DISS. ETH NO. 28105

Scaling-up Soil Quality Assessments: Efficient Infrared-spectroscopic Workflows across Space and Time

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES OF ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

PHILIPP BAUMANN

BSc. Biology, ETH Zurich

MSc. Agroecosystem Science, ETH Zurich

born on May 7, 1989

citizen of Nesslau-Krummenau, Ennetbühl

examined by

Prof. Dr. Johan Six

Prof. Dr. Raphael Viscarra Rossel

Dr. Juhwan Lee

Prof. Dr. Bas van Wesemael

2021

Summary

Modern soil science relies upon the measurement of quantitative and qualitative soil attributes, and is, whenever possible, based on data that we typically measure in the laboratory and also in the field. Soil is a critical component of farming systems and is a limited resource, but also tends to be more and more degraded and lost due to human activities. Thus, sustainable soil management practices are needed more than ever to sustain and improve chemical, physical and biological aspects of soil quality. These are related to key ecosystem functions such as supplying water and nutrient to crops, carbon sequestration, nutrient recycling, and the promotion of biodiversity. Therefore, soil systems require a lot of data so that we can understand, localize or generalize processes in soils, and accordingly manage them from environmental and agronomic viewpoints. For example, we need a better integration of knowledge of how soil management practices (i.e, tillage, fertilization) affects the dynamics of soil's properties, to better manage this precious resource, to protect life across the ecological landscapes, and feed the world's population.

Soil properties emerge in function of their mineral and organic compositional diversity, but also processes governed by living components, across space and time. With the classical methods, we have not managed to collect enough static and dynamic parameters of soils' direct physical, chemical and biological composition and properties, for basic soil research and soil decision making (i.e., land planning, construction, environmental and agricultural policy). In Switzerland, only about 10% to 15% of agricultural soil landscapes have been sufficiently characterized to make informed land use decisions. This is mainly because most of the classical methods with established measurement protocols and procedures need a lot of time and financial capacities, for example for soil sample collection and processing. To complement and scale up estimates of the soils properties, diffuse reflectance infrared (IR) spectroscopy in the mid-IR and visible near infrared (vis–NIR) range are powerful methods to rapidly integrate the chemical and physical complexity of soils. Soil spectral libraries (SSLs) are harmonized collections of spectroscopic measurements and analytical reference values. The SSL can be used to estimate soil properties on new soil samples collected, primarily within but

not necessarily limited to the geographical regions covered. For robust predictions, a gradient from global machine learning methods using all data, to local chemometric modeling methods are used.

The research of my thesis evolved around creating SSLs from collected samples and soil archives. I developed spectroscopic data processing and modeling workflows to evaluate the trade-off between accuracy and requirement of updating with analytical measurements for agronomic soil quality assessments and local estimation for temporal soil monitoring. My thesis consists of the general introduction, three chapters, each of which represent this progressive evolution of local or regional modeling, general modeling and transfer learning. Finally, there is an overall conclusion of the work done and I give a research outlook.

In the first chapter, I built cost-effective diagnostic databases of soil quality in four project regions under yam production. I focused on specific soil proxies related to different yam production landscapes that then allow to scale up for other similar landscapes in West Africa. There I sampled typical soil variation in 20 fields of smallerholder farmers in each region, and complemented the soil library with 14 samples from the Land Health Degradation Framework. The purely local calibrations across all samples that we built for total carbon (C), nitrogen (N), sulfur (S), exchangeable calcium, effective CEC, diethylenetriaminepentaacetic acid (DTPA)-extractable iron and clay content gave excellent estimates ($R^2 > 0.75$), which can be recommended for screening major soil constraints at new sites within the region. Despite the small size of the library and a gradient in inherent soil fertility (texture; organic carbon (OC)) across the soil ecoregions (humid forest to Northern Guinean savannah), the calibration with a standard multivariate linear method gave good results that are comparable to SSLs developed with higher sample number and local soil sampling densities.

In the second chapter, I led the development of a mid-IR SSL for Switzerland ($n = 4374$). It was oriented towards local soil estimation and monitoring and based on time series data and soils from the Swiss Soil Monitoring Network (NABO) from 71 agricultural monitoring sites since 1985 ($n = 596$), and single-time measurements made at 1094 sites from the Swiss Biodiversity Monitoring program (BDM; $n = 3778$). Of the 16 properties we tested, ten showed good discrimination capacity ($R^2 > 0.72$) using mid-IR measurements and rule-based predictions with the CUBIST algorithm. Of these, total C, OC, total N, pH and clay showed very good agreement with the analytical measurements ($R^2 > 0.8$), and were almost unbiased across all data. We also designed a strategy for site-local adaptation based on performance-driven selection (RS-LOCAL), which we tuned with 2 reference observations for each of the 71 modeling sites and relevant observations from the SSL. Using such a transfer approach reduced the root

mean squared error (RMSE) for total C projections of repeated time measurements more than four times (RMSE(C) = 0.7 $g\,kg^{-1}$) for a monitoring site compared to the general rules. Interestingly, we also found substantial soil heterogeneity in the respective subsets from the SSL that were selected for localized modeling. The results suggest that the relatively large and diverse SSL has good potential to facilitate temporal soil monitoring at the plot level (10×10 m).

Finally, in the third chapter, I determined the uncertainty and effectiveness of spectroscopic measurement and modeling in combination with the established Swiss mid-IR SSL for detecting soil OC changes at individual plots at a long-term experiment (LTE; n = 311; five sampling times between 2002 and 2018) for the evaluation of sustainable farming practices. The estimation of measured OC changes between all possible consecutive time points and individual combinations of plot and depth had high cross-validation accuracy with partial least squares regression (PLSR) across all data points (n = 311; RMSE(ΔC) = 1.9 $g\,kg^{-1}$). The approach was only 1.3 times more inaccurate for cluster-based and data-driven sample transfer from the SSL using the RS-LOCAL algorithm. The transfer was highly effective because it only needed 2% of the LTE samples to achieve marginally lower prediction accuracy than the purely local calibration. Despite the promising results, high small-scale soil variation in mineralogy might reduce the information transfer that is related to the functional change of SOM. To mitigate this effect, there is the need for either better representation of the soil conditions at specific LTEs in SSLs, or more innovation in the learning scheme to normalize soil changes directly via the spectra.

The establishment of the first version of the mid-IR SSL of Switzerland with the NABO and BDM collections was a key deliverable to test the suitability of SSLs for systematic soil monitoring over time. With chapters two and three, my PhD research has been one of its first country-to-plot level transfers of a large spectroscopic collection. With the findings of this thesis I conclude that the operationalization of SSLs needs adequately designed workflows for soil information transfer and modeling that are tailored to the soil conditions at the respective monitoring locations. For effective learning, it matters how we present data to predictive algorithms. Global modeling with all data in SSLs is useful if data is really scarce and can sometimes work out of the box. Therefore, it is important to continuously update SSLs with new soil records. Simultaneously, to make best use of them, I stand for that we should infuse local adaptation and also independent validation samples to extract and verify knowledge for specific uses, such as soil monitoring, digital soil mapping, agronomic soil testing or soil OC accounting.

Zusammenfassung

Die modernen Bodenwissenschaften basieren auf den Messungen quantitativer Bodenattribute, deren Daten im Labor aber auch direkt in Böden im Feld erhoben werden. Der Boden ist eine wichtige Komponente von Agrarökosystemen und eine limitierte Ressource. Beispielsweise sind viele Böden auf der Welt erosionsgefährdet und wir erleben durch menschliche Aktivitäten zunehmend Verluste an Bodensubstanz. Deshalb wird es immer wichtiger, nachhaltige Bodenbewirtschaftungspraktiken zu realisieren, um die chemischen, physikalischen und biologischen Parameter der Bodenqualität zu verbessern. Diese Parameter haben einen starken Einfluss auf ökologische Schlüsselfunktionen wie Wasser- und Nährstoffversorgung von Pflanzen, Kohlenstoffspeicherung, Nährstoffkreisläufe und die Förderung der Biodiversität. Um Bodenumweltsysteme zu verstehen, interpretieren, generalisieren, aber auch im Kontext von umweltrelevanten und agronomischen Kriterien schützen und nützen zu können, benötigen wir eine grosse Menge Daten. Diese ermöglichen eine verbesserte Integration unseres Wissens, durch welches wir das Bodenmanagement optimieren können. So können beispielsweise reduzierte Bodenbearbeitung oder standortspezifisches Düngen den Boden als Ressource stärken. Ausserdem müssen wir den Boden als wertvolle Ressource besser bewirtschaften, um sowohl Leben in und auf den ökologischen Bodenlandschaften zu schützen und zugleich die Weltbevölkerung zu ernähren.

Bodeneigenschaften werden durch die mineralogische und organische Zusammensetzung und Struktur des Bodens bestimmt, welche wiederum durch Prozesse von Bodenorganismen über Raum und Zeit beeinflusst werden. Durch traditionelle Laboranaylsemethoden können statische und dynamische Attribute der bodenphysikalischen, -chemischen und -biologischen Zusammensetzung für die Grundlagenforschung und die Anwendung von Massnahmen (z.B. Landnutzungsplanung, Bauten, Umwelt- und Agrargesetzgebung), nicht in ausreichender Menge und schnell genug generiert werden. In der Schweiz sind nur 10% bis 15% der landwirtschaftlich genutzten Böden ausreichend beschrieben, um gut kalkulierte Landnutzungsentscheidungen durchzuführen. Einer der Gründe dafür ist, dass die etablierten Messmethoden und -verfahren viel Zeit und grosse finanzielle Mittel benötigen, so wie etwa für die Beprobung und Pro-

benvorbereitung. Für die Ergänzung und Skalierung von Bodendaten bietet sich die Infrarot(IR)spektroskopie in diffuser Reflektion im mittleren Infrarotbereich (mid-IR) sowie sichtbaren und Nahinfrarot (vis–NIR) Bereich an. Spektralbibliotheken von Böden (soil spectral libraries; SSLs) sind harmonisierte Sammlungen von Spektralmessungen und laboranalytisch–bestimmten Bodeneigenschaften inklusive Metadaten. Eine SSL wird genutzt, um Eigenschaften neuer Bodenproben schnell analysieren zu können, dies hauptsächlich jedoch nicht nur innerhalb einer geographisch abgedeckten Region. Um Bodenattribute robust und präzise spektral zu bestimmen, wird eine Palette von Modellierungsansätzen verwendet. So werden Regressionsverfahren lokaler Modelle, globale Machine-Learning Verfahren, welche alle Daten nutzen oder Transfer-Modellansätze, verwendet.

Die Forschung meiner Doktorarbeit konzentrierte sich auf den Aufbau und die methodische Umsetzung von SSLs aus Bodenarchiven. Ich entwickelte spektroskopische Datenprozessierungs- und Modellierworkflows um den den Trade-Off zwischen Schätzgenauigkeit und notwendiger Probenanzahl von laboranalytischen Messwerten für agronomische Bodenqualitätsdiagnosen und lokale Schätzungen für das Bodenmonitoring zu evaluieren und zu verbessern. Meine Dissertation umfasst eine allgemeine Einleitung, drei Kapitel, welche die progressive Evolution von lokaler Kalibration, genereller Kalibration und Transfer-Lernen abdeckt. Zuletzt folgt die Schlussfolgerung und ein Forschungsausblick.

Im ersten Kapitel entwickelte ich eine kostengünstige spektral-diagnostische Datenbank für Bodenqualitätsattribute in vier Projektregionen, in welchen Yams angebaut wird. Die Fragestellung beschäftigte sich mit spezifischen Proxygrössen des Bodens in verschiedenen ökoklimatischen Produktionslandschaften, mit dem Ziel die kalibrierte Methodik für ähnliche Landschaften in Westafrika anzuwenden und zu skalieren. Dazu beprobte ich Böden mit charakteristischer Variabilität von 20 Feldern von Kleinbauern in jeder Region und ergänzte die Bibliothek mit 14 Proben aus dem Land Health Degradation Framework. Die rein lokalen spektralen Kalibrationen aller Daten für Gesamtkohlenstoff (C), Gesamtstickstoff (N), Schwefel (S), austauschbares Kalzium, die effektive Kationaustauschkapazität (KAK), Diethylentriaminpentaessigsäure (DTPA)-extrahierbares Eisen und Tongehalt ergaben gute kreuzvalidierte Schätzwerte ($R^2 > 0.75$). Diese Eigenschaften können für das Selektion von Bodenfruchtbarkeitsindikatoren an neuen Standorten in den Regionen empfohlen werden. Trotz der kleinen Probenanzahl in der Bibliothek und einem Gradient der inherenten Bodenfruchtbarkeit (Textur, organischer Kohlenstoff) über die Ökoregionen (feuchter Wald bis nördliche Guineasavanne), ergaben die Kalibrationen mit einer linear-multivariaten Standardmethode gute Resultate. Diese sind vergleichbar mit SSLs, die deutlich mehr Proben und höhere geographische Beprobungsdichten aufweisen.

Im zweiten Kapitel übernahm ich die Entwicklung einer ersten mid-IR SSL für die Schweiz (n = 4374). Das Ziel war es, standortspezifische spektrale Schätzungen von Bodeneigenschaften anhand der 71 Zeitserien des Nationalen Bodenbeobachtungsnetz NABO der Agrarstandorte seit 1985 (n = 596) und den 1094 Standorten des Biodiversistätsmonitoring Schweiz (BDM; n = 3778) zu optimieren. Von den 16 getesteten Bodeneigenschaften konnte für zehn mittels mid-IR Messungen und regelbasierter Vorhersage mit dem CUBIST Verfahren, die gemessene Variation relativ zuverlässig abbildet werden ($R^2 > 0.72$). Von diesen zehn Eigenschaften zeigten Gesamt-C, Corg, Gesamt-N, pH-Wert und der Tongehalt sehr gute Übereinstimmung mit den analytischen Messungen ($R^2 > 0.8$). Die Modelle wiesen praktisch keine systematischen Fehler über alle Datenpunkte auf. Ich entwarf zudem eine Strategie, um standortspezifische Adapation der SSL via datengetriebener Selektion (RS-LOCAL) mittels je zwei Bodenproben mit Referenzanalysen für jede der 71 Modellierungsstandorte und den jeweils relevanten Beobachtungen der SSL durchzuführen. Mit einem solchen Transfer Learning Ansatz konnte die Wurzel aus mittlerer quadratischer Abweichung (root-mean-square error; RMSE) für die jeweilige Gesamt-C Zeitserie pro NABO Standort im Schnitt im Vergleich zu den generellen Regeln um mehr als Faktor 4 reduziert werden (RMSE(C) = 0.7 $\mathrm{g\,kg^{-1}}$). Interessanterweise fanden wir erhebliche Heterogenität der Bodeneigenschaften in den jeweiligen Teilen der SSL, die für die lokale Modellierung ausgewählt wurden. Die Resultate zeigen, dass relativ grosse und bodenchemisch diverse SSLs gutes Potential haben, um das zeitlich aufgelöste Bodenmonitoring auf Plot-Ebene (10×10 m) in Zukunft zu unterstützen.

Zum Schluss ermittelte ich im dritten Kapitel den Schätzfehler und die Effektivität von mid-IR Spektroskopie und Transfer Learning in Kombination mit der etablierten Schweizer mid-IR SSL, um Änderungen in Corg in individuellen Plots eines Langzeitexperiments (n = 311; fünf Probeentnahmezeitpunkte zwischen 2002 und 2018) festzustellen. Das Langzeitexperiment untersucht nachhaltige Agrarpraktiken. Die im Labor gemessenen Corg-Werte und berechneten Änderungen, welche ich zwischen allen aufeinanderfolgenden Zeitpunkten und individuellen Kombinationen von Plot und Tiefe berechnete, konnten mit relativ hoher Genauigkeit mit mid-IR und partial least squares regression (PLSR) reproduziert werden (n = 311; RMSE(ΔC) = 1.9 $\mathrm{g\,kg^{-1}}$). Im Vergleich dazu hatte ein cluster-basierter Ansatz mit datengetriebener Probenauswahl aus der SSL mittels dem RS-LOCAL Verfahren nur eine 1.3 mal höhere Ungenauigkeit (RMSE). Dieser Transfer war hocheffektiv, zumal nur von 2 % der Proben des Langzeitversuchs benötigt wurdenn (rein lokale Kalibration). Trotz der vielversprechenden Resultate scheint eine hohe kleinräumige Variation in der Mineralogie den Informationsgehalt des Transfers im Bezug auf die funktionellen Änderungen in der bodenorganischen Substanz einzuschränken. Um diesen Effekt zu reduzieren schlage ich vor,

biophysikalischen Bodenbedingungen von Langzeitexperimenten in die SSL mit spezifischen Proben zu integrieren. Zudem sollten innovativere Lernmethoden für die Modellierung entwickelt werden, die Veränderungen im Boden über Zeit direkt über die Spektren normalisieren können.

Die erste Version einer mid-IR SSL der Schweiz, welche NABO und BDM Proben einschliesst, war eines der Hauptziele meiner Doktorarbeit. Damit konnte ich geeignete Modellansätze für den Einsatz von Spektroskopie spezifisch für das Bodenlangzeitmonitoring testen. In den Kapiteln zwei und drei konnte ich während meiner Forschungzeit im Doktorat als eine der ersten Arbeiten eine relativ grosse SSL auf nationalem Level auf Parzellen-Ebene transferieren und erreichte dabei Modellfehler, welche normalerweise bei reinen Kalibrationen auf Feldskala mit 30 oder mehr Proben zu finden sind. Aus den Ergebnissen meiner Dissertation schlussfolgere ich, dass die Operationalisierung von SSLs entsprechend der Fragestellung und den lokalen Bodenbedingungen im Monitoring, zugeschnittene Abläufe für einen gezielten spektralen Informationstransfer benötigt. Für effektives statistitisches Lernen ist es massgebend, wie wir vorhersagenden Algorithmen die Daten präsentieren. Globale Modellierungsansätze mit allen Daten von einer SSL sind hilfreich, wenn es nicht genügend Daten für neue Prognosen von Bodeneigenschaften gibt. Sie funktionieren jedoch nicht immer zuverlässig ohne Rekalibration. Es ist daher sinnvoll, eine SSL kontinuierlich mit neuen Bodenproben zu ergänzen. Gleichzeitig stehe ich dafür ein, dass wir in Zukunft lokale Adaptions- und unabhänge Validerungsproben bereitstellen sollten. Damit können wir sowohl gezielt Wissen aus SSLs extrahieren als auch spezifische Anwendungen verfizieren, wie etwa für Bodenmonitoring, digitale Bodenkartierung, agronomische Diagnostik oder Bodenkohlenstoffinventarisierung.

Acknowledgments

My journey in soil spectroscopy and predictive modeling started with my MSc thesis, that I did in the groups of Prof. Johan Six and Prof. Emmanuel Frossard. Because I expressed my interest in plant nutrition, soil health and statistics, Emmanuel proposed my to do my MSc thesis in the YAMSYS project in Burkina Faso and Côte d'Ivoire. Interestingly, I very well remember them saying that IR spectroscopy is a very established tool and partial least squares regression is the toolbox to use when it comes to the data analysis. Very early on in my work, I realized that there is much more to do, which was a lucky case for me. I would like to thank both for showing me the gate for this path.

I remember very well my first class in Biochemical Modeling that Juhwan gave in my MSc at ETH Zurich, which I was attending with great interest. A bit before, I got to know that Juhwan will be my MSc advisor, which is why I wanted get to know him and attended the lecture. During my entire PhD and beyond, Juhwan has provided me constantly with kind support and humble opinions about any life and research question he thought fitted his knowledge or was personal matter he could support me on. I enjoyed our discussions, mostly with a cup of coffee, for example at CSIRO during the exchange. Besides, I was extremely lucky that I could count on Juhwan's quick responses and constructive feedback. He would typically encourage me to only take those inputs from him that I agreed with because he did not want to change my style too much. Without Juhwan, I would not have started and continued my PhD to the end. Thanks so much for being there for me, Juhwan.

After almost being a year in my PhD, it was time to found the *Formula Uno* office together with Kenza and Will. Soon after we surprised Marijn, who became acquainted with our not so random habits and conversations when he moved into our office. Usually, there was for example the dancing Friday afternoon, where we were allowed to listen to loud music and step our feet in the legendary office in the TAN building.

My colleagues and friends in the SAE group made my days coding R, writing and conceptualizing my research enjoyable. Because it is impossible to focus eight hours straight every day in front of one's computer, I particularly enjoyed the lovely and fun conver-

sations at lunch, coffee breaks in the kitchen, random chats in the PhD office, and of course more organized gatherings.

I look back to a very exciting modeling time in the SAE group. During peak time, we had our regular meetings of the modeling group together with Marjin, Roman, Erkan, Matti, Laura, Anatol, Magda. This was a really inspiring exchange for me. Parallel to that, we also started the reproducible science initiative in the SAE group.

I was happy that I could count on the support of the SAE lab team. Matti and Britta inspired me with their solutions and it was very straight-forward. This deserves a warm thank you.

During my PhD, I particularly enjoyed to engage in teaching in capacity building. There were quite some MSc students that were doing spectroscopic analyses in our group for different research projects. This was an ideal opportunity to test and further develop my R packages for spectroscopy.

Laura and Anatol were part of the spectral hacker crew in the D-USYS. It was very inspiring for me to transfer knowledge but also get a lot more ideas back from them. I remember writing Jo an email in middle of boulders in a forest in Ticino, that he should employ Laura so that she can further explore and extend the Central African SSL that she had started. It was a lot of fun to exchange ideas together and jointly implement and improve spectroscopy workflows in the SAE group together. Anatol was a fast learner from the beginning of his MSc. It turned out that he could make good use of the Swiss mid-IR SSL for the pilot study of mapping organic soils during his MSc thesis. We had a great time together in Australia. I would like to thank Anatol for the nice collaboration and the exciting time together.

Besides soil spectroscopy, Jonas enlightened me with his compelling research on crop phenotyping using IR spectroscopy. Thanks Jonas for the interesting personal and research discussions that we usually had with one or two coffees, teas or beers. Lea approached me with an interesting application of vis–NIR spectroscopy for estimating the starch content in red clover. The topic turned out into a nice side project that we could publish. Thanks Lea for the diversified discussions and coding sessions we had. Last but not least, I would like to thank herbal Peter for his very interesting use cases for spectroscopy and his good company.

Joining SYNCOPATION CLIMBING CLUB was the perfect activity for the weekend, after too many thoughts about the PhD during the week. Friendship bowls became cult and once even a big bear called Pimly made his way back to the office. The Corona lockdown phase was also big time for work retreats. For example, we once packed all our screens of ETH and climbing gear to work and climb in a wonderful resort in Lago

di Como in Italy. I would like to express my gratitude to Travis, Laura, Will, Anna, Mana, Thomas aka Mr. Propper, the tree, Dan, Luisa, Manuel, Matti, Shiva, Praveena, and the many guest climbers for the awesome company, the epic climbs and dance parties.

The PhD was also the phase to try new operating systems and make the computational environment, the computing capabilities and graphics as smooth, fast, and elegant as possible. I would also like to thank the IT support crew for providing such a fantastic infrastructure and service to make fast computations happen. The Euler batch system provided me with a nice learning experience in parallel computing concepts. Then, in the second phase of my PhD, I was a proud user of the D-HEST high-peformance virtual machine. Andreas aka Scheff gave me admin access to a dedicated VM so that I could use the infrastructure to its full potential with Cent OS. Whenever I had some tricky configuration I could not figure out of the box, he gave a hint how to solve it whenever he could. But I first had to learn how to write cryptically, too. Not only when talking about computers, but rather to have humble research debates, I particularly enjoyed Martin's presence. Thank you for being there Martin, and keep up the good research and mentoring.

Although I came to realize my capacity limits, the PhD in the Johan Six's group gave me a tremendous learning experience that I am really grateful for. While certain things had to get done and could be replaced by more in depth spectroscopic research only towards the end, I had a lot of freedom. I was involved in other collaborations with colleagues and did some interesting research where the use of soil spectroscopy was also very promising. I want to thank Jo for his patience and also the opportunities he gave me in his group. He made me like to get things done and I now often stick to his quote to always have a plan, even if you are a couch potato; which I before sometimes lacked. When having a meeting with Jo, I liked particularly that he remembered very well what we discussed last time. When it came to the science, Jo was always straight-forward. When I had a convincing idea that I supported with arguments or outputs, he would usually trust me on taking a certain pathway.

After every second Thursday's team meetings, the evening and night often ended in delightful beer adventures. I therefore have to admit the Friday was not the most early bird and productive day for me. Therefore, I had to rely on Saturdays and do this work and invest more of my evenings during the second phase of my PhD. If if was time for Pain Perdu, Willie the Whale and Simon the Tree often joined him, and with Julia the three of us almost went to Zagreb one Friday. Kenza motivated me to keep on exploring new teaching and programming adventures during my PhD. This culminated in stickers entitled "pain perdu in the tidyverse", that Kenza drafted for me. While I tried to code cleanly, both me and my neighbor Will had not the most organized desks.

I would like to thank Astrid for her patience with me during this productive time. I really enjoyed being in the same boat as my PhD colleagues. It was a very busy and social PhD time with Elena, Manon, Laura, Betsy, Ariani, Janina, Erkan, Karl, Simon, Rafa, Charles, Omolara, Rebecca, Ben, and many others. Roman inspired me with his critical re-search in the valley of shit, and his work and life philosphy.

With Jo I enjoyed having a conversation about a societal matter or having a chat about various behavioral aspects of people, which happened mostly quite spontaneously. I am grateful for Jo's good response times and straightforward communication via email. Right before submitting my thesis prior my defense, I could count on Jo's extremely quick editing skills. Thanks so much Jo for the time in SAE.

I was in the lucky position to already know where I will work after my core academic time in the middle of my PhD. At a pedometrics conference in Thübingen, Armin asked me whether I wanted to join the Kompetenzzentrum Boden (KOBO; Swiss Competence Center for Soils) to contribute with spectroscopic method development and operationalization to improve the national soil mapping in Switzerland. I did however not quite finish my PhD at ETH Zurich when I started the position at KOBO in October 2021. Armin was really generous and allowed me to work on my PhD during my employment. I am extremely thankful for this and would like to thank Armin for his kind support.

From Raphael Viscarra Rossel I learned how to make my writing more simple and appealing. I also picked up from him to have perseverance in getting articles accepted and published. I was lucky to get his good mentoring in spectroscopy and further had a really nice collaborative visit when I was doing an exchange in his group at CSIRO in Canberra, Australia. He introduced me to the local culture in Australia. Also I always enjoyed chatting and having a beer with him and other colleagues. When my health was in bad condition, I could always count on his support.

I would further like to thank Prof. Dr. Bas van Wesemael for agreeing to be the independent examiner of my PhD defense, and Prof. Dr. Sebastian Doetterl for taking the role as chair for my PhD examination.

Norah gave me a lot of positive energy during the past two years. It was wonderful and inspiring to spend our PhD time and more importantly our lives together. In the final run of my PhD, she had a lot of patience with me and helped out where she could. I want to thank her for the creative ideas, for being such a positive person, and deep conversations, and love that we have shared together.

My parents Claudia and Fredy deserve much praise for cheering me up and always supporting me during uplifting but also less pleasant times. Because they have known

my weaknesses and strengths better and longer then most, I can imagine that this has not been a trivial task. I could always count on a calming atmosphere at my parent's place, be it for writing or to rest, talk and go for walks together in the beautiful valley where I grew up. My younger siblings, my brother Manuel and my two sisters Nicole and Livia were often surrounded by me in nerd phase. If not, I enjoyed their company, gained deep understanding of our family's psychology together and developed some cool ideas such as our the family startup plan to grow nettle and extract protein and iron supplements from it.

With that, I am thanking family, friends and colleagues — certainly also the ones that I have not explicitly listed here — for being there for me during my PhD journey and beyond.

Contents

Part 1

General Introduction

1.1 The principles and history of soil sensing

The soil community delft into visible–near-infrared (vis–NIR; 25000–4000 cm^{-1}; 400–2500 nm) spectroscopic sensing already in the 1960ties and 1970ties, with experimental work pioneered by Bowers and Hanks (1965), and follow-up research by Bowers and Smith (1972). Their research, together with other colleagues, investigated the effect of soil moisture, organic matter contents and particle size on the diffuse reflectance patterns in the vis–NIR range. They reported that reflectance increased with decreasing moisture contents, smaller particle sizes, and soil organic carbon (SOC) contents (Peterson and Baumgardner, 1981; Baumgardner et al., 1986). First key research was conducted with focus on understanding the reflectance behavior of soils and nuisance factors of the measurement process on the field and in the laboratory, in prospect to proposed applications in soil survey, soil degradation assessment, and the development of soil information systems (Stoner et al., 1980; Baumgardner et al., 1986). These studies usually focused on a handful of soil types and texture classes, and simple Pearson correlative or linear regression analysis was usually used to focus on single wavelengths that discriminated the aforementioned soil characteristics that were studied (Sinha, 1987). However, robust empirical calibrations with reference measurements selected for the estimation of soil physical and chemical properties in many new soils, based on spectra only, were not yet developed at that time.

It was not until the early 1980ties that the correlative nature of soil infrared (IR) spectra and several soil composition and properties, such as minerals and soil organic matter (SOM), could in principle be efficiently used in a predictive manner. The IR reflectance spectra of minerals were systematically measured before, to understand their fundamental structure, composition and reactions better (Farmer, 1974). The emergence

of an interdisciplinary field called "chemometrics" had quietly sparked the *"application of mathematical and statistical methods to 1) improve chemical measurement process, and 2) extract more useful chemical information from chemical and physical measurement data"* (Workman et al., 1996). In chemistry, the application of spectroscopic methods to characterize chemical structure and concentration of specific compounds — but also their joint occurrence in chemical mixtures or complex biological media (Wold and Sjöström, 1977) — has also left many scientists and practitioners in the industry with much data (typically more variables than samples), but major obstacles in their analyses. The early analytic task was for example best selecting and combining relevant variables in spectra. Moreover, methods for reducing measurement noise and environmental nuisance factor were demanded by the community (i.e., humidity, temperature). This was because common methods such as multiple linear regression or principal components regression often not worked reliably enough for generalization. *"The problem of multivariate calibration"* (Wold et al., 1983), as it was termed by applied statisticians (chemometricians), was resolved to a significant proportion by the development of partial least squares regression (PLSR). This success, along with the continuous improvement and combination of methods for signal correction (e.g, Tan and Li, 2007) and calibration transfer between instruments and environments (e.g., Shenk et al., 1985; De Noord, 1994; Bouveresse and Massart, 1996; Tan and Li, 2007), has boosted a variety of applications of IR spectroscopy in various domains, such as analytical chemistry, the food industry, the agricultural sector, the pharmaceutical industry, mining, chemical manufacturing, and natural sciences until today (Martens and Næs, 1984; Martens and Naes, 1989; Ozaki et al., 2006).

The introduction of PLSR, and also the technological advances, for example Fourier-Transform (FT) IR instrumentation that became more affordable and the rise of scientific computing, was followed by wide-spread usage of predictive analytics with IR spectroscopy. With the increasing capacity of statistical methods to deal with many noisy and highly correlated (multicollinear) variables from spectroscopic measurements, calibrations became more robust, accurate, and generally accessible. Moreover, the analysts' toolbox had been gradually extended over time with more non-linear machine learning methods, such as multivariate adaptive regression splines (MARS; Friedman, 1991) and neural networks, and others (Gemperline et al., 1991; Sekulic et al., 1993). Parallel to the development of these classical non-linear statistical learning methods that were popularized in the chemometric scene and also machine learning in the 1980ties and 1990ties, instance-based learning (Stanfill and Waltz, 1986) developed into characteristic modeling principles. It gained relatively low attraction in its beginning but was found to be very successful for IR calibrations from mid to end of the 1990ies.

The rise of methods for memory-based learning (MBL) opened the realm of more complex, globally non-linear, but locally-linear statistical modeling (Wilson and Martinez, 2000b,a). It is also known under *instance-based* or *lazy learning*. However, the approach of *"quantification of the similarity concept (SIMCA)"* (soft modeling; classes) had been developed much earlier (Wold and Sjöström, 1977). MBL approaches create explicit model sub-sets for every new sample that needs to be predicted, using distance or dissimilarity metric and thresholds for nearest neighbors, and builds an explicit statistical model for each new prediction sample with a specific selection of instances from the training set (Aha, 1992). Although first criticized (Breiman et al., 1984), most of their limitations have been addressed (Volper and Hampson, 1987; Salzberg, 1990). Thereby, instance based principles were also further elaborated and incorporated into models such as rule-driven systems such as Cubist (Quinlan, 1987, 1993). Local centering (Lorber et al., 1996), and the famous LOCAL algorithm (Shenk et al., 1997), which weighted the relative model influence of neighboring samples by a distance function, are also forms of local learning. The latter influenced the development of the spectrum-based learner (SBL) (Ramirez-Lopez et al., 2013a,b), that perfected the use of spectral relatedness for modeling based on chemical similarity in complex soil collections. Altogether, this gave better generalization capacity to calibrate and predict attributes of more complex materials and media, such as compositional contents and concentrations of chemical analytes, organic plant compounds and mineral elements, and — much later — chemically complex soils over broader and more complete physicochemical ranges at continental and country-level soil variation (Shepherd and Walsh, 2002; Viscarra Rossel and Webster, 2012; Clairotte et al., 2016).

In the 1990ies, the combination of IR spectroscopy and statistical modeling had its breakthrough for quantitative soil estimation (Nguyen et al., 1991; Janik and Skjemstad, 1995; Ben-Dor and Banin, 1995; Couillard et al., 1997). Either mid-infrared (mid-IR; 4000–500 cm^{-1}; 2500–25 000 nm)) or NIR diffuse reflectance spectroscopy was combined with standard chemometric techniques — signal correction and PLSR — to estimate multiple soil properties and interpret the predictive ability with loading vectors. The advantage of NIR for soils is that one only needs sieving and no milling to produce robust spectra, that are not influenced by physical scattering. In contrast, mid-IR measurements and modeling need fine milling for good quality spectra, but allow very detailed assessments of SOM and the mineralogical composition. Another interesting IR application in soil science was the selection of soil samples from a large population of soils from major cultivation areas of Sweden, with retention of variation in major soil properties using principal component analysis on NIR spectra (Stenberg et al., 1995). Because some non-linear effects between specific spectral features and soil attributes were observed, subsetting of the calibration set into ranges of the properties of interest

improved calibrations as shown by Janik and Skjemstad (1995). In soils, this leads to mineralogically more distinct and chemically related subsets, and thus allowed the parsimonious modeling of linearly related spectral features and variables. This and other types of subsetting was reported to make learning algorithm locally more linear, less biased and more accurate (Sekulic et al., 1993).

In the beginning, preparing milled soils with potassium bromide (KBr) as a carrier was required to reduce absorbance and yield stable mid-IR spectra. Gains in efficiency and accuracy of IR spectroscopic methods from improved spectroscopic processing and modeling methods went hand in hand with the technological advances in instrumentation (i.e., making KBr redundant). The benefits in developing of larger soil spectral libraries (SSL) were that higher soil diversity with a larger pools of predictive information can make soil assessment in new, related geographical regions within the calibration range of soil properties more efficient (Brown et al., 2006), because of the pre-trained relationships in the SSL. To handle increasing soil complexity with the substantial enlargement of SSLs, non-linear predictive modeling — machine learning — was used relatively early on (i.e., MARS: Shepherd and Walsh, 2002; boosted regression trees Friedman, 2001 – Brown et al., 2006; Viscarra Rossel and Behrens, 2010). Large SSLs and *"general" modeling* over all data points, also called *"global" modeling*, became possible with non-linear machine learning methods, that conduct model-internal partitioning. Thus, general rules could be extrapolated more straight-forwardly. Rule-based learning with CUBIST for soil spectroscopy gave satisfatory results along with the establishment of the Australian continental SSL, which yielded very interpretable soil rules that partitioned the data space (Viscarra Rossel and Webster, 2012).

Nevertheless, global modeling with large SSL had in the meantime also received double-edged perception within the soil community; in particular, it had been questioned for specific target sites, where small-sized SSLs showed optimal prediction performance (Guerrero et al., 2016). Furthermore, SSLs built within individual countries had difficulties at predicting particular regions using a general modeling approach; however they were much less biased when spiking was used (i.e, Wetterlind and Stenberg, 2010; Guerrero et al., 2010, 2014). Still, growing SSL with more soil diversity has become a major driving force for research to make spectroscopic workflows more efficient. For example, the aforementioned spectrum-based learner (SBL; Ramirez-Lopez et al., 2013b) can leverage a variety of dissimilarity concepts and uses optimized principal components to decide for chemically projected spectral similarity (Ramirez-Lopez et al., 2013a). This has again advocated locally-linear learning for knowledge extraction from chemically diverse SSLs (i.e., Clairotte et al., 2016), however it was also similarly accurate as CUBIST for the USDA NSSC-KSSL library at continental extent (Dangal et al., 2019). Lately, an instance-based transfer learning method, called RS-LOCAL (Lobsey et al.,

2017) was developed. It selects optimal calibration sets from SSLs with minimal number of local augmentation samples, through a process of repeatedly sampling training observations from the SSL and dropping those that on average yield worst PLSR test predictions for the local augmentation samples, are successively removed from the training set. Satisfactory ratios of minimal amounts of reference measurements and estimation errors were found for novel regions and also for extending local soil estimation and monitoring (Helfenstein et al., 2021; Baumann et al., 2021).

There have been many attempts at subsetting large and diverse SSLs for yielding better modeling results based on various geographical, pedological, land use, and soil management factors, that influence the physical state and chemical conditions of soils in its distribution over space, depth, and time (Moura-Bueno et al., 2020; Vohland et al., 2022). Such apparently simple stratification of diverse soil data sets are often done for SSLs collected across regional, country, or even continental extents. As the extents vary, so does the sampling density over the soil landscapes. Although overall model-spectroscopic evaluation metrics can be improved, many of the criteria used to form appropriate groups within SSLs for modeling are still far from universally applicable for new soil samples. This is mainly because they do often not always reflect soil compositional similarity or do not embrace soils' beneficial diversity directly in the modeling. Although taxonomic information (i.e., soil order) has been found to improve predictions in grouped calibrations in vis–NIR, scarcity in detailed soil survey data can however limit these approaches (Vasques et al., 2010). In fact, the better results obtained for human decision-based subset modeling is confounded by the statistical model choices and learning setup, to extract useful representations in more locally linear relationships from generally non-linear data relationships. This caveat still often attains marginal attention and thus tends to be poorly addressed in current studies.

Notwithstanding the rise of machine learning in spectroscopic applications, general calibrations or global models of large and considerably diverse SSLs were often found to be considerably biased when estimating soils within the same country or with similar soil types but could be considerably improved with the addition of local data (Sankey et al., 2008). When parameters of models are optimized with a global measure of goodness of fit obtained over all data, this approach often implies that there are better models that are able to adapt to the structure in local data spaces for defined target application domains that render the global model incomplete (Hand and Vinciotti, 2003). With the global perspective, rules trained on a defined data population with discrete distributions of the response and marginal distributions of high-dimensional data, such as spectra, are in reality from multiple sources. These multi-source properties are often not specifically accounted for when a general spectroscopic model is tuned by optimizing mean performance on all training data. These empirical lim-

itations of general learning become intuitively clear when we expand our view to a sub-domain of machine learning — *transfer learning* (Pan and Yang, 2010; Pratt et al., 1991) — and also when we consider that soils have enormous chemical diversity. Statistical transfer learning is characterized by the transfer of relevant knowledge from the source domain(s) or task(s) (i.e., large training population(s)) to the target domain (i.e., new application, new region with soils collected to estimate spectroscopically). There are various ways to transfer information from the source domain to the target domain, such as instance-based transfer (i.e., RS-LOCAL; MBL). This is why, for example, it was recommended to use an external data set for optimizing tuning parameters (Ellenbach et al., 2020). To avoid over-optimistic model assessment after optimization for off-training set errors (OTE; i.e., Wolpert, 1995 for formal implications on CV), a second data set should be available that has not been touched in modeling (as principles of training, validation and test). While this is intuitively clear, spectroscopic measuring, modeling and estimation should consider these dualistic properties of source and target domain as integral part of the localized prediction workflow.

For regional and field-level calibrations of soil collections with a few soil types, the root-mean-square errors (RMSEs) of general empiric calibrations are often reported to significantly improve from 30 to 50, and often also until 100 calibration samples, while the increase in accuracy has a much flatter curve until 200 reference used for relatively defined, farm-to-regional general spectroscopic modeling in absence of transfer or adaptation (Wetterlind et al., 2008; Ramirez-Lopez et al., 2014; Debaene et al., 2014). Further, it is important to consider the methods for calibration sampling and its properties with regard to covering the multivariate spectral space and representing the measured response(s) well in its empirical distribution(s). The use of local-only calibrations at the soil spatial extents of farms or fields has been advocated before because already as few as 25 samples were successful for calibrations (Wetterlind and Stenberg, 2010; Vagen et al., 2010). From the current perspective it seems that interoperable knowledge between partly dissimilar soil physical and chemical conditions at different locations (e.g., Schirrmann et al., 2013) was mostly limited by the size of the available SSLs and particularly influenced by the model choice. Further, the framing of the spectroscopic modeling and estimation workflow —- defining the way how data were presented to algorithms for deriving adaptive forms of spectroscopic learning — seems crucial for statistical algorithms to adapt to structures in the data.

More recent advances come from further development of sensor technologies, such as microelectromechanical system (MEMS) sensors that have miniaturized interferometers on electronic chips (Xu et al., 2017b; Karyotis et al., 2021; Wang and Liu, 2021). These technologies enable the current development of smart sensing solutions that are of pocket size, and hence can be used directly on-field in soil surveys and soil mapping,

so that less samples need to be processed and analyzed in the laboratory.

1.2 Basics of soil mid-infrared spectroscopy and predictive diagnostics

The mid-IR is characterized by the fundamental vibrations of molecular bonds that occur within that wavenumber range. The energies of molecular bonds are quantized, meaning they have defined energies for different vibration modes such as stretching or bending. Infrared spectroscopy makes the vibrational transitions of molecular bonds, which are part of the respective functional molecular groups, visible in the form of spectra. The mid-IR spectral absorption features are relatively narrow compared to the NIR. Still, for soils that are chemically complex, characteristics of soil compounds overlap in a soil diffuse reflectance spectrum. Therefore, we need to apply supervised and unsupervised statistical methods so that we can systematically interpret and reproducibly use soil spectra for classification and prediction purposes. Here we mostly use the tool to estimate emerging soil properties that we traditionally measure in the laboratory by means of chemical and physical reference methods (Shepherd and Walsh, 2007).

The interaction of soil particles and light within the mid-IR range produces optical scattering effects, which mostly follow the physics of the Lorenz-Mie theory (Kerker, 1982). The scattering happens for spherical microparticles that are smaller or about the same size as the wavelength. Downstream effects of scattering are for example peak shifts, altered baselines, additive effects, or multiplicative changes in features of spectra. Typically, there is a critical radius r for each particle, which depends strongly on the refractive index of the particle (Dazzi et al., 2013). The scattering effects complicate the analysis and modeling of mid-IR spectra because they confound or irreversibly reduce the chemical information content. Hence, milling is strongly recommended to diminuate such effects prior spectroscopic measurements (Guillou et al., 2015; Deiss et al., 2020).

To further reduce present distortions and measurement noise in spectra, signal processing on the spectral matrix X can be applied prior or as part of the modeling (Varmuza and Filzmoser, 2016; Rinnan, 2014). Savitzky-Golay filtering is a popular method to smooth the spectra and to remove baseline effects through derivatives (Savitzky and Golay, 1964). There are many other preprocessing techniques, from multiplicative scatter correction (MSC), the standard normal variate (SNV; Barnes et al. 1989) to wavelets (Bruce and Li, 2001; Viscarra Rossel and Behrens, 2010; Igne et al., 2010). Many of these come from the chemometric research. While some methods are purely empirical, others are formulated by the chemical and physical theory. The SNV is a row-wise auto-

scaling the spectrum with the mean and the standard deviation across all columns of a sample. The choice of preprocessing methods for a particular data set is typically based on validation statistics. However, there are also re-projection methods to overlay calibrated data spaces onto new data sets (Ramirez-Lopez et al., 2013a; Roger and Boulet, 2018). The optimal preprocessing of spectra differs among data contexts. Hence, there is a trade-off between reducing spectral noise, producing processing artifacts and decreasing the absorption signal of a major soil constituent. For example, finite derivatives with insufficient smoothing can introduce noise, but together with smoothing (e.g., Savitzy-Golay first derivatives) they can enhance the relevant information of the spectral features relative to the modeled response.

The exploratory data analysis is an integral part of the spectral analysis. It typically follows the preprocessing step. Because of the multivariate nature of spectra, dimensionality reduction methods, such as principal component analysis (PCA), are computed to investigate broadly the relationship between spectra and the response, the modeled soil property. Dimensionality reduction is also combined with the analysis of atypical samples by outlier detection methods (Filzmoser et al., 2008, e.g.). Although the common dimensionality reduction are linear, there are non-linear methods available, such as self-organizing maps.

The statistical modeling of soil attributes with spectra and particularly the application of such models onto new soil populations requires thorough validation and testing (Stenberg et al., 2010). Data spliting or resampling from statistical learning theory serves this purpose (Friedman et al., 2008). It is particularly important that the data splitting covers the diversity of future prediction samples and also separates dependent units such as samples along the same soil core or other repeated measures. A typical complete data split involves a proportion of the data for model fitting, which is called the training (set). A validation set is used to verify the model performance depending on the preprocessing and the choice of empirical model parameters on held out data that is unseen during training. Lastly, the test set is an independent new data set to evaluate how model generalizes to predict new data. For spectroscopy applications, a totally independent sample drawn is often not available. To maximally use the available training data but reduce the over-fit at the same time, 10-fold cross-validation works often well. It is a good compromise that delivers relatively unbiased and low-variance estimates of performance for validation (Beleites et al., 2005; Kuhn and Johnson, 2013). Figure 1.1 illustrates one repeat of a 10-fold cross-validation, which was further grouped by the respective sites. The folds Fold01 to Fold10 illustrate the ten data segments.

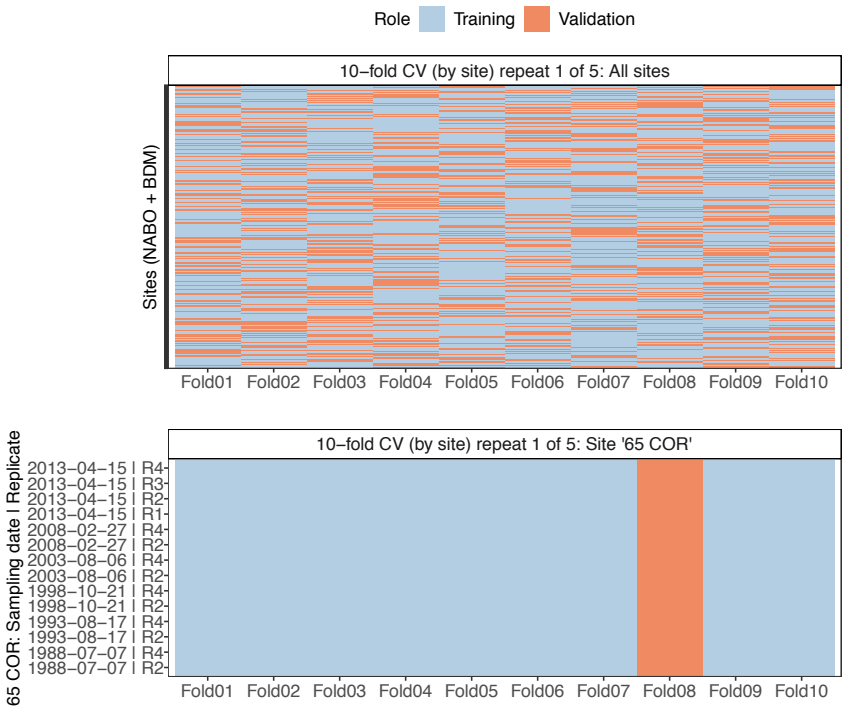

Figure 1.1: Illustration of a 10-fold cross-validation (CV) scheme that is grouped by location. *Top:* The x-axis shows the training and validation samples of the folds (splits), and the y-axis corresponding observations of individual sites. *Bottom:* Example of the data split of the first repeat for observations of an example site.

1.3 Extending quantitative soil research and information systems with soil spectral libraries

As our society evolves culturally and with the advancement of soil and related environmental domains of science, new interdisciplinary questions and challenges arise. Soils are in essence non-renewable because soil need in the range of 10000's (temperate zone) or millions of years (tropical soils) for their formation and transformation. Because soils have utmost importance for many pillars of world's ecosystems and human well-being, and nowadays are impacted more than ever by human activities causing biophysical changes, we need quantitative and objective measures to record soils' state and changes in landscapes and ecosystems (FAO, 2006; Dwivedi, 2017; Arrouays et al., 2021). Soil information systems need to be ready to fetch many attributes in systematic ways. We also need to update past, current and future description layers, and we need to be ready to answer new questions when our societal needs change, when research agendas change, or when we challenge the past concepts of co-existing with soils.

Soil information systems are the foundation to assess soil functions and to investigate fundamental questions of soil research. They help us also to understand the formation of soils and their variability in space and time. Maybe even more importantly, they enable us to focus on current environmental and agricultural questions, for example on how we evaluate and manage soil landscapes from a science perspective but also as a society (Bouma, 1989, 2014; Greiner et al., 2018). The definition of diagnostic features of soils in context of modern land use requires quantitative data on soils, which are often scarce in the context of traditional soil surveys; which is therefore augmented by the framework of digital soil mapping (DSM; Hartemink et al., 2008).

The induction of soil properties from soil-landscape relationships by data mining with large point databases (i.e., digital elevation models and terrain attributes, satellite bands, lithology, land use, etc.), serves the purpose of knowledge discovery for pedogenesis and landscape ecology, and extrapolation of soil maps within suitable extents (Bui et al., 2006). DSM has a long history of integrating both remote and proximal sensing methods together with statistical methods (Minasny and McBratney, 2016). In the beginning, pedometric research focussed on gamma spectrometry, electromagnetic induction and ground-penetrating radar (GPR) as proximal sensors for data input (Wong et al., 2010). Later on, along with the success and advancement of soil IR spectroscopy, it was increasingly used in soil landscape modeling and mapping (McCarty et al., 2010). Abundant soil profile measurements in a geographical context are needed to model the relationships of soil processes and to derive a spatial prediction map of soil properties. Hence, after soil diffuse reflectance spectroscopy and modeling had been widely estab-

lished and accepted as an alternative tool to scale up the production of key properties in the soil research scene and precision farming (Viscarrra Rossel and McBratney, 1998; McCarty et al., 2002), time was ready to propose it as a high-throughput and inexpensive technique to produce quantitative data on a handful of measurable properties that are key for soil mapping over larger areas (Rossel and McBratney, 2008). Nowadays, national soil classification are more frequently updated because of vis–NIR modeling and estimation that is integrated into a DSM approach (Teng et al., 2018; Seybold et al., 2019).

In support of soil spectroscopy, particularly in the infrared area, the data collection becomes more efficient and a larger amount of soil variability can be quantified compared to traditional wet chemistry. Proximal sensing is often primarily a data source for high-resolution DSM. Because of this, accelerating the time and decreasing the cost of analytical measurements and even sensing by improving computational methods has been and will advance in soil mapping in practice. Furthermore, an important component to improve the accuracies and the uncertainties of the soil maps is to develop better methods that can extract stored knowledge from SSLs more efficiently and accurately. With more sampling points and proxy measurements, the resolution of soil maps can be drastically improved. Further, good data quality of input attributes is important to calibrate accurate DSM models, based on rapid sampling and proximal measurement methods. This way more sampling sites and soil depth can be included.

Because vis–NIR and particularly mid-IR spectra contain a footprint of the mineralogy in soils, this information is frequently used to produce useful maps at different scales and extents, even over continents. For example, a map of soil types was made, which was supported by vis–NIR predictions from the Australian SSL. IR-based soil analytics brings significantly improved spatial support to DSM approaches, and increases the update frequency of such maps (Teng et al., 2018).

Previously, possibilities of the rapid estimation capacity of soil properties with diffuse reflectance spectroscopy had been elaborated for soil agricultural, environmental and engineering applications. Together with the development and use of soil spectral libraries (SSL), this enabled the prediction of new soil samples from the target area (Shepherd and Walsh, 2002). In the beginning, when initiatives and research institutions just started to build larger SSLs with relatively diverse physical, chemical, and biological attributes, to leverage the capacity to characterize new soils with broader ranges of attributes, Shepherd and Walsh (2002) emphasized that "global models may be more robust than local models in terms of ability to predict new samples". This was a strong statement for fostering the worldwide capacity building in SSLs, but as we will see later, this perspective on general vs. local modeling is not a question of "either-or".

Various research institutes and larger research organizations and networks have invested in building SSLs with standardized reference analysis methods and spectral measurement protocols. France has for example a national Vis–NIR and mid-IR spectral library that was developed with samples collected on a 16 km × 16 km grid across the country Clairotte et al. (2016). The LUCAS (Land Use/Cover Area frame statistical Survey; Tóth et al., 2013) open-access SSL (Stevens et al., 2013) constitutes about 18000 soil samples from 23 countries in the European Union (EU). There is also complementary sub-collections of the LUCAS SSL in the mid-IR (Leenen et al., 2022). The Australian soil vis-NhIR infrared spectroscopic database contains about 20000 samples with 24 measured soil properties that were compiled from the national inventories (Viscarra Rossel and Webster, 2012). In the United States, the USDA-NRCS Kellogg Soil Survey Laboratory has been assembling a mid-IR SSL with over 80,000 soil samples called KSSL (Dangal et al., 2019). The Brazilian vis–NIR SSL has started with a community effort and now contains about 40,000 samples with reference measurements (Demattê et al., 2019). The Chinese SSL contains about 4000 soil reference and Vis–NIR observations from 19 provinces. The Africa Soil Information Service (AfSIS) and the World Agroforestry Centre have grown a continental mid-IR SSL to deliver soil decision support to countries in sub-saharan Africa Sila et al. (2016); Vågen et al. (2020). With the 1800 soil samples in the central African SSL, the estimation of soil properties in the soils of the Congo Basin was improved because of the soil diversity in ten distinct geoclimatic regions (Summerauer et al., 2021). Finally, researchers have also compiled a collaborative vis–NIR SSL to enable predictions on the global scale (Viscarra Rossel et al., 2016). More recently, research capacity building has also shifted towards web platforms that provide spectral analysis and prediction services (e.g., Shepherd et al., 2022). Joint library collections have the advantage of being more flexible to new applications because they contain more pedological knowledge. At the same time, harmonization of the soil analysis methods and sometimes also the standardization of spectral measurements remains challenging. Altogether, these and other SSLs are joint efforts to make the analysis of soil properties more efficient and attractive for many research, decision making and operational contexts. This effort further requires reproducibility.

Optimizing biogeochemical processes such as soil C sequestration and nutrient cycling in agroecosystems requires frequent quantitative information on soil properties in large amounts and in relatively high spatial densities (Paustian et al., 2016). Characterizing soils' attributes in their spatial and temporal variation is key to establish a decision-making and intervention system in order to sustain soil functions and related ecosystem services, including agricultural production potentials. There is a manifold of studies which have framed general concepts that embrace soil function aspects, in particular soil fertility, over the last decades. For example, Haines-Young and Potschin

(2008) proposed a cascading framework for mapping relations between soil properties and processes, and the resulting services that the soil systems provide. Based on this cascading framework, Greiner et al. (2017) not only discuss soil function assessment methods that are adequate to describe multi-functionalities of soils and various ecosystem services, but also provide key soil property data and pedotransfer functions required to determine individual soil function assessment criteria. Last but not least, such soil physical, chemical or biological transfer functions are increasingly provided through spectroscopic modeling (Xu et al., 2017a; Yang et al., 2022; Baumann et al., 2022).

Soil monitoring networks and long-term experiments (LTEs) are an important backbone for providing soil inputs and process-level information of soil dynamics into modeling platforms. They deliver quality-assured measurements, controlled and well-described records on environmental and management factors. Both, experimental research and modeling in the realms of environment and agricultu re form contexts with high demand of quantitative soil data. For example, processed based models — e.g., soil-plant-climate biogeochemical models (Del Grosso et al., 2006; Lee et al., 2020) — need inputs of basic soil data such as texture, which become more readily available with IR spectroscopic predictions and larger SSLs.

1.4 The chemical nature of soil organic matter and its transformation processes in light of IR spectroscopy

Soil organic matter is — against to former beliefs of persistent recalcitrant humic substances — a vastly diverse collection of organic compounds that is under continuous decomposition through physical, biological and chemical transforming forces. The myth of humic substances was busted long time ago. For example, as Greenland et al. (1992) stated: *""Humus" as an identifiable substance that can be separated from other constituents of soil organic matter is a myth that has confounded soil scientists in the temperate as well as tropical zone for too many years."*. The gradient of functional complexity of soil organic matter spans from fresh plant and animal residues to biopolymers of different sizes until highly oxidized C compounds (Lehmann and Kleber, 2015).

Molecular diversity and functional complexity is currently understood as one of the main drivers of OC persistence (Lehmann et al., 2020). The three facets of complexity of SOM are molecular diversity, spatial heterogeneity, and temporal variability (Lehmann et al., 2020). Microscale spatial heterogeneity, namely the patchiness of SOM, is according to our current understanding of soil processes at high molecular level one of the

main factors for limited decomposition, which is caused by distance of substrate and enzymes of microbial decomposers. From a continental process perspective, lignin and carbonyl-C abundance had a positive relationship, while lignin and protein, and also carbohydrates and char-like aromatics, had a negative one (Hall et al., 2020). The stabilizing link between oxalate-extractable aluminium and carbonyl-C was not found for protein, which is also believed to cycle faster between living and dead microbial biomass. Thereby, major functional groups of O-alkyl, alkyl and aromatic C varied in their relative importance across a relatively variable set of soils (Hall et al., 2020). While biogeochemical description of controls on SOM are important from a coarse view across soil ecosystems over large areas of the world, the distinction of SOM into particulate organic matter (POM) and mineral-associated organic matter (MAOM) describes the association to size fractions and aggregates, and characterizes the vulnerability of SOM to decomposition (Cambardella and Elliott, 1992; Lugato et al., 2021). For the European continent, arable land use had higher proportions of MAOM than POM, while coniferous forest had the highest POM storages (Lugato et al., 2021). POM is less stabilized by mineral protection, and its fine fraction has even higher turnover rates, so that it is mostly preserved in no-till systems (Six et al., 1998).

The surplus C principle (Prescott et al., 2021) assumes that regenerative agricultural practices with optimal rates of nitrogen and phosphorus nutrient supply, which should promote root exudates hence diverse microbial biomass, can lead to the accumulation of mineral-associated organic matter (MAOM). However, for improving soil quality and resilience, to contribute to the mediation of climate change in medium-term by increasing C, we also need more biomass input from both aboveground and belowground plant residues. Most importantly, we must experimentally evaluate whether and when the proposed principle holds true.

Most of the aforementioned functional aspects of OC, apart from their microscale spatial arrangement, can principally be assessed by diffuse reflectance mid-IR spectroscopy, SSLs, and predictive modeling. For example, the IR reflectance behavior has been studied in detail through experimental work in the laboratory, with mixes of pure minerals and spiking of defined components of SOM (Calderón et al., 2013). The functional groups of SOM are relatively well separated and well described in the literature (Calderón et al., 2011). For example, aliphatic C−H stretching occurs at 2950 and 2870 cm^{-1}, vibrations of the C=C bonds in aromatic groups and C=O in ketonic group are located at 1660 cm^{-1}. In addition, OC contents in specific size fractions were predicted well with both NIR and mid-IR spectroscopy, across relatively diverse soils (i.e., MAOM; POM; Leifeld, 2006; Zimmermann et al., 2007; Bornemann et al., 2010; Viscarra Rossel and Hicks, 2015; Ramírez et al., 2021).

1.5 The importance of soil spectroscopy for future environmental monitoring, modeling, and for agricultural management

Soil infrared reflectance spectroscopy has traditionally been used as an integrative measure of soil quality, and hence has been proposed as tool to monitor chemical and biological aspects of soil quality (Terhoeven-Urselmans et al., 2008; Cécillon et al., 2009). Soil IR spectroscopy has also been proposed to derive more holistic soil health indicators — under the umbrella term *"evidence-based diagnostic surveillance approach"* (Shepherd and Walsh, 2007) — which is an emergent property of its capability to sense many aspects of the characteristic links of soil composition, chemical properties, and correlated biological proxies, to fundamental IR vibrations in functional groups. In the tropics, many soil properties are proportional to the levels of SOM or OC contents (Foster, 1981; Tiessen et al., 1994; Six et al., 2002). Its presence has for example a positive influence on the cation exchange capacity (CEC), because SOM has high surface binding capacity for cations relative to clay minerals, which are generally more weathered and have lower exchange capacity than in temperate soil (1:1 vs. 2:1 layer clay minerals).

Deriving and interpreting the measured and modeled trajectory of the functional diversity of soil OC compounds and their total contents requires the capacities to measure the response in fluctuating soil biological, chemical and physical properties. With the classical laboratory methods alone, it is unlikely that we will manage to verify a representative amount of soil effects after site-specific improvement measures. To study, monitor, and model soils, and to, for example, answer questions about impacts of climate change in different environmental contexts, it becomes more and more important to filter the relevant indicators that drive soil changes at specific locations individually and from systems perspective (Smith et al., 2012). The same applies for upscaling the conditions from local observational extents over larger areas. Spectroscopy, especially mid-IR spectroscopy modeling, can quantify a great deal of the functional diversity of SOM and the mineral components. It can hence be considered as a holistic fingerprinting solution that entails many soil health indicators in one measurement.

It is postulated that we could sequester more C in the form of soil OC compounds in the short to medium term (Paustian et al., 2016). SOM is key to soil functioning and health in general, which maybe an even better argument for sustainable management practices (Bai et al., 2018; Bünemann et al., 2018). At the moment, soil measurement, monitoring, reporting, and verification (MRV) is being heavily promoted as inter-operable science and societal framework, especially in the context of climate mitigation and soil

status improvement. This includes many components such as measurement and monitoring tools, protocols, decision frameworks, sustainable soil management practices, as well as data and software platforms to manage and use the tremendous knowledge collected. The mission statement for such harmonized frameworks across the world is concisely summarized as follows by the Soil Global Partnership of FAO: "In today's world, there is a strong demand for standardized, robust, reliable, cost-effective, and easily applicable MRV platforms to measure SOC change and GHG removals related to different agricultural systems." (FAO, 2020) . The compilation of these guidelines in FAO (2020) was initialized based on the Global Symposium on SOC (GSOC17), which was conducted in Rome in March 2017. There many follow-up programmes such as RECSOIL for the global recarbonisation of soils to maintain C rich soils and also reduce the degradation of croplands by sustainable soil management practices centered around SOC (FAO, 2021).

While the efforts in such international protocols are pivotal for scientific validation of soil responses in managed agroecosystems, and forming policies around farming practices, the protocol was only released under advisory nature. The *"protocol for measurement, monitoring, reporting and verification of SOC in agricultural landscapes"* (FAO, 2020) treats spectroscopic measurements mainly in the context of reducing the measurement uncertainties that stem from high spatial variability in SOC contents. Further, soil spectroscopic methods are advised only when an evaluation of the appropriateness of the calibration in the agroecological zone(s) the project is embedded into is given.

Such broad and general recommendations on soil OC reporting can be useful for the adaption and verification of sustainable farm management practices tailored to specific intervention areas (i.e., fields) at individual or multiple farms. Improving such spectroscopy-based soil diagnostic systems, so that they can become an effective component in MRV systems, is a key deliverable of active soil spectroscopic research. In the context of addressing knowledge gaps in soil science to resolve major environmental challenges, for which integrated soil monitoring and modeling will play an important role, members of the soil research community are calling out for more detailed testing of spectroscopic methods to reliably determine SOC change (Evans et al., 2021). As a matter of fact, ongoing research activities in the soil spectroscopic community reflect this comprehensible call from interdisciplinary teams of scientists that are pulling their forces together to advance the capacities of soil change assessments. We should further explore the use of advanced mobile IR sensors on-field (internet of things (IoT devices)), and find better ways to correct for homogeneous sample surface under field conditions, and find effective ways to correct for moisture (i.e., develop calibration transfer for it).

Direct measurements of soil OC stock changes require bulk density measurements across the depths, but also as many measurements of contents of different forms of C. Of these forms, at least roughly organic and inorganic C should be discriminated in the measurements. Particularly, the assessment across depth seems important because reduced vs. conventional tillage can have heterogeneous effects on OC storage depending on depth (Olson and Al-Kaisi, 2015). Spectroscopic methods can, broadly speaking, deliver rapid, cheaper and faster measurements than the reference methods, but they come at the cost of lower accuracy, often require adequate expertise in data mining, and (eventually) extensive calibrations. According to Paustian et al. (2019), major complications of direct measurement methods for determining SOC stock changes can be addressed with *"designing effective sampling methods and reducing the time and effort in sample processing and analysis"*. The same can be said about proxy measurements such as infrared spectroscopy, that greatly facilitate and scale the effect of classical measurements in combination statistical learning, and ultimately need less but representative reference analyses.

1.6 Outline of the thesis

The research of my dissertation was centered around the development of SSLs and their efficient use by statistical modeling and localized soil estimation workflows with state-of-the-art methods in the soil spectroscopy community. I developed strategies to maximize the general soil knowledge for local estimation suitable for agricultural diagnostics, long-term soil monitoring, and experimental soil research, so that we can minimize the need of new laborious and expensive laboratory measurements. Thereby, I focused on deriving 1) regional calibrations with purely local methods to assess many properties describing soil quality (chapter 1), 2) predictive models to estimate key soil properties with general and interpretable rule-based learning with incorporation of harmonized data from the Swiss long-term monitoring network and soil data of the biodiversity monitoring (chapter 2), and 3) adaptive modeling and a new data-driven method that combine relevant chemical knowledge in the diverse national SSL with minimal local analytical reference data from soil monitoring sites and a LTE to spectroscopically assess soil changes over space and time (chapters 2 and 3). Figure 1.2 depicts the spectroscopic modeling tasks and selected strategies that I have applied for the three chapters of my dissertation.

All three chapters show detailed performance assessments that compare the modeled mid-IR estimates of soil properties with their analytical reference measurements. Further, each research chapter includes an interpretation of important spectral features, and also the criteria of application, for example in the context of agronomic soil qual-

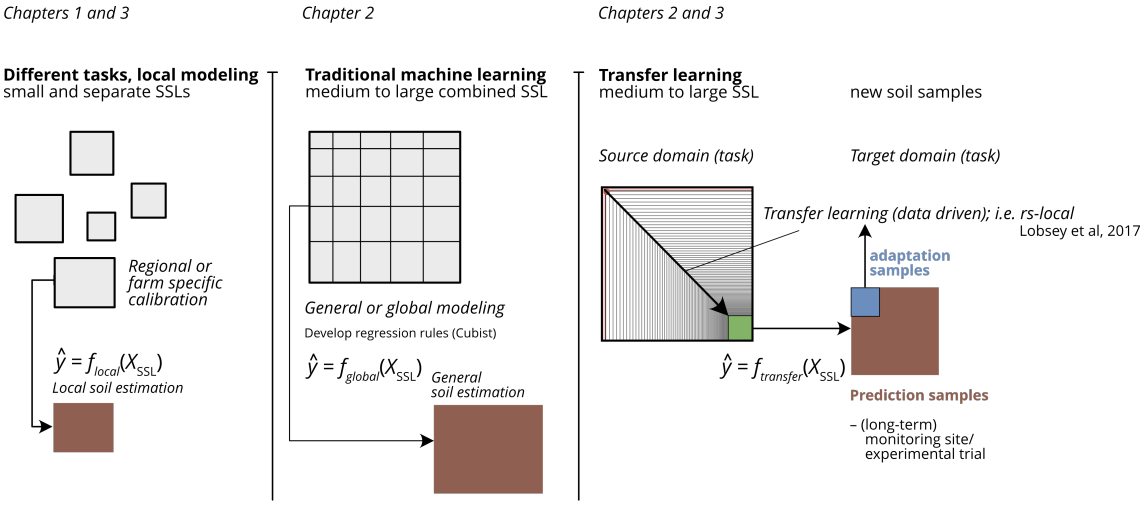

Figure 1.2: Conceptual modeling schemes for soil spectral libraries (SSLs) that I used for the three chapters of my PhD thesis.

ity assessments, are critically discussed.

Chapter 1: Estimation of soil properties with mid-infrared soil spectroscopy across yam production landscapes in West Africa

In chapter one, I addressed the following sets of related research questions around the development of a project-specific SSL for agricultural soil diagnostics in Ivory Coast and Burkina Faso:

1. How reliable are mid-IR regression models for a wide set of soil properties that are specifically developed for landscape-level soil variation in farmers' fields within four climatically distinct soil zones suitable for yam growth in West Africa?

2. What are the biophysical and statistically linked model mechanisms that allow the spectroscopic estimation of properties related to inherent soil status but also properties that regulate or are an agronomic proxy for nutrient availability?

3. What are the caveats of rapid soil diagnostics with the relatively small but agronomically targeted mid-IR SSL, and where are its benefits to derive region-specific and farm-adapted nutrient management strategies?

Chapter 2: Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring

In chapter two, where I and my collaborators developed a mid-IR SSL for broad quantitative soil characterization and local soil monitoring, was centered around the follow-

ing research aims and questions:

1. What is the predictive potential of including soil organo-mineral diversity when building a harmonized mid-IR SSL for Switzerland that includes soils from many selected geographical locations and also time-series data from national soil monitoring? How reliable can rule-based modeling extract general information for each of the 16 calibrated soil properties?

2. How useful is the developed SSL in combination with data-driven modeling to extract specific subsets of the SSL to perform localized estimation and monitoring of measured soil changes at the temporal monitoring sites? How does the accuracy of local transfer compare to the general modeling done under objective one?

3. What are the underlying spectroscopic features of the SSL and their associations to functional groups of characteristic organic and mineral substances of soils? How does their model contribution to the general model of total C differ in relative importance? What are the predictive mechanism behind the local transfer?

Chapter 3: Determining management induced SOC changes at the plot level based on soil infrared spectroscopy

Chapter three aimed at critically assessing the capacities of mid-IR spectroscopy for detecting management-induced changes of OC contents at plot-level over time for a LTE under organic agriculture. Previous research on this LTE has shown management-induced changes of OC, which was mainly due to reduced tillage practice. Local-only modeling vs. localized modeling with optimized subsets of the Swiss mid-IR SSL were tested for accuracy and analytical sampling efficiency. The research objectives were:

1. How does the local-only workflow with both standard PLSR and Cubist modeling on all available observations from the LTE perform in comparison to cluster-based similarity-driven and performance-driven transfer learning compare in terms of trade-off between the accuracy and number of local samples need for estimating OC contents?

2. What proportion of the measured differences at individual experimental plots and depths can be inferred with mid-IR spectroscopy and the three tested predictive modeling workflows? Are there particular trends in uncertainty for treatments?

3. What are the physico-chemical mechanisms that are involved in the estimation of site-local contents and changes in OC with mid-IR modeling? What are the rel-

ative importance of mineral and organic substance groups in soils inferred from the purely local calibration of OC, and to what degree can we explore and confirm the soil chemical variability in the experimental design with mid-IR spectra?

Part 2

Chapter 1: Estimation of soil properties with mid-infrared soil spectroscopy across yam production landscapes in West Africa

Philipp Baumann[1,*], Juhwan Lee,[2], Emmanuel Frossard[3], Laurie Paule Schönholzer[3], Lucien Diby[4], Valérie Kouamé Hgaza[5,6], Delwende Innocent Kiba[3,7], Andrew Sila[8], Keith Sheperd[8], and Johan Six[1]

[1]Group of Sustainable Agroecosystems, Institute of Agricultural Sciences, ETH Zurich, 8092 Zürich, Switzerland
[2]Department of Smart Agro-industry, Gyeongsang National University, Jinju, 52725, Republic of Korea
[3]Group of Plant Nutrition, Institute of Agricultural Sciences, ETH Zurich, 8315 Lindau, Switzerland
[4]World Agroforestry Centre (ICRAF), Côte d'Ivoire Country Programme, BP 2823 Abidjan, Côte dIvoire
[5]Centre Suisse de Recherches Scientifiques en Côte dIvoire, 01 BP 1303 Abidjan, Côte d'Ivoire
[6]Département d'Agrophysiologie des Plantes, Université Peleforo Gon Coulibaly, BP 1328 Korhogo, Côte d'Ivoire
[7]Département Gestion des Ressources Naturelles et Systèmes de Production, Centre National de la Recherche Scientifique et Technologique, Institut de lEnvironnement et Recherches Agricoles, 01 BP 476 Ouagadougou, Burkina Faso
[8]Land Health Decisions, World Agroforestry Centre (ICRAF), Nairobi, Kenya

[*]Corresponding author Email: baumann-philipp@protonmail.com

This chapter is a reprint of the paper published in SOIL, 7, 717731, 2021,
https://doi.org/10.5194/soil-7-717-2021

Abstract

Low soil fertility is challenging the sustainable production of staple crops in the yam belt of West Africa. Quantitative soil measures are needed to assess soil fertility decline and to improve crop fertilization management in the region. We developed and tested a mid-infrared (mid-IR) soil spectral library to enable timely and cost-efficient assessments of soil properties. Our collection included 80 soil samples from four landscapes (10 km×10 km) and 20 fields/landscape across a gradient from humid forest to savannah and 14 additional samples from one landscape that had been sampled within the Land Health Degradation Framework. We derived partial least squares regression models to spectrally estimate soil properties. The models produced accurate cross-validated estimates of total carbon, total nitrogen, total sulfur, total iron, total aluminum, total potassium, total calcium, exchangeable calcium, effective cation exchange capacity,and diethylenetriaminepentaacetic acid (DTPA) extractable iron and clay content ($R^2 > 0.75$). The estimates of total zinc, pH, exchangeable magnesium, bioavailable copper and manganese were less predictable ($R^2 > 0.50$). Our results confirm that mid-IR spectroscopy is a reliable and quick method assess the regional-scale variation in most soil properties, especially the ones closely associated with soil organic matter. Although the relatively small mid-IR library shows satisfactory performance, we expect that frequent but small model updates will be needed to adapt the library to the variation of soil quality within individual fields in the regions and their temporal fluctuations.

2.1 Introduction

Yam (*Dioscorea* spp.) is an important food and cash crop in West Africa. The yam belt of West Africa spans across the central zone of coastal countries in West Africa, located across the humid forest zone and northern Guinean savanna. It contributes to about 92 % of total world yam production, e.g. a total yield of 72×10^6 t in 2017 (FAO, 2014a). The cropping area in the West African yam belt has been expanded with accelerated population growth, which has in many places caused soil degradation. Furthermore, there is a trend of shortened fallow periods in the cropping areas of West Africa over the last decades, which has further exacerbated the decline in soil fertility across the yam belt. Traditionally, yam is grown without external input in these areas. Therefore, the production of yam and other crops grown in the region depends on soil organic matter

(SOM) status (Padwick, 1983), which serves as a main pool of plant-available nutrients and provides cation exchange surfaces for soil nutrients (Syers et al., 1970; Soares and Alleoni, 2008). Particularly, a strong positive relationship between high organic matter stocks and yam productivity is reported after fallow and when no fertilizer is added (Diby et al., 2009; Kassi et al., 2017). Thus, maintaining or increasing SOM and available nutrient levels is of utmost importance for sustainable production of yam and other crops in West Africa (Carsky et al., 2010). Furthermore, linking soil properties and yam yields (Frossard et al., 2017) and accounting for soil macro- and micronutrient status (O'Sullivan and Jenner, 2006) are fundamental to improving crop yields and soil management strategies.

Soil fertility is an integrative measure of soil attributes and their interactions that support the long-term agricultural production potential. Soil fertility is commonly decomposed into three main components, the physical, chemical and biological (Abbott and Murphy, 2007). Here, it is important to interpret soil fertility in the form of soil conditions and functions at an adequate resolution over time and space, and in relation to the crop of interest. For yam, low tuber yields are often attributed to an unbalanced ratio of essential nutrients (i.e. N, P, K) available in the soil (Enyi, 1972) and a fast mineralization and hence depletion of organic matter (Carsky et al., 2010; Hgaza et al., 2011). Yet, the relationship between soil properties and tuber yield is not fully understood (Frossard et al., 2017). The reason is that the response of yam to mineral fertilization is highly variable because of confounding environmental and management variables, such as climate, soil type, micronutrient deficiencies, seed tuber quality and planting density or disease pressure across the yam belt (Kang and Wilson, 1981; O'Sullivan and Jenner, 2006; Cornet et al., 2016). Further, there are no soil fertility recommendations specific for yam under West African conditions. For this reason, establishing yam field trials designed with different organic and mineral fertilization strategies within different yam-growing regions is required to optimize yam fertilization targeting regional soil and environmental conditions (Frossard et al., 2017). Despite the importance of soil fertility, it is challenging to quantify soil measures at sufficient temporal and spatial resolution to relate them to yam productivity together with other management effects.

In order to quickly assess key soil properties, such as soil organic carbon (SOC) and cation exchange capacity (CEC), we need more cost- and time-efficient methods in addition to the traditional wet chemistry laboratory analyses that are often cost-intensive and time consuming. Proximal sensing is a method that can provide reliable soil measurements rapidly and inexpensively (UNEP, 2012). Soil visible and near-infrared (vis-NIR) and mid-infrared (mid-IR) diffuse reflectance spectroscopy has gained popularity over the past 30 years to assess soil properties to complement conventional laboratory analytical methods (Nocita et al., 2015a). For model development and calibration

but, importantly, also for validation purposes, soil IR spectroscopy requires laboratory reference analysis data. Previous studies have shown successful spectroscopic predictions of soil properties, such as organic C, texture, cation exchange capacity (CEC), and exchangeable K (Viscarra Rossel et al., 2006; Cécillon et al., 2009; Nocita et al., 2015a; Sila et al., 2016). Many soil chemical and physical properties, such as soil mineralogy, and the concentration, forms, and distribution of SOM, are closely associated with IR spectral diversity. Nevertheless, soil IR spectroscopy often needs laboratory reference analysis data for model development and calibration. Further, a library that includes a broad range of soil biophysical conditions found in the region in which it is used needs to be established. Depending on the study scale — field (e.g., Cambou et al., 2016), region, country (e.g., Clairotte et al., 2016), continent (e.g., Sila et al., 2016)), world (e.g., Viscarra Rossel et al., 2016) — various statistical predictive modeling strategies are typically employed to account for regional variability in soil properties and determine empirical relationships between spectra and soil attributes. However, particular regions in spectra are characteristic for functional groups of soil components and thus, elucidating spectral features that are important for the prediction of a particular soil attribute helps to understand and validate the mechanisms based on which the empirical-derived models predict the soil properties.

Thus the main objectives of this study are to (1) develop and evaluate mid-IR spectroscopic models to estimate soil properties for selected landscapes representing major soil and climatic conditions in the West African yam belt, (2) to determine important spectral features for specific soil properties, and (3) to build a new soil spectral library in four landscapes of the West African yam belt for soil prediction and assessment.

2.2 Materials and methods

Landscapes and soil sampling

Our study area covered the climatic and soil biophysical conditions representative of the West African yam belt. We selected four landscapes, two in Côte d'Ivoire and two in Burkina Faso. Each landscape (10 km x 10 km) represents a diverse geographic ecoregion. The landscapes cover a gradient between humid forest and the northern Guinean savannah. Specifically, the landscape Liliyo in Côte d'Ivoire is at 5.88°N and in the humid forest zone. The predominant soil type is Ferralsol (FAO, 2014b). The landscape Tieningboué in Côte d'Ivoire is at 8.14° N and belongs to the forest savannah transitional zone. The soils are dominated by Nitisols and Lixisols (FAO, 2014b). The landscape Midebdo is at 9.97°Năand in the sub-humid savannah of Burkina Faso. Its dominant soil types include Lixisols, Gleysols, and Leptosols (FAO, 2014b). The landscape Léo is at 11.07°N and in the northern Guinean savannah of Burkina Faso and has Lixisols and Vertisols as the dominant soil type (FAO, 2014b). The mean annual rainfall were approximately 1300mm in Liliyo, and 900mm in Tiéningboué, Midebdo, and Léo.

During July and August 2016, we sampled the soil from a total of 80 fields under yam cultivation across the four landscapes, i.e. 20 yam fields in each landscape. The fields were selected in advance by taking into account visual variation in soil color and texture across the landscape. The yam fields selected contained the maximum soil variability based on soil color and cropping history, taking into account both local farmers' knowledge on soil fertility and agronomic extension expertise. Yam is typically planted on soil mounds, ranging from 5000 to 10000 mounds per hectare with a single yam plant per mound. Within each field, we sampled the soil at four adjacent mounds in square arrangement, which were spaced between 0.5 and 2 m. At each mound, six to eigth auger cores (2.5 cm in diameter) to the 30 cm depth were taken at a radius between 15 and 30 cm away from the center of a mound, depending on the size of the mounds. Then the soils from the four mounds were combined into one composite sample per field (around 500 to 1000 g of soil).

An additional set of 14 composite soil samples was collected by the International Center for Research in Agroforestry (ICRAF) at Liliyo from one sentinel site called "Petit-Bouaké" (UNEP, 2012). Sampling took place between 25 and 29 August, 2015 at positions that were previously selected for the Land Degradation Surveillance Framework (LDSF) in a spatially stratified manner (Vagen et al., 2010). The soil samples received from ICRAF were within the same landscape as the sampled soils in Liliyo within YAMSYS, but sampled from different positions. All soil samples were air-dried and stored in plastic bags until further analysis.

Soil reference analyses

The air-dried soil samples were crushed and sieved at 2 mm. About 60 to 70 g of the sieved soil was oven-dried at 60 °C for 24 hours, of which 20 g was ball-milled. All chemical analyses except soil pH were conducted both on the soils sampled in yam fields ($n = 80$) and the LDSF soils obtained from ICRAF ($n = 14$).

The milled soils were analyzed for total C and macronutrient (N and S) concentrations using an elemental analyzer (vario PYRO cube, Elementar Analysensysteme GmbH, Germany). For each of the four landscapes, two soils were selected and analyzed based on three analytical replicates for quantifying within-sample variance of the elemental analysis. For the remaining samples, the analysis was not repeated. Sulfanilamide was used as a calibration standard for the dry combustion. For pH determination 10 g of air-dried soil per sample was placed in a 50 mL Falcon tube and 20 mL of de-ionized water was added. The samples were shaken in a horizontal shaker for 1.5 hours and measured for pH using a pH electrode (Benchtop pH/ISE meter model 720A, Orion Research Inc., USA).

Bioavailable micronutrient (Fe, Mn, Zn, and Cu) concentrations in soils were determined with the diethylenetriaminepentaacetic acid (DTPA) extraction method, as described in Lindsay and Norvell (1978). The extracting solution consisted of 0.0005 M DTPA, 0.01 M $CaCl_2$, and 0.1 M triethanolamine. Briefly, 10 g of the sieved <2 mm) soils was extracted with 20 mL of DTPA solution. Micronutrient concentrations in the filtrates were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES; using a Shimandzu ICPE-9820 plasma atomic emission spectrometer). Final DTPA-extractable concentrations of Fe, Mn, Zn, and Cu were calculated back to per kilogram dry soil. For each landscape, two soils were selected and analyzed in triplicate to assess analytical errors. For the remaining soils the analysis was not repeated.

For each sample, the concentrations of total element (Fe, Si, Al, K, Ca, P, Zn, Cu, and Mn) in the soil was assessed by energy dispersive X-ray fluorescence spectrometry (ED-XRF) measurements on 4 g of the milled soil with a SPECTRO XEPOS instrument (SPECTRO Analytical Instruments GmbH, Germany). The soil was mixed with an equal amount of wax using a ball mill and pressed into pellets. Exchangeable cations (Ca^{2+}, Mg^{2+}, K^+, Na^+, and Al^{3+}) were determined with the $BaCl_2$ method (Hendershot and Duquette, 1986). About 2 g of the air-dried soil (<2 mm) was extracted by shaking for 2 hours with 30 mL of 0.1 M $BaCl_2$ on a horizontal shaker (120 cycles min^{-1}). The suspension was filtered through no. 40 filter paper (Whatman, Brentford, UK). For each landscape, two soils were analyzed in analytical triplicates. The concentrations of exchangeable cations in the $BaCl_2$ extract were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES, Shimandzu Plasma Atomic Emission

Spectrometer ICPE-9820). Different $BaCl_2$ extract dilutions were used in order to obtain an optimal signal intensity for the quantification of specific elements across all samples. Concentration of H^+ per kilogram of dry soil was calculated based on the pH measured in the $BaCl_2$ extractant. The $BaCl_2$ extraction does only slightly modify pH and is therefore an appropriate method to calculate effective CEC (CEC_{eff}) at native soil pH. Using the concentrations of the $BaCl_2$-extractable cations (i.e., Ca^{2+}, Mg^{2+}, K^+, Na^+, Al^{3+} and H^+), CEC_{eff} was calculated as the sum of exchangeable cations in centimoles (cmol) of cation charge per kilogram of dry soil. Exchangeable acidity was defined by the sum of exchangeable Al^{3+} and H^+. Base saturation in percent was calculated as a ratio of the sum of basic cations (Ca^{2+}, Mg^{2+}, K^+, Na^+) in cmol(+) per kilogram of soil to the CEC_{eff} multiplied by 100.

Particle size analysis was conducted by IITA in Cameroon as described in Bouyoucos (1951). Briefly, 50 g of dried 2 mm sieved soil was stirred with 50 mL 4 % sodium hexametaphosphate and 100 mL of deionized water in a mixer, for breaking down the aggregates into individual particles. Readings with a hydrometer (ASTM 152 H, Thermco, New Jersey, USA) were taken after letting it stand in the suspension for 30 minutes. The silt content was calculated by subtracting the measured proportion of sand and clay from 100 %.

Spectroscopic measurements

The milled soils ($n = 94$) were measured on a Bruker ALPHA DRIFT spectrometer (Bruker Optics GmbH, Ettingen, Germany), which was equipped with a ZnSe optics device, a KBr beamsplitter, and a DTGS (deuterated triglycine sulfate) detector. Mid-IR spectra were recorded between 4000 cm^{-1} and 500 cm^{-1} with a spectral resolution of 4 cm^{-1} and a sampling resolution of 2 cm^{-1}. Reflectance (R) spectra were transformed to apparent absorbance (A) using $A = \log_{10}(1/R)$ and corrected for atmospheric CO_2 using macros within the OPUS spectrometer software (Bruker Corporation, US). The spectra were referenced to a IR-grade fine ground potassium bromide (KBr) powder spectrum, which was measured prior to the first soil sample and measured every hour again. All spectra were recorded by averaging 128 measurements for each of the three sample repetitions per soil.

Spectroscopic modeling

Processing of soil spectra

Three replicates of spectra were averaged for each sample. The spectra were transformed by using a Savitzky–Golay-smoothed first derivative using a third-order polynomial and a window size of 21 points (42 cm^{-1} at spectrum interval of 2 cm^{-1}) (Sav-

itzky and Golay, 1964). Prior to spectral modeling, Savitzky–Golay-preprocessed spectra were further mean-centered and scaled (divided by standard deviation) at each wavenumber.

Model development and validation

The measured soil properties were modeled by applying partial least squares regression (PLSR) (Wold et al., 1983) with the preprocessed spectra as predictors. The models were fitted using the orthogonal scores PLSR algorithm. A 10-fold cross-validation, repeated five times, was performed to provide unbiased and precise assessment of PLSR model performance (Molinaro et al., 2005; Kim, 2009). For each individual soil property, the number of factors for the most accurate PLSR model was tuned separately. For each soil property model, the sample set was repeatedly randomly split into $k = 10$ (approximately) equally sized subsets without replacement for all repeats $r = 1, 2, .., 5$ and all candidate values in the tuning grid with the number of PLSR factors (ncomp) $= 1, 2, ..., 10$. Within each of the $r \times \text{ncomp} = 5 \times 10 = 50$ resampling data set splits, each of the 10 possible held-out and model fitting set combinations (folds) was subjected to candidate model building at the respective ncomp, using $k-1 = 9$ out of 10 subsets and remaining held-out samples were predicted based on the fitted models. The root mean square error (RMSE, eq. (2.1)) of the held-out samples was calculated by aggregating all repeated K-fold cross-validation predictions ($\hat{y}_i$) and corresponding observed values (y_i) grouped by ncomp, which resulted in a cross-validated performance profile RMSE vs. ncomp.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{2.1}$$

Based on this performance profile, the minimal ncomp among the models whose performance was within a single standard error ("one standard error rule", (Breiman et al., 1984)) of the lowest numerical value of RMSE was selected.

Model assessment was done with the best factors for each property using cross-validation hold outs. We reported the cross-validated measures RMSE, R^2 (coefficient of determination) obtained via linear least squares regression, and ratio of performance to deviation (RPD), after averaging predictions across repeats. The RPD index is the ratio of the chemical reference data standard deviation to the RMSE of prediction.

$$\text{RPD} = \frac{s_y}{\text{RMSE}} \tag{2.2}$$

Besides calculating the above listed performance measures, the uncertainty of spectral estimates was graphically reported for each soil sample, using prediction means and 95% confidence intervals derived from cross-validation repeats ($n = r = 5$; Eq. 2.3 and 2.4).

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{\hat{y}_i}\right)^2 \tag{2.3}$$

$$\overline{\hat{y}_i} \pm t(n-1, 1-\alpha/2)\frac{S_n}{\sqrt{n}}; \alpha = 0.05 \tag{2.4}$$

In order to cover the full training data space in the models for future sample predictions, the final PLSR models were rebuilt using the entire training set and the respective values of optimal final number of PLSR components determined by the procedure described above.

Model interpretation

The mid-IR spectra contain complex information about soil composition and properties. To establish a predictive relationship, statistical models need to find relevant spectral features for each soil property. Model interpretation requires a variable importance assessment to decide on the contribution of spectral variables to prediction and to explain spectral mechanisms. Therefore, we conducted model interpretation based on the variable importance in projection (VIP) method (Wold et al., 1993; Chong and Jun, 2005), using the model at respective best number of factors (ncomp). The VIP measure v_j was calculated for each wavenumber variable j as

$$v_j = \sqrt{p\sum_{a=1}^{A}\left[\text{SS}_a\left(w_{aj}/\|w_{aj}\|\right)^2\right] / \sum_{a=1}^{A}(\text{SS}_a)} \tag{2.5}$$

where w_{aj} are the PLSR weights for the a^{th} component for each of the wavenumber variables and SS_a is the sum of squares explained by the a^{th} component:

$$SS_a = q_a^2 \, t_a^\top \, t_a \tag{2.6}$$

where q_a are the scores of the predicted variable y and t_a are the scores of the predictors X. These VIP scores account for multicollinearity found in spectra and are considered to be a robust measure to identify relevant predictors. Important wavenumbers were classified with a VIP score above 1. A variable with VIP above 1 contributes more than average to the model prediction. For model interpretation, we only computed VIP at the respective finally chosen number of PLS (partial least squares) components a_{final} for each considered model. We focused on a selection of three well-performing models with $R^2 \geq 0.8$ (RPD ≥ 2.3) to illustrate model interpretation. These were total C, total N and clay content.

Statistical software

The entire analysis was performed using the R statistical computing language and environment (version 3.6.0) (R Core Team, 2017). We used the pls (Mevik et al., 2019) package for PLSR, as described by Martens and Naes (1989). Cross-validation resampling, model tuning, and assessment was done using the caret package (Kuhn et al., 2019). Custom functions from the simplerspec package were used for spectroscopic modeling (Baumann, 2019). All data and code to reproduce the results of this study are available online via Zenodo (Baumann, 2020).

2.3 Results

Measured properties and mid-IR estimates of yam soils

The distribution of soil properties at the yam fields showed a wide variation across the landscapes (Figure 2.1). Total C concentrations across all fields ranged from 2.4g C kg^{-1} to 24.7g C kg^{-1}. Total C values at the landscape scale were the lowest (median) in Léo and the highest in Tiéningboué. Soils from yam fields in the two landscapes from Côte d'Ivoire (13.0 ± 5.4 g C kg^{-1} soil; mean ± standard deviation) had relatively higher total C compared to the fields in the landscapes in Burkina Faso (6.1 ± 3.6 g C kg^{-1} soil). The median value and variation of CEC_{eff} exhibited similar patterns across the landscapes to total C. Total N concentrations across all fields ranged from 0.18g N kg^{-1} to 2.48g N kg^{-1}. Total N within and across the four landscapes exhibited a similar pattern as total C. Generally, the landscapes in Burkina Faso were low in total N compared to those from Côte d'Ivoire (0.44 ± 0.24 g N kg^{-1} soil vs. 1.09 ± 0.46 g N kg^{-1} soil). Median

total N concentrations were almost identical for Liliyo and Tiéningboué ($1.1\,g\,N\,kg^{-1}$). Total S concentrations varied between $41\,mg\,S\,kg^{-1}$ to $242\,mg\,S\,kg^{-1}$ across all fields, and showed a similar pattern as total C and N. The yam fields in the landscapes of Bukina Faso had on average more than two times higher total S than the other landscapes. Total P concentrations were in a similar range for the landscapes Léo, Midebdo, and Liliyo. In Tiéningboué, total P values were on average almost two times higher than the other fields ($817\,mg\,S\,kg^{-1}$ vs. $453\,mg\,S\,kg^{-1}$), with more within-landscape variation.

Total Fe, total Al, total Ca, total Zn, and total Cu concentrations in the soil tended to be higher for the landscapes in Côte d'Ivoire than in Burkina Faso (Figure 2.1). To give an example, median concentrations of total Ca were $2.16\,g\,Ca\,kg^{-1}$ in fields sampled from the Tiéningboué region, and similar in Liliyo (= $1.90\,g\,Ca\,kg^{-1}$), while they were markedly lower in Léo and Midebdo (= 0.90 vs. $1.26\,g\,Ca\,kg^{-1}$). In general, the ranges for total micronutrient contents were more variable in the landscapes of Côte d'Ivoire (e.g., range = $14-57\,mg\,Zn\,kg^{-1}$ in Liliyo; lowest range in Léo = $12.2-19.7\,mg\,Zn\,kg^{-1}$). Total K concentration was highly variable within and across the landscapes (overall range = $0.5-34.1\,g\,K\,kg^{-1}$), and lowest in Midebdo (range = $0.9-8.9\,g\,K\,kg^{-1}$), while the highest total K median was measured for yam fields in Léo (range = $4.1-25.0\,g\,K\,kg^{-1}$).

Soil resin-extractable P concentrations varied between $0.8\,mg\,P\,kg^{-1}$ to $33.1\,mg\,P\,kg^{-1}$. In Tiéningboué, resin-extractable P was on average higher than in soils of the other landscapes (Figure 2.1). Median extractable Fe and its interquartile ranges were comparable across the landscapes (see Figure 2.1). However, there were some fields where extractable Fe reached values higher than $100\,mg\,Fe\,kg^{-1}$. Median extractable Zn values showed a similar pattern as total C, with the highest median values and interquartile range in Tiéningboué and had the lowest in Léo. In comparison, the highest median values and interquartile range of extractable Cu and Mn were found in Liliyo. For extractable Zn, Cu, and Mn median values and interquartile range were higher in the two landscapes in Côte d'Ivoire than the two landscapes in Burkina Faso.

Across all samples and landscapes, soil pH varied between 4.7 and 8.4. Median pH was comparable in Tiéningboué (i.e., 6.4), Liliyo (i.e., 6.5), and Midebdo (i.e., 6.5). Median pH of yam fields in Léo (i.e., 6) was lower than in the other landscapes. Exchangeable K, Ca, and Mg concentrations showed similar patterns across the four landscapes. In Burkina Faso, each of the exchangeable cations showed relatively low median concentrations across the fields and less landscape-level variation than in Côte d'Ivoire. In general, the highest median and variation of exchangeable cations among the landscapes were measured for the yam field soils in Tiéningboué. Median exchangeable Al values were comparable among the landscapes, although there were some outliers with exchangeable Al > $20\,mg\,kg^{-1}$ for Midebdo, Liliyo, and Tiéningboué. The CEC_{eff}

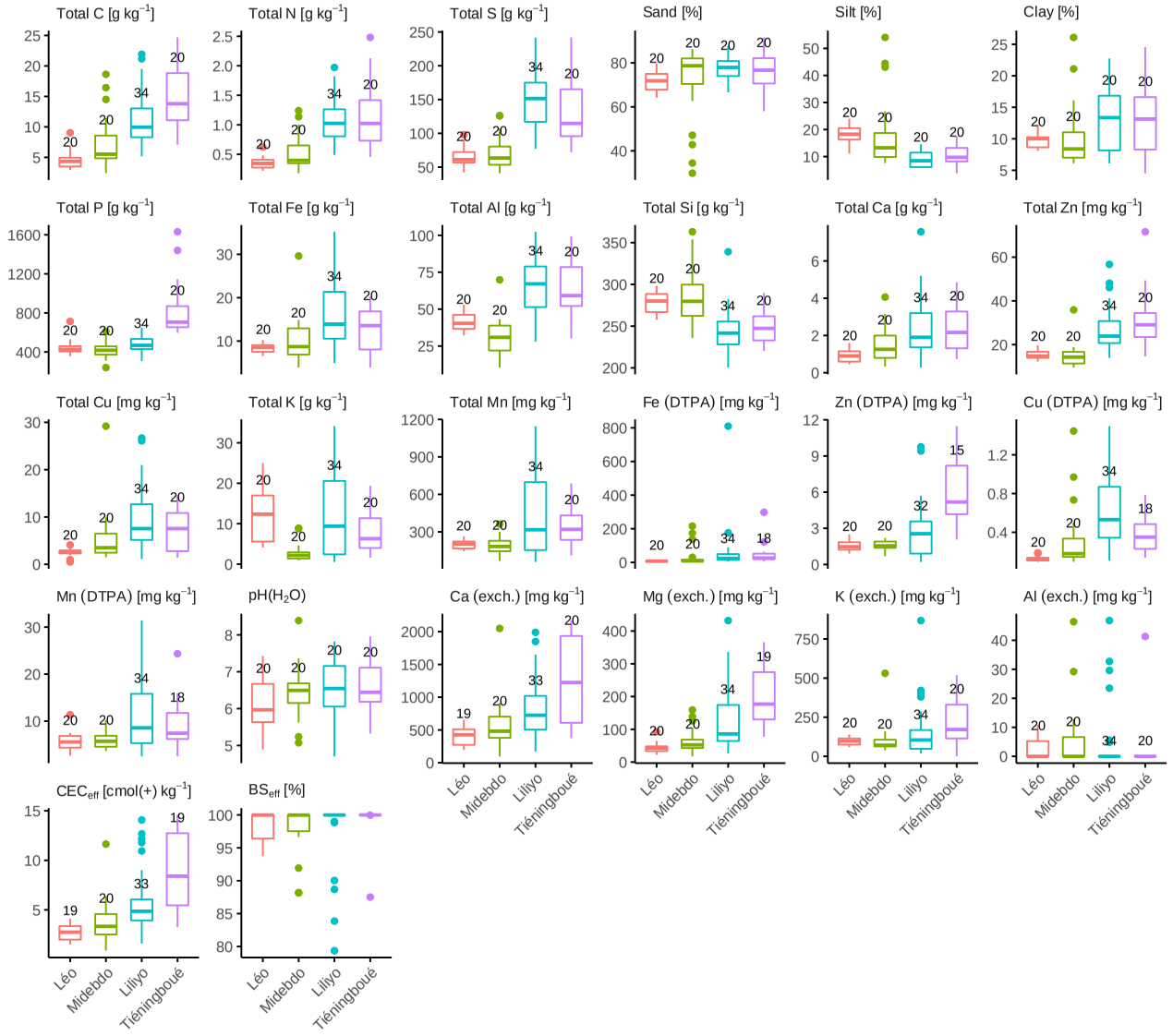

Figure 2.1: Reference measurements of soil chemical properties. Léo and Midebdo are two yam-growing regions in Burkina Faso, and Liliyo and Tiéningboué are in Côte d'Ivoire. The chemically analyzed soils (n = 94) originated from 20 yam fields per landscape and 14 additional soils from the Liliyo region were provided by the World Agroforestry Center (ICRAF). C = carbon, N = nitrogen, P = phosphorus, Fe is iron, Al is aluminum, Si is silicon, Ca is calcium, Zn is zinc, Cu is copper, K is potassium, Mn is manganese. Bioavailable micronutrients were measured by diethylenetriaminepentaacetic acid (DTPA extraction). Ca(exch.), Mg(exch.), K(exch.), and Al(exch.) signify exchangeable elements determined with $BaCl_2$ extraction. CEC_{eff} is the effective cation exchange capacity, BS_{eff} = effective base saturation. The number of soils analyzed for each individual property is indicated above the 75 % percentile.

ranged from 0.9cmol(+) kg^{-1} to 14.6cmol(+) kg^{-1} across all fields and landscapes. Median CEC_{eff} tended to decreases in the following order across landscapes: Léo > Midebdo > Liliyo > Tiéningboué. The interquartile range of CEC_{eff} was also the greatest in Tiéningboué and the smallest in Léo.

Reference measurements for total N, S, exchangeable Ca, exchangeable Mg and $CEC_{eff.}$ were highly correlated to total C (Figure 2.2;ă0.71 ≤ r ≤ 0.92 ($CEC_{eff.}$)). Also, total Ca, Al, and clay content correlated strongly to total C ($r > 0.70$). Clay contents were weakly related to silt ($r = 0.21$), while sand had a markedly negative relationship to silt ($r = -0.89$). Bioavailable Cu and Zn . Resin-extractable P was moderately correlated to total C ($r = 0.53$) and pH ($r = 0.38$). Bioavailable Zn (DTPA) was co-varying with both $CEC_{eff.}$ ($r = 0.58$) and total Zn ($r = 0.59$). Bioavailable Cu (DTPA) had a strongly positive association to total Cu ($r = 0.90$). Exchangeable K ($BaCl_2$) had the strongest relationship to total C and $CEC_{eff.}$ ($r = 0.63$, and $r = 0.64$).

Soil mid-IR spectroscopic models

Among the measured soil properties, mid-IR PLSR models for total K ($R^2 = 0.96$) and total Al ($R^2 = 0.97$) performed best (Table 2.1). Out of a total of 27 soil attributes, 11 were well quantified by the models when considering categorization judged upon on an $R^2_{cv} \geqslant 0.75$ criterion (Figure 2.3). The confidence intervals derived from cross-validation prediction were very narrow, showing all PLSR models were stable. Within this group of stable models, four soil attributes are directly related to the mineralogy (total Fe, Al, K and Ca), three are related to soil organic matter (total C, N and S), one is related to texture (clay fraction), one is related to plant nutrition (exchangeable Fe), and two are related to mineralogy and plant nutrition (exchangeable Ca and CEC_{eff}). More specifically, total C was accurately predicted, with an R^2 of 0.92 and a RMSE of 1.6g C kg^{-1}. The models were also able to predict total N well ($R^2 = 0.89$; RMSE = 0.16g N kg^{-1}). Prediction accuracy of total S was slightly lower than for total C, but its goodness-of-fit and RMSE suggest that the model was reliable for prediction. However, exchangeable K ($R^2 = 0.28$) and BS_{eff} ($R^2 = 0.24$) were poorly predicted (Table 2.1). Predictions for percent clay were reliable ($R^2 = 0.81$; RMSE = 2.1%), whereas predictions for percent sand ($R^2 = 0.45$; RMSE = 8.1%) and percent silt ($R^2 = 0.41$; RMSE = 6.5%) were not accurate. Finally chosen models of all soil attributes had between one and nine PLSR components.

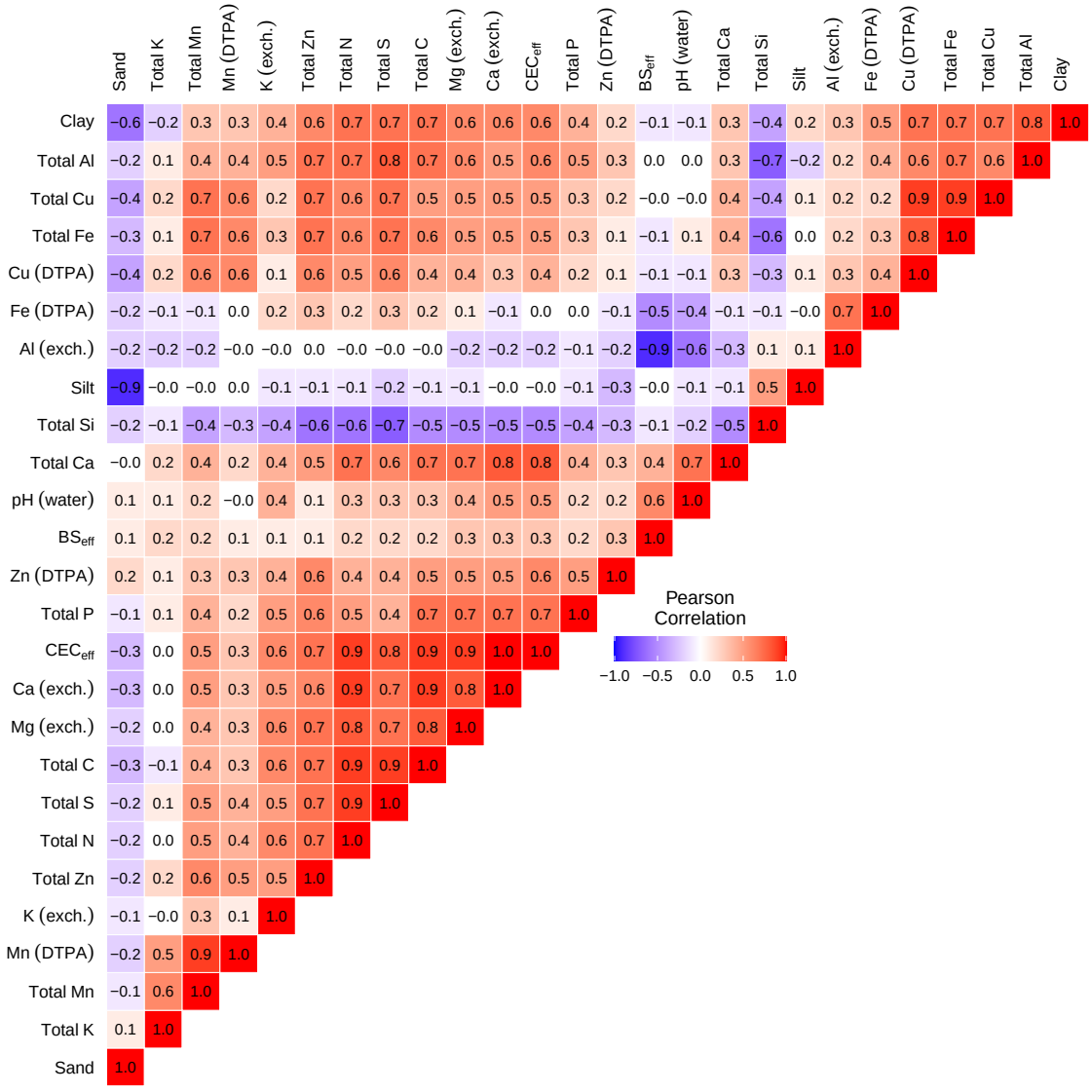

Figure 2.2: Correlation matrix of soil properties measured on each 20 soils sampled from individual yam fields per landscape, and 14 additional agricultural soils received from the World Agroforestry Center ($n = 94$; see Figure 2.1 for further details and abbreviated chemical properties). Pearson correlation coefficients (r) were rounded to 1 digit.

Table 2.1: Descriptive summary of measured (meas.) soil reference data (shown in Figure 2.1) and evaluation results of cross-validated PLSR models. All samples across the four landscapes were aggregated into a single model per respective soil property. Model evaluation was done on held-out predictions of 10-fold cross-validation (cv) repeated five times at the finally selected number of PLSR components (ncomp). CV is the coefficient of variation, RMSE is the root mean square error, RPD is the ratio of performance to deviation. C is carbon, N is nitrogen, P is phosphorus, Fe is iron, Al is aluminum, Si is silicon, Ca is calcium, Zn is zinc, Cu is copper, K is potassium, Mn is manganese. Bioavailable micronutrients were measured by diethylenetriaminepentaacetic acid (DTPA) extraction. Ca(exch.), Mg(exch.), K(exch.), and Al(exch.) signify exchangeable elements determined with $BaCl_2$ extraction. CEC_{eff} is the effective cation exchange capacity, BS_{eff} is the effective base saturation.

Soil attribute	n	$Min_{meas.}$	$Max_{meas.}$	$Med_{meas.}$	$Mean_{meas.}$	$CV_{meas.}$	ncomp	$RMSE_{cv}$	$R^2{}_{cv}$	RPD_{cv}
Total C [g kg^{-1}]	94	2.4	24.7	8.5	9.9	58	6	1.6	0.92	3.6
Total N [g kg^{-1}]	94	0.18	2.48	0.72	0.81	61	6	0.16	0.89	3.0
Total S [mg kg^{-1}]	94	41	242	99	111	46	2	20	0.85	2.6
Sand [%]	80	29.8	91.6	75.6	74.2	14	2	8.1	0.42	1.3
Silt [%]	80	3.9	54.1	12.0	14.1	60	2	6.5	0.41	1.3
Clay [%]	80	4.5	26.1	10.1	11.6	42	2	2.1	0.81	2.3
Total P [mg kg^{-1}]	94	240	1631	467	530	40	3	131	0.61	1.6
Total Fe [g kg^{-1}]	94	4	35	10	12	54	5	3	0.81	2.3
Total Al [g kg^{-1}]	94	10	102	48	53	42	5	4	0.97	6.0
Total Si [g kg^{-1}]	94	200	363	262	262	12	3	20	0.59	1.6
Total Ca [g kg^{-1}]	94	0.3	7.6	1.4	1.9	70	5	0.6	0.78	2.2
Total Zn [mg kg^{-1}]	94	9.5	71.6	19.1	22.6	49	1	6.7	0.63	1.7
Total Cu [mg kg^{-1}]	94	0.5	29.2	4.7	6.8	87	7	3.2	0.71	1.9
Total K [g kg^{-1}]	94	0.5	34.1	5.8	9.5	91	7	1.7	0.96	5.1
Total Mn [mg kg^{-1}]	94	59.2	1146.0	221.5	308.0	74	5	116.4	0.74	2.0
log(P resin) [mg kg^{-1}]	92	-0.2	3.5	1.4	1.4	57	2	0.6	0.40	1.3
log(Fe(DTPA)) [mg kg^{-1}]	92	1.0	6.7	2.7	2.9	38	9	0.5	0.77	2.0
Zn (DTPA) [mg kg^{-1}]	87	0.2	11.5	1.9	2.8	89	3	2.1	0.25	1.1
Cu (DTPA) [mg kg^{-1}]	92	0.1	1.5	0.2	0.4	89	6	0.2	0.74	2.0
Mn (DTPA) [mg kg^{-1}]	92	2.5	31.4	6.5	8.6	69	3	4.0	0.55	1.5
pH_{H_20}	80	4.7	8.4	6.4	6.4	11	8	0.5	0.61	1.6
Ca (exch.) [mg kg^{-1}]	92	98	2170	604	774	70	5	237	0.81	2.3
Mg (exch.) [mg kg^{-1}]	93	18	432	76	113	84	3	58	0.62	1.6
K (exch.) [mg kg^{-1}]	94	0	868	104	145	95	1	120	0.28	1.2
Al (exch.) [mg kg^{-1}]	94	0	47	0	4	258	2	9	0.21	1.1
CEC_{eff} [cmol(+) kg^{-1}]	91	0.9	14.6	4.2	5.3	67	6	1.4	0.84	2.5
BS_{eff} [%]	91	79	100	100	98	4	2	3	0.24	1.1

Model interpretation

A large proportion of absorptions had VIP > 1 for each the total C, total N and clay models (Figure 2.4). Important wavenumbers (VIP > 1) for total C were mostly between 3140cm^{-1} and 1230cm^{-1}. Besides clear absorption peaks, there were relatively continuous spectral features that were important to the models. For example, the relatively continuous and smooth spectral region between the alkyl C–H vibrations at 2855cm^{-1} and 2362cm^{-1} had a comparable contribution to the model as peak regions associated with total C prediction. The VIP patterns across wavenumbers were almost identical for total C and N models, and its reference measurements were strongly correlated (r = 0.94; Figure 2.2). In contrast, the clay content model deviated from the total C model in particular regions, for example around the kaolinite OH^- feature at 3620cm^{-1} or at kaolinite Al–O–H vibrations at 934cm^{-1} and 914cm^{-1}.

2.4 Discussion

Accuracy of mid-IR spectroscopy for agronomic diagnostics

Timely and accurate estimates of multiple soil properties are required to better understand and predict soil constraints across the yam belt in West Africa. The soil spectral library from our study, which includes four landscapes of the yam belt, can be practical to diagnose and monitor (and eventually manage) soil fertility that is considered to be low and therefore being a major constraint for yam production in West Africa. Specifically, our results show that properties closely related to organic matter — total amount of C, (micro)-nutrients, and exchangeable cations — can be accurately estimated using mid-IR spectra and in the selected yam growing landscapes (Figure 2.3). Soil organic matter plays a crucial role during vegetative growth and tuber formation phases of yam, as it guarantees among many other functions the storage and availability of essential nutrients needed for yam and tuber growth throughout the season, and as well prevents soil erosion due to it's structural stabilization capacity. Fertilizers are becoming more essential to replenish mineral nutrients for prolonged cropping; however, soil organic matter is at high risk of depletion in the regions because of the increasing land use frequencies and shorter fallows to restore the soil organic C pools. Thus, when testing sustainable soil and crop management options, for example to derive region-specific and farm-adapted nutrient management strategies, validated quantitative statements of soil organic carbon concentrations are of huge importance.

Fertilizers are becoming more essential to replenish mineral nutrients for prolonged cropping. Nevertheless, soil organic matter is at high risk of depletion in the these regions because of the increasing land use frequencies and shorter 20 fallows to restore

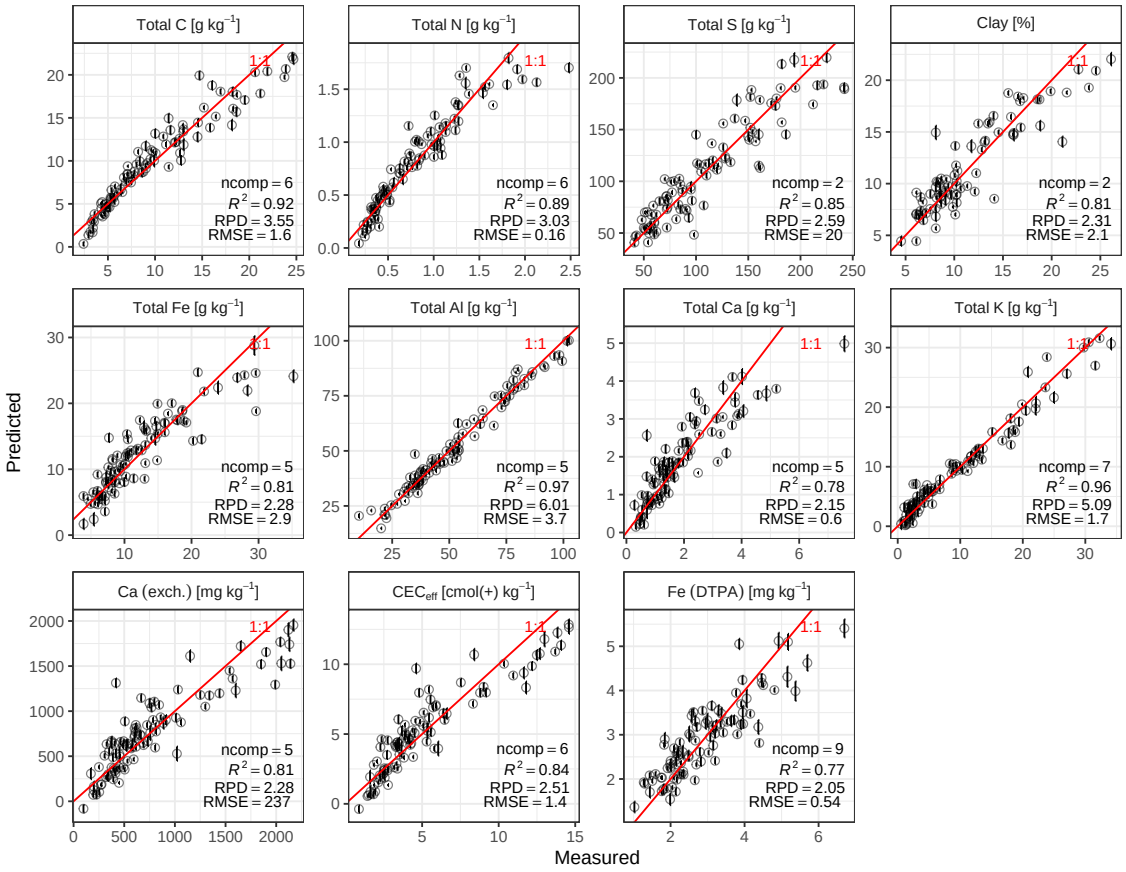

Figure 2.3: Cross-validated predictions of soil properties (y axis) derived from best mid-IR partial least squares regression (PLSR) models vs. laboratory reference measurements (x axis; see Figure 2.1). Average estimates, their confidence intervals (error bars), and evaluation metrics were derived with 5 × repeated 10-fold cross-validation. ncomp is the number of PLSR components of most accurate final models, RSME is the root mean square error, RPD is the ratio of performance to deviation. Only soil properties modeled with $R^2 > 0.75$ are shown. CEC_{eff} is the effective cation exchange capacity. Exchangeable (exch.) elements were determined with $BaCl_2$. Bioavailable Fe was determined diethylenetriaminepentaacetic acid (DTPA) extraction.

the soil organic C pools. While it is pivotal to develop innovative crop and soil management solutions to this problem (O'Sullivan and Jenner, 2006; Frossard et al., 2017; Kiba et al., 2020), it is also crucial to perform a separate but complementary activity to give feedback on potential soil changes: developing and applying soil conventional and proximal sensing methods. When testing sustainable soil and crop management options, for example to derive region- specific and farm-adapted nutrient management strategies, putting both validated quantitative statements on the status 30 of soil organic carbon and local farmers soil knowledge into the equation is crucial (Wawire et al., 2021). Inevitably, both determining the inherent soil status (i.e., soil texture and organic carbon) and measuring the chemical and physical environment that regulates nutrient availability at trial sites (e.g., pH) , is of agronomic and environmental importance (Foster, 1981). Maintaining and improving soil quality attributes will be paramount to sustaining soils ecosystem functions and crop yields over time. Activities to maintain and improve soil properties can for example be oriented towards fostering nutrient recycling.

Quick and reasonably accurate soil estimates derived from mid-IR spectra and empiric models as for example outlined in this study can inform the site-adapted timing, placing and form of nutrient supply based on local soil conditions. Specifically, to give an example, light-textured soils can attain high tuber yields but at a high risk of losing large proportions of applied N and K (e.g., O'Sullivan and Australian Centre for International Agricultural Research, 2010) — which are both demanded in relatively large quantities by yam — to the environment (e.g., Diby et al., 2011). Therefore, spectral estimates of texture can give an indication that applying larger amounts of N and K at once would 20 not improve yield potential under such situations. Hence, more frequent and local mineral applications of these nutrients after crop emergence, eventually combined with organic mulch, could improve the fertilizer efficiency and mitigate negative environmental impacts under these soil conditions. To estimate the availability of specific (micro)nutrients, however, more efforts need to be made to measure them at fine temporal and spatial resolution.

The mid-IR model accurately estimated C (RMSE = 1.6g kg^{-1}; Table 2.1; Figure 2.3). Mostly, only field-scale spectroscopic models achieve such accuracy (Nocita et al., 2015a; Guerrero et al., 2016), whereas the predictive accuracy reported for larger scale application of spectroscopic models is lower than for our model (Viscarra Rossel and Webster, 2012; Stevens et al., 2013; Sila et al., 2016). Models covering a wide geographical range of soils often result in high prediction errors (Stenberg and Rossel, 2010). Despite different soil types and climate regimes across a wide geographic spacing between the calibration fields, we achieved an accurate spectroscopic estimation of total C. The model was also able to reliably estimate a range of other important soil properties than to-

tal C. Specifically, other soil variables eligible for a mid-IR quantification include total N, total S, total Ca, total K, total Al, exchangeable Ca, Fe DTPA, $CEC_{eff.}$, and clay content ($R^2 > 0.75$). The high correlations of total C to N, S, exchangeable Ca, exchangeable Mg, $CEC_{eff.}$, total Ca, Al, and clay content (Figure 2.2) are consistent with Johnson et al. (2019), who reported very similar associations of clay content and exchangeable cations (Ca, Mg, K) as well as $CEC_{eff.}$ in soils from rice fields ($0.54 \leq r \leq 0.65$) — nevertheless they spectrally modeled a considerable soil variability (20 countries in sub-Saharan Africa; 42 study sites) and a larger sample size ($n = 285$) using PLS regression. At the same time, the measured range and the error in spectral estimates of CEC were larger compared to ours (RMSE = 6.7 cmol(+) kg^{-1} vs. 1.4 cmol(+) kg^{-1}; range = 1.9–66.5 cmol(+) kg^{-1} vs. 0.9–14.6 cmol(+) kg). Even though, total K and Fe(DTPA) were poorly correlated to total C, their spectroscopic estimates were relatively accurate. This suggests that the mid-IR prediction of other soil properties is largely based on their correlation with total C as well as other absorption features of many organic and mineral soil components having a specific IR adsorption.

We also found reasonable prediction accuracy for Cu(DTPA) ($R^2 = 0.74$) and Mn(DTPA) ($R^2 = 0.55$), despite that soil nutrients that are extraction-based or dependent on surface chemistry usually have variable predictive performance (Janik et al., 1998). Since relationships between soil composition and soil matrix exchange processes are typically complex, some properties may not be represented in the models in a straightforward manner (Janik et al., 1998; Nocita et al., 2015a).

Although total elements are not necessarily a direct proxy for plant-available nutrients – with the exception of total C from organic matter — they can be related to mineralogical status, which is influenced by weathering and nutrient supply. For example, total Fe from iron oxides can be important in controlling the availability of P (Parfitt et al., 1975), and total P can be correlated to available P in other cases. For yam which is an understudied crop with a relatively large yield gap fertilizer response to N, P, and K are often absent on soils that have been under long fallow periods (O'Sullivan and Australian Centre for International Agricultural Research, 2010). Even more importantly, the number of thoroughly conduced yam fertilizer trials in a region and for distinct soil types is not sufficient for the specific calibration of soil tests with regard to fertilizer response and recommendations (OSullivan and Jenner, 2006)

Interpretation of spectral features

All mid-IR spectra that we measured for soils in the four landscapes exhibited a similar pattern of absorbance (Figure 2.4). The O – Si – O absorptions in quartz at 1080 cm^{-1}, 800–780 cm^{-1} and 700 cm^{-1} were a prominent feature in the spectra due to relatively

high sand contents across the landscapes (range 30% to 92%, median 76%). Our spectra further had hydroxyl (OH) absorptions that are typical for kaolin minerals, at 3695 cm^{-1} (surface OH), 3620 cm^{-1} (inner OH), 914 cm^{-1} (inner OH), and 936 cm^{-1} (surface OH) (Madejová et al., 2002). The spectral pattern between the hydroxyl bands at 3695 cm^{-1} and 3620 cm^{-1} was relatively consistent and the intensity ratio of these flanking peaks was close to 1. This is typical for halloysite (0.8–0.9), while the ratio for kaolinite is often higher (1.2–1.5) and dickite lower (0.6–0.8) (Lyon and Tuddenham, 1960). The two weak intermediary stretching absorptions at around 3657 cm^{-1} and 3670 cm^{-1} indicate surface hydroxyls. Together with the absorption at 936 cm^{-1}, the spectra would suggest the presence of rather well-ordered prismatic halloysite (Hillier et al., 2016). This aligns well with the spectral patterns of soils that were assigned to the Halloysite archetype through similarity mapping (by comparison to the pure mineral spectra) by Sila et al. (2016). Our spectra confirm the presence of kaolin minerals, which reflects the advanced state of mineral weathering in these tropical soil types.

Our accurate predictions, which are comparable to field-scale calibrations, are most likely because of the relatively uniform mid-IR spectra we obtained for our samples. This suggests a relatively homogeneous soil chemical composition, particularly with regard to the mineralogy in the sampled soils. Still, the data set presented here is relatively small and no randomized spatial sampling strategy was used for selecting field locations. Therefore, we propose to the implementation of a spectroscopy-driven approach to diagnose soils in more yam-growing areas, as an effort to broaden the library to achieve better spatial coverage of soil variability.

2.5 Conclusions

We developed models with mid-IR spectra to estimate soil chemical and physical properties relevant to the production of yam and other staple crops in four landscapes in the yam belt of West Africa. We tested the models for the important soil properties that are applied widely for agronomic performance evaluation. We showed that mid-IR spectroscopy models have the potential for cost-effective and rapid determination of the distribution and variability of important soil properties across highly variable yam production landscapes in West Africa. Specifically, total C, total N, total S, total Fe, total Al, total K, total Ca, exchangeable Ca, CEC_{eff}, Fe(DTPA), and clay content can be quantified with RPD > 2 and $R^2 > 0.75$ when aiming to predict in the range of soil property values found in the environmental conditions covered by this study. We achieved spectral estimates with quite small uncertainties, that are typically reported for libraries at the scale of a field or farm. The correlation analysis of measured values together with spectral inference helps improve our understanding of how soil properties are interre-

lated with soil functional composition. This study delivered parsimonious, unbiased and accurate mid-IR spectroscopy-based models to monitor and predict soil quality and to manage crop nutrition. Hence, we envision this pilot study as being a starting point to continuously update and adapt the mid-IR model library for more efficient site-specific and agronomically relevant soil estimates in the West African yam belt. This can create a better capacity to diagnose and monitor soils in the long term compared with traditional wet chemistry and will hopefully ameliorate the soil conditions for sustainably meeting the demand of yam and other important staple crops in the regions.

Code and data availability. All data and code to reproduce the results of this publication are publicly available under GNU General Public License v3.0 and can be accessed via the Zenodo archive and the corresponding GitHub public repository (Baumann, 2020)

Author contributions. Philipp Baumann carried out the research and analysis (soil sampling, sample preparation, soil chemical analysis, infrared spectroscopy, statistical modeling) under continuous support of the YAMSYS project team, and took the lead in writing the manuscript. All co-authors helped to improve the manuscript. Johan Six, Juhwan Lee and Emmanuel Frossard framed the idea of delivering validated models for soil properties relevant for yam growth in the four pilot regions in Burkina Faso and Côte d'Ivoire. Valérie Hgaza and Delwende Kiba contributed to the selection of representative yam fields that were sampled for our work.

Acknowledgments. This study has been done within the YAMSYS project (www.yamsys.org) funded by the food security module of the Swiss Programme for Research on Global Issues for Development (r4d programme; www.r4d.ch) (SNF project no. 400540_152017 / 1). We would like to express gratitude to the site managers — Marie Leance Kouassi, Marcel Soma, and Augustin Kangah Nda — and the 80 farmers that strongly supported our sampling endeavor and the idea of developing diagnostic and management innovations for improved yam growth. We would like to thank Nestor Pouya and Carole Werdenberg who assisted with soil sampling. Our thanks also go to Dr. Bahar Aciksöz, Michele Wyler, and Patricia Schwitter, who helped with sample milling, weighting, and acid digestion. We would also like to thank Dr. Federica Tamburini for support with the CNS analysis and Björn Studer for the opportunity to perform XRF analyses on soils. We thank Raphael Viscarra Rossel and Marijn Van de Broek for their valuable feedback, which helped us to improve the manuscript.

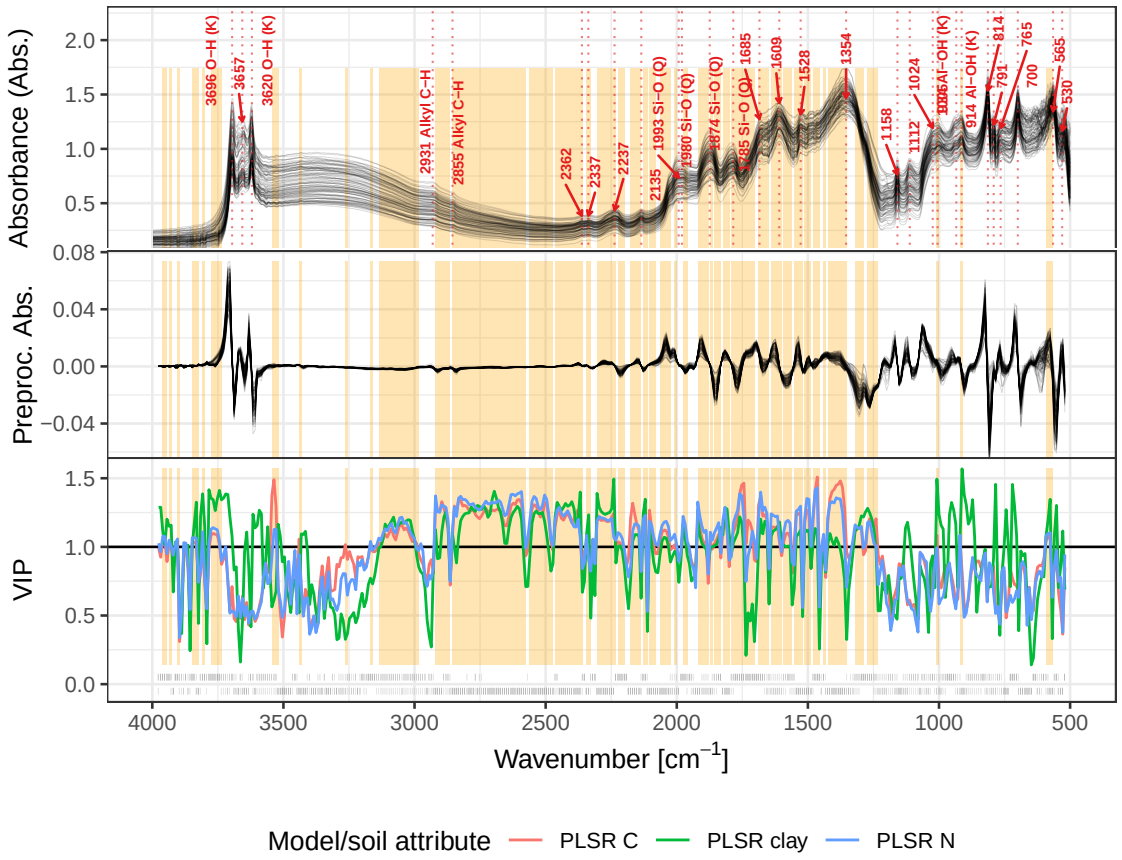

Figure 2.4: Variable importance analysis of partial least squares regression (PLSR) models for total soil C, total N and % clay , including overlaid raw and preprocessed spectra. Top panel shows resampled mean sample absorbance spectra ($n = 94$). Prominent peaks were identified as local maxima with a span of 10 points 20 cm^{-1}) for the selected wavenumbers. Fundamental mid-IR vibrations that are well described in the literature (e.g., Madejová et al., 2002; Rossel et al., 2010; Stevens et al., 2013) were added as labels when identified peaks matched literature assignments. (Q) stands for quartz and (K) for kaolinite. The middle panel depicts preprocessed spectra (Savitzky-Golay first derivative with a window size of 21 points (42 cm^{-1}); 3rd order polynomial fit). The bottom panel shows variable importance in the projection (VIP) for three selected well performing PLSR models (total C, total N and % clay; $R^2 > 0.81$). The black horizontal line at VIP = 1 indicates the threshold above where absorbance at the wavenumbers explain more than average to the prediction of a certain soil property. Dashed points closely below the $y = 0$ line of the VIP graph visualize positive (above $y = 0$) and negative (below $y = 0$) PLSR β coefficients.

Part 3

Chapter 2: Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring

Philipp Baumann[1,*], Anatol Helfenstein[1,2], Andreas Gubler[3], Armin Keller[4], Reto Giulio Meuli[3], Daniel Wächter[3], Juhwan Lee[5], Raphael Viscarra Rossel[6], and Johan Six[1]

[1]Group of Sustainable Agroecosystems, Institute of Agricultural Sciences, ETH Zurich, 8092 Zürich, Switzerland
[2]Soil Geography and Landscape Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands
[3]Swiss Soil Monitoring Network (NABO), Agroscope, Reckenholzstrasse 191, 8046 Zürich, Switzerland
[4]Swiss Competence Center for Soils (KOBO), School of Agricultural, Forest and Food Sciences HAFL, Bern University of Applied Sciences BFH, Zollikofen, Switzerland
[5]Department of Smart Agro-industry, Gyeongsang National University, Jinju, 52725, Republic of Korea
[6]Soil and Landscape Science, School of Molecular and Life Sciences, Curtin University, GPO Box U1987, Perth WA 6845, Australia

[*]Corresponding author Email: baumann-philipp@protonmail.com

This chapter is a reprint of the paper published in SOIL, 7, 525546, 2021,
https://doi.org/10.5194/soil-7-525-2021

Abstract

Information on soils' composition and physical, chemical and biological properties is paramount to elucidate agroecosystem functioning in space and over time. For this purpose we developed a national Swiss soil spectral library (SSL; n = 4374) in the mid-infrared (mid-IR), calibrating 16 properties from legacy measurements on soils from the Swiss Biodiversity Monitoring Program (BDM; n = 3778; 1094 sites) and the Swiss long-term Soil Monitoring Network (NABO; n = 596; 71 sites). General models were trained with the interpretable rule-based learner CUBIST, testing combinations of $\{5, 10, 20, 50, and\, 100\}$ ensembles of rules (committees) and $\{2, 5, 7, and\, 9\}$ nearest-neighbors used for local averaging with repeated 10-fold cross-validation grouped by location. To evaluate the information in spectra to facilitate long-term soil monitoring at a plot level, we conducted 71 model transfers for the NABO sites to induce locally relevant information from the SSL, using the data-driven sample selection method RS-LOCAL. In total, 10 soil properties were estimated with discrimination capacity suitable for screening ($R^2 \geq 0.72$; ratio of performance to interquartile distance (RPIQ) ≥ 2.0), out of which total carbon (C), organic C (OC), total nitrogen (N), pH, and clay showed accuracy eligible for accurate diagnostics ($R^2 > 0.8$; RPIQ ≥ 3.0). CUBIST and the spectra estimated total C accurately with the root mean square error (RMSE) = 8.4 g kg^{-1} and the RPIQ = 4.3 while the measured range was 1 – 583 g kg^{-1} and OC with RMSE = 9.3 g kg^{-1} and RPIQ = 3.4 (measured range 0 – 583 g kg^{-1}). Compared to the general statistical learning approach, the local transfer approach — using two respective training samples — on average reduced the RMSE of total C per site fourfold. We found that the selected SSL subsets were highly dissimilar compared to validation samples, in terms of both their spectral input space and the measured values. This suggests that data-driven selection with RS-LOCAL leverages chemical diversity in composition rather than similarity. Our results suggest that mid-IR soil estimates were sufficiently accurate to support many soil applications that require a large volume of input data, such as precision agriculture, soil C accounting and monitoring and digital soil mapping. This SSL can be updated continuously, for example, with samples from deeper profiles and organic soils, so that the measurement of key soil properties becomes even more accurate and efficient in the near future.

3.1 Introduction

Soils provide a manifold of functions within terrestrial ecosystems, many of which are vital for humankind. To quantify these functions from the soils' composition and properties, one typically relies on physical, chemical and biological laboratory analytical measurements. Doing this consumes both financial resources and time. For exam-

ple, repeated measurements are needed to describe soil functioning in changing environments, for example in response to agronomic management. Soil visible (vis) and infrared (IR) spectroscopic measurements and modeling have become indispensable tools to gather quick, relatively accurate, and inexpensive estimates of soil properties, both on the field and in the laboratory (Nocita et al., 2015a; Viscarra Rossel et al., 2016, 2017). Once soil chemical and physical properties are calibrated to the spectra, a single mid-IR (mid-infrared; 4000 – 500 cm^{-1}; 2500 – 25000 nm) or vis–NIR (visibe near infrared; 25000 – 4000 cm^{-1}; 400 – 2500 nm) measurement can be used to estimate multiple soil properties of new samples. Soil is a complex matrix with many organic and mineral components. This yields spectra with absorptions that overlap and contain many and often highly correlated variables. Hence, to successfully develop calibrations and make predictions for attributes related to soil composition on more samples, statistical learning methods are needed to find and use relationships between these variables and measured attributes. It is important to consider that the diversity in spectral characteristics typically reflects the soils' chemical and physical composition. Since the soil composition is influenced by the soil-forming factors — soil parent material, climate, topography, organisms and age of soils (Dokuchaev, 1899; Jenny, 1941) — these factors provide further means of causally interpreting and judging the applicability of the method for a particular set of soils. Compared to the NIR, mid-IR offers a more accurate characterization of soils' chemistry since this region contains the fundamental vibrations with more defined peaks (Janik et al., 1998; Viscarra Rossel et al., 2006).

A soil spectral library (SSL) can be defined as a well-ordered and harmonized collection of soil samples, their spectra, analytical reference measurements, contextual information, and additional metadata on samples and methods used. A central question behind the development of large SSLs is how to achieve accurate local predictions based on established collections of soil information — for example, within a new landscape, ecosystem, farm, field, or plot in a new region — where reference data of only a few observations are available. More recently, SSLs that span large geographical extents are being developed (Sila et al., 2016; Viscarra Rossel et al., 2016; Padarian et al., 2019b; England and Viscarra Rossel, 2018; Briedis et al., 2020; Angelopoulou et al., 2020; Dangal et al., 2019). These efforts are motivated by the prospect that soil spectroscopy can supplement many conventional methods of soil analysis. A range of predictive modeling strategies and algorithms have been tested for soil spectral analysis, among others involving tools from chemometrics (e.g., partial least squares, PLS, regression); Janik and Skjemstad, 1995), traditional machine learning (e.g., regression tree methods; Viscarra Rossel and Webster, 2012, to convolutional neural networks (CNNs; Padarian et al., 2019a,b; Tsakiridis et al., 2020).

There are two main strategies for estimating properties of new soils using spectra. The

first one is to calibrate one general or global model that is applied to predict new samples, and the other is to derive local calibrations by conditioning on a specific set of observations and features of the SSL to new data based on soil knowledge and/or algorithms. However, empirical evaluations of local and global methods are needed in different contexts where data on soil attributes are needed (i.e. soil studies and soil mapping projects). Such studies or applications should consider the "no-free-lunch" theorems for machine learning and optimization (Wolpert, 1996; Wolpert and Macready, 1997); i.e., there is no single algorithm combination that works best under all situations or applications.

General statistical learning makes use of all available training data to construct one parametric model. In contrast, local learning methods combine different learning methods, supervised and/or unsupervised, and, together with domain knowledge produce more modular forms of learning (Solomatine, 2008). The available training set can be a subset and algorithmic submodels, can, thereby be optimized to more accurately predict new single observations or groups of them. Local learning has also been termed *transfer learning*. Transfer learning is a general expression for adapting previous knowledge gained from existing data (i.e., model representation) for a new prediction case (Pratt et al., 1993; Pratt and Thrun, 1997; Thrun and Pratt, 1998). It has been defined as a transfer from knowledge in the source task(s) or domain(s)—here an SSL—to a target domain (Pan and Yang, 2010), and,thus comprises soils from new locations in this case.

The soil spectroscopy community has suggested several approaches to achieve local calibrations based on an established SSL and its feature space. One example is augmenting (spiking) SSLs with a few unweighted (Guerrero et al., 2010; Seidel et al., 2019) or extra-weighted (Guerrero et al., 2014, 2016) local samples. Other studies calibrated separate models on partitions of training data that were derived from applying certain criteria (i.e., geographical region, terrain attributes, parent material, soil type, land use, spectra-based clustering) (Sila et al., 2016; Ogen et al., 2019). Still others used memory-based learning based on spectral similarity, extracting useful information from compositional relatedness of soils (Ramirez-Lopez et al., 2013b; Clairotte et al., 2016; Hong et al., 2019; Dangal et al., 2019) or additionally geographic proximity (Tziolas et al., 2019). These all produce individual models for each sample to be predicted. Memory-based learning combines both *lazy learning*, where a subset of stored samples are only retrieved and modeled when new samples are predicted, and *local learning* principles, where modeled subsets define points within a local neighborhood (Dietterich et al., 1993). The spectrum-based learner developed by Ramirez-Lopez et al. (2013b) is a prominent memory-based method for which each new prediction sample forms its own target domain. The selection of source instances is governed by spectral similarity. Therefore, the spectrum-based learner is also considered a transfer learning method.

Another approach used by Padarian et al. (2019a) was retraining weights within specific layers of a deep CNN using local (target) sets, which were spectral soil data sets per country (parameter transfer approach). Finally, the selection of matching SSL samples using the resampling-based selection RS-LOCAL algorithm has also been used (Lobsey et al., 2017). Lobsey et al. (2017) showed that this data-driven transfer approach outperforms most other current methods for deriving local estimates. Still, despite these promising learners, transferring the useful information contained within large and diverse SSLs, and their resulting calibrations onto new, local target areas with unique soil characteristics remains challenging due to soil complexity.

RS-LOCAL obtains locally-relevant information by selecting specific rows (instances) from the training set and transferring them to the prediction set. RS-LOCAL is an example of an instance or sample transfer approach. It heavily relies on sampling and performance-driven reduction of the library, yielding a subset of samples that can accurately estimate the properties of soils in the local target task. We wanted to investigate this promising new method for local soil estimation and monitoring in Switzerland because it makes no prior assumptions on which samples from the library could be useful. This makes it potentially more accurate and also more flexible to new local soil contexts than when creating constraints with similarity measures. A further advantage for large SSLs is that it removes samples that might be spectrally similar but cause inaccurate calibrations (i.e., erroneous measurements or spectra with confounding effects). Such a local approach, however, requires an well-established and sufficiently diverse SSL in order to extract useful soils that are locally relevant.

Thus, our first goal was to develop a national mid-IR SSL with reference measurements for Switzerland to deliver 16 key chemical and physical soil proxies. This SSL includes soils and their analysis data from the long-term Swiss Soil Monitoring Network (NABO; 71 agricultural sites with time series measurements, $n = 596$) and the Swiss Biodiversity Monitoring (BDM) network (1094 grid locations, $n = 3778$). This is the first operational SSL for Switzerland in the mid-IR that allows means for spectral estimation with sufficient existing soil diversity. The second goal was to develop general rule-based models for all available soil properties using the CUBIST algorithm. Furthermore, we wanted to infer important spectral regions in the models and their chemical associations, which we illustrated with the estimation of total carbon (C) contents.

For soil monitoring and also for determining C stocks, it is crucial to obtain locally accurate spectral estimates of key soil properties such as organic C contents, from high soil variability in large SSLs and over time. This was our motivation to design a predictive transfer workflow that was adaptive to soils' composition and properties for each long-term monitoring site. Hence, our third goal was to leverage the SSL with its spa-

tial and temporal variability in soils to derive local calibrations by transfer learning with RS-LOCAL. Specifically, we aimed at showing local models' capacity to reproduce time series measurements (starting from 1985) of soil C at the Swiss agricultural long-term monitoring sites based on spectral analyses and two calibration samples per site. To the best of our knowledge, there is no study yet that has evaluated the usefulness of a large and diverse SSL for systematic soil monitoring. We, therefore, wanted to design a local calibration strategy using transfer learning, that would be effective in reducing (conditional) errors at monitoring plots compared to the general rules derived in the first aim. Furthermore, we had a strong interest in identifying the mechanisms, considering both soil knowledge and data distributions, of how such a local transfer would induce locally adaptive soil estimation.

In brief, our work addresses the following three objectives: (1) developing a national SSL, (2) building general prediction models using CUBIST, and (3) building site-specific (local) prediction models using RS-LOCAL.

3.2 Material and methods

Soils and data sets

To establish the Swiss SSL, we obtained soil samples and reference data from two different sources: 1) the Swiss Soil Monitoring network (NABO), and 2) the Swiss Biodiversity Monitoring (BDM) program (BAFU, 2014; Fig. 3.1). The NABO currently consists of 108 sites where soils have been continuously measured every 5 years since 1985 for long-term soil monitoring. Out of the 108 sites, we selected 71 sites under agricultural management—comprising arable land (33 sites), permanent grassland (26 sites) and special crops (11 sites; horticulture)—and one protected area. For the mid-IR SSL, we used 596 NABO soil samples from six campaigns conducted between 1985 and 2015.

The plots at the NABO sites covered 10 m× 10 m each. These were repeatedly sampled for 0–20 cm soil depth. In total, four replicate samples were taken by stratified random sampling and bulking 4 × 25 cores from 100 subareas of 1 m^2 to account for small-scale soil variability. Desaules et al. (2010) and Gubler et al. (2019) detailed the sample collection and data harmonization process of the measurements. The soils of the BDM were sampled at 0–0.2 m depth from positions on a regular grid of 6 km ×4 km laid over Switzerland (a total of 1094 locations). The points that were not sampled were inaccessible; these were mostly in the alpine regions. Each sampled location included four subsamples that were taken at the intersection of the four cardinal directions from the center point and the circumference of a circle with a radius of 3 m to 3.5 m (Meuli et al., 2017). Due to its design which covers all major geographic regions in Switzerland—the

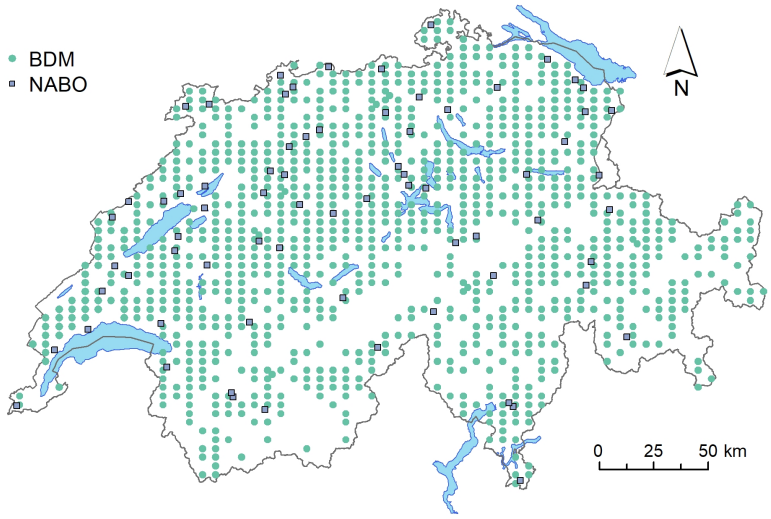

Figure 3.1: Swiss map with sampling locations of mid-infrared spectral library including the sites of the Biodiversity Monitoring Program (BDM; 6 × 4 km; n = 1094) and the National Soil Monitoring Network (NABO; n = 71). 71 NABO sites (10 m× 10 m) were sampled with a grid-based stratified design. 1094 BDM samples were obtained from single sampling events. The NABO sites have been continuously sampled and measured in five-year intervals since 1985.

Jura Mountains, the Central Plateau and the Alps—the BDM sampling campaign comprises a major part of the biogeochemical diversity of soils and predominant land use types in Switzerland. The wide coverage of soil conditions are an important source of soil chemical variability.

Chemical reference analysis

Data on chemical and physical soil properties were previously measured and provided by the NABO group. All laboratory soil analyses for the 16 properties were based on the protocols of the Swiss standard method (Agroscope, 1996). Mineral elements were determined by extraction with 1:10 ammonium acetate–EDTA solution (AAE10; method following, Agroscope, 1996). The measured properties were total C, organic C (OC), total nitrogen (N), pH ($CaCl_2$), $CaCO_3$, clay, silt, sand, CEC_{pot}, P(AAE), K(AAE), Ca(AAE), Mg(AAE), Cu(AAE), Zn(AAE), and Fe(AAE). For samples of BDM and for the more recent NABO sampling campaigns five and six (years 2009 – 2014), the total C and N measurements were done with dry combustion (LECO TruSpec). For campaigns one through four (years 1985 – 2014), the OC contents determined with wet oxidation using a modified Walkley–Black method were transformed into dry combustion equivalents, using site-specific robust linear regressions (complementary data of campaigns five and six; Gubler et al., 2018). Carbonates were determined by volumetric calcimetry using hydrochloric acid (HCl) for digestion. Organic C was obtained by difference in total C

and carbonate C when pH was greater than 6.5. Inorganic (carbonate) C was calculated with 0.12 × $CaCO_3$. The texture was determined by the pipette method. The pH was measured in $CaCl_2$ using a 1:2 volumetric ratio of soil to water. For CEC_{pot}, the exchangeable elements were extracted with a 0.05 N-0.025 N HCl-H_2SO_4 solution, which was buffered with triethanolamine for soil samples with pH > 5.9. All soil properties were referenced to dry weight by water correction after drying at 105 °C. All chemical analyses of NABO soils were done on four bulked replicates per site and sampling event. For BDM locations, four spatial replicates were measured each.

Measuring and processing spectra

All milled soil samples from the NABO and the BDM archive (n = 4374; with a particle size below 100 µm) were measured with the VERTEX 70v Fourier transform spectrometer from Bruker (Bruker Optik GmbH, Ettlingen, Germany) at ETH Zurich, using a high-throughput accessory (HTS-XT) and custom 24-well plates tailored to diffuse reflectance measurements. The mid-IR spectrometer was equipped with a KBr beam splitter and a mercury cadmium telluride (MCT) detector, which was permanently cooled with liquid nitrogen during the measurements. The reflectance spectra were acquired between 7500 cm^{-1} (1333.3 nm) and 600 cm^{-1} (16666.7 nm) at an effective resolution of 2 cm^{-1} and trimmed to the mid-IR range between 3996 cm^{-1} and 600 cm^{-1} before further processing (see below).

Each soil sample was measured twice. The soil surface was flattened evenly and without compression by the thin round middle part of the spatula. The first measurement position of the 24-well plate contained a gold (Au) reference surface, which produced a single reflectance spectrum for normalizing the reflectance of the 23 following soil measurements. The "atmospheric compensation" routine, implemented in the Bruker OPUS software was used to eliminate unwanted absorptions of H_2O vapor continuum and CO_2 gas in the measurement chamber, based on the single channel reference spectrum measured once on each plate. All single channel reflectance spectra were obtained by averaging 32 internal measurements.

The resulting reflectance spectra (R; background referenced) were converted to apparent absorbance (A) by $A = log_{10}(1/R)$. Then, an average spectrum per sample was produced by calculating the mean of all spectral variables for the measured replicates. Finally, the spectrum offset and further scatter effects were reduced and the features were transformed with a Savitzky–Golay (Savitzky and Golay, 1964) first derivative smoother using a window size of 35 variables (70 cm^{-1}) and third-order polynomial fit. Finally, we selected every eighth spectral variable to reduce redundancy in the spectra (collinearity) and produce more parsimonious spectral estimates of soil properties.

This resulted in 209 variables between 634 cm^{-1} and 3962 cm^{-1}, which formed the predictors for the subsequent general and local transfer modeling.

Data processing and statistical computing

All spectral and reference data were processed and modeled with the R software environment for statistical computing and graphics (version 3.6.0; R Core Team, 2019. We used the caret (**?**) R package to streamline the statistical learning process. Basic data transformations, such as data preparation and aggregation, were done using the tidyverse (Wickham, 2019) set of packages and data.table (Dowle and Srinivasan, 2019). The spectral data were handled and processed with the simplerspec (Baumann, 2019) and prospectr (Stevens and Ramirez-Lopez, 2013) packages.

General soil estimation: rules for the entire SSL

The general soil estimation was done by rules trained with the CUBIST (Quinlan, 1993) learner, separately developed for each analytical soil measure. We chose this algorithm since it has shown excellent performance for modeling soil information and developing SSLs with rather large soil variability and multicollinear spectral variables (Bui et al., 2006; Viscarra Rossel and Webster, 2012; Stevens et al., 2013; Miller et al., 2015; Peng et al., 2015; Viscarra Rossel et al., 2016; Dangal et al., 2019; Padarian et al., 2019b), and because its interpretation is mechanistically more intuitive as it is a form of data partitioning (simple conditions and linear equations). CUBIST first forms model trees using basic mechanisms of M5 (Quinlan and others, 1992). CUBIST is a form of a rule-based decision tree with piecewise linear models. Wang and Witten (1996) outlined detailed principles behind the construction of the model trees and derivation of rules, and Viscarra Rossel and Webster (2012) described it for soil spectroscopic modeling.

A CUBIST prediction rule is a unique set of conditions, i.e., "if, then" logical statements, together with the associated ordinary linear regression model. During training, the condensed regression equations are made for samples in the terminal nodes. All preceding split variables are potentially allowed for regression in a final node; however, some of them are pruned or combined in the rules. The smoothed regression equation with the selected variables allows one to predict an individual, new observation. CUBIST features two empirical parameters that can improve predictions, namely committees and neighbors. Committees are ensembles of rules that are created by successive construction of trees, which correct predictions of preceding rules and, thereby, lower predictive errors by averaging. When neighbors are used (maximum nine), a new training sample is predicted, using both unweighted or weighted averages of the measured values of the nearest neighbors, using all features in the training set and the prediction

of the new sample using the training rule(s).

Model development and validation

We tested a full-factorial combination of $\{5, 10, 20, 50, and 100\}$ committees of rules and $\{2, 5, 7, and 9\}$ neighbors to tune the CUBIST models. To obtain realistic estimates of the models' general performance, we defined a grouped ten-fold cross-validation scheme that treated the entire site (e.g., for total C: NABO –71 sites; BDM –1079 sites) as independent in the modeling data sets. This made all observations from a site the unit of prediction, making the procedure equivalent to external cross-validation.

To reduce the bias variance trade-off in the assessment, we repeated the grouped ten-fold cross-validation (CV) procedure five times (Friedman et al., 2008; Kuhn and Johnson, 2013). The division into training and validation proportions of the data was done in consistent and repeatable manner (pseudo random number generation). We considered this site grouping factor as prior information when cross-validation segments were created, so that samples from a particular site were only present within one segment (fold) of a cross-validation split. This grouped assignment prevented that the relationships were trained on the model fitting sets and prevented a particular site from leaking into the testing segments, yielding reliable generalization errors.

We tested the correspondence of mid-IR and model-derived predictions ($\hat{x_i}$) and measured standard reference measurements (x_i) with common regression metrics. We cross-validated the inaccuracy of the models with the root mean square error (RMSE). The mean squared error (MSE) was further decomposed into mean error (ME) or bias and the standard deviation of the error (SDE) or imprecision, so that $\text{RMSE}^2 = \text{ME}^2 + \text{SDE}^2$ (Viscarrra Rossel and McBratney, 1998). To describe the linear dependency between measurements and modeled values and give a relative goodness of fit, the coefficient of determination (R^2) from linear regression was also reported. All metrics were aggregated from five estimates from independent resampling repeats. We reported mean values and standard deviations to provide uncertainties of the estimates.

Deriving important spectral variables

The importance of each spectral variable was assessed based on its usage in the rule conditions and the model for CUBIST. We used the recursive feature elimination (RFE) method, a backwards variable-selection algorithm described by Guyon et al. (2002), to test whether the modeling can be simplified and to find most important spectral features. Soil reflectance spectra typically contain many correlated and potentially redundant variables. Therefore, constraining them to relevant subsets that feed into the modeling can further improve predictive accuracy and reduce computation time and

storage for model updates. We recursively eliminated subsets of variables with low CUBIST variable importance, calculated as the average relative usage frequencies of a particular variable in split conditions and regressions.This step-wise variable reduction was based on the following predefined subset sizes S_i, starting with the full set at $i = 1$ and ending with the most important predictor at $i = 30$:

$$S_i = \{209, 150, 120, 105, 90, 75, 60, 50, 40, 35, 30, 25, 20, 17, 14, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1\} \tag{3.1}$$

The dropped variables at each specific reduction step received identical importance ranks, from 30 (least important variables) to 1 (most important variable). Importance ranks were determined with step-wise variable reduction because model-based importance of a given input variable can substantially change when some correlated variables occur more frequently than others. Otherwise, using CUBIST importance measure on the entire spectrum would confound the importance of relevant but highly correlated variables. Since RFE is a wrapper method of variable selection, external test sets (resampling) were needed to exclude selection bias in estimating subset performance (RMSE) (Kuhn and Johnson, 2013). For this purpose, we nested another inner layer of resampling for RFE within the five times repeated 10-fold CV scheme. Importance ranks of variables and outer test RMSEs were averaged from the 50 CV folds. To decrease computation time, we conducted the RFE with five CUBIST committees. The RFE procedure and the resampling setup is explained further in the appendix 3.3.

Local soil estimation for plot-level monitoring

We defined a local soil estimation scenario where a new long-term monitoring site was initiated at time zero (t_0). Each one of the 71 NABO sites was assumed to be novel, while the remaining ones were established with spectral and reference data records. We, therefore, conducted 71 separate sample selections from the SSL, each yielding different transfer subsets of the SSL, to test spectral-based soil monitoring using the Swiss mid-IR SSL presented here.We calibrated models at each site using two local samples per given site and a relevant subset of the remaining Swiss SSL (see description below).

The two local samples were chosen from pooled samples at t_0 (first two out of a maximum of four replicates), or in addition at t_1 if there was only one sample in t_0. Figure 3.2 illustrates the local modeling workflow. All other samples per given site besides the two chosen during calibration (in other words, the successive time series measurements at a monitoring plot) were used as local validation samples (N_{site}). The respec-

tive samples from the remaining SSL included spectra and reference measurements from all BDM samples and NABO samples, excluding the ones from the respective target site. We used only two calibration samples per NABO site to capture the predictive mechanisms at site level because we wanted to avoid overoptimistic local assessment; both local calibration and validation samples were repeated soil measurements, and are otherwise — if not adequately handled in the calibration sampling strategy — at risk of overfitting when soils' composition and relevant properties show constant trends over time.

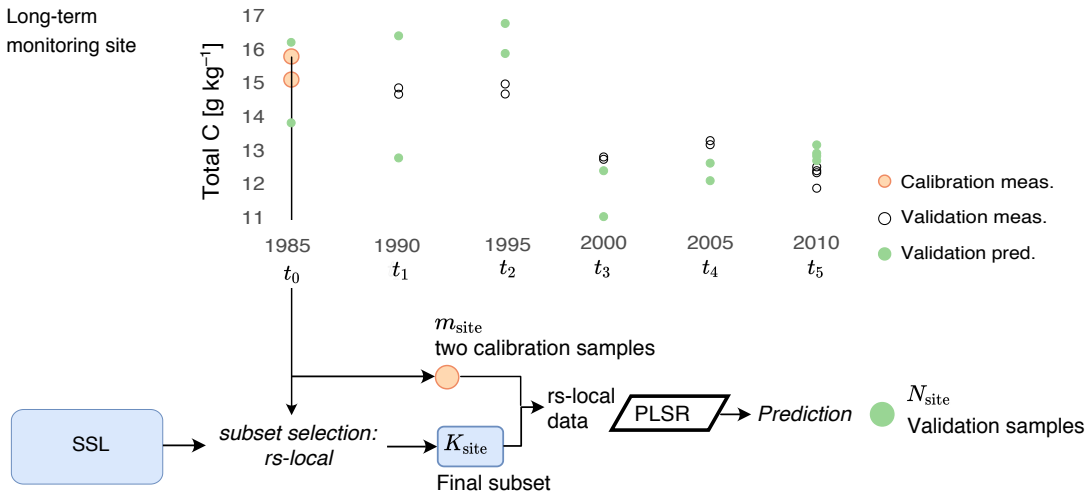

Figure 3.2: Conceptual scheme illustrating the transfer of the soil spectral library (SSL) to a long-term monitoring site using the RS-LOCAL approach. The local calibration samples and a subset of the SSL are used to calibrate a partial least squares regression (PLSR) model, which predicts the local validation samples.

For each of the 71 sites, the spectral relevant samples from the remaining Swiss SSL were selected using the RS-LOCAL algorithm (see Lobsey et al. (2017)). The site-specific samples (m_{site}) denote local calibration samples from a NABO plot. The recursive reductions of the initial training data, which determined the finally yielded subsets (K_{site}) were driven by model performance (RMSE) for the two local calibration samples. For each NABO site, the corresponding K_{site} set was spiked with the two local calibration samples. On this combined $m_{site} + K_{site}$ data set, a final partial least squares regression (PLSR) model, locally adapted for the monitoring plot by optimization on the calibration samples, was developed using 10-fold cross-validation. Finally, the local validation spectra (N_{site}) were predicted using the most accurate calibration model.

The search algorithm RS-LOCAL has three empirical parameters to control the samples that are selected for the local transfer from the SSL (Lobsey et al., 2017). Parameter k is both the number of samples drawn from the original and reduced library without replacement and the number of samples of the returned SSL subset. Parameter b is the number of times k samples are randomly drawn from the remaining data at iteration i of the performance-driven library reduction. Parameter r is the proportion of samples, which are consistently in weakest models, that are removed at it each reduction step. The configuration of the RS-LOCAL search was optimized for each NABO site. For each site, we ran separate RS-LOCAL runs, testing a full-factorial combination of empirical parameter sets $k = \{30, 50, 150\}$, $b = \{10, 20, 50\}$, $r = \{0.05, 0.1, 0.2\}$. The RS-LOCAL

procedure is based on PLSR (Wold et al., 1983). For the RS-LOCAL tuning during the subset selection procedure and final calibrations, we tested 1 to 10 PLSR components. The finally selected optimal subset per site yielded the smallest RMSE on the two local calibration samples, and was therefore used to predict the local validation samples.

Uncertainty of spectral monitoring uncertainty: CUBIST vs. RS-LOCAL transfer

To compare the performance of the CUBIST approach and RS-LOCAL transfer, errors and concordance of both methods were conditionally assessed per individual NABO (n = 71) site. For CUBIST, grouped cross-validation holdouts were used. Thereby, the two respective local calibration samples m_{site} were excluded, so that the test configuration was identical to the local transfer scenario. In addition to the mentioned assessment statistics, the ratio of performance to interquartile distance (Bellon-Maurel et al., 2010; RPIQ; 75th and 25th percentiles) was used for relative comparisons between the local transfer and rule-based model because it is robust to non-normal and skewed distributions of measured values.

Evaluating the predictive mechanisms behind the local transfer

For each of the 71 statistical transfers at a plot level, we quantified the similarity between the selected data sources K_{site} (from SSL) and the respective local target domain $\{N_{\text{site}}\}$ (local validation) by multivariate distances across the spectral input variables. The distance of single observations within $\{K_{\text{site}}; N_{\text{site}}\}$ was referenced to the center of all data, which led to two respective distributions of distance measures for these sets of points and per site. This procedure involved two steps, namely (1) compressing the input data to reduce the "curse of dimensionality" (Bellman, 1962) and being able to discriminate similarity with spectra (with many dimensions, distance to nearest neighbor becomes similar to distance to farthest neighbor) and (2) calculating the Mahalanobis distance using a robust method (see below; Varmuza and Filzmoser, 2016) so that the location and scatter were influenced by the main data rather than by atypical observations.

To condense the spectral information over the entire SSL, Savitzky–Golay preprocessed spectra that included all observations with C elemental measurements were mean centered, scaled and then transformed by principal component analysis (PCA) using singular value decomposition. Dimensionality reduction was necessary to avoid computationally singular values during the subsequent calculation of the covariance matrix (for the Mahalanobis distance). The first t10 principal components that explained 86.5 % of the variation in preprocessed spectra were kept for distance calculations. Finally, the Mahalanobis distance of all the observations to their center was computed with robust

estimates for both the center and the covariance matrix of the selected PCA scores, using the minimum covariance determinant (MCD) estimator (Rousseeuw, 1984; Hubert and Debruyne, 2010).

3.3 Results

Summary of reference measurements

The samples from the Swiss Soil Monitoring Network (NABO) exhibited the highest variability across samples for total C and OC (n = 592; Table 3.1). Organic C ranged from 1 g kg^{-1} to 583 g kg^{-1}. The texture of the soils varied considerably. The pH had values between 3.5 and 7.6 and the soils were slightly acidic overall with a median of 5.8. Compared to the NABO data set, the soils from the BDM program covered a wider set (n = 3723 for total C) and range of measured soil properties. The measured range of total C for BDM (1 – 583 g kg^{-1}) extended further than that of the NABO. The distribution of pH values was similar in the NABO and BDM sets. The BDM data also included the available cations extracted by AAE (see Table 3.1). The median CEC_{pot} (potential cation exchange capacity) was almost equivalent to the value of the NABO sites (24 vs. 23 cmol(+) kg^{-1}). Exchangeable Ca showed the largest coefficient of variation (CV = 1.56) among the measured properties of the BDM set. All soil properties, except pH and CEC_{pot} were positively or neutrally (sand) skewed, for both NABO and BDM data sets, respectively.

General soil estimation with CUBIST modeling

For most of the properties, minimal cross-validated errors were achieved with 100 committees and nine neighbors. The rule-based models explained a large proportion of the variation ($R^2 > 0.9$) in properties that typically have a strong link to total C (organic C and N; Table 3.3; Fig. 3.3). Clay was accurately estimated (RMSE = 47 g kg^{-1}; RPIQ = 3.0; range = 0–602 g kg^{-1}), whereas sand and silt were less accurately estimated. The pH was accurately estimated (RMSE = 0.3; RPIQ = 6.5). Our models discriminated a large proportion in the measured variation of Ca and Mg (ammonium acetate–EDTA) in the mid-IR (R^2 = 0.97 and 0.79; RPIQ = 2.4 and 1.2). Reference values of potential cation exchange capacity ranged from 0 to 136 cmol(+) kg^{-1} and were modeled with an RMSE of 7 cmol(+) kg^{-1} (R^2 = 0.72; RPIQ = 2.0). However, the extractable nutrients P, K, Cu and Zn were insufficiently explained by mid-IR spectral rules (R^2 = 0.05 – 0.1; RPIQ = 0.4 – 0.9). Nonetheless, the rules achieved nearly unbiased property estimates over all measurements. We found marginal local bias at the largest values, mostly for variables with positively skewed distributions such as total C (Table 3.3; Fig. 3.3).

Table 3.1: Summary statistics of the measured soil properties of the Swiss soil spectral library derived from the sample archive of the Swiss Soil Monitoring Network (NABO) and the Swiss Biodiversity Monitoring (BDM) program. Total C is total carbon, OC is organic carbon, CEC_{pot} is potential cation exchange capacity.

	n	No. of Locations	Min	Max	Median	Mean	SD	Skewness
NABO								
Total C [$g\,kg^{-1}$]	572	71	11	273	33	40	35	3.84
OC [$g\,kg^{-1}$]	592	71	11	273	30	37	34	4.03
N [$g\,kg^{-1}$]	572	71	1.1	19.9	3.2	3.6	2.6	3.12
pH	574	71	3.5	7.6	5.7	5.8	0.9	0.09
Clay [$g\,kg^{-1}$]	80	55	30	590	220	231	105	0.73
Silt [$g\,kg^{-1}$]	81	55	150	800	380	397	125	0.95
Sand [$g\,kg^{-1}$]	80	55	40	820	400	371	166	0.01
CEC_{pot} [$cmol(+)\,kg^{-1}$]	121	58	7	136	23	26	17	4.10
BDM								
Total C [$g\,kg^{-1}$]	3723	1079	1	583	41	55	49	4.04
OC [$g\,kg^{-1}$]	3652	1068	0	583	37	50	48	4.55
N [$g\,kg^{-1}$]	3724	1079	0.0	26.4	3.2	3.8	2.5	2.91
pH	3765	1094	2.6	8.0	5.6	5.6	1.3	−0.10
$CaCO_3$ [$g\,kg^{-1}$]	1565	455	0	885	36	107	144	1.80
Clay [$g\,kg^{-1}$]	787	785	5	602	194	213	108	0.71
Silt [$g\,kg^{-1}$]	787	785	105	708	303	309	95	0.74
Sand [$g\,kg^{-1}$]	787	785	5	817	419	411	168	−0.08
CEC_{pot} [$cmol(+)\,kg^{-1}$]	674	190	0	94	24	26	12	1.39
P (AAE) [$g\,kg^{-1}$]	417	417	1	1047	19	40	77	7.99
K (AAE) [$g\,kg^{-1}$]	417	417	22	1255	106	136	115	4.06
Ca (AAE) [$mg\,kg^{-1}$]	417	417	141	96250	3927	12226	19127	2.20
Mg (AAE) [$mg\,kg^{-1}$]	417	417	16	3196	161	232	259	5.35
Cu (AAE) [$mg\,kg^{-1}$]	417	417	2	73	6	8	5	5.32
Zn (AAE) [$mg\,kg^{-1}$]	417	417	1	131	4	6	9	8.52
Fe (AAE) [$mg\,kg^{-1}$]	417	417	84	1640	342	387	194	1.93

Cross-validated mid-IR estimates of 18 soil properties derived from rule-based CUBIST models that were developed using all available data. The 10-fold cross-validation procedure was grouped by the site and repeated 5 times to achieve many test-train data combinations and provide realistic model assessment with generalization. #C and #N denote the number of committees and neighbors for CUBIST. Total C = total carbon , OC = organic carbon, N = total nitrogen, and CEC_{pot} = potential cation exchange capacity.

	n	Range	#C	#N	#Rules	RMSE	ME	SDE	R^2	RPIQ
Total C [g kg^{-1}]	4295	1583	100	9	626	8.4 ś 0.1	0.3 ś 0.0	8.4 ś 0.1	0.97 ś 0.00	4.3 ś 0.0
OC [g kg^{-1}]	4244	0583	100	9	424	9.3 ś 0.2	0.0 ś 0.0	9.2 ś 0.2	0.96 ś 0.00	3.4 ś 0.1
N [g kg^{-1}]	4296	0.026.4	100	9	921	0.55 ś 0.00	0.00 ś 0.00	0.55 ś 0.00	0.95 ś 0.00	4.3 ś 0.0
pH	4339	1.08.0	100	9	418	0.3 ś 0.0	0.0 ś 0.0	0.3 ś 0.0	0.93 ś 0.00	6.5 ś 0.0
$CaCO_3$ [g kg^{-1}]	1565	0885	10	9	15	6.6 ś 0.0	0.1 ś 0.0	6.6 ś 0.0	0.79 ś 0.00	2.6 ś 0.0
Clay [g kg^{-1}]	867	5602	100	9	26	47 ś 1	1 ś 0	47 ś 1	0.81 ś 0.00	3.0 ś 0.0
Silt [g kg^{-1}]	868	100800	100	9	114	71 ś 1	−0 ś 0	71 ś 0	0.51 ś 0.01	1.7 ś 0.0
Sand [g kg^{-1}]	867	5820	100	9	16	89 ś 1	−1 ś 0	89 ś 1	0.72 ś 0.00	2.8 ś 0.0
CEC_{pot} [cmol(+) kg^{-1}]	795	0136	50	9	110	7 ś 0	0 ś 0	7 ś 0	0.72 ś 0.01	2.0 ś 0.0
P (AAE) [mg kg^{-1}]	417	11047	50	9	18	77 ś 1	−3 ś 1	77 ś 1	0.05 ś 0.00	0.4 ś 0.0
K (AAE) [mg kg^{-1}]	417	11255	100	9	112	111 ś 2	−3 ś 1	111 ś 2	0.10 ś 0.01	0.8 ś 0.0
Ca (AAE) [mg kg^{-1}]	417	196250	100	9	614	3326 ś 121	−60 ś 56	3325 ś 121	0.97 ś 0.00	2.4 ś 0.1
Mg (AAE) [mg kg^{-1}]	417	13196	100	9	113	123 ś 5	4 ś 1	123 ś 5	0.79 ś 0.02	1.2 ś 0.0
Cu (AAE) [mg kg^{-1}]	417	173	50	9	17	5 ś 0	−0 ś 0	5 ś 0	0.10 ś 0.01	0.9 ś 0.0
Zn (AAE) [mg kg^{-1}]	417	1131	100	9	113	9 ś 0	−0 ś 0	9 ś 0	0.06 ś 0.01	0.4 ś 0.0
Fe (AAE) [mg kg^{-1}]	417	11640	50	5	17	167 ś 3	1 ś 1	167 ś 3	0.28 ś 0.02	1.2 ś 0.0

Overall, out of the 16 available soil properties, total C, total N, total $CaCO_3$, Ca and Mg (ammonium acetate–EDTA), OC, CEC_{pot}, pH, sand and clay (10) were modeled with relatively good discrimination capacity in the measured ranges (Fig. 3.3).

Model interpretation and filtering with variable importance

Figure 3.4 shows that the test RMSE of total C first slightly decreased and then steadily increased from all (209) to less spectral variables using CUBIST and RFE. The lowest error ($RMSE_{test}$ = 8.10 g kg^{-1} total C) of spectroscopic estimation was achieved with the spectra with 105 variables. For the subsequent variable reduction steps, model performance steadily dropped until one wavenumber was left ($RMSE_{test}$ = 18.8 g kg^{-1} total C).

The spectral feature between 1786 cm^{-1} and 1754 cm^{-1} was the most important one for the estimation of total C with CUBIST (Fig. 3.4). The 12 spectral variables with the best importance ranks across all RFE iterations and test sets derived from the subset sizes were (starting with the best) 1754 cm^{-1} (mean(rank) = 1.04) , 1786 cm^{-1} , 1770 cm^{-1} , 2010 cm^{-1}, 2506 cm^{-1}, 1850 cm^{-1}, 1370 cm^{-1}, 2522 cm^{-1}, 1818 cm^{-1}, 1866 cm^{-1}, 2058 cm^{-1}, and 1386 cm^{-1} (mean(rank) = 12.7; Fig. 3.4).

Accuracy of the local transfer models compared to the general model

For the example site 65 COR, the best performance of RS-LOCAL was achieved with 55 samples from the SSL (K), 10 sampling events (B) of size K at each iteration, and 10 % reduction (r) at each iteration (Fig. 3.5). Therefore, 55 transfer samples from the SSL were combined with two site calibration samples previously used to supervise the selection from the data source, to form a PLSR calibration model for the estimation of the site validation samples (see Fig. 3.5 panel *a* right). Compared to the target observations from the site (right part of Fig. 5*a* and *b*; measured range = 11.9–16.0 g kg^{-1} C), the selected instances were heterogeneous with regard to their characteristic patterns in raw spectra, their preprocessed feature space, and their measurements (range = 8.7–97.7 g kg^{-1} C). The selected instances covered a significant proportion of the first two components in the feature space of the entire SSL.

The RMSE on the site validation samples ($RMSE_{N_{site}}$) at the final subsets varied between 0.01 g kg^{-1} C and 10.73 g kg^{-1} C and for all tuning parameter combinations and sites, and between 0.01 and 3.02 g kg^{-1} C for the best subsets per site (Fig. 3.8).

The local approach reduced the error of the rule-based approach on average by factor 4.4 (Fig. 3.6; mean($RMSE_{RS\text{-}LOCAL}$) = 0.7 g kg^{-1} C; mean($RMSE_{CUBIST}$) = 3.1 g kg^{-1} C). The local transfer was more accurate for the majority of NABO sites (69 out of 71 sites).

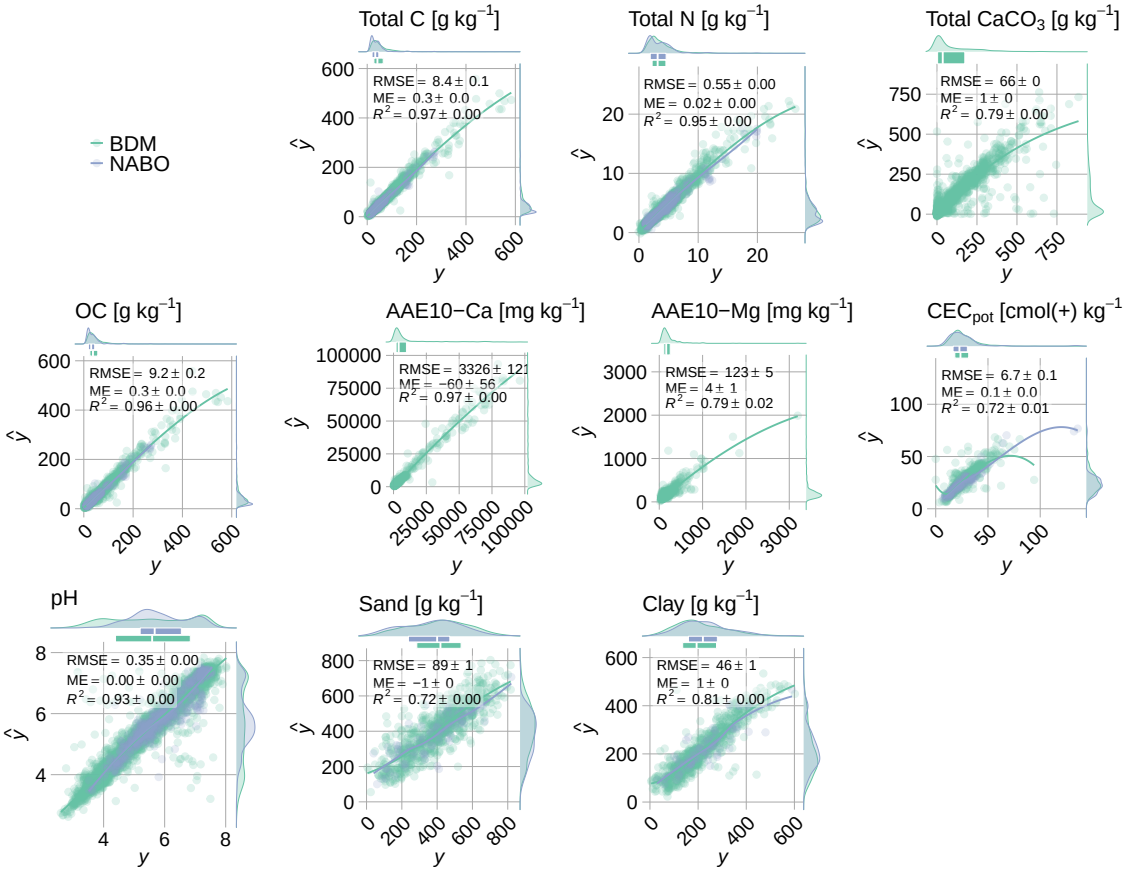

Figure 3.3: Agreement between measured and mid-IR predicted values that were obtained from CUBIST models. Models' performance was assessed by site-grouped cross-validation hold-outs (five times repeated 10-fold). The lines obtained with local regression (LOESS) smoothing indicate the trends in predictions. Models of soil properties with $R^2 \geq 0.72$ are shown (see Table 3.3 for more detailed model summaries).

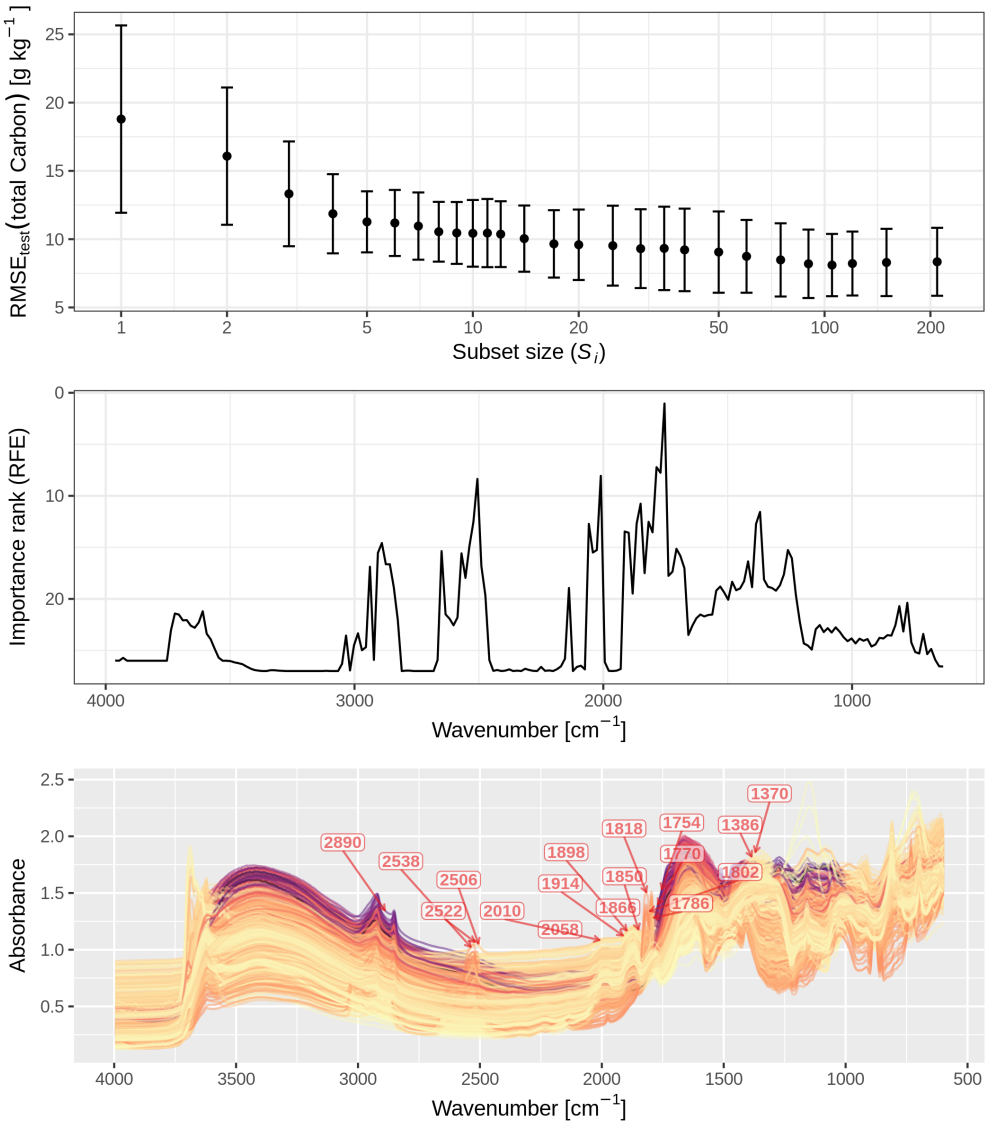

Figure 3.4: *Top:* Root mean square error (RMSE) of mid-IR estimates of total C that CUBIST produced at the respective subsets of spectral variables. The performance profile was obtained with a recursive feature elimination (RFE) procedure. The error bars represent the standard deviations of the test RMSE derived with nested cross-validation (n = 50). *Middle:* Average importance ranks across the spectrum. Lower rank values indicate higher importance for the estimation of total C. Ranks were determined with RFE. *Bottom:* Mid-IR absorbance spectra of the Swiss soil spectral library (n = 4295; with corresponding total carbon (C) measurements determined by dry combustion). The unprocessed absorbance spectra are annotated with the 17 most influential spectral variables (wavenumbers) in the CUBIST model (average importance rank < 15); these had the highest mean importance ranking determined by the recursive feature elimination procedure.

The linear dependency between modeled and measured values was higher for the local transfer compared to the general model (53 out of 71 sites). Moreover, RS-LOCAL produced on average 1.3 times less biased estimates of total C per site for 52 out of 69 sites in terms of absolute values ($|\,\mathrm{ME}\,| = 0.1\,\mathrm{g\,kg^{-1}}$ C vs. $0.5\,\mathrm{g\,kg^{-1}}$ C). The ratio of performance to interquartile distance (RPIQ) confirmed that local learning in the mid-IR was able to better discriminate developments of total C over time, relative to its measured distribution. Overall, local learning with two local calibration samples and targeted SSL selections allowed for better estimations than the generic CUBIST approach on average (RPIQ = 3.08 vs. 1.00; RPIQ larger for 66 out of 71 sites). Across all validation data points of the NABO set, the RS-LOCAL transfer was 5.6 times more accurate for total C than the general rules in terms of RMSE and RPIQ (RMSE = $0.9\mathrm{g\,kg^{-1}}$ C; RPIQ = 31.7)

Predictive mechanisms behind the local transfer

The samples used for the transfer process (RS-LOCAL data) of the example site COR 65 showed high spectral dissimilarity along the first 2 PCs (principal components), explaining 39.8 % of the preprocessed spectral variance (Fig. 3.5). Compared to the entire SSL with total C measurements available (the source domain prior selection; range of PC1 – from −41.4 to 13.0; range of PC2 – from −19.0 to 30.0), the selected transfer samples of this site occupied a region of major variation in the PC space (range of PC1 – from −15.4 to 11.4; range of PC2 – from −10.2 to 10.9). The two local calibration samples and the 12 validation samples on the upper right corner were close to each other in the PC1-PC2 subspace (Fig. 3.5, panel *a*, *left* and *right*; range PC1: 9.2 to 11.0; range PC2: 4.9 to 7.5). Not only the absorbance spectra but also the corresponding C reference values were highly variable compared to the exemplary NABO site (Fig. 3.5, panel *b*; 7.3–117.8 $\mathrm{g\,kg^{-1}}$ C for $\mathrm{K_{RS\text{-}LOCAL}}$, and 11.9–16.0 $\mathrm{g\,kg^{-1}}$ C for the plot of this site). This particular target monitoring site indicated that RS-LOCAL selected soils from the SSL with a relatively large spectral diversity and a wide range of total C.

The instances selected by RS-LOCAL filled a substantial proportion of the SSL's feature space (Fig. 3.7), confirming the trend of site *65 COR*. We found that RS-LOCAL yielded quite a wide selection of relevant samples from the SSL with reference to both the total C range and a wide coverage of spectral features expressed with robust multivariate locations. The spectral estimations of the site validation sets that resulted from RS-LOCAL-based transfers neither showed trends in the mode or spread for distributions of C measurements nor in the ones from their spectral distances. The measured distributions of K_{site} SSL subsets and N_{site} local validation samples for further key soil properties related to the chemical composition (OC, pH, $\mathrm{CEC_{pot}}$, clay and $\mathrm{CaCO_3}$) were also markedly different, confirming the local transfer of quite heterogeneous soils (Table 3.4). For example, standard deviations of the 0 %, 25 %, 50 %, and 75 % percentile dif-

Table 3.4: Standard deviations (SDs) of the absolute differences in percentiles ($P_0, P_{25}, P_{50}, P_{75}$, and P_{100}) of final RS-LOCAL subsets (K_{site}) and corresponding site validation samples (N_{site}) the across 71 long-term monitoring sites. The aggregated values for six measured soil properties are shown. Total C is total carbon, OC is organic carbon, CEC_{pot} is potential cation exchange capacity.

	$\text{SD}\Big(\lvert P_X(K_{\text{site}}) - P_X(N_{\text{site}})\rvert\Big)$				
	$\text{SD}\Big(\lvert\Delta P_0\rvert\Big)$	$\text{SD}\Big(\lvert\Delta P_{25}\rvert\Big)$	$\text{SD}\Big(\lvert\Delta P_{50}\rvert\Big)$	$\text{SD}\Big(\lvert\Delta P_{75}\rvert\Big)$	$\text{SD}\Big(\lvert\Delta P_{100}\rvert\Big)$
Total C [g kg^{-1}]	18	22	25	31	61
OC [g kg^{-1}]	25	23	22	22	66
pH	0.8	0.7	0.5	0.6	0.9
CEC_{pot} [cmol(+) kg^{-1}]	9.8	10.3	10.6	10.6	23.1
Clay [g kg^{-1}]	110	97	85	68	90
$CaCO_3$ [g kg^{-1}]	74	79	89	98	178

ferences between the transfer sets selected the SSL and the samples from the respective NABO site were on average between 18 g kg^{-1} and and 66 g kg^{-1} for measured C and OC, respectively. Further, the measured clay and $CaCO_3$ contents were markedly different between the RS-LOCAL selection and the local validation sets (mean absolute median differences of 85 g kg^{-1} clay and 89 g kg^{-1} $CaCO_3$). This findings correspond with the dissimilar selection compared to the local target samples found in the PCA space of preprocessed spectra.

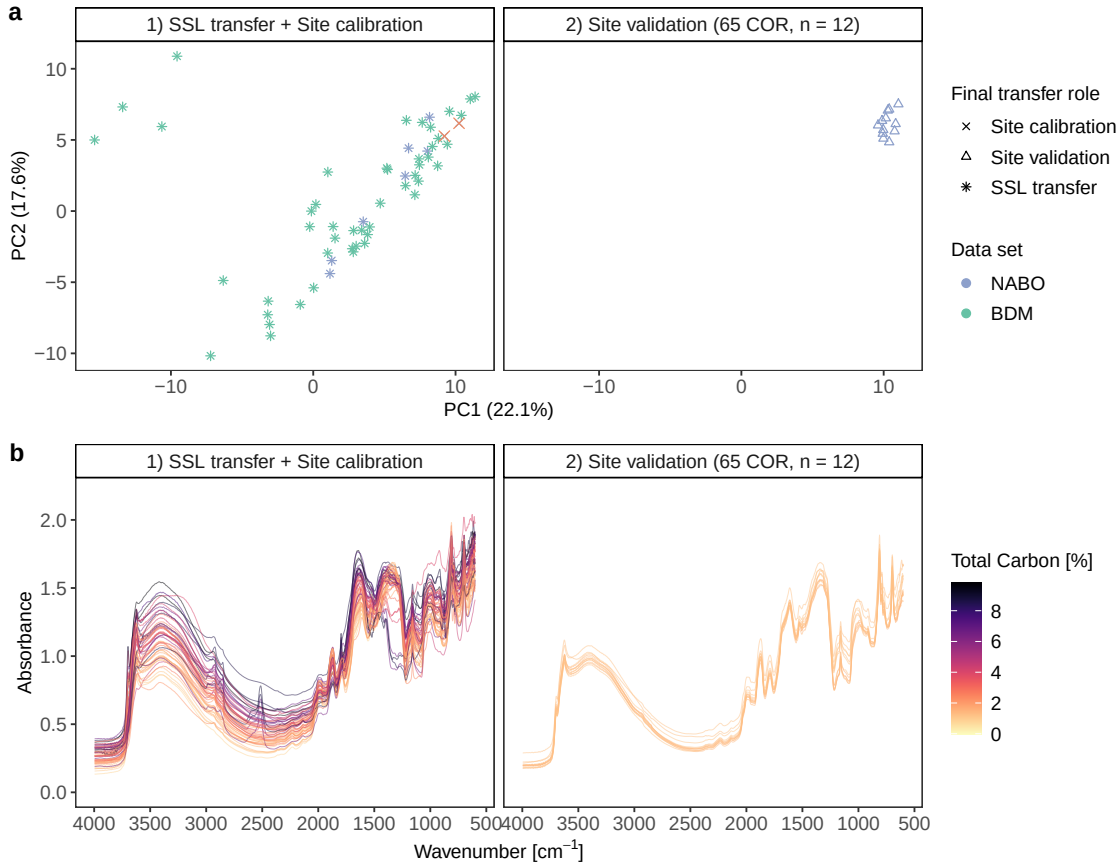

Figure 3.5: Illustration of the site-specific transfer modeling of total carbon (C), using RS-LOCAL for the example site 65 COR of the Swiss Soil Monitoring Network (NABO). Panel *a* contains the principal components subspace (PC1 and PC2) of the Savitzky–Golay first derivative mid-IR spectra, and panel *b* outlines the corresponding absorbance spectra (unprocessed for illustration), which are colored by the total C content. The left subplots show the SSL transfer samples (n = 55) that were selected from the soil spectral library (n = 4281; excluding all NABO calibration samples). This subset was most accurate when predicting the two calibration samples under the mechanisms RS-LOCAL and their optimal tuning configuration for the site ($\{K = 50; B = 10; r = 0.1\}$). The right panels shows the time series data for the validation samples of the NABO site called 65 COR.

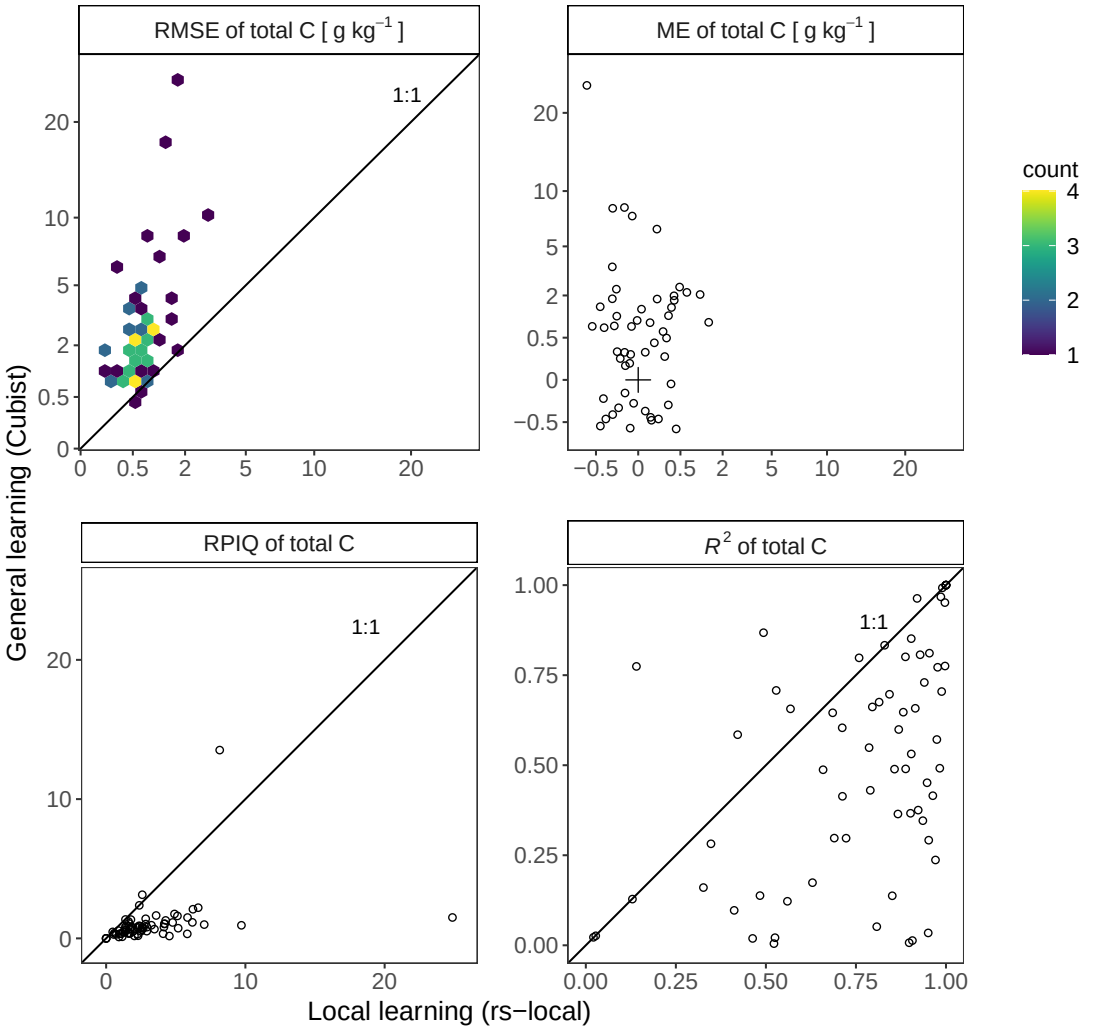

Figure 3.6: Model assessment of the estimated total carbon (C) of 71 NABO sites for the general learning with CUBIST (y axis) vs. local learning transfer with RS-LOCAL (x axis). The four panels depict the root-mean-square-error (RMSE), the mean error (ME), the ratio of performance to interquartile distance (RPIQ) and R^2. The 1:1-line emphasizes the difference between the two approaches.

Figure 3.7: Analyzing the mechanisms behind the individual adaptive transfer realized with RS-LOCAL. Panel *a*: The left horizontal bars show the root mean square error (RMSE) of mid-infrared predictions of the temporal validation set N_{site} of time series of total carbon (C) for each of the 71 NABO sites, which was calculated without the two respective calibration samples. The blue density plots depict the distribution of the site-specific validation samples, and the brown vertical bars show the measured values of C for the final subsets of SSL used for the transfer (K_{site}). Panel *b*: The distribution of the robust distances from the PCA center of Savitzky-Golay preprocessed spectra of the entire soil spectral library compared to the subset of instances involved in the individual transfer modeling (K_{site}) and validation samples (N_{site}) (similarity in site-specific vs. final RS-LOCAL selection), computed with the Minimum Covariance Determinant (MCD) estimator.

3.4 Discussion

General soil estimation with the Swiss SSL

Many of the chemical properties with distinct links to soil organic matter and the key minerals (e.g., clays and quartz) were discriminated well with mid-IR CUBIST models (Table 3.3; Fig. 3.3). Specifically, the models estimated total C, OC, N, pH, texture, AAE10-Ca and AAE10-Mg with $R^2 > 0.7$. This suggests that the majority of developed models are useful for applications that require soil proxies in order to manage land resources. For example, $CEC_{pot.}$ (RMSE = 7.0 cmol(+) kg^{-1}) and pH (RMSE = 0.3), have high ecological importance for nutrient availability in ecosystems. In agriculture, both measures are key factors for soil fertility and nutrient recommendations.

The accuracy of our estimates for the properties that have direct chemical links, through compound-associated absorptions, were mostly comparable to established continental or country-specific mid-IR SSLs. For example, Clairotte et al. (2016) achieved RMSE = 2 g kg^{-1} for OC using mid-IR and the spectrum-based learner for local predictions, while Sila et al. (2016) reported RMSE = 4 g kg^{-1}. The accuracy of our general OC estimates was lower (RMSE = 9.3 g kg^{-1}; RPIQ = 3.4), which we explain with the relatively large range of measured values and variable mineralogy (Stenberg and Rossel, 2010). We found that total C had more CUBIST rules per committee than OC (Table 3.3), indicating that total C, which also included inorganic C (mostly $CaCO_3$), leverages more chemical constituents and latent absorptions for its estimation. In spite of lower parsimony, slightly more accurate estimates of total C were achieved (RMSE = 8.4 g kg^{-1}; RPIQ = 4.3).

The majority of soil properties were most accurately estimated with the maximum tested 100 committees and nine neighbors. Instance-based correction with similar data in training set, (nearest) neighbors, yielded considerably higher accuracy for total C (e.g., RMSE = 8.9 g kg^{-1} for 20 committees and two neighbors vs. RMSE = 8.1 g kg^{-1} for 20 committes and nine neighbors; model evaluation across cross-validation folds; results not shown). The number of rules give a first proxy for model complexity and the complementary of spectral features that are involved in prediction. The range in number of rules across the ensembles was widest for total C (6 – 26), similar for OC (4 – 24), medium wide for CEC_{pot} (1 –10), and very narrow for $CaCO_3$ (1 – 5) to give specific examples. Viscarra Rossel and Webster (2012) report comparably fewer rules (medium – 21; range of all properties – 5 – 64) for OC and relatively similar number of rules for CEC (15). Nonetheless, such comparisons have to be done with care because the NIR range has a less pronounced representation of functional groups than the mid-IR range, and because temperate soils have fundamental differences in chemical composition com-

pared to more weathered tropical soils. For our mid-IR SSL, we were surprised that the rules for OC were complex, similar to the ones for total C; in fact, we also could not find any clear partitioning in the rules with respect to measured ranges and spectral patterns (exploratory analysis not shown), which is in contrast to Viscarra Rossel and Webster (2012). In fact, this is different from the general patterns found by Viscarra Rossel and Webster (2012), where the rules clearly partitioned the data into distinct measured distributions. Last but not least, the diversity in rules for total C and OC of the general estimation approach makes the soil diversity selected from the library and what we found for site-specific local transfer even less exotic (see Sect. 3.4).

The variable importance assessment of the spectroscopic models revealed five major regions of features with particularly high predictive influence for total C –, i.e., 2890 cm^{-1}, 2522 cm^{-1}, 2010 cm^{-1}, 1754 cm^{-1} and 1370 cm^{-1} (Fig. 3.4). We attribute the two absorption peaks near 2890 cm^{-1} to C – H stretching vibrations of organic matter (Skjemstad and Dalal, 1987), which were also relatively important for estimating C in other studies (e.g., Janik and Skjemstad, 1995; Viscarrra Rossel and McBratney, 1998). The important variable at 2522 cm^{-1} is indicative of C=O absorption due to the carbonyl group present in carbonates (e.g., calcite; Nguyen et al., 1991; Soriano-Disla et al., 2014). The three important absorptions between 2010 cm^{-1} and 1786 cm^{-1} result from three consecutive Si – O – Si (overtone and combination) absorptions, which are indicative of quartz. However, the most important absorptions near 1754 cm^{-1} showed no distinct peak but an edge feature. This is in accordance with Sila et al. (2016), who identified this region as being most relevant for estimating total C with a (general) random forest model developed from the SSL of the Africa Soil Information project. This region is close to the C=O stretching vibration of the carboxyl group that occurs around 1725 – 1720 cm^{-1} (Madari et al., 2006), which is further confirmed by the high importance of these vibrations found by Janik and Skjemstad (1995). The last relatively important region around 1370 cm^{-1} was also an edge feature with no distinctly visible peak of chemical group assigned, which, however, might be influenced by the adjacent carboxylate (COO^-) or – CH absorptions at 1400 – 1350cm^{-1} of aliphatic compounds such as humic acids (Madari et al., 2006; Parikh et al., 2014). In summary, the CUBIST-RFE variable importance analysis enabled us to link characteristic absorptions of typically prominent functional groups of soil organic and inorganic C compounds, and as well quartz absorptions as indirect correlative features of predictive relevance, with our general model-based estimates of total C.

Because the rule-based models we developed can estimate 10 soil properties reasonably well ($R > 0.6$; RPIQ > 2.0; Fig. 3.3), the Swiss SSL will be useful for new soils when new reference measurements for model adaptation are relatively scarce or not available. Thereby, the Swiss SSL will be cost and time efficient for characterizing soils of

similar composition in the near future. The new predictions can further be augmented with straightforward model interpretation, which allows chemical inference of pedological aspects to provide means of model applicability. Although the combined BDM and NABO set comprises a large soil variability in Switzerland, the diversity of subsoils at depths greater than 20 cm — mostly in terms of the mineral composition — as well as peat and forest soils are probably not yet represented sufficiently in the SSL. We therefore must continuously update the present SSL with more and deeper soil horizons in the near future.

Local transfer from the SSL for soil monitoring at plot-scale

The local estimates of total C that were derived with RS-LOCAL selection were substantially better on average (RMSE = 0.7 g kg^{-1} C) as those derived using all of the data and general CUBIST models (RMSE = 3.1 g kg^{-1} C; Fig. 3.6). The data-driven estimation at plot-scale further considerably reduced bias and increased R^2 compared to the general CUBIST rules.

Our third goal was to analyze the characteristics of soils that were selected from the SSL and used for establishing locally-adaptive models tailored to the respective long-term monitoring sites. Surprisingly, the RS-LOCAL subsets selected from the SSL had rather dissimilar spectra in the robust PCA space (Fig. 3.5; Fig. 3.7); their distances to the center had a wide distribution compared to the local samples. The K_{site} subsets accordingly covered a large proportion of the spectral input space. The likely dissimilar chemical composition of soils was also reflected in the reference measurements of total C. We conducted a broader analysis to interpret the soil context of the selected samples with further soil compositional covariates (OC, pH, CEC_{pot}, clay, $CaCO_3$), which also did not resemble the soil characteristics of the local monitoring sites (see Fig. 3.4). These findings together with the accurate validation results clearly indicate that dissimilarity and diversity in soils can also provide the means for fitting locally-adaptive models.

Nevertheless, we can yet only speculate about how and why such diverse calibration sets are able to leverage accurate local calibrations. One hypothesis is that by increasing the range and variability in spectral variables and measurements a model can become quite stable in the central range of local reference measurements because a larger range of input variables is considered; thereby, the RS-LOCAL subsets that are selected from the SSL and used for PLS regression would stabilize and reduce the errors of the local samples. We imagine that we leverage a similar mechanism as in simple linear regression, where narrowing the range of the independent variable (x) in the training samples would decrease the accuracy of intermediate values of the independent

variable. We therefore need to further look into the details of spectral dissimilar learning, for example, also investigating the relevance of specific spectral features for local spectral transfers. The inherent working principles of RS-LOCAL are in contrast to the spectrum-based learner (SBL) or other forms of memory-based learning that utilize similar samples to infer sample-specific predictions based on existing training data (Lin and Vitter, 1994; Ramirez-Lopez et al., 2013b). Our approach could describe a data-driven phenomenon, which implies that spectra can help to estimate a set of unrelated new soils. Another possibility is that there is in fact a pedological explanation that could be elucidated with more soil covariates such as mineralogy.

Local soil characterization is simpler, quicker and cheaper when a large proportion of properties of new soils are estimated by spectroscopy. Our results suggest the importance of optimizing the transfer of relevant information present in large SSLs to minimize the required amount of conventional laboratory analyses of new soils. Soil chemical and physical heterogeneity can be substantial in large SSLs. Therefore, such data variation can be beneficial for future predictions of properties of soils. However, learning a single general model over a heterogeneous training set, and obtaining parameter estimates optimized with a global measure of goodness of fit can introduce bias and inaccuracy to local (soil) estimation (Hand and Vinciotti, 2003; Ramirez-Lopez et al., 2013b; Lobsey et al., 2017). Although the highest estimation accuracy could be achieved with only soils of the target study area (Stenberg and Rossel, 2010; Guerrero et al., 2016), it is impractical and inefficient to derive a single spectral prediction model with those. It requires 1) a large volume of reference measurements for a reasonably accurate multivariate calibration, and 2) it does not utilize already existing soil information.

Currently, the Swiss long-term soil monitoring uses a spatially representative sampling and then bulking the soils into four replicates for reference measurements (Desaules et al., 2010; Gubler et al., 2019). When the long-term monitoring would be augmented with mid-IR spectroscopy, one could make spectral measurements on all subsamples, rather than only on bulked samples, which would deliver spatially-explicit information and reduce nuisance factors from different sampling conditions. If not constrained economically (separate drying, sieving, and milling of sub-samples), a spectral workflow could thus allow to account for small-scale soil variability and reduce bias in measurements to robustly estimate temporal soil changes. For example, there is currently a relatively large variability in C measurements between the bulked replicate samples at one time point (Gubler et al., 2019). Our results suggest that unbiased spectral measurements eventually mediate such inconsistencies. Although Gubler et al. (2019) reported only minor changes for the ensemble of permanent cropland or cropland-meadow monitoring sites (30), there were four sites with declining trends and nine

sites with increasing trends in OC (−11% to +16% relative change per decade, respectively). Here, the trend of spectroscopic predictions could be investigated with respect to specific research questions on agronomic management-induced changes, also with further physicochemical soil characterization (e.g., OC fractions).

Relatively precise and unbiased geographically-local estimates of soil properties from diverse and large SSLs can be achived by a handful of data-driven statistical approaches that are currently popular in the soil science community (Viscarra Rossel and Webster, 2012; Ramirez-Lopez et al., 2013b; Guerrero et al., 2014; Lobsey et al., 2017; Tsakiridis et al., 2020). Among the methods, we tested RS-LOCAL (Lobsey et al., 2017) in our local soil monitoring scenario. Compared to memory-based learning, such as SBL (Ramirez-Lopez et al., 2013b), RS-LOCAL does not precondition the choice of useful subsets based on similarity in the input dimensions, here spectra, when performing the selection of SSL samples. The RS-LOCAL method is applied to exhaustively sample instances from the SSL without replacement, while it preferably selects those that perform well on the local target set, using PLS regression. An advantage of the method is that it can deal better with erroneous spectra as well as inaccurate and imprecise analytical reference measurements in the SSL, because it filters them as irrelevant instances. Besides chemometric and classical machine learning approaches, convolutional neural networks are being popularized for modeling SSLs with large soil variability (e.g., Liu et al., 2018; Padarian et al., 2019a,b; Tsakiridis et al., 2020). There seems to be a small performance gain of a multi-output CNN with a similarity-based error correction using neighbors compared to the SBL (Tsakiridis et al., 2020; RMSE = 11 g kg^{-1} vs. 12 g kg^{-1} for OC). Despite the current development of interpretation methods in deep learning, CUBIST and PLSR modeling employed in both in the SBL and RS-LOCAL offer easier interpretation with comparable accuracy to CNNs.

Transfer learning or local learning introduces a new paradigm to supervised learning: model building is governed by the intended model application and thus coupled to it (Hand and Vinciotti, 2003). This contrasts general-model application, where the inference process is separated from the prediction of new data. Including local samples and their local data characteristics is necessary in order that a combined search and learning algorithm has a chance to capture predictive mechanisms. At the same time, the selection process and the partial data dependence within the predictive unit, the site, requires a careful assessment scheme to prevent a potential selection bias in the assessment of the approach. To account for this, we kept the respective site-specific local tuning and calibration set — whose hold-out performance directed the iterative search process and the reduction of the SSL — at minimum size of two observations at t_0 or in addition t_1 when only one measurement was available from the first sampling (see Fig. 3.2).

Future applications and updates of the SSL

We found that data-driven modeling with selection of spectral dissimilar soils (see Fig. 3.7) is accurate for inducing local predictions of total C (Fig. 3.6). Hence, there is the need to further improve data-driven selection using RS-LOCAL, i.e., by further optimizing the current version of the algorithm. To address this need, we could use combined memory-based or lazy learning strategies (Stanfill and Waltz, 1986; Lin and Vitter, 1994) to optimize with more data-driven transfer methods (Pan and Yang, 2010) in terms of reducing the time needed to evaluate suitable subsets of the SSL for a new application. To give an example, some similarity criteria or clustering before doing calibration sampling could be used as prior information for reducing the SSL size to obtain the final subsets. In principle, the sample reduction could also be done with algorithms that can deal with non-linear relationships between spectra and soil properties, such as random forest or CUBIST. Another extension is to further filter spectral features and to do data compression to make the local modeling faster and even more adaptive to local conditions.

Our results showed that a transfer of the SSL to individual monitoring sites yielded very low bias and was accurate. This indicates that mid-IR spectroscopy and SSLs have the potential to give quick and relatively precise soil property estimates for soil monitoring. Nevertheless, the sites of the NABO long-term-monitoring program has not undergone substantial changes in OC (Gubler et al., 2019). Up to now, although major changes in C content and organic composition should yield a spectral response, spectral changes in OC have mostly been reported along chronosequences (i.e., Awiti et al. (2008)), and only rarely for changes within individual plots over time (Deng et al., 2013). Hence, to address this, we propose to further investigate to what extent mid-IR spectroscopy can detect changes of OC considering small-scale variability and different agronomic management practices. This could for example be achieved with a study using soils from a long-term field trial, that shows sufficient temporal changes to be detected with spectroscopy.

The current SSL includes soils that contain between 0 and 583 $g\,kg^{-1}$ total C and OC (Table 3.1). Because organic soils can have up to 500 $g\,kg^{-1}$ OC, and because more than 98% of the samples are mineral soils, organic soils are underrepresented in the current Swiss SSL. For this reason, Helfenstein et al. (2021) evaluated the present Swiss SSL for a regional transfer based on new organo-mineral soils from two peat land regions in Switzerland. Although the range of total C measured was large (14 – 520 $g\,kg^{-1}$ C) and the soils were diverse, as few as 5 or 10 site-specific tuning samples were sufficient to estimate the validation samples with reasonably accurately (RMSE = < 30 $g\,kg^{-1}$ C; RPIQ > 3.4); this was comparable to a local-only calibration with 50 samples. Helfen-

stein et al. (2021) found considerably lower conditional prediction errors ($< 10\,g\,kg^{-1}$) when considering measurements of $< 100\,g\,kg^{-1}$; this suggests that increasing the amount and compositional complexity of organic soils in the library has potential for more accurately characterizing diverse soil ecoregions with soil high organic matter contents.

Our results suggest that the present mid-IR SSL has great potential for applications that require soil data in high temporal and spatial coverage (i.e., for deriving quantitative indicators of soil quality for spatial planning or for soil-related environmental research). Mid-infrared spectral modeling was able to estimate many soil properties accurately with rather large variation in measurements explained (Fig. 3.3), making them suitable for agronomic diagnosis and the assessment of soil functions in various landscapes. Currently, fine grained soil information of properties and function across agricultural lands in Switzerland is still scarce and often challenging to harmonize (i.e., measurement methods) because legacy maps are at varying levels of detail and quality (Keller et al., 2018; Grêt-Regamey et al., 2018). For example, only 13 % (127000 ha) of soil in agricultural land has been mapped with soil attributes of sufficient quality to evaluate its potential for crop production (Rehbein et al., 2020). Soil properties are also insufficiently mapped nationwide from point into space, depth and over time to regionally model soil processes, or to evaluate site-specific effects of agricultural practices on soils (i.e., soil C dynamics). Therefore, we suggest to couple infrared spectral estimation with traditional soil surveys and digital soil mapping to speed up the collection of soil information in Switzerland and elsewhere. This will offer means to test and further extend this SSL, so that only minimal amounts of costly and time consuming traditional laboratory analyses will be needed for characterizing and mapping soils' properties and functions in the next decades.

3.5 Conclusions

We developed the Swiss mid-IR SSL (n = 4374), using legacy soils and reference measurements of 16 properties, from 71 long-term monitoring sites (national soil monitoring; NABO) and 1094 locations sampled from a regular grid over Switzerland (biodiversity monitoring program; BDM). The trained CUBIST models — a general modeling approach using all data — were able to explain a relatively large proportion ($R^2 \geq 0.72$; RPIQ ≥ 2.0) of measured variance for ten of the properties. Total C, OC, total N, pH, CEC_{pot}, and clay content were estimated with high discrimination capacity ($R^2 > 0.8$; RPIQ > 3.0). Total C was estimated with a cross-validated RMSE = $8.4\,g\,kg^{-1}$ at a measured range of $0 - 583\,g\,kg^{-1}$, and OC with RMSE = $9.3\,g\,kg^{-1}$ at the same measured range. Compared to the general CUBIST approach, the local transfer yielded on average 4.4 times more accurate estimates of total C with the mean RMSE = $0.7\,g\,kg^{-1}$ C,

which is a substantial improvement of local estimates at plot-scale. Our similarity analysis revealed that local learning with subset selection based on RS-LOCAL produced a chemically diverse calibration set rather than narrowing down soil diversity for local modeling, as it is for example the case in memory-based learning. The developed national mid-IR SSL offers rapid soil estimates which are key inputs for many applications requiring soil information, such as digital soil mapping, agronomic diagnostics and precision farming, soil C accounting and monitoring, etc. The created mid-IR SSL and both local and general models can be updated with new soil records, which will allow to cover more soils conditions and will require less and less soil laboratory reference measurements in relation to spectral measurements for monitoring, mapping and modeling new soils.

Appendix A: Figures and tables in appendices

A1 Recursive feature elimination for interpreting general soil estimation with CUBIST

The recursive feature elimination (RFE) procedure started with the initial set of $S_1 = 209$ predictive variables that resulted after processing the spectra (see section 3.2). The following subset sizes S_i representing the number of spectral variables that are retained after each i^{th} variable elimination step were defined and evaluated within the RFE procedure:

$$S_i = \{209, 150, 120, 105, 90, 75, 60, 50, 40, 35, 30, 25, 20, 17, 14, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1\} \tag{3.2}$$

The first variable elimination step ($i = 1$) started with tuning a full CUBIST model derived from $S_1 = 209$ possible predictors using 10-fold cross-validation, then calculating the CUBIST model usage statistics for all predictors, next sorting all predictors from highest to lowest importance, and lastly dropping $S_1 - S_2 = 59$ of the least important predictors. For the next iteration ($i = 2$) and the following ones, we repeated this model fitting and variable reduction procedure with $S_2 = 150$ predictors and the preceding subsets, until the most important predictive variable ($S_{30} = 1$) was left at the last iteration ($i = 30$).

Variable selection is in addition prone to overoptimistic model assessment when resampling subsets (i.e., cross-validation) are used for two purposes, here model build-

ing and selection. This selection bias due to data leakage is well-documented for so-called *wrapper methods* of variable selection like RFE (Ambroise and McLachlan, 2002; Kuhn and Johnson, 2013), and occurs if these two tasks are not sufficiently separated by using independent data sets for each of them; this becomes especially more important when many predictive variables in relation to to relatively few observations are used, as it the case for our spectra.

To provide realistic predictive generalization of the RFE method, the aforementioned iterative selection procedure was done within an internal cross-validation scheme so that independent data were used to test the performance of the variable selection on the outer data segments. These outer cross-validation segments served external validation. To quantify the uncertainty of the models using the reduced variable sets and specifically variable selection, the outer cross-validation layer that served cross-validation was repeated five times, leading to five independent estimations per sample.

Tuning profile of the RS-LOCAL parameters for local predictive transfers

The most relevant samples from the SSL at each respective NABO long-term monitoring plot were empirically selected at the RS-LOCAL configuration that yielded the lowest RMSE on two calibration samples per plot (Fig. 3.8; performance profile). Time-series validation on the remaining samples of each site was separated from the optimization in the transfer workflow (see Fig. 3.2).

Coda and data availability. The data that were used to produce the results of this paper are available upon reasonable request.

Author contributions. B prepared the draft, conceptualized the spectral modeling scenario, carried out the data analysis and wrote the code. JL and JS developed the concept of a Swiss Spectral Library. All co-authors contributed to writing the paper. DW and RGM contributed to writing the methods section, AH contributed to the introduction, results and the discussion, and AK provided inputs to writing the introduction and the conclusion. AG created the map of locations and contributed to writing the methods and the discussion. RVR helped guide the spectral modeling experiments and provided the implementation of RS-LOCAL in R.

Competing interests. The authors declare no competing interests.

Acknowledgments. We express our thanks to Leonardo Ramirez-Lopez, Laura Summerauer and Jonas Anderegg, for their valuable inputs that helped to improve the pa-

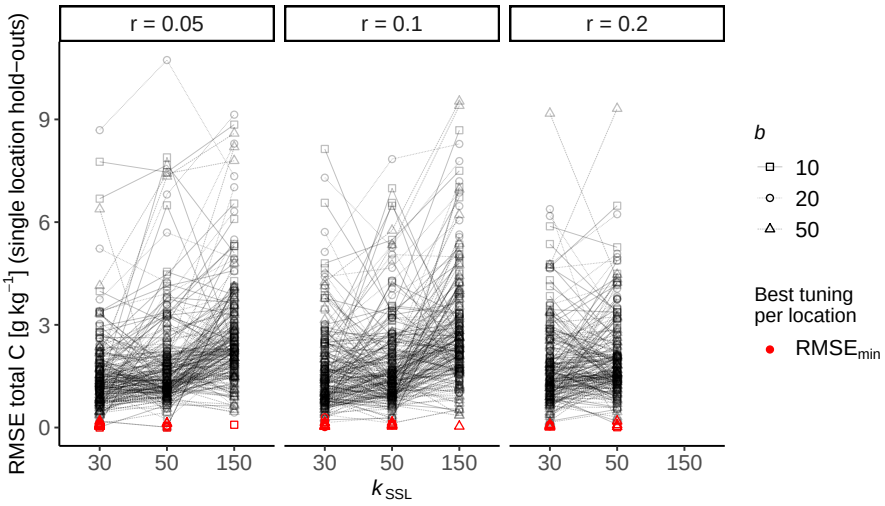

Figure 3.8: Performance profile of the 27 empirical parameter combinations of RS-LOCAL tested on each of the 71 NABO sites. The root mean square error (RMSE) of the plot-level transfer was assessed with the first two calibration samples for each time series of total carbon (C) (see Fig. 3.1 for an illustration of the setup of the local predictive transfer)

per. This research was funded by ETH grants. We also would like to express our gratitude to the Federal Office of Environment for commissioning and funding the soil analyses conducted within the BDM. NABO provided the milled soil archive of NABO and BDM, together with the chemical reference data and metadata.

Part 4

Chapter 3: Detecting management induced changes in soil organic carbon at the plot level with mid-infrared spectroscopy

Philipp Baumann[1,*], Maike Krauss[2], Paul Mäder[2], Markus Steffens[2], Juhwan Lee[3], Jonas Anderegg[4], Raphael Viscarra Rossel[5], and Johan Six[1]

[1]Group of Sustainable Agroecosystems, Institute of Agricultural Sciences, ETH Zurich, 8092 Zurich, Switzerland
[2]Department of Soil Sciences, Research Institute of Organic Agriculture (FiBL), 5070 Frick, Switzerland
[3]Department of Smart Agro-industry, Gyeongsang National University, Jinju, 52725, Republic of Korea
[4]Plant Pathology Group, Department of Environmental Systems Science, ETH Zurich, Zürich, Switzerland
[5]Soil and Landscape Science, School of Molecular and Life Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia.

Abstract

Sustainable soil management practices should maintain or increase present levels of soil organic matter (SOM) and organic carbon (OC) stocks in farmlands. SOM is inherently important for the storage and continuous provision of soil nutrients, promotion of soil structure to prevent erosion and sequester OC in the short to medium term.

The verification of soil OC changes and its controls needs frequent, spatially-explicit and reliable measurements of bulk density and OC contents across soil depths. Because the traditional measurements are too costly and laborious to do this at sufficiently high and financially affordable throughput, soil spectral libraries (SSLs), minimal local reference measurements and information transfer to local conditions might provide an efficient solution to satisfy this high demand of data. Thus, we aimed at determining the soil OC changes at plot level and across soil depths for short-range measurements between consecutive time points (ΔOC) through mid-IR spectroscopic calibrations of OC contents (n = 311) for an agronomic long-term experiment (LTE), which showed primarily tillage-mediated increases of OC contents over five sampling years between 2002 and 2018. Thereby, we compared purely local mid-infrared modeling (partial least squares regression (PLSR); CUBIST), and both local performance-driven (RS-LOCAL) vs. similarity-driven (spectrum-based learner; SBL) extraction of knowledge stored in the mid-IR SSL of Switzerland (n = 4244), together with only 10 LTE observations with analytical measurements for targeted selection and/or spiking. The purely local cross-validated point estimates were relatively accurate with a root mean squared error (RMSE) = 1.4 $g\,kg^{-1}$ and a ratio of performance to interquartile range (RPIQ) = 4.8, and were almost unbiased. The SBL approach produced an 1.9-fold higher RMSE compared to the local PLSR calibration, while the RS-LOCAL approach with CUBIST modeling across the finally selected SSL subset had only 1.5-times higher RMSE. The spectroscopically estimated ΔOC were less accurate for the best purely local model (RMSE(ΔOC) = 1.9 $g\,kg^{-1}$; RPIQ = 1.7), and was relatively better for RS-LOCAL (RMSE(ΔOC) = 2.5 $g\,kg^{-1}$; RPIQ =1.3) compared to SBL (RMSE(ΔOC) = 3.2 $g\,kg^{-1}$; RPIQ = 1.0). For SBL, we also found higher non-unity slopes (b = 0.68) for the regression of measured vs. estimated compared to both RS-LOCAL (b = 0.71) and purely local PLSR (b = 0.88). We have also found that the gradient in clay contents and contents of inorganic carbon clustered well with the spectra, and that the spatial blocking of the experiment very well accounted for small-range spatial soil chemical heterogeneity. We conclude that a locally optimized data-driven selection of beneficial diversity in SSL with only 3% of novel local reference measurements drastically increases the efficiency of reproducing plot level ΔOC with the spectroscopic transfer approach. Marginal slope effects, probably due to unaccounted functional OC complexity in the transfer, call for more research to improve temporal adaptation of the modeling approaches.

4.1 Introduction

Soil organic matter (SOM) is a heterogeneous mixture of organic substances and one of the most complex chemical and biological media on earth. SOM supports a plethora of soil functions that sustain life and ecosystem services, such as carbon (C) seques-

tration, nutrient cycling, water dynamics, production of food, and providing habitats and thus enhancing biodiversity. SOM is primarily made of carbon and soil organic carbon is quantitatively and qualitatively important for many different soil processes, which foster ecological interactions and transformations (Six et al., 1999). Organic compounds influence the crop productivity in agroecosystems because they store and and release water and mineral nutrients. Soil management practices can alter soil conditions and the distribution of contents and forms of organic carbon in spatially and temporally-explicit manners, e.g., along the soil profile, at micro-scale arrangement (patchiness of SOM), and in different locations over time. It is therefore critical to monitor its dynamics over time and in space — in context of the soil and environmental conditions under investigation, and with regard to the extent of agronomic interventions.

The present global challenges such as climate change and soil degradation, which are a threat to soil and food security, are urging us to better manage our soils as they are a precious resource. Thus, we need to monitor changes in soil attributes to evaluate and verify the effects of our interventions, so that we can maintain, restore or even improve soils rather than continue their current global trajectory of soil degradation (Komatsuzaki and Ohta, 2007; FAO, 2020). Agricultural soil carbon crediting is one of the proposed instruments to steer action towards better soil conservation and improved soil health. Here, for example, not only records of land use and management changes, but also a large amount at a high frequency of quantitative data on soils are needed to do soil monitoring over space and time at low minimal detectable difference (Post et al., 2001; Conen et al., 2003; Smith, 2004; Viscarra Rossel et al., 2011). Hence, gaining more data is the only way to ultimately validate positively (or negatively) soil biophysical changes that occur in terms of ground-truth measurements.

The traditional methods in the laboratory (wet chemistry) are well established and precise enough for deriving the baseline and change of organic carbon (OC) contents over time. However, the concept of minimal detectable change (MDC) with required samples and time periods for reliably measuring changes needs consideration (Necpalova et al., 2014; Deluz et al., 2020) if given the mean changes are normally distributed (Saby et al., 2008). The correct determination of changes is still very challenging because of the interaction of soil C inputs, climate, vegetation and physical, chemical and biological soil status, causing large spatial heterogeneity (Post et al., 2001; Spencer et al., 2011; Arrouays et al., 2012; Smith et al., 2020). Therefore, traditional analyses alone (i.e., dry combustion before and after adification for OC) are clearly too costly and laborious to meet the demand of soil data to support most of environmental monitoring endeavors with those complexities in mind. Alternatively, a practical solution is to utilize spectroscopic measurements in the lab and possibly on-field, and intelligently model and in-

terpret OC contents and also carbon fractions with predictive analytics (Cremers et al., 2001; Francioso et al., 2005; Zimmermann et al., 2007; Gehl and Rice, 2007; Stevens et al., 2006; England and Viscarra Rossel, 2018). Mid-infrared (mid-IR) spectroscopy is particularly useful to describe and model functional diversity and changes in soil organic matter (McCarty et al., 2002; Calderón et al., 2013). For this, well harmonized soil, data collections with reference measurements, and a minimal new data set can be equally important. The crucial part is to maintain these soil archives well and to make them compatible with quality-controlled measurements from the past, that fit into the scope of the new diagnostic application. The soil sensing approach roots two major principles that greatly improve quantitative soil assessment: 1) the ability to conduct more measurements in local and more coarse study contexts, geographical extents, and sampling densities, and 2) the parsimonious estimation and interpretation of spectral signals through statistical modeling (Viscarra Rossel and Bouma, 2016; Viscarra Rossel et al., 2017). There is greater variance of spectroscopic proxy estimates that arises from measurement noise and modeling errors compared to chemical and physical reference analyses. Still, these errors are more than paid off by many more measurements that are pivotal to deriving robust and relatively unbiased estimates of changing soil attributes, such as OC. Hence, the sensing paradigm gives statistical power to quantifying biophysical changes of soils by massively scaling-up the collection of soil data, so site-specific projections of overall temporal trends of soil OC can become more feasible.

Established statistical techniques such as partial least squares regression (PLSR; Wold et al., 1983, 2001) together with other chemometric tools (i.e., signal processing; Delwiche and Reeves III, 2010) usually provide accurate calibrations when the spectral intensities are fairly linear to the physical, chemical or biological measurements of interest (Martens and Næs, 1984). The concept of linear dimensionality reduction by latent projections have also proven to be successful in rather defined soil (data) contexts, as collected from individual sites up to regions (Fernández Pierna and Dardenne, 2008). This has in turn facilitated the adoption of local(-only) calibration methods for soil IR spectroscopy. Custom-made calibrations for isolated areas hosting one or a few chemically distinct soil types, for example at the extents of individual farms (RamirezLopez et al., 2019), or for subsets of larger spectral libraries are simple to realize in terms of the standard-modeling approaches (Vohland et al., 2022), but more tricky for application in new soil contexts. This approach of very (context-)specific calibrations, however, has the major drawback that the expensive and time-intensive collection of laboratory reference data and modeling and validation has to be repeated over and over again for a novel study or project. Further, there is the cost of generally lower accuracy or completely failing calibrations when new data is too dissimilar in terms of chemical

soil composition compared to already calibrated soils in a soil spectral library (SSL). Although there are good guideline recommendations for sound soil spectroscopic practice (e.g., Nocita et al., 2015b), there is no general golden rule for the amount of reference data needed for calibration as this is typically context-specific (size of data set, soil variability, geographical extent and sampling density, pedogenetic factors, etc.). Nevertheless, the number of samples needed for a stable local-only calibration, e.g. at field-scale or also within smaller landscapes (i.e., land health degradation framework; Vågen et al., 2012), is typically 25 calibration samples or more (Wetterlind and Stenberg, 2010); this is the magical threshold of local-only calibrations below which spectroscopic transfer from other related and previously characterized soil collections can be useful.

There is shareable knowledge available in large soil spectral libraries (SSLs), which comes along with high soil compositional variability (Vasques et al., 2015; Viscarra Rossel et al., 2016; Sila et al., 2016). To bring such general soil spectroscopic data sources into the equation and deal with but also profit from the increasing complexity of soils, machine learning approaches are often needed for better accuracy (Brown et al., 2006; Rossel et al., 2010; Viscarra Rossel and Webster, 2012). The reason is that the increasing complexity of soils and relationships to measured properties comes with highly diverse spectroscopic feature spaces and predominantly non-linear relationships to the measured spectra. Despite this, spectral libraries have grown at and over continental extents (e.g., Viscarra Rossel et al., 2016; Sila et al., 2016; Dematt̂e et al., 2019; Dangal et al., 2019), not all issues with data representativeness and generalization capacities have been addressed with highly predictive and often also interpretable machine learning approaches (i.e., CUBIST; Quinlan, 1993) and memory-based learning, particularly with the spectrum-based learner (SBL; Ramirez-Lopez et al. (2013b,a)). Such general or global calibrations on all available training samples can be useful in situations of data scarcity, for example, when more reference measurements are not affordable. For example, relatively large errors at new local target sites, that typically spread an area of few hectares to square kilometers, were problems that were encountered with considerably large SSLs and global calibrations (Guerrero et al., 2016). Recent method development has lead to a novel data-driven modeling method — ReSampling-local (RS-LOCAL; Lobsey et al., 2017). RS-LOCAL addresses the main limitations in usability of SSLs, which is that they often bring biased prior information that is explicitly used in modeling, e.g. soil type, geography, depth, or spectral similarity. Because both RS-LOCAL and the SBL select specific observations (instances) of the source domain (i.e., SSL), to targeted models for new samples to predict (target domain), they can both be considered as instance-based transfer learning methods (Pan and Yang, 2010). Whereas the SBL selects a unique set of neighbors from the SSL to build specific models

for each new prediction sample, RS-LOCAL determines one set of optimally performing samples from the SSL for PLSR modeling and prediction for each local target set. Empirical evidence shows that the previous methods do often not improve local predictions. In contrast to this, RS-LOCAL can principally leverage relevant predictive information from both (spectral) similarity-based (instance-based) and diversity-leveraging (learning) paradigms. Hence, this makes SSLs more universally usable. It has also shown success in spectroscopic estimation when modeling in proximity fails because of missing training data in the feature space of SSLs or erroneous reference measurements. Results are in particular promising to achieve both site-local (Baumann et al., 2021) and regional-level adaptation (Helfenstein et al., 2021) with optimal calibration sets with a selection of chemical diversity that benefits the new prediction tasks. The reported mid-IR estimation errors were significantly lower than those of large SSLs combined with general machine learning across all data, but only slightly higher than purely local calibrations with high data densities for particular soil chemistries.

For soils, spectroscopic diagnostics are prominent in spatial applications, but the collection of soil samples and measurements is often not repeated over time. Hence, the time component of soil biophysical changes has mostly been limited to chronosequences and large-scale land health assessments (Awiti et al., 2008; Vågen et al., 2012; Deng et al., 2013; Zheng et al., 2016), and only rarely been addressed based on repeated analyses over time at individual locations at small soil-geographical extents. Recent pedometric work and also an earlier study on experimental amendment with SOM compounds has shown more direct evidence that mid-IR spectroscopy could be efficient and accurate for detecting shifts in OC contents (Calderón et al., 2013; Sanderman et al., 2021). However, few have critically assessed the generation of an adaptive statistical learning framework that keeps track of changes in SOM quality and in particular soil OC contents over time under different soil mineralogical contexts, and have also critically discussed how its opportunities and limitations now and back then have changed (Stevens et al., 2008). Hence, more detailed analysis around the efficient and accurate transfer of SSLs to field and plot level are still almost absent. A successful application of mid-IR technology for soil monitoring implies that both spectral measurements and modeling are responsive to a change in contents of characteristic functional groups — likely due to enhanced plant residue input and partial decomposition — at individual agroecological locations. It remains also crucial that spectral interpretation is done for validating effective biophysical changes in soils. Last but not least, proof-of-concept requires data sets of sites that have sufficient fluctuations of OC contents compared to the modeled errors. Since transfer learning with RS-LOCAL at the plot resolution has not been evaluated in full depth so far in the soil spectroscopic community, but indicated potential as shown in Baumann et al. (2021), we wanted to test this approach for

its potential in soil monitoring applications.

In this work, we firstly aimed to address the research gap and evaluate the efficiency of mid-IR spectroscopy to address changes in soil OC contents at short-range spatial (plot) within field-scale variation based on an agronomic long-term experiment (LTE), across different agronomic management conditions (experimental treatments), and changes per plot over time. Secondly, we wanted to elucidate the spectroscopic and soil compositional (mechanistic) relationships that enable the projection of changes in SOC over time. Thirdly, we aimed to test the efficiency of transferring relevant information from the current Swiss mid-IR SSL to field-scale monitoring in terms of the cost-accuracy trade-off for site-specific reference measurements compared to model-spectroscopic estimation errors.

4.2 Material and methods

Experimental site and soil data

We used soils and reference measurements from the archive of the long-term experimental field trial in Frick, Switzerland that has been running since 2002 (Berner et al., 2008; Fontana et al., 2015; Krauss et al., 2010, 2020). The soils and the soil reference data from five sampling years (2002, 2005, 2008, 2011, and 2018) was provided by the Research Institute of Organic Agriculture (FiBL). The Frick long-term experiment addresses the impact of reduced tillage on indicators of soil fertility and crop yields under organic farming management. The trial has been conducted with a split-split plot design, whereby subplots are nested within large strip-like main plots ($n = 4$). The experimental field is on a stagnic eutric cambisol with loamy clay texture. The clay content ranged from 420 to 580 $g\,kg^{-1}$ (Fontana et al., 2015). The clay mineralogy is characterized by smectite and illite, and small amount of poorly cristalline kaolinite (Fontana et al., 2015). For details of the trial layout, more detailed soil conditions, and the management history, we refer the reader to Berner et al. (2008) and Krauss et al. (2020).

Soil total OC was measured with wet oxidation by a modified Walkley-Black method, following the protocols of the Swiss standard method (Agroscope, 1996; Berner et al., 2008). Corresponding measurements of OC by dry combustion were determined by measuring total carbon (TC) with a elemental analyzer (Vario Max Cube CHN, Elementar Analysensysteme GmbH, Langenselbold, Germany) and subtracting inorganic carbon (IC) on a different sample aliquot that was heated to 500 °C prior to measurement with the same instrument.

Spectroscopic measurements

The milled soils were measured on a Bruker Vertex 70v mid-infrared (mid-IR) Fourier-Transform (FT) spectrometer. We used 24-well plates for high-throughput measurements, in which we measured each soil in two analytical replicates by filling the sample cups evenly with soils and gently smoothing the surface with the ridge of a spatula. The first sample position was a gold standard, which was used to reference the reflectance (R) of the spectra via $R = R_{soil}/R_{gold}$. The reflectance was transformed into absorbance (A) using $A = log(1/R)$ and saved on disk in binary format via the Bruker OPUS measurement software. The correction for atmospheric water and CO_2 was also performed in the OPUS software. The spectrometer setting and the measurement procedure was identical as the measurements done for the establishment of the first version of the mid-IR SSL for Switzerland Baumann et al. (2021).

General processing and modeling of spectra

We resampled the spectra from a measured resolution of 1.4 cm^{-1}, to an effective resolution of 2 cm^{-1}. To reduce multiplicative and additive noise we did Savitzky-Golay filtering (Savitzky and Golay, 1964) with a set of window sizes and the first and second derivatives prior to the statistical modeling. To reduce the redundancy in Savitzky-Golay spectra, we selected every 8th variable, which mapped the mid-IR information at an apparent resolution of 16 cm^{-1} for modeling. For spectrosopic data pretreatments, we used the prospectr R package (Stevens and RamirezLopez, 2014).

For local modeling with all samples from the Frick trial, we used a 10-fold cross-validation, to select optimized preprocessing and to tune and assess the respective best models. The cross-validation segments were grouped by the plots of the field and repeated 10 times. By grouping the cross-validation by the experimental plot, we ensured the independence of training and validation samples. Specifically, we wanted to avoid that any spatially and temporally autocorrelated and therefore confounding effects could produce over-fits in the training segments due to inadequate prior information, which in turn would produce overly optimistic model evaluation in test segments. We describe the validation strategies for model fitting and evaluation procedures used for the transfer approaches in the corresponding sections below.

We only show results with the locally-best preprocessing optimizations and models. The principles of empirically testing and the statistical properties of those methods are generally well elaborated in chemometric and broader machine learning research, and also well tested on soil data sets.

Local modeling with samples from the long-term trial

To do local learning with generalized patterns, we modeled all preprocessed sets of spectra with CUBIST and Partial Least Squares Regression (PLSR). Both methods are highly interpretable and accurate, and therefore highly appreciated and widely used in the soil spectroscopy modeling community. While CUBIST performs generally better for extracting rules from highly diverse calibration sets such as SSLs, PLSR can be trained with very few samples, and is particularly robust and accurate for small sample sets with high densities in the local data space. CUBIST was tested with all combinations of the two main empirical parameters (hyperparameters) — model ensembling with 5 and 10 committees, which are sets of rules derived from tree fitting, and 2, 5, 7, and 9 neighbors (N) to adjust the prediction with the n-nearest neighbors. We computed the CUBIST regression models with the Cubist R package.

To uncover the physical and chemical mechanisms that make the detection of SOC

changes possible, a model variable importance analysis was performed. The relative importance of spectral variables and regions was derived with relevant importance metrics for the respective models. We only used the variable importance in the projection (VIP) metric for PLSR, because PLSR yielded more accurate mid-IR estimates of OC than CUBIST, consistently for both wet oxidation and dry combustion reference measurements.

General procedure for transfer learning

The data set of the current mid-IR Swiss SSL that contained corresponding total OC measurements derived with dry combustion ($n = 4244$) was used as information source for transferring knowledge relevant to the soil conditions at the Frick long-term trial. This library was the basis for the sample transfer using RS-LOCAL and modeling in optimized neighborhoods of prediction samples by memory-based learning, specifically with the SBL. Both transfer learning strategies were fine-tuned accordingly to harvest efficient, robust and accurate OC estimates. The data setup and workflow for transfer was influenced by findings from recently published work of the authors, which we explain in the next paragraph and illustrate in Fig. 4.1. We present identical data setups to both transfer methods because we want to provide an adequate efficiency comparison of similarity-driven and completely performance-driven usage of the SSL for plot-scale soil monitoring to the soil research community.

The transfer scenario aimed to select the most relevant information in the current mid-IR SSL of Switzerland specifically for the Frick trial, to capture field-level information from national soil variation. Because the findings in Baumann et al. (2021) indicated that performing sample transfer learning with RS-LOCAL was quite successful for 71 tested point soil monitoring sites of Switzerland, we here further elaborated this approach for this long-term experimental site. Our previous results indicated that it was principally possible to detect temporal trends at decades-time measurement series when the local sets (target domains) were rather homogeneous in terms of chemical soil composition and also their mid-IR PCA space (one transfer per single NABO monitoring site each). Along this line, since the Frick trial data indicated a bimodal distribution of measured carbon, and formed distinct clusters in the mid-IR PCA space, we decided to apply theRS-LOCAL sample selection algorithm for clusters of spectral observations from the Frick trial.

We defined an operational scheme for plot-scale transfer for a long-term experimental site that involves measuring 10 characteristic reference samples from the first year sampled. The assertive number of samples that would require reference analysis was chosen because the main motivation to do similarity- or data-driven transfer from a di-

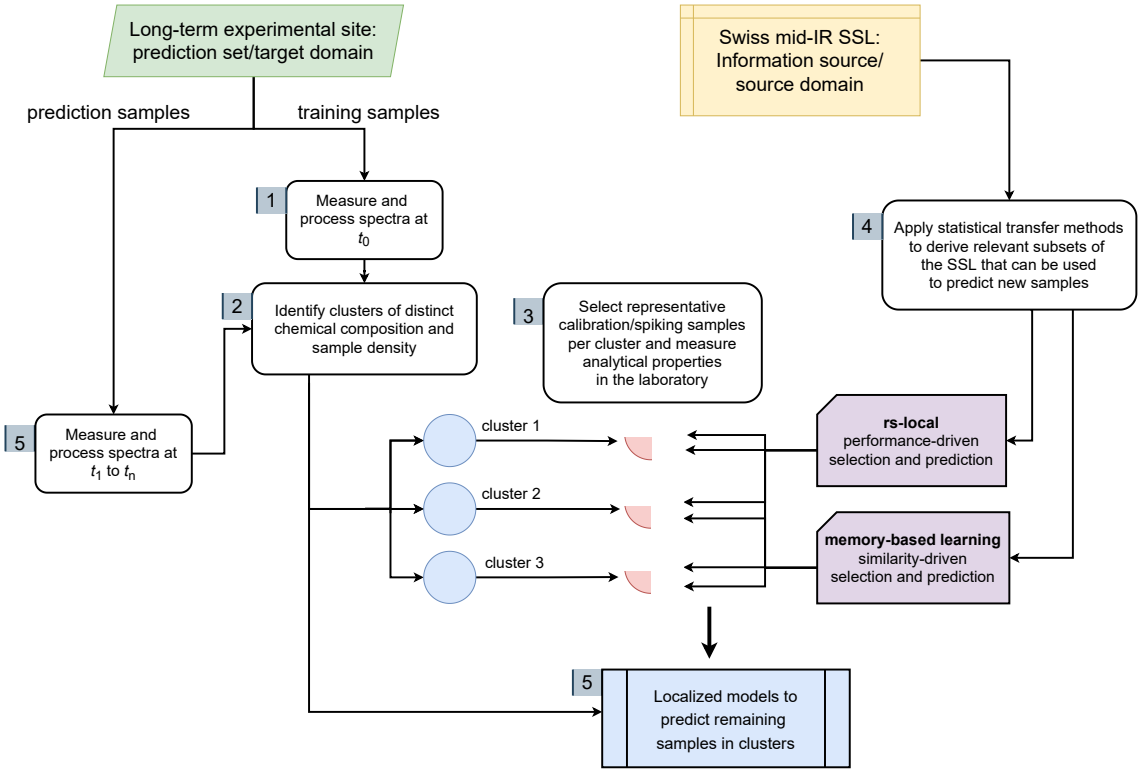

Figure 4.1: General workflow of transferring information from SSLs to localized modeling at a long-term experimental (LTE) site. This workflow can significantly increase the amount of estimated organic carbon (OC) contents for new prediction samples collected. The SSL-to-LTE transfer workflow for OC monitoring was designed for high efficiency in site-specific spectroscopic baseline and temporal change assessments, with minimal amounts of representative samples that undergo chemical reference analysis for calibration and spiking. The majority of predictive information is thereby sourced from the harmonized SSL built with soils of high diversity in spectra (composition) and large range in OC reference measurements, so that 10 representative local reference samples (m) from the first year of the LTE, that are selected under a hierarchical density-based clustering (HDBSCAN algorithm) and standard chemometric calibration sampling methods (Kennard-Stone), can improve the trade-off between mid-IR estimation accuracy and analytical measurement effort.

verse library would be to save time and financial costs for repeated chemical analyses. These ten points will serve as local calibration or transfer optimization samples, and at the same time they are spiking samples for local modeling. We will refer to them as m set. We use the terminology proposed by the developers of RS-LOCAL and the SBL. The m site-specific samples will be used as spiking samples in the extended SBL (see also Fig. 4.1). The recently introduced spiking-functionality in SBL will be explained in the section of memory-based transfer. For the Frick trial, the local calibration samples, that are used to steer the selection of the samples from the SSL, were restricted to the ones sampled in the first year (2002). To capture the main variance in spectra and hence soil chemical variability evenly for calibration sampling, we used spectra in optimal preprocessing found in the local modeling scenario and performed a principal component analysis (PCA) on these. This further facilitated the direct comparison of local-only and transfer-modeling conditions.

The PCA scores that explained at least 75% of variance in preprocessed spectra (PCs 1–4) were subjected to hierarchical density-based clustering with HDBSCAN (Campello et al., 2013, 2015). Using a density-based approach that robustly identified clusters on denoised and dimensionality reduced spectra allowed us to find related data points that had distinct shapes (sample densities), which were assumed to be associated with biophysical soil properties. Within each of the two robust hierarchical clusters found, we assigned the 10 calibration samples proportionally to the total number of samples present in the density clusters. To obtain a representative calibration set within the individual cluster spaces, spanning both atypical (edge) and central points equally in the score space, we again conducted a separate PCA per cluster. Thereby, the calibration samples were selected with the Kennard-Stones algorithm — to account best for soil chemical particularities in the local data structures in the modeling procedure. Specifically, we selected local calibration samples using the first 6 principal components (PCs; singular value decomposition) of the optimal preprocessing determined for local modeling that explained at least 0.95% of variance in spectra.

Sample transfer from the mid-IR Swiss SSL with rs-local

The parameter k is the final number of samples selected from the SSL and also the number of samples randomly drawn b-times from the library at each iteration of RS-LOCAL.

To reduce eventual overfitting in the data-driven optimization of SSL subsets based on a few local samples per cluster, PLSR fitting in the RS-LOCAL reduction procedure was constrained within 1 to maximum 10 PLS components. We used a speed-optimized version of the RS-LOCAL algorithm for tuning the models in the B-loops (PLSR). This

implementation was based on leave-group-out cross-validation that was derived with stratified sampling on the distribution of the response, measured OC (25% validation; 75% calibration; 10 sampling repeats across 10 strata). The speed-optimized version of RS-LOCAL allowed faster and therefore more exhaustive tuning of its controlling parameters compared to the original implementation used in (Lobsey et al., 2017), within reasonable times of computation. In this approach, we tested full-factorial combinations (j) of $r = \{0.95; 0.98\}$, $b = \{50; 100; 150\}$, and $k = \{50, 100, 150; 300; 500\}$. The sampling rates ($b$) of the library and its subsets were above the recommendations that are provided in Lobsey et al. (2017) because further experiments have indicated slight performance gains when doing so (personal communication with Raphael Viscarra Rossel).

After reduction of the SSL and the selection of the final best subsets $K_{\text{cluster};i;\text{tune},j}$ that led to minimal root mean squared error (RMSE) on the local $m_{\text{cluster},i}$ samples, PLSR models were trained on all finally returned RS-LOCAL data sets ($K_{\text{cluster},i;\text{tune},j} + m_{\text{cluster};i}$ (spiking)) using 5-fold cross-validation and ncomp = 1–5 for tuning and assessment. The RS-LOCAL configuration j for which the final model set indicated the best cross-validated RMSE was chosen. Then, with the best combination j for each cluster i, the RS-LOCAL subset selection from the SSL was repeated with the original version of the algorithm that used PLSR with 10-fold cross-validation in the iterations of the B-loop. This was to principally reduce the bias component in cross-validated performance metrics, because it delivered a higher training-to-test set ratio compared to the stratified cross-validation done earlier. The prediction of the test samples $V_{\text{cluster},i}$, which summed up to 301 observations of the Frick long-term experiment in total, was realized with the best models of the cluster-specific RS-LOCAL training sets.

Sample transfer from the mid-IR Swiss SSL with the Spectrum-based Learner

For memory-based learning, the available training set from the Swiss mid-IR with OC measurements was identical to the one used in the RS-LOCAL sample transfer. The entire workflow for localized modeling and prediction was computed per density-based cluster. Here, the site identifiers of the national soil monitoring network (NABO) and biodiversity monitoring program of Switzerland — which indicated repeated measurements over time and/or space — were used as grouping factor for internal group-leave-out cross-validation in the SBL. The optimal number of components (PCA or PLS, respectively) used for dissimilarity calculation, which controls the assembly of the neighborhood with local training samples with these projection methods, was done with optimized principal component (OPC) selection developed by Ramirez-Lopez et al. (2013a,b). We tested from 1 to 20 components for computing dissimilarity matrices.

The local modeling in neighborhoods of Frick samples was done with the weighted average PLS method (Shenk et al., 1997; Ramirez-Lopez et al., 2013b), which was tuned between 1 and 18 PLSR components.

A custom multi-step similarity selection procedure was elaborated to retrieve optimal sets of neighborhood samples from the national SSL for each observation of the Frick prediction set (N_{cluster}; $n_{\text{tot}}(N_{\text{cluster}}) = 310$). It was inspired by the geographical prediction analysis done within the scope of establishing the Congo SSL by Summerauer et al. (2021). We adapted the procedure and added more tuning of dissimiliarity for site-level chemical and physical variability and soil and management context at the Frick LTE.

The procedure started with filtering the training samples of the SSL based on computing the moving window correlation dissimilarity threshold between SSL and samples of the Frick trial (N). This was to yield pre-filtered training sets that excluded irrelevant SSL observations for every local prediction sample. To empirically choose the optimal correlation dissimilarity threshold and the moving window size (w) specified in spectral points (variables), the combination of these parameters that yielded the lowest (aggregated) RMSE of the grouped nearest-neighbor cross-validation statistics in the training subsets (SSL; k-neighbors) derived with the SBL was chosen. This means that at this stage we did only plug in local pre-processed spectra, and no local measurements of OC were used in the SBL fitting. Specifically, we tested (tuned) combinations of $w = \{3;5;11;23;47;59;93\}$ and a sequence of $\rho_{\text{diss.;cutoffs}}$ between 0.025 to 0.4 in increments of 0.025. We limited the range of neighbors that were allowed to be retrained ($k_{\text{mbl;step1}}$) around each predicted observation for the optimal correlation dissimilarity to 20 to 500. Then, all samples from the SSL that have never been selected as neighbors of any local observation were removed from the SSL to be put into the second round of MBL (filtering based on unique neighbors).

In the second MBL fitting round with the reduced SSL, several different dissimilarity metrics were tested in combination with fixed number of neighbors starting from 25 and ending with 400 samples in increments of 25. The m_{cluster} local samples were added to the reduced library for the purpose of spiking, and forced into the neighborhood as part of a recent extension to the resemble R package. The tested dissimilarity methods were PCA distance (Mahalanobis distance computed on principal component projection), PLS distance (Mahalanobis distance computed on PLS projection), Euclidean distance, cosine distance, and the spectral information divergence. The final SBL training model was selected based on best RMSE of grouped nearest-neighbor cross-validation of the training set, based on which the local test samples in the cluster (V_{cluster}) were predicted.

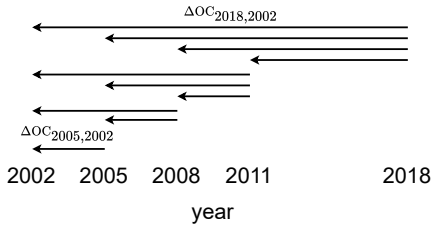

Figure 4.2: Scheme for calculating the differences of soil organic carbon (OC; ΔOC) between the sampling years (2002, 2005, 2008, 2011, 2018) for each plot and soil depth. For each year after 2002, we calculated the measured differences in OC contents within each experimental plot and depth of the time point, to all of the respective preceding years separately. We also calculated the same ΔOC differences for the mid-infrared estimates from the modeling, and assessed the agreement between measured and modeled estimates for purely local modeling, transfer with the spectrum-based learner (SBL), and transfer with RS-LOCAL

Diagnosing changes of organic carbon contents over time with mid-IR

For assessing changes in OC over time at plot-depth level, we defined an approach that involved calculating all consecutively measured and mid-IR estimated OC differences. The OC contents obtained for all time points from t_0 (year 2002) to t_4 (year 2018) were subtracted from the possible values measured earlier in time (t_1–t_0), which was done separately for each plot and depth combination (64 unique spatial, experimental units) separately (see Fig. 4.2). The calculation of differences in OC contents was done for the reference measurements (meas. ΔOC) and their corresponding estimates obtained with mid-IR modeling (pred. ΔOC). For this, the results of predicted changes in OC contents were separately assessed for local-only modeling (PLSR), SBL memory-based learning, and RS-LOCAL local adaption. We also evaluated the performance of spectroscopic estimates in total OC changes for unique treatment combinations that had significant effects on measured OC contents (Fontana et al., 2015; Krauss et al., 2020). These were conventional tillage (CT) vs. reduced tillage (RT), and liquid cattle slurry (SL) vs. manure compost and liquid cattle slurry (MC).

We assessed the quality and performance of the models with the RMSE, R^2, the ratio of performance to the interquartile range (RPIQ) (Bellon-Maurel and McBratney, 2011), and the mean error (bias). Further, for computing and visually comparing the y-intercept (a) and slope (b) of the linear consistency between observed (O; measured) and predicted values (P), OP regressions (x-axis: P; y-axis: O) were used. The same applied for visual displays of assessing the changes in OC over time. Although R^2 is not affected by this swap of PO to OP regressions, a and b are significantly different. There is both formal proof and empirical evidence that the only correct way to present and analyze predictive aspects of the slope without spurious effects, that can confound cor-

rect conclusions, is to do so with OP regressions Piñeiro et al. (2008). For evaluating the best cross-validated preprocessing and modeling combination, we used the RMSE calculated across all hold-out predictions, and aggregated the performance metrics with means across all 5 repetitions of cross-validation.

4.3 Results

General performance of local-only models

Among the general approaches tested, PLSR (ncomp = 5) in combination with a first derivative Savitzky-Golay filter with a window size of 23 points (44 cm^{-1}) and polynomial of second degree, estimated SOC best (R^2 = 0.89; RMSE = 1.4 $g\,kg^{-1}$; RPIQ = 4.8; Figure 4.3). For organic C determined with the WB method, the PLSR model (ncomp = 4) with first derivative SG spectra at a window size of 35 points, which corresponded to a resolution of 70 cm^{-1} yielded optimal results (R^2 = 0.83; RMSE = 1.6 $g\,kg^{-1}$; RPIQ = 3.8). Cubist had a slightly lower performance for both the WB and DC methods.

General performance for SSL transfer and localized modeling

The PCA-HDBSCAN procedure with the preprocessed spectra of soil samples from 2002 produced two stable clusters of almost equal size when applied on all data with dry combustion measurements of OC contents (n = 311; $n(\text{cluster}_1)$ = 156; $n(\text{cluster}_2)$ = 155)). Five local transfer and spiking samples for each cluster were selected for optimizing the transformation of relevant spectra and reference measurements from the SSL ($m(\text{cluster}_1) = m(\text{cluster}_2) = 5$), so that the remaining N(cluster) of samples from n(cluster) were the test or prediction set ($N(\text{cluster}_1) = 151$; $N(\text{cluster}_1) = 156$).

The SSL transfer workflow with localized modeling predicted the N = 301 test observations better with RS-LOCAL than with the SBL, with a roughly 1.5-fold increase in RMSE compared to local-only modeling vs. a 1.9-fold increase for the SBL (Fig. 4.3; RMSE(RS-LOCAL) = 2.1 $g\,kg^{-1}$; R^2(RS-LOCAL) = 0.77; RMSE(SBL) = 2.7 $g\,kg^{-1}$; R^2(SBL) = 0.66). The RS-LOCAL approach delivered slightly higher RPIQ values compared to SBL (RPIQ = 2.6 vs. 2.3). Both data-driven and similarity-driven transfer delivered relatively unbiased estimates of OC contents with similar tendency in marginally over-predicting OC contents (ME(RS-LOCAL) = $-5\,g\,kg^{-1}$; ME(SBL) $-4\,g\,kg^{-1}$).

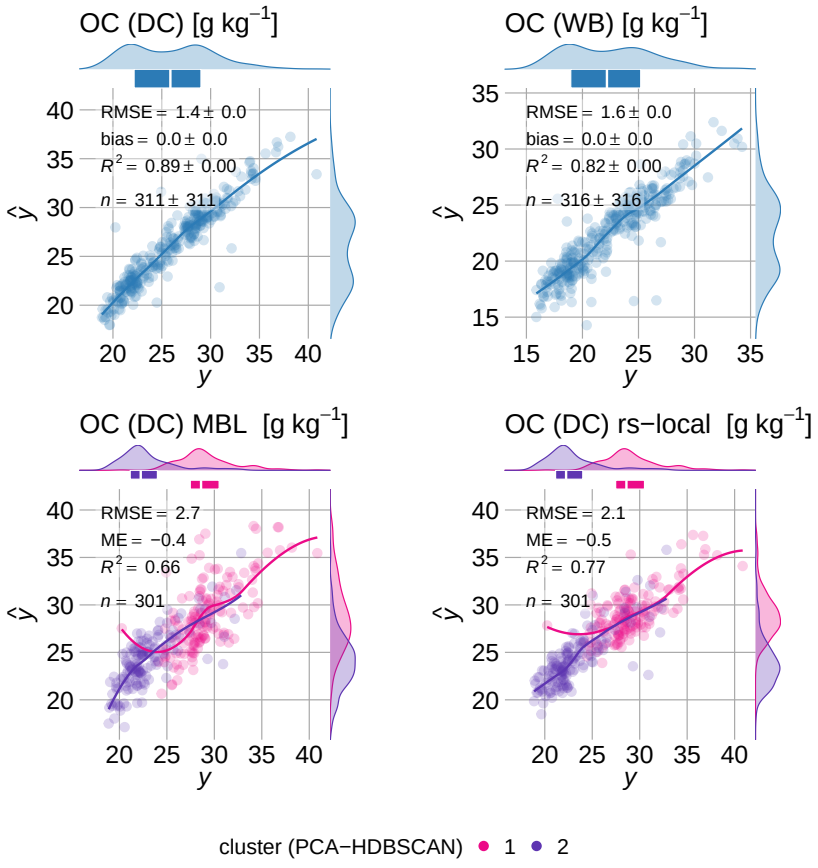

Figure 4.3: Evaluation of best local-only mid-infrared partial least squares regression (PLSR) calibrations (all samples used in calibration; cross-validated) and estimates derived with both performance-driven (RS-LOCAL) and similarity-driven (spectrum-based learner; SBL) selection and prediction workflows (see also Fig. 4.1). *Top:* For local modeling both organic carbon (OC) contents determined with dry combustion (DC; panel *left*; n = 311) and a Walkley-Black wet oxidation variant (WB; panel *right*; n = 316) were used. For both, PLSR was the best approach and performed better than CUBIST (not shown). The OC contents with dry combustion were best estimated with first derivative Savitzky-Golay smoothed spectra (window size – 23 points; polynomial degree – 2) and 5 PLSR components (ncomp), and for OC with Walkley-Black the best performing preprocessing was a first derivative Savitzky-Golay with a window size of 35 points and PLSR ncomp = 4. *Bottom:* Local transfer results with localized modeling workflow done within clusters, which was derived with principal component analysis (PCA) scores and Hierarchical Density-based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm. For both RS-LOCAL and two-stage memory-based learning approaches, we only used 10 spiking and/or calibration samples in total (5 per cluster), and determined optimal calibration subsets with spectra and OC measurements (n = 4244) from the current mid-IR soil spectral library (SSL) of Switzerland (Baumann et al., 2021)

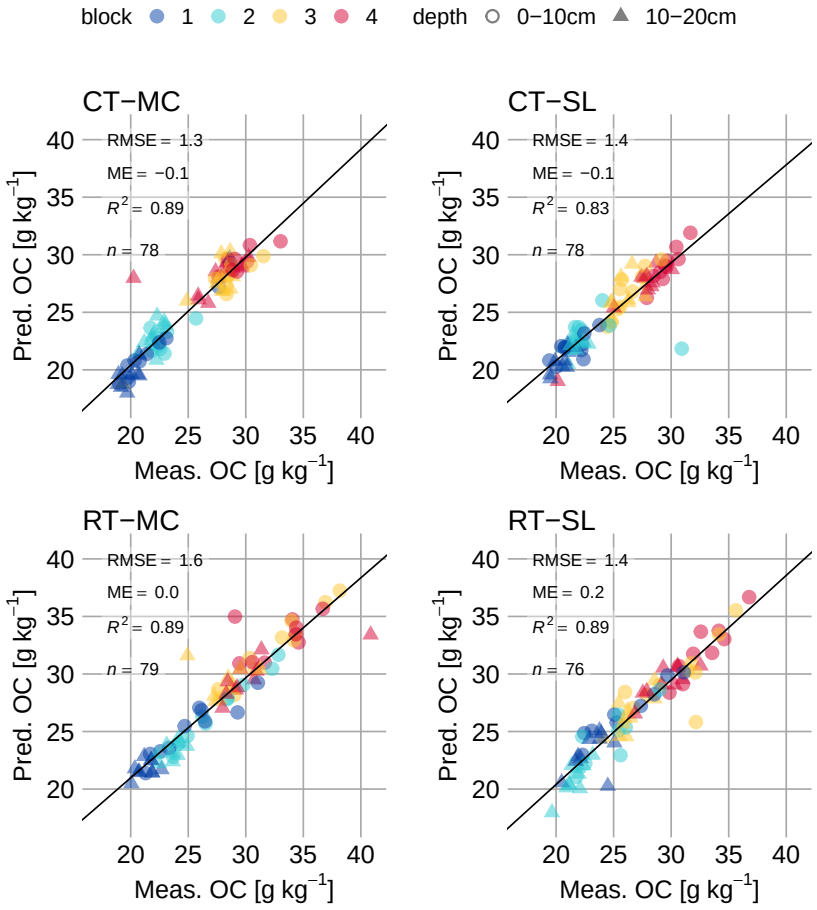

Figure 4.4: Cross-validated assessment of the best local mid-infrared partial least squares regression (PLSR) model of organic carbon (OC; dry combustion) contents measured with dry combustion, conditional on the treatment combination. All points of all repeated measured per year are discriminated by the block number in the experimental trial that they were sampled from. CT is conventional tillage, RT is reduced tillage, SL is fertilization with liquid slurry, and MC is fertilization with manure compost and liquid slurry.

Spectroscopic estimates for carbon trends within treatments and across individual plots over time

The spectroscopic prediction errors from the best performing local-only PLSR modeling were relatively similar for the four main treatment combinations (CT-MC; CT-SL; RT-MC; RT-SL; RMSE 1.3–1.6 $g\,kg^{-1}$). These treatment combinations showed significant impact on levels of OC across the years (see also Berner et al. (2008); Krauss et al. (2020)). For the best local-only model (n_{cv} = 311; RMSE = 1.4 $g\,kg^{-1}$), the trends in measured per plot-depth level changes in OC contents of the different time points were relatively well reproduced from the mid-IR-PLSR estimates (n = 665; range($\Delta OC_{meas.}$) = −12.5–11.7 $g\,kg^{-1}$; R^2 = 0.62; RMSE = 1.9 $g\,kg^{-1}$; Fig. **??**).

For local-only PLSR, the median and mean estimates reproduced the measured plot-depth changes between 2018 and 2002 (Median($\Delta OC_{2018;2002}$) = 1.5 $g\,kg^{-1}$; Mean($\Delta OC_{2018;2002}$) = 2.1 $g\,kg^{-1}$; n = 58) with an absolute deviation of 0.1 $g\,kg^{-1}$ and −0.2 $g\,kg^{-1}$, and a relative deviation of 9.4% and 6.4%, respectively (Table). The absolute deviations between mid-IR PLSR estimates and measurements for the first and third quartile were only slightly higher (0.4 $g\,kg^{-1}$; 0.3 $g\,kg^{-1}$).

Spectroscopic variability and mechanisms of detecting changes in soil organic carbon at field and plot level

Samples with higher soil OC contents tended to exhibit constantly higher absorbance values in the untreated mid-IR spectra. (Fig. 4.7). The clustering into two distinct absorbance curvatures was pronounced in the preprocessed spectra, and even more so in the exploration of the PCA score space of the first four loading axes (Fig. 4.6). Many of the samples with generally lower spectroscopic baselines had the C=O double peaks around 2520 cm^{-1}, while samples with high carbon contents had more distinct peaks at the C−H stretch absorption lines (around 2920 cm^{-1} and 2855 cm^{-1}). In addition, samples with less OC and more IC had more narrow (see preprocessed spectra; Fig. 4.7) and generally lower absorbance peaks at the three consecutive peaks between around 1950 and1750 cm^{-1}. In the organic fingerprint region, visual differences were most pronounced between 1500 and 1250 cm^{-1}

For the prediction of OC contents with PLSR, the eleven mid-IR variables (wavenumbers) of highest importance were distributed in five main regions (Fig. 4.7). These were at 2966 cm^{-1}, 2982 cm^{-1}, 2998 cm^{-1}, 2950 cm^{-1} (region 1; VIP = 2.1–1.6); 1750 cm^{-1}, 1766 cm^{-1}, 1734 cm^{-1} (region 2; VIP = 1.7 –1.5); 1654 cm^{-1} (region 3; VIP = 1.4); 1478 cm^{-1} (region 4, VIP = 1.4); and 3574 cm^{-1} (region 5; VIP = 1.4).

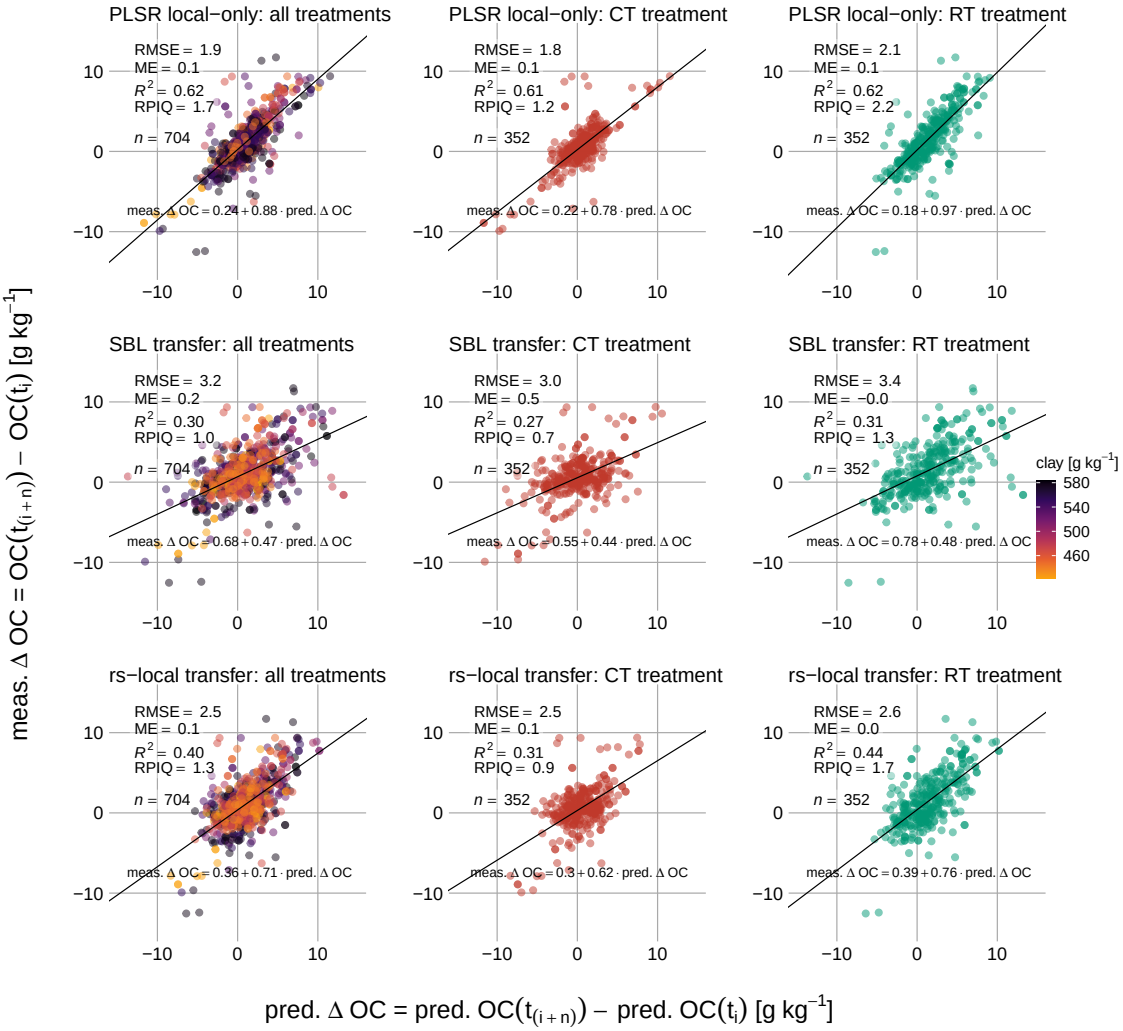

Figure 4.5: Differences in organic carbon contents (ΔOC) of dry combustion measurements (meas.) vs. mid-infrared preditions (mid-IR; pred.) of the Frick long-term trial within defined plot-depth sample pairs across consecutive years: values at $t_0 := 2002$ to $t_4 := 2018$ (t_i) were substracted from later time points of sampling ($t_{(i+n)} := t_1$ to t_4; $i = 0$, $n > 1$; see Fig. 4.2). Three mid-IR estimation strategies were tested. *Top* panels: partial least squares regression (PLSR) with cross-validation on the entire local training set (*PLSR local-only*). *Middle* panels: similarity-driven transfer from the current national mid-IR soil spectral library (SSL) of Switzerland (Baumann et al. (2021)) with the spectrum-based learner (SBL) in a cluster-wide two-stage dissimilarity approach including local spiking (SBL *transfer*; $m = 10$ local spiking; see also Fig. 4.1). *Bottom* panels: Data-driven transfer with RS-LOCAL conducted within the two density-based clusters ($m = 10$ local adaptation and spiking samples). *Left* panels: assessment over all test data of the Frick long-term trial; observations are colored by their soil analytical clay contents. *Middle* (horizontal) panels: conditional model assessment on ΔOC within the conventional tillage (CT) treatment. *Right* panels: conditional model assessment on ΔOC within the reduced tillage (RT) treatment. The RMSE is the root mean squared error, ME is the mean error (bias), R^2 is the coefficient of determination of the linear regression of measured against predicted values. The black lines represent the linear fits, that are annotated by their respective equation with intercept and slope.

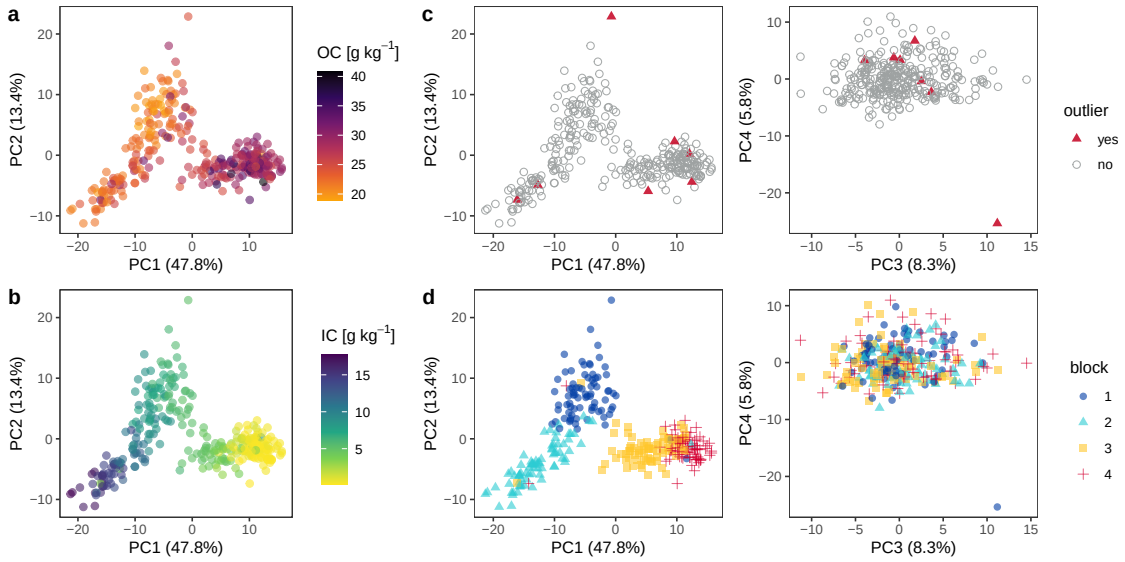

Figure 4.6: Explorative analysis of the mid-IR variability of the samples of the Frick trial. The score plots of the principle component analysis of the preprocessed spectra of the best model training for organic carbon (OC) show characteristic patterns in soil biophysical properties and error-based outliers.*a:* PC2-vs-PC1 colored by measured organic carbon (OC) contents. *b:* PC-vs-PC1 colored by measured inorganic carbon (IC) contents. *c:* PC2-vs-PC1 and PC4-vs-PC3 biplots of soil samples are marked with the top 5 residual errors with a mean absolute deviation > 4.2 g kg^{-1}. *d:* PC2-vs-PC1 and PC4-vs-PC3 biplots annotated with the spatial blocking from the experimental design.

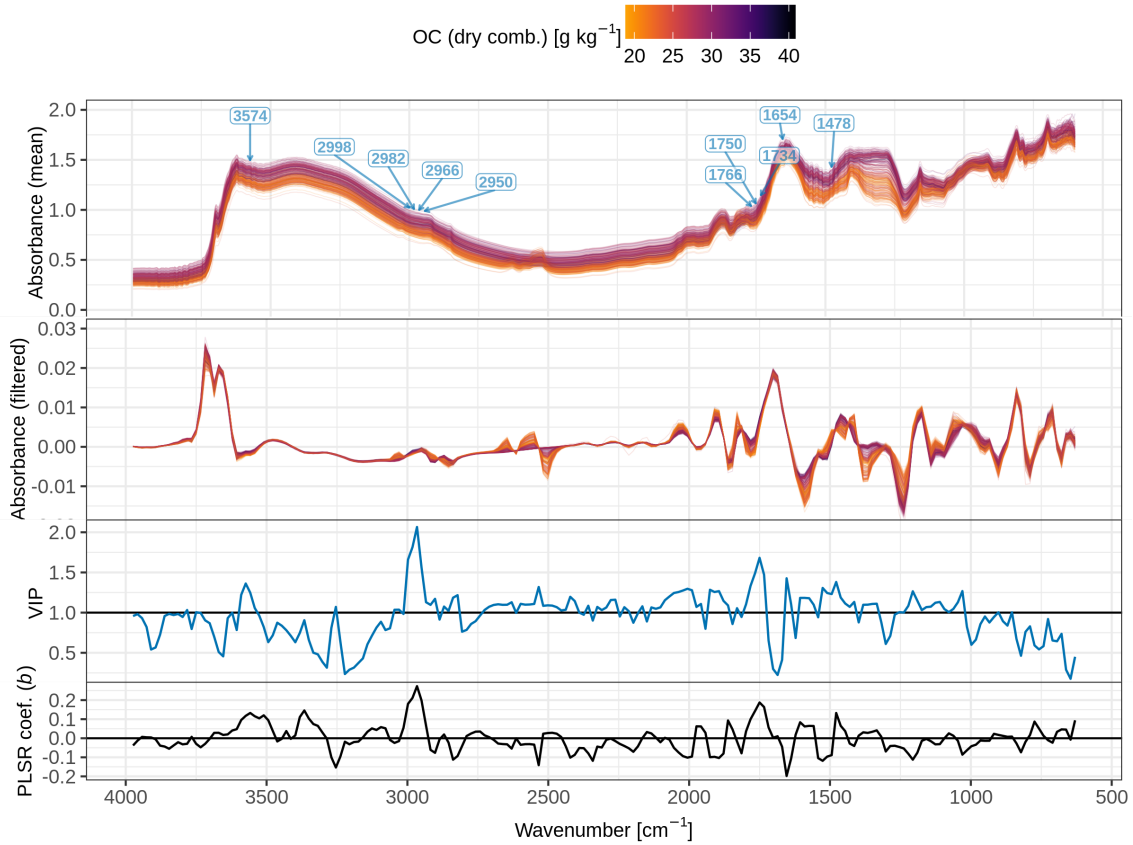

Figure 4.7: Evaluating mechanisms for estimating OC contents and their changes over time. Mean absorbance mid-infrared spectra, optimal Savitzky-Golay filtered spectra (first derivative; window size of 23 points) used for the final partial least squares regression (PLSR) model of OC (ncomp = 5). The variable importance in the projection (VIP) delineates mid-infrared regions that have a high contribution in the PLSR for OC, whereby VIP > 1 indicates that a mid-IR variable contributes more than the average mid-IR variable does to the prediction of OC contents.

4.4 Discussion

General comparison between LTE-only and SSL-to-LTE transfer workflow calibrations

The local modeling with all available Frick LTE samples (n = 316 (WB), n = 311 (DC)) was motivated by yielding empirically best spectroscopic calibrations under optimal data conditions — which is a scenario for research managed trails because of the extensive measurements of OC and other fundamental soil properties. Here, we were able to reproduce the primarily tillage-induced plot-depth-level changes in OC contents (DC) between the consecutively sampled years (2002, 2005, 2008, 2011, and 2018; 10 combinations of differences) with moderate explanatory and predictive power (R^2 = 0.62 and RPIQ = 1.7), but in almost unbiased manner (ME = 0.1 $g kg^{-1}$). We took maximal leverage from the information content in mid-IR spectra, so that the PLSR model coefficients captured OC in its functional composition at experimental plot, block, time, and depth resolution. The uncertainties in the mid-IR PLSR point estimates that were obtained at all sampling times (n = 311), at all plots (32), and their depths (0–10 cm; 10–20 cm) were 74% smaller than their projections for ΔOC over consecutive differences (RMSE = 1.4 $g kg^{-1}$ vs. 1.9 $g kg^{-1}$). This is an effect of uncertainty propagation, because OC estimates with spectroscopic measurement, reference analysis and modeling errors at two points are used to calculate each ΔOC.

With only 10 representative reference samples for calibration and spiking within the transfer scenario (see also Fig. 4.1), the majority of predictive modeling information was sourced from the harmonized mid-IR SSL of Switzerland (spectra and OC reference measurements). Data-driven selection and calibration with RS-LOCAL worked better for estimating both OC contents and also their ΔC values over time, compared to dissimilarity-based modeling (SBL) (RPIQ(OC; RS-LOCAL) = 3.2; RPIQ(OC; SBL) = 2.6; RPIQ(ΔC; RS-LOCAL) = 1.3; RPIQ(ΔOC; SBL) = 1.0)

Soil chemical variability and interpretation of influential spectroscopic features in local modeling

The spectra of soils with higher OC contents and generally lower contents of IC (carbonates) indicated a clear offset towards higher absorbance values. The soils with an opposite inverse relationship of these two measured compositional properties had the opposite effect, which is an albedo effect, which was pronounced by physical scattering effects in the mid-IR. Soils with higher OC contents tended to have broader absorption features between 1950 and 1750 cm^{-1}, which can be attributed to Si – O – Si features in quartz (Janik et al., 1998), that is predominately found in the sand and silt fractions.

Coinciding with this, we observed lower intensities of C–H double peaks (2920 cm^{-1} and 2850 cm^{-1}; Calderón et al. (2013)), and distinct peaks for the C=O vibration found at 2520 cm^{-1}. This absorption was due to carbonates, probably predominantly from calcite. Overall, the chemical and textural variability and also the ranges in measured OC (18.9–40.8 $g kg^{-1}$) and IC (IC = 0.2–17.7 $g kg^{-1}$) of topsoil in the field of the Frick LTE (0.68 ha) were relatively large for the geographic range. The first PC (47.8 %) strongly discriminated this aspect of soil heterogeneity, whereby low scores indicated low contents of IC, and high scores high IC. In addition, the experiment had a pronounced clay gradient from north (block 4) to south (block 1) (Fontana et al., 2015) — which explains the same structure along PC1 when displaying the block in our spectroscopic explorative analysis (Fig. 4.6). That means that density-based clustering in the transfer scenario effectively separated the spatial blocks of the LTE and reproduced the spatial structure of soil heterogeneity (cluster 1: blocks 3 and 4; cluster 2: blocks 1 and 2).

The WB method is known for variable recovery rates of OC. Usually, the deviations from more accurate DC methods are caused by soil mineralogy (i.e., carbonates; texture) and different oxidation state of OC that confound its correct determination (Meersmans et al., 2009; Gubler et al., 2018; Visconti and de Paz, 2021). The correlative analysis between WB and DC (not shown) revealed that there are non-negligible differences between those methods (RMSE = 4.1 $g kg^{-1}$; R^2 = 0.81); however the regression line had a slope of almost 1 (b = 0.99). We found that the local-only PLSR model for OC determined with WB had only a small performance drop compared to the model with DC (RMSE = 1.6 $g kg^{-1}$ vs. 1.4 $g kg^{-1}$), because the mid-IR spectra and modeling allowed to adapt to the differences in average formal oxidation states of OC in these soils, which is according to Visconti and de Paz (2021) the main reason that results based on wet dichromate oxidation (WB) deviates from those with DC. Because the results were more accurate with OC measurements obtained with DC, we continued the transfer modeling and also the assessment of ΔOC contents at plot-depth level with the modern, de-facto standard DC method.

The analysis of important model features in the final mid-IR PLSR model clearly showed that mid-IR regions with soil compositional links to soil OC contributed most strongly to the prediction of OC in local PLSR (Fig. 4.7). The models were also able to capture a considerable proportion of variance of OC contents that changed between the years for each plot and depth (Figure 4.5). Hence, the spectroscopic information imprints the soil biogeochemical changes and is effectively captured with PLSR modeling. For example, the important feature at 1478 cm^{-1} is in vicinity of 1460–1440 cm^{-1}, which has amongst others, been attributed to the C–H bending oscillations in polysaccharides and proteins, and also N–H and C–N stretching (Calderón et al., 2013). The important spectral variable found at 1654 cm^{-1} can be linked to aromatic C=C and C=O as

well as C=O present in amides (Leifeld, 2006). The second most important region for the prediction of OC contents was an edge feature at 1750 cm^{-1}, which is influenced by the C=O stretch vibrations of the carboxy group located at around 1725 cm^{-1}. Its interference at the right edge of the three consecutive Si–O–Si vibrations explain the edge feature. This most important feature for the local-only prediction of OC was almost identical to the most important feature at 1754 cm^{-1} that explained the prediction of total C in the (global) CUBIST rules developed for the establishment of the Swiss mid-IR SSL (Baumann et al., 2021). These strong model-induced vibrational associations are mostly linked to moieties in prominent soil organic molecules — for example carbohydrates, proteins, or (partly) microbiologically decomposed organic matter derived from cellulose and lignin. We therefore conclude that absorbance signals from aliphatic, carboxyl, and aromatic groups are specifically represented in the local PLSR calibration.

Although a large proportion of important mid-IR variables are directly linked to functional groups of OC, there are also purely mineralogical links from which the best local-only PLSR model learned. These include the signal at 3574 cm^{-1}, which is influenced by clay mineralogy and possibly inter-lattice water, and was one of the top 11 predictors. We also found that the presence of carbonates was negatively correlated with OC because of the negative PLSR coefficients at the characteristic band at 2520 cm^{-1} (Fig. 4.7). Altogether, the mid-IR measurements and the modeling indicate the capacity of mid-IR to predict OC contents. This capacity emerges in direct and proportionally significant mid-IR signals that change in response to dominant absorption features of a complex mixture of compounds that make up organic matter — and not only with signatures from the mineralogy. Hence, our data set and analysis supports that mid-IR predictive modeling works efficiently and robustly for estimating contents and changes at plot-resolution and under considerable functional complexity of OC, although this mingles with a gradient in clay fraction (range = 420–580 $g\,kg^{-1}$) and carbonate contents (range(IC) = 0.2–17.7 $g\,kg^{-1}$) in the soil matrix.

Similarity driven vs. local-performance driven prediction based on soil spectral libraries

The results on transferring relevant samples from the current Swiss mid-IR SSL indicated that mid-IR and data-driven modeling was the better strategy to estimate plot-depth-level differences in OC between the evaluated time differences compared to dissimilarity-driven transfer (Fig. 4.5; RPIQ = 1.3 vs. 1.0). It yielded a 1.5-fold higher RMSE compared to LTE-only local modeling for time point estimates of OC (Fig. 4.3) and it was relatively more accurate at the level of predicted ΔOC with an 1.3-fold in-

crease compared to the local-only approach. However, RS-LOCAL transfer imposed non-unity slope effects (b = 0.71) that were stronger than for PLSR local-only modeling (b = 0.88), but markedly better than the two-stage SBL (b = 0.47; Fig. 4.5). For local-only PLSR modeling and also the RS-LOCAL and SBL approaches, the RT treatment yielded less confounded slopes and improved RPIQ values (RPIQ(local-only PLSR = 2.2; RPIQ(RS-LOCAL) = 1.7; RPIQ(RS-LOCAL) = 1.3).

We were able to do the transfer with only 3% new local reference analyses required to estimate OC for 311 samples of the Frick LTE from plots with different treatments and repeats over the five sampling years. Thus, our workflow for SSL transfer and localized modeling had a really efficient analytical measurement to spectroscopic model error ratio. The approach was moreover sufficient to capture locally-relevant relationships in the SSL, so that the estimation of the remaining samples of the Frick trial was sufficiently accurate and relatively unbiased to reconstruct temporal fluctuations in OC over time at plot level. We can, however, explain the success of the transfer by the unique soil mineralogical composition of the soil at the Frick long-term trial, which is to a large degree already represented in the the current mid-IR Swiss SSL.

Despite the overall closely unbiased agreements between measured and PLSR local-derived estimates in OC contents for time point-plot estimates and also by treatment combination (Fig. 4.3; Fig. 4.4), the mid-IR model tended to over-predict ΔOC contents for soil samples with rather high increases in OC contents, which was most pronounced in the RT treatment. This can be a sign of unaccounted response of spectral regions that change purely due to functional changes in OC compounds. This perspective gates novel perspectives that are probably needed for specifically training functional carbon compounds independently of soil mineral matrix effects (cross-sectional correlations). This aspect is almost completely neglected in current soil spectroscopic research. Furthermore, we made some interesting discoveries within the cluster-based searches of best performing subsets of the SSL selected with RS-LOCAL. Despite that spectroscopic clustering should principally decrease non-linearity in the localized modeling, and RS-LOCAL inherently should select for linear relationships in the latent variable spaces (PLSR), we found considerable bias on the test set after we applied a final PLSR model (RS-LOCAL set). We have also found significant soil chemical diversity for plot level transfer of the Swiss long-term soil monitoring sites (NABO; Baumann et al., 2021), but there we allowed RS-LOCAL to learn in explicitly locally-linear contexts because we did separate transfers over sites at very small extents (10 × 10 m); therefore, PLSR was sufficient. In the Frick data set, however, we could eliminate the majority of bias on the Frick test set by training the spiked final selection ($K + m$) set in CUBIST. Our conclusion is thus that with more temporal applications of soil spectroscopy and potential limitations, concepts of RS-LOCAL can even be further devel-

oped and improved in the non-linear direction of instance-based transfer learning.

We found systematic but only small under-prediction of the actual observed differences between the years (Fig. 4.5), although the ensemble of predictions had negligible bias — which was close to zero. Our interpretation of this is that there is not only direct spectral information on concentrations of organic C compounds but also mineralogical context, e.g. related to carbonate contents or clay mineralogy, that has a correlative relationship to OC contents. The latter contribution is, in contrast to the first one, not expected to change too much in short-term over the duration of the experiment. For under-predictions of the OC differences between the years, the small-scale mineralogical context that helps to predict the OC content for a particular soil sample might be dominant over the spectral response in changing organic moieties; or at least measurable change in mid-IR spectra is not captured specifically in the general model built across all samples. The variable importance analysis is supporting such a granular view on distinct (major VIP peaks linked to groups of organic molecules) and more correlative chemical predictors of OC contents (mineralogy) distributed along the mid-IR spectra.

Accountability of the bias-variance trade-off for estimating soil OC changes with mid-infrared spectroscopy

The application of soil sensing is guided by the principles of gaining more measurements to characterize local soil and environmental conditions. Sensor measurements, particularly visible-near and mid-infrared spectrometers resolve the chemical and physical composition and those attributes that are strongly linked to them, very rapidly and inexpensively. The lower accuracy of spectroscopic modeling compared with classical laboratory reference measurement methods is compensated by many more spectroscopic measurements and estimates of multiple soil attributes (speed and low cost) through the predictive workflow. Large spatial and temporal heterogeneity in soils that arises from inherently complex interactions of soil, environmental (biotic and abiotic), land use and management factors, demand a huge scaling-factor in soil sampling and determination of soil properties Saby et al. (2008). We therefore believe that local-only calibrations, that need relatively high data densities in locally-linear soil contexts for every new application (typically from 30 to 150 samples with reference measurements), are not the main way forward for verifying sustained and sustainable soil use and soil protection across large geographical extents. There are specialized applications (i.e., experimental soil research), where this approach is clearly useful. However, the majority of societies' stakeholders that have a high demand of quantitative soil data across landscapes are mostly interested in improvement of locally unbiased transfer meth-

ods. This is because expanding SSLs continuously with diverse soils reduces the dependence on calibration samples measured with traditional lab methods.

4.5 Conclusions

Until now, despite the significant advances in soil spectroscopic research over the last two decades, there has been little direct proof-of-evidence that soil spectroscopy modeling can significantly enhance and support OC monitoring through detecting management-induced temporal changes in OC for long-term field experiments, via plot level transfer of predictively useful information from large SSLs. With this work we deliver conclusive results that confirm that we can predict and explain a significant amount in the trajectory of changes of OC content at plot level, that are on average slightly higher than the typical estimation errors reported in studies with high-density local calibrations (field, farm, and regional level) and many lab reference measurements. A detailed, site-local, spatially explicit and temporally integrated predictive analysis in combination with a diverse SSL of medium size at national extent is a novelty. We found that we only needed very few additional OC reference measurements because these local calibration and augmentation samples ($n = 10$) could be optimally used. The results were significantly more accurate for completely data-driven modeling with RS-LOCAL compared to similarity-driven modeling with the SBL method, but relatively unbiased for both localized modeling approaches across all data. Nevertheless, we also still found considerable non-linearity aspects in transfer sets selected from the current mid-IR SSL of Switzerland, which we addressed with CUBIST. Further, non-unity slope effects, possibly due to the mineralogical context, need to be taken into account in the assessment of OC changes. Our work brings first and foremost credible evidence to the soil science community that spectroscopy with diverse SSLs and localized modeling workflows can not only scale-up to many more measurements of OC over spatial extents, but is also relatively effective in addressing the temporal component. Moreover, it remains that we need to update SSLs continuously to allow for adaptation in new pedological contexts, and also create locally-linear knowledge (increased sample density). With this, we hope that infrared spectroscopy can contribute to soil measurement/-modeling, verification and reporting initiatives, and still further support experimental research in soils and farming.

Part 5

Research findings and conclusions

Chapters 1 and 2 tackled the development of regional and country-level mid-IR SSLs. The first chapter was about establishing a relatively small SSL for project-specific agronomic purposes but across four very diverse soil agroecological zones in West Africa (Ivory Coast and Burkina Faso). The project landscapes spanned a climatic gradient between humid forest and northern Guinean savannah. This SSL was constructed for soils with yam growth potential and targeted the assessment of regional-level variation in properties relevant to soil quality and nutrient supply. The second chapter targeted a considerably larger and more variable soil samples across Switzerland for the establishment of a soil mid-IR data and model collection with legacy data. The Swiss mid-IR SSL is the compilation of country-level soil variation at gridded locations of the Swiss biodiversity monitoring (BDM) and the agricultural sites from the Swiss long-term monitoring (NABO). The Swiss mid-IR SSL was first established in regard to local monitoring and soil survey purposes, but also to build soil research capacities in general. Specifically, this work involved developing general rules for 16 soil properties and also showcasing the potential of local adaptation with relevant samples from the SSL for estimating temporal trends of soil C for each of the 71 monitoring sites individually. Finally, in chapter three I had the opportunity to test the compiled national mid-IR SSL as a data source for transferring both targeted, spatially and temporally explicit information for the prediction of samples of an LTE field at plot level. In this chapter, I also included a detailed analysis of the uncertainties and the soil conditions. Specifically, I tested the scaling possibilities of mid-IR workflows to estimate management-induced changes of OC contents at individual experimental plots.

5.1 Good practice for developing spectroscopic libraries and models

I consider the careful study of properties of statistical data splitting and resampling methods central to the problem of soil IR calibrations (as in any other domain using statistical learning; e.g., Stone, 1974; Varma and Simon, 2006; Friedman et al., 2008). For this, Stenberg and Rossel (2010) provide some background specific to soil spectroscopy, for example on independent validation. The model evaluation procedure the basis for deriving suitable, stable empirical model parameters and producing a generalized evaluation of model uncertainties, at least within the data context at hand. This firstly avoids the injection of auto-correlations that can cause overfitting on data in regression context. Resampling should be further considered at any stage of the model selection process. The resampling scheme — for model parameter tuning, preprocessing, and evaluation — is in my opinion at least as important as getting acquainted with soils and their chemistry together with knowledge about soil formation processes.

According to statistical learning theory, the model error (i.e., mean squared error; MSE) can be decomposed into systematic and asymptomatic error terms (i.e., due to noise or overfit; precision), which are *bias* and *variance* (Friedman et al., 2008). The so-called *bias-variance tradeoff* of the different resampling methods and learning algorithms still seems to receive little attention in the soil spectroscopic community overall. For example, there is good reason to disadvise single test-train splits for model development evaluation in small and highly variable data sets ($n < 100$), and use CV instead. I will later in this chapter elaborate why this also needs particular attention for modeling larger spectroscopic data sets over space and time because of auto-correlation caused in such *profile data*. For now, a simple explanation is that multiple realizations modeling and assessment partitions increases the bias in the assessment metrics through principles of averaging, which lowers bias in those estimates. In fact, the use of 5-fold or 10-fold CV has been encouraged in machine learning and also for sample sparse but high-dimensional data in chemometric calibrations to be specific (Kohavi, 1995; Molinaro et al., 2005; Beleites et al., 2005).

To further reduce — as we have done thoroughly throughout the modeling in chapters one, two and three — the high variance (but low bias) that generally comes with k-fold CV, we can repeat the resampling procedure multiple times. I usually repeat my standard 10-fold CV scheme five times. Further realizations of the training population by splitting the data set in fixed segments therefore produce further means to report uncertainties of the model metrics. If the k-fold cross-validation or other resampling schemes are repeated, multiple holdouts of the same observation can also help to assess the robustness of the chosen resampling and modeling scheme. This version of

repeated CV is in my opinion much more adequate than bootstrapping (Efron, 1983) is for deriving modeling uncertainties, particularly when many more repeats are created in CV. This is because bootstrapping violates the independence of training and test instances, since 63.2% of training samples are present in multiple copies, which is too dangerous high-dimensional spectroscopic applications (Friedman et al., 2008; i.e., over-fit), regardless of sample size (overfit in machine learning). Bootstrap resampling and also the derived out-of-bag estimates (36.8% unique samples) are, in contrast, useful for giving low variance estimates for tuning empirical model parameters (hyperparameters) separately from model selection and evaluation (Efron and Tibshirani, 1997; Kuhn and Johnson, 2013). It would be certainly best practice if the additional test set is entirely structurally independent from the soil training samples (Stenberg and Rossel, 2010).

From a soil spatial modeling perspective, a complete independence of validation from the training can however only be guaranteed when making another probability sampling and design-based inference; thus, by collecting additional soil descriptive and measurement data that is independent of those sampled as inputs for mapping (Brus, 2019; Wadoux and Brus, 2021). Besides, spatially stratified k-fold CV has been advised in environmental mapping, but it has been recently challenged due to its controversies and its use has even been entirely disadvised in pedometrics (Wadoux et al., 2021). The enforcement of vast geographical separation in clustered assessments is believed to disrupt context-dependent learning of small-scale local spatial relationships that occur within the typical ranges of the variograms in soil landscapes (Wadoux et al., 2021). This is in my opinion indeed a valid point made, however, there is no comparable work for soil spectroscopic analyses yet. I think its useful to compare those aspects of soil infrared modeling evaluations with aspects of uncertainty in mapping with the factors of soil formation in its multi-scale nature (Behrens et al., 2019). At the same time, I believe it is important to further delineate where soil spectroscopic modeling has its particularities and in what points data differs from soil spatial modeling. I will in the following reason about paradigms of chemical complexity of soils in chemical sense. It seems also important to me to have a perspective on spectroscopic information contents with regard to mid-IR in statistical learning. In chapter one, however, no precautionary grouping measures in CV were necessary, because the measurements used for modeling were from one pooled soil sample originating from spatially distant and biophysically relatively independent properties of farmers' fields (80 smallholder farmers, one sample per field of max. a few hectares in size) in the four landscapes.

Because of the purely chemical and physical aspects of soil composition seen from the lens of spectroscopy, the intensity responses in infrared spectra are to a certain degree context-independent. For example, the same clay minerals can be present in

different soil types. Therefore, soil spectra also at least partly transferable for univariate responses of interest (measured properties), across different sets of soils. I refer to this as *inter-soil knowledge*. This type of knowledge is the basis that transfer learning approaches and intelligently designed workflows can address the useful knowledge often better for new target soil domains than training one general model over all data of an SSL. The matrix of minerals and a multitude of overlapping vibrations of important functional groups in molecules builds the backbone of shareable and transferable *inter-soil knowledge*. Even though soils are probably the most complex media on earth, infrared fingerprints of its individual components (e.g., sand, clay minerals, functional groups of organic molecules in SOM) are characteristic and relatively unique. Hence, I hypothesize that the compositional information content in spectra that can be transduced by modeling across soil samples from different geographical locations is generally higher for soil infrared spectroscopy than for simulation modeling of soil processes across landscapes on very variable scales. Thus, this makes the spectroscopic modeling approach also more susceptible to over-fit than soil spatial modeling in DSM.

My most important concern for statistical resampling in the realm of soil spectroscopy is, therefore, that we should consider spatial or temporal auto-correlation. We should do so particularly when the training data, for example the SSL, contains repeated measurements at the exact same sampling location or plot and over time. This can be soil sampled in proximity multiple samples in depth along the soil profile or also repeated measurements over time. I have experienced such particular situations that require attention in the modeling process, for example the BDM set (1094 locations each with sampling radii of 3–3.5 m from a grid of 6 × 4 km) but also the long-term soil monitoring sites of Switzerland (NABO set; 71 sites; subsampling from 10 × 10 m area) have local replicates with measurements every 5 years since 1985. Thus, for deriving and testing general rules on all these different, partly overlapping, sources of training data with principally high degree of short-range local and temporal autocorrelation, we used a grouped 10-fold CV approach. By considering the site identifier as a grouping information, I could entirely avoid that local spectroscopic information on the modeled response would otherwise be used in training and testing at the same time, which would inarguably result in over-optimistic model assessment for rather adaptive non-linear methods such as CUBIST, but also PLSR. In chapter 3, I followed the same line of argument when defining the grouped CV, but this time I have done this by the experimental plot. This also allowed to tackle the probably strong cross-correlation between the two depths within the plowing pan (0–10 cm, 10–20 cm). In summary, I recommend to follow two main aspects of resampling and testing for good practice in spectroscopic modeling:

1. Identify the independent experimental unit in the data set. Consider the group

hierarchy that identifies this unit and use the labels for constructing resamples where all instances of a particular group are assigned to either analysis/modeling or testing sets. Depending on the group size, this is also known under group-leave-out CV.

2. Tune, train, and select best statistical models with a 5-fold or 10-fold CV with grouping identified above, and repeat the procedure at least 5 times (i.e., a total of at least 50 or 100 pairs of training and validation data partitions). If possible, use an independent test set in addition if the modeling scenario is more complex and involves several data-dependent steps (e.g., certain preprocessing methods that depend on the remaining data). To make best use of all data and further separate the tasks of model selection and assessment in the data, use nested CV (also called double CV; if computationally feasible; see Varma and Simon, 2006; Filzmoser et al., 2009; Krstajic et al., 2014; Bates et al., 2021).

5.2 Limitations of soil IR spectroscopy and opportunities to be seized

Soils are very complex in their chemical composition, and therefore the modeling of IR spectra can be a complex task for larger libraries. The challenge is further that the target soil properties can be indirectly or directly related to spectroscopic signals. There are also physical scattering effects that reduce the information content and complicate statistical prediction, which are at least partly mediated by preprocessing (Barnes et al., 1989).

Both PLSR and CUBIST have excellent prediction and interpretation capacities with accessible variable importance measures (e.g., Wold et al., 1993; Chong and Jun, 2005) for linear and non-linear data contexts, respectively. Soil IR spectroscopy in combination with these methods are generally more reliable for properties that are strongly related to SOM and the more distinct attributes of the mineralogy (e.g., characteristic IR vibrations of clays). A further requirement is often that the soil data contexts and learning workflow need to match the data distributions in the new application as well as training contexts (Shen et al., 2022). Mid-infrared spectroscopy comparably better discriminates signals of SOM compounds from the mineralogy than NIR spectroscopy because it contains the fundamental vibrations that are more sharp and less overlapping (Madejová et al., 2017).

In chapter one, a purely local PLSR modeling approach with pooling all field samples across the four regions was successful and not expected from the literature, because of major biophysical differences in inherent soil fertility in the four regions. The tropical

soils that we used for calibrations — the four regions are known for presence of Lixisols, Gleysols, Leptosols, Vertisols and Ferralsols — confirmed that those properties that are well linearly correlated to SOM and some chemical properties related to mineralogy (e.g., total C, N, S; $CEC_{eff.}$, pH, Fe(DTPA), Cu(DTPA)) can be estimated with reasonable accuracy ($R^2 > 0.6$). Hence, we concluded that the established small and targeted SSL is a first foundation for screening the variation in soil fertility for new locations in the regions. Nevertheless, despite that the calibrated models gave accuracy that are typical for small-area calibrations (i.e. field of several hectares), it would probably need small but frequent updates. I suggest to test RS-LOCAL for this purpose because PLSR worked well.

In weathered tropical soils with yam-growth potential (chapter 1), DTPA-extractable Fe and Cu ($R^2 = 0.77$ and 0.74), and also $BaCl_2$-extractable Ca and Mg ($R^2 = 0.81$ and 0.62; $n = 92$ and 93) were relatively well discriminated with mid-IR spectroscopy and PLSR modeling, while available K ($BaCl_2$) and Zn (DTPA) were poorly estimated ($R^2 = 0.28$ and 0.25; $n = 94$ and 87). A similar tendency was found for the Swiss mid-IR SSL (chapter 3), where the rules developed on the available nutrient contents (AAE; $n = 417$) of a subset of BDM sites showed excellent and good results for extractable Ca and Mg ($R^2 = 0.97$ and 0.79), but they were unreliable to estimate extractable K, Cu, Zn, Fe and also P ($R^2 = 0.10$; 0.10; 0.06; 0.28; 0.05). In this regard, the tendencies to predict agronomically important extractable elements for plant nutrition were comparable for chapter one and two, despite the very different weathering status, size of SSLs, and sampling extents and densities. For the calibrations with collected samples from yam fields, measured total C had a strong positive relationship with measured exchangeable Ca and Mg ($r = 0.9$ and 0.8), while we observed weaker relationship to the measured clay content ($r = 0.6$) and weak negative relationship to sand contents ($r = -0.3$ and -0.2). Terhoeven-Urselmans et al. (2010) have reported similar validation success in terms of linear agreements between mid-IR predicted contents and measurements of exchangeable Ca and Mg measured by ammoniumacetate (NH_4OAc; pH < 7) and natriumacetate extractions (NaOAc; pH > 7) ($n_{cal.} = 679$; $n_{val.} = 292$; $R^2_{val.} = 0.68$ and 0.61, respectively). Because we have used identical soils and sample numbers from the same set of grid sites in Switzerland ($n = 417$) and non-linear modeling for different plant relevant macro- and microelements, I argue that the general relationship made about sample size and R^2 of the mid-IR evaluation by Sanderman et al. (2020) does not hold true in Switzerland. We found that extractable Mg and Ca were performing well in the SSL while other key plant macro- and micronutrients did not; their statement might therefore be confounded, simply because these other micro- and macroelements are less directly (linearly) correlated to SOM, and the mineral matrix that could otherwise provide such links is vastly complex. Because microelement extraction analyses are

more laborious and expensive, they might have been less readily accessible to SSLs. Rather than talking about requirements of more samples (size of SSL) in general, I think we should be more honest about soil spectroscopy being limited in it's capacity to predict available P, K, Zn and Cu, whose values involve dependencies to complex surface chemistry and exchange processes in the soil solution (extracts) (see also the review by Soriano-Disla et al., 2014).

Because of the aforementioned complications as well as the generally high errors (variance) in soil extraction reference methods, combining general modeling and IR spectroscopy alone may be often not an adequate strategy to model extraction-based soil nutrients. For this, it is arguable whether we gain a much better understanding about soils using this IR technology alone, or if we better focus on making soil IR spectroscopy more efficient for the domains it is successful already (i.e.,pH, OC, texture, and cation exchange capacity, carbonates, contents of specific clay minerals, selected total elements), and couple IR spectroscopy to other spectroscopic approaches and modeling workflows to deepen the understanding about the often hidden complexity (X-ray diffraction analysis, portable XRF, laser-induced breakdown spectroscopy (LIBS), etc.; see e.g., Hillier et al., 2016; Butler et al., 2020; VillasBoas et al., 2020). If there are prospects in spectroscopic approaches, then I think we probably need either 1) locally linear approximations — which comes with the extensive requirement of sufficient local sample and sampling densities to cover same soil mineralogy and a functional gradient in SOM in the neighborhood of future samples, or 2) effective transfer learning approaches (Pan and Yang, 2010; Lobsey et al., 2017). I hypothesize that the latter approach is the cheapest and more promising one, although we further need to enlarge SSLs in this regard for sure.

In chapter three, I conducted on of the first detailed studies on model-spectroscopic diagnosis of changing OC contents over time in an LTE under organic management together with a large country-level SSL. This taught me that a diverse SSL can save a lot of new reference measurements. The approach was still relatively accurate in comparison to purely local calibrations with many more samples. At the same time, when we deploy mid-IR spectroscopy in such an adaptive learning workflow, it seems that the proportion of changes in OC we can explain with spectra in function of selective responses to the functional diversity of SOM is limited. Here, the results suggest that the direct information content of the spectra on SOC becomes smaller with strong gradients in the mineralogical composition, due to relatively short-range, local, soil variation. This observation at a single location can either mean that this challenges the current approaches we take for spectroscopic modeling, or that it would be useful to diversify the current mid-IR SSL of Switzerland. I take these comprehensible findings as lesson that we can improve SSLs even further by updating them, and that the con-

ceptual modeling approach, i.e., how we present data to models to learn from specific changes in spectra and specific organic constituents, matters, and needs innovation.

5.3 Soil grouping, similarity-based, and transfer principles for localized spectroscopic modeling

In chapter two, the general estimation capacity of the established mid-IR SSL has been shown for all 16 soil properties, of which 10 have potential for diagnostic screening (RPIQ ≥ 2). The general estimation of total C with CUBIST rules had an average RMSE per site of 3.1 $g\,kg^{-1}$ for the 71 NABO sites, and was substantially more accurate for when using transfer learning with 2 spiking observations (RMSE = 0.7 $g\,kg^{-1}$). Later in chapter three, similar principles of data-driven selection were more successful than memory-based approaches, for integrating the diverse SSL into the assessment of soil OC changes for an agricultural LTE with minimal spiking (RMSE(ΔC) = 2.5 vs. 3.2 $g\,kg^{-1}$). The approach with RS-LOCAL only needed 3% of the analytical measurements of the LTE prediction set to estimate ΔOC of 301 samples over time with only 1.3 times higher RMSE . Our results from localized transfer modeling in chapters 2 and 3 were very close to a range of uncertainties that the literature typically suggests for purely local calibrations at the level of agricultural fields or also areas of farms (few hectares to km^{-1}).

Both chapters 2 and 3 conclude that a data-driven way of selecting suitable instances can utilize the knowledge stored in the SSL in the most efficient and universal way possible in terms of new applications. Data-driven ways of selecting optimal samples, i.e., RS-LOCAL, narrow down soil chemical and physical complexity so that the relevant compositional aspects are filtered with regard to a new target application. This approach is not constrained by spectroscopic and soil compositional similarity per se. Thus, the selected samples are not necessarily restricted in the multivariate (spectroscopic) feature space, so that the relevant part of soil chemical diversity of an SSLs remains for estimating soil samples in new locations. Memory-based learning, in contrast, enforces strong compositional and chemical relatedness of the prediction samples and the selected subsets of the SSL, which limits its accuracy if there is not not enough samples in the library that resemble the prediction samples chemically. I note that performance-driven instance-based transfer learning emphasizes the *inter-soil knowledge* more than other popular approaches for spectroscopic modeling currently do. In chapter 2, I found evidence for leverage effects in the final regression model based on the RS-LOCAL set. That means that the prediction set is fitted by means of a relative wide training set (local samples and SSL subset), which is then also evident in the predicted vs. observed plots. Hence, data outside of the measured range in the prediction set are rather influential. The data distributions in the finally selected subsets

from the training population had a much large range in carbon and also diversity in the spectra, compared to the target, the test population. RS-LOCAL is currently a method of its own, with regard to both the overlapping machine learning and chemometrics fields in general, and soil spectroscopy in specific. Its working principles heavily draw from repeatedly sampling observations from the available training set and optimizing the off-training set error on a few local spiking samples by steadily dropping observations that are not matching in performance.

The grouping by texture (i.e., Moura-Bueno et al., 2020) is popular for showcasing the representativity of an SSL for soil landscapes. These laboratory analytical data is often only available in research scenarios, hence it may not be useful for model subsetting. In fact, in applied contexts of IR spectroscopy, prediction samples (with exception of local adaptation (calibration) and validation samples) have probably no laboratory results on texture, which are rather predicted, too. In my opinion, subsetting an SSL for better predictive knowledge extraction with human domain knowledge is also less optimal than algorithms can do (in data-driven manner from spectra) for novel soil environmental contexts. Here, best possible predictions from as few new reference measurements and adaptive training workflows would be needed. There is various reasons that bring my conclusion forward. First, the mineral and organic soil constituents produce broader, less intense, and more overlapping overtones and combination bands in the NIR (mirrored from the mid-IR) than the fundamental absorptions in the mid-IR. Thus, for soil OC estimation, mid-IR spectroscopy gives generally more information on the chemical composition and is, combined with global modeling, mostly more accurate than NIR spectroscopy for soils. Secondly, which is related to the first argument, there have been also concerns that the concept of grouping by spectral dissimilarity or relatedness for modeling does not always exactly mirror useful chemical constraints for modeling. Modeling by spectral similarity could 1) produce heterogeneous mineralogy in the neighborhood due to nonspecific chemical signals in the NIR (i.e., Reeves et al., 1999; McCarty et al., 2002), and 2) can imply that sufficient training samples are needed in the local (spectral) input spaces for both NIR and mid-IR (i.e., concept of "neighborhoods" or "exemplars" in memory-based learning aka instance-based learning; see Aha, 1992; Wilson and Martinez, 2000b). The third reason to step away from these type of sub-library modeling of established SSL for new projects is because some strata can have highly variable OC, clay contents and mineralogy, and spectra, which decreases performance compared to modeling all data with CUBIST (e.g., Moura-Bueno et al., 2020). Lastly, most of the criteria used for partitioning, except clustering based on spectra in the modeling workflow to reduce non-linearity, are not universally transferable to new applications because either auxiliary data is not available or the characteristics of soils in the new study region is not (yet) covered by the existing SSL. This

is consequently reason to test transfer and update of the soil information stored with new spectroscopic and laboratory information. The results of my dissertation and especially earlier work that has lead to the development of RS-LOCAL (see Lobsey et al., 2017) suggest that data-driven selection makes the concepts of traditional soil grouping (i.e., by geographic region, land cover and use, soil types)— although still advised — fairly obsolete in the context of optimizing spectroscopic estimation for site-specific estimation and temporal soil monitoring.

I conclude that soil grouping in SSLs with clustering by other soil or environmental covariates are good ways to 1) show representativeness of existing libraries with regard to likely future application contexts, and 2) to reveal data gaps. However, we should not use those criteria directly in the modeling — and use rather spectral clustering as part of the workflow because it works with chemically distinct data (e.g, chapter three; also Summerauer et al., 2021). Further, we can also work with model-based partitioning (i.e., CUBIST), or use more elaborate transfer methods, because SSLs need anyway adaptation to work robustly and give accurate estimates in new (uncovered) use cases. This is not always attributed and there is the need to be more specific here. I also think that we currently haven't yet realized the full potential in soil spectroscopy in new contexts. Transfer learning is a rapidly growing field of machine learning (not limited to deep(er) learning only), which allows for adaptation of partial shifts in the marginal distribution of the training domain (source; stored knowledge; SSL) towards a new target domain (new application) (e.g., instance-based, model-based, feature-based; Pan and Yang, 2010; Wang et al., 2019). This will make the extraction of knowledge from larger and chemically diverse SSLs more straight-forward and more universally useful in calibration engines.

5.4 Framing efficient soil spectroscopic workflows

The main motivation to do soil spectroscopic research is to accelerate the collection of soil chemical, biological, and physical information, so that we can harvest possibly all stored relationships in targeted manner. This means extrapolating the relevant part of information for predicting selected attributes and emerging properties. However, one of the main challenges of soil IR spectroscopy and existing SSLs, particularly when they cover a substantial amount of soil diversity, is that the estimation workflow for soils need adaption and fine-tuning with regard to data processing and modeling. This varies depending on expected soil variation within new geographical regions or individual sites of small area, but also depends on the study context, soil conditions, and associated multivariate data at hand. Hence — as the results of my thesis outline — often neither a univariate perspective on global or general statistical learning on all data

points of existing SSLs nor purely local calibrations for particular soil sets are entirely true. Because general learning extracts predictive relationships that minimize errors across all data points based on a loss function (cost function), the models are therefore adapted to the data context and the distributions of the population sampled. Novel samples can be either outside of the calibration range in terms of their properties or the compositional space, or the model does not have parameterized relationships specific enough in defined subregions (i.e. varying local density; non-linearities). Purely local (local-only) calibrations might adapt the prediction for new target soils, but the general benefits of SSLs are not given if new calibrations with analytical measurements have to be constructed over and over again. As I have shown before, transfer learning is maybe one of the main solutions to the problem.

In all three chapters, I have laid careful attention to design, train and test effective workflows that take (hopefully; we can only falsify) maximum information gain from SSLs at minimal reference analyses in the different application contexts. Because spectroscopy does not make the important reference analytical measurements obsolete, efficient soil spectroscopic workflows in real-world scenarios need both adaptation (model transfer) and validation samples. I hereby state that IR-spectroscopic workflows are then efficient across space and time, when the errors in the resulting spectroscopic estimates are low in comparison to the amount of novel local (classical) analytical measurements that are required for new target contexts, specifically in comparison to the efforts that would have been needed to produce the same accuracy by doing purely local analyses and calibrations.

Part 6

Research outlook

6.1 Advancement of a spectroscopy modeling platform towards more spatial and temporal soil information contexts

The BDM data set brought an important source of topsoil variation across Switzerland for future mid-IR estimation workflows, which generally broadens the scope of spectroscopic estimation of soil properties within the established SSL (Baumann et al., 2021). We have shown that this data source together with the NABO set has both general and local estimation capacity for soil monitoring, with minimal requirements of reference measurements in the spectroscopic feature space that it already covers. The SSL in its current version has also produced excellent results for estimating total C in the range of 14–522 $g\,kg^{-1}$, across 26 locations of two regions in Switzerland with drained peatlands in a recent study (Helfenstein et al., 2021). Only five to ten new reference samples were necessary, in addition to the SSL, to predict the 122 samples from different soil horizons with an RMSE of only 3.2 to 2.7 $g\,kg^{-1}$. Because the national mid-IR SSL currently under-represents organic soils in their different decomposition stages and mineral matrices with presently only 2% organic soils (Helfenstein et al., 2021; Baumann et al., 2021), it is clear that we need to expand the library more towards these organic soils. Both the general models and localized models for OC and total C, respectively, were more inaccurate for the higher ranges of measured values than the lower ones where more data are available, because the reference measurements in the SSL were positively skewed with regard to those measurements (Helfenstein et al., 2021; Baumann et al., 2021).

I believe that the established mid-IR SSL of Switzerland has more potential to be ex-

tended with soil chemical and physical diversity. If it is updated further, it can represent the soil ecological contexts in Switzerland better than a finite, coarse grid-sampling approach. The soil mapping in Switzerland, which is now of higher importance in the political agenda, is a promising application of mid-IR spectroscopy, that feeds new training relationships into the SSL. It can thereby also greatly profit from an SSL and modeling infrastructure. The current and historic speed of soil survey has not been sufficient to keep up soil information with the spatial planning demands and soil protection measures in Switzerland, and also anywhere else in the world. We currently, mostly, have topsoils in the mid-IR SSL of Switzerland. Since the mineral complexity is much higher in deeper layers of soil profiles, the SSL needs to account better for the depth component. The main focus of this work was on agricultural soils, whereas we can also expect enormous soil complexity in forest soils. Forest soils span much wider ecological zones on diverse geological substrates and large-scale gradients in environmental controlling factors (i.e., expanding to the alpine altitudes). In a hybrid form of classical and digital soil mapping, my hypothesis is that mid-IR spectroscopy with continuous updates of and learning from the SSL can replace 95% to 99% of the classical reference measurements for soil pH, OC, and texture. With that, we can replace the outdated texture by feel. Further, a well-maintained, quality-controlled, spectroscopic measurement and fully automated modeling system (i.e., pre-processing, calibration sampling, adaptive statistical learning, outlier analysis, model-based interpretation), including a centralized laboratory that provides sufficient but minimal, representative validation and measurements for model adaptation, will deliver a drastic throughput increase. With a high-throughput module, a trained person can measure up to 200 milled soils with two replicates each on a mid-IR spectrometer. Because a lot of capacities will be needed to sample soils on the field and process them in the laboratory (drying, weighing, crushing, sieving, milling) before measuring their IR spectra, future soil mapping in Switzerland (and elsewhere) will not only rely on mid-IR spectroscopy in the laboratory, but also on other proximal sensing approaches. Besides the NIR-range, which will be useful because it needs no milling, many sensing technologies within the electromagnetic ranges of radiation will be probably included in the classical measurement, modeling and estimation workflow. This will include both, sensors on field, and in the laboratory. Hence, we need to develop a tightly linked and modular set of components that have their own scaling-up role. The emerging diagnostic soil information system will not only be data-driven to make best use of SSL, but also be oriented towards pedological knowledge integration and expansion (soil forming factors). It will also reflect the current state-of-the art knowledge in soil chemistry and soil physics.

Spectroscopy-enhanced grouping of soils is a way of augmenting the purely qualitative pedological classifications systems. They can, besides information of the environmen-

tal and management conditions (i.e., climate, land cover and use, vegetation) and terrain, profit from a higher amount of quantitative characteristics of soils' composition and their attributes (e.g, Dotto et al., 2020). Quantitative soil grouping systems that are oriented towards IR spectroscopic fingerprinting have been found to produce heterogeneous clusters that can contain multiple soil classes (Vasques et al., 2014; Dotto et al., 2020). This calls for alternative concepts such as soil series — as practiced by the Soil Survey staff of USDA, which is mentioned in Dotto et al. (2020) — based on hierarchically structured resemblance in landscape, climate, and soil characteristics. Purely spectral information, for example collected in the NIR range, also showed promising results to quantify soil types using unsupervised and supervised data classification methods Xie et al. (2015). However, depending on the complexity of present soil forming factors, the classification level input for spectroscopic modeling for soil surveying and mapping purposes needs to be adjusted according to the affordable analytical measurements, the soil variation detected in the spectra. The current Swiss mid-IR SSL seems already quite useful for many soil properties related to soil quality, but we do not have the complete soil (profile) description, soil types, and morphological descriptions within the current Swiss classification system, so that we can fully estimate its representativeness for pedological landscapes within the country. With the prospect to include legacy soil collections of past soil survey campaigns, and also forest inventories, together with the incorporation of soil-landscape relationships (i.e., digital soil mapping with remote sensing to capture soil variation on systems level; e.g. Stumpf et al., 2018), I hope that we can arrive soon at a more complete assessment of soil variability with IR spectroscopy. At the same time, we can also make informed decisions to choose more, novel systematic soil monitoring sites. We will be able to do this based on relevant soil covariates and chemical information in the IR spectrum.

6.2 Detecting the model-spectroscopic response in soil quality attributes due to management changes

Spectroscopy performs relatively well on single time point estimates, to keep track of soil variability across broad geographical extents, and also to link the biogeochemical controls of soil OC. There are high complementary and scaling possibilities of spectroscopic measurements and data mining for estimating spatial patterns in soil quality attributes. This is useful to derive more data in classical soil surveying, and in digital soil mapping. The potential has been explored and realized relatively early on, but mostly in research scenarios because its capacity needed empirical evidence in the soil research community **?**. We are about to make first good steps to getting SSLs, ensembles of different measurement technologies in the lab and in the field, and sophisti-

cated sensor data processing and predictive modeling ready, so that they can effectively address soil biophysical changes at local small-range soil variation. I hypothesize that soil spectroscopy will soon expand further the time domain. This will be of utmost importance to cost-effectively and quickly describe the soils' chemical variability in many soil forming contexts. Moreover, to meet the growing demand of verification of sustainable management practices in soil to enhance soil quality and possibly to soil C sequestration under favorable soil conditions, such a scaling up with proxy tools and also further modeling (i.e., biogeochemical modeling) integration needs to be made.

6.3 The lesser known foundations of performance-driven learning and its conceptual extensions in future

Performance-driven learning by the RS-LOCAL algorithm works by discarding those samples from the SSL that are on average not performing well on the usually small set of site-specific samples with analytical measurements. Each stage of library reduction deploys extensive sampling of the library and modeling with an implementation of PLSR, which is based on linear latent projections with the SIMPLS method (de Jong, 1993). These linear restrictions are therefore important to consider when aiming for deriving generic calibrations and hence accurate prediction in highly non-linear data sets. Without further partitioning the spectral space and modeling in locally-linear subsets of the data, a linear method typically performs poorer on complex data sets. In the RS-LOCAL selection procedure, those samples that have consistent, towards linearity oriented loading vectors that produce, are consequently kept. That means that training set scores should get linearly oriented towards the locally projected scores (m samples). However, in my opinion, this seems ot be the ideal behavior of RS-LOCAL because of its working principles. In particular, the procedure iteratively removes a specific proportion of non-linearity in those features that are part of the respective PLSR loading vectors at the decisions of training drop-out (related to concept of off-training set error). However, there are situations where the final RS-LOCAL set shows substantial heterogeneity in soil conditions and is hence better modeled with non-linear methods such as CUBIST.

The concept of multi-scale features is an important aspect in the scale-space theory (Lindeberg, 1994), which integrates pieces from signal processing, information theory, and computer vision. For spectroscopy in general, and soil spectroscopy specifically, inner scales map to the resolution at which peaks of distinct vibrations of functional chemical groups start to appear, while outer scales map to the energy range (window; wavenumber range; wavelength range) that completely contains the spectral feature or groups thereof. In analogy to geographical maps and the euclidean concepts of dis-

tances, a holistic analysis of structures requires the ability to increase and decrease the inner scale of the observation depending on the situation and knowledge to be generated — to simplify zooming in and out. Linear filters in computer vision perform the task of selection by preprocessing of image data as a first step in object detection, which often needs non-linear operations later, while the analogous job that RS-LOCAL does is linearization and feature selection as part of the modeling and reduction of the library, which is not constrained by (dis)similarity and multivariate distance.

In chapter three, I first found substantial bias (RMSE(OC) $\approx 2\,\mathrm{g\,kg^{-1}}$) on the test prediction of the Frick trial (301 samples), when I applied the trained model on the best RS-LOCAL selection with local spiking observations ($K+m$). I have tested considerably high sampling rates from the SSL (up to $b = 150$). After a lot of testing, PLSR with and without prior centering or scaling of the preprocessed spectra, and also restricting the final prediction model to only 3 to 5 components, there was still substantial bias. The target set (Frick LTE) was clustered for the analysis to even further reduce non-linearity. I tested CUBIST on the RS-LOCAL selection, to remove the majority of bias across the IR estimated OC contents of the LTE test set. Thus, it seems that even by drastically reducing sample numbers, there is a point where linear methods of modeling cannot improve prediction results anymore. In conclusion, it seems that we might also incorporate some non-linear machine learning methods to the core of the RS-LOCAL algorithm. I have carefully scanned the literature on instance-based transfer learning and have found only one method that resembles RS-LOCAL at least partly in its working principles: *"Double-bootstrapping source data selection for instance-based transfer learning"* (Lin et al., 2013). The method reduces the impact of irrelevant or erroneous samples in the source domain (i.e., SSL) by drawing bootstrap samples from the target domain, and selecting those observations in the source for which the target model is performing well.

6.4 Opinion piece on how research in soil spectroscopy will be evolving and applied by different soil stakeholders

I imagine a world of soil spectroscopy where the coverage of SSLs is critically evaluated and discussed in function of the chemical data distributions, density and connectivity represented in the multivariate spectroscopic space. Further, we should also discuss more about geographic sampling density and soil landscapes and types covered. My recommendation is to use these concrete information criteria and metrics, instead of using rather loose and, in my opinion, often marketing terminology of dif-

ferent "scales", such as field scale or regional scale, etc.. Scale concepts have their justification in soils, and more so for the linkage between microscopic and macroscopic processes and emerging properties, especially in soil physics (e.g., Miller and Miller, 1956; Sadeghi et al., 2016). However, because soil ecosystems are drastically variable in their short-range variation in composition and properties (e.g., Zhang and Hartemink, 2021) and also across larger geographical extents (e.g., Loescher et al., 2014), sample size and sampling spacing in the field should be mentioned for the assessment of spectroscopic modeling and estimation of soils' properties and characterization of mineral and organic composition.

So far, I do not know of any published research in soil spectroscopy that has systematically tested the caveats of cross-validation for different combinations of sample size and soil variation (see also section 5.1). It would be therefore beneficial for the community's awareness if a soil science journal would feature a case-study in this regard. It would be good to illustrate and explain the effect different training set sizes and variability — e.g., small data set with high soil variability vs. large data set with small soil variability with profile data — under different gross resampling strategies. This should be done with and without appropriate grouping. Moreover, it could include linear and non-linear modeling approaches that are useful and established in the soil spectroscopy community. I hypothesize that these choices will have a significant effect on how accuracy and bias of the modeling results is reported.

I also hypothesize that data-driven transfer learning approaches such as RS-LOCAL will be soon on the rise. Regardless of the model choices and learning mantras, which are sometimes also personal preference, we will probably need continuous learning and growing SSLs. This is mainly because the world of soils is inherently complex in many aspects of the soil forming factors. This continuous learning process with efficient IR-spectroscopic workflows and growing SSLs across space and time will hopefully already soon deliver a lot more key chemical, physical, and also biological characteristics of soils than we now can imagine. This will be key for soil monitoring, agronomic soil testing, environmental mapping and decision making. We urgently need this basic soil information to implement informed soil decision making, and to foster sustainable soil management practices that preserve and improve soil health conditions for our and future generations.

Bibliography

Abbott, L. K. and Murphy, D. V., editors (2007). *Soil Biological Fertility: A Key to Sustainable Land Use in Agriculture*. Springer Netherlands.

Agroscope (1996). Schweizerische Referenzmethoden der Forschungsanstalten Agroscope. `https://www.agroscope.admin.ch/agroscope/de/home/themen/umwelt-ressourcen/monitoring-analytik/referenzmethoden/standortcharakterisierung.html` [Accessed: 2022-05-25].

Aha, D. W. (1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36(2):267–287.

Ambroise, C. and McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, 99(10):6562–6566.

Angelopoulou, T., Balafoutis, A., Zalidis, G., and Bochtis, D. (2020). From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon EstimationA Review. *Sustainability*, 12(2):443.

Arrouays, D., Marchant, B., Saby, N., Meersmans, J., Orton, T., Martin, M., Bellamy, P., Lark, R., and Kibblewhite, M. (2012). Generic Issues on Broad-Scale Soil Monitoring Schemes: A Review. *Pedosphere*, 22(4):456–469.

Arrouays, D., Mulder, V. L., and Richer-de Forges, A. C. (2021). Soil mapping, digital soil mapping and soil monitoring over large areas and the dimensions of soil security A review. *Soil Security*, 5:100018.

Awiti, A. O., Walsh, M. G., Shepherd, K. D., and Kinyamario, J. (2008). Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence. *Geoderma*, 143(1-2):73–84.

BAFU (2014). Biodiversitätsmonitoring Schweiz BDM, Koordinationsstelle BDM 2014: Biodiversitätsmonitoring Schweiz BDM. Beschreibung der Methoden und Indika-

toren. Technical Report Umwelt-Wissen Nr. 1410, Bundesamt für Umwelt BAFU, Bern, Switzerland.

Bai, Z., Caspari, T., Gonzalez, M. R., Batjes, N. H., Mäder, P., Bünemann, E. K., de Goede, R., Brussaard, L., Xu, M., Ferreira, C. S. S., Reintam, E., Fan, H., Miheli, R., Glavan, M., and Tóth, Z. (2018). Effects of agricultural management practices on soil quality: A review of long-term experiments for Europe and China. *Agriculture, Ecosystems & Environment*, 265:1–7.

Barnes, R. J., Dhanoa, M. S., and Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5):772–777.

Bates, S., Hastie, T., and Tibshirani, R. (2021). Cross-validation: what does it estimate and how well does it do it? *arXiv:2104.00673 [math, stat]*. arXiv: 2104.00673.

Baumann, P. (2019). philipp-baumann/simplerspec: Beta release simplerspec 0.1.0 for zenodo. `https://doi.org/10.5281/zenodo.3303637` [Accessed: 2022-05-25].

Baumann, P. (2020). philipp-baumann/yamsys-soilspec-publication: Open data and code (manuscript submission): "Estimation of soil properties with mid-infrared soil spectroscopy across yam production landscapes in West Africa".

Baumann, P., Helfenstein, A., Gubler, A., Keller, A., Meuli, R. G., Wächter, D., Lee, J., Viscarra Rossel, R., and Six, J. (2021). Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring. *SOIL*, 7(2):525–546.

Baumann, P., Lee, J., Behrens, T., Biswas, A., Six, J., McLachlan, G., and Viscarra Rossel, R. A. (2022). Modelling soil water retention and water-holding capacity with visiblenear-infrared spectra and machine learning. *European Journal of Soil Science*, 73(2):e13220. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.13220.

Baumgardner, M. F., Silva, L. F., Biehl, L. L., and Stoner, E. R. (1986). Reflectance Properties of Soils. In Brady, N. C., editor, *Advances in Agronomy*, volume 38, pages 1–44. Academic Press.

Behrens, T., Viscarra Rossel, R. A., Kerry, R., MacMillan, R., Schmidt, K., Lee, J., Scholten, T., and Zhu, A.-X. (2019). The relevant range of scales for multi-scale contextual spatial modelling. *Scientific Reports*, 9(1):14800.

Beleites, C., Baumgartner, R., Bowman, C., Somorjai, R., Steiner, G., Salzer, R., and Sowa, M. G. (2005). Variance reduction in estimating classification error using sparse datasets. *Chemometrics and Intelligent Laboratory Systems*, 79(1-2):91–100.

Bellman, E., R. (1962). *Adaptive Control Processes: A Guided Tour*. Princeton University Press, Princeton, New Jersey.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., and McBratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry*, 29(9):1073–1081.

Bellon-Maurel, V. and McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils Critical review and research perspectives. *Soil Biology and Biochemistry*, 43(7):1398–1410.

Ben-Dor, E. and Banin, A. (1995). Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. *Soil Science Society of America Journal*, 59(2):364–372.

Berner, A., Hildermann, I., Fliesbach, A., Pfiffner, L., Niggli, U., and Mader, P. (2008). Crop yield and soil fertility response to reduced tillage under organic management. *Soil and Tillage Research*, 101(1-2):89–96.

Bornemann, L., Welp, G., and Amelung, W. (2010). Particulate Organic Matter at the Field Scale: Rapid Acquisition Using Mid-Infrared Spectroscopy. *Soil Science Society of America Journal*, 74(4):1147.

Bouma, J. (1989). Using soil survey data for quantitative land evaluation. *Advances in soil science*, 9(1989):177–213.

Bouma, J. (2014). Soil science contributions towards Sustainable Development Goals and their implementation: linking soil functions with ecosystem services. *Journal of Plant Nutrition and Soil Science*, 177(2):111–120.

Bouveresse, E. and Massart, D. L. (1996). Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*, 32(2):201–213.

Bouyoucos, G. J. (1951). A recalibration of the hydrometer method for making mechanical analysis of soils. *Agronomy journal*, 43(9):434–438.

Bowers, S. A. and Hanks, R. J. (1965). Reflection of Radiant Energy from Soils. *Soil Science*, 100(2):130–138.

Bowers, S. A. and Smith, S. J. (1972). Spectrophotometric Determination of Soil Water Content 1. *Soil Science Society of America Journal*, 36(6):978–980.

Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis.

Briedis, C., Baldock, J., de Moraes Sá, J. C., dos Santos, J. B., and Milori, D. M. B. P. (2020). Strategies to improve the prediction of bulk soil and fraction organic carbon in Brazilian samples by using an Australian national mid-infrared spectral library. *Geoderma*, 373:114401.

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

Bruce, L. and Li, J. (2001). Wavelets for computationally efficient hyperspectral derivative analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1540–1546.

Brus, D. J. (2019). Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma*, 338:464–480.

Bui, E. N., Henderson, B. L., and Viergever, K. (2006). Knowledge discovery from models of soil properties developed through data mining. *Ecological Modelling*, 191(3):431–446.

Butler, B. M., Palarea-Albaladejo, J., Shepherd, K. D., Nyambura, K. M., Towett, E. K., Sila, A. M., and Hillier, S. (2020). Mineralnutrient relationships in African soils assessed using cluster analysis of X-ray powder diffraction patterns and compositional methods. *Geoderma*, 375:114474.

Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T. W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J. W., and Brussaard, L. (2018). Soil quality A critical review. *Soil Biology and Biochemistry*, 120:105–125.

Calderón, F., Haddix, M., Conant, R., Magrini-Bair, K., and Paul, E. (2013). Diffuse-Reflectance Fourier-Transform Mid-Infrared Spectroscopy as a Method of Characterizing Changes in Soil Organic Matter. *Soil Science Society of America Journal*, 77(5):1591–1600.

Calderón, F. J., Reeves, J. B., Collins, H. P., and Paul, E. A. (2011). Chemical Differences in Soil Organic Matter Fractions Determined by Diffuse-Reflectance Mid-Infrared Spectroscopy. *Soil Science Society of America Journal*, 75(2):568–579.

Cambardella, C. A. and Elliott, E. T. (1992). Particulate Soil Organic-Matter Changes across a Grassland Cultivation Sequence. *Soil Science Society of America Journal*, 56(3):777–783.

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C., and Barthès, B. G. (2016). Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field. *Geoderma*, 261:151–159.

Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In Pei, J., Tseng, V. S., Cao, L., Motoda, H., and Xu, G., editors, *Advances in Knowledge Discovery and Data Mining*, Lecture Notes in Computer Science, pages 160–172, Berlin, Heidelberg. Springer.

Campello, R. J. G. B., Moulavi, D., Zimek, A., and Sander, J. (2015). Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Transactions on Knowledge Discovery from Data*, 10(1):5:1–5:51.

Carsky, R. J., Asiedu, R., and Cornet, D. (2010). Review of soil fertility management for yam-based systems in west africa. *African Journal of Root and Tuber Crops*, 8(2):1.

Chong, I.-G. and Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1):103–112.

Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N. P., Bernoux, M., and Barthès, B. G. (2016). National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. *Geoderma*, 276:41–52.

Conen, F., Yakutin, M. V., and Sambuu, A. D. (2003). Potential for detecting changes in soil organic carbon concentrations resulting from climate change. *Global Change Biology*, 9(11):1515–1520.

Cornet, D., Sierra, J., Tournebize, R., Gabrielle, B., and Lewis, F. I. (2016). Bayesian network modeling of early growth stages explains yam interplant yield variability and allows for agronomic improvements in West Africa. *European Journal of Agronomy*, 75:80–88.

Couillard, A., Turgeon, A. J., Shenk, J. S., and Westerhaus, M. O. (1997). Near Infrared Reflectance Spectroscopy for Analysis of Turf Soil Profiles. *Crop Science*, 37(5):cropsci1997.0011183X003700050024x.

Cremers, D. A., Ebinger, M. H., Breshears, D. D., Unkefer, P. J., Kammerdiener, S. A., Ferris, M. J., Catlett, K. M., and Brown, J. R. (2001). Measuring Total Soil Carbon with

Laser-Induced Breakdown Spectroscopy (LIBS). *Journal of Environmental Quality*, 30(6):2202–2206.

Cécillon, L., Barthès, B. G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A., and Brun, J. J. (2009). Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *European Journal of Soil Science*, 60(5):770–784.

Dangal, S. R. S., Sanderman, J., Wills, S., and Ramirez-Lopez, L. (2019). Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library. *Soil Systems*, 3(1):11.

Dazzi, A., Deniset-Besseau, A., and Lasch, P. (2013). Minimising contributions from scattering in infrared spectra by means of an integrating sphere. *The Analyst*, 138(14):4191.

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263.

De Noord, O. E. (1994). Multivariate calibration standardization. *Chemometrics and Intelligent Laboratory Systems*, 25(2):85–97.

Debaene, G., Niedwiecki, J., Pecio, A., and urek, A. (2014). Effect of the number of calibration samples on the prediction of several soil properties at the farm-scale. *Geoderma*, 214-215:114–125.

Deiss, L., Culman, S. W., and Demyan, M. S. (2020). Grinding and spectra replication often improves mid-DRIFTS predictions of soil properties. *Soil Science Society of America Journal*, 84(3):914–929. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.1002/saj2.20021.

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Walsh, M. K., Ojima, D. S., and Thornton, P. E. (2006). DAYCENT National-Scale Simulations of Nitrous Oxide Emissions from Cropped Soils in the United States. *Journal of Environmental Quality*, 35(4):1451–1460.

Deluz, C., Nussbaum, M., Sauzet, O., Gondret, K., and Boivin, P. (2020). Evaluation of the Potential for Soil Organic Carbon Content Monitoring With Farmers. *Frontiers in Environmental Science*, 8:113.

Delwiche, S. R. and Reeves III, J. B. (2010). A Graphical Method to Evaluate Spectral Preprocessing in Multivariate Regression Calibrations: Example with SavitzkyGolay Filters and Partial Least Squares Regression. *Applied spectroscopy*, 64(1):73–82.

Demattê, J. A. M., Dotto, A. C., Paiva, A. F. S., Sato, M. V., Dalmolin, R. S. D., de Araújo, M. d. S. B., da Silva, E. B., Nanni, M. R., ten Caten, A., Noronha, N. C., Lacerda, M. P. C., de Araújo Filho, J. C., Rizzo, R., Bellinaso, H., Francelino, M. R., Schaefer, C. E. G. R., Vicente, L. E., dos Santos, U. J., de Sá Barretto Sampaio, E. V., Menezes, R. S. C., de Souza, J. J. L. L., Abrahão, W. A. P., Coelho, R. M., Grego, C. R., Lani, J. L., Fernandes, A. R., Gonçalves, D. A. M., Silva, S. H. G., de Menezes, M. D., Curi, N., Couto, E. G., dos Anjos, L. H. C., Ceddia, M. B., Pinheiro, . F. M., Grunwald, S., Vasques, G. M., Marques Júnior, J., da Silva, A. J., Barreto, M. C. d. V., Nóbrega, G. N., da Silva, M. Z., de Souza, S. F., Valladares, G. S., Viana, J. H. M., da Silva Terra, F., Horák-Terra, I., Fiorio, P. R., da Silva, R. C., Frade Júnior, E. F., Lima, R. H. C., Alba, J. M. F., de Souza Junior, V. S., Brefin, M. D. L. M. S., Ruivo, M. D. L. P., Ferreira, T. O., Brait, M. A., Caetano, N. R., Bringhenti, I., de Sousa Mendes, W., Safanelli, J. L., Guimarães, C. C. B., Poppiel, R. R., e Souza, A. B., Quesada, C. A., and do Couto, H. T. Z. (2019). The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. *Geoderma*, 354:113793.

Deng, F., Minasny, B., Knadel, M., McBratney, A., Heckrath, G., and Greve, M. H. (2013). Using Vis-NIR Spectroscopy for Monitoring Temporal Changes in Soil Organic Carbon:. *Soil Science*, 178(8):389–399.

Desaules, A., Ammann, S., and Schwab, P. (2010). Advances in long-term soil-pollution monitoring of Switzerland. *Journal of Plant Nutrition and Soil Science*, 173(4):525–535.

Diby, L. N., Hgaza, V. K., Tie, T. B., Assa, A., Carsky, R., Girardin, O., and Frossard, E. (2009). Productivity of yams (Dioscorea spp.) as affected by soil fertility. *Journal of Animal & Plant Sciences*, 5(2):494–506.

Diby, L. N., Tie, B. T., Girardin, O., Sangakkara, R., and Frossard, E. (2011). Growth and Nutrient Use Efficiencies of Yams (Dioscorea spp.) Grown in Two Contrasting Soils of West Africa.

Dietterich, T. G., Wettschereck, D., Atkeson, C. G., and Moore, A. W. (1993). Memory-based methods for regression and classification. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in neural information processing systems 6, [7th NIPS conference, denver, colorado, USA, 1993]*, pages 1165–1166. Morgan Kaufmann.

Dokuchaev, V. (1899). Report to the Transcaucasian statistical committee on land evaluation in general and especially for the Transcaucasia. Horizontal and vertical soil zones. (In Russian.) Off. Press Civ. *Affairs Commander-in-Chief Cacasus, Tiflis, Russia*.

Dotto, A. C., Demattê, J. A. M., Viscarra Rossel, R. A., and Rizzo, R. (2020). Soil environment grouping system based on spectral, climate, and terrain data: a quantitative branch of soil series. *SOIL*, 6(1):163–177.

Dowle, M. and Srinivasan, A. (2019). data.table: Extension of 'data.frame'. `https://CRAN.R-project.org/package=data.table` [Accessed: 2022-05-25].

Dwivedi, R. S. (2017). Soil Information Systems. In Ravi Shankar, D., editor, *Remote Sensing of Soils*, pages 359–398. Springer, Berlin, Heidelberg.

Efron, B. (1983). Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *Journal of the American Statistical Association*, 78(382):316–331.

Efron, B. and Tibshirani, R. (1997). Improvements on Cross-Validation: The 632+ Bootstrap Method. *Journal of the American Statistical Association*, 92(438):548–560.

Ellenbach, N., Boulesteix, A.-L., Bischl, B., Unger, K., and Hornung, R. (2020). Improved Outcome Prediction Across Data Sources Through Robust Parameter Tuning. *Journal of Classification*.

England, J. R. and Viscarra Rossel, R. A. (2018). Proximal sensing for soil carbon accounting. *SOIL*, 4(2):101–122.

Enyi, B. a. C. (1972). Effect of staking, nitrogen and potassium on growth and development in lesser yams: Dioscorea esculenta. *Annals of Applied Biology*, 72(2):211–219.

Evans, D. L., Janes-Bassett, V., Borrelli, P., Chenu, C., Ferreira, C. S. S., Griffiths, R. I., Kalantari, Z., Keesstra, S., Lal, R., Panagos, P., Robinson, D. A., Seifollahi-Aghmiuni, S., Smith, P., Steenhuis, T. S., Thomas, A., and Visser, S. M. (2021). Sustainable futures over the next decade are rooted in soil science. *European Journal of Soil Science*, n/a(n/a).

FAO (2006). *Guidelines for soil description*. Food and Agriculture Organization of the United Nations, Rome, 4th ed edition. OCLC: ocm71825863.

FAO (2014a). FAOSTAT.

FAO (2014b). *World reference base for soil resources 2014: international soil classification system for naming soils and creating legends for soil maps*. FAO, Rome. OCLC: 1004475995.

FAO (2020). *A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes : GSOC-MRV Protocol*. FAO, Rome, Italy. `https://www.fao.org/documents/card/en/c/cb0509en` [Accessed: 2022-05-25].

FAO (2020). Release of the GSOC MRV Protocol! ă|ăGlobal Soil Partnershipă|ăFood and Agriculture Organization of the United Nations. https://www.fao.org/global-soil-partnership/resources/highlights/detail/en/c/1308261/ [Accessed: 2022-05-25].

FAO (2021). Recarbonization of global agricultural soils (RECSOIL). https://www.fao.org/global-soil-partnership/areas-of-work/recarbonization-of-global-soils/en/ [Accessed: 2022-05-25].

Farmer, V. C., editor (1974). *The Infrared Spectra of Minerals*. Mineralogical Society of Great Britain and Ireland, Londan.

Fernández Pierna, J. A. and Dardenne, P. (2008). Soil parameter quantification by NIRS as a Chemometric challenge at Chimiométrie 2006. *Chemometrics and Intelligent Laboratory Systems*, 91(1):94–98.

Filzmoser, P., Liebmann, B., and Varmuza, K. (2009). Repeated double cross validation. *Journal of Chemometrics*, 23(4):160–171.

Filzmoser, P., Maronna, R., and Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics & Data Analysis*, 52(3):1694–1711.

Fontana, M., Berner, A., Mäder, P., Lamy, F., and Boivin, P. (2015). Soil Organic Carbon and Soil Bio-Physicochemical Properties as Co-Influenced by Tillage Treatment. *Soil Science Society of America Journal*, 79(5):1435–1445.

Foster, H. L. (1981). The Basic Factors Which Determine Inherent Soil Fertility in Uganda. *Journal of Soil Science*, 32(1):149–160.

Francioso, O., SánchezCortés, S., Corrado, G., Gioacchini, P., and Ciavatta, C. (2005). Characterization of Soil Organic Carbon in LongTerm Amendment Trials. *Spectroscopy Letters*, 38(3):283–291.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). *The elements of statistical learning*. Springer series in statistics Springer, Berlin, second edition edition.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1):1–67. Publisher: Institute of Mathematical Statistics.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232. Publisher: Institute of Mathematical Statistics.

Frossard, E., Aighewi, B. A., Aké, S., Barjolle, D., Baumann, P., Bernet, T., Dao, D., Diby, L. N., Floquet, A., Hgaza, V. K., Ilboudo, L. J., Kiba, D. I., Mongbo, R. L., Nacro, H. B., Nicolay, G. L., Oka, E., Ouattara, Y. F., Pouya, N., Senanayake, R. L., Six, J., and Traoré, O. I. (2017). The Challenge of Improving Soil Fertility in Yam Cropping Systems of West Africa. *Frontiers in Plant Science*, 8.

Gehl, R. J. and Rice, C. W. (2007). Emerging technologies for in situ measurement of soil carbon. *Climatic Change*, 80(1-2):43–54.

Gemperline, P. J., Long, J. R., and Gregoriou, V. G. (1991). Nonlinear multivariate calibration using principal components regression and artificial neural networks. *Analytical Chemistry*, 63(20):2313–2323.

Greenland, D. J., Wild, A., and Adams, D. (1992). Organic Matter Dynamics in Soils of the TropicsFrom Myth to Complex Reality. *Myths and Science of Soils of the Tropics*, sssaspecialpubl(mythsandscience):17–33.

Greiner, L., Keller, A., Grêt-Regamey, A., and Papritz, A. (2017). Soil function assessment: review of methods for quantifying the contributions of soils to ecosystem services. *Land Use Policy*, 69:224–237.

Greiner, L., Nussbaum, M., Papritz, A., Fraefel, M., Zimmermann, S., Schwab, P., Grêt-Regamey, A., and Keller, A. (2018). Assessment of soil multi-functionality to support the sustainable use of soil resources on the Swiss Plateau. *Geoderma Regional*, 14:e00181.

Grêt-Regamey, A., Kool, S., Bühlmann, L., and Kissling, S. (2018). *Eine Bodenagenda für die Raumplanung*. Thematische Synthese TS3 des Nationalen Forschungsprogramms ńNachhaltige Nutzung der Ressource Bodenż (NFP 68). Swiss National Science Foundation (SNF), Bern.

Gubler, A., Wächter, D., and Schwab, P. (2018). Homogenisation of series of soil organic carbon: harmonising results by wet oxidation (Swiss Standard Method) and dry combustion. *Agroscope Science*, 62. Zürich-Reckenholz.

Gubler, A., Wächter, D., Schwab, P., Müller, M., and Keller, A. (2019). Twenty-five years of observations of soil organic carbon in Swiss croplands showing stability overall but with some divergent trends. *Environmental Monitoring and Assessment*, 191(5):277.

Guerrero, C., Stenberg, B., Wetterlind, J., Viscarra Rossel, R. A., Maestre, F. T., Mouazen, A. M., Zornoza, R., Ruiz-Sinoga, J. D., and Kuang, B. (2014). Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset: Spiking and extra-weighting to improve soil organic carbon predictions with NIR. *European Journal of Soil Science*, 65(2):248–263.

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A. M., Gabarrón-Galeote, M. A., Ruiz-Sinoga, J. D., Zornoza, R., and Viscarra Rossel, R. A. (2016). Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? *Soil and Tillage Research*, 155:501–509.

Guerrero, C., Zornoza, R., Gómez, I., and Mataix-Beneyto, J. (2010). Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. *Geoderma*, 158(1):66–77.

Guillou, F. L., Wetterlind, W., Viscarra Rossel, R. A., Hicks, W., Grundy, M., and Tuomi, S. (2015). How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? *Soil Research*.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46(1):389–422.

Haines-Young, R. and Potschin, M. (2008). England's Terrestrial Ecosystem Services and the Raionale for an Ecosystem Approach. Full Technical Report to Defra.

Hall, S. J., Ye, C., Weintraub, S. R., and Hockaday, W. C. (2020). Molecular trade-offs in soil organic carbon composition at continental scale. *Nature Geoscience*, 13(10):687–692.

Hand, D. J. and Vinciotti, V. (2003). Local Versus Global Models for Classification Problems: Fitting Models Where it Matters. *The American Statistician*, 57(2):124–131.

Hartemink, A. E., McBratney, A. B., and Mendonça-Santos, M. d. L., editors (2008). *Digital soil mapping with limited data*. Springer, Dordrecht ; London. OCLC: ocn231590444.

Helfenstein, A., Baumann, P., Viscarra Rossel, R., Gubler, A., Oechslin, S., and Six, J. (2021). Quantifying soil carbon in temperate peatlands using a mid-IR soil spectral library. *SOIL*, 7(1):193–215.

Hendershot, W. H. and Duquette, M. (1986). A simple barium chloride method for determining cation exchange capacity and exchangeable cations. *Soil Science Society of America Journal*, 50(3):605–608.

Hgaza, V. K., Diby, L. N., Tié, T. B., Tschannen, A., Aké, S., Assa, A., and Frossard, E. (2011). Growth and distribution of roots of Dioscorea alata L. do not respond to mineral fertilizer application. *Open Plant Science Journal*, 5:14–22.

Hillier, S., Brydson, R., Delbos, E., Fraser, T., Gray, N., Pendlowski, H., Phillips, I., Robertson, J., and Wilson, I. (2016). Correlations among the mineralogical and physical properties of halloysite nanotubes (HNTs). *Clay Minerals*, 51(3):325–350.

Hong, Y., Chen, S., Liu, Y., Zhang, Y., Yu, L., Chen, Y., Liu, Y., Cheng, H., and Liu, Y. (2019). Combination of fractional order derivative and memory-based learning algorithm to improve the estimation accuracy of soil organic matter by visible and near-infrared spectroscopy. *CATENA*, 174:104–116.

Hubert, M. and Debruyne, M. (2010). Minimum covariance determinant. *WIREs Computational Statistics*, 2(1):36–43.

Igne, B., Reeves, III, J., McCarty, G., Hively, W., Lund, E., and Hurburgh, Jr, C. (2010). Evaluation of spectral pretreatments, partial least squares, least squares support vector machines and locally weighted regression for quantitative spectroscopic analysis of soils. *Journal of Near Infrared Spectroscopy*, 18(1):167.

Janik, L. J. and Skjemstad, J. O. (1995). Characterization and analysis of soils using mid-infrared partial least-squares .2. Correlations with some laboratory data. *Australian Journal of Soil Research*, 33(4):637–650.

Janik, L. J., Skjemstad, J. O., and Merry, R. H. (1998). Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture*, 38(7):681.

Jenny, H. (1941). *Factors of soil formation*. McGraw-Hill Book Co., New York.

Johnson, J.-M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K. D., and Saito, K. (2019). Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. *Geoderma*, 354:113840.

Kang, B. T. and Wilson, J. E. (1981). Effect of mound size and fertilizer on white Guinea yam (Dioscorea rotundata) in Southern Nigeria. *Plant and Soil*, 61(3):319–327.

Karyotis, K., Angelopoulou, T., Tziolas, N., Palaiologou, E., Samarinas, N., and Zalidis, G. (2021). Evaluation of a Micro-Electro Mechanical Systems Spectral Sensor for Soil Properties Estimation. *Land*, 10(1):63. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

Kassi, S.-P. A., Koné, A. W., Tondoh, J. E., and Koffi, B. Y. (2017). Chromoleana odorata fallow-cropping cycles maintain soil carbon stocks and yam yields 40 years after conversion of native- to farmland, implications for forest conservation. *Agriculture, Ecosystems & Environment*, 247:298–307.

Keller, A., Franzen, J., Knüsel, P., Papritz, A., and Zürrer, M. (2018). *Bodeninformations-Plattform Schweiz (BIP-CH)*. Thematische Synthese TS4 des Nationalen Forschungsprogramms ńNachhaltige Nutzung der Ressource Bodenż (NFP 68). Swiss National Science Foundation (SNF), Bern.

Kerker, M. (1982). LorenzMie Scattering by Spheres: Some Newly Recognized Phenomena. *Aerosol Science and Technology*, 1(3):275–291.

Kiba, D. I., Hgaza, V. K., Aighewi, B., Aké, S., Barjolle, D., Bernet, T., Diby, L. N., Ilboudo, L. J., Nicolay, G., Oka, E., Ouattara, F. Y., Pouya, N., Six, J., and Frossard, E. (2020). A Transdisciplinary Approach for the Development of Sustainable Yam (Dioscorea sp.) Production in West Africa. *Sustainability*, 12(10):4016. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Kim, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*, 53(11):3735–3745.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*, IJCAI'95, pages 1137–1143, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Komatsuzaki, M. and Ohta, H. (2007). Soil management practices for sustainable agro-ecosystems. *Sustainability Science*, 2(1):103–120.

Krauss, M., Berner, A., Burger, D., Wiemken, A., Niggli, U., and Mäder, P. (2010). Reduced tillage in temperate organic farming: implications for crop management and forage production. *Soil Use and Management*, 26(1):12–20.

Krauss, M., Berner, A., Perrochet, F., Frei, R., Niggli, U., and Mäder, P. (2020). Enhanced soil quality with reduced tillage and solid manures in organic farming a synthesis of 15 years. *Scientific Reports*, 10(1):1–12.

Krstajic, D., Buturovic, L. J., Leahy, D. E., and Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, 6.

Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York, New York, NY.

Kuhn, M., Wing, J., Weston, S., A., W., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., and Hunt, T. (2019). caret: Classification and regression training.

Lee, J., Necpálová, M., Calitri, F., and Six, J. (2020). Simulation of a regional soil nitrogen balance in Swiss croplands. *Nutrient Cycling in Agroecosystems*, 118(1):9–22.

Leenen, M., Pätzold, S., Tóth, G., and Welp, G. (2022). A LUCAS-based mid-infrared soil spectral library: Its usefulness for soil survey and precision agriculture. *Journal of Plant Nutrition and Soil Science*, n/a(n/a). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jpln.202100031.

Lehmann, J., Hansel, C. M., Kaiser, C., Kleber, M., Maher, K., Manzoni, S., Nunan, N., Reichstein, M., Schimel, J. P., Torn, M. S., Wieder, W. R., and Kögel-Knabner, I. (2020). Persistence of soil organic carbon caused by functional complexity. *Nature Geoscience*, 13(8):529–534.

Lehmann, J. and Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*, 528(7580):60–68.

Leifeld, J. (2006). Application of diffuse reflectance FT-IR spectroscopy and partial least-squares regression to predict NMR properties of soil organic matter. *European Journal of Soil Science*, 57(6):846–857.

Lin, D., An, X., and Zhang, J. (2013). Double-bootstrapping source data selection for instance-based transfer learning. *Pattern Recognition Letters*, 34(11):1279–1285.

Lin, J.-H. and Vitter, J. S. (1994). A Theory for Memory-Based Learning. *Machine Learning*, 17(2):143–167.

Lindeberg, T. (1994). Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1-2):225–270.

Lindsay, W. L. and Norvell, W. A. (1978). Development of a DTPA soil test for zinc, iron, manganese, and copper. *Soil science society of America journal*, 42(3):421–428.

Liu, L., Ji, M., and Buchroithner, M. (2018). Transfer Learning for Soil Spectroscopy Based on Convolutional Neural Networks and Its Application in Soil Clay Content Mapping Using Hyperspectral Imagery. *Sensors (Basel, Switzerland)*, 18(9).

Lobsey, C. R., Viscarra Rossel, R. A., Roudier, P., and Hedley, C. B. (2017). rs-local data-mines information from spectral libraries to improve local calibrations. *European Journal of Soil Science*, 68(6):840–852.

Loescher, H., Ayres, E., Duffy, P., Luo, H., and Brunke, M. (2014). Spatial Variation in Soil Properties among North American Ecosystems and Guidelines for Sampling Designs. *PLOS ONE*, 9(1):e83216.

Lorber, A., Faber, K., and Kowalski, B. R. (1996). Local centering in multivariate calibration. *Journal of Chemometrics*, 10(3):215–220.

Lugato, E., Lavallee, J. M., Haddix, M. L., Panagos, P., and Cotrufo, M. F. (2021). Different climate sensitivity of particulate and mineral-associated soil organic matter. *Nature Geoscience*, 14(5):295–300.

Lyon, R. J. P. and Tuddenham, W. M. (1960). Infra-Red Determination of the Kaolin Group Minerals. *Nature*, 185(4716):835–836. Number: 4716 Publisher: Nature Publishing Group.

Madari, B. E., Reeves, J. B., Machado, P. L., Guimarães, C. M., Torres, E., and McCarty, G. W. (2006). Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma*, 136(1-2):245–259.

Madejová, J., Gates, W. P., and Petit, S. (2017). Chapter 5 - IR Spectra of Clay Minerals. In Gates, W. P., Kloprogge, J. T., Madejová, J., and Bergaya, F., editors, *Developments in Clay Science*, volume 8 of *Infrared and Raman Spectroscopies of Clay Minerals*, pages 107–149. Elsevier.

Madejová, J., Kekés, J., Pálková, H., and Komadel, P. (2002). Identification of components in smectite/kaolinite mixtures. *Clay Minerals*, 37(2):377–388.

Martens, H. and Naes, T. (1989). *Multivariate Calibration*. Wiley Chichester.

Martens, H. and Næs, T. (1984). Multivariate Calibration. In Kowalski, B. R., editor, *Chemometrics: Mathematics and Statistics in Chemistry*, NATO ASI Series, pages 147–156. Springer Netherlands, Dordrecht.

McCarty, G., Hively, W., Reeves, J., Lang, M., Lund, E., and Weatherbee, O. (2010). Infrared Sensors to Map Soil Carbon in Agricultural Ecosystems. In Viscarra Rossel, R. A., McBratney, A. B., and Minasny, B., editors, *Proximal Soil Sensing*, Progress in Soil Science, pages 165–176. Springer Netherlands, Dordrecht.

McCarty, G. W., Reeves, J. B., Reeves, V. B., Follett, R. F., and Kimble, J. M. (2002). Mid-Infrared and Near-Infrared Diffuse Reflectance Spectroscopy for Soil Carbon Measurement. *Soil Science Society of America Journal*, 66(2):640–646.

Meersmans, J., Wesemael, B. V., and Molle, M. V. (2009). Determining soil organic carbon for agricultural soils: a comparison between the Walkley & Black and the dry combustion methods (north Belgium). *Soil Use and Management*, 25(4):346–353.

Meuli, R. G., Wächter, D., Schwab, P., Kohli, L., and Zimmermann, R. (2017). Connecting Biodiversity Monitoring with Soil Inventory Data A Swiss Case Study. *BGS Bulletin*, 38:3.

Mevik, B.-H., Wehrens, R., and Liland, K. H. (2019). pls: Partial least squares and principal component regression.

Miller, B. A., Koszinski, S., Wehrhan, M., and Sommer, M. (2015). Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks. *SOIL*, 1(1):217–233.

Miller, E. E. and Miller, R. D. (1956). Physical Theory for Capillary Flow Phenomena. *Journal of Applied Physics*, 27(4):324–332.

Minasny, B. and McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, 264:301–311.

Molinaro, A. M., Simon, R., and Pfeiffer, R. M. (2005). Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, 21(15):3301–3307.

Moura-Bueno, J. M., Dalmolin, R. S. D., Horst-Heinen, T. Z., ten Caten, A., Vasques, G. M., Dotto, A. C., and Grunwald, S. (2020). When does stratification of a subtropical soil spectral library improve predictions of soil organic carbon content? *Science of The Total Environment*, 737:139895.

Necpalova, M., Anex, R. P., Kravchenko, A. N., Abendroth, L. J., Del Grosso, S. J., Dick, W. A., Helmers, M. J., Herzmann, D., Lauer, J. G., Nafziger, E. D., Sawyer, J. E., Scharf, P. C., Strock, J. S., and Villamil, M. B. (2014). What does it take to detect a change in soil carbon stock? A regional comparison of minimum detectable difference and experiment duration in the north central United States. *Journal of Soil and Water Conservation*, 69(6):517–531.

Nguyen, T., Janik, L., and Raupach, M. (1991). Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Soil Research*, 29(1):49.

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., 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., and Wetterlind, J. (2015a). Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. In *Advances in Agronomy*, volume 132, pages 139–159. Elsevier.

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., and Wetterlind, J. (2015b). Chapter Four - Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. In Sparks, D. L., editor, *Advances in Agronomy*, volume 132, pages 139–159. Academic Press.

Ogen, Y., Zaluda, J., Francos, N., Goldshleger, N., and Ben-Dor, E. (2019). Cluster-based spectral models for a robust assessment of soil properties. *Geoderma*, 340:175–184.

Olson, K. R. and Al-Kaisi, M. M. (2015). The importance of soil sampling depth for accurate account of soil organic carbon sequestration, storage, retention and loss. *CATENA*, 125:33–37.

O'Sullivan, J. N. and Australian Centre for International Agricultural Research (2010). *Yam nutrition nutrient disorders and soil fertility management*. Australian Centre for International Agricultural Research, Canberra, A.C.T.

O'Sullivan, J. N. and Jenner, R. (2006). Nutrient Deficiencies in Greater Yam and Their Effects on Leaf Nutrient Concentrations. *Journal of Plant Nutrition*, 29(9):1663–1674.

Ozaki, Y., McClure, W. F., and Christy, A. A., editors (2006). *Near-Infrared Spectroscopy in Food Science and Technology: Ozaki/Near-Infrared Spectroscopy in Food Science and Technology*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Padarian, J., Minasny, B., and McBratney, A. B. (2019a). Transfer learning to localise a continental soil vis-NIR calibration model. *Geoderma*, 340:279–288.

Padarian, J., Minasny, B., and McBratney, A. B. (2019b). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional*, 16:e00198.

Padwick, G. W. (1983). Fifty Years of *Experimental Agriculture* II. The Maintenance of Soil Fertility in Tropical Africa: A Review. *Experimental Agriculture*, 19(4):293–310.

Pan, S. J. and Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.

Parfitt, R. L., Atkinson, R. J., and Smart, R. S. C. (1975). The Mechanism of Phosphate Fixation by Iron Oxides. *Soil Science Society of America Journal*, 39(5):837–841. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj1975.03615995003900050017x.

Parikh, S. J., Goyne, K. W., Margenot, A. J., Mukome, F. N. D., and Calderón, F. J. (2014). Chapter One - Soil Chemical Insights Provided through Vibrational Spectroscopy. In Sparks, D. L., editor, *Advances in Agronomy*, volume 126, pages 1–148. Academic Press.

Paustian, K., Collier, S., Baldock, J., Burgess, R., Creque, J., DeLonge, M., Dungait, J., Ellert, B., Frank, S., Goddard, T., Govaerts, B., Grundy, M., Henning, M., Izaurralde, R. C., Madaras, M., McConkey, B., Porzig, E., Rice, C., Searle, R., Seavy, N., Skalsky, R., Mulhern, W., and Jahn, M. (2019). Quantifying carbon for agricultural soil management: from the current status toward a global soil information system. *Carbon Management*, 10(6):567–587.

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., and Smith, P. (2016). Climate-smart soils. *Nature*, 532(7597):49–57.

Peng, Y., Xiong, X., Adhikari, K., Knadel, M., Grunwald, S., and Greve, M. H. (2015). Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra. *PLOS ONE*, 10(11):e0142295.

Peterson, J. and Baumgardner, M. (1981). Use Of Spectral Data To Estimate The Relationship Between Soil Moisture Tensions And Their Corresponding Reflectances. *Annu. Rep. OWRT Purdue Univ.*, pages 1–18.

Piñeiro, G., Perelman, S., Guerschman, J. P., and Paruelo, J. M. (2008). How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216(3-4):316–322.

Post, W. M., Izaurralde, R. C., Mann, L. K., and Bliss, N. (2001). Monitoring and Verifying Changes of Organic Carbon in Soil. *Climatic Change*, 51(1):73–99.

Pratt, L. and Thrun, S. (1997). Guest Editors' Introduction. *Machine Learning*, 28(1):5–5.

Pratt, L. Y., Mostow, J., and Kamm, C. A. (1991). Direct transfer of learned information among neural networks. In *Proceedings of the ninth National conference on Artificial intelligence - Volume 2*, AAAI'91, pages 584–589, Anaheim, California. AAAI Press.

Pratt, L. Y., Pratt, L. Y., Hanson, S. J., Giles, C. L., and Cowan, J. D. (1993). Discriminability-Based Transfer between Neural Networks. In *Advances in Neural Information Processing Systems 5*, pages 204–211. Morgan Kaufmann.

Prescott, C. E., Rui, Y., Cotrufo, M. F., and Grayston, S. J. (2021). Managing plant surplus carbon to generate soil organic matter in regenerative agriculture. *Journal of Soil and Water Conservation*, 76(6):99A–104A.

Quinlan, J. (1993). Combining Instance-Based and Model-Based Learning. In *Machine Learning Proceedings 1993*, pages 236–243. Elsevier.

Quinlan, J. R. (1987). Generating production rules from decision trees. In *Proceedings of the 10th international joint conference on Artificial intelligence - Volume 1*, IJCAI'87, pages 304–307, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Quinlan, J. R. and others (1992). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence*, volume 92, pages 343–348. tex.organization: World Scientific.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2019). R: A language and environment for statistical computing. `https://www.R-project.org/` [Accessed: 2022-05-25].

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Rossel, R. V., Demattê, J., and Scholten, T. (2013a). Distance and similarity-search metrics for use with soil visNIR spectra. *Geoderma*, 199:43–53.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J. A. M., and Scholten, T. (2013b). The spectrum-based learner: A new local approach for modeling soil visNIR spectra of complex datasets. *Geoderma*, 195-196:268–279.

Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J. A. M., and Scholten, T. (2014). Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma*, 226-227:140–150.

RamirezLopez, L., Wadoux, A. M. J.-C., Franceschini, M. H. D., Terra, F. S., Marques, K. P. P., Sayão, V. M., and Demattê, J. a. M. (2019). Robust soil mapping at the farm scale with visNIR spectroscopy. *European Journal of Soil Science*, 70(2):378–393.

Ramírez, P. B., Calderón, F. J., Haddix, M., Lugato, E., and Cotrufo, M. F. (2021). Using Diffuse Reflectance Spectroscopy as a High Throughput Method for Quantifying Soil C and N and Their Distribution in Particulate and Mineral-Associated Organic Matter Fractions. *Frontiers in Environmental Science*, 9:153.

Reeves, J. B., McCarty, G. W., and Meisinger, J. J. (1999). Near Infrared Reflectance Spectroscopy for the Analysis of Agricultural Soils. *Journal of Near Infrared Spectroscopy*, 7(3):179–193.

Rehbein, K., Sprecher, C., and Keller, A. (2020). *Übersicht Stand Bodenkartierung in der Schweiz. Ergänzung des Bodenkartierungskataloges Schweiz um Bodeninformationen aus Meliorationsprojekten*. Servicestelle NABODAT, Agroscope, Zürich.

Rinnan, . (2014). Pre-processing in vibrational spectroscopy when, why and how. *Analytical Methods*, 6(18):7124.

Roger, J.-M. and Boulet, J.-C. (2018). A review of orthogonal projections for calibration. *Journal of Chemometrics*, 32(9):e3045. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cem.3045.

Rossel, R. V., Lark, R., and Ortega, A. (2010). Using Wavelets to Analyse Proximally Sensed VisNIR Soil Spectra. In Viscarra Rossel, R. A., McBratney, A. B., and Minasny, B., editors, *Proximal Soil Sensing*, Progress in Soil Science, pages 201–210. Springer Netherlands, Dordrecht.

Rossel, R. V. and McBratney, A. (2008). Diffuse Reflectance Spectroscopy as a Tool for Digital Soil Mapping. In Hartemink, A. E., McBratney, A., and Mendonça-Santos, M. d. L., editors, *Digital Soil Mapping with Limited Data*, pages 165–172. Springer Netherlands, Dordrecht.

Rousseeuw, P. J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79(388):871–880.

Saby, N. P. A., Bellamy, P. H., Morvan, X., Arrouays, D., Jones, R. J. A., Verheijen, F. G. A., Kibblewhite, M. G., Verdoodt, A., Üveges, J. B., FREUDENSCHUSS, A., and Simota, C. (2008). Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Global Change Biology*, 14(10):2432–2442. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2486.2008.01658.x.

Sadeghi, M., Ghahraman, B., Warrick, A. W., Tuller, M., and Jones, S. B. (2016). A critical evaluation of the Miller and Miller similar media theory for application to natural soils: A CRITICAL EVALUATION OF MILLER-MILLER SIMILAR MEDIA THEORY. *Water Resources Research*, 52(5):3829–3846.

Salzberg, S. L. (1990). *Learning with nested generalized exemplars*. Kluwer Academic Publishers, Boston. OCLC: 645073856.

Sanderman, J., Savage, K., and Dangal, S. R. S. (2020). Mid-infrared spectroscopy for prediction of soil health indicators in the United States. *Soil Science Society of America Journal*, 84(1):251–261.

Sanderman, J., Savage, K., Dangal, S. R. S., Duran, G., Rivard, C., Cavigelli, M. A., Gollany, H. T., Jin, V. L., Liebig, M. A., Omondi, E. C., Rui, Y., and Stewart, C. (2021). Can Agricultural Management Induced Changes in Soil Organic Carbon Be Detected Using Mid-Infrared Spectroscopy? *Remote Sensing*, 13(12):2265.

Sankey, J. B., Brown, D. J., Bernard, M. L., and Lawrence, R. L. (2008). Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma*, 148(2):149–158.

Savitzky, A. and Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639. Publisher: American Chemical Society.

Schirrmann, M., Gebbers, R., and Kramer, E. (2013). Performance of Automated Near-Infrared Reflectance Spectrometry for Continuous in Situ Mapping of Soil Fertility at Field Scale. *Vadose Zone Journal*, 12(4):vzj2012.0199. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/vzj2012.0199.

Seidel, M., Hutengs, C., Ludwig, B., Thiele-Bruhn, S., and Vohland, M. (2019). Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: Spectral libraries and spiking vs. local calibrations. *Geoderma*, 354:113856.

Sekulic, S., Seasholtz, M. B., Wang, Z., Kowalski, B. R., Lee, S. E., and Holt, B. R. (1993). Nonlinear Multivariate Calibration Methods in Analytical Chemistry. *Analytical Chemistry*, 65(19):835A–845A. Publisher: American Chemical Society.

Seybold, C. A., Ferguson, R., Wysocki, D., Bailey, S., Anderson, J., Nester, B., Schoeneberger, P., Wills, S., Libohova, Z., Hoover, D., and Thomas, P. (2019). Application of Mid-Infrared Spectroscopy in Soil Survey. *Soil Science Society of America Journal*, 83(6):1746–1759. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2019.06.0205.

Shen, Z., Ramirez-Lopez, L., Behrens, T., Cui, L., Zhang, M., Walden, L., Wetterlind, J., Shi, Z., Sudduth, K. A., Baumann, P., Song, Y., Catambay, K., and Viscarra Rossel, R. A. (2022). Deep transfer learning of global spectra for local soil carbon monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:190–200.

Shenk, J. S., Westerhaus, M. O., and Berzaghi, P. (1997). Investigation of a LOCAL Calibration Procedure for near Infrared Instruments. *Journal of Near Infrared Spectroscopy*, 5(4):223–232.

Shenk, J. S., Westerhaus, M. O., and Templeton Jr., W. C. (1985). Calibration Transfer Between near Infrared Reflectance Spectrophotometers1. *Crop Science*, 25(1):cropsci1985.0011183X002500010038x.

Shepherd, K. D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J., and Ge, Y. (2022). A global soil spectral calibration library and estimation service. *Soil Security*, 7:100061.

Shepherd, K. D. and Walsh, M. G. (2002). Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Science Society of America Journal*, 66(3):988–998.

Shepherd, K. D. and Walsh, M. G. (2007). Infrared SpectroscopyEnabling an Evidence-Based Diagnostic Surveillance Approach to Agricultural and Environmental Management in Developing Countries. *Journal of Near Infrared Spectroscopy*, 15(1):1–19.

Sila, A. M., Shepherd, K. D., and Pokhariyal, G. P. (2016). Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemometrics and Intelligent Laboratory Systems*, 153:92–105.

Sinha, A. K. (1987). Variations in soil spectral reflectance related to soil moisture, organic matter and particle size. *Journal of the Indian Society of Remote Sensing*, 15(2):7–11.

Six, J., Conant, R. T., Paul, E. A., and Paustian, K. (2002). Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant and Soil*, 241(2):155–176.

Six, J., Elliott, E., Paustian, K., and Doran, J. W. (1998). Aggregation and Soil Organic Matter Accumulation in Cultivated and Native Grassland Soils. *Soil Science Society of America Journal*, 62(5):1367–1377.

Six, J., Elliott, E. T., and Paustian, K. (1999). Aggregate and soil organic matter dynamics under conventional and no-tillage systems. *Soil Science Society of America Journal*, 63(5):1350–1358.

Skjemstad, J. and Dalal, R. (1987). Spectroscopic and chemical differences in organic matter of two vertisols subjected to long periods of cultivation. *Soil Research*, 25(3):323.

Smith, P. (2004). How long before a change in soil organic carbon can be detected? *Global Change Biology*, 10(11):1878–1883.

Smith, P., Davies, C. A., Ogle, S., Zanchi, G., Bellarby, J., Bird, N., Boddey, R. M., McNamara, N. P., Powlson, D., Cowie, A., van Noordwijk, M., Davis, S. C., Richter, D. D. B., Kryzanowski, L., van Wijk, M. T., Stuart, J., Kirton, A., Eggar, D., Newton-Cross, G., Adhya, T. K., and Braimoh, A. K. (2012). Towards an integrated global framework to assess the impacts of land use and management change on soil carbon: current capability and future vision. *Global Change Biology*, 18(7):2089–2101.

Smith, P., Soussana, J.-F., Angers, D., Schipper, L., Chenu, C., Rasse, D. P., Batjes, N. H., van Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J. E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro Fuentes, J., Sanz-Cobena, A., and Klumpp, K. (2020). How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Global Change Biology*, 26(1):219–241. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14815.

Soares, M. R. and Alleoni, L. R. F. (2008). Contribution of Soil Organic Carbon to the Ion Exchange Capacity of Tropical Soils. *Journal of Sustainable Agriculture*, 32(3):439–462.

Solomatine, D. (2008). Combining Machine Learning and Domain Knowledge in Modular Modelling. In Abrahart, R. J., See, L. M., and Solomatine, D. P., editors, *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications*, Water Science and Technology Library, pages 333–345. Springer, Berlin, Heidelberg.

Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., Macdonald, L. M., and McLaughlin, M. J. (2014). The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. *Applied Spectroscopy Reviews*, 49(2):139–186.

Spencer, S., Ogle, S. M., Breidt, F. J., Goebel, J. J., and Paustian, K. (2011). Designing a national soil carbon monitoring network to support climate change policy: a case example for US agricultural lands. *Greenhouse Gas Measurement and Management*, 1(3-4):167–178.

Stanfill, C. and Waltz, D. (1986). Toward memory-based reasoning. *Communications of the ACM*, 29(12):1213–1228.

Stenberg, B. and Rossel, R. V. (2010). Diffuse Reflectance Spectroscopy for High-Resolution Soil Sensing. In Viscarra Rossel, R. A., McBratney, A. B., and Minasny, B., editors, *Proximal Soil Sensing*, Progress in Soil Science, pages 29–47. Springer Netherlands, Dordrecht.

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., and Wetterlind, J. (2010). Visible and Near Infrared Spectroscopy in Soil Science. In *Advances in Agronomy*, volume 107, pages 163–215. Elsevier.

Stenberg, B. O., Nordkvist, E., and Salomonsson, L. (1995). Use of near infrared reflectance spectra of soils for objective selection of samples. *Soil Science*, 159(2):109–114.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., and van Wesemael, B. (2013). Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE*, 8(6):e66409.

Stevens, A. and Ramirez-Lopez, L. (2013). An introduction to the prospectr package.

Stevens, A. and RamirezLopez, L. (2014). An introduction to the prospectr package.

Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., and Ben-Dor, E. (2008). Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma*, 144(1):395–404.

Stevens, A., Wesemael, B. v., Vandenschrick, G., Touré, S., and Tychon, B. (2006). Detection of Carbon Stock Change in Agricultural Soils Using Spectroscopic Techniques. *Soil Science Society of America Journal*, 70(3):844–850.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, pages 111–147.

Stoner, E. R., Baumgardner, M. F., Weismiller, R. A., Biehl, L. L., and Robinson, B. F. (1980). Extension of Laboratory-measured Soil Spectra to Field Conditions1. *Soil Science Society of America Journal*, 44(3):572.

Stumpf, F., Keller, A., Schmidt, K., Mayr, A., Gubler, A., and Schaepman, M. (2018). Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. *Agriculture, Ecosystems & Environment*, 258:129–142.

Summerauer, L., Baumann, P., Ramirez-Lopez, L., Barthel, M., Bauters, M., Bukombe, B., Reichenbach, M., Boeckx, P., Kearsley, E., Van Oost, K., Vanlauwe, B., Chiragaga, D., Heri-Kazi, A. B., Moonen, P., Sila, A., Shepherd, K., Bazirake Mujinya, B., Van Ranst, E., Baert, G., Doetterl, S., and Six, J. (2021). The central African soil spectral library: a new soil infrared repository and a geographical prediction analysis. *SOIL*, 7(2):693–715.

Syers, J. K., Campbell, A. S., and Walker, T. W. (1970). Contribution of organic carbon and clay to cation exchange capacity in a chronosequence of sandy soils. *Plant and Soil*, 33(1-3):104–112.

Tan, C. and Li, M. (2007). Calibration transfer between two near-infrared spectrometers based on a wavelet packet transform. *Analytical sciences*, 23(2):201–206.

Teng, H. T., Viscarra Rossel, R. A., Shi, Z., and Behrens, T. (2018). Updating a national soil classification with spectroscopic predictions and digital soil mapping. *CATENA*, 164:125–134.

Terhoeven-Urselmans, T., Schmidt, H., Georg Joergensen, R., and 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(5):1178–1188.

Terhoeven-Urselmans, T., Vagen, T.-G., Spaargaren, O., and Shepherd, K. D. (2010). Prediction of Soil Fertility Properties from a Globally Distributed Soil Mid-Infrared Spectral Library. *Soil Science Society of America Journal*, 74(5):1792.

Thrun, S. and Pratt, L., editors (1998). *Learning to Learn*. Springer US, Boston, MA.

Tiessen, H., Cuevas, E., and Chacon, P. (1994). The role of soil organic matter in sustaining soil fertility. *Nature*, 371(6500):783–785.

Tsakiridis, N. L., Keramaris, K. D., Theocharis, J. B., and Zalidis, G. C. (2020). Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. *Geoderma*, 367:114208.

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., and Zalidis, G. (2019). A memory-based learning approach utilizing combined spectral sources and geographical proximity for improved VIS-NIR-SWIR soil properties estimation. *Geoderma*, 340:11–24.

Tóth, G., Jones, A., and Montanarella, L. (2013). The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environmental Monitoring and Assessment*, 185(9):7409–7425.

UNEP (2012). Land Health Surveillance: An Evidence-Based Approach to Land Ecosystem Management. Illustrated with a Case Study in the West Africa Sahel.

Vagen, T.-G., Shepherd, K. D., Walsh, M. G., Winowiecki, L., Desta, L. T., and Tondoh, J. E. (2010). AfSIS technical specifications: Soil Health Surveillance.

Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7:91.

Varmuza, K. and Filzmoser, P. (2016). *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press, 0 edition.

Vasques, G. M., Demattê, J. A. M., Viscarra Rossel, R. A., Ramírez-López, L., and Terra, F. S. (2014). Soil classification using visible/near-infrared diffuse reflectance spectra from multiple depths. *Geoderma*, 223-225:73–78.

Vasques, G. M., Demattê, J. A. M., Viscarra Rossel, R. A., Ramírez López, L., Terra, F. S., Rizzo, R., and De Souza Filho, C. R. (2015). Integrating geospatial and multi-depth laboratory spectral data for mapping soil classes in a geologically complex area in southeastern Brazil: Geospatial and spectral soil class mapping. *European Journal of Soil Science*, 66(4):767–779.

Vasques, G. M., Grunwald, S., and Harris, W. G. (2010). Spectroscopic Models of Soil Organic Carbon in Florida, USA. *Journal of Environment Quality*, 39(3):923.

VillasBoas, P. R., Franco, M. A., MartinNeto, L., Gollany, H. T., and Milori, D. M. B. P. (2020). Applications of laser-induced breakdown spectroscopy for soil analysis, part I: Review of fundamentals and chemical and physical properties. *European Journal of Soil Science*, 71(5):789–804.

Viscarra Rossel, R., Adamchuk, V., Sudduth, K., McKenzie, N., and Lobsey, C. (2011). Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. In *Advances in Agronomy*, volume 113, pages 243–291. Elsevier.

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

Viscarra Rossel, R., Walvoort, D., McBratney, A., Janik, L., and Skjemstad, J. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131(1-2):59–75.

Viscarra Rossel, R. A. and Bouma, J. (2016). Soil sensing: A new paradigm for agriculture. *Agricultural Systems*, 148:71–74.

Viscarra Rossel, R. A. and Hicks, W. S. (2015). Soil organic carbon and its fractions estimated by visiblenear infrared transfer functions. *European Journal of Soil Science*, 66(3):438–450.

Viscarra Rossel, R. A., Lobsey, C. R., Sharman, C., Flick, P., and McLachlan, G. (2017). Novel Proximal Sensing for Monitoring Soil Organic C Stocks and Condition. *Environmental Science & Technology*, 51(10):5630–5641.

Viscarra Rossel, R. A. V. and Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158(1):46–54.

Viscarra Rossel, R. A. V. and Webster, R. (2012). Predicting soil properties from the Australian soil visiblenear infrared spectroscopic database. *European Journal of Soil Science*, 63(6):848–860.

Viscarrra Rossel, R. A. V. and McBratney, A. B. (1998). Soil chemical analytical accuracy and costs: implications from precision agriculture. *Australian Journal of Experimental Agriculture*, 38(7):765.

Visconti, F. and de Paz, J. M. (2021). Estimation of the carbon valence from its average formal oxidation state in the soil organic matter. *European Journal of Soil Science*, 72(5):2225–2230.

Vohland, M., Ludwig, B., Seidel, M., and Hutengs, C. (2022). Quantification of soil organic carbon at regional scale: Benefits of fusing vis-NIR and MIR diffuse reflectance data are greater for in situ than for laboratory-based modelling approaches. *Geoderma*, 405:115426.

Volper, D. J. and Hampson, S. E. (1987). Learning and using specific instances. *Biological Cybernetics*, 57(1):57–71.

Vågen, T.-G., Davey, F. A., and Shepherd, K. D. (2012). Land Health Surveillance: Mapping Soil Carbon in Kenyan Rangelands. In Nair, P. R. and Garrity, D., editors, *Agroforestry - The Future of Global Land Use*, Advances in Agroforestry, pages 455–462. Springer Netherlands, Dordrecht.

Vågen, T.-G., Winowiecki, L. A., Desta, L., Tondoh, E. J., Weullow, E., Shepherd, K., and Sila, A. (2020). Mid-Infrared Spectra (MIRS) from ICRAF Soil and Plant Spectroscopy Laboratory: Africa Soil Information Service (AfSIS) Phase I 2009-2013. Type: dataset.

Wadoux, A. M.-C. and Brus, D. J. (2021). How to compare sampling designs for mapping? *European Journal of Soil Science*, 72(1):35–46.

Wadoux, A. M. J. C., Heuvelink, G. B. M., de Bruin, S., and Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling*, 457:109692.

Wang, H. and Liu, X. (2021). Light People: Tarik Bourouina. *Light: Science & Applications*, 10(1):165, s41377–021–00602–w.

Wang, T., Huan, J., and Zhu, M. (2019). Instance-Based Deep Transfer Learning. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 367–375. ISSN: 1550-5790.

Wang, Y. and Witten, I. H. (1996). Induction of model trees for predicting continuous classes. Working Paper 96/23, University of Waikato, Department of Computer Science, Hamilton, New Zealand.

Wawire, A. W., Csorba, ., Kovács, E., Mairura, F. S., Tóth, J. A., and Michéli, E. (2021). Comparing farmers soil fertility knowledge systems and scientific assessment in Upper Eastern Kenya. *Geoderma*, 396:115090.

Wetterlind, J. and Stenberg, B. (2010). Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. *European Journal of Soil Science*, 61(6):823–843.

Wetterlind, J., Stenberg, B., and Söderström, M. (2008). The use of near infrared (NIR) spectroscopy to improve soil mapping at the farm scale. *Precision Agriculture*, 9(1-2):57–69.

Wickham, H. (2019). tidyverse: Easily install and load the 'tidyverse'. `https://CRAN.R-project.org/package=tidyverse` [Accessed: 2022-05-25].

Wilson, D. R. and Martinez, T. R. (2000a). An Integrated Instance-Based Learning Algorithm. *Computational Intelligence*, 16(1):1–28.

Wilson, D. R. and Martinez, T. R. (2000b). Reduction Techniques for Instance-Based Learning Algorithms. *Machine Learning*, 38(3):257–286.

Wold, S., Johansson, E., and Cocchi, M. (1993). PLS-partial least squares projections to latent structures. *3D QSAR in drug design*, 1:523–550.

Wold, S., Martens, H., and Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. In *Matrix pencils*, pages 286–293. Springer.

Wold, S. and Sjöström, M. (1977). SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. In *Chemometrics: Theory and Application*, volume 52 of *ACS Symposium Series*, pages 243–282. American Chemical Society. Section: 12.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130.

Wolpert, D. and Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82.

Wolpert, D. H. (1995). Off-training set error and a priori distinctions between learning algorithms. Technical report.

Wolpert, D. H. (1996). The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation*, 8(7):1341–1390.

Wong, M., Wittwer, K., Oliver, Y., and Robertson, M. (2010). Use of EM38 and Gamma Ray Spectrometry as Complementary Sensors for High-Resolution Soil Property Mapping. In Viscarra Rossel, R. A., McBratney, A. B., and Minasny, B., editors, *Proximal Soil Sensing*, Progress in Soil Science, pages 343–349. Springer Netherlands, Dordrecht.

Workman, J. J., Mobley, P. R., Kowalski, B. R., and Bro, R. (1996). Review of Chemometrics Applied to Spectroscopy: 1985-95, Part I. *Applied Spectroscopy Reviews*, 31(1-2):73–124.

Xie, H., Zhao, J., Wang, Q., Sui, Y., Wang, J., Yang, X., Zhang, X., and Liang, C. (2015). Soil type recognition as improved by genetic algorithm-based variable selection using near infrared spectroscopy and partial least squares discriminant analysis. *Scientific Reports*, 5:10930.

Xu, C., Xu, X., Liu, M., Liu, W., Yang, J., Luo, W., Zhang, R., and Kiely, G. (2017a). Enhancing pedotransfer functions (PTFs) using soil spectral reflectance data for estimating saturated hydraulic conductivity in southwestern China. *CATENA*, 158:350–356.

Xu, D., Wang, Y., Xiong, B., and Li, T. (2017b). MEMS-based thermoelectric infrared sensors: A review. *Frontiers of Mechanical Engineering*, 12(4):557–566.

Yang, Y., Shen, Z., Bissett, A., and Viscarra Rossel, R. A. (2022). Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. *SOIL*, 8(1):223–235. Publisher: Copernicus GmbH.

Zhang, Y. and Hartemink, A. E. (2021). Quantifying short-range variation of soil texture and total carbon of a 330-ha farm. *CATENA*, 201:105200.

Zheng, G., Jiao, C., Zhou, S., and Shang, G. (2016). Analysis of soil chronosequence studies using reflectance spectroscopy. *International Journal of Remote Sensing*, 37(8):1881–1901.

Zimmermann, M., Leifeld, J., and Fuhrer, J. (2007). Quantifying soil organic carbon fractions by infrared-spectroscopy. *Soil Biology and Biochemistry*, 39(1):224–231.