.B. and Harris, W. 2007. Aggregation and aggregate carbon in a forested southeastern coastal plain Spodosol. Soil Science Society of America Journal, 71, 1779-1787.

Schlichting, E., Blume, H.P. and Stahr, K. 1995. Bodenkundliches Praktikum—eine Einführung in pedologisches Arbeiten für ökologen, insbesondere Land- und Forstwirte, und für Geowissenschaftler. Blackwell Wiss.-Verl., Berlin; Wien u.a. in: Bruckman, V.J. and Wriessnig, K. 2013. Improved soil carbonate determination by FT-IR and X-ray analysis. Environmental Chemistry Letters, 11, 65-70.

Schulz, E. and Körschens, M. 1998. Characterization of the decomposable part of soil organic matter (SOM) and transformation processes by hot water extraction. Eurasian Soil Science, 31, 809-813.

Senesi, N., D’Orazio, V. and Ricca, G. 2003. Humic acids in the first generation of EUROSOILS. Geoderma, 116, 325-344.

Shenk, J.S. and Westerhaus, M.O. 1993. Near infrared reflectance analysis with single and multiproduct calibrations. Crop science, 33, 582-584.

Shepherd, K.D. and Walsh, M.G. 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal, 66, 988-998.

Shepherd, K.D. and Walsh, M.G. 2004. Diffuse Reflectance Spectroscopy for Rapid Soil Analysis, In R. Lal, ed. Encyclopedia of Soil Science. Marcel Dekker, Inc.

Shepherd, K.D. and Walsh, M.G. 2007. Infrared spectroscopy - enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. Journal of Near Infrared Spectroscopy, 15, 1-20.

Six, J., Conant, R.T., Paul, E.A. and Paustian, K. 2002. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. Plant Soil, 241, 155-176.

Smith D.M., Griffin, J.J. and Goldberg, E.D. 1975. Spectrometric method for the quantitative determination of elemental carbon. Analytical Chemistry, 47, 233-238.

Socrates, G. 1980. Infrared Characteristic Group Frequencies. John Wiley and Sons, Chichester, UK.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M. and McLaughlin, M.J. 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Applied Spectroscopy Reviews, 49, 139-186.

Soriano-Disla, J.M., Janik, L., McLaughlin, M.J., Forrester, S., Kirby, J. and Reimann, C. 2013. The use of diffuse reflectance mid-infrared spectroscopy for the prediction of the concentration of chemical elements estimated by X-ray fluorescence in agricultural and grazing European soils. Applied Geochemistry, 29, 135-143.

Spaccini, R. and Piccolo, A. 2007. Molecular characterization of compost at increasing stages of maturity. 1. Chemical fractionation and infrared spectroscopy. Journal of Agricultural and Food Chemistry, 55, 2293-2302.

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

Stevens, A., Van Wesemael, B., Vandenschrick, G., Toure, S. and Tychon, B. 2006. Detection of carbon stock change in agricultural soils using spectroscopic techniques. Soil Science Society of America Journal, 70, 844-850.

Stevenson, F.J. 1982. Humus Chemistry: Genesis, Composition, Reactions. John Wiley & Sons, Inc., New York.

Stevenson, F.J., 1994. Humus Chemistry: Genesis, Composition, Reactions. John Wiley & Sons, Inc, New York.

Stoner, E.R. and Baumgardner, M.F. 1981. Characteristic variations in reflectance from surface soils. Soil Science Society of America Journal, 45, 1161-1165.

Stuart B. 2004. Infrared spectroscopy: Fundamentals and applications. New York: Wiley.

Tans, P.P., Fung, I.Y. and Takahashii, T. 1990. Observational Constraints on the Global Atmospheric CO_2 Budget. Science, 247, 1432-1438.

Tatzber, M., Stemmer, M., Spiegel, H., Katzlberger, C., Haberhauer, G. and Gerzabek, M. 2007. An alternative method to measure carbonate in soils by FT-IR spectroscopy. Environmental Chemistry Letters, 5, 9-12.

Tatzber, M., Mutsch, F., Mentler, A., Leitgeb, E., Englisch, M. and Gerzabek, M.H. 2010. Determination of organic and inorganic carbon in forest soil samples by midinfrared spectroscopy and partial least squares regression. Applied Spectroscopy, 64, 1167-1175.

Tesfahunegn, G.B. , Tamene, L. and Vlek, PL..G. 2011. A participatory soil quality assessment in northern Ethiopia's Mai-Negus catchment Catena, 86, 1-13.

Tittonell, P., Vanlauwe, B., de Ridder, N. and Giller. K.E. 2007. Heterogeneity of crop productivity and resource use efficiency within smallholder Kenyan farms: Soil Soil fertility gradients or management intensity gradients? Agricultural Systems, 94, 376-390.

Udelhoven, T., Emmerling, C. and Jarmer, T. 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. Plant and Soil, 251, 319-329.

UN. 1998. Kyoto Protocol to the United Nations Framework Convention on Climate Change. Technical report.

Varmuza, K. and Filmoser, P. 2009. Introduction toMultivariate Statistical Analysis in Chemometrics. Taylor & Francis: Boca Raton, FL.

Vasques, G.M., Grunwald, S. and Sickman, J.O. 2009. Modeling of soil organic carbon fractions using visible–near-infrared spectroscopy. Soil Science Society of American Journal, 73, 176-184.

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

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

Viscarra Rossel, R. and Lark, R. 2009. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. European Journal of Soil Science, 60, 453-464.

Vohland, M., Ludwig, M., Thiele-Bruhn, S. and Ludwig, B. 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma, 223-225, 88–96.

Wander, M.M., Traina, S.J., Stinner, B.R and Peters, S.E. 1994. Organic and conventional management effects on biologically active soil organic matter pools. Soil Science Society of American Journal, 58, 1130-1139.

Wang, Y., Shao, M. and Liu, Z. 2010. Large-scale spatial variability of dried soil layers and related factors across the entire Loess Plateau of China. Geoderma, 159, 99-108

Wang, Y., Zhang, X. and Huang, C. 2009. Spatial variability of soil total nitrogen and soil total phosphorus under different land uses in a small watershed on the Loess Plateau, China. Geoderma, 150, 141-149.

Wang, Q. and Wang, S. 2011. Response of labile soil organic matter to changes in forest vegetation in subtropical regions. Applied Soil Ecology, 47, 210-216.

Webster, R. and Oliver, M.A. 2001. Geostatistics for Environmental Scientists. John Wiley, Chichester, U.K

Webster, R. and Oliver, M.A. 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science, 43, 177-192

Wei, J., Xiao, D., Zeng, H. and Fu, Y. 2008. Spatial variability of soil properties in relation to land use and topography in a typical small watershed of the black soil region, northeastern China. Environmental Geology, 53, 1663-1672.

Wetterlind, J., Stenberg, B. and Söderström, M. 2008. The use of near infrared (NIR) spectroscopy to improve soil mapping at the farm scale. Precision Agriculture, 9, 57-69.

White, J.L. and Roth, C.B. 1986. Infrared spectrometry, In: Klute, A. (Eds.), Methods of Soil Analysis: Part 1. Physical and Mineralogical Methods. Soil Science Society of America, Madison, pp. 291-330.

Wilding, L.P. and Lin, H. 2006. Advancing the frontiers of soil science towards a geoscience. Geoderma, 131, 257–274.

Wold S., Kettaneh-Wold, N. and Skagerberg, B. 1989. Nonlinear PLS modeling. Chemometrics and Intelligent Laboratory Systems, 7, 53-65.

Workman Jr, J.J. 1992. NIR spectroscopy calibration basics. In: Burns, D.A. and Ciurczak, E.W. (Eds.). Handbook of near-infrared Analysis. Marcel Dekker, New York, pp. 247-280.

WRB .2007. World Reference Base for Soil Resources 2006: first update 2007. A Framework for International Classification, Correlation and Communication, 103. FAO, Rome.

Yang, H. and Mouazen, A.M. 2012. Vis/Near- and Mid- Infrared Spectroscopy for Predicting Soil N and C at a Farm Scale, Infrared Spectroscopy - Life and Biomedical Sciences, Prof. Theophanides Theophile (Ed.), ISBN: 978-953-51-0538-1, InTech, Shanghai.

Zaehle, S., Bondeau, A., Carter, T.R., Cramer, W., Erhard, M., Prentice, I.C., Reginster, I., Rounsevell, M.D.A, Sitch, S., Smith, B., Smith, P.C. and Sykes, M. 2007. Projected changes in terrestrial carbon storage in Europe under climate and land-use change, 1990-2100. Ecosystems, 10, 380-401.

Zhang, S., Huang, Y., Shen, C., Ye, H. and Du, Y. 2012. Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. Geoderma, 171–172, 35-43.

Zhang, Z., Hu, B. and Hu, G. 2014. Spatial Heterogeneity of Soil Chemical Properties in a Subtropical Karst Forest, Southwest China. Hindawi Publishing Corporation , The Scientific World Journal, Article ID 473651, 1-9.

Zhu, A.X., Hudson, B., Burt, J., Lubich, K. and Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of American Journal, 65, 1463-1472.

Zhu, A.X. 1997. A similarity model for representing soil spatial information. Geoderma, 77, 217-242.

Zimmermann, M., Leifeld, J. and Fuhrer, J. 2007. Quantifying soil organic carbon fractions by infrared-spectroscopy. Soil Biology and Biochemistry, 39, 224-231.

Peterson, S.A., Urquhart, N.S. and Welch, E.B. 1999. Sample representativeness: A must for reliable regional lake condition estimates. Environ. Sci. Technol. 33, 1559–1565.

Luck, M. and J. G. Wu. 2002. A gradient analysis of the landscape pattern of urbanization in the Phoenix metropolitan area of USA. Landscape Ecology17, 327-339.

7. Abstract

Changing climate conditions and land-use change severely affect key ecosystem processes in soils. Hence, regular monitoring of essential soil properties such as soil carbon and nitrogen pools, nutrient status and pH are required to implement appropriate soil management in agro-ecosystems. However characterizing soil properties at different spatial scales remains challenging as it requires a large amount of geo-referenced data by intensive sampling. Therefore a more efficient, rapid and low-cost approach is needed to develop soil databases for spatial analysis and monitoring at regional scale and for the calibration of soil C and nutrient turnover models at the regional scale. The application of mid-infrared diffuse reflectance infrared Fourier transform spectroscopy (midDRIFTS) could play an important role in this regard. This dissertation aimed to assess midDRIFTS in combination with partial least square regression (PLSR) as a rapid-throughput method to quantify soil properties, to assess soil quality, to model spatial patterns, and to create soil property maps at the regional scale of two agro-ecological areas, Kraichgau (K) and Swabian Alb (SA), (40 by 40 km) in southwest Germany.

As a first step, a pre- sampling at the regional scale was done to develop the most efficient midDRIFTS-PLSR prediction models by testing two different calibration procedures, i.e. cross-validation and calibration with independent validation, to quantify total carbon (TC), organic carbon (TOC), inorganic carbon (TIC), nitrogen (TN), mineral N (N_{min}), C:N ratio, hot water extractable C and N (C_{HWE}, N_{HWE}), microbial biomass C and N (C_{mic}, N_{mic}), pH, bulk density, clay, silt and sand content with 126 sample points across the two regions. Generic calibration models using either calibration approach were obtained from a set of 84 samples, while the remaining 42 samples were used for model validation. Cross-validation was further used for region specific midDRIFTS-PLSR models from a set of 42 and 21 samples for model calibration and validation, respectively. Based on calibrated models, most soil properties were predicted successfully using either calibration approach (residual prediction deviation (RPD) >3 and R^2 >0.9), except for Nmin, C/N ratio, pH, bulk density and sand. Application of the models on "unknown" samples revealed that the generic cross-validation models (GCV) accuracy was considerably lower than generic independent validation models (GIC). Similarly, accuracies of region specific cross-validated models in prediction of "unknown" samples were also lower in both regions when compared to the GIC models. In general the independent validation approach

was the more robust approach for predicting soil properties of “unknown” samples without any further laboratory measurements.

The next step was the integration of midDRIFTS-PLSR with geostatistics to facilitate spatial analysis and regional soil property mapping. Based on a regular and random sampling design, a total of 1170 soil samples were taken in the two regions and measured via midDRIFTS. Then previously generic calibrated and validated midDRIFTS-PLSR models (GIC) were used to predict TC, TIC, TOC and soil texture contents (clay, silt and sand) of the 1170 soil samples. The midDRIFTS-PLSR models accurately predicted all soil properties.. These predicted values were then used for spatial analysis, which showed in the K region moderately weak spatial dependency (Nugget-to-sill ratios > 50%) for soil TC, TIC and TOC with ranges of 30, 18 and 22 km, respectively, whereas very and moderately strong spatial dependencies (ratio $\leq$ 50%) were obtained for sand, silt and clay with ranges of 21, 20 and 28 km, respectively. Soil properties in the SA region showed moderately strong spatial dependency with the range of spatial autocorrelation varying from 8.7 km for sand up to 24 km for TIC. Ordinary kriging resulted in accurate soil property maps with a root mean square standardized error (RMSSE) close to 1, except for sand in K and TIC in SA region. While the new soil organic matter (SOM) map in SA and the texture map in K corresponded very well with of the existing maps, the new SOM map in K and the texture map in SA showed discrepancies as compared to the existing maps. These discrepancies were attributed to very low resolution and to the preparation methods of the existing maps. MidDRIFTS-PLSR provided high quality soil data for integration with geostatistics to create maps of soil carbon and texture with high resolution at the regional scale which are an improvement over the existing maps.

As a further development of midDRIFTS approaches for soil quality assessment, spectral-based indexes for characterizing SOM quality and quantifying carbonate at regional scale were explored. Spectral peak areas were calculated and plotted against biochemical data (C_{mic}, N_{mic}, C_{HWE}, N_{HWE}, C:N ratio, TOC, TIC and TN) of bulk soil across two regions. midDRIFTS peak areas corresponding to SOM functional groups (2930, 1620, 1520 and 1159 cm^{-1}) were assessed to study the composition of SOM. Results of the relative distribution of peaks between two regions showed that the relative peak area at 2930 cm^{-1} was larger in SA than K which was related to soil type and organic C content and quality. In contrast, larger peak areas at 1620, 1520 and 1159 cm^{-1} was obtained for soils from K region. The peak assigned for aliphatic C-H bond (2930 cm^{-1}) was an appropriate index to investigate SOM fractions if the interference of carbonates was taken into consideration. Regression performance obtained between the peak at 2930 cm-1 and SOM fractions (e.g., R^2 = 0.31 for Cmic) increased to R^2 = 0.65 when high carbonate containing samples (total inorganic carbon > 1%) were excluded. A significant nonlinear relationship was found between the ratio of aliphatic C-H to aromatic C=C peaks (1620:2930) and C_{mic}, C_{HWE} and C/N ratio (R^2 = 0.4, 0.56 and 0.5, respectively; $P < 0.05$), but, not with the ratio of more stable to labile C fraction (TOC-C_{HWE})/C_{HWE}. The most accurate spectral index for carbonate was the peak area at 713 cm^{-1} when relating to TIC obtained by Scheibler's method (R^2 = 0.98). The results showed the feasibility of the midDRIFTS-based spectral indexes to be used as an indicator of SOM quality and composition if mineral interference avoided as well as to quantify TIC contents.

In conclusion, it was demonstrated that midDRIFTS-PLSR is a rapid-throughput method for providing high-quality predictions of soil properties to update regional digital soil property mapping by integration with geostatistics. It opens a new possibility to gain high resolution data coverage of soil C and N pools, which is relevant for the application of SOM simulation models on a regional scale.. However, to up-scale the approach for extended geographical areas, further efforts are needed to establish a national level spectral library by considering standardization of sampling, analytical reference analyses and midDRIFT spectroscopy techniques.

8. Zusammenfassung

Veränderte klimatischen Bedingungen und Änderungen in der Landnutzung üben einen erheblichen Einfluss auf ökologische Prozesse in Böden aus. Aus diesem Grund ist eine regelmäßige Überprüfung wichtiger Bodeneigenschaften, wie der Kohlenstoff- (C) und Stickstoff- (N) Vorräte, des Nährstoffhaushalts und des Boden-pHs notwendig um nachhaltige Bearbeitungs- und Pflegemaßnahmen in Agrarökosystemen empfehlen zu können. Eine Charakterisierung von Bodeneigenschaften auf verschiedenen räumlichen Skalen ist jedoch weiterhin schwierig da sie eine intensive GPS-referenzierte Probenentnahme voraussetzt. Ein effizienterer und kostengünstigerer Ansatz ist erforderlich um Bodendatenbanken für räumliche Auswertungen, Überprüfungen von Bodeneigenschaften auf regionaler Ebene und die Kalibrierung computergestützter Bodenmodelle auf regionaler Ebene zu ermöglichen. Der Einsatz der diffusen Reflexions-Fouriertransformations-Infrarotspektroskopie im mittleren Infrarotbereich (midDRIFTS) könnte in diesem Bereich von wichtiger Bedeutung sein. Diese Studie hat das Ziel, den Einsatz von midDRIFTS in Kombination mit einer Regressionsanalyse mit partiellen kleinsten Quadraten (PLSR) als Hoch-Durchsatz Methode für die Quantifikation von Bodeneigenschaften und Bodenqualität zu überprüfen und auf Grundlage dieser Daten die räumliche Verteilung der Bodeneigenschaften zu modellieren und Bodenkarten auf regionaler Ebene zu erstellen. Hierfür wurden zwei Agrarökosysteme herangezogen, das Kraichgau (K) und die Schwäbische Alb (SA), (40 x40 km) in Südwestdeutschland.

Im ersten Teil dieser Studie wurde eine Probenentnahme auf regionaler Ebene durchgeführt um das effizienteste midDRIFTS-PLSR Prognosemodel zu erstellen. Zwei Kalibrationsansätze, (1) die Kreuzvalidierung und (2) die Kalibration mit unabhängiger Validierung wurden auf die Möglichkeit untersucht, gesamt-C (TC), gesamt organisches C (TOC), mineralisches C (TIC), gesamt-N (TN), mineralisches N (N_{min}), das C:N Verhältnis, heißwasserlösliches C und N (C_{HWE}, N_{HWE}), C und N der mikrobiellen Biomasse (C_{mic}, N_{mic}), Boden-pH, Lagerungsdichte und Korngrößenverteilung in 126 Proben der untersuchten Regionen zu bestimmen. Regionsübergreifende Kalibrationsmodelle wurden mit beiden Kalibrationsansätzen anhand der Daten von 84 Proben erstellt, die verbleibenden 42 Proben wurden zur unabhängigen Modellvalidierung verwendet. Außerdem wurde die Kreuzvalidierung regionenspezifischer midDRIFTS-PLSR Modelle mit 42 Proben für die Kalibrierung und 21 Proben für eine zusätzliche unabhängige Validierung durchgeführt. Der Großteil der Bodeneigenschaften, mit Ausnahme von Nmin, dem C/N Verhältnis, pH, Lagerdichte und dem Sandanteil konnte anhand beider Kalibrationsmodelle vorhergesagt werden (residual prediction deviation (RPD) >3 und R^2 >0.9). Die Anwendung der Modelle an „unbekannten“ Proben zeigte, dass die Genauigkeit des regionsübergreifenden Kreuzvalidierungsmodels (GCV) weniger zuverlässig als die des allgemeinen unabhängig validierten Models (GIC) war. Ebenso wiesen die regionenenpezifischen Kreuzvalidierungsmodelle in der Vorhersage „unbekannter“ Proben eine geringere Genauigkeit im Vergleich zum unabhängig validierten regionsübergreifenden Model auf. Insgesamt war das unabhängig validierte Model der robustere Ansatz für die Vorhersage der Eigenschaften „unbekannter“ Proben.

Der zweite Schritt dieser Arbeit war die Integration des midDRIFT-PLSR Models in geostatistische Modelle, um die räumliche Analyse und regionale Bodenkartierungen zu erleichtern. Während einer regelmäßigen, randomisierten Probenentnahme wurden 1170 Bodenproben in den zwei Regionen entnommen und mittels midDRIFT analysiert. Im Anschluss wurde das midDRFIFT-PLSR (GIC) Model angewandt um TC, TIC, TOC und Korngrößenverteilung (Ton, Schluff, Sand) der 1170 Proben zu bestimmen. Die midDRIFT-PLSR Modelle sagten alle Bodeneigenschaften mit ausreichender Genauigkeit vorher. Diese Prognosen wurden dann für die räumliche Analyse verwendet. Für die Region K zeigten TC, TIC und TOC (30, 18 und 22 km) eine geringe räumliche Abhängigkeit (Nugget-to-sill ratios > 50%) und Sand, Schluff und Ton (21, 20 and 28 km) eine starke bis moderate räumliche Abhängigkeit (Nugget-to-sill ratios ≤ 50%). Die Bodeneigenschaften in der Region SA zeigten verhältnismäßig starke räumliche Abhängigkeiten. Die räumliche Autokorrelation variierte zwischen 8.7 km für Sand und 24 km für TIC. Durch geostatistisches Kriging der Daten wurden Karten der Bodeneigenschaften mit einem mittleren quadratischen Standardfehler (RMSSE) von fast 1 erstellt. Ausnahmen waren die Sandgehalte in der Region K und TIC in der Region SA. Die neu erstellten SOM Karten der Region SA und die Texturkarten der Region K wiesen eine hohe Übereinstimmung mit bereits existierenden Bodenkarten auf. Jedoch zeigte die neu erstellte SOM Karte der Region K und die Texturkarte der Region SA deutliche Unterschiede zu den existierenden Bodenkarten auf. Diese Unterschiede beruhen vermutlich auf der geringen Auflösung und der schwachen Datengrundlage für die existierenden Karten. Im Gegensatz dazu stellt midDRIFTS-PLSR qualitativ hochwertige Daten für die Integration in geostatistische Verfahren der C- und Texturkartierung auf regionaler Ebene zur Verfügung, die eine deutliche Verbesserung gegenüber existierender Bodenkarten darstellen.

In einem weiteren Schritt wurde die Weiterentwicklung des midDRIFT Ansatzes im Hinblick auf die Untersuchung der Bodenqualität an Hand spektral-basierter Indexe für die Charakterisierung der SOM Qualität und der C-Quantifizierung auf regionaler Ebene untersucht. Peaks relevanter Spektralbereiche wurden quantifiziert und gegen biochemischen Daten (C_{mic}, N_{mic}, C_{HWE}, N_{HWE}, C:N ratio, TOC, TIC und TN) von Böden der zwei Regionen geplottet. Peak-Bereiche von midDRIFTS die mit funktionellen Gruppen der SOM übereinstimmen (2930, 1620, 1520 and 1159 cm^{-1}) wurden untersucht um die Zusammensetzung der SOM zu bestimmen. Die relative Verteilung der Peaks in Proben der zwei Regionen zeigte, dass der Peak bei 2930 cm^{-1} in der Region SA größer war als der der Region K. Dies steht im Zusammenhang mit dem Bodentyp, dem Gehalt an organischem C und der Bodenqualität. Im Gegensatz dazu wurden bei 1620, 1520 und 1159 cm^{-1} größere Peaks für Böden der Region K ermittelt. Der Peak welcher der aliphatischen C-H Bindung (2930 cm^{-1}) zugewiesen wird, war ein zuverlässiger Index um SOM-Fraktionen zu charakterisieren sofern der Carbonatanteil der Böden in Betracht gezogen wurde. Die Bestimmtheitsmaß der Regressionsanalyse zwischen den Peaks bei 2930 cm^{-1} und den SOM Fraktionen (z.B., $R^2 = 0.31$ for C_{mic}) erhöhte sich auf $R^2 = 0.65$ wenn carbonathaltige Proben von der Regressionsanalyse ausgeschlossen wurden. Eine signifikante, nicht-lineare Korrelation wurde zwischen dem Verhältnis der aliphatischen C-H-Peaks zu den aromatischen C=C-Peaks (1620:2930) und C_{mic}, C_{HWE} sowie dem C/N Verhältnis (R^2 = 0.4, 0.56 und 0.5; $P < 0.05$) festgestellt, jedoch nicht zu den labilen und stabileren C Fraktionen (TOC-C_{HWE})/C_{HWE}. Der genaueste Spektral-Index für Carbonat mit Bezug zu TIC nach Scheibler ($R^2 = 0.98$) war der Peak bei 713 cm^{-1}.

Die Ergebnisse zeigen die Möglichkeiten der Anwendung von midDRIFT-Spektral-Indizes als Indikatoren für SOM Qualität und Zusammensetzung, sowie der Quantifizierung der TIC Gehalte. Diese Studie zeigte dass midDRIFT-PLSR eine Hochdurchsatzmethode für die Erstellung qualitativ hochwertiger Vorhersagen zu Bodeneigenschaften ist, die ein schnelles und genaues Verfahren zur digitalen Bodenkartierung durch geostatistische Verfahren ermöglicht. Das Verfahren eröffnet neue Perspektiven in Hinblick auf die Erstellung kompletter Datensätze zu C und N pools im Boden. Dies ist für die Kalibrierung und Anwendung computergestützter SOM Modelle auf regionaler Ebene von größter Bedeutung. Um diesen Ansatz auf größere Regionen auszuweiten sind jedoch Untersuchungen notwendig um durch die Standardisierung der Probennahme, der analytischen Bestimmungen und der midDRIFT-Spektroskopie Techniken eine nationale Spektral-Bibliothek zu erstellen.