Marie-Isabel Ludwig

Dem Fachbereich VI der Universität Trier
(Geographie/Geowissenschaften)
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
eingereichte Dissertation

Assessing Soil Organic Matter by Combining Non-Invasive (Spectroscopic) and Invasive Methods

Betreuender: Prof. Dr. agr. habil. Sören Thiele-Bruhn

20. Juni 2017

Erstgutachter: Prof. Dr. Sören Thiele-Bruhn

Zweitgutachter: Prof. Dr. Christoph Emmerling

Externer Drittgutachter: Prof. Dr. Michael Vohland

“the real voyage of discovery consists not in seeking new lands but in seeing with new eyes“

Marcel Proust

dedicated to

the soil –

the living, the dead, and the never dead

marie-isabel ludwig

Contents

Abstract ..... **i**

Kurzzusammenfassung ..... **ii**

List of Abbreviations ..... **iii**

Preliminary Remarks ..... **iv**

1 Introduction **1**

2 State of the Art **3**

2.1 Soil Organic Matter (SOM) in Soil Assessment 3

2.1.1 Definition, Composition and Distribution 3

2.1.2 Dynamics, Pools and Assessment Methods 4

2.1.3 Characteristics and Application as a Parameter of Soil Quality 6

2.2 Spectroscopy in Soil Assessment 7

2.2.1 Background and Advantages 7

2.2.2 Types and Procedures 8

2.2.3 Vis-NIR Spectroscopy Applications and Challenges 10

2.3 Resulting Objectives and Hypotheses 11

2.4 References (Chapters 1 & 2) 13

3 Using variable selection and wavelets to exploit the full potential of visible–near infrared spectra for predicting soil properties **19**

4 Spatial and temporal changes of the soil microbial community and labile organic matter fractions affect the Vis-NIR spectroscopic quantification of organic matter in arable topsoils **43**

5 Microbial contribution to soil organic matter (SOM) quantity and quality in density fractions of temperate arable soils **70**

6 Synthesis and Conclusions **97**

7 Summary and Outlook **102**

7.1 References (Chapters 6 & 7) 105

Selbstständigkeitserklärung ..... **v**

Danksagung ..... **vi**

Curriculum vitae ..... **viii**

Abstract

Soil organic matter (SOM) and its contained carbon and nutrients drive fundamental nano- to global-scale biogeochemical processes and influence carbon-climate feedbacks. The current challenge of ecosystem assessment is to identify the significant drivers of SOM formation and persistence. To this effect non-invasive methods are required that can act on a large-scale to assess these SOM pools as bioindicators of soil quality and to include them in the global soil monitoring of earth observation programs. Remote sensing by Vis-NIR spectroscopy has been shown to be an effective technique of large-scale soil monitoring. However, its potential especially regarding the assessment of biological and microbial soil parameters still largely remains undefined. Therefore, the aim of this research was to investigate temperate arable topsoils for a qualitative and quantitative assessment of SOM pools by applying diverse methods and sites with a varying spatio-temporal resolution. On the basis of the data obtained, from laboratory experiments and field data, the potential of Vis-NIR spectroscopy for the assessment of selected SOM parameters was evaluated and to furthermore improve predictive models by transformation of spectral data.

Therefore, a comprehensive local inventory of soil spectra and associated soil parameters of a heterogeneous soil region in SW-Germany was established that presented a diverse and broad range of site-specific values. On this basis, data processing was optimized and from there the accuracy of common spectrally monitored SOM parameters (SOC, N) from soil reflectance spectra of hyperspectral laboratory measurements could be improved. This was reached by pre-processing of spectral data using variable selection and wavelet transformation. Then an additional data set comprising microbial SOM parameter and in situ data was studied to evaluate the predictive capacity of soil spectra. The spectral potential of SOM assessment of labile and biological or microbial SOM parameter such as microbial biomass carbon (MBC), dissolved organic carbon (DOC), hot water extractable carbon (HWEC), chlorophyll α (Chl α) and phospholipid fatty acids (PLFAs) was examined. A medium prediction accuracy providing the possibility to distinguish between high and low values was obtained for MBC and DOC depending on soil depth and season. In contrast, predictions were not successful for OC, $PLFA_{tot}$, and PLFA markers addressing algae, bacteria and fungi. Highest prediction accuracy was achieved for surface (0-1cm soil depth) changes induced by green soil algae, which is assumed to trigger the modelling of MBC. Instead, the total SOM content, measured as total OC, could not be predicted, which was largely attributed to microbial interference. Hence, the highly dynamic microbial SOM pools at the soil surface are assumed not to be representative and to interfere and impede accurate estimations of total bulk SOM content. The optimization strategies to evaluate

spectral data were successfully applied to evaluate an extended data set of soil parameters. Notwithstanding the predictve accuracy was only slightly improved by pre-processing of the situ spectral data.

The approach to resolve the origin of SOM (plant vs. microbial input) and its contribution to SOM quantity and quality showed that microbial necromass and soil surface algae can be significant sources for SOM formation and persistence. Overall, results indicated that microbial contribution to temperate arable topsoil OM is higher than generally assumed. The microbial contribution to SOM affected SOM quantity and in particular the mineral associated OM and the quality of SOM with microbial derived carbohydrates contributing to SOM stability. An exact quantification of the contribution of SOM pools to soil quality and its estimation by use of spectral data cannot finally be determined and requires further studies.

Kurzzusammenfassung

Die organische Bodensubstanz (OBS) ist eine fundamentale Steuergröße aller biogeochemischen Prozesse und steht in engem Zusammenhang zu Kohlenstoffkreisläufen und globalem Klima. Die derzeitige Herausforderung der Ökosystemforschung ist die Identifizierung der für die Bodenqualität relevanten Bioindikatoren und deren Erfassung mit Methoden, die eine nachhaltige Nutzung der OBS in großem Maßstab überwachen und damit zu globalen Erderkundungsprogrammen beitragen können.

Die fernerkundliche Technik der Vis-NIR Spektroskopie ist eine bewährte Methode für die Beurteilung und das Monitoring von Böden, wobei ihr Potential bezüglich der Erfassung biologischer und mikrobieller Bodenparameter bisher umstritten ist. Das Ziel der vorgestellten Arbeit war die quantitative und qualitative Untersuchung der OBS von Ackeroberböden mit unterschiedlichen Methoden und variierender raumzeitlicher Auflösung sowie die anschließende Bewertung des Potentials non-invasiver, spektroskopischer Methoden zur Erfassung ausgewählter Parameter dieser OBS.

Dafür wurde zunächst eine umfassende lokale Datenbank aus chemischen, physikalischen und biologischen Bodenparametern und dazugehörigen Bodenspektren einer sehr heterogenen geologischen Region mit gemäßigten Klima im Südwesten Deutschlands erstellt. Auf dieser Grundlage wurde dann das Potential der Bodenspektroskopie zur Erfassung und Schätzung von Feld- und Geländedaten ausgewählter OBS Parameter untersucht. Zusätzlich wurde das Optimierungspotential der Vorhersagemodelle durch statistische Vorverarbeitung der spektralen Daten getestet.

Die Güte der Vorhersagewahrscheinlichkeit gebräuchlicher fernerkundlicher Bodenparameter (OC, N) konnte für im Labor erhobene Hyperspektralmessungen durch statistische Optimierungstechniken wie Variablenselektion und Wavelet-Transformation verbessert werden. Ein zusätzliches Datenset mit mikrobiellen/labilen OBS Parametern und Felddaten wurde untersucht um zu beurteilen, ob Bodenspektren zur Vorhersage genutzt werden können. Hierzu wurden mikrobieller Kohlenstoff (MBC), gelöster organischer Kohlenstoff (DOC), heißwasserlöslicher Kohlenstoff (HWLC), Chlorophyll α (Chl α) und Phospholipid-Fettsäuren (PLFAs) herangezogen. Für MBC und DOC konnte abhängig von Tiefe und Jahreszeit eine mittlere Güte der Vorhersagewahrscheinlichkeit erreicht werden, wobei zwischen hohen und niedrigen Konzentration unterschieden werden konnte. Vorhersagen für OC und PLFAs (Gesamt-PLFA-Gehalt sowie die mikrobiellen Gruppen der Bakterien, Pilze und Algen) waren nicht möglich. Die beste Prognosewahrscheinlichkeit konnte für das Chlorophyll der Grünalgen an der Bodenoberfläche (0-1cm Bodentiefe) erzielt

werden, welches durch Korrelation mit MBC vermutlich auch für dessen gute Vorhersagewahrscheinlichkeit verantwortlich war. Schätzungen des Gesamtgehaltes der OBS, abgeleitet durch OC, waren hingegen nicht möglich, was der hohen Dynamik der mikrobiellen und labilen OBS Parameter an der Bodenoberfläche zuzuschreiben ist. Das schränkt die Repräsentativität der spektralen Messung der Bodenoberfläche zeitlich ein. Die statistische Optimierungstechnik der Variablenselektion konnte für die Felddaten nur zu einer geringen Verbesserung der Vorhersagemodelle führen.

Die Untersuchung zur Herkunft der organischen Bestandteile und ihrer Auswirkungen auf die Quantität und Qualität der OBS konnte die mikrobielle Nekromasse und die Gruppe der Bodenalgen als zwei mögliche weitere signifikante Quellen für die Entstehung und Beständigkeit der OBS identifizieren. Insgesamt wird der mikrobielle Beitrag zur OBS höher als gemeinhin angenommen eingestuft. Der Einfluss mikrobieller Bestandteile konnte für die OBS Menge, speziell in der mineralassoziierten Fraktion der OBS in Ackeroberböden, sowie für die OBS Qualität hinsichtlich der Korrelation von mikrobiellen Kohlenhydraten und OBS Stabilität gezeigt werden. Die genaue Quantifizierung dieser OBS Parameter und ihre Bedeutung für die OBS Dynamik sowie ihre Prognostizierbarkeit mittels spektroskopischer Methoden ist noch nicht vollständig geklärt. Für eine abschließende Beurteilung sind deshalb weitere Studien notwendig.

List of Abbreviations

BC	Black Carbon
CARS	Competitive Adaptive Reweighted Sampling
CEC_{eff}; CEC_{pot}	Cation Exchange Capacity (effective; potential)
Chl α	Chlorophyll α
CO_2	Carbon Dioxide
DOC	Dissolved Organic Carbon
FAO	Food and Agriculture Organization of the United Nations
HWEC	Hot Water Extractable Carbon
ITPS	Intergovernmental Technical Panel on Soils
IRENA	International Renewable Energy Agency
K	Potassium
MBC	Microbial Biomass Carbon
N	Nitrogen
Nano-SIMS	Nano-Scale Secondary Ion Mass Spectrometry
NEXAFS	Near-Edge X-ray Adsorption Fine Structures
NIR	Near Infrared region (700–1100 nm)
(^{13}C-) NMR	Solid-State ^{13}C Nuclear Magnetic Resonance Spectrometry
OC	Organic Carbon
P	Phosphorus
PLFAs	Phospholipid Fatty Acids
PLSR	Partial Least Squares Regression
POM (fPOM; oPOM)	Particulate organic matter (free; occluded)
PyFIMS	Pyrolysis-field ionization mass spectrometry
S	Sulfur
SDGs	Sustainable Development Goals
SEM	Scanning Electron Microscopy
SOC	Soil Organic Carbon
SOM	Soil Organic Matter
SPT	Sodium Polytungstate (Na6(H2W12O40))
STXM	Scanning X-ray Transmission Microscopy
SWIR	Shortwave Infrared Region (1100–2500 nm)
TRFLP	Terminal Restriction Fragment Length Polymorphism
UN	United Nations
VIS	Visible Region (400–700 nm)
WC	Water Content

Preliminary Remarks

This PhD-thesis was carried out in the interdisciplinary research project “SomScapes” of the Department of Soil Science, University of Trier, and the Department of Geoinformatics and Remote Sensing, University of Leipzig. Funding of this project was provided by the DFG (SOMScapeS, TH678/12-1 and Priority Program 1315, “Mi598/2-2”). This cumulative PhD thesis is a collection of already published papers in peer-reviewed journals (Chapters 3 and 5) and a submitted paper (Chapter 4). All results were presented to a broad scientific community at national and international congresses. Journal contributions and presentations are specified below.

Author contributions:

Dipl.-Umweltwiss. Marie-Isabel Ludwig developed the study design of all conducted studies of the PhD-thesis with scientific advice of Prof. Dr. Sören Thiele-Bruhn and Prof. Dr. Christoph Emmerling of the department of soil science at Trier University. The comprehensive field- and lab-work from 2010 to 2013 was conducted or supervised by Marie-Isabel Ludwig.

The Pyrolysis-field ionization mass spectrometry (Py-FIMS) measurements and data processing (Chapter 5) were realized by Prof. Dr. Peter Leinweber and Kai-Uwe Eckhardt from the University of Rostock. Dr. Anja Miltner and Dr. Jan Achtenhagen from the UFZ Leipzig provided analysis by scanning electron microscopy (SEM) and help evaluating the study results (Chapter 5).

Prof. Dr. Michael Vohland, Prof. Dr. Thomas Udelhoven, Christian Bossung and Monika Harbich supervised the remote sensing work package of the interdisciplinary project and contributed significantly to study design and evaluation. Prof. Dr. Michael Vohland conducted and published feasibility analysis to exploit the full potential of Vis-NIRS for predicting soil variables (Chapter 4).

Christian Bossung’s collaboration enabled the intensive cross-disciplinary field work of our one-year soil monitoring and Prof. Dr. Thomas Udelhoven helped significantly with data processing and statistics (Chapter 4).

Marie-Isabel Ludwig wrote the manuscripts (Chapters 4 and 5) and contributed soil analysis and interpretation of soil analysis results to the manuscript of Vohland et al. (Chapter 3).

List of Publications in peer-reviewed Journals

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

Ludwig, M., Achtenhagen, J., Miltner, A., Eckhardt, K.-U., Leinweber, P., Emmerling, C., Thiele-Bruhn, S. (2015): Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils. Soil Biology and Biochemistry, 81, p.311-322

Vohland, M., Harbich, M., Ludwig, M., Emmerling, C., Thiele-Bruhn, S. (2016) Quantification of soil variables in a heterogeneous soil region with VIS-NIR-SWIR data using different statistical sampling and modeling strategies. IEEE J Selected Topics Appl Earth Observ Remote Sens 9(9), p. 4011-4021

Vohland, M., Ludwig, M., Harbich, M., Emmerling, C., Thiele-Bruhn, S. (2016) Using variable selection and wavelets to exploit the full potential of visible-near infrared spectra for predicting soil properties. J Near Infrared Spectrosc 24(3), p.255-269

Ludwig, M., Harbich, M., Vohland, M., Emmerling, C., Thiele-Bruhn, S. (2017): Spatio-temporal changes of soil biofilms and their impact on VIS-NIR spectroscopy. European Journal of Soil Science (submitted)

List of Presentations

Ludwig, M., Harbich, M., Vohland, M., Emmerling, C., Thiele-Bruhn, S. (2011): Evaluation of soil sample pretreatment for VIS-NIR spectroscopy. Poster DBG Jahrestagung, Berlin, September 2011

Ludwig, M., Hens, S., Leinweber, P., Thiele-Bruhn, S. (2012): Quantitative and qualitative characteristics of SOM-fractions obtained from chemical and density fractionation. Vortrag EUROSOIL 2012 Bari, Italien, Juli 2012

Ludwig, M., Emmerling, C., Bossung, C., Vohland, M., Thiele-Bruhn, S. (2012): Spatio-temporal changes of soil algae and their impact on VIS-NIR spectroscopy. Posterbeitrag EUROSOIL 2012 in Bari, Italien, Juli 2012 Ausgezeichnet mit dem „Best Poster Award Eurosoil 2012“ des Journals Biology and Fertility of Soils

Ludwig, M., Harbich, M., Vohland, M., Emmerling, C., Thiele-Bruhn, S. (2013): The potential of VIS-NIR spectroscopy to determine spatio-temporal changes of soil biofilms. Vortrag European Microbiology Congress FEMS 2013, in Leipzig, Juli 2013

Ludwig, M., Harbich, M., Vohland, M., Emmerling, C., Thiele-Bruhn, S. (2013): Saisonale und vertikale Dynamik bodenmikrobieller Gemeinschaften und ihre Erfassbarkeit durch VIS-NIR-Spektroskopie. Poster DBG Jahrestagung, Rostock, September 2013

Ludwig, M., Harbich, M., Vohland, M., Emmerling, C., Thiele-Bruhn, S. (2013): Spatio-temporal changes of soil algae and their impact on VIS-NIR spectroscopy. Vortrag 20th International Conference on Environmental Indicators, in Trier, September 2013

Ludwig, M., Harbich, M., Vohland, M., Emmerling, C., Thiele-Bruhn, S. (2014): Changes of soil biofilms as a hotspot of SOM formation and their determination by VIS-NIR spectroscopy. Vortrag DFG Workshop, München, April 2014

Ludwig, M., Pagel, H., Poll, C., Kandeler, E., Emmerling, C., Thiele-Bruhn, S. (2015): Soil algae as a source of soil organic matter in arable topsoils. Poster DBG Jahrestagung, München, Oktober 2015

Ludwig, M., Pagel, H., Poll, C., Kandeler, E., Emmerling, C., Thiele-Bruhn, S. (2016): Soil algae as a source of soil organic matter in arable topsoils. Poster GfÖ-Workshop Soil Food Webs, Berlin, Februar 2016.

1 Introduction

Soils are the junction of the water-energy-food nexus, providing ecosystem services on which humanity depends (MEA, 2005; Müller and Weigelt, 2013). Soils are the most significant non-renewable geo-resource, functioning not only in the production of food and fiber but also in the maintenance of global sustainability and environmental balance (Muscolo et al., 2015). This balance is threatened. Global change and human activities are already causing unacceptable environmental change by overstepping planetary boundaries (Rockström et al., 2009). The main drivers of global soil change are climate change and demographic factors (FAO and ITPS, 2015; Stockmann et al., 2015). Soil threats resulting from the synergy of these drivers include global warming, soil organic matter decline, industrial and nuclear pollution, soil compaction, soil erosion, soil sealing, soil salinization, the use of genetically modified organisms (GMOs) in agriculture, land use change and intensive human exploitation (FAO and ITPS, 2015; Orgiazzi et al., 2016). Furthermore, the forecasted increase in world population of up 9.7 billion in 2050 (UN, 2015) will require an additional food production of 60-100% as well as energy and clean water demands of additional 100% and at least 55% compared to today (Valin et al., 2014; IRENA, 2015). The challenge facing a growing population is the increased and intensified use of the ecosystem services provided by soils (Blum, 2005). The retention of soil organic matter (SOM), and especially its ‘backbone’, the soil organic carbon (SOC) is fundamental for both, to sustain food security and mitigate the effects of climate change (Lal, 2004; Lal, 2010). The global soil carbon pool is approximately three times larger than the contemporary atmospheric pool (Fig. 1), therefore even minor changes to its integrity may have major implications for atmospheric CO2 concentrations (Erhagen et al., 2013).That makes soil a tipping element in global change (Wall, 2007; Lenton et al., 2008) and its management crucial to remain within a safe operating space for humanity and maintain global sustainability (Rockström et al., 2009; Weigelt et al., 2014). Soil management is therefore identified as a key element in the realization of the Sustainable Development Goals (SDGs) recently adopted by the United Nations (Keesstra et al., 2016). Towards a sustainable soil management, soil monitoring is essential for the early detection of changes in soil quality (Morvan et al., 2008). The total SOM content has the most widely recognized influence on soil quality (Muscolo et al., 2015) and total SOM content or SOM fractions are used to indicate changes in soil quality due to soil management (Pulleman et al., 2000; Haynes, 2005; Leifeld and Kögel-Knabner, 2005). To monitor the global soil status, there is a growing demand of regional to worldwide databases providing spatio-temporal high-resolution data (Grunwald et al., 2011). Hence, assessment methods are needed that work large-scale, but at the same time are selective

and sensitive enough to determine and quantify different pools of SOM quality. This is a balancing act between on the one hand to “zoom in” into the soil ecosystem in order to evaluate soil quality, on the other hand “zoom out” in order to be able to assess soil quality on a landscape-scale to global-scale level. Large-scale, sensitive assessment methods to identify the different SOM pools are currently under development. Soil spectroscopy is a promising approach, because over the past two decades it could be shown that the VIS (400–700 nm), NIR (700–1100 nm), and SWIR (1100–2500 nm) spectral regions to serve as powerful tools for recognizing soils qualitatively and quantitatively (Viscarra Rossel et al., 2006; Ben-Dor et al., 2009). It also represents a quick and cheap scanning method that can be used on large-scales and as an alternative to more expensive standard soil analysis procedures (Vohland et al., 2014). But its potential to assess especially the soil biological or microbiological properties yet remains unclear.

Accordingly, our motivation was to zoom into arable soils and assess SOM pools quantitatively and qualitatively by a huge variety of methods and sites in a varying spatio-temporal resolution. Moreover, we wanted to exploit the potential of Vis-NIR spectroscopy to zoom out the soil ecosystem and assess selected SOM parameters in situ and non-invasively, extending the focus to also microbiological SOM pools.

2 State of the Art

2.1 Soil Organic Matter (SOM) in Soil Assessment

2.1.1 Definition, Composition and Distribution

SOM is derived from two main groups of input materials: plant residues (and exudates) and microbial residues (and exudates) (Kogel-Knabner, 2002). The total SOM content is the balance of inputs versus losses via such pathways as mineralization and leaching (Campbell and Paustian, 2015) and represents a continuum of progressively decomposing organic compounds (Lehmann and Kleber, 2015).

The main constituents of SOM are altered and relatively unaltered aliphatic polymers, polysaccharides (e.g., cellulose), lignin and lignin degradation products, fats, proteins, pectines and cutins (Schaumann, 2006). The carbon content of SOM is about 50%, which enables to account for the total SOM content by multiplying the total soil organic carbon (SOC) content of the soil with 2,0. Many other studies referring to a SOC content of 58% of SOM multiply with 1,724, respectively (Blume et al., 2010). In most soils, SOM is only a small percentage of the soil mass, for example ranging from minima of <1% to maxima of 8% to 9% in mineral temperate soils under agricultural use (Campbell and Paustian, 2015). Depending on land use, SOM contents of temperate European soils are on average about 4% for arable soils, 6% for grassland soils and about 10% for forest soils (Hiederer, 2009). Almost half of the terrestrial OM is sequestered in the first meter of the soil (Fig. 1).

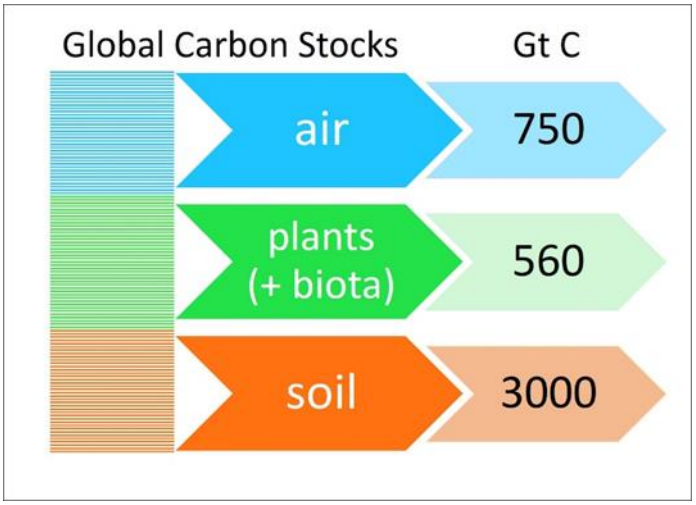


Figure 1: The global distribution of carbon stocks given in Gigatons carbon (equal to Petagram Pg). The main part is terrestrially sequestered, while up to 750 Gt C of this part are located in the first 30cm and up to 1400 Gt C in the first 1m of the soil, respectively (data from Stockmann et al., 2013; Lal, 2015)

2.1.2 Dynamics, Pools and Assessment Methods

The controlling factors for the formation and persistence of SOM were assumed in the first instance to be the amount of plant litter, its molecular composition and its properties (Kogel-Knabner, 2002). However, the rate of OM decomposition in a soil seems strongly determined by both the indigenous microbial community and the environmental conditions (e.g., temperature, pH, soil water capacity, etc.), which govern the biogeochemical activities of the microorganisms (Schmidt et al., 2011; Schimel and Schaeffer, 2012). There is controversy to which extent the initial litter chemistry is affecting the microbial decomposition (Kleber, 2010a; Kleber, 2010b; von Lützow and Kögel-Knabner, 2010). Nevertheless, the crucial role of microorganisms in the formation and persistence of SOM is unquestioned.

The microbial biomass is assumed to be a comparably small SOM pool of about 50-2000 µg C^{-1} g soil (Blagodatskaya and Kuzyakov, 2013) which on average represents 2-3% of the total organic C content (Anderson and Domsch, 2010). In parallel, indications of seriously underestimated microbial biomass contributions to SOM can be found, ranging up to >50% of total extractable SOM and >80% of soil N (Simpson et al., 2007). Yet it has been generally accepted that the microbial contribution to SOM is almost completely metabolic and contributed by the living biomass, i.e. by the active or potentially active microorganisms. The dead microorganisms are mostly regarded as a very dynamic and inconsistent fraction due to permanent re-utilization of microbial C (Blagodatskaya and Kuzyakov, 2013; Bradford et al., 2013). However, several studies emphasized the underestimation of microbial necromass as a SOM source (Simpson et al., 2007; Liang et al., 2010; Liang and Balser, 2011). Recent studies showed that microbial necromass directly contributes to non-living SOM and may provide a major part of the molecular SOM structures of up to 40% recovered in non-living SOM (Kindler et al., 2009; Miltner et al., 2009; Miltner et al., 2012; Schurig et al., 2013). Therefore, it is necessary to partition and analyse SOM pools in more detail than just divide it into living and non-living SOM. Approaches for that can be manifold.

Biological approaches in this context often aim at investigating the microbiology of soils and result in pools that elucidate the microbial contribution by the microbial biomass (MBC) to total SOM, the microbial community composition profiles by fingerprinting (e.g. TRFLP) or biomarker (e.g. phospholipid fatty acid analysis) techniques. The microbial performance can be derived from parameter like basal respiration (RB), ecophysiological ratios like C/N or MBC/OC and the metabolic quotient (qCO_2). The new “multi-omics” approach aims at linking soil biodiversity and functioning due to analysis of microbial genomes in comparison to their proteome, transcriptome and metabolome (Roume et al., 2015; Jansson and Baker, 2016).

Chemical approaches mostly result in total SOM elemental composition analysis by total contents of foremost carbon (TC), nitrogen (TN), phosphorus (P), potassium (K), sulfur (S)

and others. The chemical characterization of different SOM pools or fractions can for instance describe relevant pools of carbon like the soil organic and inorganic carbon (SOC; IC) or the the continuum of black carbon (BC). The characterization of the nutrient status and its bioavailability is possible by the pool of effective cation exchange capacity (CECeff of K, Na, Fe, Mg, Ca, Al, Mn; plant-available K and P) and parameter like pH and base saturation. An example for more specific fractions would be the extractation of pedogenic oxides (such as cristalline Mn_d; Fe_d or oxalate-extractable Mn_{ox}; Fe_{ox}).

Analyses of soil spectra from sensitive mass spectrometric methods are also frequently used to describe the molecular composition and properties of different SOM or SOM pools. Pyrolysis-field ionization mass spectrometry (PyFIMS) can be used to derive the quantity, polydispersity and thermostability of each main SOM compound class including carbohydrates, phenols, lignin, lipids, alkylaromatics, N-containing compounds, sterols, suberin, peptides and free fatty acids. Other powerful tools to analyze SOM pools can be the solid-state 13C nuclear magnetic resonance spectrometry (13C-NMR) (Knicker et al., 2005; Leifeld et al., 2012) or the nano-scale secondary ion mass spectrometry (Nano-SIMS) (Mueller et al., 2013; Schaeffer et al., 2015). NMR allows the characterization of insoluble heterogeneous mixtures which are widely excluded from analysis by means of wet chemical approaches (Knicker et al., 2005), while Nano-SIMS can be used to visualize and characterize mineral-associated soil organic matter at the submicrometer scale (Mueller et al., 2012; Remusat et al., 2012)

Physical approaches to describe SOM vary from traditional field methods to high-end optical or spectral methods. The analysis of soil type, soil texture or soil color give rough estimates of total SOM contents. Fractionation procedures by contrast can directly generate physical SOM pools, differed through their density. For example, the density fractionation with Sodium polytungstate (SPT; Na6(H2W12O40)) is generally used to generate three or more density fractions, including free and occluded particulate organic matter (fPOM and oPOM of different densities) and mineral associated organic matter (MOM > 2.0 g cm-3). To further differentiate the origin of compounds in the SOM or SOM fractions with non-invasive techniques, visualizing techniques like scanning electron microscopy (SEM) (Miltner et al., 2012) or scanning X-ray transmission microscopy (STXM) can be used (Schaeffer et al., 2015). The latter is even more promising when combined with near-edge X-ray adsorption fine structures (NEXAFS), which then can visualize filament-like microbial agglomerates structures on the micro- and nano-scale and detect the single origins (bacterial or fungal) of the involved cells (Solomon et al., 2012a; Solomon et al., 2012b).

2.1.3 Characteristics and Application as a Parameter of Soil Quality

SOM has significant influence on all soil functions and plays a central role in the global carbon cycle (Blume et al., 2010). SOM is an important sorbent of inorganic and organic compounds due to its negative load and its relatively high specific surface, it increases the cation exchange capacity (CEC), is carbon and energy source for soil biota, is nutrient supplier for plants due to mineralization by soil microorganisms and contributes to soil formation and aggregation (Blume et al., 2010). Despite its great significance, there is controversy about its usefulness or relevance by means of detecting just the total SOM content as a parameter of soil quality.

A variety of definitions for the term soil quality has been proposed (Parr et al., 1992; Doran and Parkin, 1994; Karlen et al., 1997), but still there is an ongoing debate about how quality can be exactly defined in terms of soil. Whilst the majority of countries have criteria to evaluate the quality of the air and water, the same does not occur for the quality of the soil (Muscolo et al., 2015). Quality assessment of air and water usually consists of the analysis of certain contaminants. Soils have no well-defined ecosystem thresholds (Loveland and Webb, 2003) or show just one ideal environmental state. Instead, they show an almost limitless number of environmental scenarios resulting from their complex interactions and dynamic chemical, physical, biological and ecological properties (Bastida et al., 2008). Some authors therefore state that soil quality is indefinable (Sojka and Upchurch, 1999) and propose to instead of further attempting to define the term, we should speak of the “concept of soil quality” (Bastida et al., 2008). Traditionally, soil quality was associated with soil productivity, resulting in yield-orientated monitoring of parameters such as soil texture, infiltration rate, aggregation, bulk density, pH, forms of C and N, microbial biomass and available nutrients (Parr et al., 1992; Karlen et al., 1997). Recently soil quality has more been defined in terms of sustainability (Tóth et al., 2007), as the capacity of the soil to absorb, store and recycle water, minerals and energy in such a way that the production of the crops can be maximized and environmental degradation minimized (Muscolo et al., 2015). This reflects the crucial role of soils in the water-energy-food nexus and illustrates the need of finding parameters that work on a meta-level of information, representing the multifunctionality of soils (Bradford et al., 2014). Hence, to meet the requirements of a sustainable management approach it is necessary to not only monitor random soil parameter but to find relevant bioindicators for monitoring soil quality. Bioindicators are measurable surrogates for environmental end points that are in themselves too complex to assess or too difficult to interpret in terms of ecological significance (Pulleman et al., 2012). Indicators for agro-ecosystems, either biological, physical or chemical, give information about the state, trends and the seriousness of the situation by complex interactions between agriculture and environment (COM, 2000). Generally, soil quality has been related to soil parameters such as the total content of SOM,

total C and N (and C/N ratio) as well as microbial biomass and activity (Karlen et al., 1997; Murphy et al., 2011; Laishram et al., 2012). However, the usefulness of SOM total content as a general indicator is not clear because of its insensitivity in assessing soil quality changes in the short-term (Muscolo et al., 2015). Recently, the living component of SOM and its species and functional diversity are gaining significance as a bioindicators of soil quality (Velasquez et al., 2007; Ritz et al., 2009; Pulleman et al., 2012), leading to a huge output of studies throughout the last decade (Bastida et al., 2008; Havlicek, 2012; Stone et al., 2016). This is due to the assumption that soil functioning to a large extent depends upon soil biodiversity (Nielsen et al., 2011; Griffiths et al., 2016). Accordingly, there has developed a general agreement that the soil biochemical, microbiological and biological properties are more suitable than physical or chemical properties for the purpose of estimating alterations in soil quality (Paz-Ferreiro and Fu, 2016).

2.2 Spectroscopy in Soil Assessment

2.2.1 Background and Advantages

Since soil management is crucial for the food-energy-water nexus, the acquisition of larger amounts of accurate soil data is more important now than ever before (Viscarra Rossel et al., 2006). Efficient and standardized large-scale monitoring of soil quality is thus required (Cécillon et al., 2009). Spectroscopic techniques like mass spectroscopy (MS), nuclear magnetic resonance (NMR), visible (Vis), near infrared (NIR) and mid infrared (MIR) spectroscopy are being considered as adequate alternatives to extend or replace conventional laboratory methods of soil analysis (Viscarra Rossel et al., 2006; Croft et al., 2012). Especially remote sensing by Vis-NIR spectroscopy has been shown to be an effective alternative, encompassing a larger proportion of the landscape by being a non-destructive scanning method that allows the preservation of the basic integrity of the soil system (Viscarra Rossel et al., 2006; Archer et al., 2014). Furthermore, the use of spectroscopy offers the potential to characterize various soil parameter simultaneously by a single spectrum (Askari et al., 2015), thus requiring less time and expensive sampling procedures and laboratory analysis (Vohland et al., 2014). This advantage of using spectroscopy becomes especially evident when focusing on soil quality, which requires large amounts (in terms of sample numbers and number of parameters measured) of soil data across management systems (Askari et al., 2015).

2.2.2 Types and Procedures

For the acquisition of spectral reflectance data, different sensor types can be mounted on either airborne or spaceborne platforms (remote sensing) or ground-based (in situ) measurements are performed on-site by portable field-spectrometers (proximal sensing) (Viscarra Rossel et al., 2011; Croft et al., 2012; Archer et al., 2014). The recent technical development from multispectral to hyperspectral sensors provides new opportunities for environmental assessment. Already successfully used in many disciplines, including geology and marine and vegetative studies, this technique holds new capabilities and especially opens new frontiers in soil applications (Ben-Dor, 2011). Hyperspectroscopy is characterized by many additional spectral channels, which expand the spectral information of the sensed material to be for the first time analysed in terms of quantity (Ben-Dor, 2011). Yet, these sensors are available for laboratory analyses, as portable devices for the field and mounted onboard an aircraft. Hyperspectral satellites are planned to be launched soon.

Soil reflectance varies according to biogeochemical factors, such as soil mineralogy, soil moisture and SOM content, influenced by physical factors such as surface roughness, particle size and micro shadows (Croft et al., 2012). This is due to the principle of reflectance spectroscopy, which generally relies on the reflectance of chemical substances depending on their individual molecular bond vibration and related incidental radiation absorption (Rodionov et al., 2016). Spectral absorption features arise from the vibrational stretching and bending of structural groups of atoms and the electronic transitions of atoms (Croft et al., 2012). Fundamental bond vibrations (stretching and bending) related to soil constituents occur in the MIR region (2500 - 25,000 nm) while weaker and broader signals from vibration overtones and combination bands occur in the Vis (400 – 780 nm) and NIR (780 - 2500 nm) region (Croft et al., 2012). These broad bands result from many overlapping peaks (referred to as ‘multicollinearity’), which make the absorption wavelengths in the NIR less specific than in the MIR. Vis and NIR region differ in their main molecular vibrational frequencies. These frequencies can occur for relatively light vibrating C-H, N-H, and O-H groups containing hydrogen, as well groups of “heavier” atoms C-O, C-N, N-O, C-C in organic materials as well as Al-O, Fe-O, and Si-O in minerals (Soriano-Disla et al., 2013). The NIR is dominated by overtone and combination vibrations of the relatively light atoms involving hydrogen, while electronic transitions rather absorb in the Vis region. The MIR region absorbs a combination of frequencies derived from the light vibrating groups and heavier atomic groups (Soriano-Disla et al., 2013). But the conversion of the signal detected by the remote sensor into a meaningful soil parameter may involve considerable uncertainty (Cécillon et al., 2009; Archer et al., 2014) because both Vis-NIR and MIR spectra provide large sets of predictor variables that normally are strongly collinear and noisy (Vohland et al., 2014). Hence the assignment of the signals to certain soil constituents is difficult, particularly for functional groups in soil

organic matter (Rodionov et al., 2016). The lack of specificity makes it necessary to correlate spectral features with soil parameters using multivariate calibration procedures. Partial least squares regression (PLSR) is probably the most popular approach and manages big, complex datasets such as Vis-NIR spectra (Stenberg et al., 2010; Vohland et al., 2014; Rodionov et al., 2016). PLSR can be used to reduce the dimensionality of variables by projecting the spectral data into a low-dimensional space formed by a set of orthogonal latent variables (Vohland et al., 2014). These variables are then used to develop prediction models for soil parameter. To further improve the prediction ability (accuracy) of these models, a wide range of pre-processing techniques can be applied (Askari et al., 2015), e.g. by selecting the most informative spectral variables instead of using the full spectrum. This excludes further complexity and noise which in total increases the robustness of the whole model (Vohland et al., 2014). Several approaches exist for selecting such an optimal subset of spectral wavelengths or regions in the model calibration process such as, for example, competitive adaptive reweighted sampling (CARS) or variable importance for projection (VIP)(Vohland et al., 2014).

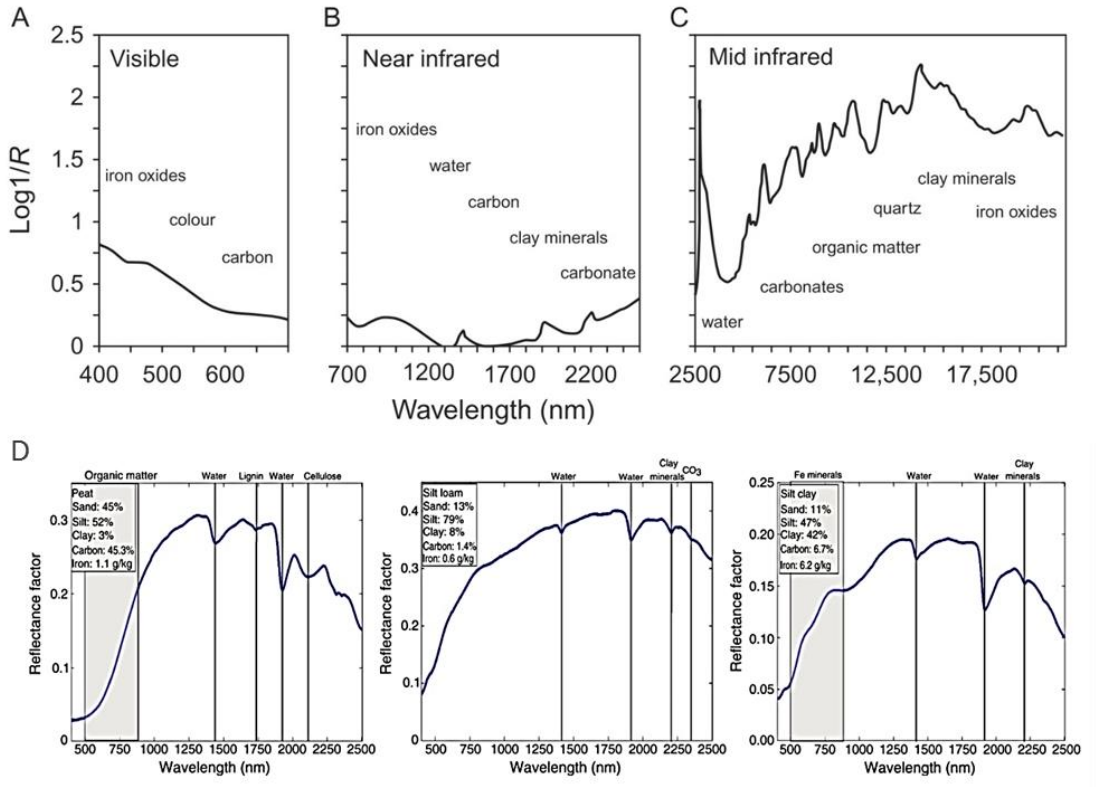


Figure 2: Typical soil spectrum in the (A) visible, (B) near infrared and (C) mid infrared portions of the EM spectrum and (D) reflectance spectra from three different soil types, showing examples of spectral absorption features with various associated soil biochemical parameters. Broad absorption features can be seen in the visible region, associated with iron oxides and soil organic matter (from Viscarra Rossel et al., 2011; Croft et al., 2012)

2.2.3 Vis-NIR Spectroscopy Applications and Challenges

The most common soil parameter investigated by Vis-NIR spectroscopy are the soil water content (WC), soil organic carbon (SOC), total nitrogen (N), total soil organic matter (SOM), metals (trace and heavy metals), total or exchangeable nutrients, cation exchange capacity (CEC), soil pH and physical soil parameter such as soil texture (proportions of clay, silt and sand) and bulk density (Cécillon et al., 2009; Stenberg et al., 2010; Bellon-Maurel and McBratney, 2011; Soriano-Disla et al., 2013). Far fewer studies are available for the spectral assessment of soil biological parameters such as microbial biomass carbon (MBC), basal respiration or specific enzyme activities (Heinze et al., 2013).

The absorption features of every specific soil parameter are derived from different bands or wavelength regions of the spectra (see Fig. 2) (Ben Dor et al., 1999). Bands around 1100, 1600, 1700–1800, 2000, and 2200–2400 nm have for example been identified as being particularly important for soil organic matter and total N determination (Armenta and de la Guardia, 2014). But some parameters are not expected to have direct spectral absorption features in the Vis-NIR region. If a prediction of these parameters is still possible, they are assumed to be correlated to spectrally active soil constituents (Ben Dor et al., 1999; Armenta and de la Guardia, 2014). Examples for soil parameter which do not provide a direct spectral response are plant nutrients, such as P, K, Ca, Fe, Na and Mg as well as soil pH (Armenta and de la Guardia, 2014). Biological parameter like microbial biomass, basal respiration or enzyme activities are also expected to be predicted only indirectly as a consequence of high correlations with total soil organic matter quantity and quality (Johnson et al., 2003; Zornoza et al., 2008). Moreover, many absorption features overlap so that absorptions related to one soil parameter can be masked, distorted or shifted (Nocita et al., 2015). For example can two soil samples with the same SOC content but different sand content result in significantly different SOC results derived from the soil spectra, because an increase of sand content increases the SOC absorption depth, which can easily be confused with an increase in SOC (Stenberg et al., 2010; Stevens et al., 2013). That means, spectral features of soils are known to be not unique (Nocita et al., 2015), but rather weak, narrow and mixed (Ben-Dor et al., 2009). Overall, the complexity of both SOM biochemistry and SOM spectral response hardly allows to assign absorption features to specific SOM parameter or functional groups, which results in a highly variable use of wavelengths in prediction models (Ladoni et al., 2009; Croft et al., 2012).

Besides the challenge of linking (these very site-specific) SOM parameter to spectral features, there are other general challenges of applying reflectance spectroscopy to soils. Challenges associated with the procedure of spectroscopy in this context can be summarized as 1) atmospheric attenuation, 2) spectral resolution and number of channels, 3)

signal to noise ratio, 4) pixel size and sampling techniques and 5) measurement geometry. Challenges associated with the soil are mainly due to 1) vegetation cover, 2) soil crusts and 3) surface conditions (water, roughness, management measures) (Ben-Dor, 2011). For studying soils of temperate regions this explicitly means that the window of opportunity for remote sensing to measure soil surfaces is generally reduced because i) bare soil exists mainly for short periods of the year directly after tillage of arable soils, ii) the cloud cover is often preventing clear day skies, which are necessary for remote sensors to provide spectral reflectance data of soil parameters, iii) soil moisture is very variable, including aspects of freeze/thaw and snow in winter, which have a large effect on spectral reflectance and iv) the angle of the sun is low in winter, which also affects spectral response of remote sensors (Archer et al., 2014). Also for proximal sensing, the presence of surface disturbances such as eolian deposit, rock outcrops, biogenic or physical surface crusts, the possible disturbance of the A horizon by ploughing and associated micro-shadows due to increased roughness of the surface can additionally affect the soil spectral measurements negatively (Ben-Dor et al., 2009).

2.3 Resulting Objectives and Hypotheses

In the first phase of the presented interdisciplinary research, the objective was to assemble a comprehensive local soil library of soil spectra and associated parameters covering a diverse and broad range of site-specific values. On this basis, optimization strategies of data processing were investigated first. In the second phase, the potential of an extended set of SOM parameter (focus on biological SOM parameter; in situ data) to be assessed by soil spectroscopy was examined. Additionally, the transferability of the identified optimization strategies to these extended spectral data sets were tested. In this second phase, the focus was also to monitor SOM dynamics and to evaluate in this context if the soil surface can be spectrally representative for the layers below. The third phase of the research aimed at studying SOM origin (plant vs. microbial input) and its impact on SOM quantity and quality. This led to the formulation of the following central hypotheses (H1 – H5):

H1: Prediction accuracy of common SOM parameter (SOC, N) from soil reflectance spectra of hyperspectral laboratory measurements can be improved by pre-processing of spectral data. This refers to Chapter 3: Using variable selection and wavelets to exploit the full potential of visible-near infrared spectra for predicting soil properties.

H2: The spectral potential of SOM assessment of other than the common spectrally monitored parameter (SOC; N), with emphasis on biological or labile SOM parameter (microbial carbon (MBC), hot water extractable carbon (HWEC), dissolved organic carbon (DOC), chlorophyll α (Chl α) and phospholipid fatty acids (PLFAs)) is greater than generally

assumed. This refers to Chapter 4: Spatial and temporal changes of the microbial community and labile organic matter fractions affect the Vis-NIR spectroscopic characterization of arable topsoil SOM.

H3: Prediction accuracy improvement by pre-processing of spectral data (H1) can also be obtained for in situ data from hyperspectral field measurements. This refers to Chapter 4: Spatial and temporal changes of the microbial community and labile organic matter fractions affect the Vis-NIR spectroscopic characterization of arable topsoil SOM.

H4: The highly dynamic microbial community of the soil surface affects the labile C topsoil fractions and the in situ spectroscopic measurements of the soil surface. This refers to Chapter 4: Spatial and temporal changes of the microbial community and labile organic matter fractions affect the Vis-NIR spectroscopic characterization of arable topsoil SOM.

H5: The relative contribution of microbial vs. plant-derived organic matter to SOM in different soils and pools of stability is higher than generally assumed. This refers to Chapter 5: Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils.

2.4 References (Chapters 1 & 2)

Anderson, T.-H., Domsch, K.H., 2010. Soil microbial biomass: The eco-physiological approach. Soil Biology and Biochemistry 42, 2039-2043.

Apostel, C., Dippold, M.A., Bore, E., Kuzyakov, Y., 2017. Sorption of Alanine changes microbial metabolism in addition to availability. Geoderma 292, 128-134.

Archer, N., Rawlins, B., Grebby, S., Marchant, B., Emmett, B., 2014. Identify the opportunities provided by developments in earth observation and remote sensing for national scale monitoring of soil quality. In: Survey, B.G. (Ed.), British Geological Survey Internal Report, p. 32.

Armenta, S., de la Guardia, M., 2014. Vibrational spectroscopy in soil and sediment analysis. Trends in Environmental Analytical Chemistry 2, 43-52.

Askari, M.S., Cui, J., O'Rourke, S.M., Holden, N.M., 2015. Evaluation of soil structural quality using VIS–NIR spectra. Soil and Tillage Research 146, 108-117.

Basler, A., Dippold, M., Helfrich, M., Dyckmans, J., 2015. Microbial carbon recycling: an underestimated process controlling soil carbon dynamics – Part 2: A C₃-C₄ vegetation change field labelling experiment. Biogeosciences 12, 6291-6299.

Bastida, F., Zsolnay, A., Hernández, T., García, C., 2008. Past, present and future of soil quality indices: A biological perspective. Geoderma 147, 159-171.

Bellon-Maurel, V., 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.

Ben-Dor, E., 2011. Quantitative Remote Sensing of Soil Properties. In: Sparks, D.L. (Ed.), Advances in Agronomy.

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

Ben-Dor, E., Irons, J.R., Epema, J.F., 1999. Soil reflectance. In: Ryerson, R.A., Rencz, A.N. (Eds.), Manual of Remote Sensing, Remote Sensing for the Earth Sciences. Wiley, New York, pp. 111-188.

Bird, J.A., Kleber, M., Torn, M.S., 2008. 13C and 15N stabilization dynamics in soil organic matter fractions during needle and fine root decomposition. Organic Geochemistry 39, 465-477.

Blagodatskaya, E., Kuzyakov, Y., 2013. Active microorganisms in soil: Critical review of estimation criteria and approaches. Soil Biology and Biochemistry 67, 192-211.

Blum, W.E.H., 2005. Functions of Soil for Society and the Environment. Reviews in Environmental Science and Bio/Technology 4, 75-79.

Blume, H.P., Brümmer, G.W., Scheffer, F., Horn, R., Kandeler, E., Schachtschabel, P., Kögel-Knabner, I., Welp, G., Kretzschmar, R., Thiele-Bruhn, S., 2010. Scheffer/Schachtschabel: Lehrbuch der Bodenkunde. Spektrum Akademischer Verlag.

Bradford, M.A., Keiser, A.D., Davies, C.A., Mersmann, C.A., Strickland, M.S., 2013. Empirical evidence that soil carbon formation from plant inputs is positively related to microbial growth. Biogeochemistry 113, 271-281.

Bradford, M.A., Wood, S.A., Bardgett, R.D., Black, H.I.J., Bonkowski, M., Eggers, T., Grayston, S.J., Kandeler, E., Manning, P., Setälä, H., Jones, T.H., 2014. Discontinuity in the responses of ecosystem processes and multifunctionality to altered soil community composition. Proceedings of the National Academy of Sciences 111, 14478-14483.

Campbell, E.E., Paustian, K., 2015. Current developments in soil organic matter modeling and the expansion of model applications: a review. Environmental Research Letters 10, 123004.

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

Chodak, M., Niklińska, M., Beese, F., 2007. The Use of near Infrared Spectroscopy to Quantify Lignite-Derived Carbon in Humus-Lignite Mixtures. Journal of Near Infrared Spectroscopy 15, 195-200.

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

COM, C.o.t.E.C., 2000. Indicators for the integration of environmental concerns into the common agricultural policy. Communication from the Commission to the Council and the European Parliament. COM 20 final.

Croft, H., Kuhn, N.J., Anderson, K., 2012. On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. Catena 94, 64-74.
de Paul Obade, V., Lal, R., 2016. A standardized soil quality index for diverse field conditions. Sci Total Environ 541, 424-434.
Dippold, M.A., Kuzyakov, Y., 2013. Biogeochemical transformations of amino acids in soil assessed by position-specific labelling. Plant and Soil 373, 385-401.
Doelman, P., Eijsackers, H.J.P., 2004. Vital Soil: Function, Value and Properties. Elsevier Science.
Doran, J.W., Parkin, T.B., 1994. Defining and Assessing Soil Quality. In: Doran, J.W., Coleman, D.C., Bezdicek, D.F., Stewart, B.A. (Eds.), Defining Soil Quality for a Sustainable Environment. Soil Science Society of America and American Society of Agronomy, Madison, WI.
Doran, J.W., Parkin, T.B., 1996. Quantitative Indicators of Soil Quality: A Minimum Data Set. In: Doran, J.W., Jones, A.J. (Eds.), Methods for Assessing Soil Quality. Soil Science Society of America, Madison, WI, pp. 25-37.
Erhagen, B., Oquist, M., Sparrman, T., Haei, M., Ilstedt, U., Hedenstrom, M., Schleucher, J., Nilsson, M.B., 2013. Temperature response of litter and soil organic matter decomposition is determined by chemical composition of organic material. Glob Chang Biol 19, 3858-3871.
FAO, ITPS, 2015. Status of the World's Soil Resources (SWSR) – Main Report Rome, Italy.
Feld, C.K., Martins da Silva, P., Paulo Sousa, J., de Bello, F., Bugter, R., Grandin, U., Hering, D., Lavorel, S., Mountford, O., Pardo, I., PÃ¤rtel, M., RÃ¶mbke, J.r., Sandin, L., Bruce Jones, K., Harrison, P., 2009. Indicators of biodiversity and ecosystem services: a synthesis across ecosystems and spatial scales. Oikos 118, 1862-1871.
Gleixner, G., 2013. Soil organic matter dynamics: a biological perspective derived from the use of compound-specific isotopes studies. Ecological Research 28, 683-695.
Gregorich, E.G., Monreal, C.M., Carter, M.R., Angers, D.A., Ellert, B.H., 1994. Towards a minimum data set to assess soil organic matter quality in agricultural soils. Canadian Journal of Soil Science 74, 367-385.
Griffiths, B.S., Römbke, J., Schmelz, R.M., Scheffczyk, A., Faber, J.H., Bloem, J., Pérès, G., Cluzeau, D., Chabbi, A., Suhadolc, M., Sousa, J.P., Martins da Silva, P., Carvalho, F., Mendes, S., Morais, P., Francisco, R., Pereira, C., Bonkowski, M., Geisen, S., Bardgett, R.D., de Vries, F.T., Bolger, T., Dirilgen, T., Schmidt, O., Winding, A., Hendriksen, N.B., Johansen, A., Philippot, L., Plassart, P., Bru, D., Thomson, B., Griffiths, R.I., Bailey, M.J., Keith, A., Rutgers, M., Mulder, C., Hannula, S.E., Creamer, R., Stone, D., 2016. Selecting cost effective and policy-relevant biological indicators for European monitoring of soil biodiversity and ecosystem function. Ecological Indicators 69, 213-223.
Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. Soil Science Society of America Journal 75, 1201.
Gunina, A., Kuzyakov, Y., 2015. Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. Soil Biology and Biochemistry 90, 87-100.
Havlicek, E., 2012. Soil biodiversity and bioindication: From complex thinking to simple acting. European Journal of Soil Biology 49, 80-84.
Haynes, R.J., 2005. Labile Organic Matter Fractions as Central Components of the Quality of Agricultural Soils: An Overview. Advances in Agronomy. Academic Press, pp. 221-268.
Heinze, S., Vohland, M., Joergensen, Rainer G., Ludwig, B., 2013. Usefulness of near-infrared spectroscopy for the prediction of chemical and biological soil properties in different long-term experiments. Journal of Plant Nutrition and Soil Science 176, 520-528.
Hiederer, R., 2009. Distribution of Organic Carbon in Soil Profile Data. In: Communities, O.f.O.P.o.t.E. (Ed.), EUR 23980 EN. European Commission Joint Research Centre Institute for Environment and Sustainability, Luxembourg, p. 126.
IRENA, 2015. Renewable Energy in the Water, Energy & Food Nexus.
Jansson, J.K., Baker, E.S., 2016. A multi-omic future for microbiome studies. Nat Microbiol 1, 16049.
Johnson, M.J., Lee, K.Y., Scow, K.M., 2003. DNA fingerprinting reveals links among agricultural crops, soil properties, and the composition of soil microbial communities. Geoderma 114, 279-303.
Kallenbach, C.M., Frey, S.D., Grandy, A.S., 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nat Commun 7, 13630.
Karlen, D.L., Mausbach, M.J., Doran, J.W., Cline, R.G., Harris, R.F., Schuman, G.E., 1997. Soil Quality: A Concept, Definition, and Framework for Evaluation (A Guest Editorial). Soil Science Society of America Journal 61, 4-10.
Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., Pachepsky, Y., van der Putten, W.H., Bardgett, R.D., Moolenaar, S., Mol, G., Jansen, B., Fresco, L.O., 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. Soil 2, 111-128.

Kindler, R., Miltner, A., Thullner, M., Richnow, H.-H., Kästner, M., 2009. Fate of bacterial biomass derived fatty acids in soil and their contribution to soil organic matter. Organic Geochemistry 40, 29-37.

Kleber, M., 2010a. Response to the Opinion paper by Margit von Lützow and Ingrid Kögel-Knabner on 'What is recalcitrant soil organic matter?' by Markus Kleber. Environmental Chemistry 7, 336-337.

Kleber, M., 2010b. What is recalcitrant soil organic matter? Environmental Chemistry 7, 320.

Kleber, M., Johnson, M.G., 2010. Advances in Understanding the Molecular Structure of Soil Organic Matter. 106, 77-142.

Knicker, H., Totsche, K.U., Almendros, G., González-Vila, F.J., 2005. Condensation degree of burnt peat and plant residues and the reliability of solid-state VACP MAS 13C NMR spectra obtained from pyrogenic humic material. Organic Geochemistry 36, 1359-1377.

Kogel-Knabner, I., 2002. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. Soil Biology and Biochemistry 34, 139-162.

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept & review. Soil Biology and Biochemistry 83, 184-199.

Ladoni, M., Bahrami, H.A., Alavipanah, S.K., Norouzi, A.A., 2009. Estimating soil organic carbon from soil reflectance: a review. Precision Agriculture 11, 82-99.

Laishram, J., Saxena, K.G., Maikhuri, R.K., Rao, K.S., 2012. Soil quality and soil health: a review. International Journal of Ecology and Environmental Sciences 38, 19-37.

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

Lal, R., 2010. Managing Soils and Ecosystems for Mitigating Anthropogenic Carbon Emissions and Advancing Global Food Security. BioScience 60, 708-721.

Lal, R., Negassa, W., Lorenz, K., 2015. Carbon sequestration in soil. Current Opinion in Environmental Sustainability 15, 79-86.

Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J.M., Heurich, M., Jung, A., Klenke, R., Neumann, C., Pause, M., Rocchini, D., Schaepman, M.E., Schmidtlein, S., Schulz, K., Selsam, P., Settele, J., Skidmore, A.K., Cord, A.F., 2016. Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. Ecological Indicators 70, 317-339.

Lehmann, J., Kleber, M., 2015. The contentious nature of soil organic matter. Nature 528, 60-68.

Leifeld, J., Kögel-Knabner, I., 2005. Soil organic matter fractions as early indicators for carbon stock changes under different land-use? Geoderma 124, 143-155.

Leifeld, J., Steffens, M., Galego-Sala, A., 2012. Sensitivity of peatland carbon loss to organic matter quality. Geophysical Research Letters 39, n/a-n/a.

Lenton, T.M., Held, H., Kriegler, E., Hall, J.W., Lucht, W., Rahmstorf, S., Schellnhuber, H.J., 2008. Tipping elements in the Earth's climate system. Proceedings of the National Academy of Sciences 105, 1786-1793.

Liang, C., Balser, T.C., 2011. Microbial production of recalcitrant organic matter in global soils: implications for productivity and climate policy. Nat Rev Microbiol 9, 75; author reply 75.

Liang, C., Cheng, G., Wixon, D.L., Balser, T.C., 2010. An Absorbing Markov Chain approach to understanding the microbial role in soil carbon stabilization. Biogeochemistry 106, 303-309.

Linsler, D., Sawallisch, A., Höper, H., Schmidt, H., Vohland, M., Ludwig, B., 2017. Near-infrared spectroscopy for determination of soil organic C, microbial biomass C and C and N fractions in a heterogeneous sample of German arable surface soils. Archives of Agronomy and Soil Science, 1-11.

Loveland, P., Webb, J., 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. Soil and Tillage Research 70, 1-18.

Ludwig, B., Vormstein, S., Niebuhr, J., Heinze, S., Marschner, B., Vohland, M., 2017. Estimation accuracies of near infrared spectroscopy for general soil properties and enzyme activities for two forest sites along three transects. Geoderma 288, 37-46.

McCormick, R., Kapustka, L.A., 2016. The Answer is 42...What is THE question? Journal of Environmental Studies and Sciences 6, 208–213.

MEA, 2005. Millennium Ecosystem Assessment: Ecosystems and Human Well-being: Synthesis. In: Institute, W.R. (Ed.), Washington, DC.

Miltner, A., Bombach, P., Schmidt-Brücken, B., Kästner, M., 2012. SOM genesis: microbial biomass as a significant source. Biogeochemistry 111, 41-55.

Miltner, A., Kindler, R., Knicker, H., Richnow, H.-H., Kästner, M., 2009. Fate of microbial biomass-derived amino acids in soil and their contribution to soil organic matter. Organic Geochemistry 40, 978-985.

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

Mueller, C.W., Kölbl, A., Hoeschen, C., Hillion, F., Heister, K., Herrmann, A.M., Kögel-Knabner, I., 2012. Submicron scale imaging of soil organic matter dynamics using NanoSIMS – From single particles to intact aggregates. Organic Geochemistry 42, 1476-1488.

Mueller, C.W., Weber, P.K., Kilburn, M.R., Hoeschen, C., Kleber, M., Pett-Ridge, J., 2013. Advances in the Analysis of Biogeochemical Interfaces. 121, 1-46.

Müller, A., Weigelt, J., 2013. Governance for a Land Degradation Neutral World. IISD, IISD Land Policy and Practice Knowledge Database.

Murphy, D.V., Cookson, W.R., Braimbridge, M., Marschner, P., Jones, D.L., Stockdale, E.A., Abbott, L.K., 2011. Relationships between soil organic matter and the soil microbial biomass (size, functional diversity, and community structure) in crop and pasture systems in a semi-arid environment. Soil Research 49, 582.

Muscolo, A., Settineri, G., Attinà, E., 2015. Early warning indicators of changes in soil ecosystem functioning. Ecological Indicators 48, 542-549.

Nielsen, U.N., Ayres, E., Wall, D.H., Bardgett, R.D., 2011. Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity-function relationships. European Journal of Soil Science 62, 105-116.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. 132, 139-159.

Orgiazzi, A., Panagos, P., Yigini, Y., Dunbar, M.B., Gardi, C., Montanarella, L., Ballabio, C., 2016. A knowledge-based approach to estimating the magnitude and spatial patterns of potential threats to soil biodiversity. Sci Total Environ 545-546, 11-20.

Parr, J.F., Papendick, R.I., Hornick, S.B., Meyer, R.E., 1992. Soil quality: Attributes and relationship to alternative and sustainable agriculture. American Journal of Alternative Agriculture 7, 5-11.

Paz-Ferreiro, J., Fu, S., 2016. Biological Indices for Soil Quality Evaluation: Perspectives and Limitations. Land Degradation & Development 27, 14-25.

Poirier, N., Sohi, S.P., Gaunt, J.L., Mahieu, N., Randall, E.W., Powlson, D.S., Evershed, R.P., 2005. The chemical composition of measurable soil organic matter pools. Organic Geochemistry 36, 1174-1189.

Pronk, G.J., Heister, K., Kögel-Knabner, I., 2015. Amino sugars reflect microbial residues as affected by clay mineral composition of artificial soils. Organic Geochemistry 83-84, 109-113.

Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Pérès, G., Rutgers, M., 2012. Soil biodiversity, biological indicators and soil ecosystem services-an overview of European approaches. Current Opinion in Environmental Sustainability 4, 529-538.

Pulleman, M.M., Bouma, J., van Essen, E.A., Meijles, E.W., 2000. Soil Organic Matter Content as a Function of Different Land Use History. Soil Science Society of America Journal 64, 689.

Rasche, F., Marhan, S., Berner, D., Keil, D., Kandeler, E., 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.

Remusat, L., Hatton, P.J., Nico, P.S., Zeller, B., Kleber, M., Derrien, D., 2012. NanoSIMS study of organic matter associated with soil aggregates: advantages, limitations, and combination with STXM. Environ Sci Technol 46, 3943-3949.

Rinnan, R., Rinnan, Å., 2007. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. Soil Biology and Biochemistry 39, 1664-1673.

Ritz, K., Black, H.I.J., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: A framework for balancing scientific and technical opinion to assist policy development. Ecological Indicators 9, 1212-1221.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., Wit, C.A.d., Hughes, T., Leeuw, S.v.d., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. Nature 461, 472-475.

Rodionov, A., Pätzold, S., Welp, G., Pude, R., Amelung, W., 2016. Proximal field Vis-NIR spectroscopy of soil organic carbon: A solution to clear obstacles related to vegetation and straw cover. Soil and Tillage Research 163, 89-98.

Rodriguez-Caballero, E., Paul, M., Tamm, A., Caesar, J., Budel, B., Escribano, P., Hill, J., Weber, B., 2017. Biomass assessment of microbial surface communities by means of hyperspectral remote sensing data. Sci Total Environ.

Roume, H., Heintz-Buschart, A., Muller, E.E.L., May, P., Satagopam, V.P., Laczny, C.C., Narayanasamy, S., Lebrun, L.A., Hoopmann, M.R., Schupp, J.M., Gillece, J.D., Hicks, N.D., Engelthaler, D.M., Sauter, T., Keim, P.S., Moritz, R.L., Wilmes, P., 2015. Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. npj Biofilms and Microbiomes 1, 15007.

Rousk, J., Bengtson, P., 2014. Microbial regulation of global biogeochemical cycles. Front Microbiol 5, 103.

Schaeffer, A., Nannipieri, P., Kästner, M., Schmidt, B., Botterweck, J., 2015. From humic substances to soil organic matter–microbial contributions. In honour of Konrad Haider and James P. Martin for their outstanding research contribution to soil science. Journal of Soils and Sediments 15, 1865-1881.

Schaumann, G.E., 2006. Soil organic matter beyond molecular structure Part I: Macromolecular and supramolecular characteristics. Journal of Plant Nutrition and Soil Science 169, 145-156.

Schimel, J.P., Schaeffer, S.M., 2012. Microbial control over carbon cycling in soil. Front Microbiol 3, 348.

Schmidt, M.W., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kogel-Knabner, I., Lehmann, J., Manning, D.A., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49-56.

Schurig, C., Smittenberg, R.H., Berger, J., Kraft, F., Woche, S.K., Goebel, M.-O., Heipieper, H.J., Miltner, A., Kaestner, M., 2013. Microbial cell-envelope fragments and the formation of soil organic matter: a case study from a glacier forefield. Biogeochemistry 113, 595-612.

Simpson, A.J., Simpson, M.J., Smith, E., Kelleher, B.P., 2007. Microbially Derived Inputs to Soil Organic Matter: Are Current Estimates Too Low? Environmental Science & Technology 41, 8070-8076.

Sojka, R.E., Upchurch, D.R., 1999. Reservations regarding the soil quality concept. Soil Science Society of America Journal 63, 1039-1054.

Solomon, D., Lehmann, J., Harden, J., Wang, J., Kinyangi, J., Heymann, K., Karunakaran, C., Lu, Y., Wirick, S., Jacobsen, C., 2012a. Micro- and nano-environments of carbon sequestration: Multi-element STXM–NEXAFS spectromicroscopy assessment of microbial carbon and mineral associations. Chemical Geology 329, 53-73.

Solomon, D., Lehmann, J., Wang, J., Kinyangi, J., Heymann, K., Lu, Y., Wirick, S., Jacobsen, C., 2012b. Micro- and nano-environments of C sequestration in soil: a multi-elemental STXM-NEXAFS assessment of black C and organomineral associations. Sci Total Environ 438, 372-388.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M.J., 2013. 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.

St. Luce, M., Ziadi, N., Zebarth, B.J., Grant, C.A., Tremblay, G.F., Gregorich, E.G., 2014. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. Geoderma 232-234, 449-458.

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and Near Infrared Spectroscopy in Soil Science. 107, 163-215.

Stevens, A., Nocita, M., Toth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. PLoS One 8, e66409.

Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S.Y., Rawlins, B.G., Field, D.J., 2015. Global soil organic carbon assessment. Global Food Security 6, 9-16.

Stone, D., Ritz, K., Griffiths, B.G., Orgiazzi, A., Creamer, R.E., 2016. Selection of biological indicators appropriate for European soil monitoring. Applied Soil Ecology 97, 12-22.

Terhoeven-Urselmans, T., Schmidt, H., Georg Joergensen, R., Ludwig, B., 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pre-treatment. Soil Biology and Biochemistry 40, 1178-1188.

Tóth, G., Stolbovoy, V., Montanarella, L., 2007. Soil quality and sustainability evaluation – an integrated approach to support soil-related policies of the European Union. In: Communities, O.f.O.P.o.t.E. (Ed.), EUR 22721 EN, Luxembourg, p. 40.

UN, 2015. World population prospects: the 2015 revision. . In: Nations, U. (Ed.). Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf.

Valin, H., Sands, R.D., van der Mensbrugghe, D., Nelson, G.C., Ahammad, H., Blanc, E., Bodirsky, B., Fujimori, S., Hasegawa, T., Havlik, P., Heyhoe, E., Kyle, P., Mason-D'Croz, D., Paltsev, S., Rolinski, S., Tabeau, A., van Meijl, H., von Lampe, M., Willenbockel, D., 2014. The future of food demand: understanding differences in global economic models. Agricultural Economics 45, 51-67.

Velasquez, E., Lavelle, P., Andrade, M., 2007. GISQ, a multifunctional indicator of soil quality. Soil Biology and Biochemistry 39, 3066-3080.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, N., 2011. Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. In: Sparks, D.L. (Ed.), Advances in Agronomy pp. 237-282.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Science Reviews 155, 198-230.

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

Vohland, M., Ludwig, M., Thiele-Bruhn, S., 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.

von Lützow, M., Kögel-Knabner, I., 2010. Response to the Concept paper: 'What is recalcitrant soil organic matter?' by Markus Kleber. Environmental Chemistry 7, 333.

von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., Marschner, B., 2007. SOM fractionation methods: Relevance to functional pools and to stabilization mechanisms. Soil Biology and Biochemistry 39, 2183-2207.

Wall, D.H., 2007. Global change tipping points: above- and below-ground biotic interactions in a low diversity ecosystem. Philosophical Transactions of the Royal Society B: Biological Sciences 362, 2291-2306.

Weathers, K.C., Groffman, P.M., Van Dolah, E., Bernhardt, E., Grimm, N.B., McMahon, K., Schimel, J., Paolisso, M., Maranger, R., Baer, S., Brauman, K., Hinckley, E., 2016. Frontiers in Ecosystem Ecology from a Community Perspective: The Future is Boundless and Bright. Ecosystems 19, 753-770.

Weigelt, J., Müller, A., Beckh, C., Töpfer, K., 2014. Soils in the Nexus. A Crucial Resource for Water, Energy and Food Security. Oekom verlag, München

Wickings, K., Grandy, A.S., Reed, S.C., Cleveland, C.C., 2012. The origin of litter chemical complexity during decomposition. Ecol Lett 15, 1180-1188.

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K.M., Arcenegui, V., Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. Soil Biol Biochem 40, 1923-1930.

Zornoza, R., Mataix-Solera, J., Guerrero, C., Arcenegui, V., Garcia-Orenes, F., Mataix-Beneyto, J., Morugan, A., 2007. Evaluation of soil quality using multiple lineal regression based on physical, chemical and biochemical properties. Sci Total Environ 378, 233-237.

3 Using variable selection and wavelets to exploit the full potential of visible–near infrared spectra for predicting soil properties

M. Vohland, M. Ludwig, M. Harbich, C. Emmerling, S. Thiele-Bruhn

Journal of Near Infrared Spectroscopy 24 (2016) 255–269

DOI:10.1255/jnirs.1233

Abstract

In soil spectroscopy a series of strategies exist to optimise multivariate calibrations. We explore this issue with a set of topsoil samples for which we estimated soil organic carbon (OC) and total nitrogen (N) from visible–near infrared (vis–NIR) spectra (350–2500 nm). In total, 172 samples were collected to cover the soil heterogeneity in our study area located in western Rhineland-Palatinate, Germany. There, soils with varying properties developed from very diverse parent materials, e.g., ranging from very acidic sandstone to dolomitic marl. We defined four sample sets each of a different size and heterogeneity. Each set was subdivided into a calibration and a validation set. The first strategy that we tested to improve prediction accuracies was spectral variable selection using competitive adaptive reweighted sampling (CARS) and iteratively retaining informative variables (IRIV), both in combination with partial least squares regression (PLSR). In addition, continuous wavelet transformation (CWT) with the Mexican Hat wavelet was applied to decompose the measured spectra into multiple scale components (dyadic scales 2^1–2^5) and thus to represent the high and low frequency features contained in the spectra. CARS was then applied to select wavelet coefficients from the different scales and to introduce them in the PLSR approach (CWT-CARS-PLSR). Regarding prediction power, CWT-CARS-PLSR outperformed the other approaches. For the smallest data set with 30 validation samples, prediction accuracy for OC increased from approximately quantitative with full spectrum-PLSR (r^2 = 0.81, residual prediction deviation (*RPD*) = 2.27) to excellent when using wavelet decomposition and CARS-PLSR (r^2 = 0.93, *RPD* = 3.60). For N, predictions improved from unsuccessful (r^2 = 0.63, *RPD* = 1.36) to approximately quantitative (r^2 = 0.84, *RPD* = 2.03). In case of OC, predictions were worst for the largest dataset with 57 validation samples: CWT-CARS-PLSR achieved approximately quantitative predictions (r^2 = 0.82, *RPD* = 2.31), whereas full spectrum-PLSR provided estimates that allowed only separating between high and low values (r^2 = 0.72, *RPD* = 1.88). Accuracy of N estimation for this dataset using CWT-CARS-PLSR was also approximately quantitative. Concerning the tested spectral variable selection techniques, both methods provided similar results in the prediction. The application of IRIV was limited due to long processing times.

1. Introduction

Near-infrared (NIR) spectroscopy (wavelength range: 700–2500 nm) and visible to near-infrared (vis–NIR) spectroscopy (400–2500 nm) are reported in many studies to be useful for the estimation of a series of chemical, physical and biological soil properties (for reviews see e.g. Cécillon *et al.*[1] and Soriano-Disla *et al.*[2]). Both NIR and vis–NIR spectroscopy provide large sets of spectral predictors that normally are strongly collinear, which may affect the success of a quantitative derivation of soil properties from the measured spectra. To partly compensate for these effects, reduction of dimensionality is a standard approach, which is realised e.g. by partial least squares regression (PLSR), with a projection of the spectral data into a low-dimensional space formed by a set of orthogonal latent variables.[3] Furthermore, many studies found that selecting informative spectral variables (or eliminating uninformative

or noisy variables) can further improve the accuracy of PLSR. Complexity of the model might be reduced, which may lead to a more robust calibration in terms of its predictive ability for independent validation samples. Additionally, interpretation of the model may be enhanced by selecting informative or key variables which allow an insight in the underlying spectral predictive mechanisms.[4-6]

The methods yet developed for variable selection may be subdivided into two categories.[7] In the first category, statistical features of the variables are used, such as regression coefficients or t-statistics; spectral variables are selected individually without considering joint effects. Uninformative variable elimination (UVE)[8], competitive adaptive reweighted sampling (CARS)[9] or successive projection algorithm (SPA)[10] are examples for this strategy. The second category comprises all methods that consider the interaction of variables in the search space, as for example the use of a genetic algorithm (GA; e.g. Leardi and González[11]) or simulated annealing (SA)[12]. Approaches of this category have to compromise between an exhaustive search through the possible combinations of variables and run time, whereas the approaches of the first category are computationally efficient and thus may be applied easily to datasets with some thousands of variables.[7]

With wavelet analysis the original spectra can be decomposed into multiple scales that correspond to different frequencies. This alternative representation may convey additional information such as the location and nature of high frequency features (narrow absorption features, sharp peaks and also noise) or the magnitude and shape of the reflectance (or absorbance) continuum at higher scales. Wavelet analysis of a spectrum is performed by scaling and shifting a wavelet function to produce wavelet coefficients, which then can be used as independent factors to construct predictive models for spectrally sensitive chemical, physical or biological constituents.[13,14] A range of studies exist for using wavelet coefficients as predictors of vegetation properties, e.g. pigment concentrations[15,14], foliar nitrogen contents[16], leaf area index[17,18], leaf water contents[19] or leaf mass per area for different plant species[20]. For the retrieval of soil properties in chemometric approaches with wavelet analysis, the number of available studies is more limited. Peng *et al.*[21] have shown that the decomposition of vis-NIR spectra with wavelet transforms could correctly reflect the variation of soil moisture; prediction success depended on the kind of mother wavelet and the decomposition levels combined. For soil clay content, Ge *et al.*[22] also demonstrated the capability of wavelet analysis to distinguish specific narrow and broad spectral absorptions by selecting wavelet regressors at different scales, which – in addition to parsimonious predictive models – facilitated the physical interpretation of the underlying spectral predictive mechanisms. Viscarra-Rossel and Lark[23] used wavelets and a variable selection technique to improve calibrations of soil clay and organic carbon (OC) from Vis–NIR and mid-infrared spectra. Most recently, Lin *et al.*[24] used wavelets to decompose spectra in different levels, identified the optimal decomposition level for soil organic matter (SOM) estimation by correlation analysis and integrated selected bands in a combined Wavelet-Correlation-PLSR approach. Another example for using wavelets in the field of soil spectroscopy, although with another focus, is the study of Zhang *et al.*[25]; they used wavelet transformed spectra to test classification strategies for the discrimination of different soil texture classes.

To apply wavelet analysis for calibration purposes one has to decide on the basic transform strategy, i.e., either discrete or continuous wavelet transform (DWT vs. CWT). The DWT was, for example, used in the soil studies of Ge *et al.*[22], Peng *et al.*[21] and Viscarra Rossel and Lark[23], whereas CWT was favoured in the vegetation studies mentioned above. In short, DWT chooses a subset of positions (and scales) usually based on powers of two (dyadic), which is computationally efficient. In CWT, the analysing wavelet is shifted smoothly over the full domain of the signal and the correlation between both is calculated at each position; thus, wavelet coefficients of CWT provide, at each scale, a vector of the same length as the spectrum, and each CWT wavelet coefficient is directly comparable to one reflectance band. As a drawback, CWT is computationally more intensive and generates a large amount of redundant data which could complicate regression analysis.[15,22] Blackburn and Ferwerda[15] applied both DWT and CWT to hyperspectral data for leaf chlorophyll analysis. They found CWT superior to DWT, which was attributed to the greater amount of spectral detail that was retained generally across all scales and especially in the higher levels of the continuous decomposition. However, in the CWT approach of that study stepwise regressions were performed only on coefficients from each scale separately. Thus, possible refinements of the approach could, *inter alia*, result from an integration of wavelet coefficients from multiple scales and the use of other approaches for the selection of wavelet coefficients during model calibration.[15]

Against this background, we used both vis–NIR spectra with and without wavelet decomposition combined with variable selection for a multivariate calibration procedure. Our complete sample set comprised 172 topsoil samples from arable sites that were collected in a region of approximately 600 km² with widely differing parent materials. We compiled four datasets of different size (n = 90, 124, 138 and 172 samples) and heterogeneity based on different geographic sub-regions. All samples were spectrally measured in the laboratory and analysed for total soil organic carbon and total soil nitrogen (N). We investigated the ability to predict these properties from the spectra, using PLSR as the currently most common multivariate calibration method in soil spectroscopy.[26,27] To this end, PLSR was combined with CARS and the iteratively retaining informative variables method (IRIV)[7], representing the two conceptually different categories of spectral variable selection mentioned above. For wavelet analysis, we used CWT with the Mexican hat as mother wavelet. With these methods, our objectives were (i) to study possible benefits of spectral feature selection (combined with PLSR) for the estimation of OC and N from spectra that were measured for sample sets of different size and heterogeneity, (ii) to identify key wavelengths for the prediction and to compare them for the different sample sets and (iii) to explore whether the decomposition of spectra with wavelets, in addition to variable selection, could further improve prediction accuracies.

2. Material and methods

Study site, soil samples and their analytical properties

A set of 172 soil samples was taken from topsoils (0–10 cm) of arable sites in the greater region of Trier (Rhineland-Palatinate, Germany). The field work was conducted in September 2010 and March 2011. The study region with a total area of approximately 600 km² can be subdivided in four geographic regions; one part was located in the northern Hunsrück (48 samples were taken from there), another in the southwest Eifel region (n = 78), the third in the middle Moselle valley (n = 12) and the fourth in eastern Luxembourg (n = 34) (Figure 1).

For the Hunsrück region, soil types ranged mainly from Stagnic Cambisols to Haplic Stagnosols derived from Devonian slate or tertiary deposits. Haplic to Dystric Leptosols derived from Lias sandstone were the main soil types for the Luxembourg region while for the Moselle valley Fluvisols derived from alluvial deposits of the Moselle river were found. Leptic and Haplic Cambisols derived from Triassic siltstone, sandstone and limestone were characterising the Eifel region. Depending on the morphological dynamics soils were locally characterised by air-blown silt in the top horizon. As a consequence of the high heterogeneity of parent materials, soil types and landscape dynamics, the sample set comprised a wide range of soil textures reaching from silty sand over silty loamy sand and silty loam to clayey loam.

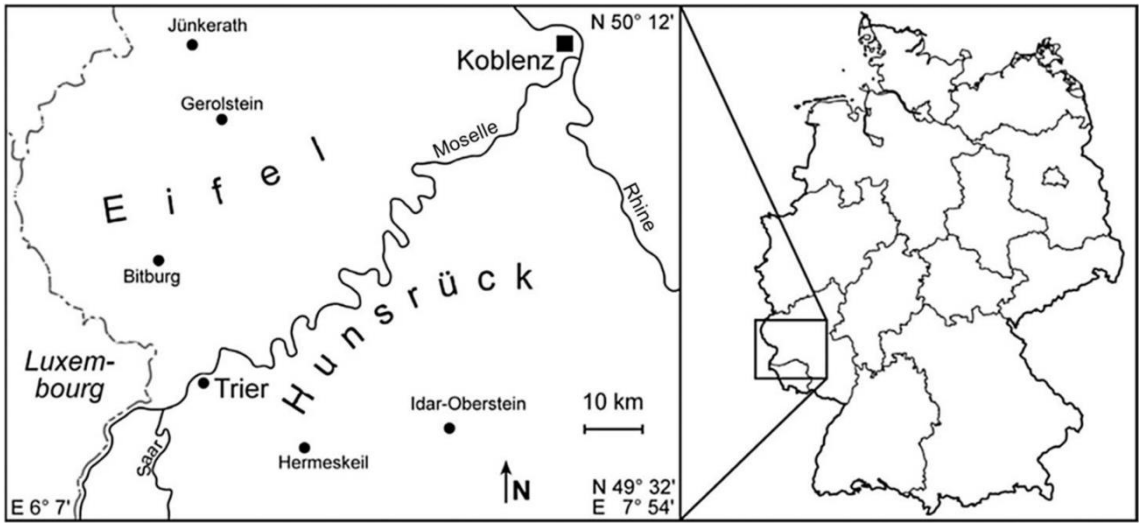


Figure 1: Location of the study area in Germany and eastern Luxembourg.

Soil samples were taken by five positions per site, i.e., at one central point and four surrounding spots in a distance of 1.5 m from the central point in each cardinal direction; the collected soil material was afterwards pooled to one composite sample.

In the laboratory, samples were sieved ≤ 2 mm, homogenised, air-dried and then, for OC and N analysis, finely ground using an agate mortar. The total contents of C and N were measured by gas chromatography after combustion at 1100°C using a EuroEA elemental analyzer (HekaTech, Wegberg, Germany). Soil samples containing free carbonate were additionally analysed using an IC-KIT 3100M (HekaTech, Wegberg, Germany) for inorganic carbon (IC) determination. There, carbonate-C is released from soil samples after reaction

with 50% H_3PO_4 in a closed chamber. The CO_2 released from carbonate was measured with the EuroEA elemental analyser. For samples containing IC, OC was calculated as OC = total-C – IC. Goodness of analyses was verified by eightfold repeated analyses of standard soils and compounds, respectively. Standard deviation of repeated measurements of a standard soil (No. HE 338601009; HekaTech, Wegberg, Germany) was 0.061% for OC and 0.004% for N. Reproducibility of IC analysis was tested using Na_2CO_3 (Nacalai Tesque Inc., Kyoto, Japan), revealing a standard deviation of 0.411%.

Statistics of the analysed soil properties are provided for both the total set and the four spatial subsets in Table 1. Reported values are typical for agricultural soils, but also indicate a relatively broad range for the complete dataset and the Eifel and Hunsrück subsets.

Table 1: Descriptive statistics of soil parameters analytically determined for the different sample sets.

		Mean	Median	Range	Standard deviation	Skewness[a]
OC (%)	Total	1.94	1.76	0.98-4.01	0.65	1.28
	Eifel	2.18	1.94	1.17-4.01	0.72	0.88
	Hunsrück	1.84	1.65	0.98-3.42	0.60	0.90
	Moselle	1.68	1.68	1.10-2.90	0.47	1.52
	Luxembourg	1.64	1.57	1.04-2.44	0.33	0.55
N (%)	Total	0.184	0.172	0.085-0.347	0.058	0.77
	Eifel	0.204	0.191	0.098-0.339	0.054	0.46
	Hunsrück	0.186	0.163	0.085-0.347	0.065	0.94
	Moselle	0.147	0.137	0.110-0.210	0.032	0.87
	Luxembourg	0.147	0.141	0.087-0.234	0.037	0.59
C/N	Total	10.7	10.3	7.5-15.8	1.56	0.78
	Eifel	10.7	10.4	7.5-15.8	1.72	0.71
	Hunsrück	10.0	9.8	7.6-11.9	0.89	-0.08
	Moselle	11.4	10.7	9.5-14.8	1.98	0.76
	Luxembourg	11.3	11.2	8.8-14.7	1.36	0.30
pH (CaCl2)	Total	5.34	5.99	3.87-7.49	0.83	7.89[b]
	Eifel	5.35	6.70	3.87-7.49	0.90	7.03[b]
	Hunsrück	5.54	5.82	4.60-6.85	0.48	3.52[b]
	Moselle	6.19	6.82	5.48-7.33	0.59	2.24[b]
	Luxembourg	5.08	5.40	4.44-6.55	0.58	1.55[b]

[a] $$Skewness = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3$$

[b] calculated for activity of hydrogen $a(H^+) = 10^{-pH}$

Spectra

All 172 samples were spectrally measured in the laboratory with a spectroradiometer (FieldSpec II fibre-optics, ASD Inc., Boulder, CO, USA), which performs measurements in the spectral range from 350 nm to 2500 nm. The spectral resolution of this instrument is 3 nm at 700 nm and 10 nm at 1400 and 2100 nm. The sampling interval is 1.4 nm in the range from 350 to 1000 nm and 2 nm at longer wavelengths; spectra are provided with 1 nm increments (2151 channels). Spectra were measured as percentage of reflectance with

respect to a Spectralon® (Labsphere, Inc., North Sutton, USA) reference panel. As the reflectivity of this panel is known, the measurements could be converted to absolute bidirectional reflectance values. The samples were illuminated with a 1000 W quartz-halogen lamp at 30° zenith angle and measured with the optical head mounted on a tripod from nadir position.

For each sample, 50 readings were averaged, afterwards transformed to apparent absorbances by log(reflectance^{-1}) and resampled to an increment of 5 nm. The increment of 5 nm was selected as we found it appropriate to preserve all spectral details and to reduce at the same time the large number of highly correlated 2151 spectral variables. To exclude noisy parts of the spectra the wavelength ranges below 400 and above 2400 nm were removed. The spectra with now 401 spectral data points were used as input for the subsequent chemometric modelling.

Continuous wavelet transforms

The pre-processed spectra were directly used for the multivariate calibration or beforehand subjected to CWT using the *cwt* MATLAB function (The Math Works Inc., Natick, MA, USA). In CWT, the mother wavelet is shifted smoothly over the full domain of the spectrum. Wavelet coefficients, that represent the correlation between the wavelet and the respective segment of the spectrum, are then calculated at each possible position. After the shift is completed to the end of the spectrum, one therefore obtains a vector with wavelet coefficients that is of the same length as the vector with spectral values. Then, the wavelet is stretched (scaled) and again shifted over the spectrum. Accordingly, CWT provides many wavelet coefficients as a function of scale and position. Scale refers to the width of the wavelet; as the scale increases and the wavelet gets wider, the finer details get smeared out.

A great number of mother wavelets with different shapes exist that may be applied for a transformation of the measured spectra. However, as absorption features are similar to Gaussian function or a combination of multiple Gaussian functions we applied the Mexican hat wavelet (Figure 2), as it is proportional to the second derivative of the Gaussian probability density function. The effective support range of Mexican Hat wavelet without scaling (scale = 2^0) is [-5, 5] in terms of sample intervals (Figure 2).

Generally, CWT can operate at every scale. However, to avoid a huge volume of output data, we applied CWT at some selected (discretisized) scales large enough to cover the existing small- and large-scale spectral features contained in the measured spectra. We specified five scales in an increment of 2^a (with $1 \leq a \leq 5$), which resulted in five continuous vectors of wavelet coefficients for each spectrum.

Due to the continuous shift of the wavelet function along the signal CWT will lead to artefacts at the edges, i.e., at the beginning and the end of the spectrum, where parts of the wavelet function especially at larger scales will be outside the signal. One way to avoid these effects of missing data is padding each end of the spectrum.[28] After a test of different padding methods (e.g. zero-padding or linear padding), we performed a polynomial extrapolation of the spectrum at either end; the polynomial fit was done at each end with 20 data points using a second degree polynomial. All spectra extended in that way were then decomposed with

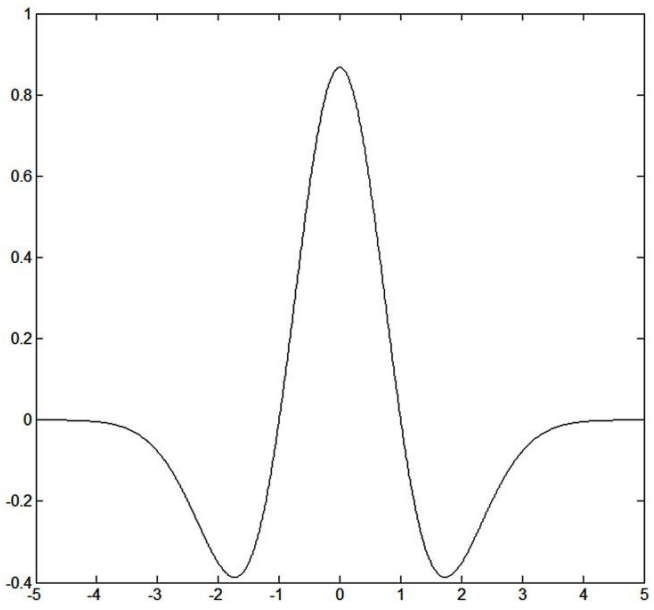


Figure 2: Shape of the Mexican hat wavelet with an [-5, 5] effective support range.

CWT to retrieve the final vectors of wavelet coefficients at all scales and positions in the 400-2400 nm feature space.

For the multivariate modelling, we followed two strategies: We used the data from each scale separately as input, and we also combined these data so that coefficients from different scales could be selected simultaneously for model calibration.

Multivariate calibration: PLSR and variable selection techniques

PLSR employs, similar to principal component regression, statistical rotations to overcome the problems of high-dimensionality and multicollinearity. Different from PCA, PLSR considers the relationship and the structure of both datasets, **X** (predictor variables) and **y** (dependent variable), to explain a maximum of covariance between both.[3] We used the nonlinear iterative partial least squares (NIPALS) algorithm for computing PLS regression components (i.e. factors or latent variables). For each constituent, the optimum number of latent variables was identified by performing a “leave-one-out” (LOO)-cross validation procedure. The minimum of the root mean squared error (RMSE) in the cross validation was used as decision criterion for the optimum number of latent variables. However, with a maximum number of 115 calibration samples (see next section, Definition of sample sets) the number of latent variables was restricted in all cases to a maximum of 12.

We combined PLSR with IRIV[7] and CARS[9] as variable selection strategies. Both methods were implemented as described in brief here.

The IRIV method intends to retain only informative spectral variables and performs a comparatively exhaustive selection strategy. Given the spectral data matrix with n samples and p variables a binary matrix with either 1 or 0 (in a ratio of 1:1) is generated; the dimension of this matrix is $k \times p$ with k as the number of random combinations of variables. The concrete dimension of k is defined depending on the number of considered spectral

variables. Each row of this binary matrix with a randomly generated sequence of 0 (variable is switched off) and 1 (variable is switched on) determines which variables are considered for the modelling with PLSR; cross validated RMSE ($RMSE_{cv}$) is used to evaluate the performance of each subset (row). At the end of this first step, a vector of $RMSE_{cv}$ values (k × 1) is obtained. In the next step, again all subsets defined by the rows of the binary matrix are considered for modelling, but systematic changes are performed to assess the importance of each variable. Let us consider the first column (spectral variable) of the binary matrix; formerly 0, the value is now changed to 1 (and vice versa), while the state of all other variables in column 2, 3, …, p is not changed. Again, each subset (now changed at the first position) is used with PLSR and another vector of $RMSE_{cv}$ values is obtained. At the end of this step, we therefore retrieve k pairs of $RMSE_{cv}$, each pair is related to one row of the binary matrix, one time including (case A), one time excluding (case B) the first spectral variable. The average $RMSE_{cv}$ is calculated for all case A- and all case B-models. If the mean $RMSE_{cv}$ (A) is less than the mean $RMSE_{cv}$ (B), the variable is considered informative, otherwise uninformative or interfering. The described procedure is performed for all variables (columns) of the binary matrix. At the end, all informative variables are retained, while all other variables are removed. Then, a new binary matrix with a reduced dimension p is generated and the selection procedure starts again. Many rounds are performed until only informative variables exist. For reasons of computation time, a five-fold (instead of LOO) cross validation is performed at all steps of the IRIV approach.

At the very last step, a backward elimination is conducted with the retained informative variables to retrieve a final list of variables to be included in the modelling approach. At this step, $RMSE_{cv}$ is calculated with all retained (j) variables. Then it is tested whether leaving out another variable i (with i = 1,2, …, j) provides a new minimum of $RMSE_{cv}$ being smaller than $RMSE_{cv}$ with all variables. If yes, this variable i is left out and the elimination strategy starts again (with j changed to j-1); otherwise no further variable is removed.

CARS aims at selecting key wavelengths with a rigorous and computationally efficient procedure. It performs m sampling runs to select m sets of spectral variables. In each run, two successive steps of wavelength selection are performed: In a first step, an exponentially decreasing function (EDF) is used for an enforced removal of wavelengths with relatively small PLS regression coefficients. The number of wavelengths to be kept from the original p wavelengths is defined in this step by $r_i \times p$ with

$$r_i = a \times \exp(-b \times (i+1) \quad (1)$$

$$b = \ln(0.5 \times p) \times (m\text{-}1)^{-1} \quad (2)$$

with i = sampling run = 1, 2, 3, …, m *and* $a = (0.5 \times p)^{1/(m-1)}$.

This means that from a data set with for example p = 401 spectral variables (as in our case) and with m = 50 sampling runs a total of 360 spectral variables is kept in the first step of the first run (with r_1 = 0.8975 from EDF); 41 variables – those with the smallest absolute values of regression coefficients – are removed.

After this step the complete probability space is made up of 360 subspaces in the [0,1] interval; the size of each subspace (i.e. the weight of each retained variable) is determined by the ratio of the absolute value of the respective regression coefficient to the total sum of absolute values of all regression coefficients. Next, an adaptive reweighted sampling of variables is employed to further eliminate wavelengths in a competitive way. In our example, 360 random numbers over the interval [0,1] are generated to pick variables; variables with greater weights will be selected with higher frequency. At the end of this step, only those variables are kept that have selection frequencies of one or more.[5,9]

In the second sampling run the selection is more rigorous as based on a smaller value of *r*. In the end, m sets of selected spectral variables are then used for PLSR with the complete sample set. The lowest RMSE in the LOO cross validation indicates the optimal subset of variables.

Due to the use of random combinations of variables in the binary matrix (IRIV) and random numbers in the second selection step (CARS) both variable selection methods do not provide unique solutions. Thus we repeated the complete cycle of both CARS- and IRIV-PLSR 30 times; for each sample the obtained 30 estimates were averaged for the final estimates.

Both selection strategies were used to identify spectral key variables for OC and N, which was based on the respective selection frequencies obtained in all selection runs. As key variables we defined variables that amounted at least 2.5% to all selections for the respective soil property.

With pseudo-absorbance spectra, the search space consisted of 401 spectral variables, which allowed us to apply both selection techniques. Processing time with IRIV, however, is large compared to CARS and it increases markedly with an increasing number of variables. As the search space with wavelet-transformed spectra increased to 2005 variables (401 variables for each of the five different scales), only CARS was applied to select appropriate wavelet coefficients in this large search space which then were used for the model calibration with PLSR. In this CWT-CARS-PLSR approach, again 50 sampling runs and 30 repetitions were performed to obtain the final estimates for OC and N.

Multivariate modelling results were evaluated with the coefficient of determination (R^2 for cross validation and r^2 for validation samples), the residual prediction deviation (*RPD*, defined as the ratio of standard deviation of the reference values to standard error of the predicted values), and the *rRMSE* (relative *RMSE* = *RMSE* × measured arithmetic mean^{-1}). Results were also ranked according to the guideline of Saeys *et al.*[29] with the following order: excellent (RPD > 3.0 and R^2 > 0.91); good (RPD between 3.0-2.5 and R^2 between 0.91-0.82); approximate quantitative (RPD between 2.5-2.0, R^2 between 0.81-0.66); and being able to distinguish between high and low values (RPD between 2.0-1.5 and R^2 between 0.65-0.50). All smaller values of RPD and R^2 indicate an unsuccessful estimation. In cases where R^2 and RPD suggested a different accuracy level, we chose the weaker one.

Definition of sample sets

Studied soil parameters (Table 1) and also measured spectra (Figure 3) indicate rather strong overlaps between the four geographically stratified regions (Eifel, Moselle, Hunsrück, Luxembourg). For the spectral data, a principal component analysis (PCA) was performed, in which the first two principal components explained 92.6 % of the total variance. In this feature space, the scores of the Eifel and Moselle samples provided a patch being more homogeneous than that of the Hunsrück or the Luxembourg samples (Figure 3). Based on these results, we compiled four different datasets each with a sufficient number of samples to be analysed in a separate calibration and validation approach. The first spectrally most homogeneous set comprised Eifel and Moselle samples (total number of samples n = 90), the second set additionally included the Luxembourg samples (n = 124), the third comprised the samples from Eifel, Moselle and Hunsrück (n = 138) and a fourth most heterogeneous set covered all 172 samples.

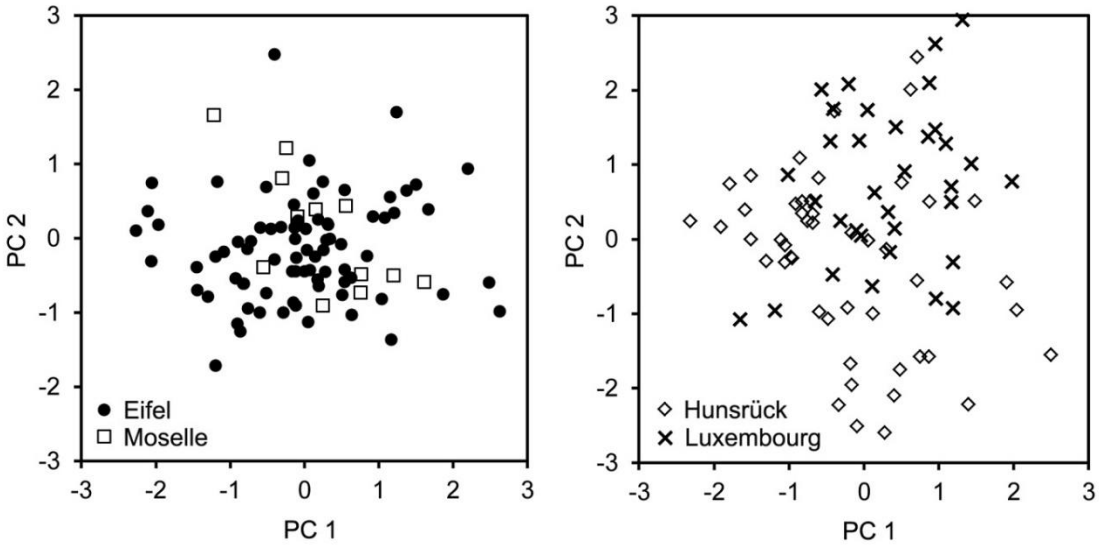


Figure 3: Plot of scores on the first (PC 1) and the second (PC 2) principal component derived from the spectra of all studied soil samples (n = 172), illustrated for the studied subregions of Eifel, Moselle, Hunsrück and Luxembourg.

Each set was split into a calibration set (with two-thirds of samples) and a validation set (one-third of samples) using the Kennard-Stone (KS) algorithm.[30] KS performs a sequential selection of samples to realise a uniform coverage over the feature space (in our case the spectral space made up by five PCs which explained 99.0 % of the total variance) and to include samples on the boundary of the data set. It starts with the selection of a pair of samples for which the Euclidean distance is the largest. At each subsequent stage, the sample with the largest minimum distance to any sample already selected is added until the pre-defined number of samples is reached.[31,32] Accordingly, the Eifel-Moselle set was split to 60 calibration and 30 validation samples, the Eifel-Moselle-Luxembourg set consisted of 83 calibration samples and 41 validation samples, the Eifel-Moselle-Hunsrück set contained 92 samples in the calibration set and 46 samples in the validation set and the complete set with all data was split into 115 calibration and 57 validation samples.

For calibration sets with markedly skewed soil data (skewness > 1.0), values were transformed with the natural logarithm prior to model calibration. In these cases (OC values of the Eifel-Moselle-Luxembourg and the Eifel-Moselle-Hunsrück-Luxembourg datasets), this yielded logarithmically transformed estimates that were back-transformed to the original scale for accuracy assessment.

3. Results

Results with spectral domain-PLSR and variable selection methods

OC and N were estimated for the differently sized datasets first with full spectrum-PLSR based on pseudo-absorbances and 401 spectral values per spectrum. For all datasets, prediction accuracies obtained for OC were higher than those for N. *RPD* for OC was > 2.0 (in the validation) for the Eifel-Moselle, the Eifel-Moselle-Luxembourg and the Eifel-Moselle-Hunsrück dataset, R^2 ranged for these data from 0.77 to 0.81 (approximate quantitative predictions). For the largest dataset, OC validation results dropped to RPD = 1.88 and R^2 = 0.72 (possibility to distinguish between high and low values). The best validation results for N were obtained with the Eifel-Moselle-Luxembourg set (RPD = 1.84 and R^2 = 0.73); for the other sets, validation results for N ranged from 1.36 to 1.72 (RPD) and 0.63 to 0.69 (R^2) (Table 2).

Table 2: Obtained accuracies with full spectrum-PLSR in the LOO-cross validation (cv) and the outer validation (val).

		OC[a]		N[b]	
		cv	val	cv	val
Eifel–Moselle	R^2	0.75	0.81	0.63	0.63
	RPD	1.99	2.27	1.62	1.36
	rRMSE	0.19	0.15	0.18	0.19
Eifel–Moselle–Luxembourg	R^2	0.77	0.77	0.66	0.73
	RPD	2.05	2.08	1.69	1.84
	rRMSE	0.19	0.17	0.19	0.15
Eifel–Moselle–Hunsrück	R^2	0.78	0.80	0.58	0.69
	RPD	2.10	2.22	1.51	1.72
	rRMSE	0.19	0.15	0.21	0.17
Eifel–Moselle–Hunsrück–Luxembourg	R^2	0.76	0.72	0.58	0.65
	RPD	2.03	1.88	1.54	1.62
	rRMSE	0.19	0.17	0.21	0.19

[a] number of latent variables l.V.: 11 (Eifel-Moselle, Eifel-Moselle-Luxembourg), otherwise 12

[b] l.V.: 12 for all datasets

R^2/r^2: coefficient of determination for cross validation (cv) and validation (val)

RPD: ratio of standard deviation of the reference values to standard error of the predicted values

r*RMSE*: relative root mean square error (*RMSE* × measured arithmetic mean^{-1})

With the additional application of CARS or IRIV, the majority of estimation results improved slightly in the validation and more markedly in the cross validation. With one exception of N

with IRIV for the Eifel-Moselle-Hunsrück dataset, LOO-cross validated estimates were superior to validation results in terms of statistical measures (Table 3). Validation results improved most distinctly for OC in the Eifel-Moselle-Luxembourg dataset when using CARS, with *RPD* now greater than 2.50 and r^2 at 0.85 (indicating a good instead of an approximate quantitative prediction). For N, *RPD* improved to values > 1.50 in the Eifel-Moselle data (with both CARS and IRIV) and, with CARS, to a value > 2.0 in the Eifel-Moselle-Luxembourg dataset (Table 3). For the Eifel-Moselle-Hunsrück-Luxembourg dataset, estimates for OC could not even slightly be improved with CARS or IRIV.

Validation accuracies obtained with CARS and IRIV were very similar. Marked differences were only found for N in the Eifel-Moselle-Luxembourg data (CARS > IRIV) and, less pronounced, in the Eifel-Moselle-Hunsrück-Luxembourg data (IRIV > PLSR) (Table 3). IRIV and CARS also selected a similar number of spectral variables, ranging from 17 to 41 at maximum. The number of latent variables decreased slightly with both approaches compared to full spectrum-PLSR (Table 3).

Selection frequencies and selection patterns were studied in detail for both selection approaches. In case of OC (Figure 4), we found for the example of the Eifel-Moselle data some spectral regions that were prominent with both methods (even if peaks were pronounced differently). These were some wavelengths in the visible (most striking at 420 nm), the region from 1830 to 1905 nm (near the water combination band at 1915-1940 nm) and between 2205 and 2270 nm (near the hydroxyl band at 2200 nm). At some additional prominent peaks for IRIV (e.g. at 970 and 1680 nm) variables were also selected by CARS but markedly less often.

Selection patterns shifted when we changed the sample set. For the complete data (Figure 4), spectral values in the visible range and near the water and hydroxyl bands were again relevant. However, CARS showed also pronounced peaks in the region beyond 2250 nm which had no importance for OC in the smaller Eifel-Moselle dataset. For IRIV, a prominent additional peak was now found at about 1535 nm (Figure 4, Annex 1). For all datasets the most important (key) wavelengths (as identified by CARS and IRIV) are in detail listed for both OC and N in Annex 1. Here we found clear overlaps between OC and N, most pronounced at 420 nm, 1870-1875 nm and 2210-2220 nm.

Results with wavelet transformed spectra

Possible benefits of the wavelet approach relate to the spread of relevant information over different scales, which – after the wavelet transformation of the spectra – may be selected in the multi-level search space with an appropriate algorithm and then be used for the multivariate calibration. To analyse the actual spread of information, we correlated the calculated wavelet coefficients at all different scales with OC values for the example of the calibration set with 115 samples (Eifel-Moselle-Hunsrück-Luxembourg) (Figure 5; for clarification, Pearson correlation coefficients were squared). Highest r^2 values were thus found in the longwave region beyond 2250 nm at the first two scales. The other scales, however, also showed prominent regions with high values (e.g. nearby 2200 nm at the 2^2

Table 3: Validation and cross validation (cv) results obtained with both applied variable selection strategies (CARS-PLSR vs. IRIV-PLSR).

Soil Group		OC CARS	OC IRIV	N CARS	N IRIV
Eifel–Moselle	R^2	0.84	0.85	0.76	0.77
		0.86 (cv)	0.89 (cv)	0.85 (cv)	0.87 (cv)
	RPD	2.47	2.48	1.80	1.82
		2.68 (cv)	3.04 (cv)	2.63 (cv)	2.78 (cv)
	rRMSE	0.14	0.14	0.14	0.14
		0.14 (cv)	0.12 (cv)	0.11 (cv)	0.11 (cv)
	var/l.V.a	35/9.9	32/10.8	39/11.0	32/11.7
Eifel–Moselle– Luxembourg	R^2	0.85	0.82	0.82	0.67
		0.89 (cv)	0.86 (cv)	0.83 (cv)	0.79 (cv)
	RPD	2.57	2.35	2.34	1.74
		2.63 (cv)	2.63 (cv)	3.47 (cv)	2.19 (cv)
	rRMSE	0.14	0.15	0.12	0.16
		0.15 (cv)	0.15 (cv)	0.13 (cv)	0.14 (cv)
	var/l.V.a	17/10.8	19/8.8	28/11.6	24/9.3
Eifel–Moselle– Hunsrück	R^2	0.82	0.82	0.71	0.72
		0.86 (cv)	0.87 (cv)	0.74 (cv)	0.71 (cv)
	RPD	2.36	2.40	1.74	1.81
		2.80 (cv)	2.91 (cv)	1.97 (cv)	1.85 (cv)
	rRMSE	0.14	0.14	0.16	0.16
		0.14 (cv)	0.13(cv)	0.16 (cv)	0.17 (cv)
	var/l.V.a	32/10.4	31/11.1	27/11.0	28/10.9
Eifel–Moselle–Huns-rück–Luxembourg	R^2	0.70	0.72	0.65	0.71
		0.85 (cv)	0.85 (cv)	0.73 (cv)	0.74 (cv)
	RPD	1.80	1.91	1.67	1.84
		2.56 (cv)	2.57 (cv)	1.93 (cv)	1.98 (cv)
	rRMSE	0.18	0.17	0.18	0.17
		0.15 (cv)	0.15 (cv)	0.17 (cv)	0.16 (cv)
	var/l.V.[a]	23/10.6	41/10.5	25/11.0	35/11.9

[a]var= number of selected spectral variables, l.V. = number of latent variables (both averaged over 30 runs; var: rounded value)

CARS-PLSR: competitive adaptive reweighted sampling-partial least squares regression
IRIV-PLSR: iteratively retaining informative variables-partial least squares regression
R^2 / r^2: coefficient of determination for cross validation (cv) and validation (val)
RPD: ratio of standard deviation of the reference values to standard error of the predicted values
r*RMSE*: relative root mean square error (*RMSE* × measured arithmetic mean^{-1})

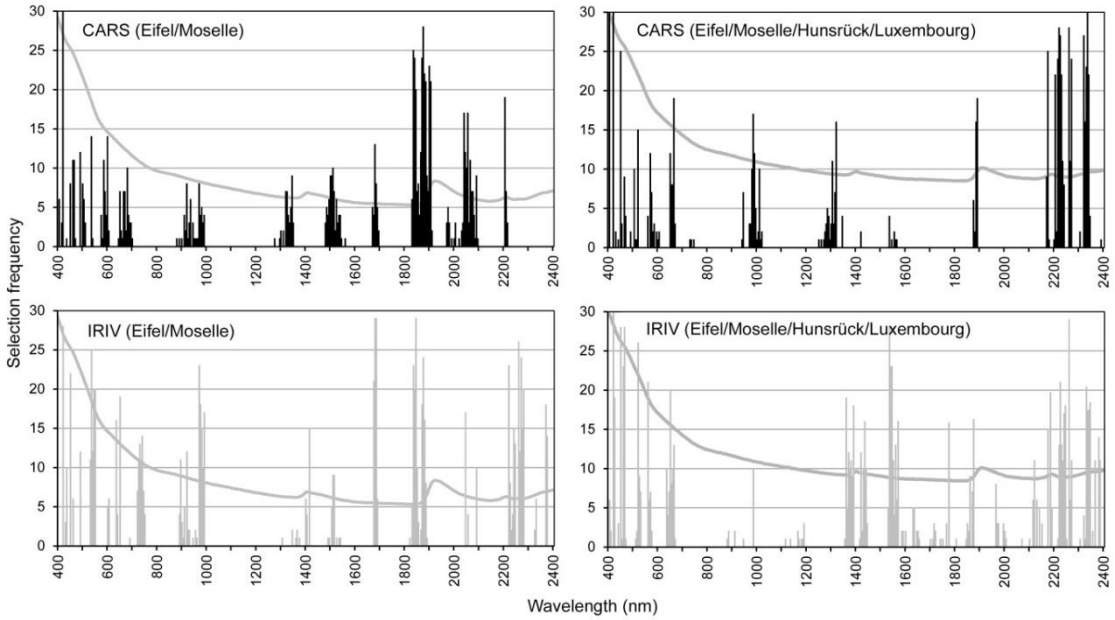


Figure 4: Selection frequencies of CARS-PLSR and IRIV-PLSR for estimating OC in two different sample sets (mean spectra for each sample set in the background and scaled for illustration).

and 2^3 scale or at about 1800 nm at the 2^5 scale) indicating information being potentially useful when assessing OC with decomposed spectra. This indicates, although redundancy between the different scales was also visible, the necessity for a multi-level selection procedure.

In the following, CARS was applied to individual scales separately, and also to the coefficients of several and all scales stacked together. Table 4 documents the main results of this CWT-CARS-PLSR analysis for OC.

In the cross validation, best results were obtained for all geographic datasets with the coefficients of the first (2^1) scale; accuracies decreased in all cases with an increase of the level. CARS-PLSR provided, when applied to the 2^1 scale coefficients, R^2 values generally ≥ 0.90 and RPD values > 3.0, i.e., good to excellent results. The results distinctly outperformed the cross validation results of full-spectrum PLSR (RPD values between 1.99 and 2.10, Table 2) and CARS- or IRIV-PLSR (2.80 and 3.04 as highest RPD values, Table 3).

In the validation, highest accuracies were achieved with a multi-level selection of coefficients (Table 4). Only for the Eifel-Moselle-Hunsrück set, accuracy obtained with 2^2 scale coefficients equalled that with 2^1–2^2 data. Obtained accuracies decreased from the smallest and most homogenous dataset (Eifel-Moselle with excellent predictions; R^2 = 0.93, RPD = 3.60) to the largest and most heterogeneous set (Eifel-Moselle-Hunsrück-Luxembourg, approximate quantitative predictions; R^2 = 0.82, RPD = 2.31).

Compared to accuracies achieved with full spectrum-PLSR, better ranking categories were obtained for all datasets, i.e., excellent or good instead of approximate quantitative and approximate quantitative instead of the possibility to discriminate between high and low values (Table 2, Table 4). Improvements towards CARS-PLSR and IRIV-PLSR were also

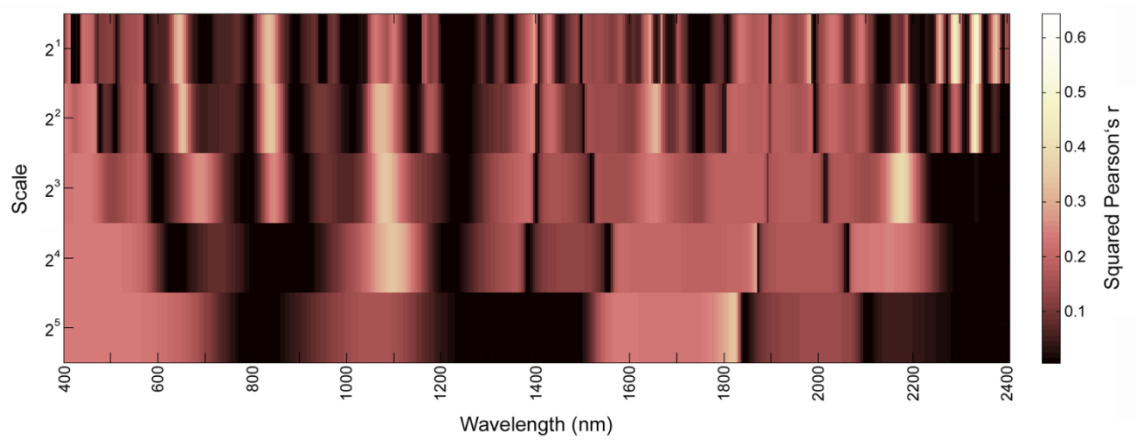


Figure 5: Squared Pearson's *r* between CWT coefficients and OC (n = 115; Eifel-Moselle-Hunsrück-Luxembourg calibration set) for each wavelength and scale.

marked, as the effect of both methods had been limited compared to full spectrum-PLSR. Smallest improvements with CWT-CARS-PLSR were obtained with the Eifel-Moselle-Hunsrück data (RPD = 2.80 compared to 2.40 with IRIV-PLSR; r^2 = 0.87 instead of 0.82), similar improvements were achieved with the largest dataset (Eifel-Moselle-Hunsrück-Luxembourg). Improvements were most marked with the Eifel-Moselle data (RPD = 3.60 instead of 2.48 with IRIV-PLSR, r^2 = 0.93 instead of 0.85) (Table 3, Table 4).

Results are illustrated for the smallest and the largest dataset (Figure 6). For the Eifel-Moselle data, scattering around the 1:1-line decreased distinctly for the samples with less than 3% OC when CWT and CARS were used instead of full-spectrum PLSR. Although visible, these effects were limited with the Eifel-Moselle-Hunsrück-Luxembourg data. However, the labelling of all data points according to their geographic origin did not indicate one of the regions to be out of line with the general trend.

For N, CWT-CARS-PLSR outperformed all other approaches in the cross validation with *RPD* values > 2.0 for all sample sets (at least approximate quantitative estimates, good and excellent estimates for Eifel-Moselle-Luxembourg and Eifel-Moselle data) (Table 5). Similar to OC, the data of the 2^1 scale or a combination of the first scales provided the best results. In the validation, CWT-CARS-PLSR outperformed all other approaches for the Eifel-Moselle and the Eifel-Moselle-Hunsrück-Luxembourg data and achieved approximate quantitative predictions. However, for the Eifel-Moselle-Hunsrück data, all approaches reached only an accuracy which allows distinguishing between high and low values (Table 2, Table 3 and Table 5). With the exception of the largest dataset, modelling results for N with CWT-CARS-PLSR were at least one category weaker than for OC (Table 4, Table 5).

Although the number of predictor variables increased fivefold by the wavelet decomposition, the CARS procedure again resulted in a rather low number of selected variables (the maximum for both soil parameters was 63 variables averaged over 30 repetitions). In terms of extracted factors, the maximum equalled 11.0 latent variables.

Table 4: Validation and cross validation (cv) results obtained for OC using coefficients of different scales (2^1–2^5) of wavelet decomposition combined with CARS-PLSR (best validation and cv results for each regional dataset marked in bold italic and in bold letters)

	Scale:	2^1	2^2	2^3	2^4	2^5	$2^1 – 2^2$	$2^1 – 2^3$	$2^1 – 2^5$
Eifel–Moselle	R^2	0.84	0.86	0.81	0.78	0.84	0.87	0.93 b	0.80
		0.94 (cv)	0.91 (cv)	0.84 (cv)	0.81 (cv)	0.72 (cv)	0.92 (cv)	0.90 (cv)	0.86 (cv)
	RPD	2.07	2.66	2.27	2.11	2.46	2.63	3.60 b	2.23
		4.21 (cv)	3.34 (cv)	2.51 (cv)	2.33 (cv)	1.89 (cv)	3.58 (cv)	3.17 (cv)	2.74 (cv)
	rRMSE	0.17	0.13	0.15	0.17	0.14	0.13	0.10 b	0.16
		0.09 (cv)	0.11 (cv)	0.15 (cv)	0.16 (cv)	0.20 (cv)	0.11 (cv)	0.12 (cv)	0.14 (cv)
	var/l.V[a]	35/10.4	47/11.0	55/10.1	44/9.5	60/10.7	44/9.5	63/10.9	24/10.5
Eifel–Moselle– Luxembourg	R^2	0.85	0.83	0.86	0.84	0.86	0.86	0.89	0.90
		0.92 (cv)	0.90 (cv)	0.86 (cv)	0.81 (cv)	0.75 (cv)	0.91 (cv)	0.92 (cv)	0.89 (cv)
	RPD	2.48	2.12	2.62	2.45	2.71	2.45	2.90	3.10
		3.63 (cv)	3.16 (cv)	2.67 (cv)	2.31 (cv)	2.01 (cv)	3.37 (cv)	3.54 (cv)	3.05 (cv)
	rRMSE	0.15	0.17	0.14	0.15	0.13	0.15	0.12	0.12
		0.11 (cv)	0.12 (cv)	0.15 (cv)	0.17 (cv)	0.19 (cv)	0.12 (cv)	0.11 (cv)	0.13 (cv)
	var/l.V[a]	37/10.6	29/10.6	44/9.6	48/9.6	15/10.7	37/10.4	33/10.1	32/9.1
Eifel–Moselle– Hunsrück	R^2	0.83	0.87	0.80	0.71	0.75	0.87	0.81	0.84
		0.90 (cv)	0.88 (cv)	0.87 (cv)	0.80 (cv)	0.75 (cv)	0.89 (cv)	0.89 (cv)	0.86 (cv)
	RPD	2.40	2.80	2.22	1.82	1.95	2.80	2.27	2.48
		3.26 (cv)	3.04 (cv)	2.84 (cv)	2.33 (cv)	2.10 (cv)	3.21 (cv)	3.08 (cv)	2.85 (cv)
	rRMSE	0.14	0.12	0.15	0.18	0.17	0.12	0.15	0.13
		0.12 (cv)	0.13(cv)	0.14 (cv)	0.17 (cv)	0.19 (cv)	0.12 (cv)	0.13 (cv)	0.14 (cv)
	var/l.V[a]	33/10.0	28/10.1	31/9.6	53/10.0	15/9.3	42/10.3	37/10.2	62/9.6
Eifel–Moselle–Hunsrück–Luxembourg	R^2	0.75	0.81	0.80	0.72	0.69	0.82	0.81	0.78
		0.90 (cv)	0.86 (cv)	0.83 (cv)	0.75 (cv)	0.71 (cv)	0.88 (cv)	0.87 (cv)	0.83 (cv)
	RPD	1.95	2.19	2.26	1.89	1.79	2.31	2.25	2.13
		3.11 (cv)	2.67 (cv)	2.42 (cv)	2.02 (cv)	1.85 (cv)	2.83 (cv)	2.77 (cv)	2.45 (cv)
	rRMSE	0.17	0.15	0.14	0.17	0.18	0.14	0.14	0.15
		0.13 (cv)	0.15 (cv)	0.16 (cv)	0.19 (cv)	0.21 (cv)	0.14 (cv)	0.14 (cv)	0.16 (cv)
	var/l.V[a]	41/10.7	33/10.5	23/9.9	18/8.8	36/10.5	41/10.8	28/10.1	19/10.1

[a]var= number of selected spectral variables, l.V. = number of latent variables (both averaged over 30 runs; var: rounded value)
[b]Eifel–Moselle: similar validation accuracies were obtained with $2^1 – 2^2$, 2^5

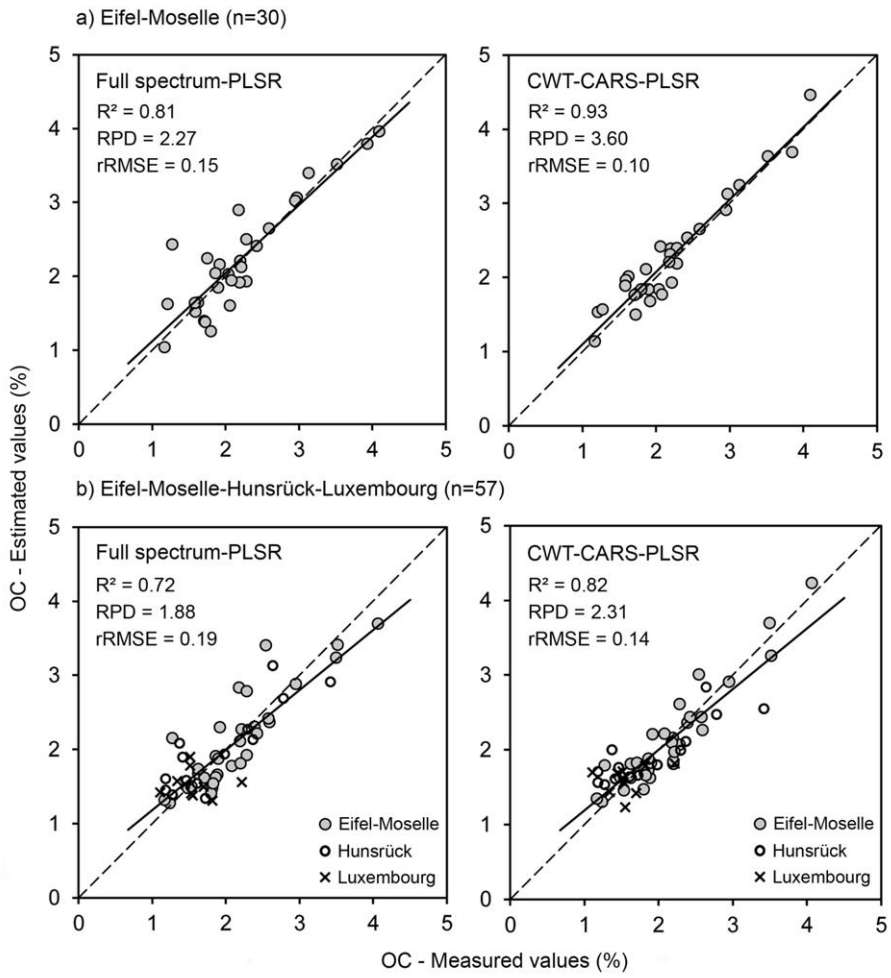


Figure 6: Estimated vs. measured values obtained with full spectrum-PLSR and CWT-CARS-PLSR for the validation samples of the Eifel-Moselle (a) and the Eifel-Moselle-Hunsrück-Luxembourg (b) set.

4. Discussion

A large number of studies exist that also used vis-NIR data to calibrate and predict OC of soils based on dried and ground samples. Results of these studies were summarized, for example, by Bellon-Maurel and McBratney[33], who reported a range of *RPD* values from 1.2 to 5.4. The accuracies that we finally obtained here (Table 4) were in the middle range of these comparative values. However, direct comparisons are complicated by the fact that, in general, calibration results vary depending on the considered spatial scale, the heterogeneity of the included soil samples, the size of the sample sets and the type of collected soil samples (topsoil samples vs. soil cores). The predictions that we obtained for N were generally less accurate than those for OC.

To improve estimation accuracies for both OC and N, we tested variable selection techniques with limited success, at least for the validation results. At this point we have to be aware that selections with both methods were optimised in the cross validation procedure, so that improvement of the cross validated results was more marked than that in the validation. CARS and IRIV provided very similar results, none of the methods outperformed the other regularly. Different from this, Jung *et al.*[34] found IRIV-PLSR to perform slightly better than

CARS-PLSR when assessing OC, N, hot water-extractable carbon and clay content from soil spectra; their analysis, however, was restricted to a cross validation approach. Yun *et al.*[7] also tested both approaches with different benchmark datasets (diesel fuels, corn and soy data) and found IRIV to outperform CARS in both the cross validation and the outer validation, which was attributed to the IRIV strategy to consider the synergetic effects among the spectral variables.

Table 5: Best CWT-CARS-PLSR modelling results obtained for N in the validation and cross validation (cv) for each regional dataset.

	Eifel–Moselle		Eifel–Moselle–Luxembourg		Eifel–Moselle–Hunsrück		Eifel–Moselle–Huns-rück–Luxembourg	
	cv	val	cv	val	cv	val	cv	val
R^2	0.92	0.84	0.87	0.80	0.80	0.73	0.78	0.78
RPD	3.56	2.03	2.80	2.22	2.25	1.93	2.16	2.13
rRMSE	0.08	0.13	0.11	0.13	0.14	0.15	0.15	0.14
var/l.V[a]	40/10.8	45/10.9	31/9.6	31/9.6	36/10.0	26/10.0	34/10.1	26/10.6
scale	2^1	2^2	$2^1 – 2^2$	$2^1 – 2^2$	$2^1 – 2^3$	$2^1 – 2^2$	2^1	$2^1 – 2^2$

[a]var= number of selected spectral variables, l.V. = number of latent variables (both averaged over 30 runs; var: rounded value)

CWT-CARS-PLSR: continuous wavelet transformation-competitive adaptive reweighted sampling-partial least squares regression
r^2/R^2: coefficient of determination for validation (val) and cross validation (cv)
RPD: ratio of standard deviation of the reference values to standard error of the predicted values r*RMSE*: relative root mean square error (*RMSE* × measured arithmetic mean^{-1})

In our study, we identified the processing time to be the most important difference between both methods. Processing with CARS-PLSR (and 401 spectral variables) needed about one minute to complete 30 selection cycles (when using a desktop PC with a 64-Bit-CPU, 3.4 GHz and 12 GB RAM), whereas IRIV took some hours. This is a considerable limitation for the application of IRIV, at least for datasets with some hundreds of variables.

Concerning the selected key variables, we found e.g. the hydroxyl band to be relevant for estimating OC for all geographic datasets. Bands in the visible (at 420 nm, for example) or near the main water absorption band at 1915 nm were also selected often. These findings coincide with other studies.[35-37] However, the overall selection pattern changed from one dataset to another. We attribute this to the different sub-regional provenance and, linked to this, the different heterogeneity of the studied sample sets. Spectrally active soil constituents interact in a complex way or may mask each other. Therefore spectral predictive mechanisms vary depending on factors such as soil texture, soil color, the content and humification degree of soil organic matter or the content of carbonates.[33,38] .This also partially explains, why different studies with different sample sets (and an often different data pre-processing) have identified a wide variety of wavelengths associated with OC (see overview e.g. in Cécillon *et al.*[1]).

As also reported in literature[5,39], key wavelengths of OC and N showed a strong overlap. This is consistent, considering that numerous bonds between C and O, N or H absorb in the NIR region[1] and that N and OC are usually highly correlated, which also applies for our data. The OC and N of all 172 samples were highly significantly correlated with a Pearson correlation coefficient of 0.88.

The CWT-CARS-PLSR approach improved predictions especially for the Eifel-Moselle (OC and N) and the Eifel-Moselle-Luxembourg (OC) datasets. The selection of coefficients from different scales was beneficial for the modelling (instead of using only one scale). This coincides with available soil studies using DWT[21-23], in which coefficients from the different decomposition levels (detail and/or approximation components) were integrated in the respective multivariate modelling approaches. Viscarra Rossel and Lark[23] estimated OC from vis-NIR spectra for a sample set with a relatively wide range of OC contents (0.01–13.90%). Using wavelets, absolute *RMSE* reduced from 1.08% (with spectral domain PLSR) to 0.86%, which equals a reduction of about 20%. For SOM, Lin *et al*. reached a reduction of *RMSE* from 7.74 g × kg^{-1} to 5.85 g × kg^{-1} (improvement of about 24%). The results we obtained were similar to these values. In the best case (Eifel-Moselle), absolute RMSE reduced from 0.34% to 0.22% (improvement of 35%); smallest reduction was 19% (0.28% absolute RMSE instead of 0.35%) for the Eifel-Moselle-Hunsrück-Luxembourg data

Peng *et al*. listed both calibration and validation results for a series of models using different inputs from the wavelet analysis. Similar to our study, the best model in the cross validation (for OC it was, in our case, 2^1 scale model) did not correspond to the model that was best suited for prediction purposes. This is of course crucial as prediction values are normally unknown and cannot be used for model selection. Thus, the calibration procedure should be, if possible, modified with a split into three sets, a training set for model definition, a validation set and a prediction set.[40] However, this demands a sufficiently large number of available samples. On this point, the large spectral databases of continental or global soil libraries[41,42] could be of great use.

5. Conclusion

We studied four soil sample sets of different size and heterogeneity sampled in a region with an area of approximately 600 km^2. For the prediction of OC and N from vis–NIR spectra we found PLSR combined with spectral variable selection to be only slightly superior to full spectrum-PLSR. However, IRIV was, in terms of operationality, markedly more limited than CARS due to very long processing times. Thus, CARS is more appropriate as a fast and “easy-to-use” variable selection strategy for multivariate modelling approaches with spectroscopic data.

For the accuracy of predictions it was beneficial to transform spectra with CWT and to apply CARS-PLSR to the obtained wavelet coefficients instead of original spectral values. To achieve marked improvements, it was necessary to select coefficients from different scales and to use them in the PLSR calibration approach. Best results with excellent predictions were obtained for OC in the smallest and most homogeneous data set. For the largest set we

retrieved only approximate quantitative prediction. This may refer to heterogeneity of sample sets as critical factor for the estimation of soil properties from spectral data.

Acknowledgements

This study was funded by the Deutsche Forschungsgemeinschaft (DFG VO 1509/3-1, TH 678/12-1). Calculations were carried out in Matlab (R2014b), using also the freely available Matlab codes CARS (Copyright Hongdong Li, 2011; http://code.google.com/p/carspls/) and IRIV (IRIV_1.1, Yong-Huan Yun et al., 2013; http://code.google.com/p/multivariate-calibration/ downloads/list). We would like to thank Christopher Hutengs (Geoinformatics and Remote Sensing, Leipzig University) for his support in data processing.

References

1. L. Cécillon, B.G. Barthès, C. Gomez, D. Ertlen, V. Genot, M. Hedde, A. Stevens and J.J. Brun, “Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS)”, Eur. J. Soil Sci. 60, 770 (2009). doi: 10.1111/j.1365-2389.2009.01178.x
2. J.M. Soriano-Disla, L.J. Janik, R.A. Viscarra Rossel, L.M. MacDonald and M.J. McLaughlin, M.J., “The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties”, Appl. Spectrosc. Rev. 49, 139 (2014). doi: 10.1080/05704928.2013.811081
3. S. Wold, M. Sjöström and L. Eriksson, "PLS-regression: a basic tool of chemometrics," Chemometr. Intell. Lab. Syst. 58, 109 (2001). doi: 10.1016/S0169-7439(01)00155-1
4. B.-C. Deng, Y.-H. Yun, P. Ma, C.-C. Lin, D.-B. Ren and Y.-Z. Liang, “A new method for wavelength interval selection that intelligently optimizes the locations, widths and combinations of the intervals”, Analyst 140, 1876 (2015). doi: 10.1039/C4AN02123A
5. M. Vohland, M. Ludwig, S. Thiele-Bruhn and B. Ludwig, “Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection”, Geoderma 223-225, 88 (2014). doi: 10.1016/j.geoderma.2014.01.013
6. Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes and M. Hanpin, “Variables selection methods in near-infrared spectroscopy”, Anal. Chim. Acta 667, 14 (2010). doi: 10.1016/j.aca.2010.03.048
7. Y.-H. Yun, W.-T. Wang, M.-L. Tan, Y.-Z. Liang, H.-D. Li, D.-S. Cao, H.-M. Lu and Q.-S. Xu, “A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration”, Anal. Chim. Acta 807, 36 (2014). doi: 10.1016/j.aca.2013.11.032
8. V. Centner, D.-L. Massart, O. De Noord, S. De Jong, B.M. Vandeginste and C. Sterna, “Elimination of uninformative variables for multivariate calibration”, Anal. Chem. 68, 3851 (1996). doi: 10.1021/ac960321m
9. H. Li, Y. Liang, Q. Xu and D. Cao, “Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration”, Anal. Chim. Acta 648, 77 (2009). doi: 10.1016/j.aca.2009.06.046
10. M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame and V. Visani, “The successive projections algorithm for variable selection in spectroscopic multicomponent analysis”, Chemometr. Intell. Lab. Syst. 57, 65 (2001). doi: 10.1016/S0169-7439(01)00119-8
11. R. Leardi and A.L. González, “Genetic algorithms applied to feature selection in PLS regression: how and when to use them”, Chemometr. Intell. Lab. Syst. 41, 195 (1998). doi: 10.1016/S0169-7439(98)00051-3

12. J.H. Kalivas, N. Roberts and J. Sutter, “Global optimization by simulated annealing with wavelength selection for ultraviolet-visible spectrophotometry”, Anal. Chem. 61, 2014 (1989). doi: 10.1016/0169-7439(92)80022-V.
13. A.Banskota, R.H. Wynne and N. Kayastha, “Improving within-genus tree species discrimination using the discrete wavelet transform applied to airborne hyperspectral data”, Int. J. Rem. Sens. 32, 3551 (2011). doi: 10.1080/01431161003698302
14. J.F. Huang and G.A. Blackburn, “Optimizing predictive models for leaf chlorophyll concentration based on continuous wavelet analysis of hyperspectral data”, Int. J. Rem. Sens. 32, 9375 (2011). doi: 10.1080/01431161.2011.558130
15. G.A. Blackburn and J.G. Ferwerda, “Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis”, Remote Sens. Environ. 112, 1614 (2008). doi: 10.1016/j.rse.2007.08.005
16. J.G. Ferwerda and S. Jones, “Continuous wavelet transformations for hyperspectral feature detection”, Proceedings of the 12th International Symposium on Spatial Data Handling, 12-14 July, University of Vienna, Austria (2006). doi: 10.1007/3-540-35589-8_11
17. R.L. Pu and P. Gong, “Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping”, Remote Sens. Environ. 91, 212 (2004). doi: 10.1016/j.rse.2004.03.006
18. Y. Huang, Q. Tian, l. Wang, J. Geng and C. Lyu, “Estimating canopy leaf area index in the late stages of wheat growth using continuous wavelet transform”, J. Appl. Remote Sens. 8, 083517 (2014). doi: 10.1117/1.JRS.8.083517
19. T. Cheng, B. Rivard and G.A. Sánchez-Azofeifa, “Spectroscopic determination of leaf water content using continuous wavelet analysis”, Remote Sens. Environ. 115, 659 (2011). doi: 10.1016/j.rse.2010.11.001
20. T. Cheng, B. Rivard, G.A. Sánchez-Azofeifa, J.-B. Féret, S. Jaquemoud and S.L. Ustin, “Deriving leaf mass per area (LMA) from foliar reflectance across a variety of plant species using continuous wavelet analysis”, ISPRS J. Photogramm. 87, 28 (2014). doi: 10.1016/j.isprsjprs.2013.10.009
21. J. Peng, H. Shen, S.W. He and J.S. Wu, “Soil moisture retrieving using hyperspectral data with the application of wavelet analysis”, Environ. Earth Sci. 69, 279 (2013). doi: 10.1007/s12665-012-1955-x
22. Y. Ge, C.L.S. Morgan, J.A. Thomasson and T. Waiser, “A new perspective to near-infrared reflectance spectroscopy: a wavelet approach”, Trans. ASABE 50, 303 (2007). doi: 10.13031/2013.22394
23. R.A. Viscarra Rossel and R.M. Lark, “Improved analysis and modelling of soil diffuse reflectance spectra using wavelets”, Eur. J. Soil Sci. 60, 453 (2009). doi: 10.1111/j.1365-2389.2009.01121.x
24. L. Lin, Y. Wang, J. Teng and X. Wang, “Hyperspectral analysis of soil organic matter in coal mining region using wavelets, correlations, and partial lest squares regression”, Environ. Monit. Assess. 188, 97 (2016). doi: 10.1007/s10661-0165107-8
25. X. Zhang, N.H.Younan and C.G. O’Hara, “Wavelet domain statistical hyperspectral soil texture classification”, IEEE T. Geosci. Remote 43, 615 (2005). doi: 10.1109/TGRS.2004.841476
26. G.M. Vasques, S. Grunwald and J.O. Sickman, “Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra”, Geoderma 146, 14 (2008). doi: 10.1016/j.geoderma.2008.04.007
27. T. Shi, Y. Chen, Y. Liu and G. Wu, “Visible and near-infrared reflectance spectroscopy—An alternative for monitoring soil contamination by heavy metals”, J. Hazard. Mat. 265, 166 (2014). doi: 10.1016/j.jhazmat.2013.11.059
28. P.S. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance. CRC Press (2002). doi: 10.1201/9781420033397
29. W. Saeys, A.M. Mouazen and H. Ramon, “Potential for onsite and online analysis of pig manure using visible and near infrared spectroscopy”, Biosyst. Eng. 91, 393 (2005). doi: 10.1016/j.biosystemseng.2005.05.001

30. R.W. Kennard and L.A. Stone, “Computer aided design of experiments”, Technometrics 11, 137 (1969). doi: 10.2307/1266770
31. W. Wu, B. Walczak, D.L. Massart, S. Heuerding, F. Erni, I.R. Last and K.A. Prebble, “Artificial neural networks in classification of NIR spectral data: design of the training set”, Chemometr. Intell. Lab. Syst. 33, 35 (1996). doi: 10.1016/0169-7439(95)00077-1
32. R.K.H. Galvão, M.C.U. Araújo, G.E. José, M.J.C. Pontes, E.C. Silva and T.C.B. Saldanha, “A method for calibration and validation subset partitioning”, Talanta 67, 736 (2005). doi: 10.1016/j.talanta.2005.03.025
33. V. Bellon-Maurel and A. McBratney, “Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives”, Soil Biol. Biochem. 43, 1938 (2011). doi: 10.1016/j.soilbio.2011.02.019
34. A.Jung, S. Thiele-Bruhn and M. Vohland, “Use of a portable camera for proximal soil sensing with hyperspectral image data”, Remote Sens. 7, 11434 (2015), doi: 10.3390/rs70911434
35. D. Cozzolino and A. Morón, “Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions”, Soil Till. Res. 85, 78 (2006). doi: 10.1016/j.still.2004.12.006
36. R.A. Viscarra Rossel and T. Behrens, “Using data mining to model and interpret soil diffuse reflectance spectra”, Geoderma 158, 46 (2010). doi: 10.1016/j.geoderma.2009.12.025
37. M. Vohland and C. Emmerling, “Determination of total soil organic C and hot water-extractable C from VIS-NIR soil reflectance with partial least squares regression and spectral feature selection techniques”, Eur. J. Soil Sci. 62, 598 (2011). doi: 10.1111/j.1365-2389.2011.01369.x
38. E. Ben-Dor, J.A. Irons and A. Epema, “Soil spectroscopy”, in Manual of Remote Sensing, Ed. By A. Rencz. John Wiley & Sons, Inc., New York, p-111 (1999).
39. R.C. Dalal and R.J. Henry, “Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry”, Soil Sci. Am. J. 50, 120 (1986). doi: 10.2136/sssaj1986.03615995005000010023x
40. K. Varmuza and P. Filzmoser, Introduction to multivariate statistical analysis in chemometrics. CRC Press, Boca Raton, FL, USA (2009). doi: 10.1201/9781420059496.
41. A.Stevens, M. Nocita, G. Tóth, L. Montanarella and B. van Wesemael, “Prediction of soil organic carbon at the European scale by visible and near infared reflectance spectroscopy”, PLOS ONE 8, e66409 (2013).doi:http://dx.doi.org/10.1371/journal.pone.0066409
42. R.A. Viscarra Rossel, T. Behrens, E. Ben-Dor, D.J. Brown, J.A.M. Dematte, K.D. Shepherd, Z. Shi, B. Stenberg, A. Stevens, V.A. Adamchuk, H. Aichi, B.G. Barthes, H.M. Bartholomeus, A.D. Bayer, M. Bernoux, K. Bottcher, L. Brodsky, C.W. Du, A. Chappell, Y. Fouad, V. Genot, C. Gomez, S. Grunwald, A. Gubler, C. Guerrero, C.B. Hedley, M. Knadel, H.J.M. Morras, M. Nocita, L. Ramirez-Lopez, P. Roudier, E.M. Rufasto Campos, P. Sanborn, V.M. Sellitto, K.A. Sudduth, B.G. Rawlins, C. Walter, L.A. Winowiecki, S.Y. Hong and W. Ji, “A global spectral library to characterize the world’s soil”, Earth-Sci. Rev. 155, 198 (2016). doi: http://dx.doi.org/10.1016/j. earscirev.2016.01.012

Annex Table 1: Key wavelengths found by CARS and IRIV for the assessment of OC and N (semicolons separate different spectral regions, visible, nIR ≤ 1300 nm and long wavelength nIR > 1300 nm; overlaps between CARS and IRIV in bold)

Region		Method	Key wavelengths
OC	Eifel–Moselle	CARS	**420**; –; **1835-1840**, **1870-1885**, 1900-1905
		IRIV	**420**, 450, 535; 970; 1675-1685, **1835-1845**, **1875**, 2260-2270
	Eifel–Moselle–Luxembourg	CARS	445-450, 465, 500, 560, 570; –; 1880-1885, 2090, 2215, **2330**-2335
		IRIV	520; 880, 1140-1145, 1280-1285; 1860, 2175, 2205, **2330**
	Eifel–Moselle–Hunsrück	CARS	–; –; 1875, 2205-**2220**, **2330-2345**
		IRIV	575, 650; –; 1665, 2055, **2220**, 2260-2270, **2335**, 2375
	Eifel–Moselle–Hunsrück–Luxembourg	CARS	**405**, **420**, 445-**450**, 665; –; **2175**, 2205-2230, **2260**-2270, **2320-2340**
		IRIV	400-**405**, **420**, **450**, 465, 520, 560; –; 1535-1540, **2175**, **2260**, **2320**, **2340**
N	Eifel–Moselle	CARS	**420**, 465; –; **1870**
		IRIV	**420**, 460, 535, 560; –; 1540, **1870**, 2000, 2045, 2075, 2215, 2355
	Eifel–Moselle–Luxembourg	CARS	420, 445-450, 465, 500-505; –; **1870**, 1885-1895, 2070, 2080-**2105**, **2395**
		IRIV	695; –; 1385, 1415, 1595, 1855-**1870**, **2105**, 2125, **2395**
	Eifel–Moselle–Hunsrück	CARS	445, 665; –; **1870**, 1890-1895, **2210-2220**, **2355-2360**
		IRIV	420, 505, 650; –; **1870**, 2105, 2125-2130, 2175, **2210-2220**, **2355-2360**
	Eifel–Moselle–Hunsrück–Luxembourg	CARS	405, **420**, 665; –; 1890-1895, **2260**
		IRIV	**420**, 475, 560; –; 2215, **2260**-2290, 2310, 2335

4 Spatial and temporal changes of the soil microbial community and labile organic matter fractions affect the Vis-NIR spectroscopic quantification of organic matter in arable topsoils

M. Ludwig, M. Vohland, M. Harbich, T. Udelhoven, C. Emmerling, S. Thiele-Bruhn

European Journal of Soil Science (2017) submitted

Abstract

Knowledge about the quality of soil organic matter (SOM) is essential for managing agro-ecosystems. Hence, there is widespread interest to assess biochemical parameters and bioindicators with quick and cheap scanning methods such as visible to near infrared (Vis-NIR) spectroscopy of soil surfaces. However, living and non-living OM are not constant but contents show variation over time and with soil depth, which may bias findings derived from analysis of the soil surface. Hence, this study focused on the spatial and temporal variability of selected OM pools in arable soils as affected by microbial community variation in time and with soil depth. More specifically it was aimed to assess the potential of Vis-NIR spectroscopy to determine these pools and vice versa to identify the impact of microbial variations on Vis-NIR spectral estimates. A one-year monitoring of SOM at 21 temperate arable sites was performed; organic C and N, microbial C (MBC), hot water extractable C (HWEC), dissolved organic carbon (DOC), chlorophyll α (Chl α) and phospholipid fatty acids (PLFA) were analyzed separately for three soil depths (0-1 cm; 1-5 cm; 5-10 cm). The monthly sampling was accompanied by in situ Vis-NIR spectroscopic measurements at all three soil depths. Results show that soil algae are the most significant spatio-temporally changing SOM pool, which is correlated to short-term changes of labile carbon fractions such as MBC and DOC. The prediction accuracy of MBC and DOC from spectral data depended on algal abundance, which served as a spectral proxy for the prediction. Algal correlated maxima of labile SOM fractions at the soil surface were not indicative for the SOM contents of the layers below; spectroscopic determination of total OC and PLFA markers was not successful. The potential of Vis-NIR to estimate SOM and SOM pools is yet not fully exploited. In case of biocrusts at the soil surface, the potential is indicated to be limited regarding predictions of total SOM content, but high regarding short-term changes of topsoil OM pools.

1. Introduction

There is a growing demand of regional to worldwide approaches to monitor the global soil status (Grunwald et al. 2011; Nocita et al, 2015). Especially the contents and properties of soil organic matter (SOM) must be recorded, since SOM is essential for maintaining soil fertility and other ecosystem services of soil. Yet a decline in the content and quality of SOM progresses in the last decades in temperate arable soils (Schulze and Freibauer 2005). Hence, accurate, high-resolution soil data is essential for environmental monitoring and soil quality assessment across spatial and temporal scales (Viscarra Rossel et al. 2006; Grunwald et al. 2011). Assessment methods are needed that work large-scale and that are at the same time selective and sensitive enough to determine, localize and quantify SOM and SOM fractions (Bellon-Maurel and McBratney 2011). To this end, visible to near infrared (Vis-NIR) spectroscopy is increasingly applied as a quick and cost-effective scanning method, it can be used on large-scales by enabling non-invasive analysis of the soil surface.

However, there are still relatively few examples of its routine use, despite its potential of rapid low-cost soil analysis (Nocita et al, 2015; Clairotte et al. 2016). This may partly be due to the fact that the potential to predict SOM contents of whole horizons and soil profiles from spectral analysis of soil surfaces is questionable, because largest environmental fluctuations of temperature, moisture and light make the highly dynamic soil surface a rather unique sub-compartment of soil profiles (Hillel 1982; Jeffery et al. 2009). Thus, a strong variability in quality and quantity of SOM with soil depth as well as with time has been reported (Moll et al. 2015). Especially for arable soils the very first centimeter of the soil surface is of high functional relevance, being the biologically most active zone with increased microbial biomass concentrations (Nunan et al. 2003; Jeffery et al. 2007) that tend to aggregate and form biological soil crusts (Hodge et al. 1998; Ekschmitt et al. 2005; Kuzyakov and Blagodatskaya 2015). Microbial communities at soil surfaces are functionally important regarding water infiltration, gas exchange, erodibility and contribute significantly to nutrient and SOM pools and availability (Harper and Belnap 2001; Auzet et al. 2004; Jeffery et al. 2009; Bowker et al. 2010). Particularly soil surfaces dominated by soil algae and other photoautotrophic microorganisms that tend to form biological soil crusts even in temperate climates (Belnap et al. 2001; Schulz et al. 2016) are assumed to be a significant carbon and nitrogen source for temperate arable soils (Shimmel and Darley 1985; Zancan et al. 2006; Schmidt et al. 2016) since they can directly fix carbon dioxide and nitrogen from the atmosphere (Belnap et al. 2001; Elbert et al. 2012). Yet, the algal contribution to SOM and soil functioning is still widely unknown (Langhans et al. 2009) and the spatial and temporal distribution near the soil surface of microorganisms in general have been poorly studied for temperate, arable soils (Jeffery et al. 2009; Büdel et al. 2014). Consequently, the spatial and

temporal variation of SOM content, properties and depth distribution and its dependence on the microbial community needs to be determined to provide soil quality status information required for sustainable agro-ecosystem management (Franklin and Mills 2009; Scharroba et al. 2015; De Paul Obade and Lal 2016).

Studies investigating the potential of spectroscopic methods to assess soil microbial properties and related labile SOM fractions are promising. For example, Barthès et al. (2011) found NIR spectroscopy to be useful for the prediction of nematofauna abundance and composition. The study of Zornoza et al. (2008) indicates excellent results from NIR spectroscopy for a series of soil microbial parameters such as microbial biomass carbon (MBC), basal soil respiration, selected enzyme activities, total biomass from phospholipid fatty acids (PLFA) and PLFA biomarkers for bacteria and actinomycetes. Heinze et al. (2013) showed that contents of microbial biomass C were predicted with good accuracy using Vis-NIR spectroscopy, while prediction was markedly less successful for other parameters such as contents of microbial N and basal respiration. Regarding such discrepancies, it has been generally stated that rather soil mineralogical and sum parameters such as the total content of SOM may be estimated well with Vis-NIR data, while most others are largely predicted through co-variations with water, SOM and/or minerals (Stenberg et al. 2010; Soriano-Disla et al. 2014). Furthermore, the potential of Vis-NIR to estimate SOM, when it is differently influenced by spatio-temporally variable labile, microbial SOM pools that especially affect the soil surface is left unclear.

Therefore, the objectives of this study were (i) to monitor the spatio-temporal variability of SOM and SOM pools of arable soils (ii) with special focus on the soil surface and its microbial and labile SOM pools, and to investigate the potential of diffuse Vis-NIR spectroscopy (iii) for (in situ) assessment of these SOM pools and (iv) whether findings obtained for the soil surface are representative for the soil layers below. We hypothesize that labile and microbial SOM pools in topsoils show high spatio-temporal variability, with marked differences between soil surfaces and layers below. Consequently, spectral responses differ among topsoil layers, which implies that diffuse in situ Vis-NIR spectroscopy of soil surfaces is not representative for mean SOM pool values of whole topsoil horizons;

This was investigated for topsoils from 21 temperate arable sites that were monthly measured in situ using VIS-NIR spectroscopy and sampled at three selected depths (0-1 cm; 1-5 cm; 5-10 cm) over one year.

2. Materials and methods

Study soils

The study area of the Kenner and Ehranger Flur has a size of 5 km² and is located on lower terraces of the Mosel River located 125 m a.s.l. in the north-east of Trier, Rhineland-Palatinate, West Germany. The study area is agriculturally used and the site conditions (such as topography and climate) are largely similar. Climate is temperate with a mean annual temperature of 10.5 °C and a mean annual precipitation of 761 mm (period 1997-2010; agricultural meteorological station Trier Riol). Topsoils of 21 arable sites were monitored that were cultivated in crop rotation with cereals (Triticum aestivum and x Triticosecale; n=7), rapeseed (Brassica napus; n=7) and maize (Zea mays; n=7), representing typical crop plants in Central Europe. Information on the specific plant cultivated in the study period is listed in supplementary Table S1. By combining the data from all 21 sites, we aimed to eliminate influences of the specific crops on the investigated SOM pools. Main soil types were Haplic to Stagnic Fluvisols and Fluvic Cambisols with the soil texture varying from loamy sand to sandy or clayey loam (detailed texture information is given in Supplementary Table S1). Soil sampling was done once a month at three soil depths (0-1 cm, 1-5 cm and 5-10 cm) after preceding in situ measurements using Vis-NIR, during the period from October 2011 to October 2012 (for detailed information about soil climatic conditions at all 12 sampling dates see Supplementary Table S2).

Soil analysis

Soil temperature was determined in the field using a Digital Soil Thermometer HI 145 (Hanna Instruments, Kehl, Germany). In situ Vis-NIR diffuse reflectance spectra were taken in the field at five randomly chosen, replicate points within a 25□25 cm² square. This was done for each of the three soil depths separately, by carefully digging open with a blade one soil layer after the other and taking in situ reflectance measurements using an ASD FieldSpec II spectroradiometer (Analytical Spectral Devices, Boulder, Colorado, USA). The spectral resolution of this instrument is 3 nm at 700 nm and 10 nm at 1400 and 2100 nm; the sampling interval is 1.4 nm in the VIS–NIR range from 350 to 1000 nm and 2 nm in the SWIR range. Based on the high spectral sampling interval, spectra are provided with 1 nm increments over the complete 0.35–2.5 µm wavelength range; 50 measurements were taken for each spectrum.

In order to collect in situ reflectance measurements independent of sun illumination, spectra were taken with an external illumination and measuring device (ASD high intensity reflectance probe) with a circular measuring area (diameter 7 cm). Spectra were taken before each soil sampling, except for January when ice in frozen soil impeded spectroscopic

measurement. Raw spectra were converted to absolute bidirectional reflectances by measuring a Spectralon© standard with known reflectivity just before the soil measurements. Additionally, wavelengths below 0.4 and above 2.4 µm were cut due to high noise. For the multivariate approaches spectra were interpolated to 5 nm and normalized to vector unity. The spectrally measured soil material was subsequently collected at each sample point for further analysis in the laboratory.

Field moist soil samples were sieved ≤ 2 mm and immediately frozen for subsequent biological analysis or air-dried. The stone content was about 15% for all investigated soils. If necessary, soil was finely ground using an agate mortar before further analysis. The water content was determined by weight loss upon soil drying at 105 °C for 48 h. The pH was analyzed potentiometrically in 0.01 M CaCl2 with a glass electrode. The content of total soil organic carbon (OC) was measured using a EuroEA elemental analyzer (HekaTech, Wegberg, Germany). Determination of hot water extractable C (HWEC) followed the method of Landgraf et al. (2003). Briefly, fresh soil, equal to 10 g dry soil mass, was suspended with 50 ml distilled H2O. After boiling for 1 h using a Kjeldatherm block digestion system (Gerhardt, Bonn, Germany) and subsequent cooling in a water bath, suspensions were centrifuged (2000 g for 3 min) and decanted supernatants were used for TOC analysis (Shimadzu TOC-V-analyzer, Duisburg, Germany). The DOC measurement was realized by extracting 25 g (dry mass) field moist soil with 100 ml 0.01 M CaCl2. After agitation on an oscillating shaker for 1 h at 120 rpm, suspensions were filtered through cellulose-acetate 0.45 µm filters (Ø185 mm, Schleicher & Schuell, Dassel, Germany) and DOC was measured using a Shimadzu TOC-V-analyzer (Duisburg, Germany).

Microbial biomass C (MBC) was analyzed using the chloroform fumigation extraction method according to Vance et al. (1987) with 100 ml 0.01 M $CaCl_2$ as an extractant for 25 g (dry mass) field moist soil. MBC was determined using a Shimadzu TOC-V-analyzer (Duisburg, Germany). For the calculation the factors KEC = 0.45 and a KEN = 0.40 were used, correcting for the extractable part of microbial biomass C (Wu et al., 1990; Joergensen, 1995) and N (Joergensen, 1996). Soil samples for chlorophyll analysis were carefully hand sorted to eliminate remaining residues of plants, moss or lichen. Chlorophyll α (Chl α) as a measure of soil algae and cyanobacteria (Reisser 2012) was extracted from 1 g of soil with 3 ml 90% acetone (Fluka, Buchs, Switzerland). After agitation on an oscillating shaker for 5 h at 120 rpm, suspensions were centrifuged (2000 g for 5 min) and Chl α absorption measurement was done in the decanted supernatants using a UV-Vis spectrophotometer (Shimadzu UV-1650PC, Duisburg, Germany). The PLFAs were extracted according to the protocol of Zelles and Bai (1993) from fresh soil equal to 10 g dry mass. An Agilent 6890A

(G1530A) gas chromatograph (Agilent, Böblingen, Germany), equipped with a 50 m x 0.25 mm x 0.25 µm, polyethylene glycol 2-nitroterephthalate column (Optima® FFAPplus, Macherey-Nagel, Düren, Germany) and a mass spectrometer (Agilent MSD 5973, Agilent, Böblingen, Germany) were used for analysis. The helium carrier gas had a flow rate of 1.2 ml min-1. The oven temperature initially was 60 °C with 2 min static time, ramped at 8 °C min^{-1} to 140 °C and held for 2 min, ramped at 5 °C min^{-1} to 200°C and held for 15 min, ramped at 20 °C min^{-1} to 240 °C and held for 25 min. The identification of individual PLFA markers was performed as reported in Helgason et al. (2010): i14:0, i15:0, a15:0, i16:0, a17:0, i17:0 (Gram-positive bacteria), 16:1ω7t, 16:1ω9c, 18:1ω7c, cy17:0, and cy19:0 (Gram-negative bacteria), 16:1ω5c (arbuscular mycorrhiza), 18:2ω6,9 and 18:1ω9c (saprophytic fungi). Additionally, the PLFA markers for phototrophic microorganisms 18:3ω3 (green algae), 20:5ω3 (diatoms) and 22:5ω3 (cyanobacteria) were investigated according to Boschker et al. (2005).

Data from individual months were aggregated to the four seasons, i.e. October, November and December for autumn, January, February and March for winter, April, May and June for spring, July, August and September for summer. Contents and fingerprints of PLFA were determined for a reduced set of representative months and sites, though, i.e. October (autumn), January (winter), April (spring) and July (summer).

Data analysis

All results are presented as arithmetic means (± SE). The statistical analyses were carried out using the SPSS 20.0 software package. To compare between seasons and soil depths, repeated measures analysis of variance (rmANOVA) was calculated. In case of violation of the assumption of sphericity, the Greenhouse-Geisser correction was used (estimate of sphericity $\varepsilon<0.75$). Subsequently, the post-hoc Scheffé tests were performed. Kruskal-Wallis H-test and U-test were used if the pre-conditions for rmANOVA were not fulfilled. In each case, statistical significance was indicated with *$p < 0.05$; **$p < 0.01$ and ***$p < 0.001$. To analyze spatio-temporal relationships between the parameter, bivariate correlation analyses of data from all 12 sampling months (as a total data set as well as for each season separately) were conducted. The Chl α content was obtained by calculation of the chlorophyll units (CU) after Hoyt (1966):

$$CU/100\ g\ soil = \frac{Absorption\ \times extractant\ (cm^3)}{light\ path\ of\ cell\ (cm)} \times \frac{100}{soil\ mass\ (g)}$$

with the absorption of Chl α being calculated as:

$$Chl\ \alpha\ absorption = 665\ nm - \frac{630\ nm + 750\ nm}{2}$$

Partial least squares regression (PLSR) was used as multivariate modelling tool to obtain prediction estimates from the measured spectroscopic data (Wold et al. 2001). Model calibration was performed by leave-one-out cross validation; the number of latent variables was determined by the model with the smallest root mean squared error (RMSE) and in general restricted to a maximum of 12 factors.

In addition to PLSR using full spectra we also combined PLSR with a spectral variable selection technique. Here, we used the competitive adaptive reweighted sampling strategy (CARS; Li et al. 2009), which is a method that needs very low processing times and has been shown to be effective for model parsimony and accuracy of obtained estimates for key soil properties such as OC or N (Vohland et al. 2014, 2016). In each CARS run a two-step selection procedure based on the PLS regression coefficients is followed, to reach the selection of soil property-specific key wavelengths. The optimal subset of variables is chosen according to the lowest RMSE in the cross validation. Since a weighted randomness is part of the the selection procedure no unique best solution exists. We thus performed 50 runs with CARS and aggregated results by averaging all 50 estimates for each sample.

CARS-PLSR was performed in order to select the relevant (informative) wavelengths for the multivariate modelling approaches. These selections were studied for the different soil parameters to identify overlaps and spectral predictive mechanisms behind each parameter. With this aim, CARS-PLSR was performed for the complete dataset with n = 560 and n = 150 samples for PLFAtot.

3. Results

Vertical and seasonal changes of SOM

The 21 soils from different arable fields within the sampling area covered moderate bandwidths in general soil properties with pH ranging from 5.1 to 7.0 and clay content from 11.3% to 25.2% (Supplementary Table S1). During the one-year study, soils were exposed to largely changing soil climatic conditions that resulted in gravimetric soil water contents ranging from 2.4% (September) to 27.5% (March) and soil temperature ranging from -7.4 °C (February) to 26.6 °C (September). Differences between respective extreme values were largest at the soil surface (0-1 cm) and declined with soil depths, i.e. layers 1-5 cm and 5-10 cm (Supplementary Table S2). The rmANOVA revealed that both soil temperature and moisture were significantly different over the time course and between the three soil layers ($p<0.05$).

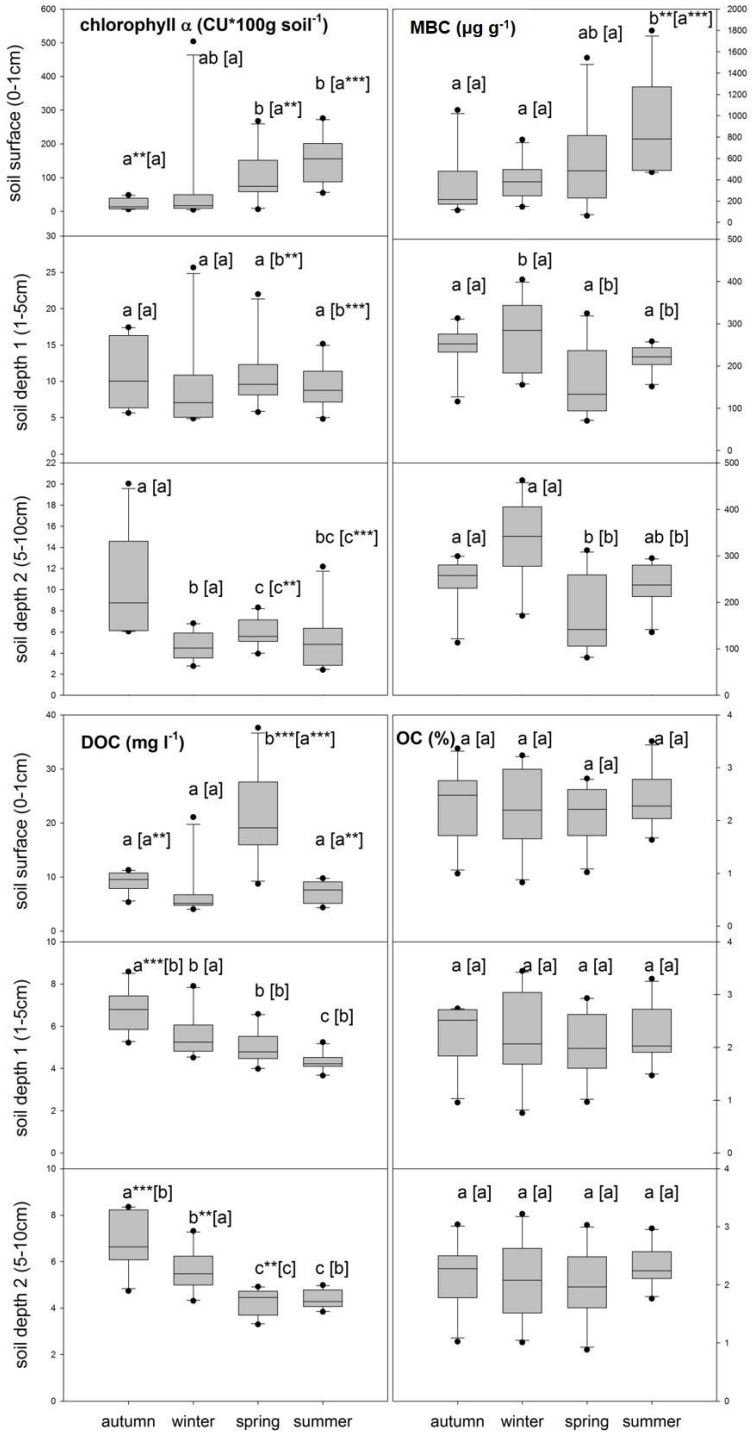


Figure 1: Vertical and seasonal change of the SOM parameter chlorophyll α, microbial biomass carbon (MBC), dissolved organic carbon (DOC) and organic carbon (OC). Significant differences between the four seasons are labeled with lower case letters and between the three soil depths with lower case letters in square brackets (Kruskal-Wallis H-Test and T-test with $p<0.05$; **$p<0.01$ and ***$p<0.001$). Note the in part different scaling of the Y-axis

Total OC did not reveal any significant changes neither between seasons nor between soil depths (Fig. 1). In contrast, the labile SOM component and fractions Chl α, MBC, HWEC and DOC showed significant differences between soil layers. The soil surface showed significant maxima for all labile SOM pools. Vertical differences were especially large in spring and summer due to increased contents of Chl α, MBC, HWEC and DOC at the soil surface (Fig. 1; Supplementary Table S3). The contents at the soil surface differed with high significance from those in the soil depths 1-5 cm and 5-10 cm, while contents were mostly not different from each other in the two deeper layers, despite of Chl α that was significantly dissimilar between all three soil layers (Fig. 1).

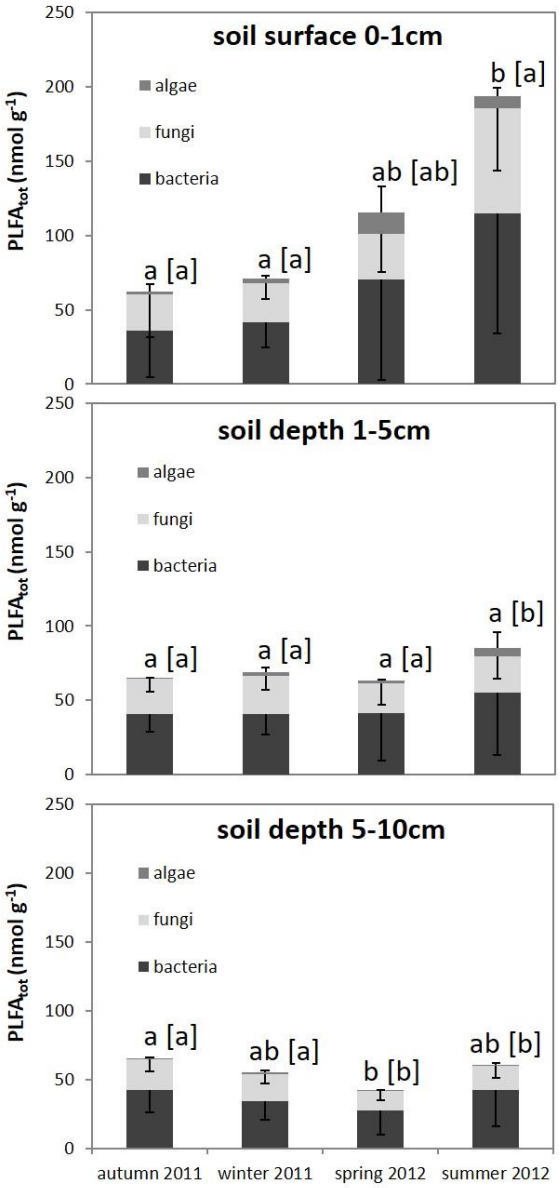


Figure 2: Vertical and seasonal change of total extracted PLFA ($PLFA_{tot}$) and distribution among PLFA markers for bacteria, fungi and algae (consisting of PLFA markers for green algae and diatoms). Significant differences of PLFAtot between the four seasons are indicated by different letters and

between the three soil depths by different letters in square brackets (ANOVA resp. Kruskal-Wallis H test with $p < 0.05$; n= 10). For reasons of clarity standard deviation is displayed only as a negative value for bacteria-and fungi, and as a positive value for algae

In addition, vertical differences between topsoil layers were obtained for PLFAs. Differences were mostly not significant during autumn and winter but became significant in spring and summer, when total PLFA contents at the soil surface largely increased (Fig. 2). This resulted in spring and summer in total PLFA contents declining in the sequence 0-1 cm > 1-5 cm > 5-10 cm, corresponding to the results from MBC. However, due to in part very large standard deviations of individual PLFA and among replicate soil surface samples from different sampling sites, the significance of differences somewhat deviated from that sequence (see letters indicating significant differences in Fig. 2). The low total PLFA content of the deeper topsoil in spring was reflected by low PLFA markers for all three microbial groups investigated, i.e. bacteria, fungi and algae (Fig. 2). In summer, the highly significant increase of bacteria and fungi differentiated the soil surface (0-1 cm) from the layers below, while the maximum for algal PLFA, representing green algae and diatoms, was reached in spring. The PLFA marker for cyanobacteria was not detected in any soil sample.

Seasonal changes were determined for several labile SOM pools by in part significant differences in contents between the four seasons (Fig. 1 and 2). The content of Chl α (in CU 100 g^{-1} soil) ranged on average from 16 (December) to 133 (July), for MBC (in µg g^{-1} soil) from 265 (November) to 770 (July), for HWEC (in µg g^{-1} soil) from 544 (March) to 1060 (August) and for DOC (in mg g^{-1} soil) from 6.0 (March) to 26.9 (September). Significant shifts in contents were also found within the seasons, almost between each single month (data not shown). In fact, a strong variability of each individual parameter was even determined among replicate samples, so that consistent trends could not be derived. The variability is indicated by standard deviations ranging for Chl α from 20±9 to 50±111, for MBC from 371±96 to 770±350 and for DOC from 7.3±1.4 to 26.9±14.2. In contrast, changes were small and not significant for OC with contents ranging from 2.1±0.5% in December to 2.7±0.7% in May (standard deviation range: 2.1±0.5 to 2.3±0.8), which corresponds to the range of contents presented for the vertical distribution of OC. The seasonal changes in the labile SOM fractions at the soil surface were not reflected by changes in the two deeper layers; while at the soil surface maxima of Chl α and MBC were reached in summer and of DOC and HWEC in spring, contents were maximum in the deeper layers in autumn (Chl α, DOC and HWEC) and winter (MBC). The carbon contents in the labile carbon fractions were not related to general soil properties. For example, HWEC showed strong temporal and spatial variations, while its concentrations, either testing averaged or individual data for each month and depth, were not correlated with soil properties such as clay content and pH.

Also total PLFA contents showed seasonal shifts that were dissimilar among soil layers from different depths. Sequences of increasing PLFA contents (nmol g-1; letters showing significant differences) were at the soil surface: autumn (63a) < winter (72a) < spring (116ab) < summer (195b). In the deeper topsoil (5-10 cm) the range of contents was much smaller and the order was: spring (43a) < winter (55b) < summer (61b) < autumn (66b) (see Fig. 2). Yet, standard deviations were much larger for PLFA from summer and autumn compared to winter and spring. Consequently, the latter were significantly different but PLFA from summer and autumn were not. Regarding the main microbial groups identified by PLFA markers for bacteria, fungi and algae, significant seasonal changes occurred at the soil surface (Fig. 2). For the soil depth 1-5 cm only the group of algae significantly increased in winter and summer, resulting in the order (in % of total PLFA content): autumn (0.3%) < spring (2.2%) < winter (4.1%) < summer (6.6%).

Table 1: Spearman rank correlation coefficient (r) of bivariate correlation analyses of green algae determined by Chl α analysis and other microbial groups determined by PLFA analysis with selected SOM parameters of the total data set (n=120) and for each of the four investigated seasons (each n=30, except winter n=28). The level of significance is given by superscript asterisks: *p ≤0.05; **p ≤0.01; ***p ≤0.001

Correlation of	with	Total	Autumn	Winter	Spring	Summer
Algae (Chl α)	MBC[1]	0.52**	0.64**	0.05	0.65**	0.79**
	DOC[2]	0.33**	0.34**	0.19	0.76**	0.80**
	HWEC[3]	0.26**	0.01	0.00	0.63**	0.36**
	OC[4]	0.01	0.05	0.01	0.03	0.01
Algae (PLFA)	MBC	0.25	0.02	0.21	0.00	0.01
	DOC	0.32	0.12	0.01	0.01	0.03
	HWEC	0.09	0.01	0.10	0.01	0.01
	OC	0.01	0.01	0.03	0.01	0.03
	algae (Chl α)	0.16	0.24	0.04	0.03	0.00
Bacteria (PLFA)	MBC	0.41**	0.58**	0.00	0.41**	0.39**
	DOC	0.10	0.02	0.04	0.36**	0.34**
	HWEC	0.11	0.21*	0.03	0.27**	0.22**
	OC	0.00	0.16*	0.01	0.00	0.04
Fungi (PLFA)	MBC	0.48**	0.77**	0.01	0.30**	0.58**
	DOC	0.06	0.06	0.00	0.30**	0.50**
	HWEC	0.09	0.16*	0.00	0.21*	0.25**
	OC	0.02	0.11	0.01	0.02	0.00

[1]MBC = microbial biomass carbon; [2]DOC = dissolved organic carbon; [3]HWEC = hot water extractable carbon; [4]OC = total organic carbon

Table 2: Coefficient of determination (R^2)[a], root mean squared error of prediction (RMSEP) and ratio of standard error of prediction to sample standard deviation (RPD) of the PLSR prediction model (cross validated results) for the parameter chlorophyll α (Chl a), organic carbon (OC), microbial biomass carbon (MBC), dissolved organic carbon (DOC), the phospholipid fatty acids for bacteria, fungi and algae as well as their total content (PLFAtot) for the total sample set (n=560; n=150 for PLFA parameters). Subdivision into the three soil depths (n=187; n=30 for PLFA parameters) and the different seasons (n=30; n=30 for PLFA parameters; for algae models, n varied due to measured 0-values, which were not included in model building). The number of factors used is given in brackets.

		Chl a	OC	MBC	DOC	PLFA			
						bacteria	fungi	algae	total
					R^2				
total		0.79 (12)	0.26 (12)	0.66 (12)	0.58 (12)	0.26 (3)	0.33 (3)	0.29 (4)	0.31 (5)
vertical	surface 0-1 cm	0.67 (7)	0.11 (6)	0.58 (6)	0.48 (7)	0.42 (3)	0.50 (3)	0.21 (4)	0.33 (2)
change	depth 1-5 cm	0.15 (5)	0.49 (10)	0.25 (5)	0.43 (6)	mnp[b]	mnp	mnp	mnp
	depth 5-10 cm	0.26 (9)	0.46 (10)	0.33 (10)	0.37 (6)	mnp	mnp	np[c]	mnp
seasonal	autumn	0.86 (10)	0.50 (8)	0.33 (5)	0.45 (6)	mnp	mnp	np	0.51 (5)
change	winter	snp[d]	snp	snp	snp	snp	snp	mnp	snp
	spring	0.82 (4)	0.43 (9)	0.49 (3)	0.74 (3)	0.17 (3)	0.16 (3)	mnp	0.22 (3)
	summer	0.84 (4)	mnp	0.59 (2)	0.72 (8)	0.36 (2)	0.55 (2)	mnp	0.46 (2)
					RMSEP				
total		23.3	0.59	128	4.16	39.7	23.6	6.70	65.6
vertical	surface 0-1 cm	41.2	0.64	195	6.79	53.5	27.0	10.3	78.9
change	depth 1-5 cm	4.33	0.51	87.2	1.52	mnp	mnp	mnp	mnp
	depth 5-10 cm	3.11	0.56	79.1	1.10	mnp	mnp	mnp	mnp
seasonal	autumn	10.1	0.54	111	1.22	mnp	mnp	1.91	28.0
change	winter	snp	snp	snp	snp	snp	snp	snp	snp
	spring	24.7	0.56	185	4.92	42.7	17.0	9.10	71.3
	summer	28.6	mnp	211	0.82	48.7	23.2	7.44	54.55
					RPD				
total		2.17	1.16	1.70	1.55	1.16	1.22	1.23	1.20
vertical	surface 0-1 cm	1.74	1.04	1.54	1.38	1.02	1.28	0.91	1.18
change	depth 1-5 cm	1.09	1.38	1.16	1.33	mnp	mnp	mnp	mnp
	depth 5-10 cm	1.15	1.22	1.20	1.26	mnp	mnp	mnp	mnp
seasonal	autumn	2.58	1.43	1.22	1.33	mnp	mnp	1.49	1.44
change	winter	snp	snp	snp	snp	snp	snp	snp	snp
	spring	2.37	1.28	1.41	1.96	1.09	1.08	1.28	1.04
	summer	2.50	mnp	1.57	1.89	1.27	1.50	1.01	1.79

[a]according to Saeys et al. (2005), an accuracy with R2 of 0.5 – 0.65 and RPD of 1.5 – 1.9 indicates that it is possible to distinguish between high and low values; r^2 between 0.66 – 0.81 and RPD between 2.0 – 2.5 is an approximate quantitative prediction; r^2 between 0.82 – 0.9 and RPD between 2.5 – 3.0 is a good prediction and R2 > 0.9 and RPD > 3 is an excellent prediction; values of r^2 < 0.5 and RPD < 1.5 give unsuccessful predictions.
[b]mnp = model (PLSR) could not be performed
[c]np = model building was not possible due to insufficient sample set size
[d]snp = spectroscopic measurement could not be performed

In general, the results show that vertical and seasonal changes of microbial SOM constituents were related to each other, with seasonal shifts being dissimilar among different topsoil depths. Thereby, the group of algae showed the most significant changes regarding

their relative contribution to the microbial biomass. Significant differences between soils that were cultivated with different crop plants (maize, rapeseed, cereals) were not determined.

Bivariate correlation analyses of data from all 12 sampling months (as a total data set as well as for each season separately) yielded high correlations of soil algae (determined from Chl α) with MBC, DOC and HWEC (Table 1). Coefficients of correlations for algal Chl α with the dynamic, labile SOM pools were higher than corresponding correlations of PLFA markers for bacteria, fungi or algae. Analyzing data sets for each season separately, highest correlations were found between soil algal Chl α and MBC, DOC and HWEC for spring and for summer. A similar effect of seasons was found for PLFA indicating bacteria and fungi, yet with lower correlation coefficients. In contrast, the soil algal PLFA marker was not significantly correlated to any labile SOM parameter nor to the soil algae determined from Chl α. In contrast to the labile SOM fractions, OC was poorly and despite one single exemption not significantly correlated to any soil microbial parameter (Table 1).

Prediction accuracy with Vis-NIR data

Full spectrum-PLSR was used to obtain estimates for the studied soil parameters from the spectral in situ measurements. For the complete dataset (n = 560), the highest accuracy was obtained for Chl α with R^2 at 0.79 and RPD at 2.17 (Table 2). These values indicate approximate quantitative predictions according to the assessment scheme of Saeys et al. (2005; see footnote to Table 2). The possibility to distinguish between high and low values was obtained for MBC and DOC, whereas predictions were not successful for OC, PLFAtot, and PLFA markers for bacteria, fungi, and algae (Table 2). A subdivision of microbial biomass into bacteria, fungi and algae was not considered due to insufficient model results in full-spectrum PLSR for the individual microbial groups. With CARS-PLSR, results improved for all soil parameters and the calibrated models were more parsimonious in terms of latent variables (Table 3).

Table 3: Cross-validated results of the spectral estimation approach obtained with CARS-PLSR for the complete dataset (n = 560, n = 150 for PLFAtot)

	Chl [a]	OC	MBC	DOC	PLFAtot
r^2	0.82	0.39	0.71	0.63	0.38
RMSE	21.6	0.54	117	3.89	61.6
RPD	2.34	1.27	1.86	1.65	1.27
l.V.a	9.4	11.3	9.9	11.0	3.9

[a]number of latent variables, averaged from 50 runs/model

Regarding the results for the different soil depths, successful modelling was only possible for the surface sample set. There, we obtained highest accuracies for Chl α with R^2 = 0.67 (RPD = 1.74) and MBC with R^2 = 0.58 (RPD = 1.54) (Fig. 3). For all other parameters at the surface, RPD was less than 1.50 (Fig. 3 for the example of OC, Table 2). Modelling for the other soil depths (1-5 cm and 5-10 cm) was not successful or even not possible (Table 2).

Parameters could be described more differentiated, when data were grouped into the four seasons (it must be noted that field spectroscopic measurements were incomplete for the winter season, and thus omitted from this analysis). For Chl α all generated models, meaning data from autumn (R^2 = 0.86), summer (R^2 = 0.84) and spring (R^2 = 0.82) led to good predictions. Best OC models were found for autumn, whereas for summer data models could not be built at all. Acceptable results for DOC were obtained only for spring and summer and for MBC in summer (with the possibility to distinguish between high and low values). PLFAtot modelling provided best results for summer and autumn (maximum R^2 = 0.51, maximum RPD = 1.79), but in spring modelling was unsuccessful. PLFA markers for bacteria, fungi and algae could not be modeled with success for any of the seasons (Table 2).

4. Discussion

Vertical and seasonal changes of SOM

The results showed that in arable topsoils, well mixed from regular tillage, microbially influenced, labile SOM fractions were highly variable in depth and during annual seasons, while total OC was not. Significant vertical and temporal changes of Chl α, MBC, HWEC and DOC became especially evident from strongest shifts, different time series and highest contents of these SOM constituents and fractions at the soil surface (0-1 cm) compared to the deeper soil layers (1-5 and 5-10 cm). Previous studies confirm that the quality and quantity of SOM fractions vary with soil depth and among seasons (e.g., Leinweber et al. 1994; Moll et al. 2015). This goes along with shifts in microbial biomass and community composition (e.g., Smalla et al. 2001; Linsler et al. 2015). Especially for arable soils, the uppermost centimeter of the soil surface is of high functional relevance. Within this centimeter-depth, microbial communities exist that differ significantly from each other, even within millimeter distance, and that decrease in biomass content with soil depth (Nunan et al. 2003; Jeffery et al. 2007). From the algal PLFA marker we found that soil algae are the spatio-temporally most dynamic microbial group. Yet it must be noted that the PLFA 18:3ω3 might merely represent a fractional amount of the soil algal content, while the PLFA signature of photoautotrophic soil organisms is much more complex (Buse et al., 2013). Hence, we focused on Chl α as a more general indicator of the soil algal content. The clear decrease of

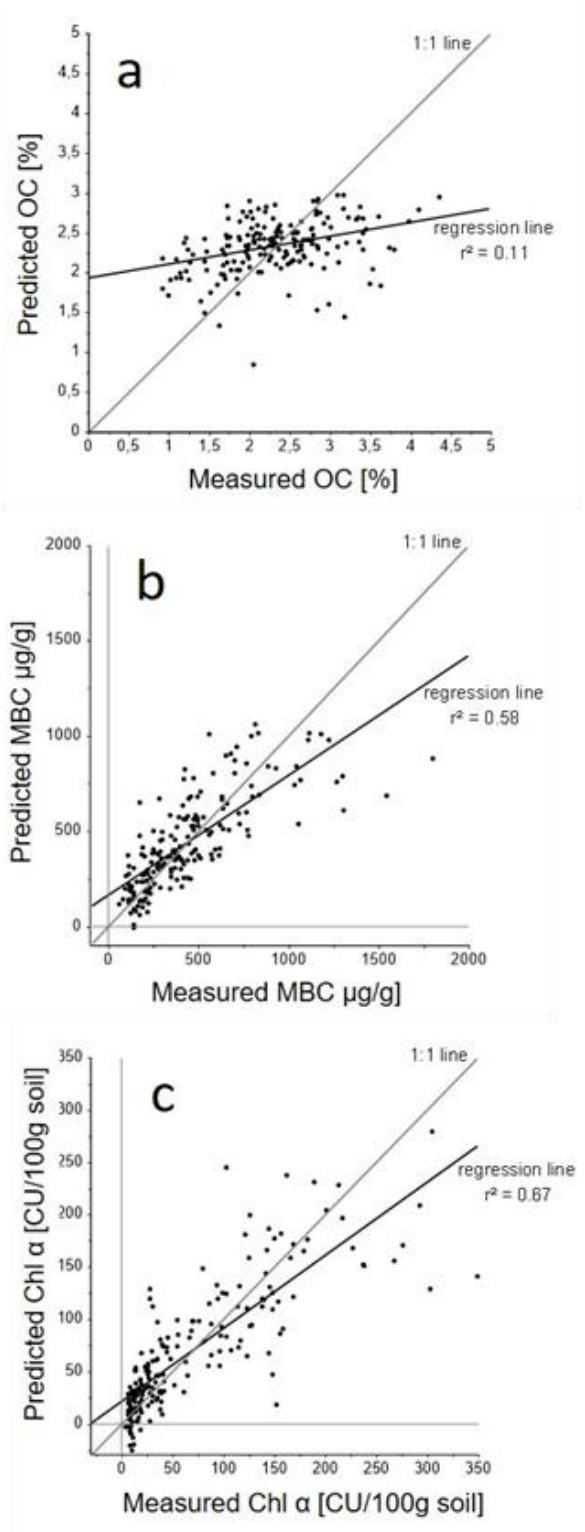


Fig. 3 Predicted values from full-spectrum PLSR (cross-validated results) versus reference values from chemical laboratory analyses for a) organic carbon (OC), b) microbial biomass carbon (MBC) and c) Chlorophyll α (Chl α) for the surface dataset of 186 samples

Chl α contents with soil depth confirmed the typical distribution of algae in soil with highest numbers at the soil surface and much lower abundance down to depths of 1 m (Reisser 2012). From the monthly data it is assumed that differences among the four depth gradients that had been calculated for the four seasons were considerably influenced by short-term changes of SOM parameters that occurred in response to single short-term weather events such as heavy rainfall, strong solar radiation, drought or frost. This was indicated by large variations of soil parameters even between close sampling sites at the same sampling date (see standard deviations presented in the Results section) as well as by abrupt maxima in the content of several SOM parameters and especially of the green soil algae even in the winter period (see maximum values in Fig. 1). The effect of such weather events on MBC, DOC, HWEC and PLFA is explained by the sensitivity of these pools to moisture and temperature (Debosz et al., 1999; Cosentino et al., 2006; Franklin and Mills, 2009; Kramer et al., 2013). Total OC did not show significant vertical or seasonal changes nor was it correlated to any of the other SOM parameters. This is assumed to be either due to an overall dynamic equilibrium in the quantitatively small, short-term C input-output processes, acting on the large OC stock of the arable soils investigated, or due to the relatively short monitoring period of one year, which might not provide sufficient data to correlate these pools (Doane and Horwáth 2004). In comparison, shifts in OC were determined in studies on long-term experiments (Leinweber et al. 1994; Sleutel et al. 2006; Leifeld and Mayer 2015).

Correlation analysis showed that the labile SOM fractions (MBC, DOC, HWEC) are stronger linked to the abundance of algae as determined with Chl α (r^2_{MBC}= 0.52**; r^2_{DOC}=0.33**; r^2_{HWEC}=0.26**) than to the abundance of bacteria (r^2_{MBC}= 0.41**; r^2_{DOC}=0.10; r^2_{HWEC}=0.11) or fungi (r^2_{MBC}= 0.48**; r^2_{DOC}=0.06; r^2_{HWEC}=0.09). This is probably due to the ability of photoautotrophic soil algae to fix carbon dioxide and nitrogen directly from the atmosphere, thus supplying it to the soil biosphere (Elbert et al. 2012). This contribution to SOM is assumed to be high, with globally 3.9 Pg C a^{-1}, which is equivalent to 7% of the primary production by terrestrial vegetation, and 49 Tg N a^{-1}, equivalent to 50% of terrestrial biological N fixation (Elbert et al. 2012). Particularly the N fixation is assumed to be crucial because it makes the associated soil microorganisms within the BSC independent from soil resources (Rousk et al. 2013) and benefits C sequestration by plants (Elbert et al. 2012). A former approach estimated a direct algal contribution to OC in temperate arable soils (as most closely related to the product of temperature and light) of 39 g C m^{-2} a^{-1} (Shimmel and Darley 1985). Since the net annual input rate for temperate arable soils is estimated at 300 to 325 g C m^{-2} a^{-1} (Schlesinger 1991; Körschens 1992), algal C contribution accounts for 12-13% of annual arable OC input. This will first of all influence labile SOM pools. However,

further studies would be necessary to more precisely quantify the contribution of algae (and cyanobacteria) to the different SOM pools.

Prediction accuracy using Vis-NIR data

For arable soil surfaces we could show the strong correlation of microbial biomass to algal abundance. Due to the algal chlorophyll, which produces dominant spectroscopic signals in the visible domain (Karnieli et al. 2001; Jensen 2014), these arable soil surfaces are especially sensitive to Vis-NIR spectroscopy. Consequently, prediction of the soil algal biomass applying PLSR models on Vis-NIR spectra was most successful. The prediction of other living, microbial or non-living SOM constituents and fractions varied largely with the season or soil depth considered. Contents of PLFA markers, indicating the microbial groups of algae, bacteria and fungi, could not be predicted successfully. This may be explained by the fact that single molecular markers might not be able to represent the whole biomass of microbial groups (Frostegård et al. 2011).

Even more, prediction of total OC, which is known as a spectrally active soil variable (Stenberg et al. 2010), was insufficient by using PLSR modelling of Vis-NIR spectra. In contrast to previous studies without such time and soil-depth resolved sampling (e.g. Stenberg et al., 2010), obtained accuracies for OC were surprisingly low, especially for the soil surface (Table 2). The results suggest that the large dynamics of Chl α and of the respective algal biomass at the soil surface has impeded the spectral estimation of OC.

Table 4: Key wavelengths identified from CARS–PLSR

	VIS (400-780 nm)	NIR (785-2400 nm)
OC	455, 505	1085, 1795-1800, 2080-2085, 2135-2155, 2195, 2250, 2325-2330, 2370-2380
Chl α	525, 570-575, 595, 630, 705-710, 755	845-855, 920, 950, 2110-2220, 2295-2300
MBC	525, 575-585, 630-640, 665, 710	955, 1900, 2080-2090, 2160-2165, 2220-2230, 2295-2300, 2320
DOC	515, 550	1365, 1915, 1945, 2190, 2245, 2285, 2400
PLFA	420-430, 445, 505, 535, 630-635	1670, 2820-1825, 2155-2160

As CARS selections identify informative wavelength regions for each parameter (i.e., those with the highest usefulness for the respective modelling approach), the analysis of selection patterns may be used to investigate similarities of predictive mechanisms or indirect predictions triggered by other spectrally active constituents. On this point, one has to be aware that the specificity of single Vis-NIR wavelengths to single compounds or chemical

bonds may change from one sample set to another, as only electronical transitions, combination bands and overtones can be measured in the Vis-NIR region (and not fundamental bands as in the middle infrared). According to the retrieved CARS selections, we found only two key wavelengths to be indicative for OC in the visible part of the spectrum (455 and 505 nm, Table 4), although numerous other wavelengths in the visible range have been reported to be relevant as OC key wavelengths (e.g. Cozzolino and Morón 2006; Vohland et al. 2014). Thus, we hypothesize that chlorophylls and possibly other photosynthetic pigments may have masked OC in the visible part of the spectrum; accordingly, no key wavelengths were found to be relevant for both OC and Chl α (Table 4). Yet, it is controversially debated whether the visible part of the spectrum really plays a key role for assessing OC spectrally (Viscarra Rossel et al. 2006; Stenberg at al. 2010), so that, from our findings, we cannot unequivocally conclude that a spectral masking induced by Chl α is the main factor for the poor quantification of OC.

We identified several common wavelengths for Chl α and MBC (Table 4) and a close correlation coefficient of 0.77 between the measured values of Chl α and MBC. Hence, it is supposed that the abundance of green algae largely influences the modelling of MBC from spectral data. In contrast, the spectral prediction of DOC is not influenced by any of the observed parameters, as there are no overlapping wavelengths. In general, the selection pattern and key wavelengths are distinctly different from the other studied properties, which impedes an explanation of the spectral mechanisms of DOC. The same applies to PLFA data with only minor overlaps with OC, Chl α and MBC (Table 4). In total, successful prediction models could only be built to estimate labile SOM fractions at the soil surface but not for deeper soil layers.

The superior relevance of soil microbial chlorophyll for the spectral quantification of labile and highly dynamic SOM constituents and fractions is further illustrated in Figure 4, as it documents the close and highly significant meta-correlation (overall r of 0.79**, $p<0.01$) between the squared Spearman rank correlation for selected labile SOM fractions with green algae (data taken from Table 3) plotted against the obtained prediction accuracy. Figure 4 says that the more significant the correlation between algae and microbial SOM parameters is, the better is the predictability of algal abundance by Vis-NIR spectral data. This again suggests that labile and dynamic SOM fractions and their spectral behavior are largely related to specific microbial groups and their population dynamics. Nevertheless, the identification of algae as the driving factor of SOM dynamics at the soil surface is not fully clear, since a confounder effect of the overall correlation to another possibly not considered variable in this approach cannot be excluded completely. Anyhow, it can be reasoned that

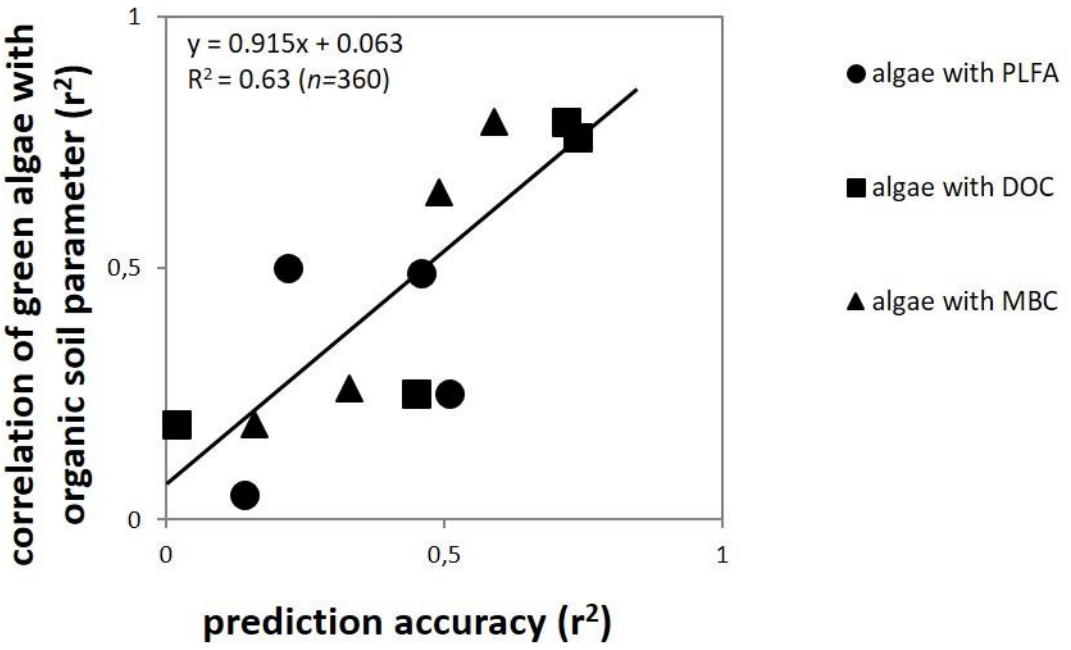


Figure 4: Meta-correlation of obtained prediction accuracies versus the squared Spearman rank correlation for selected labile SOM fractions and green algae, derived from squared Spearman rank correlation coefficients of green algae with the microbial biomass (MBC and PLFAtot) and dissolved organic carbon (DOC), separated for each season (autumn, winter, spring, summer; each n=30) with the corresponding prediction accuracy of the regarded samples (in total n=12×30=360).

spectral information on the soil surface is rather specific, especially in case of algal growth or the abundance of biological soil crusts with phototrophic microorganisms. Spectral information on the soil surface is, thus, not representative for the whole topsoil and not suited to derive contents of SOM and SOM pools of the whole topsoil.

5. Conclusion

This study showed the specific microbial influence on SOM pools and dynamics at the soil surface. All the labile SOM pools investigated in this study showed strong vertical gradients, with significant maxima at the soil surface. Gradients differed, depending on the season of the year. The Vis-NIR spectroscopy applied turned out to be suitable for estimating the spatio-temporal variability of SOM pools from the spectral data. Best prediction accuracy was achieved for surface changes induced by green soil algae, which is also assumed to trigger the modelling of MBC. The group of soil algae was the most significantly changing microbial group, and was closely related to short-term changes of dynamic carbon fractions such as

MBC and DOC. It is assumed, that soil algae can be a significant source for OM in arable topsoils and that the strong microbial community dynamics at the soil surface might trigger SOM sequestration also in deeper topsoil layers. However, to affirm this hypothesis, it would be necessary to monitor the microbial topsoil community at a higher temporal resolution, e.g. to include the impact of single short-term events such as rainfall, and to include further methods to better determine the algal C incorporated in short-term and long-term SOM pools. In contrast to the labile SOM pools, the total and largely stabilized SOM content, measured as total OC did not show spatio-temporal differences or changes. Correspondingly, total OC could not be predicted successfully from Vis-NIR spectroscopy, which was largely attributed to the interference from green soil algae, especially Chl α. This on the one hand raises concerns regarding the acquisition of representative total SOM contents when spectral measurements are solely done at the soil surface that is characterized as a highly dynamic habitat with e.g. the variable presence of soil algae or of biological soil crusts. On the other hand, it suggests a high and not yet fully exploited potential of Vis-NIR spectroscopy to monitor qualitative short-term topsoil OM changes.

Acknowledgements

The authors thank Christian Bossung from the Department of Environmental Remote Sensing and Geoinformatics (Trier University) for his profound support of the field campaigns and the spectroscopic data processing. For the application of CARS-PLSR, we used the freely available CARS Matlab package (Copyright Hongdong Li, 2011; http://code.google.com/p/carspls/), implemented under Matlab Version R2010b. This study was funded by the DFG (SOMScapeS, TH 678/12-1 and VO 1509/3-1).

References

Auzet AV, Poesen J, Valentin C (2004) Editorial: soil surface characteristics: dynamics and impacts on soil erosion. Earth Surface Processes and Landforms 29:1063-1064

Barthès BG, Brunet D, Rabary B, Ba O, Villenave C (2011) Near infrared reflectance spectroscopy (NIRS) could be used for characterization of soil nematode community. Soil Biology and Biochemistry 43:1649-1659

Bellon-Maurel V, 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

Belnap J, Büdel B, Lange OL (2001) Biological soil crusts: characteristics and distribution. In: Belnap J, Lange OL (eds) Biological soil crusts: structure, function, and management. Springer Verlag, pp 3-30

Boschker HTS, Kromkamp JC, Middleburg JJ (2005) Biomarker and carbon isotopic constraints on bacterial and algal community structure and functioning in a turbid, tidal estuary. Limnology and Oceanography 50:70-80

Bowker MA, Maestre FT, Escolar C (2010) Biological crusts as a model system for examining the biodiversity–ecosystem function relationship in soils. Soil Biology and Biochemistry 42:405-417

Bradford MA, Keiser AD, Davies CA, Mersman CA, Strickland CS (2013) Empirical evidence that soil carbon formation from plant inputs is positively related to microbial growth. Biogeochemistry 113:271-281

Brockett BFT, Prescott CE, Grayston SJ (2012) Soil moisture is the major factor influencing microbial community structure and enzyme activities across seven biogeoclimatic zones in western Canada. Soil Biology and Biochemistry 44:9-20

Budel B, Colesie C, Green TG, Grube M, Lazaro Suau R, Loewen-Schneider K, Maier S, Peer T, Pintado A, Raggio J, Ruprecht U, Sancho LG, Schroeter B, Turk R, Weber B, Wedin M, Westberg M, Williams L, Zheng L, (2014) Improved appreciation of the functioning and importance of biological soil crusts in Europe: the Soil Crust International Project (SCIN). Biodivers Conserv 23:1639-1658

Buse T, Ruess L, Filser J (2013) New trophic biomarkers for Collembola reared on algal diets. Pedobiologia 56:153-159.

Clairotte M, Grinand C, Kouakoua E, Thébault A, Saby NPA, Bernoux M, Barthès BG (2016) National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. Geoderma 276:41-52

Cosentino, D., Chenu, C., Le Bissonnais, Y., 2006. Aggregate stability and microbial community dynamics under drying–wetting cycles in a silt loam soil. Soil Biology and Biochemistry 38, 2053-2062

Cozzolino D, Morón A (2006) Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. Soil Tillage Research 85:78–85

Debosz, K., Rasmussen, P.H., Pedersen, A.R., 1999. Temporal variations in microbial biomass C and cellulolytic enzyme activity in arable soils: effects of organic matter input. Applied Soil Ecology 13, 209-218

De Paul Obade V, Lal R (2016) A standardized soil quality index for diverse field conditions. Sci Total Environ 541:424-434

Doane TA, Horwáth WR (2004) Annual dynamics of soil organic matter in the context of long-term trends. Global Biogeochemical Cycles 18:GB3008 3001-3011

Elbert W, Weber B, Burrows S, Steinkamp J, Büdel B, Andreae MO, Pöschl U (2012) Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. Nature Geoscience 5:459-462

Ekschmitt K, Liu MQ, Vetter S, Fox O, Wolters V (2005) Strategies used by soil biota to overcome soil organic matter stability – why is dead organic matter left over in soil? Geoderma 128:167-176

Franklin RB, Mills AL (2009) Importance of spatially structured environmental heterogeneity in controlling microbial community composition at small spatial scales in an agricultural field. Soil Biology and Biochemistry 41:1833-1840

Frostegård Å, Tunlid A, Bååth E (2011) Use and misuse of PLFA measurements in soils. Soil Biology and Biochemistry 43:1621-1625

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

Harper KT, Belnap J (2001) The influence of biological soil crusts on mineral uptake by associated vascular plants. J Arid Environ 47:347-357

Heinze S, Vohland M, Joergensen R, Ludwig B (2013) Usefulness of near-infrared spectroscopy for the prediction of chemical and biological soil properties in different long-term experiments. Journal of Plant Nutrition and Soil Science 176:520-528

Helgason BL, Walley L, Germida JJ (2010) No-till soil management increases microbial biomass and alters community profiles in soil aggregates. Applied Soil Ecology 46:390-397

Hillel D (1982) Introduction to Soil Physics. Academic Press, San Diego, USA

Hodge A, Paterson E, Grayston SJ, Campbell CD, Ord BG, Killham K (1998) Characterisation and microbial utilisation of exudate material from the rhizosphere of Lolium perenne grown under CO2 enrichment. Soil Biology and Biochemistry 30:1033-1043

Hoyt PB (1966) Chlorophyll-type compounds in soil. Plant and Soil 25:167-180

Jeffery S, Harris JA, Rickson RJ, Ritz K (2007) Microbial community phenotypic profiles change markedly with depth within the first centimetre of the arable soil surface. Soil Biology and Biochemistry 39:1226-1229

Jeffery S, Harris JA, Rickson RJ, Ritz K (2009) The spectral quality of light influences the temporal development of the microbial phenotype at the arable soil surface. Soil Biology and Biochemistry 41:553-560

Jensen JR (2014) Remote sensing of the environment. An earth resource perspective. Pearson new international edition. Pearson, London, UK

Joergensen, R.G. (1995): The fumigation-extraction method to estimate soil microbial biomass: extraction with 0.01 M CaCl2. Agrobiological Research 48, pp. 319–324

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

Karnieli A, Kokaly RF, West NE, Clark RN (2001) Remote Sensing of Biological Soil Crusts. From: Belnap J, Lange OL (2001) Biological Soil Crusts: Structure, Function and Management. Ecological Studies 150:431-455. Springer, Berlin Heidelberg

Körschens M (1992) Simulationsmodelle für den Umsatz und die Reproduktion der organischen Substanz im Boden. Berichte Über Landwirtschaft 206: 140-154

Kramer, S., Marhan, S., Haslwimmer, H., Ruess, L., Kandeler, E., 2013. Temporal variation in surface and subsoil abundance and function of the soil microbial community in an arable soil. Soil Biology and Biochemistry 61, 76-85

Kuzyakov Y, Blagodatskaya E (2015) Microbial hotspots and hot moments in soil: Concept & review. Soil Biology and Biochemistry 83:184-199

Landgraf D, Böhm C, Makeschin F (2003): Dynamic of different C and N fractions in a Cambisol under five year succession fallow in Saxony (Germany). Journal of Plant Nutrition and Soil Science 166:319–325

Langhans TM, Storm C, Schwabe A (2009) Community assembly of biological soil crusts of different successional stages in a temperate sand ecosystem, as assessed by direct determination and enrichment techniques. Microbiology and Ecology 58:394-407

Leifeld J, Mayer J (2015) 14C in cropland soil of a long-term field trial – experimental variability and implications for estimating carbon turnover. SOIL 1:537-542

Leinweber P, Schulten HR, Körschens M (1994) Seasonal variations of soil organic matter in a long-term agricultural experiment. Plant and Soil 160:225-235

Li H, Liang Y, Xu Q, Cao D (2009) Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Analytica Chimica Acta 648:77-84

Linsler D, Taube F, Geisseler D, Joergensen RG, Ludwig B (2015) Temporal variations of the distribution of water-stable aggregates, microbial biomass and ergosterol in temperate grassland soils with different cultivation histories. Geoderma 241-242:221-229

Ludwig M, Achtenhagen J, Miltner A, Eckhardt KU, Leinweber P, Emmerling C, Thiele-Bruhn S (2015) Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils. Soil Biology and Biochemistry 81:311-322

Miltner A, Bombach P, Schmidt-Brücken B, Kästner M (2012) SOM genesis: Microbial biomass as a significant source. Biogeochemistry 111:41-55

Moll J, Goldmann K, Kramer S, Hempel S, Kandeler E, Marhan S, Ruess L, Krüger D, Buscot F (2015) Resource type and availability regulate fungal communities along arable soil profiles. Microbial Ecology 70:390-399

Nocita M, Stevens A, van Wesemael B, Brown DJ, Shepherd KD, Towett E, Vargas R, Montanarella L (2015) Soil spectroscopy: an opportunity to be seized. Glob Chang Biol 21:10-11

Nunan, N, Wu K, Young IM, Crawford JW, Ritz K, (2003) Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. FEMS Microbiology Ecology 44:203-215

Reisser W (2012) Soil Algae: Ecophysiological Features, Environmental Impact and Biotechnological Potential. In: Krueger D, Meyer H (eds) Algae. Nova Science Publishers, Hauppauge NY, pp 121-134

Rousk K, Rousk J, Jones DL, Zackrisson O, DeLuca TH, (2013) Feather moss nitrogen acquisition across natural fertility gradients in boreal forests. Soil Biology and Biochemistry 61:86-95

Saeys W, Mouazen AM, Ramon H (2005) Pontential for on-site and on-line analysis of hog manure using visual and near-infrared reflectance spectroscopy. Biosystems engineering 91:393-402

Scharroba A, Kramer S, Kandeler E, Ruess L, (2016) Spatial and temporal variation of resource allocation in an arable soil drives community structure and biomass of nematodes and their role in the micro-food web. Pedobiologia. In press

Schmidt O, Dyckmans J, Schrader S. (2016) Photoautotrophic microorganisms as a carbon source for temperate soil invertebrates. Biology Letters 12

Schlesinger WH (1991) Biogeochemistry - an analysis of global change. Academic Press, San Diego USA

Schulz, K., Mikhailyuk, T., Dreßler, M., Leinweber, P., Karsten, U., 2016. Biological soil crusts from coastal dunes at the Baltic Sea: Cyanobacterial and algal biodiversity and related soil properties. Microbial Ecology 71, 178-193

Schulze ED, Freibauer A (2005) Environmental science: Carbon unlocked from soils. Nature 437:205-206

Shimmel SM, Darley WM (1985) Productivity and density of soil algae in an agricultural system. Ecology 66:1439-1447

Sleutel S, De Neve S, Beheydt D, Li C, Hofman G (2006) Regional simulation of long-term organic carbon stock changes in cropland soils using the DNDC model: 1. Large-scale model validation against a spatially explicit data set. Soil Use and Management 22:342-351

Smalla, K., Wieland, G., Buchner, A., Zock, A., Parzy, J., Kaiser, S., Roskot, N., Heuer, H., Berg, G., 2001. Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: Plant-dependent enrichment and seasonal shifts revealed. Applied and Environmental Microbiology 67, 4742-4751.

Soriano-Disla JM, Janik LJ, Viscarra Rossel RA, MacDonald LM, McLaughlin MJ (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

Stenberg B, Viscarra Rossel RA, Mouazen AM, Wetterlind J (2010) Visible and near infrared spectroscopy in soil science. Advances in Agronomy 107:163-215

Vance ED, Brookes PC, Jenkinson DS (1987) An extraction method for measuring soil microbial biomass C. Soil Biology and Biochemistry 19:703-707

Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75

Vohland M, Ludwig M, Thiele-Bruhn S, 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

Wold S, Sjöström M, Eriksson, L (2001) PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems 58:109-130

Wu, J., Joergensen, R.G., Pommerening, B., Chaussod, R., Brookes, P.C. (1990): Measurement of soil microbial biomass by fumigation-extraction – an automated procedure. Soil Biology and Biochemistry 22, pp. 1167-1169

Zancan S, Trevisan R, Paoletti MG (2006) Soil algae composition under different agro-ecosystems in North-Eastern Italy. Agriculture, Ecosystems and Environment 112:1-12

Zelles L, Bai QY (1993) Fractionation of fatty acids derived from soil lipids from soil lipids by solid phase extraction and their quantitative analysis by GC-MS. Soil Biology and Biochemistry 25:495-507

Zornoza R, Guerrero C, Mataix-Solera J, Scow KM, Arcenegui V, Mataix-Beneyto J (2008) Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. Soil Biology and Biochemistry 40:1923-1930

Supplementary Table S1 Site conditions of all 21 sampling sites at the Kenner and Ehranger Flur

		pH (0.01 M CaCl2)			soil texture (%)		
		0-1cm	1-5cm	5-10cm	silt (0-10cm)	clay (0-10cm)	sand (0-10cm)
cereals	1	5.4	5.1	6.6	37.1	16.8	46.9
	2	5.6	5.8	6.4	24.0	14.0	62.0
	3	6.6	6.6	6.7	29.3	15.1	56.9
	4	6.5	6.6	6.6	26.6	13.1	63.2
	5	6.1	6.2	6.2	30.1	12.1	58.0
	6	6.4	6.5	6.6	25.3	13.3	61.4
	7	6.7	6.8	6.7	35.9	13.4	50.7
rapeseed	8	5.8	5.9	5.8	35.3	11.5	53.3
	9	5.8	5.8	5.9	25.5	11.3	62.4
	10	6.4	6.4	6.6	28.5	13.8	57.7
	11	6.7	6.9	6.9	32.0	12.1	55.9
	12	6.8	6.7	6.6	30.1	19.2	51.0
	13	6.7	6.8	6.8	42.0	25.2	33.0
	14	6.7	7.0	6.6	39.2	16.9	44.9
maize	15	6.7	6.6	6.7	44.7	18.2	37.1
	16	6.8	6.8	6.7	44.1	15.1	41.0
	17	6.6	6.3	6.7	49.1	20.0	31.2
	18	6.7	6.8	6.8	47.1	19.9	33.1
	19	6.8	6.9	6.7	47.0	20.1	32.9
	20	6.7	6.9	6.8	46.7	17.1	36.2
	21	6.6	6.6	6.6	41.9	16.1	42.1

Supplementary Table S2 Soil climatic conditions at the monthly sampling dates and three soil depths

		Soil moisture content (%)			Soil temperature (°C)		
Year	Month	0-1cm	1-5cm	5-10cm	0-1cm	1-5cm	5-10cm
2011	October	9.7±6.6	12.7±5.3	14.2±4.6	9.1±1.9	11.1±1.8	11.6±1.6
	November	9.5±3.6	12.7±3.4	13.7±3.4	5.3±1.2	7.7±0.6	8.3±0.8
	December	26.7±2.6	27.4±3.1	25.6±2.8	4.7±0.4	5.3±0.3	5.7±0.3
2012	January	26.4±5.0	26.1±3.6	24.6±3.2	3.4±1.0	4.4±0.5	4.9±0.2
	Februarya				-7.4	-1.5	0.0
	March	27.5±2.8	25.7±2.7	24.2±3.2	6.9±2.6	7.5±0.9	7.1±0.5
	April	9.9±3.8	14.2±3.2	15.4±3.1	8.9±4.0	9.8±2.4	9.4±1.6
	May	25.1±2.5	23.8±2.5	22.6±2.6	15.3±2.9	16.0±2.5	15.1±1.9
	June	22.9±1.8	20.9±2.0	20.4±2.4	18.4±2.8	15.9±1.2	15.1±1.0
	July	24.0±1.9	21.2±2.3	21.0±2.0	20.4±3.4	18.0±1.7	17.6±1.1
	August	10.6±6.6	13.2±4.4	14.0±3.0	24.5±4.5	20.1±2.6	18.5±1.6
	September	2.4±1.5	7.0±2.7	9.5±2.7	26.6±4.2	19.4±1.7	17.7±1.1
Minimum		2.4	7.0	9.5	-7.4	-1.5	0.0
Maximum		27.5	27.4	25.6	26.6	20.1	18.5
Difference		25.1	20.4	16.1	34.0	21.6	18.5
	Min	0.9	4.3	6.8	-7.4	-1.5	0.0
	Max	30.3	30.5	28.4	30.8	22.7	20.1

[a]no soil sampling possible due to frozen soil, soil temperature values were recorded by the agricultural meteorological station in Riol

Supplementary Table S3 Seasonal and vertical variation of the labile carbon pools of hot water extractble carbon (CHWE), microbial biomass carbon (MBC) and dissolved organic carbon (DOC) in comparison to total organic carbon (OC) of ten selected sampling sites (according to PLFA sampling set)

		HWEC (µg g^{-1})				MBC (µg g^{-1})				DOC (mg l^{-1})				OC (%)			
		autumn	winter	spring	summer	autumn	winter	spring	summer	autumn	winter	spring	summer	autumn	winter	spring	summer
surface (0-1cm)	1	523.8	380.1	534.4	439.3	214.6	145.0	60.5	493.7	10.7	4.0	8.7	4.5	1.7	1.4	1.8	3.5
	2	872.6	511.8	702.2	685.0	1053.4	262.1	171.8	471.7	11.3	4.8	14.1	4.3	2.8	2.2	2.5	2.4
	3	696.2	596.7	841.7	797.4	725.5	367.4	629.3	767.9	10.9	6.3	22.4	5.4	2.6	3.2	2.8	2.8
	4	519.8	522.3	1248.2	630.8	397.6	494.3	927.4	469.7	10.5	5.1	27.8	5.4	2.4	2.9	2.6	2.8
	5	612.8	557.6	1331.9	814.2	232.2	366.8	1543.8	676.8	10.1	7.6	37.6	7.1	1.7	1.7	1.7	2.0
	6	289.4	296.9	691.2	1062.4	112.4	208.6	248.7	1797.5	7.9	4.7	18.9	8.9	1.0	0.8	1.0	1.6
	7	936.5	612.1	891.0	671.2	215.1	441.4	525.2	1297.5	8.4	5.2	19.3	9.5	2.5	2.1	1.9	2.2
	8	703.6	674.8	741.6	873.2	172.0	394.1	443.8	1263.6	7.9	4.9	16.6	8.9	2.3	1.9	1.7	2.3
	9	822.2	820.1	1079.8	487.1	171.3	501.4	777.5	1031.4	9.0	6.4	27.5	8.2	2.7	3.0	2.5	2.3
	10	886.6	1086.6	785.4	785.3	195.5	776.2	436.7	798.2	5.3	21.1	18.6	9.8	2.7	2.9	2.6	2.0
depth 1 (1-5cm)	1	562.4	428.1	531.1	289.3	201.3	170.9	81.0	198.7	6.9	4.5	4.0	3.7	1.9	1.4	1.6	3.3
	2	626.6	588.5	702.1	512.8	296.3	315.1	112.0	225.0	6.3	5.2	6.4	4.6	2.6	2.2	2.9	2.1
	3	670.5	626.6	610.6	629.3	299.2	412.5	118.6	235.4	6.8	7.9	6.6	4.2	2.7	3.4	2.9	2.8
	4	447.1	479.3	615.7	544.9	265.3	299.7	150.3	239.1	5.2	5.7	4.1	5.2	2.5	2.6	2.3	2.7
	5	587.2	525.6	633.7	622.4	240.3	304.9	133.0	279.6	8.6	5.6	4.7	4.2	1.7	1.8	1.5	1.8
	6	340.2	304.8	390.9	357.9	122.9	211.8	88.7	135.6	5.9	4.8	4.8	3.9	1.0	0.8	1.0	1.5
	7	834.6	570.0	708.3	501.3	262.7	400.0	252.8	281.3	7.8	5.2	4.9	4.2	2.5	1.9	1.7	2.0
	8	596.3	620.4	605.1	544.7	249.3	403.6	278.3	294.9	7.2	4.8	4.7	4.5	2.2	2.0	1.7	2.0
	9	725.7	771.4	755.3	561.9	275.4	462.5	311.8	240.9	7.3	5.3	5.2	4.4	2.7	3.2	2.4	2.0
	10	816.9	649.8	540.7	526.8	253.2	369.0	232.9	217.1	5.8	7.2	4.6	4.2	2.7	3.0	2.5	2.5
depth 2 (5-10cm)	1	551.8	452.3	514.7	451.6	230.5	178.3	84.9	239.8	6.6	5.3	3.5	4.3	1.8	1.4	1.8	2.8
	2	579.2	566.3	663.7	586.3	242.4	310.0	147.0	243.4	6.2	5.6	4.7	4.6	2.4	2.4	2.7	2.5
	3	620.5	538.5	613.6	651.0	313.0	259.1	119.5	258.3	8.4	6.9	4.1	4.9	3.0	2.6	3.0	3.0
	4	468.5	390.8	460.0	465.1	271.4	221.9	70.1	210.4	4.7	4.6	3.3	4.8	2.3	2.5	1.9	2.4
	5	625.9	478.6	631.1	565.9	234.0	185.2	104.3	243.3	7.7	5.1	3.7	3.8	1.7	1.6	1.3	1.8
	6	363.7	283.9	373.3	388.1	155.7	155.4	97.1	151.2	5.8	6.0	4.4	5.0	1.0	1.0	0.9	2.2
	7	902.1	639.4	807.2	471.9	255.7	336.8	261.6	200.1	8.2	5.2	4.5	4.0	2.4	1.8	2.0	2.2
	8	598.9	551.8	734.2	500.1	255.6	343.9	324.6	233.7	8.3	4.3	4.9	4.1	2.3	1.8	1.7	2.2
	9	733.2	765.9	786.8	463.7	249.4	404.5	228.6	204.6	6.7	5.9	4.9	4.2	2.8	3.2	2.2	2.1
	10	719.1	606.1	593.5	542.3	290.2	343.7	223.2	205.0	6.2	7.3	4.5	4.3	1.9	2.8	2.4	2.1

5 Microbial contribution to soil organic matter (SOM) quantity and quality in density fractions of temperate arable soils

M. Ludwig, J. Achtenhagen, A. Miltner, K.-U. Eckhardt, P. Leinweber, C. Emmerling, S. Thiele-Bruhn

Soil Biology and Biochemistry 81 (2015) 311-322

DOI: 10.1016/j.soilbio.2014.12.002

Abstract

The formation of soil organic matter (SOM) very much depends on microbial activity. Even more, latest studies identified microbial necromass itself being a significant source of SOM and found microbial products to initiate and enhance the formation of long-term stabilized SOM. The objectives of this study were to investigate the microbial contribution to SOM in pools of different stability and its impact on SOM quality. Hence, four arable soils of widely differing properties were density-fractionated into free and occluded particulate organic matter (fPOM, oPOM < 1.6 g cm-3 and oPOM < 2.0 g cm-3) and mineral associated organic matter (MOM > 2.0 g cm-3) by using sodium polytungstate. These fractions were characterized by in-source pyrolysis-field ionization mass spectrometry (Py-FIMS). Main SOM compound classes of the fractions were determined and further SOM properties were derived (polydispersity, thermostability, humification degree). The contribution of microbial derived input to arable soil OM was estimated from the hexose to pentose ratio of the carbohydrates and the ratio of C4-C26 to C26-C36 fatty acids. Additionally, selected samples were investigated by scanning electron microscopy (SEM) for image evaluation of the OM origin. Results showed that, although the samples differed significantly regarding soil properties, SOM composition was comparable and almost 50% of identifiable SOM compounds of all soils types and all density fractions were assigned to phenols, lignin monomers and alkylaromatics. Most distinguishing were the high contents of carbohydrates for the MOM and of lipids for the POM fractions. Qualitative features such as polydispersity or thermostability were not in general assignable to specific compounds, density fractions or different mean residence times. Only the microbial derived part of the soil carbohydrates could be shown to be correlated with high SOM thermostability (r2=0.63**, n=39). Microbial derived carbohydrates and fatty acids were both enriched in the MOM, showing that the relative contribution of microbial versus plant derived input to arable SOM increased with density and therefore especially increased MOM thermostability. Nevertheless, the general microbial contribution to arable SOM is suggested to be high for all density fractions; a mean proportion of about 1:1 was estimated for carbohydrates. Despite biomolecules released from living microorganisms, SEM revealed that microbial necromass is a significant source for stable SOM which is also increasing with density.

1. Introduction

For the sustainable management of arable soils, the understanding of the formation and persistence of soil organic matter (SOM) needs to be enhanced (Lal, 2009). SOM is derived from two main groups of input materials: plant residues (and exudates) and microbial residues (and exudates) (Kögel-Knabner, 2002). The total microbial biomass is in general comparably small with 50-2000 µg C -1 g soil (Blagodatskaya and Kuzyakov, 2013) which on average represents 2-3% of the total organic C content (Anderson and Domsch, 2010). In parallel, indications of seriously underestimated microbial biomass contributions to SOM can be found, ranging up to >50% of total extractable SOM, about 45% of the humin fraction and>80% of soil N (Simpson et al., 2007).

Microbial OM is, furthermore, contributed by microbial biomass in different states of metabolic activity: The total microbial biomass is composed of only about 2% active microorganisms, while dormant and dead microorganisms show mean contents of 42% and 56%, respectively (Blagodatskaya and Kuzyakov, 2013). The proportion of potentially active microorganisms (activated by available substrates within hours) within the dormant fraction can vary from 10-40% (up to 60%). Yet it has been generally accepted that the microbial contribution to SOM is almost completely metabolic and contributed by the living biomass, i.e. by the active or potentially active microorganisms. The dead microorganisms are mostly regarded as a very dynamic and inconsistent fraction due to permanent re-utilization of microbial C (Bradford et al., 2013, Blagodatskaya and Kuzyakov, 2013). However, several studies emphasized the underestimation of microbial necromass to SOM formation (Simpson et al. 2011, Liang and Balser 2010; Liang et al. 2011). Recent studies showed that microbial necromass directly contributes to non-living SOM and may provide a major part of the molecular SOM structures of up to 40% recovered in non-living SOM (Kindler et al., 2009; Miltner et al., 2009; Miltner et al., 2012; Schurig et al., 2013).

Consequently, microbial biomolecules such as carbohydrates and proteins have been identified as the largest contributor to stable SOM (Knicker, 2011; Cotrufo et al., 2012). Especially microbial derived carbohydrates were shown to initiate the formation of organo-mineral associations (Dümig et al., 2012) and to be enriched in mineral associated OM (MOM) due to further preferential stabilization (Gregorich, 1994; Rumpel et al., 2010). Carbohydrates represent the most abundant compound class of SOM input materials (Miltner and Zech, 1998; Kögel-Knabner, 2002), of which over 90% are represented by the five monosaccharides glucose, galactose, mannose (all hexoses) and arabinose and xylose (both pentoses) (Gregorich, 1994). Based on the observation that a predominance of hexoses represents microbial derived materials while a predominance of pentoses

represents plant-derived materials (Guggenberger, 1994), the hexose to pentose ratio is frequently used as a proxy to determine the amount of microbial OM (Oades, 1984; Six et al., 2006). Another proxy to estimate the microbial contribution to SOM is the ratio of even C-chain fatty acids of different chain lengths. This ratio distinguishes between microbial fatty acids (C4-C26) and fatty acids (C26-C38) derived from plants and higher insects (Schnitzer et al., 1986). However, soil micro- and mesofauna organisms provide mainly easily biodegradable short C-chain fatty acids to the soil lipid pool, whereas hardly biodegradable long C-chain fatty acids originate from plant materials and organic amendments such as farmyard manure (Jandl et al., 2005). These long-chain fatty acids accumulate in the small particle-size fractions and, thus, are stabilized against microbial degradation (Jandl et al., 2004).

Density fractionation has been introduced to SOM studies in the 1960ies to differentiate among particulate, litter-derived (low-density) and mineral-associated (high-density) SOM (Christensen, 1992; Jastrow et al., 1996; Six et al., 2000). Sodium polytungstate (SPT; Na6(H2W12O40)) became the favored density fractionation agent and was used in most recent studies (Six et al., 1999; Bruun et al., 2010) because it leaves the organic matter almost unchanged, can be easily adjusted to the desired density and is nontoxic. For this study, four density fractions were separated, including free and occluded particulate organic matter (fPOM, oPOM < 1.6 g cm-3 and oPOM < 2.0 g cm-3) and mineral associated organic matter (MOM > 2.0 g cm-3). Although the fPOM fraction is assumed to mostly be free of mineral associations while the MOM fraction mainly consist of organo-mineral complexes, mineral soil can be found in all fractions, strongly increasing with density in general. However, there does not seem to be a specific density at which mineral associated OM is clearly separable from free organic matter, displaying a continuum rather than a fixed cutoff (Bruun et al., 2010). The oPOM is furthermore assumed to be stabilized by two different mechanisms, by selective accumulation due to recalcitrance and by physical protection due to occlusion. Accordingly, oPOM can vary considerably in composition or age, for instance shown by 13C analysis of John et al. (2005) with mean age of C in fPOM (22 years) < oPOM2.0 (49 years) < MOM (63 years) < oPOM1.6 (83 years).Therefore, oPOM was chosen to be represented by two density fractions to better elucidate the stabilization effects on SOM composition. And, although none of the existing fractionation methods, including the SPT density fractionation, refers to a specific stabilization mechanism and hence the obtained fractions cannot represent distinctive functional SOM pools (von Lützow et al., 2007), the fractions can be assigned to SOM of different mean residence times (MRT). For example, Dorodnikov et al. (2011) reported a MRT for carbon (MRTC) increasing in the sequence fPOM (about 2 years) < oPOM1.6 = oPOM2.0 (about 25 years) < MOM (about 45 years) and

a MRT for nitrogen (MRTN) of fPOM (about 10 years) < oPOM1.6 (about 45 years) < oPOM2.0 (about 55 years) < MOM (about 65 years). Following a definition of von Lützow et al. (2007), the POM fractions with a MRT < 10 years represent the active OM pool while the MOM contributes to the intermediate or passive OM pool. However, experiments with 13C and 15N labeled SOM indicated that some components of the MOM were transformed more rapidly than expected from the MRT of the whole fraction. This suggests that mineral associated SOM fractions are composed of at least two functional SOM pools of different turnover (Bird et al., 2008).Nevertheless, only a few studies investigated the chemical composition of organic matter in density fractions by sensitive mass spectrometric methods (Golchin et al., 1997; Schulten and Leinweber, 1999; Sleutel et al., 2007).

It appears crucial for the knowledge on and management of SOM to elucidate the contribution of microbial biomass to SOM pools of different stability. SOM stability is characterized by microbial stability and thermal stability. Microbial stability represents SOM resistance to microbial decomposition while thermal stability refers more to the physical conditions of SOM stabilization mechanisms. However, organo-mineral associations are known to increase both microbial (Helfrich et al., 2010) and thermal stability (Leinweber et al., 2009, Schulten and Leinweber, 1999), which hence are hardly separable and closely related as assumed by several studies (Helfrich et al., 2010). Therefore, this study combined soil density fractionation with subsequent fractional SOM analysis by in-source pyrolysis-field ionization mass spectrometry (Py-FIMS) and scanning electron microscopy (SEM). This enabled determining the molecular SOM composition (total contents of C and N and main SOM compound classes including carbohydrates, phenols, lignin, lipids, alkylaromatics, N-containing compounds, sterols, suberin, peptides and free fatty acids) and other qualitative SOM characteristics (mean residence time, polydispersity and thermostability) with focus on the microbial contribution to SOM. Hence, the overall aim of this study was to estimate the relative contribution of microbial vs. plant-derived organic matter to SOM in different soils and pools of stability. To further investigate structured SOM contributing to the different fractions, SEM image evaluation was involved to analyze microbial residues and plant tissue.

2. Materials and Methods

2.1 Soil samples

Soils from acidic to alkaline parent materials in the greater region of Trier (Rhineland-Palatinate, SW-Germany) were studied. In that region the mean annual temperature is 10.6°C and the mean annual precipitation is 766 mm (period 1997-2009; agricultural meteorological station Trier Riol). Samples were taken during October 2010 from arable

topsoils (0-10 cm) under wheat (Triticum aestivum LINNAEUS) and triticosecale (xTriticale TSCHERM.-SEYS. EX MÜNTZING) in rotation with rapeseed (Brassica napus LINNAEUS). Samples of four sites were chosen for this study, covering the broad range of chemical and biological soil properties caused by the widely differing parent materials of this region (Tab.1).

Table 1: Mean sample characteristics of the four soil samples such as the soil type classified after the World Reference Base (WRB), the geological substrate, the proportions of clay, silt and sand, the pH value, carbonate carbon (IC), soil organic carbon (OC), total nitrogen (N), microbial carbon (C_{mic}), hot water extractable carbon (C_{hwe}) and organic carbon thermostable at 375°C (OC_{375}).

sample name	soil type (WRB)	geological substrate	clay (%)	silt (%)	sand (%)	pH ($CaCl_2$)	IC (mg g^{-1})	OC (mg g^{-1})	N (mg g^{-1})	Cmic (mg kg^{-1})	Chwe (mg kg^{-1})	OC_{375} (mg g^{-1})
haFL	Haplic Fluvisol	Holocene fluviatile sediments, windblown silt (Loess)	13.3	30.3	57.3	6.7	0.0	14.1	1.2	285	510	9.2
haCM	Haplic Cambisol	Permian siltstone (Rotliegendes)	21.8	41.4	38.7	5.9	0.0	16.4	1.5	243	640	5.8
haST	Haplic Stagnosol	Triassic limestone (Muschelkalk)	29.6	47.3	24.9	6.6	0.0	18.6	1.8	336	718	11.2
stCM	Stagnic Cambisol	Triassic siltstone (Keuper)	41.1	57.4	8.4	7.1	14.3	47.2	2.7	378	800	13.9

2.2 Soil analysis

For sampling of the top horizons fifteen random core samples were taken and pooled to one composite. These composite samples were sieved ≤ 2 mm and immediately air-dried. For specific analyses, soil was finely ground using an agate mortar. Percentage composition of clay, silt and sand were determined by sieving and sedimentation according to Hartge and Horn (1989). The pH was determined potentiometrically in 0.01 M CaCl2 with a glass electrode. The content of total soil organic carbon (OC) and nitrogen was measured by gas chromatography after combustion at 1100°C using an EuroEA elemental analyser (HekaTech, Wegberg, Germany). Previously, soil samples containing free carbonate were pretreated prior to analysis to remove carbonate C. Briefly, 200 mg pulverized soil (dry mass) was treated with 200 ml 0.22 N HCl and then suspended (18000 U min-1) with an Ultra Turrax vortexer (IKA, Staufen, Germany) for ≥ 3 min. Non purgeable organic carbon (NPOC) was analyzed from this suspension with a Shimadzu TOC-V analyzer (Shimadzu, Duisburg, Germany). Determination of hot water extractable C followed Landgraf et al. (2003). Briefly, 10 g soil (dry mass) was suspended with 50 ml distilled H2O. After boiling for 1h using a Kjeldatherm block digestion system (Gerhardt, Bonn, Germany) and subsequent cooling in a water bath, suspensions were centrifuged (2000 g for 3 min) prior to TOC analysis

(Shimadzu TOC-V-analyzer, Duisburg, Germany). Microbial biomass C (Cmic) was analysed by the chloroform fumigation extraction method according to Joergensen (1995). Organic carbon thermostable at 375°C (OC375) was determined by measuring SOC by gas chromatography (EuroEA elemental analyzer) after oxidation of samples at 375°C for 24 h (Gélinas et al., 2001).

The density fractionation by SPT (TC-Tungsten Compounds; CAS-No. 12141-67-2) was adapted and modified from Golchin et al. (1997) and Helfrich et al. (2007). For this study, four density fractions were investigated, including the free particulate organic matter fraction (fPOM < 1.6 g cm-3), the light occluded particulate organic matter fraction (oPOM <1.6 g cm-3), the heavy occluded particulate organic matter fraction (oPOM <2.0 g cm-3) and the mineral associated organic matter fraction (MOM > 2.0 g cm-3). Briefly, 10 g air-dried soil was amended with 30 ml of the SPT solution of defined density and centrifuged for 1h at 4,816 g. To obtain the free POM, SPT solution of 1.6 g cm-3 was used. To release occluded POM of the same density, glass beads (diameter 5mm) were added to the soil pellet after the separation of the fPOM (supernatant) and vortexed before installation on a oscillating shaker for 16 h at 100-120 rpm to destroy the aggregates. After centrifugation and separation of the supernatant, the remaining soil was centrifuged with SPT solution of 2.0 g cm-3 to obtain the occluded POM < 2.0 g cm-3 (supernatant) and the mineral associated OM > 2.0 g cm-3 (soil pellet). After each centrifugation step, the obtained fraction material was vacuum filtrated through a membrane filter (<0.45 µm; Sartolon Polyamid Stedim Filter, Sartorius, Göttingen, Germany). The filter residue was rinsed into a pre-weighed glass petri dish and dried at 40°C.

Pyrolysis-field ionization mass spectrometry (Py-FIMS) was done to determine the relative molecular composition and the thermal stability of SOM in bulk soil samples and density fractions. Sample aliquots (3-5 mg; three replicates) were analyzed by thermal degradation from 110°C to 700°C in steps of 10°C using a modified Finnigan MAT 731 high performance 8 kV acceleration voltage mass spectrometer (Finnigan, Bremen, Germany). For each sample about 60 magnetic scans in a mass range of 15 to 900 m/z were combined to averaged Py-FI mass spectra. Additionally, thermograms were compiled for the total ion intensities. For each spectrum the relative abundance of 10 SOM compound classes was calculated on the basis of indicator ion signal intensities (Schulten and Leinweber, 1999). These 10 SOM compound classes comprise 1) carbohydrates, 2) phenols and lignin monomers, 3) lignin dimers, 4) lipids, alkanes, alkenes, bound fatty acids and alkylmonoesters, 5) alkylaromatics, 6) mainly heterocyclic N-containing compounds, 7) sterols, 8) peptides, 9) suberin, and 10) free fatty acids.

Two selected soil samples, haST and stCM, and the corresponding fractions were visualized and analyzed by scanning electron microscopy (SEM). The selected air dried samples were fixed on sticky carbon pads and coated with a thin layer of gold (Sputtercoater SCD 50, Balzers, Liechtenstein) to increase the electric conductivity. SEM was performed with an Ultra 55 (Carl Zeiss, Oberkochen, Germany) with the following settings: gun voltage 1 kV; system vacuum 10-6 mbar; detector secondary electron 2; noise reduction frame average/line average. Representative selections of the soil samples were taken with different magnifications ranging from 100- to 50000-fold.

2.3 Data analysis

All results are presented as arithmetic means (± SE). The statistical analyses were carried out using the SPSS 20.0 software package. To compare the fractions, one-way analysis of variance (ANOVA) and post-hoc Scheffé test were performed. Kruskal-Wallis H-test and U-test were used if the pre-conditions for ANOVA were not fulfilled. In each case, statistical significance was indicated with *p < 0.05; **p < 0.01 and ***p < 0.001.

Structural information about the SOM compounds was obtained by calculation of the polydispersity after Lattimer and Schulten (1993). The polydispersity is the ratio of the mean molecular mass averaged by weight (Mw) to the mean molecular mass averaged by number (Mn) of the respective mass spectrum calculated over:

$$\overline{M_w} = \frac{\sum_{i=1}^{m} I_i \cdot M_i^2}{\sum_{i=1}^{m} I_i \cdot M_i} \quad \text{(g mol-1)} \qquad (1)$$

$$\overline{M_n} = \frac{\sum_{i=1}^{m} I_i \cdot M_i}{\sum_{i=1}^{m} I_i} \quad \text{(g mol-1)} \qquad (2)$$

where M_i represents the ith nominal mass (m/z) and I_i its ion intensity (in 106 counts mg-1). The more Mw and Mn differ, the more the polydispersity and, therefore, the structural heterogeneity increases. The thermostability was calculated as the ratio of measured ion intensities > 410 °C proportional to the total ion intensity (in 106 counts mg-1) of the sample and is expressed as a percentage (%). The hexose to pentose ratio (H/P) was calculated from the sum of ion intensities by masses characteristic for galactose and mannose (m/z 126, 127, 144, 145, 162 and 163) divided by the sum of ion intensities by masses characteristic for arabinose and xylose (m/z 114, 115, 132 and 133). The chain-length ratio of even-numbered fatty acids (E4-26/E26-38) was calculated from the sum of ion intensities by masses of fatty acids in the range C4 to C26 divided by the sum of ion intensities by masses of fatty acids in the range C26 to C38 (Schnitzer et al., 1986).

3. Results

3.1 SOM main characteristics and molecular composition

Representing the range of geological substrates and soil properties in the study area, the main bulk soil characteristics of the samples (n=4) covered pH-values from 5.9 to 7.1, contents of OC from 14.1 to 47.2 mg g-1, of carbonate C from 0.0 to 14.3 mg g-1 and of total N from 1.2 to 2.7 mg g-1. Subfractions of SOC ranged from 285 to 378 mg kg-1 for microbial C, from 510 to 800 mg kg-1 for hot water extractable C and from 5.8 to 13.9 mg g-1for OC375. The soil texture varied from sandy loam to silty clay with clay contents from 13 up to 41% and sand contents from 8 up to 57% (Tab. 1). The total SOM contents decreased in the order stCM (8.2%) > haST (3.2%) > haCM (2.8%) > haFL (2.4%) (Tab.2). About 5(±4)% of the SOM was recovered in the fPOM fraction, 8(±7)% in the oPOM1.6 and 19(±9)% in the oPOM2.0 fraction. Most of the SOM though, was MOM derived (68(±17)%, Tab. 2).

Py-FIMS of the bulk soil samples assigned the majority of identifiable signals to phenols and lignin monomers (25.3%) and alkylaromatics (24.7%) (Tab. 2). The identified SOM molecular compounds of the MOM mainly consisted of phenols and lignin monomers (27.5%) > alkylaromatics (24.4%) > carbohydrates (15.3%) > peptides (13.0%) > N-containing compounds (8.3%). The three different POM fractions showed similar mean contents of major compound classes: phenols and lignin monomers (20.1%) = alkylaromatics (20.1%) > lipids, alkanes, alkenes, bound fatty acids and alkylmonoesters (12.2%) > peptides (9.9%) > sterols (8.6%) > carbohydrates (8.4%). So for both, POM and MOM, the majority of identified SOM consisted of phenols, lignin monomers and alkylaromatics. Together they accounted on average for about 40% of identified POM and 52% of identified MOM (Tab. 2). However, the strongest single ion intensities of bulk soil SOM as well as of MOM were recorded for signals assigned to carbohydrates (m/z 96, 110) and lignin monomers (m/z 110, 208) (Fig. 1). The single ion intensities of the POM fractions in contrast were largest for the plant-derived lipids linoleic acid (m/z 278, 280) and lignin (m/z 178, 208, 252), for the microbial or plant-derived lipids oleic acid (m/z 282) and palmitic acid (m/z 256), and for the fungal marker ergosterol (m/z 396).

Difference spectra (supplementary Fig. 1) between the bulk soils and the MOM show that in three soil samples the carbohydrates at m/z 82, 96 and 110 were relatively enriched in the MOM fraction. When the carbohydrate subunits appeared as marker signals, their origin could even be differentiated in hexose (m/z 126) as rather enriched in the MOM fraction of the haFL and pentose (m/z 114) as rather enriched in the POM fractions of the haCM. Other

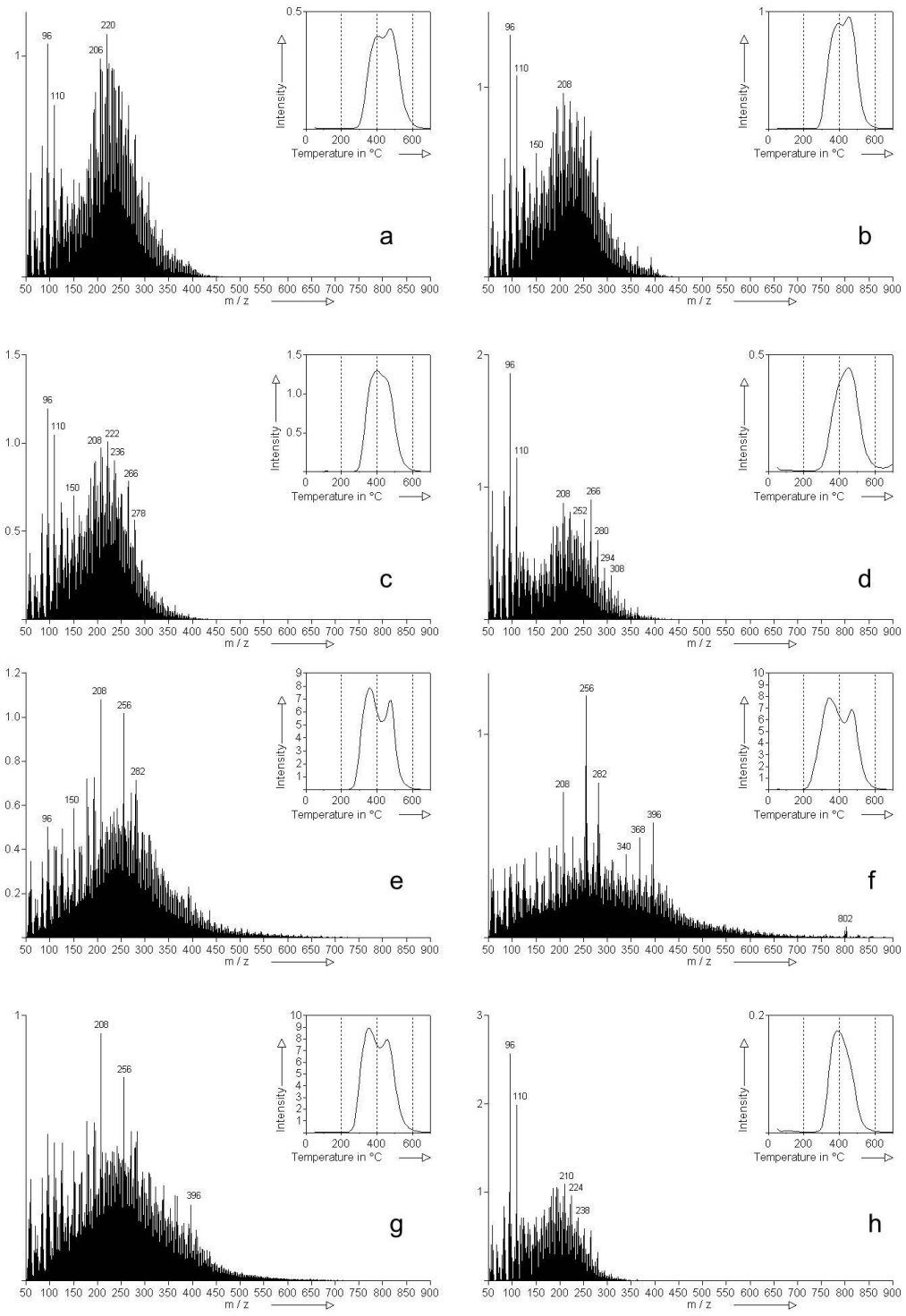


Figure 1: Pyrolysis-field ionization mass spectra and their most prominent masses (indicated by their m/z number) and the respective thermogramms of the bulk soil samples of the four soil types with (a) Haplic Fluvisol (haFL); (b) Haplic Cambisol (haCM); (c) Haplic Stagnosol (haST) and (d) Stagnic Cambisol (stCM) and exemplarily of the haST density fractions with (e) the free particulate organic matter fraction (fPOM < 1.6 g cm-3), (f) the light occluded particulate organic matter fraction (oPOM <1.6 g cm-3), (g) the heavy occluded particulate organic matter fraction (oPOM <2.0 g cm-3) and (h) the mineral associated organic matter fraction (MOM > 2.0 g cm-3)

Table 2: Mean content of soil organic matter (SOM), polydispersity, thermostability, percentage of volatilized matter and means (± SD, n=3) of SOM compound classes[a] (percentage of total ion intensity) identified in the four soil samples. Significant differences within each compound class between the four bulk soil samples are labeled with lower case letters (Kruskal-Wallis H-Test and T-test with $p < 0.05$; **$p < 0.01$ and ***$p < 0.001$).

sample	SOM (mg g^{-1})	PD[b]	TS[c]	% VM[d]	SOM compound classes (% TII)										
					CHYD	PHLM	LDIM	LIPID	ALKYL	NCOMP	STER	PEPTID	SUBE	FATTY	Sum
Haplic Fluvisol	2.4 a	1.16 a	0.79 a**	5.0 ±0.4 a	5.5 ±0.1 a	10.8 ±0.1 a	3.1 ±0.0 a	4.7 ±0.2 a	12.3 ±0.1 a	3.1 ±0.1 a	0.7 ±0.1 a	5.9 ±0.1 a	0.0 ±0.0 a	0.2 ±0.1 a	46.3 ±0.2 a
fPOM <1.6 g cm^{-3}	10.2%	1.17	0.52	28.1 ±5.2	2.9 ±0.2	7.8 ±1.5	3.3 ±1.0	5.6 ±1.8	8.2 ±2.2	1.4 ±0.6	3.2 ±0.3	3.4 ±0.1	0.5 ±0.2	6.9 ±1.4	43.3 ±7.0
oPOM <1.6 g cm^{-3}	19.3%	1.20	0.67	34.2 ±3.6	1.7 ±0.1	7.1 ±0.4	3.9 ±0.6	6.2 ±0.7	10.2 ±0.2	1.0 ±0.0	3.5 ±0.1	2.8 ±0.1	0.6 ±0.1	2.3 ±0.5	39.2 ±0.3
oPOM <2.0 g cm^{-3}	22.2%	1.21	0.57	34.8 ±3.3	3.9 ±0.1	9.4 ±0.2	3.1 ±0.2	4.9 ±0.2	9.1 ±0.1	2.2 ±0.1	2.9 ±0.2	4.6 ±0.1	0.4 ±0.1	2.3 ±0.2	42.7 ±0.4
MOM >2.0 g cm^{-3}	48.3%	1.12	0.57	5.0 ±0.5	8.9 ±0.3	16.4 ±0.6	1.0 ±0.2	3.7 ±0.4	13.1 ±0.1	4.6 ±0.4	0.1 ±0.0	7.3 ±0.2	0.0 ±0.0	0.3 ±0.1	55.2 ±0.5
Haplic Cambisol	2.8 a	1.13 b	0.68 b	8.9 ±0.3 b	6.6 ±0.0 b	12.9 ±0.2 b	1.9 ±0.2 b**	4.0 ±0.3 b	11.9 ±0.1 b	3.6 ±0.1 b	0.5 ±0.1 a	6.5 ±0.1 ab	0.0 ±0.0 a	0.3 ±0.1 a	48.2 ±0.3 a
fPOM <1.6 g cm^{-3}	6.1%	1.19	0.41	38.5 ±2.0	4.6 ±0.3	11.9 ±0.6	2.8 ±0.6	5.6 ±0.2	9.7 ±0.1	1.7 ±0.2	2.9 ±0.2	4.3 ±0.2	0.4 ±0.1	3.7 ±0.7	47.7 ±0.8
oPOM <1.6 g cm^{-3}	5.9%	1.22	0.55	33.2 ±5.1	2.6 ±0.2	5.6 ±0.5	2.7 ±0.4	5.1 ±0.5	6.6 ±0.6	1.6 ±0.1	5.5 ±0.2	3.2 ±0.1	1.0 ±0.1	3.9 ±0.1	37.8 ±1.9
oPOM <2.0 g cm^{-3}	28.0%	1.20	0.54	34.8 ±3.3	3.2 ±0.1	6.5 ±0.1	2.7 ±0.5	4.9 ±0.2	7.2 ±0.2	2.0 ±0.1	4.9 ±0.2	3.8 ±0.0	0.7 ±0.1	3.4 ±0.6	39.3 ±0.8
MOM >2.0 g cm^{-3}	60.1%	1.13	0.65	5.5 ±1.8	8.2 ±0.2	14.6 ±0.1	1.4 ±0.0	3.9 ±0.1	12.7 ±0.2	4.2 ±1.1	0.3 ±0.0	6.8 ±0.1	0.0 ±0.0	0.3 ±0.1	52.2 ±0.3
Haplic Stagnosol	3.2 a	1.13 b	0.68 b	14.9 ±0.3 bc	6.5 ±0.3 b	13.3 ±0.5 b	1.5 ±0.7 b	3.9 ±0.5 bc	12.0 ±0.2 ab	4.1 ±0.2 c	0.2 ±0.1 b	6.6 ±0.1 b	0.0 ±0.0 a	0.2 ±0.0 a	48.3 ±0.3 a
fPOM <1.6 g cm^{-3}	2.4%	1.19	0.56	27.0 ±2.0	4.2 ±0.4	9.7 ±1.0	2.8 ±0.6	4.7 ±0.1	9.0 ±0.2	2.2 ±0.3	2.2 ±0.6	4.6 ±0.2	0.4 ±0.1	2.1 ±0.5	41.8 ±1.1
oPOM <1.6 g cm^{-3}	3.8%	1.26	0.50	42.9 ±1.4	3.3 ±0.3	7.2 ±0.3	2.3 ±0.2	4.6 ±0.2	6.9 ±0.3	2.2 ±0.2	4.2 ±0.1	4.3 ±0.2	0.8 ±0.1	4.0 ±0.2	39.7 ±0.6
oPOM <2.0 g cm^{-3}	18.5%	1.21	0.55	34.9 ±1.9	4.3 ±0.1	9.4 ±0.1	2.5 ±0.2	4.3 ±0.0	8.7 ±0.0	2.6 ±0.1	2.6 ±0.1	5.1 ±0.0	0.3 ±0.1	2.1 ±0.4	41.9 ±0.3
MOM >2.0 g cm^{-3}	75.4%	1.14	0.61	6.2 ±2.1	9.8 ±0.2	17.2 ±0.1	0.5 ±0.2	3.1 ±0.1	13.8 ±0.6	6.1 ±0.2	0.0 ±0.0	7.9 ±0.1	0.0 ±0.0	0.1 ±0.0	58.5 ±0.5
Stagnic Cambisol	8.2 b**	1.26 c	0.82 a	18.5 ±2.0 c	7.5 ±0.6 b	11.0 ±1.4 ab	1.7 ±0.2 b**	3.3 ±0.3 c	10.7 ±1.0 ab	5.1 ±0.4 c	0.3 ±0.2 ab	7.5 ±0.1 c	0.0 ±0.0 a	0.1 ±0.1 a	47.2 ±2.6 a
fPOM <1.6 g cm^{-3}	2.3%	1.25	0.50	66.3 ±6.2	5.5 ±0.3	11.2 ±0.8	2.6 ±0.6	4.7 ±0.2	8.6 ±0.2	2.1 ±0.1	3.4 ±0.3	5.1 ±0.1	0.4 ±0.1	3.0 ±0.6	46.7 ±1.0
oPOM <1.6 g cm^{-3}	2.7%	1.25	0.60	58.9 ±4.7	3.0 ± 0.2	7.7 ±0.4	2.8 ±0.3	4.9 ±0.2	7.4 ±0.2	1.9 ±0.1	4.0 ±0.2	4.1 ±0.1	0.7 ±0.1	2.9 ±0.2	39.3 ±0.7
oPOM <2.0 g cm^{-3}	6.9%	1.17	0.67	35.4 ±3.0	2.4 ±0.1	6.3 ±0.1	3.7 ±0.4	5.0 ±0.2	8.3 ±0.1	1.7 ±0.1	3.4 ±0.2	3.7 ±0.2	0.4 ±0.1	1.6 ±0.3	36.4 ±0.4
MOM >2.0 g cm^{-3}	88.1%	1.16	0.87	13.1 ±2.4	5.5 ±0.2	9.8 ±0.2	3.2 ±0.3	4.6 ±0.3	12.0 ±0.1	3.8 ±0.2	0.7 ±0.1	5.5 ±0.1	0.0 ±0.0	0.1 ±0.0	45.3 ±0.3

[a]CHYD: Carbohydrates; PHLM: phenols and lignin monomers; LDIM: lignin dimers; LIPID: lipids, alkanes, alkenes, bound fatty acids and alkyl monoesters; ALKYL: alkylaromatics; NCOMP: mainly heterocyclic N-containing compounds; STER: sterols; PEPTID: peptides; SUBE: suberin; FATTY: free fatty acids (n-C16 to n-C34)

[b]polydispersity calculated using Eqs.1 and Eqs. 2

[c]thermostability calculated as the ratio of measured ion intensities > 410 °C proportional to the total ion intensity (in 106 counts mg-1) of the sample and is expressed as a percentage (%)

[d]percentage of volatilized matter during pyrolysis in means (± SD) with n=4

bulk soil POM derived groups of compounds were homologous series of shorter-chained lipids such as n-C15 to n-C21 alkenes (m/z 210-294), n-C15 to n-C19 alkanes (m/z 212-268) and C10 to C15 n-alkyl diesters (m/z 202-272) or long chained lipids such as fatty acids (m/z 312-438) of rather low ion intensities. However, regarding the distribution of either POM or MOM derived compounds, the stCM showed an inverse trend (supplementary Fig. 1).

3.2 SOM polydispersity and thermostability

The polydispersity of bulk soil OM followed the order stCM (1.26) > haFL (1.16) > haST (1.13) = haCM (1.13) (Tab. 2). When the values of Mw were related to Mn of the investigated samples, two general clusters were observed (Fig. 2). One was formed by the bulk soils and the MOM, while the other comprised all POM fractions. Comparing the distribution of these Mw to Mn values to published data of Sorge (1995), the overall POM showed closer accordance with the distribution of soil fungi, 'young' bacterial communities, grass and arable humic acids while the MOM accorded to the distribution of 'old' soil bacterial communities and arable fulvic acids. The polydispersity of the bulk soils and mean POM was strongly correlated to the ion intensities of the free fatty acids while the mean MOM showed highest values for the sterols and lignin dimers (see all correlations in supplementary Tab.1).

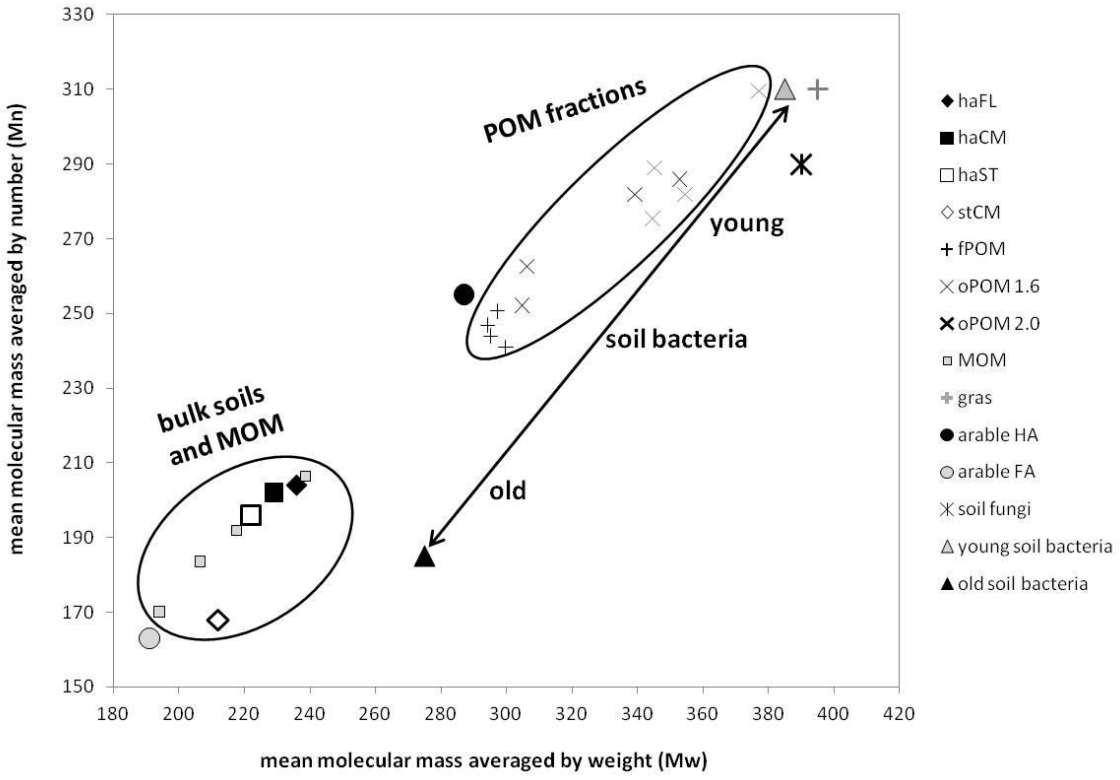


Figure 2: The ratio of the mean molecular mass averaged by number (Mn) to averaged by weight (Mw) of the bulk soils (n=4) and the mean density fractions. The ratios of 'young' and 'old' soil bacterial

communities, soil fungi, arable fulvic acid, arable humic acid and grass (Sorge, 1995) are represented as reference samples.

The thermograms of all samples and fractions revealed thermolabile (volatile < 410°C) as well as thermostable (volatile >410°C) proportions of SOM (Fig. 1). For most SOM fraction samples this was expressed in clearly bimodal thermograms (Fig. 1 e-g), in which the first peak mostly represents highly volatile compounds which are present unbound or adsorbed onto the surfaces while the second peak represents rather humified compounds present in three-dimensionally-crosslinked structures or in otherwise stabilized structures, e.g. charred organic matter. The bulk soil thermostability decreased in the same sequence as the polydispersity from stCM (82%) to haCM (68%) (Tab. 2), showing highest values for the MOM fractions. Exceptionally the haFL sample showed its significantly highest value for the oPOM1.6 fraction. The correlation of the bulk soil thermostability to certain compounds varied between the four soils, showing strongest correlations to alkylaromatics (for haFL, haST and stCM) and peptides and N-containing compounds (for haCM). The thermostability of all POM and MOM fractions correlated significantly and positively with the lignin dimers and negatively with the carbohydrates (see all correlations in supplementary Tab.1).

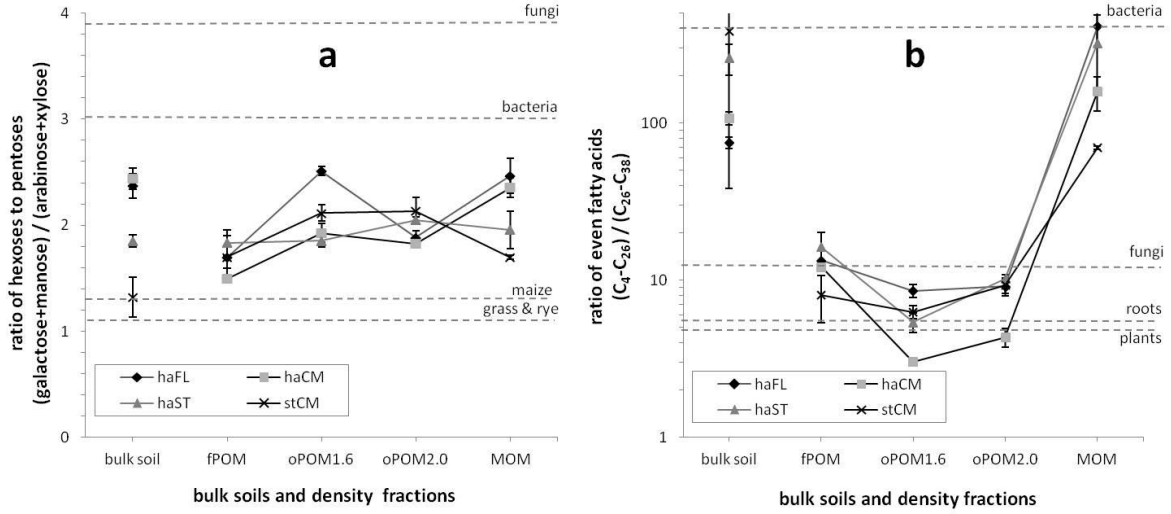


Figure 3: The ratio of a) hexoses to pentoses (H/P; with galactose and manose to arabinose and xylose) and b) even fatty acids (E4-26/E26-38; as a ratio of C4-C26 to C26-C36) of the bulk soils and their density fractions determined for the four temperate arable topsoils Haplic Fluvisol (haFL), Haplic Cambisol (haCM), Haplic Stagnosol (haST) and Stagnic Cambisol (stCM). Additionally, the mean H/P of grass (n=6), rye (n=3), maize (n=6), miscanthus (n=1), bacteria (n=3) and fungi (n=31) and the mean E/E of plants (n=9), roots (n=30), fungi (n=31) and bacteria (n=13) calculated from Py-FIMS measurements are given as reference thresholds (thin broken lines).

3.3 Microbial contribution to SOM compounds

Parameters were derived from the data indicating the contribution of microbial biomolecules to SOM. The hexose to pentose ratio (H/P) of all bulk soils and fractions ranged between 1.3 and 2.5 (Fig. 3a), which was well within the range defined by H/P ratios of plant materials and microorganisms. Respective comparative data of grass (Poa annua), rye (Secale cereale), maize (Zea mays), miscanthus (Miscanthus X giganteus) and of bacteria and fungi (Eckhardt and Leinweber, unpublished data) are additionally shown in Fig.3. The mean H/P ratios of the SOM density fractions tended to increase from fPOM (1.7±0.2) over the oPOM fractions (2.0±0.1) to the MOM fraction (2.3±0.3). As an exception stCM had the lowest H/P ratios for bulk soil SOM (1.3) and MOM (1.7). The mean ratio of even fatty acids (E4-26/E26-38) ranged from 5.8±2.3 in the oPOM1.6, to 8.2±2.6 in oPOM2.0, to 12.4±3.4 in the fPOM and up to 240.3±154.7 in the MOM fraction (Fig. 3b). Regarding respective mean comparative ratios of plant and root material (~5) and samples of fungi (~14) and bacteria (~300) (Eckhardt and Leinweber, unpublished data), the E4-26/E26-38 again indicated an increasing contribution of microbial (mainly bacterial) compounds to the MOM fractions.

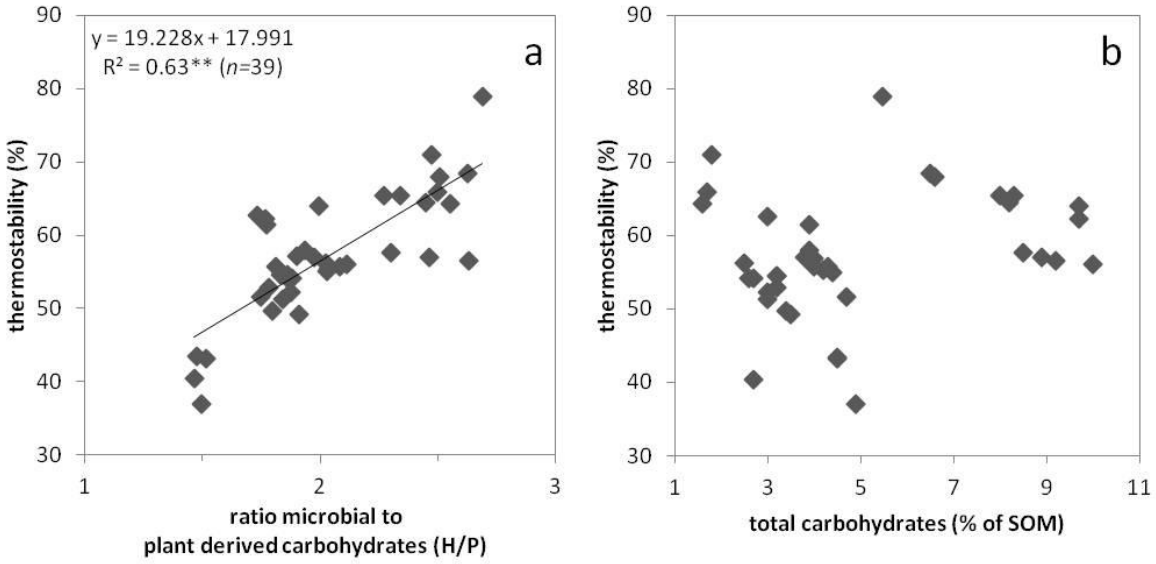


Figure 4: The correlation of the thermostability (%) with a) the ratio of microbial versus plant derived carbohydrates (H/P; hexose to pentose ratio) and with b) the total content of carbohydrates (% CHYD of total identifiable SOM). The correlation intensity is represented by the squared rank correlation coefficient r according to Spearman. The level of significance is given by superscript asterisks: *p < 0.05; **p < 0.01; ***p < 0.001.

The H/P of all soils and fractions (n=52) was significantly correlated to the polydispersity and thermostability as qualitative SOM characteristics. Correlations were substantially further improved, when the samples from stCM soil were excluded (n=39). The correlation was

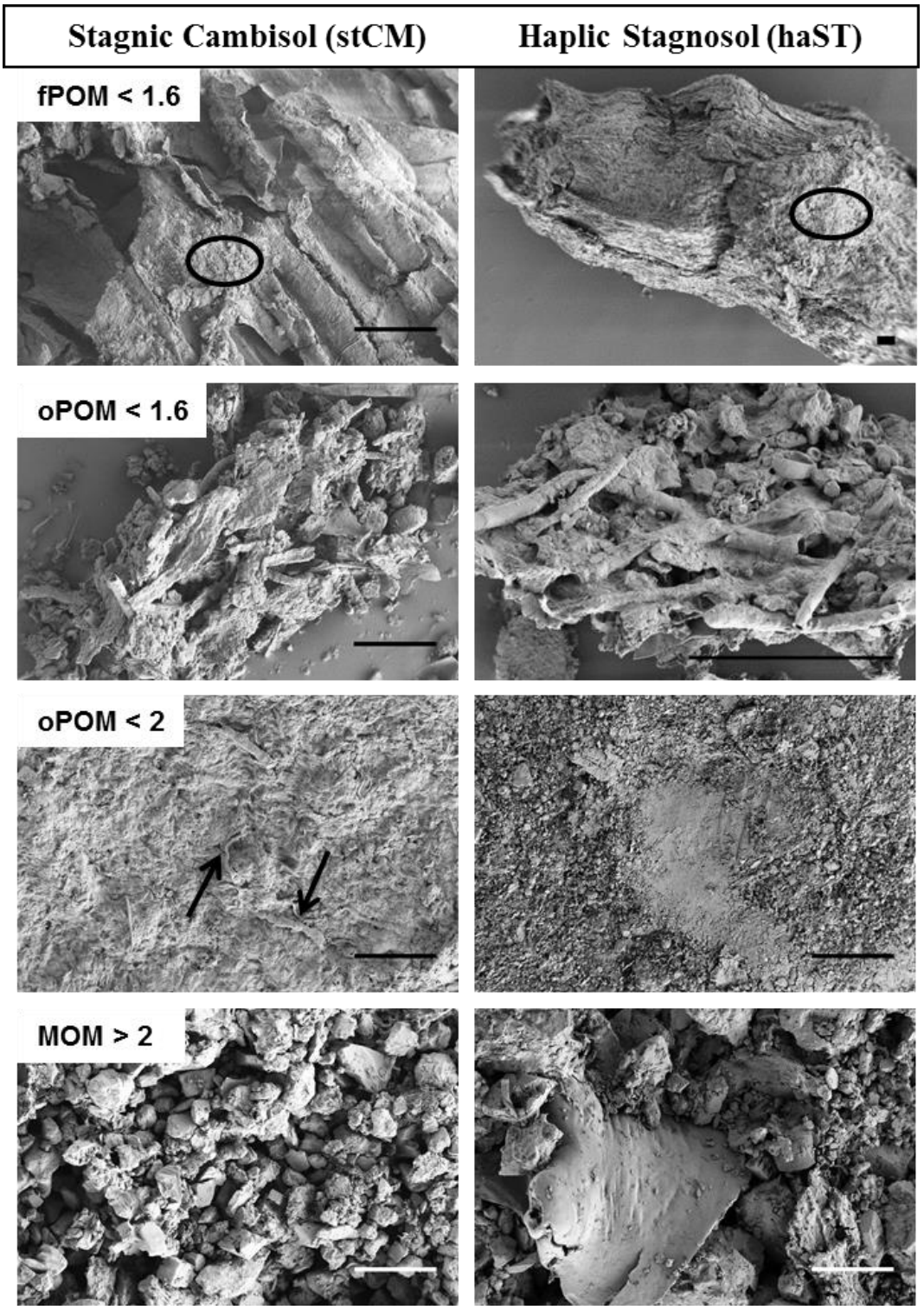


Figure 5: Scanning electron micrographs of soil samples from fractions with increasing densities (g cm-3). Occluded fungal hyphae are marked by arrows and amorphous mineral structures by circles (bar = 50 µm).

positive for the thermostability (rTS=0.33*, n=52; rTS=0.63**, n=39, Fig. 4a) and negative for the polydispersity (rPDI= -0.49**, n=52; rPDI= -0.56**, n=39; not shown). In contrast, no correlation was found for the total contribution of carbohydrates to SOM thermostability

(rTS=0.06, n=39) as shown in Fig. 4b. The E4-26/E26-38 ratios were not significantly correlated to thermostability (rTS= 0.25, n=52; rTS= 0.24, n=39; not shown).

In order to optically characterize soils and their density fractions, samples of stCM and haST were analyzed by SEM. Fig. 5 shows representative views of the different density fractions. The fPOM fractions of both soils contained plant derived structures such as partly decomposed roots, litter fragments and fragments of black carbon. Some spots of the samples were additionally covered by amorphous mineral structures (highlighted by circles in Fig. 5). The oPOM1.6 fractions were characterized by clearly visible mineral associated OM. Residues of roots, litter and fungal hyphae, embedded in a mineral matrix, were the main identifiable OM. Compared to the fPOM fractions, the organic structures seemed to be more decomposed. Plant derived residues were, in contrast, not visibly detected in the heavier oPOM2.0 fractions. Samples of this fraction were characterized by mineral structures, which were coated by organic material. In some cases, an occlusion of fungal hyphae into the organic layer was visible. The main structures of the MOM were minerals covered with organic material. Differences between the two soils in terms of structural features were not observed. For a better visualization of microbial derived residues, high magnification micrographs of the oPOM2.0 fractions and MOM were taken (Fig. 6). The SEM micrographs showed no intact bacterial cells (Fig. 6). However, the mineral surfaces of the observed soils were covered with amorphous OM and heterogeneous, flaky patchy material (some highlighted by arrows in Fig. 6). On the SEM micrographs, patchy and flaky structures with a size between 200 and 400 nm seemed to be more abundant in the MOM of haST in comparison to stCM, while the opposite was observed for the oPOM2.0 fractions. This was also observed on other micrographs (data not shown).

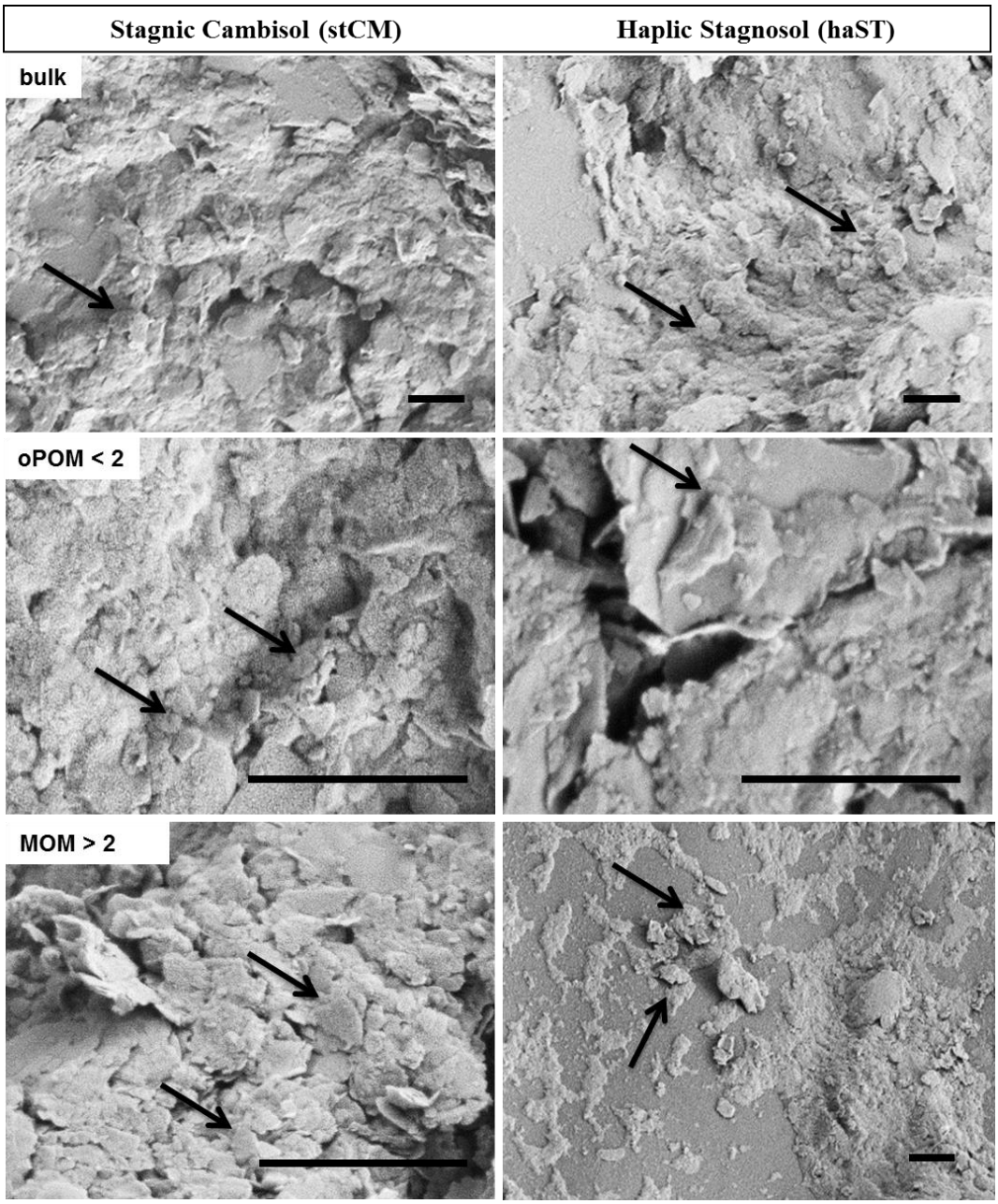


Figure 6: High magnification scanning electron micrographs of soil particles from different density fractions (g cm-3). The arrows indicate patchy organic material, which is probably microbial derived (bar = 1 µm).

4. Discussion

4.1 SOM main characteristics and molecular composition

The wide range of soil properties of the four arable soil types implied a high variability of the content and molecular composition of their SOM. Accordingly, samples and fractions from the different soil types were significantly different, especially the proportions of carbohydrates, lignin, lipids and nitriles for the bulk soils and proportions of carbohydrates for

the MOM and of lipids for the POM fractions. These compounds also showed the strongest single ion intensities in the Py-FI mass spectra, thus dominating the molecular fingerprint of the samples. This agrees with findings for density-fractionated particle-size fractions by Schulten and Leinweber (1999) who also identified the carbohydrates as key compound class for the MOM and the lipids for the POM fractions. Nevertheless, almost 50% of the identifiable SOM compounds of all soil types and all density fractions were assigned to the same two compound classes, i.e. 1) phenols and lignin monomers, and 2) alkylaromatics. Sleutel et al. (2007) reported a similar proportional distribution where these compounds accounted together for about 39 % of identified POM and 63 % of identified MOM. The similar dominance of the two compound classes is assumed to originate from the arable land-use with a common crop rotation of wheat, barley and oilseed rape resulting in the same input of plant residues at all four study sites.

4.2 SOM polydispersity and thermostability

The polydispersity as the ratio of Mw to Mn is a measure of structural molecular heterogeneity, and therefore of molecular complexity (Sorge, 1995). The Mw and Mn as well as the polydispersity of the POM fractions resembled very much corresponding values from fresh litter input, soil fungi and ‘young’ (growing) bacterial biomass, indicating a strong influence of these components on POM (Fig. 2). Hur et al. (2009) reported that the polydispersity of plant material more than halved upon microbial transformation. This would explain the slightly higher polydispersity of the POM fractions, which indicates, together with the similar molecular composition (Tab. 2), that physical occlusion but not chemical transformation discriminates the POM fractions. The contribution of plant material to the POM fractions was confirmed by SEM analysis of the fractions, where plant litter and fungal hyphae were abundant in the POM, but not observed in the MOM fractions (Fig. 5). Compared to POM, polydispersity and Mw and Mn of MOM fractions were substantially smaller (Fig. 2, Tab. 2). These values correspond to reference data for arable soil humic and fulvic acid (Fig. 2) and data published by Schulten (1987). Due to the much larger share of MOM to bulk soil SOM, similar values were determined for these samples. The MOM resembled much more ‘old’ microbial biomass analyzed by Sorge (1995), which is a cultivated microbial community in the death phase that assumably accumulated higher amounts of necromass.

The thermostability of SOM fractions increased in the sequence fPOM (49.8%) < oPOM1.6 (57.9%) ≤ oPOM2.0 (58.4%) < MOM (67.4%), which confirms previous reports on increasing SOM stability in the different density fractions (Dorodnikov et al., 2011; Kögel-Knabner et al.,

2008). However, for the haFL soil sample thermostability was highest for the oPOM 1.6 fraction. This was attributed to the specific SOM composition of this soil. The haFL contained the highest content of OC375 with 65.3% of its total OC (Tab.1). OC375 represents a thermostable fraction of OC that contributes to the continuum of black carbon (Gélinas et al., 2001). Accordingly, hard coal deposits occur in this soil from fluviatile sediments (Felten et al., 2012; Pies et al., 2008). To prove for the presence of black carbon in the Fluvisol POM fractions, an additional high resolution Py-FIMS measurement was performed (not shown). PAH marker signals were clearly identified for the oPOM fractions, which are an indication of black carbon materials (Pies et al., 2008). However, they were not detected in the MOM fraction, indicating that black carbon in this arable soils is sequestered rather due to occlusion in aggregates than to mineral association.

Given that SOM is to a great extent formed due to assimilation into microbial biomass and resulting microbial products, SOM formation mainly depends on the microbial substrate use efficiency (Cotrufo et al., 2012). The extent to which the different SOM compounds subsequently become stabilized and contribute to SOM quantity and quality is by contrast mainly determined by the interaction with the soil matrix (Six et al., 2006; Kögel-Knabner et al., 2008). For the four considered soils, the compound classes contributing in highest abundance to SOM were similar, while the overall composition of SOM was significantly different due to key compounds like carbohydrates and lipids. However, it could not be clarified in this study whether the abundance of compounds was rather due to soil matrix stabilization and/or to microbial substrate use efficiency. Furthermore, each SOM compound class showed comparable contributions to the thermolabile as well as to the thermostable section of the bimodal thermograms (Fig. 1). This indicates that even humified or mineral associated OM consists of at least two functional pools, which represent different stability and MRTs (Schulten and Leinweber, 1999; Bird et al., 2008). Altogether this emphasizes that SOM formation and persistence, which are governed by complex physicochemical and biological interactions of the surrounding environment, are accordingly rather features of the soil ecosystem than of the inherent molecular properties (Schmidt et al., 2011).

4.3 Microbial contribution to SOM

To be able to estimate the microbial contribution to the SOM contents, respective proxies were calculated for the carbohydrates and lipids with respect to their identification as key SOM compound classes of density fractions. The hexose to pentose ratio (H/P) was calculated for the carbohydrates (Fig. 3a) and the ratio of even fatty acids (E4-26/E26-38) was calculated for the lipids (Fig. 3b) to analyze their origin. Previous studies propose the

use of the H/P for the determination of either plant (<0.5) or microbial (>2) origin of carbohydrates (Oades, 1984; Guggenberger, 1994). Py-FIMS analysis of relevant plant or microbial derived reference materials revealed H/P values of ≥ 3 for microbial derived and of < 1.2 for plant derived OM (Fig. 3a). These values confirmed that higher H/P ratios indicate microbial derived carbohydrates, while lower H/P ratios indicate plant derived carbohydrates, although they deviated from the published data (Oades, 1984; Guggenberger, 1994), which is at least in part due to different analytical methods. Comparing the H/P of bulk SOM and fractions with these references, showed a substantial microbial influence with a roughly estimated, overall contribution of microbial versus plant-derived SOM of about 1:1. The OM stabilization in soil tended to increase with microbial contribution to SOM (Fig. 3a). This was further indicated by the strong positive correlation of thermostability with explicitly the microbial derived (Fig. 4a) but not with the plant derived carbohydrates. In contrast, OM thermostability of the different fractions was weakly negatively correlated with the total amount of carbohydrates. This might be explained by contribution of plant-derived celluloses with low thermal stability to the class of carbohydrates as assigned in Py-FI mass spectra. Hence, only microbial derived carbohydrates seem to have a stabilizing function within SOM.

The results of the E4-26/E26-38 ratios (Fig.3b) emphasize the relative enrichment of microbial products for the MOM fractions. According to Schnitzer et al. (1986), fatty acids in the range C4 to C26 are predominantly microbiological, whereas C26 to C38 acids are found principally in waxes of insects and plants. Thus, the more the E4-26/E26-38 ratio increases, the higher is the microbial contribution to SOM, which is also reflected by respective comparative data of plant materials and microorganisms as shown in Fig.3b. Microbial fatty acids are in contrast to the carbohydrates not hypothesized to enhance the thermostability, especially due to the high stability of plant-derived waxes, which are assumed to counteract the correlation of the E4-26/E26-38 ratio and the thermostability (see 3.3). The present study confirms Guggenberger et al. (1999), who showed a general increase of microbial biomass contribution from POM to MOM on the basis of wet-chemical analyses. Furthermore, we also confirm findings by Sleutel et al. (2009) who reported a relative increase of carbohydrate and also of peptide marker peaks for the MOM in Py-FIMS analyses. Peptides can be derived from microbial cell wall and cytosol components and therefore may indicate microbial necromass structures. They are easily stabilised in soil (Kleber et al., 2007) and afterwards highly persistent, as shown e.g. in an incubation experiment with 13C-labelled Escherichia.coli cells by Miltner et al. (2009).

From the Py-FI mass signals it was not deducible if compounds were either derived from living microbial biomass or from necromass. Therefore, SEM was used to characterize and

visualize the fractions with regards to their structural features, in particular microbial biomass. In general, no intact bacteria were detected (Fig. 5 and Fig. 6), which corresponds to the fact that the soil surface area covered by living microbes is generally far less than 1% (Chenu and Stotzky, 2002). The treatment during density fractionation may further reduce the number of intact cells, but the direct image comparison of airdried to SPT-treated samples did not indicate differences regarding the bacterial biomass. Bacterial residues found were therefore assumed to result from enriched stable necromass. High magnification micrographs of the denser fractions revealed the presence of cell wall fragments with sizes smaller than 1 µm, confirming the findings of Miltner et al. (2012), that cell wall fragments are a significant source of SOM. These fragments were highly abundant in the MOM and the oPOM2.0 fractions (Fig. 6). The fragments are very similar in shape and size to those previously shown to be microbial derived residues, i.e. multiple layers of fragmented bacterial cell-envelope (Miltner et al., 2012; Schurig et al., 2013). A recent study by Schütze et al. (2013) supports these findings with the observation that the decay of several Streptomyces strains in heavy metal contaminated soils was accompanied by the formation of patchy and flaky cell envelope fragments. The observed structures determined in this study may therefore be attributed presumably to microbial necromass, in particular cell envelope fragments. The apparently higher abundance of patchy and flaky fragments in the MOM of haST compared to stCM (Fig.6) correlates with the H/P and E/E values for these fractions (Fig. 3). The opposite findings for the oPOM2.0 fractions are also reflected in the H/P values. Assuming these ratios as indicators for microbial predominance, the correlations support the hypothesis, that the fragments are microbial derived.

Although the methods used could not exactly quantify the microbial contribution to SOM, the presumably stabilizing effect of microbial carbohydrates and the increment of microbial necromass with stable SOM is emphasizing the significance of the microbial contribution to SOM formation and persistence.

5. Conclusions

By comparison of the density fractions of four significantly different temperate arable soils, this study aimed to elucidate the microbial contribution to SOM formation and persistence. The relative microbial contribution to arable soil OM was identified as significant within every SOM fraction, and thus, assumed to be crucial for SOM formation in general. SOM persistence was related to microbial SOM origin by increased thermostability due to an increased proportion of microbial versus plant-derived carbohydrates. Although the microbial contribution versus plant derived input to arable SOM was estimated relatively high for all

fractions with a mean ratio of 1:1 regarding the carbohydrates, it generally increased with density, and thus especially correlated with MOM thermostability. Additionally, microbial necromass is assumed to significantly contribute to arable SOM quantity and quality. In conclusion, the methods to exactly quantify microbial SOM origin and to further distinguish the actually contributing microbial state (active, potentially active, dormant, dead) need to be advanced, to understand and manage the dynamic biogeochemical processes of SOM formation and persistence. The SEM analysis in this study, although using high-magnification microscopy, is only based on the visible features of the samples revealing qualitative differences with respect to the amount of fragmented cell residues. For a further investigation and quantification of the microbial fragments, an image analysis scheme could provide quantitative data and improve the validity of the findings. Another important step forward would be the analysis of the chemical composition on the scale of 100 – 500 nm, e.g. by confocal Raman imaging microscopy.

Acknowledgements

The authors thank Sabina Hens and Petra Ziegler (University of Trier) for their profound work at the laboratory and Andrea Prager (Leibniz Institute of Surface Modification, Leipzig) for her help with the scanning electron microscopy measurements. This study was funded by the DFG (SOMScapeS, TH678/12-1 and Priority Program 1315, “Mi598/2-2). The comments of two anonymous reviewers, who helped to improve the quality of the manuscript, are also gratefully acknowledged.

References

Anderson, T.H., Domsch, K.H. (2010): Soil microbial biomass: the eco-physiological approach. Soil Biology and Biochemistry 42, 2039-2043

Bird, J.A., Kleber, M., Torn, M.S. (2008): C-13 and N-15 stabilization dynamics in soil organic matter fractions during needle and fine root decomposition. Organic Geochemistry 39, 465-477

Blagodatskaya, E., Kuzyakov, Y. (2013): Active microorganisms in soil: Critical review of estimation criteria and approaches. Soil Biology and Biochemistry 67, 192-211

Bradford, M.A., Keiser, A.D., Davies, C.A., Mersman, C.A., Strickland, C.S. (2013): Empirical evidence that soil carbon formation from plant inputs is positively related to microbial growth. Biogeochemistry 113, 271-281

Bruun, S., Agren, G.I., Christensen, B.T., Jensen, L.S. (2010): Measuring and modeling continuous quality distributions of soil organic matter. Biogeosciences 7, 27-41

Chenu, C., Stotzky, G. (2002): Interactions between microorganisms and soil particles: an overview. In: Huang, P.M., Bollag, J.-M., Senesi, N. (eds) Interactions between soil particles and microorganisms. Wiley, Chichester, 3–40

Christensen, B.T. (1992): Physical fractionation of soil and organic matter in primary particle size and density separates. Advances in Soil Science 20, 1–90

Cotrufo, M.F., Wallenstein, M.D., Boot, C., Denef, K., Paul, E. (2012): The microbial efficiency-matrix stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? Global Change Biology 19, 988-995

Dorodnikov, M., Kuzyakov, Y., Fangmeier, A., Wiesenberg, G. (2011): C and N in soil organic matter density fractions under elevated CO2: Turnover vs. stabilization. Soil Biology and Biochemistry 43, 579-589

Dümig, A., Häusler, W., Steffens, M., Kögel-Knabner, I. (2012): Clay fractions from a soil chronosequence after glacier retreat reveal the initial evolution of organo-mineral associations. Geochimica et Cosmochimica Acta 85, 1-18

Felten, D., Emmerling, C. (2012): Accumulation of Miscanthus-derived carbon in soils in relation to soil depth and duration of land use under commercial farming conditions. Journal of Plant Nutrition and Soil Science 175, 661-670

Gélinas, Y., Prentice, K.M., Baldock, J.A., Hedges, J.I. (2001): An improved thermal oxidation method for the quantification of soot/graphite black carbon in sediments and soils. Environmental Science and Technology 35, 3519-3525

Golchin, A., Baldock, J.A., Clarke, P., Higashi, T., Oades, J.M. (1997): The effects of vegetation and burning on the chemical composition of soil organic matter of a volcanic ash soil as shown by 13C NMR spectroscopy. II. Density fractions. Geoderma 76 (3-4), 175-192

Gregorich, E.G., Carter, M.R., Angers, D.A., Monreal, C.M., Ellert, B.H. (1994): Towards a minimum data set to assess soil organic matter quality in agricultural soils. Canadian Journal of Soil Science 74, 367-385

Guggenberger, G., Christensen, B.T., Zech, W. (1994): Land-use effects on the composition of organic matter in particle-size seperates of soil: I. Lignin and carbohydrate signature. European Journal of Soil Science 45, 449-458

Guggenberger, G., Frey, S.D., Six, J., Paustian, K., Elliott, E.T. (1999): Bacterial and fungal cell-wall residues in conventional and no-tillage agroecosystems. Soil Science Society of America Journal 63, 1188-1198

Hartge, K. H., Horn, R. (1989): Die physikalische Untersuchung von Böden. Ferdinand Enke Verlag, Stuttgart

Helfrich, M., Flessa, H., Mikutta, R., Dreves, A., Ludwig, B. (2007): Comparison of chemical fractionation methods for isolating stable soil organic carbon pools. European Journal of Soil Science 58, 1316-1329

Helfrich, M., Flessa, H., Dreves, A., Ludwig, B. (2010): Is thermal oxidation at different temperatures suitable to isolate soil organic carbon fractions with different turnover? Journal of Plant Nutrition and Soil Science 173, 61-66

Hur, J., Park, M.H., Schlautman, M.A. (2009): Microbial transformation of dissolved leaf litter organic matter and its effects on selected organic matter operational descriptors. Environmental Science and Technology 43, 2315-2321

Jandl, G., Leinweber, P., Schulten, H.-R., Eusterhues, K. (2004): The concentrations of fatty acids in organo-mineral particle-size fractions of a Chernozem. European Journal of Soil Science 55, 459-469

Jandl, G., Leinweber, P., Schulten, H.-R., Ekschmitt, K. (2005): Contribution of primary organic matter to the fatty acid pool in agricultural soils. Soil Biology & Biochemistry 37, 1033-1041

Jastrow, J.D., Miller, R.M., Boutton, T.W. (1996): Carbon dynamics of aggregate-associated organic matter estimated by carbon-13 natural abundance. Soil Science Society of America Journal 60, 801-807

Joergensen, R.G. (1995): The fumigation-extraction method to estimate soil microbial biomass: extraction. with 0.01 M CaCl2. Agrobiological Research 48, 319–324

John, B., Yamashita, T., Ludwig, B., Flessa, H. (2005): Storage of organic carbon in aggregate and density fractions of silty soils under different types of land use. Geoderma 128, 63-79

Kindler, R., Miltner, A., Richnow, H.-H., Kästner, M. (2009): Fate of gram-negative bacterial biomass in soil - mineralization and contribution to SOM. Soil Biology and Biochemistry 38, 2860–2870

Kleber, M., Sollins, P., Sutton, R. (2007): A conceptual model of organo-mineral interactions in soils: self-assembly of organic molecular fragments into zonal structures on mineral surfaces. Biogeochemistry 85, 9–24

Knicker, H. (2011): Soil organic N – an under-rated player for C sequestration in soils? Soil Biology and Biochemistry 43, 1118-1129

Kögel-Knabner, I. (2002): The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. Soil Biology and Biochemistry 34, 139-162

Kögel-Knabner, I., Guggenberger, G., Kleber, M. (2008): Organo-mineral associations in temperate soils: integrating biology, mineralogy, and organic matter chemistry. Journal of Plant Nutrition and Soil Science 171, 61-82

Lal, R. (2009): Challenges and opportunities in soil organic matter research. European Journal of Soil Science 60, 158–169

Landgraf, D., Böhm, C., Makeschin, F. (2003): Dynamic of different C and N fractions in a Cambisol under five year succession fallow in Saxony (Germany). Journal of Plant Nutrition and Soil Science 166, 319–325

Lattimer, R.P., Schulten, H.-R. (1983): Field desorption of hydrocarbon polymers. International Journal of Mass Spectrometry and Ion Physics 52, 105-116

Leinweber, P., Jandl, G., Eckhardt, K.-U., Schlichting, A., Hofmann, D., Schulten, H.-R. (2009): Analytical pyrolysis and soft-ionization mass spectrometry. In: N. Senesi, B. Xing, P.M. Huang (eds). Biophysico-chemical processes involving natural nonliving organic matter in environmental systems. John Wiley & Sons, Inc. New York, p. 539-588.

Liang, C., Balser, T.C. (2010): Microbial production of recalcitrant organic matter in global soils: implications for productivity and climate policy. Nature Reviews Microbiology 29, 79Liang, C., Cheng, G., Wixon, D.L., Balser, T.C. (2011): An Absorbing Markov Chain approach to understanding the microbial role in soil carbon stabilization. Biogeochemistry 106, 303-309

Miltner, A., Bombach, P., Schmidt-Brücken, B., Kästner, M. (2012): SOM genesis: microbial biomass as a significant source. Biogeochemistry 111, 41-55

Miltner, A., Kindler, R., Knicker, H., Richnow, H.-H., Kästner, M. (2009): Fate of microbial biomass-derived amino acids in soil and their contribution to soil organic matter. Organic Geochemistry 40, 978–985

Miltner, A., Zech, W. (1998): Carbohydrate decomposition in beech litter as influenced by aluminium, iron and manganese oxides. Soil Biology and Biochemistry 30, 1-7

Oades, J.M. (1984): Soil organic matter and structural stability, mechanisms and implications for management. Plant and Soil 76, 319-337

Pies, C., Hoffmann, B., Petrowsky, J., Yang, Y., Ternes, T.A., Hofmann, T. (2008): Characterization and source identification of polycyclic aromatic hydrocarbons (PAHs) in river bank soils. Chemosphere 72, 1594-1601

Rumpel, C., Eusterhues, K., Kögel-Knabner, I. (2010): Non-cellulosic neutral sugar contribution to mineral associated organic matter in top- and subsoil horizons of two acid forest soils. Soil Biology and Biochemistry 42, 379-382

Schmidt, M.W.I, Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S. E. (2011): Persistence of soil organic matter as an ecosystem property. Nature 478, 49-56

Schnitzer, M., Hindle, C.A., Meglic, M. (1986): Supercritical gas extraction of alkenes and alkaloid acids from soils and humic materials. Soil Science Society of America Journal 50, 913-919

Schulten, H.-R., Leinweber, P. (1999): Thermal stability and composition of mineral-bound organic matter in density fractions of soil. European Journal of Soil Science 50, 237-248

Schulten, H.-R. (1987): Pyrolysis and softionization mass spectrometry of aquatic/terrestrial humic substances and soils. Journal of Analytical and Applied Pyrolysis 12, 149-186

Schurig, C., Smittenberg, R.H., Berger, J., Kraft, F., Woche, S.K., Goebel, M.-O., Heipieper, H.J., Miltner, A., Kästner, M. (2013): Microbial cell-envelope fragments and the formation of soil organic matter: a case study from a glacier forefield. Biogeochemistry 113, 595-612

Schütze, E., Miltner, A., Nietzsche, S., Achtenhagen, J., Klose, M., Merten, D., Greyer, M., Roth, M., Kästner, M., Kothe, E. (2013): Live and death of streptomyces in soil—what happens to the biomass? Journal of Plant Nutrition and Soil Science 176, 665–673

Simpson, A. J., Simpson, M. J., Smith, E., Kelleher, B. P. (2007): Microbially Derived Inputs to Soil Organic Matter: Are Current Estimates Too Low? Environmental Science and Technology 41, 8070-8076

Six, J., Frey, S.D., Thiet, R.K., Batten, K.M. (2006): Bacterial and fungal contributions to carbon sequestration in agroecosystems. Soil Science Society of America Journal 70, 555-569

Six, J., Elliot, E.T., Paustian, K. (2000): Soil macroaggregate turnover and microaggregate formation: a mechanism for C sequestration under no-tillage agriculture. Soil Biology and Biochemistry 32, 2099-2103

Six, J., Schultz, P.A., Jastrow, J.D., Merckx, R. (1999): Recycling of sodium polytungstate used in soil organic matter studies. Soil Biology and Biochemistry 31, 1193-1196

Sleutel, S., Leinweber, P., Begum, S.A., Kader, M.A., De Neve, S. (2009): Shifts in organic matter composition following treatment with sodium hypochlorite and hydrofluoric acid. Geoderma 149, 257-266

Sleutel, S., Kader, M.A., Leinweber, P., D´Haene, K., De Neve, S. (2007): Tillage management alters surface soil organic matter composition: a pyrolysis mass spectroscopy study. Soil Science Society of America Journal 71, 1620-1628

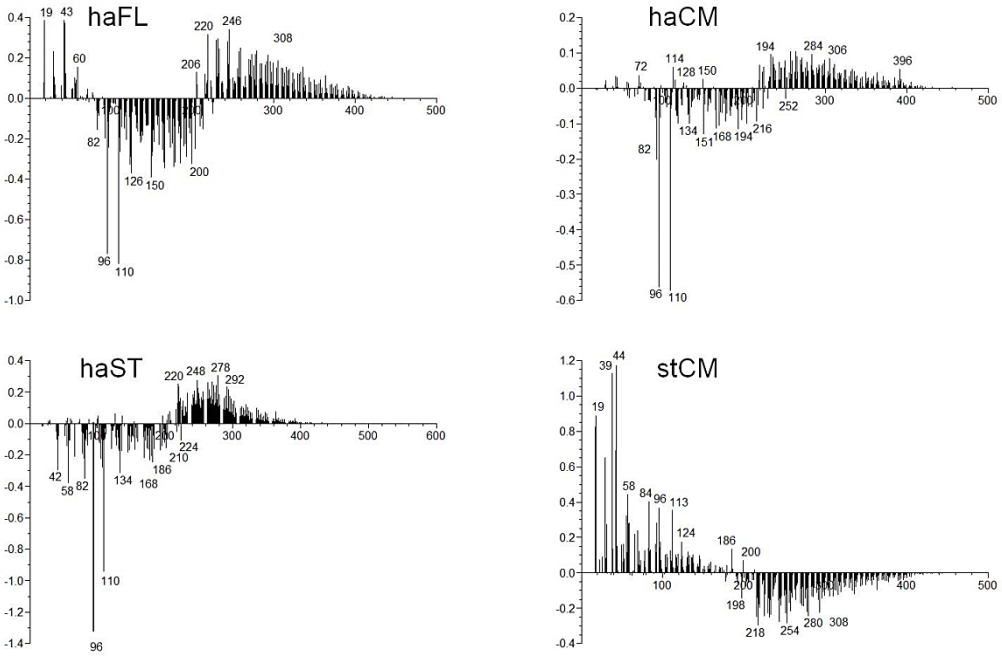
Sorge, C. (1995): Struktur der organischen Substanz in Böden und Partikelgrößenfraktionen: Pyrolyse-Gaschromatographie/Massenspektrometrie und Pyrolyse-Feldionisation Massenspektrometrie. Dissertation zur Erlangung des Doktorgrades. Schriftenreihe Institut für Pflanzenernährung und Bodenkunde Universität Kiel Nr.32 (1995) Herausgeber: H.-P. Blume, R. Horn, B. Sattelmacher

von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., Marschner, B. (2007): SOM fractionation methods: Relevance to functional pools and to stabilization mechanisms. Soil Biology and Biochemistry 39, 2183-2207

Supplementary Table 1: Results for the correlation analyses of the thermostability (TS) and the polydispersity (PDI) with each of the SOM compound classesa of the investigated soils (Haplic Fluvisol (haFL), Haplic Cambisol (haCM), Haplic Stagnosol (haST) and Stagnic Cambisol (stCM)) and the density fractions the free particulate organic matter fraction (fPOM < 1.6 g cm-3), the light occluded particulate organic matter fraction (oPOM < 1.6 g cm-3), the heavy occluded particulate organic matter fraction (oPOM < 2.0 g cm-3) and the mineral associated organic matter fraction (MOM > 2.0 g cm-3). The correlation intensity is represented by the rank correlation coefficient r according to Spearman. The level of significance is given by superscript asterisks: *p < 0.05; **p < 0.01; ***p < 0.001 or not significant (n.s.).

		CHYD	PHLM	LDIM	LIPID	ALKYL	NCOMP	STER	PEPTID	SUBE	FATTY
TS	haFL (n=15)b	n.s.	n.s.	n.s.	n.s.	0.556*	n.s.	n.s.	n.s.	n.s.	-0.633*
	haCM (n=15)	0.591*	n.s.	-0.749**	-0.935**	0.522*	0.801**	-0.565*	0.700**	n.s.	-0.849**
	haST (n=15)	0.541*	0.623*	n.s.	-0.595*	0.709**	0.536*	-0.796**	0.641*	-0.817**	-0.858**
	stCM (n=15)	n.s.	n.s.	n.s.	n.s.	0.839**	0.719**	-0.868**	n.s.	-0.723**	-0.964**
	bulk soils (n=12)c	n.s.	-0.607*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	-0.679*
	fPOM1.6 (n=12)	-0.348*	n.s.	0.637*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	oPOM1.6 (n=12)	-0.781**	n.s.	0.828**	0.615*	0.791**	-0.810**	n.s.	-0.660*	n.s.	-0.929**
	oPOM2.0 (n=12)	-0.732**	n.s.	0.916**	n.s.	n.s.	-0.608*	n.s.		n.s.	-0.736**
	POM (n=36)	-0.626**	-0.494**	0.704**	n.s.	n.s.	-0.336*	n.s.	-0.402*	n.s.	-0.557**
	MOM (n=12)	-0.937**	-0.958**	0.945**	0.830**	-0.776**	-0.637*	0.934**	-0.895**	n.s.	n.s.
PDI	haFL (n=15)b	-0.643**	-0.736**	n.s.	n.s.	-0.914**	-0.635*	0.714**	-0.681**	0.755**	0.843**
	haCM (n=15)	-0.942**	-0.896**	0.826**	0.756**	-0.964**	-0.906**	0.954**	-0.965**	0.959**	0.927**
	haST (n=15)	-0.793**	-0.867**	0.661**	0.730**	-0.926**	-0.753**	0.975**	-0.842**	0.972**	0.973**
	stCM (n=15)	n.s.	n.s.	-0.800**	-0.519*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	bulk soils (n=12)c	n.s.	-0.630*	n.s.	-0.609*	-0.839**	0.653*	n.s.	0.776**	n.s.	n.s.
	fPOM1.6 (n=12)	n.s.	n.s.	n.s.	n.s.	-0.908**	n.s.	n.s.	n.s.	n.s.	0.622*
	oPOM1.6 (n=12)	0.771**	n.s.	-0.797**	-0.745**	-0.647*	0.757**	n.s.	0.714**	n.s.	n.s.
	oPOM2.0 (n=12)	0.848**	0.658*	-0.766**	-0.586*	n.s.	0.787**	n.s.	0.651*	n.s.	n.s.
	POM (n=36)	n.s.	n.s.	-0.599**	-0.444**	-0.623**	n.s.	n.s.	n.s.	0.482**	0.544**
	MOM (n=12)	-0.644*	-0.726**	0.630*	n.s.	n.s.	n.s.	0.692*	-0.588*	n.s.	-0.654*

[a]CHYD: Carbohydrates; PHLM: phenols and lignin monomers; LDIM: lignin dimers; LIPID: lipids, alkanes, alkenes, bound fatty acids and alkyl monoesters; ALKYL: alkylaromatics; NCOMP: mainly heterocyclic N-containing compounds; STER: sterols; PEPTID: peptides; SUBE: suberin; FATTY: free fatty acids (n-C16 to n-C34)
[b]the bulk soil sample and its four respective density fraction samples with each three replicates (n=15)
[c]all four bulk soil samples with each three replicates (n=12)



Supplementary Figure 1: Difference pyrolysis-field ionization mass spectra of four temperate arable topsoils (haFL: Haplic Fluvisol, haCM: Haplic Cambisol, haST: Haplic Stagnosol, stCM: Stagnic Cambisol): spectra of the bulk soils SOM minus spectra of the mineral associated OM (MOM). Positive values indicate larger relative ion intensities in the bulk soils SOM; negative values indicate larger relative ion intensities in the MOM.

6 Synthesis and Conclusions

We are currently facing an exciting shift in both disciplines of this interdisciplinary research study, i.e. (i) SOM research and (ii) earth observation and environmental analysis using non-invasive spectroscopy. In SOM research, this shift concerns long-standing theories that are currently under revision by the soil science community, while the field of earth observation is breaking a technical frontier by switching from multi- to hyperspectral data which will improve the availability of large-scale data in terms of its depth of information and coverage of data by new satellites planned to be launched soon. Applying and combining these promoting fields may advance answering the overall question, what are most informative indicators of soil and SOM quality that might be used for global soil surveying?

The selection of appropriate indicators is the main step to a relevant and presentative assessment of soil quality (Doran and Parkin, 1996; Askari et al., 2015). However, selecting soil monitoring variables remains a challenge (Zornoza et al., 2007) as the determination of any a priori soil quality criterion or ecosystem threshold is considered as subjective since it relies only on expert opinions (Sojka and Upchurch, 1999; Velasquez et al., 2007; Cécillon et al., 2009). Therefore, the evaluation of soil quality is usually a combination of chemical, physical and biological indicators as a minimum data set (MDS) (Askari et al., 2015). The choice of the variables for this MDS is commonly derived statistically from principal component analysis (PCA), which reduces data redundancy among soil parameters (Askari et al., 2015) for identifying an essential data set in terms of soil functioning (Cécillon et al., 2009). According to the particular soil attribute of interest, these MDS can furthermore target on a specific soil function (Cécillon et al., 2009), a particular soil ecosystem service (Velasquez et al., 2007) or a key soil threat (Morvan et al., 2008; Orgiazzi et al., 2016).

The total SOM and SOC content represent the most widely recommended indicators for soil quality and related changes (Gregorich et al., 1994; St. Luce et al., 2014). However, the SOM content is unsuited to assess soil quality changes in the short-term (Muscolo et al., 2015), and especially the total SOC content could be shown to be insensitive to even significant changes of interlinked OC parameters (e.g. MBC or DOC) in the short-term (Chapter 4). However, DOC was identified to be a decisive component for understanding the net carbon balance of the ecosystem (Kindler et al., 2011). Loveland and Webb (2003) tried to find a critical level of SOM for temperate arable soils in terms of a threshold of SOC in percentage, but had to conclude that the influence of SOM on soil properties is driven by

quality and origin of SOM fractions, and not quantity. Hence practical assessment of soil quality via SOM remains a challenging task since it requires the integrated consideration of quantitative and qualitative soil parameters (Chapter 4 and 5) involved in soil functioning and their variation in space and time (Doran and Parkin, 1994; Doelman and Eijsackers, 2004). Soil functioning and spatio-temporal SOM dynamics are identified to be mainly driven by soil microorganisms (Schimel and Schaeffer, 2012; Gleixner, 2013; Rousk and Bengtson, 2014), with little or mostly indirect effects of the initial litter input chemistry and quality (Wickings et al., 2012). The formation and persistence of SOM are predominantly controlled rather by environmental and microbial properties than by inherent molecular properties (Schmidt et al., 2011; Schimel and Schaeffer, 2012; Rousk and Bengtson, 2014; Kuzyakov and Blagodatskaya, 2015). Consensus is emerging that microbial materials are key constituents of stable SOM, and new conceptual and quantitative SOM models are rapidly incorporating this view (Kallenbach et al., 2016). Accordingly, microbial necromass is identified to be a significant possible further source for stable SOM (Chapter 5). And especially microbial derived carbohydrates seem to significantly contribute to SOM stabilization (Chapter 5) which has recently been attributed to their quantity and quality (Gunina and Kuzyakov, 2015; Pronk et al., 2015). In the traditional view, both microbial necromass and carbohydrates are assumed to be decomposed quickly.

These new findings reflect the recent debate about the concept of SOM ‘stability’ which currently is under revision by the soil science community since the biogeochemical paradigm of ‘intrinsic recalcitrance’ in decomposition theory had been questioned (Kleber, 2010b; Lal et al., 2015; Lehmann and Kleber, 2015). SOM dynamics research is facing a time of radical change (Kleber and Johnson, 2010). Especially the current research over the last two decades led to increased scepticism toward the humification concept and long standing perceptions of SOM in decomposition theory. New information, which became available through advances in technology (e.g. in the ability to measure the isotope signatures of specific compounds) favors a new picture of organic matter dynamics (Gleixner, 2013). Instead of persistence of plant-derived residues like lignin, the majority of OM molecules in soil is identified to be derived from microbial synthesis or recycling (Gleixner, 2013; Basler et al., 2015; Apostel et al., 2017). Molecules can be newly synthesized by soil organisms and e.g. still being recognized as stable plant-derived molecules (Miltner et al., 2009; Gleixner, 2013). Additionally, carbon, recycled multiple times by generations of microbiota, can be old, decoupling the radiocarbon age of C atoms from the chemical or biological lability of the molecules they comprise (Gleixner, 2013). This induces the current change in view that “oldest” SOM molecules do not imply “most stabilized” but “most recycled” (Dippold and Kuzyakov, 2013).

Against this background, the fractionation of soils has be considered critically. None of the existing fractionation methods, including the SPT density fractionation, refers to a specific SOM stabilization mechanism and hence the obtained fractions cannot represent distinctive functional SOM pools (von Lützow et al., 2007). Therefore, the fractions have been characterized by their mean residence times (MRT). According to the MRT concept, SOM stability is expected to increase from light fractions (free and occluded particulate organic matter: fPOM, oPOM) to heavier fractions (mineral associated organic matter: MOM), despite differences in fractionation methods and turnover measurements (von Lützow et al., 2007). However, experiments with ^{13}C and ^{15}N labelled SOM indicated that some components of the SOM fractions were transformed more rapidly than expected from the MRT of the whole fraction (Poirier et al., 2005; Bird et al., 2008). Thus, each SOM fraction contains at least two different functional pools and two different MRTs. This is confirmed by analysis of the origin of SOM compounds. Although the hexose to pentose ratio (H/P) is no exact measure, it shows a substantial microbial influence with a roughly estimated, overall contribution of microbial versus plant-derived SOM of about 1:1 for every density fraction (Chapter 5). Together with the new insights in microbial molecule recycling (Gleixner, 2013), this questions the usefulness of the density fractionation of soils and of the MRT for describing SOM stability in general. However, direct evidence demonstrating that microbial residues account for the chemistry, stability and abundance of SOM is still lacking (Kallenbach et al., 2016). Quantification of microbial derived SOM (Chapter 5) and elucidation of its explicit origin and dynamics (Chapter 4) is necessary.

In general, the study of SOM in situ should be promoted (Schaeffer et al., 2015). There is a general trend of focusing thereby on biological soil indicators because they are more sensitive than chemical and physical indicators and can describe soil quality in a broader picture (Bastida et al., 2008). This sensitivity however requires non-invasive techniques that are non-destructive to obtain reliable results. The application of Vis-NIR spectroscopy as a non-invasive technique has been suggested to be useful for the estimation of a large number of soil properties, including biological parameter such as microbial biomass, respiration, enzymatic activities and microbial groups (Soriano-Disla et al., 2013; Ludwig et al., 2017). Nevertheless, there is controversy about the usefulness of Vis-NIR for estimating the microbial community composition of soils (Zornoza et al., 2008; Soriano-Disla et al., 2013). Studies working with phospholipid fatty acid (PLFA) analysis to characterize the microbial community concluded that due to the low concentrations of microbial markers in the soil it is unlikely to obtain direct measurable signals by soil reflectance (Cohen et al., 2005; Rinnan and Rinnan, 2007), especially for in situ data predictions are not successful (Chapter 4). For soil spectra taken from pre-processed samples in the laboratory, contents of total PLFA and

the microbial group of fungi could occasionally be predicted (Rinnan and Rinnan, 2007; Zornoza et al., 2008). MBC, also representing a measure for microbial biomass, could be predicted successfully from lab spectra (Terhoeven-Urselmans et al., 2008; Heinze et al., 2013; Soriano-Disla et al., 2013) as well as from in situ spectra (Chapter 4). However, spectral contribution from the soil microbial biomass per se is not expected to result in observable patterns in soil spectra (Cohen et al., 2005; Rinnan and Rinnan, 2007; Rasche et al., 2013). Hence, predictions are assumed to be a result of strong correlations between biological properties and the quantity and quality of SOM (Chodak et al., 2007). These properties are assumed to absorb foremost in the infrared range of the Vis-NIR spectra (Linsler et al., 2017). For in situ spectra of soil surfaces the visible range can however also contribute to microbial biomass predictions (Chapter 4). Biological crusts at the soil surface containing soil algae exhibit high contents of chlorophyll α. Spectra of biological soil crusts therefore show a shallow chlorophyll absorption with a typical peak at 680 nm (Rodriguez-Caballero et al., 2017) in the visible region. The correlation of chlorophyll contents and MBC for temperate arable soils is periodically strong and assumed to contribute to successful predictions of microbial biomass from in situ soil spectra (Chapter 4). A complicating factor about soil reflectance is its general responding to temporally invariant factors such as soil type, mineralogy and geology as well as to temporally variant factors, such as tillage, moisture, soil roughness, SOM dynamics, crop residue cover and soil surface crusts (Ben Dor et al., 1999; Ladoni et al., 2009).

The high spatio-temporal variability of several SOM parameter is affecting both, the optimal sampling density and sampling frequency (Croft et al., 2012). Some authors state that the within-field spatial variability of SOM must first be understood before temporal changes can be assessed (Croft et al., 2012). Thus it remains unclear for most soil parameters if the spectroscopic measurement that is suitable to infer the status of the indicator might also be suitable to infer changes in that property (Archer et al., 2014). This means, the potential to monitor both state and change is lacking accuracy regarding most soil parameters.

This is due to the high spatial variability of SOM parameters that can mask accumulation or depletion processes (Croft et al., 2012). Additionally, the highly dynamic soil surface OM (e.g. of biological soil crusts as microbial hotspots) can mask less dynamic parameters such as total SOC content (Chapter 4) and questions the representativity of spectral soil surface measurements regarding SOM parameter predictions for whole soil horizons or even soil profiles (Chapter 4).

On the way to use Vis-NIR spectroscopy for soil investigation, PLSR has been shown to be most robust among the common regression methods (de Paul Obade and Lal, 2016).

However, applied on its own PLSR is mostly not the optimal solution for processing soil spectra (Cécillon et al., 2009). Improvement of prediction accuracy can be obtained by applying a wide range of pre-processing techniques (Askari et al., 2015). One approach is to do variable selection as a pre-processing of spectral data, which was successfully tested for standard soil parameter and spectra from the laboratory (Chapter 3) and could as well slightly improve predictions from in situ soil spectra (Chapter 4). Another approach is using independent sample sets from spectral libraries to calibrate the model (Cécillon et al., 2009). Also for the internal and external validation it is recommended to use independent sample sets (Ludwig et al., 2017). These modified regression techniques require a larger number of samples, but clearly outperform the classical PLSR (Chapter 3), especially for large soil spectral datasets with wide ranges of values (Cécillon et al., 2009). This highlights the urgent need of establishing a universal and standardized global spectral library to characterize the world's soils (Cécillon et al., 2009; Viscarra Rossel et al., 2016). Especially in situ soil spectral data needs further improvement to increase the robustness and accuracy of the prediction models, e.g. regarding the microbial community composition, which is yet not predictable from in situ spectral data (Chapter 4). However, the pre-processing of spectral data indicates additional improvement potential even for challenging soil data sets (Chapter 4 and Chapter 5), once standardized sample libraries are available (Chapter 3).

To conclude, two microbial SOM pools have been identified to be possible further sources for SOM of temperate arable soils. Microbial necromass (Chapter 5) and the microbial group of soil algae (Chapter 4) are assumed to significantly contribute to topsoil SOM quantity, quality and dynamics. Together with the new insights about microbial SOM synthesis and recycling, this indicates a probably higher microbial contribution to SOM formation from more diverse sources than yet discovered.

However, precise quantification of these pools is still lacking and further basic research regarding microbial SOM pools and their spatio-temporal dynamics is necessary to identify drivers of SOM formation and persistence. The newly emerging concept of SOM stability has to be integrated in the interdisciplinary approach of finding efficient bioindicators for soil quality that can be included in large-scale monitoring by earth observation programs.

7 Summary and Outlook

Summary. In the first phase of the presented interdisciplinary research, the establishment of a comprehensive local soil library of soil spectra and associated parameters was conducted. Covering a diverse soilscape it provides a multifacetted and broad range of site-specific values. On this basis, optimization strategies of data processing were investigated. Prediction accuracy of common spectrally monitored SOM parameters (SOC, N) from soil reflectance spectra of hyperspectral laboratory measurements could be improved by pre-processing of spectral data using variable selection and wavelet transformation (H1).

The potential of soil spectroscopy regarding estimations of an extended set of SOM parameter (focus on biological SOM parameters; in situ data) was examined. The spectral potential of SOM assessment of other than the common (SOC; N) parameters, with emphasis on labile and biological or microbial SOM parameters (microbial biomass carbon, MBC; dissolved organic carbon, DOC; chlorophyll α, Chl α; phospholipid fatty acids, PLFAs) was controversy (H2). A mediocre prediction accuracy with the possibility to distinguish between high and low values was obtained for MBC and DOC, whereas predictions were not successful for OC, $PLFA_{tot}$, and PLFA markers for the microbial groups of algae, bacteria and fungi. Best prediction accuracy was achieved for surface changes induced by green soil algae, which is also assumed to trigger the modelling of MBC. Instead, the total SOM content, measured as total OC, could not be predicted successfully, which was largely attributed to the microbial interference. Hence, the highly dynamic microbial SOM pools at the soil surface are not representative and assumed to possibly impede total bulk SOM content estimations (H4). The transferability of the identified optimization strategies to spectral data of this extended set of soil parameters was successful (H3). Prediction accuracy could slightly be improved by pre-processing of the in situ spectral data.

The parallel focus of studying SOM origin (plant vs. microbial input) and its impact on SOM quantity and quality identified microbial necromass (H5) and soil surface algae (H4) to be possible significant further sources for SOM formation and persistence. Overall, the microbial contribution to temperate arable topsoil OM is higher than expected (H5). Effects of the microbial contribution could be shown for SOM quantity especially in mineral associated OM and for SOM quality regarding the correlation of microbial derived carbohydrates with SOM stability (H5).

Outlook. In Chapter 3 and 5, two possible further sources for microbial SOM pools have been identified. Further investigation of these possible SOM sources is necessary to validate

and determine their quantity and their quality (significance) in relation to the formation and persistence of arable SOM. In general, the in situ investigation of SOM in structured field soils should be promoted. To secure SOM integrity and produce reliable results this should be enhanced by non-invasive techniques. To investigate small-scale heterogeneity, techniques to locate active microbial cells should be further developed (Schaeffer et al., 2015). The combined use of STXM and near-edge X-ray adsorption fine structure (NEXAFS), two non-invasive techniques, can be used to show the existence of distinct micro- and nano-C agglomerates, a few micrometers apart from each other associated to mineral soil particles and determine at the same time whether these microbial agglomerates are of fungal or bacterial origin (Solomon et al., 2012a; Solomon et al., 2012b). This is essential, since the small-scale (micro- to nanometer-scale) distribution of SOM is important for our process understanding, to identify drivers of SOM stability and sustainability and therefore, efficient bioindicators of soil quality. Once this understanding has emerged, we need to translate it into large-scale assessment methods to be able to monitor it globally. Recent large-scale biodiversity assessment already uses surrogate measures like reflectance spectroscopy instead of direct measures of bioindicators (Feld et al., 2009; Lausch et al., 2016). High-resolution satellite imagery already enables for estimating animal population sizes, taxonomic and functional diversity on the macro-biological level (Lausch et al., 2016). Hence, on the macrobiological level the approach of *zooming in* to form process understanding and then *zooming out* and assess efficient bioindicators was already successful. To reach the microbiological level will be the next step. This is on the one side benefitted by new recent developments in the field of 'omics' research which will probably soon provide the necessary practical tools for this approach. On the other side, the launching of the German ENMAP satellite, which is scheduled for 2018, will for the first time offer large-scale coverage of hyperspectral data. This technical development provides many additional spectral channels that expand the spectral information of the sensed material to be analysed in terms of quantity and quality (Ben-Dor, 2011). Enlarging the spectral windows towards the ultraviolet (N300 nm) and TIR spectral regions (2500– 14000 nm) might additionally help to get more information for soil applications in the future (Ben-Dor et al., 2009). To improve the predictions of soil bioindicators from spectroscopy, more spectral soil data has to be collected in local spectral libraries. Data collection has to be standardized so that local libraries can contribute to a global soil spectral library (Viscarra Rossel et al., 2016) that can be used in earth observation to help monitoring planetary boundaries that ensure a safe operating space for humanity (Rockström et al., 2009).

Whatever the technique promoted on each side, the most important point is the collaboration between the disciplines. The nexus of food, energy and water illustrates the need for holism

and cross-disciplinary collaboration for the future of ecology research (McCormick and Kapustka, 2016; Weathers et al., 2016). To communicate between the disciplines is essential not only to share technical advances, but to benefit from each other's perspective. Ecosystem scientists should be encouraged to engage in increasingly important and new interfaces between disciplines and between science and society (Weathers et al., 2016). This interdisciplinary research project was aimed to contribute to this task in an interdisciplinary approach. It is envisioned that this interdisciplinarity is the right principle to go forward: advance science by joining the larger scientific community to solve major environmental and societal problems.

7.1 References (Chapters 6 & 7)

Anderson, T.-H., Domsch, K.H., 2010. Soil microbial biomass: The eco-physiological approach. Soil Biology and Biochemistry 42, 2039-2043.

Apostel, C., Dippold, M.A., Bore, E., Kuzyakov, Y., 2017. Sorption of Alanine changes microbial metabolism in addition to availability. Geoderma 292, 128-134.

Archer, N., Rawlins, B., Grebby, S., Marchant, B., Emmett, B., 2014. Identify the opportunities provided by developments in earth observation and remote sensing for national scale monitoring of soil quality. In: Survey, B.G. (Ed.), British Geological Survey Internal Report, p. 32.

Armenta, S., de la Guardia, M., 2014. Vibrational spectroscopy in soil and sediment analysis. Trends in Environmental Analytical Chemistry 2, 43-52.

Askari, M.S., Cui, J., O'Rourke, S.M., Holden, N.M., 2015. Evaluation of soil structural quality using VIS–NIR spectra. Soil and Tillage Research 146, 108-117.

Basler, A., Dippold, M., Helfrich, M., Dyckmans, J., 2015. Microbial carbon recycling: an underestimated process controlling soil carbon dynamics – Part 2: A C₃-C₄ vegetation change field labelling experiment. Biogeosciences 12, 6291-6299.

Bastida, F., Zsolnay, A., Hernández, T., García, C., 2008. Past, present and future of soil quality indices: A biological perspective. Geoderma 147, 159-171.

Bellon-Maurel, V., 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.

Ben-Dor, E., 2011. Quantitative Remote Sensing of Soil Properties. In: Sparks, D.L. (Ed.), Advances in Agronomy.

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

Ben-Dor, E., Irons, J.R., Epema, J.F., 1999. Soil reflectance. In: Ryerson, R.A., Rencz, A.N. (Eds.), Manual of Remote Sensing, Remote Sensing for the Earth Sciences. Wiley, New York, pp. 111-188.

Bird, J.A., Kleber, M., Torn, M.S., 2008. 13C and 15N stabilization dynamics in soil organic matter fractions during needle and fine root decomposition. Organic Geochemistry 39, 465-477.

Blagodatskaya, E., Kuzyakov, Y., 2013. Active microorganisms in soil: Critical review of estimation criteria and approaches. Soil Biology and Biochemistry 67, 192-211.

Blum, W.E.H., 2005. Functions of Soil for Society and the Environment. Reviews in Environmental Science and Bio/Technology 4, 75-79.

Blume, H.P., Brümmer, G.W., Scheffer, F., Horn, R., Kandeler, E., Schachtschabel, P., Kögel-Knabner, I., Welp, G., Kretzschmar, R., Thiele-Bruhn, S., 2010. Scheffer/Schachtschabel: Lehrbuch der Bodenkunde. Spektrum Akademischer Verlag.

Bradford, M.A., Keiser, A.D., Davies, C.A., Mersmann, C.A., Strickland, M.S., 2013. Empirical evidence that soil carbon formation from plant inputs is positively related to microbial growth. Biogeochemistry 113, 271-281.

Bradford, M.A., Wood, S.A., Bardgett, R.D., Black, H.I.J., Bonkowski, M., Eggers, T., Grayston, S.J., Kandeler, E., Manning, P., Setälä, H., Jones, T.H., 2014. Discontinuity in the responses of ecosystem processes and multifunctionality to altered soil community composition. Proceedings of the National Academy of Sciences 111, 14478-14483.

Campbell, E.E., Paustian, K., 2015. Current developments in soil organic matter modeling and the expansion of model applications: a review. Environmental Research Letters 10, 123004.

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

Chodak, M., Niklińska, M., Beese, F., 2007. The Use of near Infrared Spectroscopy to Quantify Lignite-Derived Carbon in Humus-Lignite Mixtures. Journal of Near Infrared Spectroscopy 15, 195-200.

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

COM, C.o.t.E.C., 2000. Indicators for the integration of environmental concerns into the common agricultural policy. Communication from the Commission to the Council and the European Parliament. COM 20 final.

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

de Paul Obade, V., Lal, R., 2016. A standardized soil quality index for diverse field conditions. Sci Total Environ 541, 424-434.

Dippold, M.A., Kuzyakov, Y., 2013. Biogeochemical transformations of amino acids in soil assessed by position-specific labelling. Plant and Soil 373, 385-401.

Doelman, P., Eijsackers, H.J.P., 2004. Vital Soil: Function, Value and Properties. Elsevier Science.

Doran, J.W., Parkin, T.B., 1994. Defining and Assessing Soil Quality. In: Doran, J.W., Coleman, D.C., Bezdicek, D.F., Stewart, B.A. (Eds.), Defining Soil Quality for a Sustainable Environment. Soil Science Society of America and American Society of Agronomy, Madison, WI.

Doran, J.W., Parkin, T.B., 1996. Quantitative Indicators of Soil Quality: A Minimum Data Set. In: Doran, J.W., Jones, A.J. (Eds.), Methods for Assessing Soil Quality. Soil Science Society of America, Madison, WI, pp. 25-37.

Erhagen, B., Oquist, M., Sparrman, T., Haei, M., Ilstedt, U., Hedenstrom, M., Schleucher, J., Nilsson, M.B., 2013. Temperature response of litter and soil organic matter decomposition is determined by chemical composition of organic material. Glob Chang Biol 19, 3858-3871.

FAO, ITPS, 2015. Status of the World's Soil Resources (SWSR) – Main Report Rome, Italy.

Feld, C.K., Martins da Silva, P., Paulo Sousa, J., de Bello, F., Bugter, R., Grandin, U., Hering, D., Lavorel, S., Mountford, O., Pardo, I., PÃ¤rtel, M., RÃ¶mbke, J.r., Sandin, L., Bruce Jones, K., Harrison, P., 2009. Indicators of biodiversity and ecosystem services: a synthesis across ecosystems and spatial scales. Oikos 118, 1862-1871.

Gleixner, G., 2013. Soil organic matter dynamics: a biological perspective derived from the use of compound-specific isotopes studies. Ecological Research 28, 683-695.

Gregorich, E.G., Monreal, C.M., Carter, M.R., Angers, D.A., Ellert, B.H., 1994. Towards a minimum data set to assess soil organic matter quality in agricultural soils. Canadian Journal of Soil Science 74, 367-385.

Griffiths, B.S., Römbke, J., Schmelz, R.M., Scheffczyk, A., Faber, J.H., Bloem, J., Pérès, G., Cluzeau, D., Chabbi, A., Suhadolc, M., Sousa, J.P., Martins da Silva, P., Carvalho, F., Mendes, S., Morais, P., Francisco, R., Pereira, C., Bonkowski, M., Geisen, S., Bardgett, R.D., de Vries, F.T., Bolger, T., Dirilgen, T., Schmidt, O., Winding, A., Hendriksen, N.B., Johansen, A., Philippot, L., Plassart, P., Bru, D., Thomson, B., Griffiths, R.I., Bailey, M.J., Keith, A., Rutgers, M., Mulder, C., Hannula, S.E., Creamer, R., Stone, D., 2016. Selecting cost effective and policy-relevant biological indicators for European monitoring of soil biodiversity and ecosystem function. Ecological Indicators 69, 213-223.

Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. Soil Science Society of America Journal 75, 1201.

Gunina, A., Kuzyakov, Y., 2015. Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. Soil Biology and Biochemistry 90, 87-100.

Havlicek, E., 2012. Soil biodiversity and bioindication: From complex thinking to simple acting. European Journal of Soil Biology 49, 80-84.

Haynes, R.J., 2005. Labile Organic Matter Fractions as Central Components of the Quality of Agricultural Soils: An Overview. Advances in Agronomy. Academic Press, pp. 221-268.

Heinze, S., Vohland, M., Joergensen, Rainer G., Ludwig, B., 2013. Usefulness of near-infrared spectroscopy for the prediction of chemical and biological soil properties in different long-term experiments. Journal of Plant Nutrition and Soil Science 176, 520-528.

Hiederer, R., 2009. Distribution of Organic Carbon in Soil Profile Data. In: Communities, O.f.O.P.o.t.E. (Ed.), EUR 23980 EN. European Commission Joint Research Centre Institute for Environment and Sustainability, Luxembourg, p. 126.

IRENA, 2015. Renewable Energy in the Water, Energy & Food Nexus.

Jansson, J.K., Baker, E.S., 2016. A multi-omic future for microbiome studies. Nat Microbiol 1, 16049.

Johnson, M.J., Lee, K.Y., Scow, K.M., 2003. DNA fingerprinting reveals links among agricultural crops, soil properties, and the composition of soil microbial communities. Geoderma 114, 279-303.

Kallenbach, C.M., Frey, S.D., Grandy, A.S., 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nat Commun 7, 13630.

Karlen, D.L., Mausbach, M.J., Doran, J.W., Cline, R.G., Harris, R.F., Schuman, G.E., 1997. Soil Quality: A Concept, Definition, and Framework for Evaluation (A Guest Editorial). Soil Science Society of America Journal 61, 4-10.

Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., Pachepsky, Y., van der Putten, W.H., Bardgett, R.D., Moolenaar, S., Mol, G., Jansen, B., Fresco, L.O., 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. Soil 2, 111-128.

Kindler, R., Miltner, A., Thullner, M., Richnow, H.-H., Kästner, M., 2009. Fate of bacterial biomass derived fatty acids in soil and their contribution to soil organic matter. Organic Geochemistry 40, 29-37.

Kindler, R., Siemens, J., Kaiser, K., Walmsley, D.C., Bernhofer, C., Buchmann, N., Cellier, P., Eugster, W., Gleixner, G., Grunwald, T., Heim, A., Ibrom, A., Jones, S.K., Jones, M., Klumpp, K., Kutsch, W., Larsen, K.S., Lehuger, S., Loubet, B., McKenzie, R., Moors, E., Osborne, B., Pilegaard, K., Rebmann, C., Saunders, M., Schmidt, M.W.I., Schrumpf, M., Seyfferth, J., Skiba, U., Soussana, J., Sutton, M.A., Tefs, C., Vowinckel, B., Zeeman, M.J., Kaupenjohann, M., 2011. Dissolved carbon leaching from soil is a crucial component of the net ecosystem carbon balance. Global Change Biology 17(2), 1167-1185

Kleber, M., 2010a. Response to the Opinion paper by Margit von Lützow and Ingrid Kögel-Knabner on 'What is recalcitrant soil organic matter?' by Markus Kleber. Environmental Chemistry 7, 336-337.

Kleber, M., 2010b. What is recalcitrant soil organic matter? Environmental Chemistry 7, 320.

Kleber, M., Johnson, M.G., 2010. Advances in Understanding the Molecular Structure of Soil Organic Matter. 106, 77-142.

Knicker, H., Totsche, K.U., Almendros, G., González-Vila, F.J., 2005. Condensation degree of burnt peat and plant residues and the reliability of solid-state VACP MAS 13C NMR spectra obtained from pyrogenic humic material. Organic Geochemistry 36, 1359-1377.

Kogel-Knabner, I., 2002. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. Soil Biology and Biochemistry 34, 139-162.

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept & review. Soil Biology and Biochemistry 83, 184-199.

Ladoni, M., Bahrami, H.A., Alavipanah, S.K., Norouzi, A.A., 2009. Estimating soil organic carbon from soil reflectance: a review. Precision Agriculture 11, 82-99.

Laishram, J., Saxena, K.G., Maikhuri, R.K., Rao, K.S., 2012. Soil quality and soil health: a review. International Journal of Ecology and Environmental Sciences 38, 19-37.

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

Lal, R., 2010. Managing Soils and Ecosystems for Mitigating Anthropogenic Carbon Emissions and Advancing Global Food Security. BioScience 60, 708-721.

Lal, R., Negassa, W., Lorenz, K., 2015. Carbon sequestration in soil. Current Opinion in Environmental Sustainability 15, 79-86.

Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J.M., Heurich, M., Jung, A., Klenke, R., Neumann, C., Pause, M., Rocchini, D., Schaepman, M.E., Schmidtlein, S., Schulz, K., Selsam, P., Settele, J., Skidmore, A.K., Cord, A.F., 2016. Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. Ecological Indicators 70, 317-339.

Lehmann, J., Kleber, M., 2015. The contentious nature of soil organic matter. Nature 528, 60-68.

Leifeld, J., Kögel-Knabner, I., 2005. Soil organic matter fractions as early indicators for carbon stock changes under different land-use? Geoderma 124, 143-155.

Leifeld, J., Steffens, M., Galego-Sala, A., 2012. Sensitivity of peatland carbon loss to organic matter quality. Geophysical Research Letters 39, n/a-n/a.

Lenton, T.M., Held, H., Kriegler, E., Hall, J.W., Lucht, W., Rahmstorf, S., Schellnhuber, H.J., 2008. Tipping elements in the Earth's climate system. Proceedings of the National Academy of Sciences 105, 1786-1793.

Liang, C., Balser, T.C., 2011. Microbial production of recalcitrant organic matter in global soils: implications for productivity and climate policy. Nat Rev Microbiol 9, 75; author reply 75.

Liang, C., Cheng, G., Wixon, D.L., Balser, T.C., 2010. An Absorbing Markov Chain approach to understanding the microbial role in soil carbon stabilization. Biogeochemistry 106, 303-309.

Linsler, D., Sawallisch, A., Höper, H., Schmidt, H., Vohland, M., Ludwig, B., 2017. Near-infrared spectroscopy for determination of soil organic C, microbial biomass C and C and N fractions in a

heterogeneous sample of German arable surface soils. Archives of Agronomy and Soil Science, 1-11.
Loveland, P., Webb, J., 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. Soil and Tillage Research 70, 1-18.
Ludwig, B., Vormstein, S., Niebuhr, J., Heinze, S., Marschner, B., Vohland, M., 2017. Estimation accuracies of near infrared spectroscopy for general soil properties and enzyme activities for two forest sites along three transects. Geoderma 288, 37-46.
McCormick, R., Kapustka, L.A., 2016. The Answer is 42...What is THE question? Journal of Environmental Studies and Sciences 6, 208–213.
MEA, 2005. Millennium Ecosystem Assessment: Ecosystems and Human Well-being: Synthesis. In: Institute, W.R. (Ed.), Washington, DC.
Miltner, A., Bombach, P., Schmidt-Brücken, B., Kästner, M., 2012. SOM genesis: microbial biomass as a significant source. Biogeochemistry 111, 41-55.
Miltner, A., Kindler, R., Knicker, H., Richnow, H.-H., Kästner, M., 2009. Fate of microbial biomass-derived amino acids in soil and their contribution to soil organic matter. Organic Geochemistry 40, 978-985.
Morvan, X., Saby, N.P., Arrouays, D., Le Bas, C., Jones, R.J., Verheijen, F.G., Bellamy, P.H., Stephens, M., Kibblewhite, M.G., 2008. Soil monitoring in Europe: a review of existing systems and requirements for harmonisation. Sci Total Environ 391, 1-12.
Mueller, C.W., Kölbl, A., Hoeschen, C., Hillion, F., Heister, K., Herrmann, A.M., Kögel-Knabner, I., 2012. Submicron scale imaging of soil organic matter dynamics using NanoSIMS – From single particles to intact aggregates. Organic Geochemistry 42, 1476-1488.
Mueller, C.W., Weber, P.K., Kilburn, M.R., Hoeschen, C., Kleber, M., Pett-Ridge, J., 2013. Advances in the Analysis of Biogeochemical Interfaces. 121, 1-46.
Müller, A., Weigelt, J., 2013. Governance for a Land Degradation Neutral World. IISD, IISD Land Policy and Practice Knowledge Database.
Murphy, D.V., Cookson, W.R., Braimbridge, M., Marschner, P., Jones, D.L., Stockdale, E.A., Abbott, L.K., 2011. Relationships between soil organic matter and the soil microbial biomass (size, functional diversity, and community structure) in crop and pasture systems in a semi-arid environment. Soil Research 49, 582.
Muscolo, A., Settineri, G., Attinà, E., 2015. Early warning indicators of changes in soil ecosystem functioning. Ecological Indicators 48, 542-549.
Nielsen, U.N., Ayres, E., Wall, D.H., Bardgett, R.D., 2011. Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity-function relationships. European Journal of Soil Science 62, 105-116.
Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. 132, 139-159.
Orgiazzi, A., Panagos, P., Yigini, Y., Dunbar, M.B., Gardi, C., Montanarella, L., Ballabio, C., 2016. A knowledge-based approach to estimating the magnitude and spatial patterns of potential threats to soil biodiversity. Sci Total Environ 545-546, 11-20.
Parr, J.F., Papendick, R.I., Hornick, S.B., Meyer, R.E., 1992. Soil quality: Attributes and relationship to alternative and sustainable agriculture. American Journal of Alternative Agriculture 7, 5-11.
Paz-Ferreiro, J., Fu, S., 2016. Biological Indices for Soil Quality Evaluation: Perspectives and Limitations. Land Degradation & Development 27, 14-25.
Poirier, N., Sohi, S.P., Gaunt, J.L., Mahieu, N., Randall, E.W., Powlson, D.S., Evershed, R.P., 2005. The chemical composition of measurable soil organic matter pools. Organic Geochemistry 36, 1174-1189.
Pronk, G.J., Heister, K., Kögel-Knabner, I., 2015. Amino sugars reflect microbial residues as affected by clay mineral composition of artificial soils. Organic Geochemistry 83-84, 109-113.
Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Pérès, G., Rutgers, M., 2012. Soil biodiversity, biological indicators and soil ecosystem services-an overview of European approaches. Current Opinion in Environmental Sustainability 4, 529-538.
Pulleman, M.M., Bouma, J., van Essen, E.A., Meijles, E.W., 2000. Soil Organic Matter Content as a Function of Different Land Use History. Soil Science Society of America Journal 64, 689.
Rasche, F., Marhan, S., Berner, D., Keil, D., Kandeler, E., 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.

Remusat, L., Hatton, P.J., Nico, P.S., Zeller, B., Kleber, M., Derrien, D., 2012. NanoSIMS study of organic matter associated with soil aggregates: advantages, limitations, and combination with STXM. Environ Sci Technol 46, 3943-3949.

Rinnan, R., Rinnan, Å., 2007. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. Soil Biology and Biochemistry 39, 1664-1673.

Ritz, K., Black, H.I.J., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: A framework for balancing scientific and technical opinion to assist policy development. Ecological Indicators 9, 1212-1221.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., Wit, C.A.d., Hughes, T., Leeuw, S.v.d., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. Nature 461, 472-475.

Rodionov, A., Pätzold, S., Welp, G., Pude, R., Amelung, W., 2016. Proximal field Vis-NIR spectroscopy of soil organic carbon: A solution to clear obstacles related to vegetation and straw cover. Soil and Tillage Research 163, 89-98.

Rodriguez-Caballero, E., Paul, M., Tamm, A., Caesar, J., Budel, B., Escribano, P., Hill, J., Weber, B., 2017. Biomass assessment of microbial surface communities by means of hyperspectral remote sensing data. Sci Total Environ.

Roume, H., Heintz-Buschart, A., Muller, E.E.L., May, P., Satagopam, V.P., Laczny, C.C., Narayanasamy, S., Lebrun, L.A., Hoopmann, M.R., Schupp, J.M., Gillece, J.D., Hicks, N.D., Engelthaler, D.M., Sauter, T., Keim, P.S., Moritz, R.L., Wilmes, P., 2015. Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. npj Biofilms and Microbiomes 1, 15007.

Rousk, J., Bengtson, P., 2014. Microbial regulation of global biogeochemical cycles. Front Microbiol 5, 103.

Schaeffer, A., Nannipieri, P., Kästner, M., Schmidt, B., Botterweck, J., 2015. From humic substances to soil organic matter–microbial contributions. In honour of Konrad Haider and James P. Martin for their outstanding research contribution to soil science. Journal of Soils and Sediments 15, 1865-1881.

Schaumann, G.E., 2006. Soil organic matter beyond molecular structure Part I: Macromolecular and supramolecular characteristics. Journal of Plant Nutrition and Soil Science 169, 145-156.

Schimel, J.P., Schaeffer, S.M., 2012. Microbial control over carbon cycling in soil. Front Microbiol 3, 348.

Schmidt, M.W., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kogel-Knabner, I., Lehmann, J., Manning, D.A., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49-56.

Schurig, C., Smittenberg, R.H., Berger, J., Kraft, F., Woche, S.K., Goebel, M.-O., Heipieper, H.J., Miltner, A., Kaestner, M., 2013. Microbial cell-envelope fragments and the formation of soil organic matter: a case study from a glacier forefield. Biogeochemistry 113, 595-612.

Simpson, A.J., Simpson, M.J., Smith, E., Kelleher, B.P., 2007. Microbially Derived Inputs to Soil Organic Matter: Are Current Estimates Too Low? Environmental Science & Technology 41, 8070-8076.

Sojka, R.E., Upchurch, D.R., 1999. Reservations regarding the soil quality concept. Soil Science Society of America Journal 63, 1039-1054.

Solomon, D., Lehmann, J., Harden, J., Wang, J., Kinyangi, J., Heymann, K., Karunakaran, C., Lu, Y., Wirick, S., Jacobsen, C., 2012a. Micro- and nano-environments of carbon sequestration: Multi-element STXM–NEXAFS spectromicroscopy assessment of microbial carbon and mineral associations. Chemical Geology 329, 53-73.

Solomon, D., Lehmann, J., Wang, J., Kinyangi, J., Heymann, K., Lu, Y., Wirick, S., Jacobsen, C., 2012b. Micro- and nano-environments of C sequestration in soil: a multi-elemental STXM-NEXAFS assessment of black C and organomineral associations. Sci Total Environ 438, 372-388.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M.J., 2013. 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.

St. Luce, M., Ziadi, N., Zebarth, B.J., Grant, C.A., Tremblay, G.F., Gregorich, E.G., 2014. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. Geoderma 232-234, 449-458.

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and Near Infrared Spectroscopy in Soil Science. 107, 163-215.

Stevens, A., Nocita, M., Toth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. PLoS One 8, e66409.

Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S.Y., Rawlins, B.G., Field, D.J., 2015. Global soil organic carbon assessment. Global Food Security 6, 9-16.

Stone, D., Ritz, K., Griffiths, B.G., Orgiazzi, A., Creamer, R.E., 2016. Selection of biological indicators appropriate for European soil monitoring. Applied Soil Ecology 97, 12-22.

Terhoeven-Urselmans, T., Schmidt, H., Georg Joergensen, R., Ludwig, B., 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pre-treatment. Soil Biology and Biochemistry 40, 1178-1188.

Tóth, G., Stolbovoy, V., Montanarella, L., 2007. Soil quality and sustainability evaluation – an integrated approach to support soil-related policies of the European Union. In: Communities, O.f.O.P.o.t.E. (Ed.), EUR 22721 EN, Luxembourg, p. 40.

UN, 2015. World population prospects: the 2015 revision. . In: Nations, U. (Ed.). Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf.

Valin, H., Sands, R.D., van der Mensbrugghe, D., Nelson, G.C., Ahammad, H., Blanc, E., Bodirsky, B., Fujimori, S., Hasegawa, T., Havlik, P., Heyhoe, E., Kyle, P., Mason-D'Croz, D., Paltsev, S., Rolinski, S., Tabeau, A., van Meijl, H., von Lampe, M., Willenbockel, D., 2014. The future of food demand: understanding differences in global economic models. Agricultural Economics 45, 51-67.

Velasquez, E., Lavelle, P., Andrade, M., 2007. GISQ, a multifunctional indicator of soil quality. Soil Biology and Biochemistry 39, 3066-3080.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, N., 2011. Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. In: Sparks, D.L. (Ed.), Advances in Agronomy pp. 237-282.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Science Reviews 155, 198-230.

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

Vohland, M., Ludwig, M., Thiele-Bruhn, S., 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.

von Lützow, M., Kögel-Knabner, I., 2010. Response to the Concept paper: 'What is recalcitrant soil organic matter?' by Markus Kleber. Environmental Chemistry 7, 333.

von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., Marschner, B., 2007. SOM fractionation methods: Relevance to functional pools and to stabilization mechanisms. Soil Biology and Biochemistry 39, 2183-2207.

Wall, D.H., 2007. Global change tipping points: above- and below-ground biotic interactions in a low diversity ecosystem. Philosophical Transactions of the Royal Society B: Biological Sciences 362, 2291-2306.

Weathers, K.C., Groffman, P.M., Van Dolah, E., Bernhardt, E., Grimm, N.B., McMahon, K., Schimel, J., Paolisso, M., Maranger, R., Baer, S., Brauman, K., Hinckley, E., 2016. Frontiers in

Ecosystem Ecology from a Community Perspective: The Future is Boundless and Bright. Ecosystems 19, 753-770.

Weigelt, J., Müller, A., Beckh, C., Töpfer, K., 2014. Soils in the Nexus. A Crucial Resource for Water, Energy and Food Security. Oekom verlag, München

Wickings, K., Grandy, A.S., Reed, S.C., Cleveland, C.C., 2012. The origin of litter chemical complexity during decomposition. Ecol Lett 15, 1180-1188.

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K.M., Arcenegui, V., Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. Soil Biol Biochem 40, 1923-1930.

Zornoza, R., Mataix-Solera, J., Guerrero, C., Arcenegui, V., Garcia-Orenes, F., Mataix-Beneyto, J., Morugan, A., 2007. Evaluation of soil quality using multiple lineal regression based on physical, chemical and biochemical properties. Sci Total Environ 378, 233-237.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst habe und keine anderen als die genannten Quellen und Hilfsmittel genutzt wurden sowie wörtlich und inhaltlich entnommene Stellen gekennzeichnet wurden.

Trier , der 19. Juni 2017

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.
Mein besonderer Dank gilt Prof. Dr. Sören Thiele-Bruhn für die Betreuung dieser meiner Doktorjahre, für die vielen konstruktiven Diskussionen und die Unterstützung meiner selbstständigen Schritte in diesen Projekten.
Prof. Dr. Christoph Emmerling möchte ich danken für den Glauben in mich, für eine immer offene Tür, ein offenes Ohr und ein offenes Herz für mich.
Prof. Dr. Michael Vohland danke ich für die den Austausch von Ideen und die kritische sowie starke Auseinandersetzung über unsren fächerübergreifenden Projektansatz. Ich habe beruflich sehr von dieser Interdisziplinarität profitiert.

Bei allen aktuellen wie ehemaligen Mitarbeitern der Trierer Bodenkunde bedanke ich mich für die Unterstützung und die tolle Zeit. Insbesondere danke ich Dr. Daniel Felten für die Ermunterung zum Promovieren und der vielfältigen Unterstützung, gerade auf der schwierigen ersten Etappe. Dr. Nadine Koch, Dr. Marco Hümann, Dr. Rüdiger Reichel und Dr. Christoph Müller danke ich dafür, dass sie so lustige und freundschaftliche Kollegen waren, das hat Spaß gemacht mit Euch! Großen Dank an Elvira Sieberger und Petra Ziegler für ihre technische und menschliche Unterstützung bei der Erhebung aller laborbezogenen Daten. Petra, ich habe versucht zu zählen wie viele Messwerte ich mit Deiner Hilfe durch die Jahre ermittelt habe, aber ich musste irgendwann kapitulieren! Bei Elke Biewer möchte ich mich ganz herzlich bedanken - Du bist die gute Seele der Abteilung und ich hab mich immer gut aufgehoben gefühlt bei Dir. Vielen Dank an all die vielen ehemaligen Diplomanden, Bachelor- und Masterstudenten sowie wissenschaftlichen Hilfskräfte, vor allem an Sabina Hens, Sascha Rachow, Pascal Mousel, Cathrine Sander und Michael Boetzel, für ihre Unterstützung bei den Gelände- und Laborarbeiten.

Prof. Dr. Leinweber und Kai-Uwe Eckhardt von der Universität Rostock danke ich für die tolle Kooperation und die vielfältige Unterstützung beim Auswerten der zusammen generierten Daten.

Bei Prof. Dr. Matthias Kästner, Dr. Anja Miltner und Dr. Jan Achtenhagen vom UFZ Leipzig möchte ich mich für die spannende Zusammenarbeit und ihre großzügige Hilfe bedanken.

Der Deutschen Forschungsgemeinschaft sowie der Nikolaus Koch Stiftung Trier danke ich für die finanzielle Unterstützung meiner Arbeit.

Danksagung

Ganz besonderer Dank gilt den Menschen, die mir privat den Rücken gestärkt und somit zum Gelingen dieser Arbeit beigetragen haben.

Stefan Danke für dieses tolle gemeinsame Projekt, und damit die Möglichkeit, meine eigenen wissenschaftlichen Gedanken und Konzepte weiter zu verfolgen. Das zusammen Denken, beruflich wie privat, macht Spaß und bringt mich so sehr vorwärts.

Christian Einfach Danke, Dr. proc. C. Bossung. Du weißt schon, wofür…

Rachid Danke dafür, dass Du mein größter Fan warst, das hat viel Anfangsmut gemacht.

Ivan ¡Oye, causa! Gracias por ser el primer científico en mi vida y por mostrarme la belleza de la pachamama en tiempos en los que no tenía nada sólo el poder de mis palabras. Aun cuando nuestros papeles han cambiado, tú escribes poemas y yo artículos, me gustaría pedirte - mantén esta parte de mí muy cerca de ti y déjala respirar el silencio de tu selva, tu montañas y tus costas mientras yo contemplo y protejo nuestras historias para no ahogarme en locura de hoy en día.

Stephan Du bist mir eine Inspiration und Hilfe. Bist mir Hilfe-zur-Selbsthilfe. Danke, dass ich immer auf Dich zählen kann, wenn es drauf ankommt und überhaupt!

Sebastian Danke für alles. Ohne Dich wäre das nie möglich gewesen.

Saskia Danke dafür, dass Du mich so gut kennst. Und dafür, dass Du immer da bist wenn ich Dich brauche, meine Liebe.

Greta Greta, meine Insel!!! Danke, dass ich bei Dir stranden und an Deiner Seite wieder neue Kraft tanken durfte. Du bist mein guter Ort.

Tim Danke für das Neuausrichten der Sterne. Die Koordinaten sind gesetzt. Ich freue mich auf diesen neuen Weg…

Mama Zuletzt und zuallererst: mein allumfassender Dank gilt Dir, Mama. Quelle meiner Zuversicht. Du hältst alles im dynamischen Gleichgewicht. Dies hier war nur eines unserer Abenteuer. Ich freue mich auf alle, die noch folgen. Deine Marie

Curriculum vitae

Marie Ludwig
PhD environmental sciences

Dorfstraße 30
54434 Lückenburg
+49 151 - 27107161
marieisabelludwig@gmail.com

https://www.researchgate.net/profile/Marie_Ludwig

Personal data

Born November 14th 1982
in Eubach (Germany)

Work Experience

Since 02/2016 — **Thünen-Institute of Biodiversity, Braunschweig, Germany**
Researcher

- Sustainability assessment of agricultural land-use by soil bioindicators
- Driving factors for functional biodiversity of soils

08/2010 – 02/2016 — **Trier University, Germany**
Research assistant

- Managing the interdisciplinary DFG-research project „SOMScapes“, to calibrate soil science and remote sensing (spectroscopic) techniques for evaluation of the soil organic matter (SOM)
- Coordination of project partners
- Data analysis
- Presentation of results at national and international conferences
- Introduction and instruction to laboratory work for students and support staff
- Supervision of bachelor-, master- und diploma students at the soil chemistry and soil biology division

Since 12/2011 — **Hydrosol – Hellebrand & Müller GbR, Trier, Germany**
Freelance

- Evaluation of agroecosystem biodiversity

- Consultancy in biodiversity improvement, environmental education and regional governance concepts

09/2009 – 10/2009 **Research Institute of Organic Agriculture (FiBL), Frick, Switzerland**
Trainee

- Collaboration in research of functional biodiversity in agroecosystems and biological control
- Collaboration in long term DOK-trial (Therwil, Basel District) to assess the impact of cropping systems on the microbial community and soil organic matter
- Preparation and participation in the „2nd European Seminar on European Learning Network on Functional AgroBiodiversity" (ELN-FAB)

09/2007 – 10/2007 **Leibniz Institute German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany, Trainee**

- Participation in plant virus experiments, virus transmission & biology
- Isolation of plant nucleic acid

Education

Since 08/2010 **Trier University, Germany**
PhD thesis „Assessing Soil Organic Matter (SOM) on the Landscape Scale"
Planning and implementation of sampling campaigns

04/2005 – 07/2010 **Trier University, Germany**
Diploma studies applied environmental sciences (AUW)

- Major fields of study: soil science, geobotany, biogeography, physical geography
- additional: business administration (BA), spatial development and land-use planning
- Diploma thesis: „Biodiversity of macrosaprophagous soil organisms in a transect of energy crop land-use" (1,0)
- Honours Degree (1,0)

08/2003 – 03/2005 **Trier University and Johannes Gutenberg-University Mainz, Germany**
Studium universale

06/2003 **Hulda-Pankok-Comprehansive School, Düsseldorf, Germany**
Maturation with honours (1,0)

Further Qualifications

Languages **German** (mother tongue), **English** (fluent), **Spanish** (basic), **Polish** (basic)
IT skills Excellent in MS Excel, MS PowerPoint, MS Word, SPSS, Canoco
Good knowledge of GIS
Interests hiking, jogging, biogeography, literature, contemporary art