Bachelorarbeit in Bodenwissenschaften

Wie robust können Bodeneigenschaften mit reduzierter Probenaufbereitung spektral bestimmt werden? Methodischer Vergleich der Vorhersagegenauigkeit von midIR Messungen und Machine Learning Modellen mit und ohne Mahlen der Proben.

Die Analyse von Bodenproben anhand ihrer spektralen Eigenschaften im mittleren Infrarot entwickelt sich zunehmend zu einem neuen Standard in den Bodenwissenschaften. Die Diffuse Reflektions-Spektroskopie (DRIFTS) ist kostengünstigen (Camargo *et al.*, 2022), nicht-destruktiv (Helfenstein *et al.*, 2021) und erlaubt es mit den entsprechenden Machine-Learning Modellen eine Vielzahl von Variablen zu analysieren (Ng *et al.*, 2019). In dieser Bachelorarbeit geht es um einen methodischen Vergleich der Vorhersagegenauigkeit mit und ohne Probenvorbereitung durch Mahlen, was aktuell diskutiert wird (Sanderman *et al.*, 2023).

Du wirst eine der aktuellsten Methoden der Bodenanalytik inkl. der Verwendung von Machine Learning kennenlernen. Du kannst deine Fähigkeiten in R oder Python aufpeppen. Und du wirst eine Abschlussarbeit schreiben, die tatsächlich von wissenschaftlicher Relevanz ist. Dafür erwarten dich ca. 2 Wochen Laborarbeit und einiges an Datenanalyse. Wir werden dich in alles in Ruhe einarbeiten. Du solltest jedoch etwas Interesse an Bodeneigenschaften und Machine Learning mitbringen.

Hintergrund und Problemstellung

Um steigende Anforderungen an die zeitliche und räumliche Auflösung von Bodenmonitoring-Konzepten erfüllen zu können, ist ein erhöhter Probendurchsatz nötig. Um den Zeit- und Kostenaufwand zu reduzieren, sind neue Analysetechniken daher unerlässlich. Eine vielversprechende Methode zur schnellen und kostengünstigen (Camargo *et al.*, 2022), sowie nicht-destruktiven (Helfenstein *et al.*, 2021) Analyse von Bodenproben ist die Diffuse Reflektions-Spektroskopie (DRIFTS). Diese Methode ermöglicht die simultane Bestimmung verschiedener Bodeneigenschaften (u.a. C, N, P-Gehalte, Textur, …) über multivariate oder machine-learning-basierte Kalibrationsmodelle. Dabei können sowohl der Nahinfrarotbereich (NIR) als auch der mittlere Infrarotbereich (midIR) genutzt werden (Ng *et al.*, 2019). Der midIR-Bereich bietet dabei einen deutlich höheren Informationsgehalt als der NIR-Bereich, da im midIR viele relevante primäre Absorbtionsbänder zu finden sind. Im Vergleich zur NIR-Spektroskopie ist dies jedoch auch mit einer erhöhten Störanfälligkeit verbunden (Reeves *et al.*, 2005; Vohland *et al.*, 2011). Folglich wird in der Literatur meist eine sorgfältige Probenaufbereitung empfohlen. Hierzu zählen:

a) das (schonende) Trocknen der Probe (Barthès *et al.*, 2006; Margenot *et al.*, 2016; Laub *et al.*, 2019). Wasser absorbiert ebenfalls im midIR Bereich. Daher sollte der Wassergehalt in den Proben möglichst gering sein. Ein zu aggressives Trocken kann jedoch Einfluss auf die chemische Zusammensetzung der Probe nehmen.
b) das feine Mahlen der Probe (Le Guillou *et al.*, 2015; Seybold *et al.*, 2019; Deiss *et al.*, 2020; Baumann *et al.*, 2021). Durch das feine Mahlen der Probe werden mehrere Effekte erzielt. Grobe Mineraloberflächen werden zerstört. Dies verhindert die Störung der Messung durch spekulare (direkte) Reflektion des IR-Lichts in den Detektor. Das Zerkleinern der Probe verhindert außerdem, dass nur die Oberfläche von Partikeln erfasst wird. Stattdessen kann eine möglichst homogene Probe erzeugt werden; der Einfluss von Partikelcoatings wird reduziert und das Innere von Aggregaten wird sichtbar.

Während das feine Mahlen von Bodenproben derzeit der Standard in der midDRIFTS ist, kann die Analyse aber auch mit ungemahlenen Proben durchgeführt werden. Folglich stellt sich die Frage, ob der Mehraufwand des zusätzlichen Aufbereitungsschritts durch einen signifikanten Gewinn an Vorhersagegenauigkeit gerechtfertigt ist (Sanderman *et al.*, 2023).
Des Weiteren gestaltet sich die Etablierung einer einheitlichen Mahlprozedur (Gerät, Mahldauer, -intensität etc.) über Labore / Institutionen hinweg schwierig, da selten identische Geräte zur Verfügung stehen. Ähnlich wie bei der Anwendung von Kalibrationsmodellen auf Daten die mit unterschiedlichen Spektrometern verschiedener Hersteller (vgl. Safanelli *et al.*, 2023) aufgezeichnet wurden, ist hier auch mit einem Rückgang der Vorhersagegenauigkeit zu rechnen.

Fragestellung / Aufgaben

Ziel der Bachelorarbeit soll es sein, die quantitative chemometrischen Analyse mittels midDRIFTS unter Verwendung von ungemahlenen Proben ggü. der Standardprozedur zu evaluieren.
Dafür stehen knapp 300 luftgetrocknte Feinbodenproben (≤ 2 mm) von landwirtschaftlichen Dauerbeobachtungsflächen Sachsens zur Verfügung. Deren Spektren wurden bereits nach Standardprozedur (gemahlen) gemessen. Nun sollen die midIR Spektren ohne weitere Aufbereitung (ungemahlen) gemessen und verglichen werden. Dafür ist neben der Laborarbeit (ca. 2 Wochen) vor allem Arbeit mit den Daten (Vorbereitung mit Glättung, Streuungskorrektur etc. und Verwendung in multivariaten ML Modellen wie PLS, Cubist oder MBL) nötig. Die so erzeugten Modelle sollen hinsichtlich ihrer Vorhersagegenauigkeit für ausgewählte Variablen (mindestens C_{ORG} und Textur) mit und ohne Mahlen verglichen werden.

Je nach Vorliebe kann die Datenaufbereitung und -auswertung in R oder Python erfolgen. In jedem Fall wird es eine Einweisung und Einarbeitung in die relevanten Pakete geben. Wir haben auch einige Skripte, mit denen gestartet werden kann. So können während der Bearbeitung der Bachelorarbeit die Kenntnisse in R oder Python vertieft werden.

Anforderungen

Die Arbeitspakete setzen sich neben der Literaturrecherche und dem Verfassen der Arbeit aus einem bodenanalytischen Teil und einem statistischen Teil (Chemometrie) zusammen. Für das allgemeine Verständnis und den Bodenanalytischen Teil ist Teilnahme an vertiefenden bodenkundlichen Modulen (z.B. Pedologie) vorteilhaft, jedoch keine Voraussetzung.
Die Auswertung der gewonnenen Daten erfordert zumindest solide Grundkentnisse in R oder Python. Du solltest Spaß an der Datenanalyse haben und offen für die Einarbeitung in weiterführende Methoden und Pakete (in R tidyverse, resemble, prospectr, simplerspec; bzw. äquivalente Pakete in Python) sein.

Zeitplan

Thematische Einarbeitung und Literaturrecherche	2 Wochen
Analyse der ungemahlenen Bodenproben	2 Wochen
Einarbeitung in spezifische R packages, Aufbereitung der Spektren und Erstellen von Kalibrationsmodellen	4 Wochen
Verfassen der Arbeit	4 Wochen

References

Barthès, B.G., Brunet, D., Ferrer, H., Chotte, J.-L. & Feller, C. (2006) Determination of Total Carbon and Nitrogen Content in a Range of Tropical Soils Using near Infrared Spectroscopy: Influence of Replication and Sample Grinding and Drying. *Journal of Near Infrared Spectroscopy*, 14(5), 341–348. Available from: https://doi.org/10.1255/jnirs.686.

Baumann, P., Helfenstein, A., Gubler, A., Keller, A., Meuli, R.G. & Wächter, D. et al. (2021) Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring. *SOIL*, 7(2), 525–546. Available from: https://doi.org/10.5194/soil-7-525-2021.

Camargo, L.A., Amaral, L.R., Reis, A.A., Brasco, T.L. & Magalh├úes, P.S.G. (2022) Improving soil organic carbon mapping with a field-specific calibration approach through diffuse reflectance spectroscopy and machine learning algorithms. *Soil Use and Management*, 38(1), 292–303. Available from: https://doi.org/10.1111/sum.12775.

Deiss, L., Culman, S.W. & 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. Available from: https://doi.org/10.1002/saj2.20021.

Helfenstein, A., Baumann, P., Rossel, R.V., Gubler, A., Oechslin, S. & Six, J. (2021) Quantifying soil carbon in temperate peatlands using a mid-IR soil spectral library. *SOIL*, 7(1), 193–215. Available from: https://doi.org/10.5194/soil-7-193-2021.

Laub, M., Blagodatsky, S., Nkwain, Y.F. & Cadisch, G. (2019) Soil sample drying temperature affects specific organic mid-DRIFTS peaks and quality indices. *Geoderma*, 355, 113897. Available from: https://doi.org/10.1016/j.geoderma.2019.113897.

Le Guillou, F., Wetterlind, W., Viscarra Rossel, R.A., Hicks, W., Grundy, M. & Tuomi, S. (2015) How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? *Soil Research*, 53(8), 913. Available from: https://doi.org/10.1071/SR15019.

Margenot, A.J., Calderón, F.J., Goyne, K.W., Mukome, F. & Parikh, S.J. (2016) IR Spectroscopy, Soil Analysis Applications. In: Lindon, J.C., Tranter, G.E. & Koppenaal, D. (Eds.) *Encyclopedia of Spectroscopy and Spectrometry.* Elsevier Science: San Diego, pp. 448–454.

Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R. & Bailey, S. et al. (2019) Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*, 352, 251–267. Available from: https://doi.org/10.1016/j.geoderma.2019.06.016.

Reeves, J.B., Francis, B.A. & Hamilton, S.K. (2005) Specular reflection and diffuse reflectance spectroscopy of soils. *Applied Spectroscopy*, 59(1), 39–46. Available from: https://doi.org/10.1366/0003702052940431.

Sanderman, J., Smith, C., Safanelli, J.L., Morgan, C.L., Ackerman, J. & Looker, N. et al. (2023) Diffuse reflectance mid-infrared spectroscopy is viable without fine milling. *Soil Security*, 13, 100104. Available from: https://doi.org/10.1016/j.soisec.2023.100104.

Seybold, C.A., Ferguson, R., Wysocki, D., Bailey, S., Anderson, J. & Nester, B. et al. (2019) Application of Mid-Infrared Spectroscopy in Soil Survey. *Soil Science Society of America Journal*, 83(6), 1746–1759. Available from: https://doi.org/10.2136/sssaj2019.06.0205.

Vohland, M., Besold, J., Hill, J. & Fründ, H.-C. (2011) Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma*, 166(1), 198–205. Available from: https://doi.org/10.1016/j.geoderma.2011.08.001.