An interlaboratory comparison of mid-infrared spectra acquisition: Instruments and procedures matter

54 authors, including:

José Lucas Safanelli
Woodwell Climate Research Center
56 PUBLICATIONS   1,152 CITATIONS

SEE PROFILE

Jonathan Sanderman
Woodwell Climate Research Center
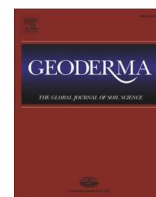154 PUBLICATIONS   11,179 CITATIONS

SEE PROFILE

Tomislav Hengl
OpenGeoHub foundation
208 PUBLICATIONS   17,410 CITATIONS

SEE PROFILE

Eyal Ben-Dor
Tel Aviv University
328 PUBLICATIONS   15,099 CITATIONS

SEE PROFILE

An interlaboratory comparison of mid-infrared spectra acquisition: Instruments and procedures matter

José L. Safanelli [a], Jonathan Sanderman [a,*], Dellena Bloom [b], Katherine Todd-Brown [b], Leandro L. Parente [c], Tomislav Hengl [c], Sean Adam [d], Franck Albinet [e,ah], Eyal Ben-Dor [f], Claudia M. Boot [g], James H. Bridson [h], Sabine Chabrillat [i,ag], Leonardo Deiss [j], José A.M. Demattê [k], M. Scott Demyan [l], Gerd Dercon [e], Sebastian Doetterl [m], Fenny van Egmond [n], Rich Ferguson [o], Loretta G. Garrett [h], Michelle L. Haddix [p], Stephan M. Haefele [q], Maria Heiling [e], Javier Hernandez-Allica [q], Jingyi Huang [r], Julie D. Jastrow [s], Konstantinos Karyotis [t], Megan B. Machmuller [p], Malefetsane Khesuoe [u], Andrew Margenot [v], Roser Matamala [s], Jessica R. Miesel [w], Abdul M. Mouazen [x], Penelope Nagel [y], Sunita Patel [h], Muhammad Qaswar [x], Selebalo Ramakhanna [u], Christian Resch [e], Jean Robertson [z], Pierre Roudier [aa], Marmar Sabetizade [ab], Itamar Shabtai [ac], Faisal Sherif [w], Nishant Sinha [ad], Johan Six [m], Laura Summerauer [m], Cathy L. Thomas [q], Arsenio Toloza [e], Beata Tomczyk-Wójtowicz [ae], Nikolaos L. Tsakiridis [t], Bas van Wesemael [ab], Finnleigh Woodings [v], George C. Zalidis [t], Wiktor R. Żelazny [af]

[a] Woodwell Climate Research Center, Falmouth, MA, USA
[b] Department of Environmental Engineering Sciences, University of Florida, Gainesville, FL, USA
[c] OpenGeoHub Foundation, Wageningen, the Netherlands
[d] TU Bergakademie Freiberg, Freiberg, Germany
[e] Soil and Water Management and Crop Nutrition Laboratory, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency, Austria
[f] Remote Sensing Laboratory, Department of Geography and Human Environment, Porter School of Environment and Earth Science, Tel Aviv University, 69978 Tel Aviv, Israel
[g] Department of Chemistry, Colorado State University, Fort Collins, CO, USA
[h] Scion, Private Bag 3020, Rotorua 3046, New Zealand
[i] Helmholtz Center Potsdam GFZ German Research Center for Geosciences, Potsdam, Germany
[j] The Ohio State University, Wooster, OH, USA
[k] Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, SP, Brazil
[l] The Ohio State University, Columbus, OH, USA
[m] Department of Environmental Systems Science, ETH Zurich, 8092 Zurich, Switzerland
[n] ISRIC World Soil Information, Wageningen, the Netherlands
[o] USDA-NRCS Kellogg Soil Survey Laboratory, Lincoln, NE, USA
[p] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA
[q] Sustainable Soils and Crops, Rothamsted Research, UK
[r] University of Wisconsin-Madison, Madison, WI, USA
[s] Environmental Science Division, Argonne National Laboratory, Lemont, IL, USA
[t] Laboratory of Remote Sensing, Spectroscopy, and GIS, Department of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[u] Department of Agricultural Research, Maseru, Lesotho
[v] University of Illinois Urbana-Champaign, Urbana, IL, USA
[w] Michigan State University, Department of Plant, Soil and Microbial Sciences, East Lansing, MI, USA
[x] Department of Environment, Ghent University, Coupure Links 653, 9000 Gent, Belgium
[y] Persistence Data Mining Inc., La Jolla, CA, USA
[z] The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK
[aa] Manaaki Whenua - Landcare Research, Private Bag 11052, Manawatū Mail Centre, Palmerston North 4442, New Zealand
[ab] Earth and Life Institute, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium
[ac] Department of Environmental Science and Forestry, Connecticut Agricultural Experiment Station, New Haven, CT, USA

* Corresponding author.
E-mail address: jsanderman@woodwellclimate.org (J. Sanderman).

[ad] *Division of Soil Physics, ICAR-IISS, Bhopal, India*
[ae] *AgroCares Golden Standard Laboratory, Care4Agro B.V., 6709PA Wageningen, the Netherlands*
[af] *Division of Crop Management Systems, Crop Research Institute Praha-Ruzyně, Praha, Czech Republic*
[ag] *Leibniz University Hannover, Institute of Soil Science, Hannover, Germany*
[ah] *Soil and Water Management Unit, KU Leuven, Leuven, Belgium*

ARTICLE INFO

ABSTRACT

Diffuse reflectance spectroscopy has been extensively employed to deliver timely and cost-effective predictions of a number of soil properties. However, although several soil spectral laboratories have been established worldwide, the distinct characteristics of instruments and operations still hamper further integration and interoperability across mid-infrared (MIR) soil spectral libraries. In this study, we conducted a large-scale ring trial experiment to understand the lab-to-lab variability of multiple MIR instruments. By developing a systematic evaluation of different mathematical treatments with modeling algorithms, including regular preprocessing and spectral standardization, we quantified and evaluated instruments' dissimilarity and how this impacts internal and shared model performance. We found that all instruments delivered good predictions when calibrated internally using the same instruments' characteristics and standard operating procedures by solely relying on regular spectral preprocessing that accounts for light scattering and multiplicative/additive effects, e.g., using standard normal variate (SNV). When performing model transfer from a large public library (the USDA NSSC-KSSL MIR library) to secondary instruments, good performance was also achieved by regular preprocessing (e.g., SNV) if both instruments shared the same manufacturer. However, significant differences between the KSSL MIR library and contrasting ring trial instruments responses were evident and confirmed by a semi-unsupervised spectral clustering. For heavily contrasting setups, spectral standardization was necessary before transferring prediction models. Non-linear model types like Cubist and memory-based learning delivered more precise estimates because they seemed to be less sensitive to spectral variations than global partial least square regression. In summary, the results from this study can assist new laboratories in building spectroscopy capacity utilizing existing MIR spectral libraries and support the recent global efforts to make soil spectroscopy universally accessible with centralized or shared operating procedures.

1. Introduction

Soil science has a data problem. There is simply not enough laboratory capacity to meet the many needs for timely and accurate information on basic soil properties. The application of diffuse reflectance spectroscopy (DRS) to soil and environmental sciences is rapidly maturing, and for soils has already contributed to generating quantitative information (Frei and MacNeil, 2019; Nocita et al., 2015; Viscarra Rossel et al., 2022). DRS can increase laboratory throughput at a fraction of the cost of traditional wet chemistry methods (Seybold et al., 2019). But in order for DRS to be successfully deployed in laboratory-based conditions, representative spectral libraries and the appropriate data measurements are necessary (Shepherd et al., 2022).

Soil spectral libraries along with corresponding laboratory measurements of several attributes (reference values) have been established worldwide, spanning from very localized datasets to national or continental databases (Ramirez-Lopez et al., 2019; Summerauer et al., 2021; Viscarra Rossel et al., 2016). These libraries often have distinct features that require significant preprocessing and harmonization in order to make the data consistent and useful before their integrated use (Francos et al., 2023; Minasny et al., 2009). Special attention to this point has been made by the Global Soil Partnership (GLOSOLAN) of the Food and Agriculture Organization (FAO) from the United Nations and its Global Soil Laboratory Network initiative on soil spectroscopy (FAO GLOSOLAN, 2023) and the Institute of Electrical and Electronics Engineers (IEEE) Standards Association (AS) P4005 working group, which aims to define the standards and protocols for soil spectroscopy (IEEE SA, 2023).

The main sources of variability when merging or working across soil spectral libraries come from the precision and reproducibility of soil analytical data used to calibrate spectral data and the collected spectra themselves. Laboratory reference data used to calibrate models from spectra are affected by cumulative issues on sample preparation, sub-sampling, and instrument readings (Shepherd et al., 2022). Many laboratories routinely monitor their analytical precision over time by using internal or external standard samples. But high variability between replicates may still remain an obstacle as reported by FAO GLOSOLAN in a global ring experiment of soil analytical determination (Suvannang and Hartmann, 2019). High intra and interlaboratory variability was found in a Forest Soil Survey project when three soil samples were shared across 52 laboratories in European countries (Cools et al., 2004). Although the same reference analytical methods were run on ring trial samples, it was not possible to link the causes of variation with the metadata provided by the participants. In a recent study, DRS was tested as a tool for routine use by commercial laboratories for internal quality control and as a complementary method for soil property estimation (Poppiel et al., 2022). The study suggested that spectral variations can be successfully employed for identifying analytical outliers with unsupervised clustering, and some laboratories can use DRS in their daily routine to reduce costs of determination, especially for soil particle size distribution and organic carbon content. Moreover, the analysis of spectral data can indicate inconsistencies in wet laboratory analyses, and thus has a quality evaluation importance.

A remaining research gap in soil spectroscopy is to quantify the variation in spectral responses across different instruments and the influence of this variation on predictions made using centralized or combined soil spectral libraries. A seminal study found large differences among the DRS measurements of three different instruments across the visible, near and shortwave infrared region (VNIR-SWIR, 400–2500 nm [25000–4000 cm^{-1}]) (Pimstein et al., 2011), which formed the bases of the Internal Soil Standard (ISS) method later recommended for this spectral region of interest (Ben-Dor et al., 2015). Gholizadeh et al. (2021) found that the ISS method was able to mostly resolve differences between four VNIR-SWIR instruments when predicting soil organic carbon content. On the other hand, although DRS with Fourier-Transform in the mid-infrared region (FT-MIR, 2500–25000 nm [4000–400 cm^{-1}]) has been promoted as an accurate method to interpret and predict soil properties (Ng et al., 2022; Sanderman et al., 2020), there is a lack of studies assessing different instruments as part of a large-scale interlaboratory comparison using FT-MIR. Previous studies have found contradicting results when transferring FT-MIR prediction models from centralized soil spectral libraries to a limited number of instruments and smaller datasets. Dangal and Sanderman (2020) found

that spectral standardization was necessary to deliver unbiased predictions for a test set of different geographical origin despite reaching good-to-excellent results without standardization. In turn, Sanderman et al. (2023) suggested that calibration transfer may not be necessary when they transferred a model from the USDA National Soil Survey Center Kellogg Soil Survey Laboratory (NSSC KSSL) FT-MIR spectral library to a localized spectral library collected with a different instrument model from the same manufacturer, with spectral preprocessing being enough to handle the spectral variability.

Thus, reducing spectral variations through spectra preprocessing is extremely important because FT-MIR laboratory measurements are still subject to particle size, light scattering, and multicollinearity issues (Barnes et al., 1989), especially when different instrumentation and operating procedures may change wavelength position, absorption shape, and albedo intensity (Pimstein et al., 2011). Spectral standardization or calibration transfer has been promoted in FT-MIR soil spectroscopy but only more recently this strategy has been further investigated (Pittaki-Chrysodonta et al., 2021). Similarly, with the widespread use of machine learning and chemometrics algorithms (Barra et al., 2021), some model architectures may be less sensitive to spectral variations due to their nature of handling large and complex datasets.

In this study, we conduct to our knowledge the first large-scale ring trial experiment to better understand the lab-to-lab variability of multiple soil spectroscopy laboratories from around the world. The goal of this study is to conduct a systematic evaluation of different spectra preprocessing and modeling algorithms for reducing instruments' variability in order to deliver better predictions from DRS FT-MIR. Simultaneously, the shared standard samples may form the basis of building calibration transfer among soil spectroscopy laboratories, especially in relation to the centralized USDA NSSC KSSL FT-MIR soil spectral library that can be leveraged to calibrate more robust prediction models.

2. Material & methods

A ring trial experiment is an interlaboratory comparison where the same set of samples are prepared using standard procedures and shared across multiple laboratories. The goal of such experimental design is to evaluate the reproducibility of an analysis (accuracy/bias) across the participant laboratories (Van Reeuwijk and Houba, 1998). However, the same set of samples can also be used for internal evaluation of the repeatability (precision/variance) as part of an quality control protocol. With the quantification of these sources of errors (bias and variance), one can ultimately propose solutions for better compatibility by sharing best practices among the laboratories network. In soil science, routine and research laboratories have been extensively tested as part of proficient tests programs (Cools et al., 2004; Rayment et al., 2000; Wolf and Miller, 1998). In soil spectroscopy, however, this type of experimental setup has not been used in routine laboratories and a few attempts are described in the literature (Ben-Dor et al., 2015; Gholizadeh et al., 2021; Pimstein et al., 2011). In the following subsections, we describe how we established a large-scale soil spectroscopy ring trial network to evaluate FT-MIR instruments and laboratories.

2.1. Soil samples

Sixty soil samples that are routinely analyzed at the USDA NSSC KSSL (hereafter KSSL) and used for process quality control were prepared as fine earth fraction (FE, i.e., air-dried and sieved < 2 mm) and a finely milled fraction (FM, which is FE run 3 min in a Retsch Mixer mill reaching < 180 μm) at Woodwell Climate Research Center for later distribution to ring trial participants. Additionally, ten samples from the Soil Science Society of America's North American Proficiency Testing Program (NAPT) were also included in the package and prepared by the same approaches. Using a riffle-splitter, aliquots of about 15 g of FM fraction were divided out for each sample into 20 ml glass scintillation

vials. These sets of 70 samples prepared to FM fraction were then shipped to the 20 participating laboratories (Fig. 1a).

In this study, OC content (gravimetric percentage, %), clay content (gravimetric percentage, %), pH (log10 units), and exchangeable potassium (K, $cmol_c$ kg^{-1}) were chosen as soil properties of interest because they usually present variable predictive model performance and have distinct relationships with absorbance patterns in the MIR spectra (Dangal et al., 2019). NAPT OC is the only property that does not necessarily have the same method as KSSL, with its reference values being defined as the median of multiple measurements and methods. For the KSSL soil set, OC was estimated by $OC = TC - 0.12CaCO3$, with TC measured by dry combustion and $CaCO_3$ determined by a manometer. Clay content was measured using the pipette method of particle size analysis. pH was measured in a 1:1 soil-to-water solution using an ion-selective electrode. Finally, exchangeable K was determined using ammonium acetate (pH 7) (Soil Survey Staff, 2022).

For data analysis, the two soil sets were pooled together to form a larger dataset (n = 70) and a potential incompatibility between KSSL and NAPT analytical results for total carbon (TC) and estimated organic carbon (OC) was verified by running the KSSL analytical procedures (Soil Survey Staff, 2022) (Fig. 1b). The results revealed a substantial correspondence of KSSL analytical procedures ($R^2 >= 0.99$) to median NAPT values for TC and OC, confirming that pooling both datasets together will not yield systematic biases in further analysis (Supporting information Fig. S1).

2.2. Instruments and spectra preprocessing

Twenty instruments belonging to different organizations returned FT-MIR spectra for the 70 ring trial samples (Fig. 1b). The instruments span diverse manufacturers, models, internal optics, and sampling accessories (Supporting information Table S1 and Table S2). These metadata were used to group similar instruments into clusters to investigate potential drivers of the instruments' variability (see the 'Statistical analysis and comparisons' section). While spectral range varied between instruments, most instruments collected spectra at 4 cm^{-1} resolution but the number of co-added scans varied from 10 to 64. With variable spectral range, format, and resolution, all spectra were transformed to absorbance ($A = log_{10}[1/R]$), truncated to 4000–650 cm^{-1}, and resampled to 2 cm^{-1} interval using splines (Stevens and Ramirez-Lopez, 2022).

MIR spectra were processed as i) original return (raw), ii) baseline offset correction (BOC), iii) Savitzky-Golay (SG) 1st derivative (SG1stDer), iv) standard normal variate (SNV), v) SNV followed by SG1stDer (SNV + SG1stDer), and vi) spectral space transformation (SST) after SNV. BOC was implemented in this paper as an operation that finds and subtracts the minimum value of each sample spectra across the whole spectral range, bringing the spectral line to the origin (zero) at the position of the minimum value. SG1stDer is a polynomial moving-window function that moves across the spectra with a specified window size and a differentiation order. In this study, it was implemented with a second-order polynomial function, a half-window size of 11 cm^{-1}, and a first-order derivative (Dotto et al., 2018). It can also work as a smoothing algorithm with a zero differentiation order. This smoothing operation was implemented before applying BOC and SNV with a half-window size of 11 cm^{-1}. Derivatives can remove both additive and multiplicative effects in the spectra, enhance absorption features, and reduce the baseline offset. SNV is a row normalization technique that centers (to mean 0) and scales (to 1 standard deviation) the spectra, not over the columns of a matrix. This preprocessing changes both the range of values and the amplitude of the curves and is intended to correct the scattering of light. The combination of SNV and SG1stDer was made to check for further improvements in handling scattering and additive/multiplicative effects. SG1stDer and SNV preprocessing strategies were implemented using the *prospectr* R package v0.2.6 (Stevens and Ramirez-Lopez, 2022), while the other pretreatments were customized within the
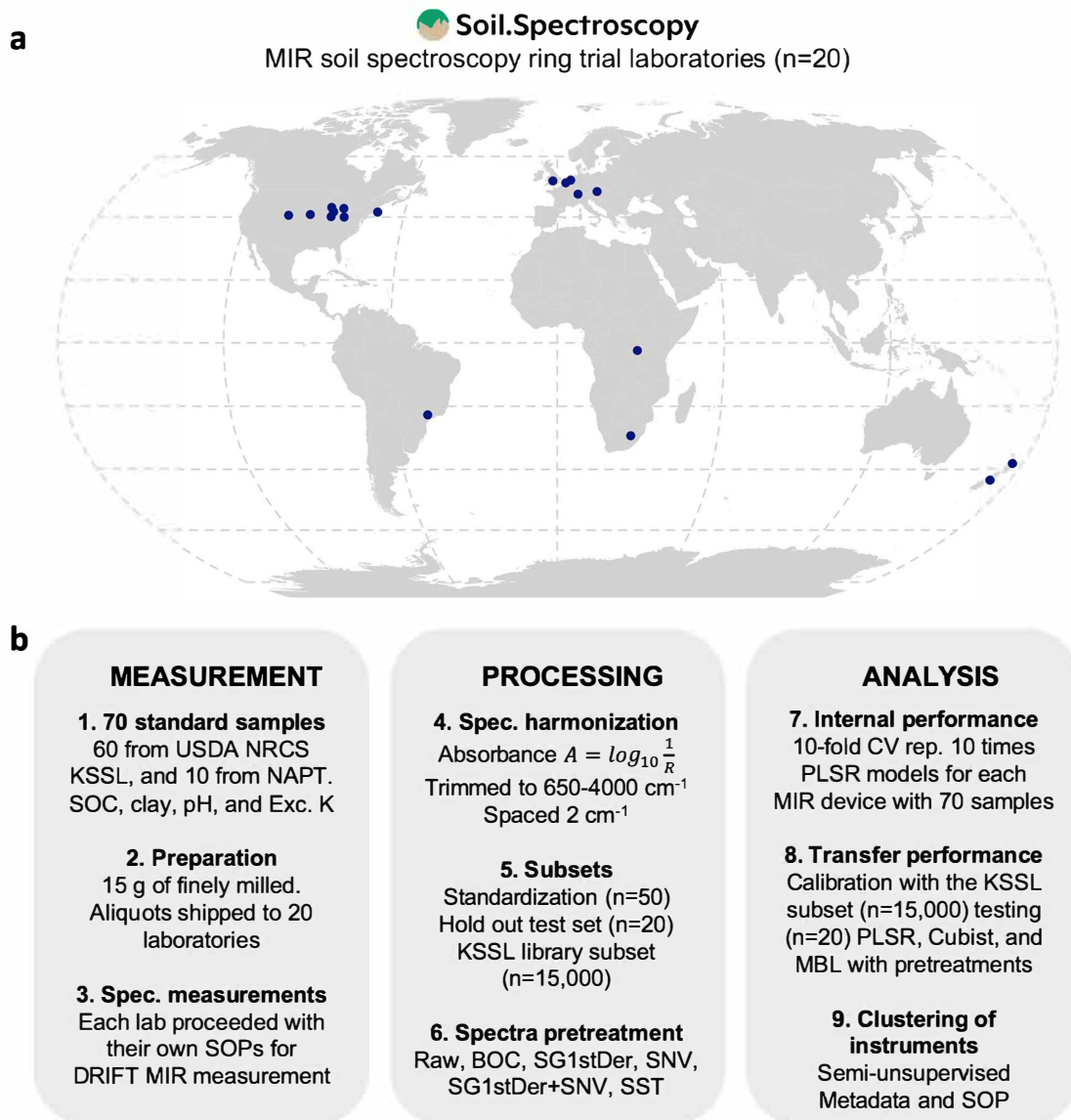
Fig. 1. A) Global laboratories participating in the diffuse reflectancemid-infrared (MIR) ring trial. b) Overview of the main steps employed in this study.

R statistical programming language v4.2.0 (R Core Team, 2022). A visualization comparison is provided in the Supporting information Fig. S3.

Different from regular spectral preprocessing, SST is a spectral standardization method that requires the same set of samples to be shared across different instruments to estimate a transformation matrix. Ring trial samples are intended to work as standard samples for calibration transfer, but in this paper the original full set (n = 70) had to be split into two subsets defined as spectral standardization (RT SST, n = 50) and holdout test (RT test, n = 20) sets (Fig. 2) for proper evaluation of modeling results. Kennard-Stone deterministic sampling algorithm (Kennard and Stone, 1969) was run on SNV KSSL spectra to subset 50 samples, an optimal number for spectral standardization defined as per the previous analysis of Sanderman et al. (2023). Before the subsetting, the KSSL SNV spectra were compressed by principal component analysis (PCA) to retain 99.99 % of the original variability.

Spectral space transformation is a relatively new method and was first described by Du et al. (2011). Its adoption in soil spectroscopy is very limited, although some recent studies have indicated that SST outperforms other spectral standardization methods, i.e., direct standardization (DS) and piecewise direct standardization (PDS), especially

when a small number of standard samples are available to be shared across laboratories or instruments (Pittaki-Chrysodonta et al., 2021; Sanderman et al., 2023). The method is based on the transformation of spectra of a secondary instrument onto the spectral space of the primary one. The transformation matrix is estimated with singular value decomposition using the same algorithm as employed in the PCA. Similar to PCA, one can determine the number of orthogonal features to create a new dataset with scores. In this study, SST was set to retain components explaining 99.99 % of the original cumulative variance and was implemented using the R statistical programming language. The primary instrument set for SST was the Bruker Vertex 70 at the KSSL (instrument #16), because we wanted to be able to use predictive models built from its large public library. All the other ring trial instruments were treated as secondary instruments and transformed to the KSSL space. Detailed information about SST can be found in Pittaki-Chrysodonta et al. (2021).

2.3. Modeling framework

The instruments were first compared regarding their internal prediction capacity using repeated cross-validation. This analysis was
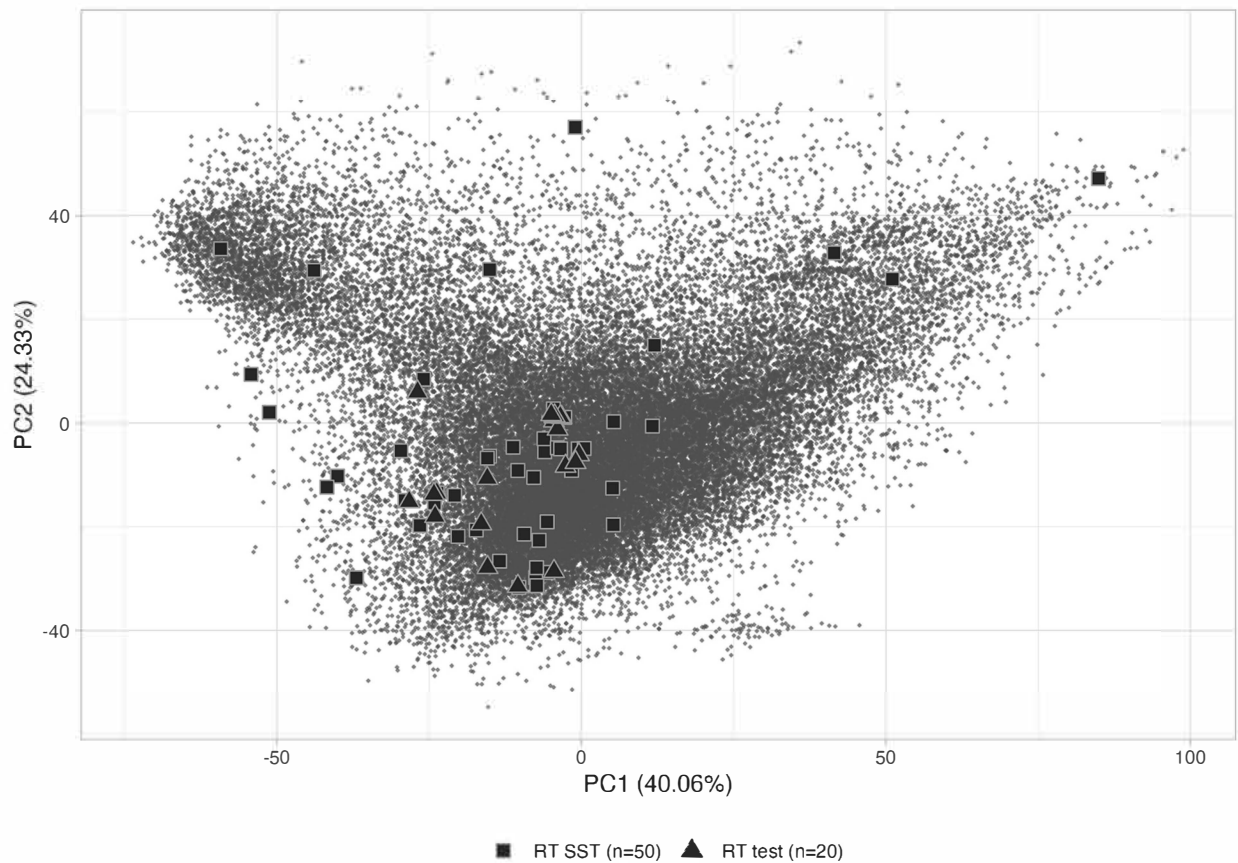
Fig. 2. Principal component analysis of 70 ring trial (RT) samples (spectral space transformation set [RT SST] or holdout test set [RT test]) projected onto the Kellogg Soil Survey Laboratory (KSSL) mid-infrared (MIR) spectral library (subset with n = 15,000, small dots) used for calibration in the transfer model mode. The spectra were preprocessed by Standard Normal Variate.

proposed to evaluate whether an instrument can deliver reliable results considering its specifications and spectral characteristics. Using the whole ring trial set (n = 70) and the standard chemometric algorithm of partial least squares regression (PLSR), each instrument had its internal performance evaluated by 10-fold cross-validation repeated 10 times using the *pls* package v2.8–1 in R (Liland et al., 2022). An outcome of this analysis is that the results may indicate a baseline performance for comparison with the model transfer mode.

The main interest of this study, in turn, was to assess the performance changes from model transfer, i.e., applying a fitted model of a primary soil spectral library onto the different spectra versions generated by each ring trial instrument. This exercise emulates the challenges of soil spectroscopy integration and the use of legacy soil spectra libraries. Model calibration was made using a subset (n = 15,000, Fig. 2) of the large KSSL MIR spectral library that has been employed in previous studies (Sanderman et al., 2021), which is the instrument #16 (Table S1 and S2). The KSSL MIR spectral library has been steadily increasing in size and spectral diversity. It is one of the largest publicly available soil spectral libraries, representing the major soil types of the USA and other territories (Ng et al., 2022; Wijewardane et al., 2018). In combination with additional datasets, it is being used to develop global estimation services of soil properties (Shepherd et al., 2022) and others are working to integrate the KSSL with regional spectral libraries already developed by the participating laboratories of this ring trial. All the subsequent analyses of this study are conducted considering the KSSL MIR as the primary soil spectral library for model calibration.

Before model fitting, soil reference data were checked for asymmetrical distribution using skewness and kurtosis metrics. Soil properties falling outside the range of −3 and 3 for skewness or kurtosis were natural-log transformed to match a normal probability distribution

function (PDF), which was the case of OC and exchangeable K. Prediction performance metrics were calculated in the natural-log space for these soil properties, therefore error metrics were not displayed in the original measurement unit. A statistical description of the ring trial soil samples is provided in Supporting information Table S3 and Table S4, while the PDFs of the three subsets (KSSL calibration set, RT SST set, and RT test set) are visualized in Supporting information Fig. S2.

Combining with the four assessed soil properties and six spectral preprocessing steps, three modeling algorithms (PLSR, memory-based learning (MBL), and Cubist) were evaluated in terms of dealing with the shifts between calibration and application domains in the model transfer mode. PLSR is widely used in the chemometrics field as it is a multidimensional method that handles the collinear nature of spectral measurements (Barra et al., 2021; Soriano-Disla et al., 2014). It is considered a global fitting method that maximizes both the target and covariate variance by linearly combining decomposed orthogonal latent features, which can be optimized by defining the number of factors explaining much of the original variance. In addition, with PLSR it is possible to estimate the variable importance in the projection (VIP) metric to verify the influence of specific wavelengths across the whole modeling spectrum (Chong and Jun, 2005; Rossel and Behrens, 2010). MBL, on the other hand, is a local fitting method that searches for the closest subset of samples (neighbors) of a given calibration set before model fitting, with each sample having its model fitted by an algorithm (Ramirez-Lopez et al., 2013; Saul and Roweis, 2003).

In this paper, MBL was implemented with the aid of the *resemble* R package v2.2.1 (Ramirez-Lopez et al., 2022) by setting the MBL's dissimilarity threshold (k_diss) ranging from 0.5 to 3.0, with 0.5 increments, resulting in 6 options that are truncated by a minimum and a maximum number of neighbors as 50 and 200 (k_range). MBL's local fit

was made by weighted-average PLSR (local_fit_wapls) with a minimum and a maximum number of orthogonal factors defined between 5 and 20 (min_pls_c = 5, max_pls_c = 20), respectively, and being internally optimized by leave-nearest-neighbor-out cross-validation (validation_type = "NNv" within mbl_control). Three different dissimilarity methods (diss_method) were used: 'cor', 'pls', and 'pca'. Considering the large number of hyperparameters that can be set for MBL, especially from the combination of diss_method and k_range, the resulting predictions from 18 models were ultimately averaged. This operation can be viewed as an ensemble MBL as the hyperparameters impact only the subsetting of samples selected as neighbors (resample), as the model optimization is automatically done by NNv.

Lastly, the tree-based algorithm Cubist (Quinlan, 1992, 1993) was also tested as it has been routinely used in soil spectroscopy studies (Barra et al., 2021; Li et al., 2022; Sharififar et al., 2019). This algorithm takes advantage of a decision-tree splitting method but fits linear regression models at each terminal leaf. It also uses a boosting mechanism (sequential trees adjusted by weights) that allows the growth of a forest by tuning the number of committees. In addition, it corrects the final prediction by the nearest neighbors' influence. To control potential model overfitting due to multicollinearity and large dimensionality before fitting the models, principal component loadings fitted with the KSSL calibration library that retained 99.99 % of the original cumulative variance were applied to all ring trial instruments' preprocessed spectra to reduce the original feature size.

All the prediction algorithms were fine-tuned by hyperparameter grid searching of the lowest root mean square error (RMSE) using 10-fold cross-validation. The modeling pipeline was implemented in the R software (version 4.2.0) following *tidymodels* framework (Kuhn and Wickham, 2020) and several packages, i.e., *pls* v2.8–1, *Cubist* v0.4.1, *resemble* v2.2.1, *recipes* v1.0.1, and *yardstick* v1.0.0 (Kuhn and Quinlan, 2022; Kuhn and Vaughan, 2022; Kuhn and Wickham, 2022; Liland et al., 2022; Ramirez-Lopez et al., 2022). Internal performance parameters of the KSSL library evaluated by 10-fold cross-validation with their fine-tuned hyperparameters (PLSR and Cubist) are provided in Supporting information Table S5.

2.4. Statistical analysis and comparisons

Variability across instruments was first assessed with exploratory data analysis. For this, spectral visualizations of selected instances measured by the 20 different instruments allowed their comparison before and after preprocessing, indicating if preprocessing can help reduce the instruments' variability. The spectra were also analyzed in terms of spectral dissimilarity by defining the KSSL ring trial spectra as the reference and calculating the Euclidean distance, which was estimated with the aid of the *resemble* R package v2.2.1 using the 'f_diss' function by centering and scaling the spectral matrices (Ramirez-Lopez et al., 2022). Each sample yielded a dissimilarity value, which was visually compared across instruments.

The performance of both internal cross-validation and model transfer was evaluated by goodness-of-fit metrics including Lin's concordance correlation coefficient (CCC), root mean square error (RMSE), average error (bias), and the ratio of performance to the interquartile range (RPIQ) using the *yardstick* R package v1.0.0 (Kuhn and Vaughan, 2022). However, the visual comparisons and statistical analysis were made using solely Lin's CCC, with the other metrics being provided in the Supporting information. Lin's CCC is a robust metric that not only encompasses overall accuracy but also the bias or deviation of a model from the perfect 1:1 line between observed and predicted values. In addition, the visualizations based on Lins' CCC are standardized with a scale ranging from −1 to 1 (Lin, 1989).

For interpreting the effects of preprocessing and model selection, pairwise comparisons were employed by using permutation tests of Lin's CCC values. Permutation tests are robust because they do not require rigid statistical assumptions (distribution-free) and allow estimating the

effect size of any statistics (Hollander et al., 2013). In this analysis, Lin's CCC values from all the modeling combinations were pooled together and grouped by the categories of a factor of interest. For a given pair of categories being tested (e.g., SNV versus SG1stDer preprocessing), the original effect size (true difference between a statistic of the two samples) is compared with several estimates of effect size after randomly shuffling the pair of samples. The null hypothesis of getting by chance the effect holds if the final proportion (p-value) of the permuted statistic is higher than a significance level. The effect size was measured on the median statistics using 10,000 simulations. With a matrix of all pairwise p-values assessed at a 5 % significance level, the factor levels were categorized by compact letter displays (Graves et al., 2019).

Lastly, a metadata clustering analysis was conducted to investigate possible associations of instruments' characteristics to their final prediction performance. K-means clustering was first performed on raw spectra compressed by principal component analysis by testing up to 20 clusters. The optimal number of clusters was determined using the Elbow method with the Akaike Information Criterion (Dotto et al., 2020; Safanelli et al., 2021). The proportion of samples belonging to a cluster was estimated for each instrument and the cluster with the majority was defined as representing the instrument. After that, two instruments were manually adjusted based on instrument characteristics and the overall performance from the transfer model approach, as they have fallen into separate groups not sharing similar performance characteristics. To help in the interpretation, correspondence analysis was employed to explore the relationship between metadata characteristics and the clusters (both qualitative variables) (Viscarra Rossel et al., 2016). The chi-square test evaluated whether there was a significant association between the categories of the two variables at a 5 % significance level. The relationships were summarized and interpreted with asymmetrical biplots by displaying metadata information (columns) over the cluster space (rows) (Greenacre, 2016).

3. Results

3.1. Spectral dissimilarity

Spectral response varied considerably across the 20 instruments from this ring trial experiment (Fig. 3a-c). Most of the differences are visible in terms of baseline offset, with slight changes over the spectral absorption features. However, a few instruments (#9 Thermo Fisher Nicolet, and #13 Perkin Elmer FT-IR II) had contrasting spectral variations, with their spectral curve being more compressed compared to the other instruments (Fig. 3a). When the samples were preprocessed by SNV, much of the spectral dissimilarities seen with raw were largely reduced (Fig. 3b). Despite being a promising correction algorithm, the remaining visible dissimilarity of some instruments with SNV demanded further processing to properly align the spectral curves (Fig. 3c).

The variations previously identified with raw or SNV were largely reduced by SST (Fig. 3c). Considering that instrument #16 (Bruker Vertex 70) is used as the reference spectra, the Euclidean dissimilarity between each instrument and the KSSL first assessed in the raw spectra (overall median dissimilarity of 1.53) decreased by about 147 % after applying SNV (overall median dissimilarity of 0.62) and 247 % after applying SST (overall median dissimilarity of 0.44) (Fig. 3d-f). The decrease from SNV to SST, however, varied from 150 % (instrument 8) to −30 % (instrument 10). Comparing the improvement between the raw spectra to SST pretreated spectra, the dissimilarity for instrument #1 (Bruker Tensor II) increased, while instruments #1 (Bruker Tensor II), #10 (Bruker Vertex 70), #17 (Bruker Invenio R, and #20 (Bruker Alpha I) showed a negative effect from SNV to SST. Similarly, some soil samples were misaligned after applying SST (Fig. 3f), with differing sample ids found for those having an Euclidean distance higher than 1.5. Other preprocessing visualizations are provided in Supporting information Fig. S4.
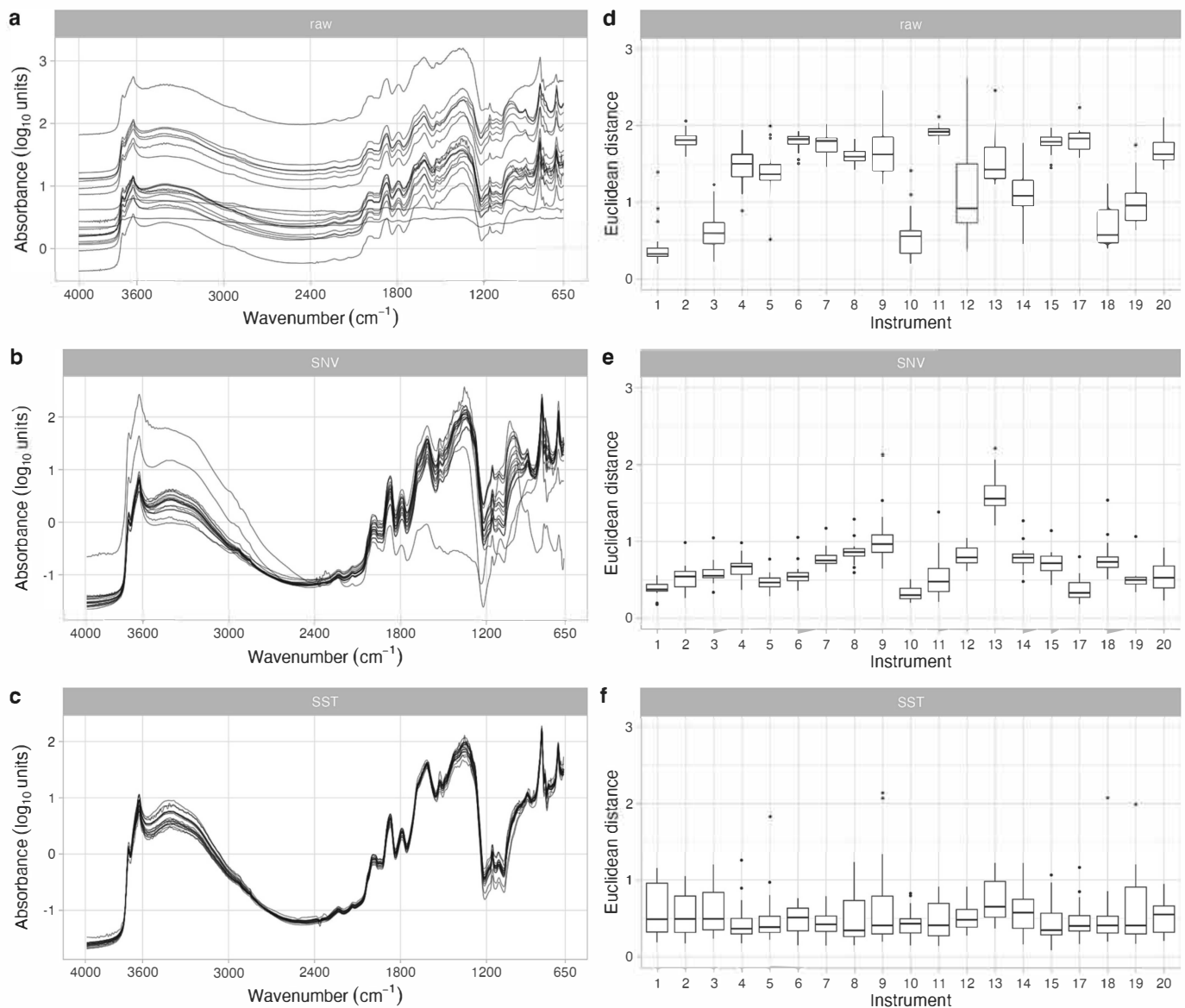
Fig. 3. Spectral variation of sample 19 for the raw (a), standard normal variate (SNV) (b), and spectral space transformation (SST) (c) spectra. Each spectral curve represents one ring trial participant laboratory. Euclidean dissimilarity of ring trial test samples (n = 20) for raw (d), SNV (e), and SST (f) pretreatments considering instrument 16 (Bruker Vertex 70 from the Charles E. Kellogg Soil Survey Laboratory - KSSL) as the reference spectra.

3.2. Internal calibration performance

Soil organic carbon was the soil property that achieved the best internal cross-validation performance with PLSR (Lin's CCC, overall median 0.95 ± IQR 0.02) (Fig. 4). Soil pH ranked second and reached a stable performance regardless of the MIR instrument and preprocessing employed (overall median 0.91 ± IQR 0.05). This stable performance among instruments was also observed for exchangeable K (overall IQR 0.05), although its results were the worst among the tested soil properties (overall median Lin's CCC 0.73). Clay content had the highest variable performance among the MIR instruments, being also more sensitive to spectra preprocessing (overall median 0.81 ± IQR 0.14).

There were only small differences in OC model performance for the different preprocessing methods, except for BOC where the performance for five instruments was substantially lower than for other preprocessing techniques (Fig. 4). This pattern was also observed for soil pH and exchangeable K, albeit with different magnitudes. Clay was the most sensitive soil property to preprocessing, where SNV and SNV followed by SG1stDer (overall median for both techniques being 0.86) were the

best methods for improving model performance. Among all combinations tested, it seems that BOC was the preprocessing step that had the lowest overall performance (median 0.84 ± IQR 0.19).

Despite good overall performance across the 20 instruments, there was substantial variation especially for clay content (Fig. 4). Instrument #13 had the lowest Lin's CCC for OC (median 0.91 ± IQR 0.01) and clay (median 0.67 ± IQR 0.03). Another interesting outcome from this experiment was that instrument #16, the KSSL Bruker Vertex 70 used as the primary instrument in calibration transfer mode, outperformed most of the other instruments for OC (median 0.96 ± IQR 0.02), clay (median 0.92 ± IQR 0.01), and exchangeable K (median 0.79 ± IQR 0.03). The variation among preprocessing techniques for this instrument was particularly low for OC and clay, indicating that other instruments may be more affected by spectral composition and quality assurance during scanning conditions.

3.3. Model transfer performance

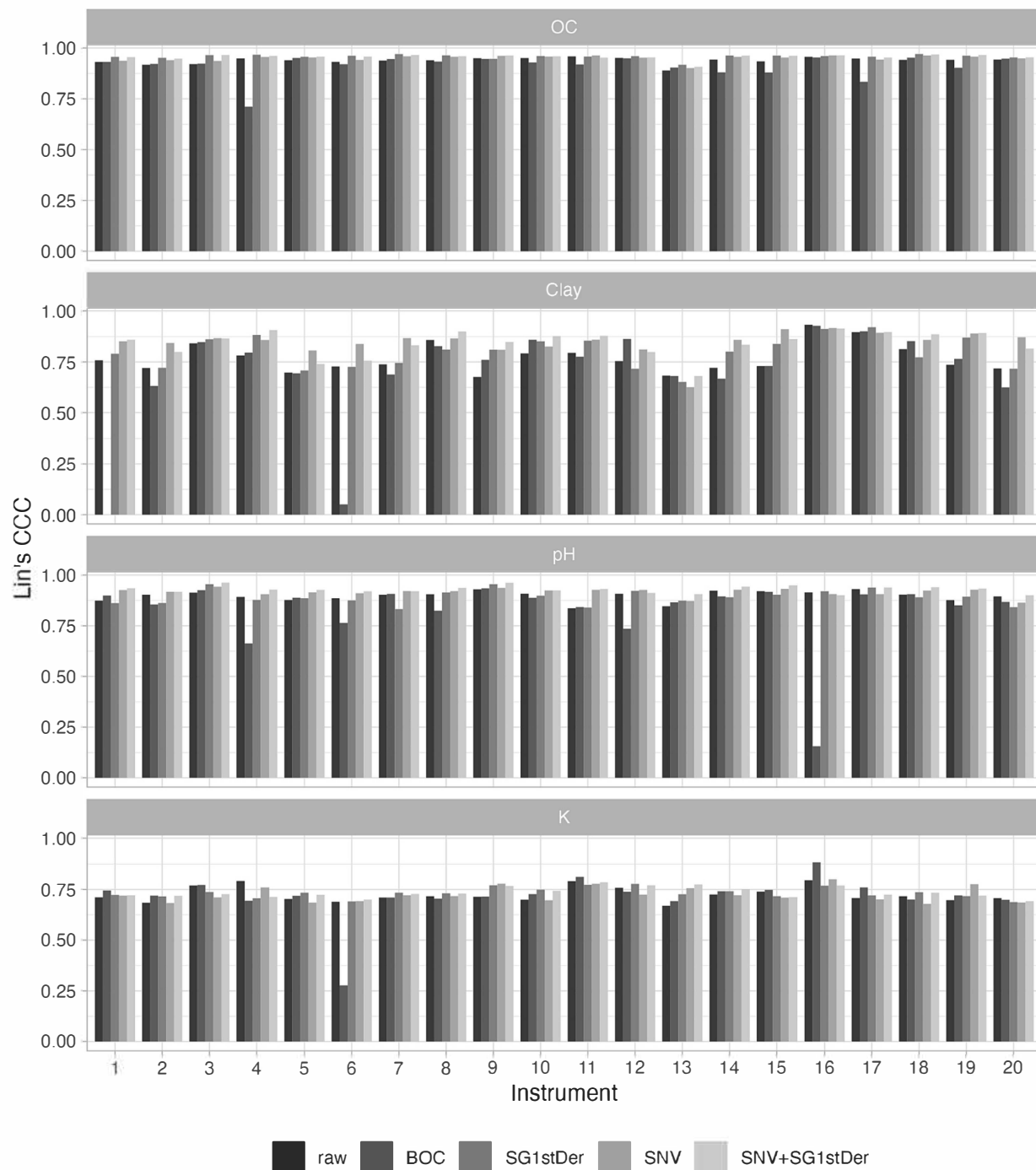One of the main interests of this study was to explore whether lab-to-

Fig. 4. Internal calibration performance (Partial Least Square Regressions with a 10 times repeated 10-fold cross-validation) of all ring trial instruments (n = 20) through Lin's concordance correlation coefficient (CCC) with different spectral preprocessing.

lab variability could be reduced by spectral preprocessing before executing model transfer and prediction. Model transfer from the KSSL library to spectra from the ring trial instruments resulted in OC achieving the best performance (Lin's CCC, overall median 0.90) (Fig. 5a). Soil pH ranked second (overall median 0.81) and clay yielded an overall performance of 0.76. The result for exchangeable K was the worst among the tested soil properties, i.e., an overall median for Lin's CCC of 0.71. In comparison, the overall median performance of the KSSL instrument #16 with the ring trial test samples (n = 20), with pooled model types and preprocessing strategies, reached a Lin's CCC of 0.96, 0.84, 0.92, and 0.83 for OC, clay, pH, and exchangeable K, respectively (Supporting information Table S6).

An important result from this experiment is that among the most common preprocessing without spectral transformation, SNV always ranked amongst the best for all soil properties (Fig. 5a). For SNV, the 10th percentile was the highest among the possible preprocessing when excluding SST, which indicates that the dissimilarity can be reduced to a certain degree when standard samples are not available, except for some very contrasting instruments. The median and 10th percentiles (in parenthesis) of SNV for OC, clay, pH, and K were 0.90 (0.75), 0.77 (0.31), 0.81 (0.52), and 0.77 (0.42), respectively.

For most soil properties (OC, pH, and K), spectral space transformation (SST) largely reduced the spectral dissimilarities between the calibration and application domains (Fig. 4a). The median Lin's CCC
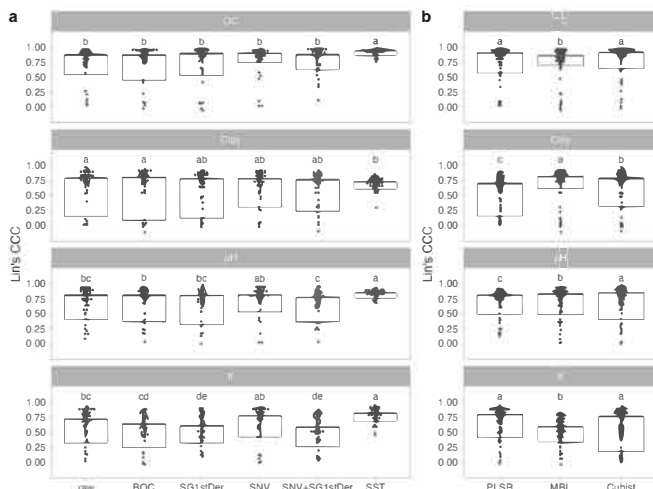
Fig. 5. Lin's concordance correlation coefficient (CCC) for (a) different pre-processing and (b) model types, after performing calibration transfer from the KSSL soil spectral library on ring trial instruments' spectra. Left panels include all model types and instruments. Right panels include all spectral treatments and instruments. Medians not sharing any letter are significantly different by permutation test at the 5% significance level. The top box notch refers to the median, while the box bottom notch represents the 10th percentile.

values for these soil properties were 0.93, 0.85, and 0.81, respectively. Similarly, the 10th percentile was the highest among all preprocessing techniques, indicating that the SST algorithm makes a robust correction before delivering the predictions even for the more extreme cases of spectra dissimilarity.

An analysis of the sensitivity of model types (Fig. 5b) to spectral dissimilarity revealed that despite PLSR achieving a comparable performance to Cubist for OC (median 0.90) and K (median 0.78), the 10th percentile indicates that this model type is more sensitive to extreme variations of calibration and application domain for OC (10th percentile 0.57) and clay (10th percentile 0.15). Cubist models ranked first for OC (median 0.91), pH (median 0.84), and K (median 0.76), while MBL outperformed only for clay (median 0.81). Nevertheless, MBL was the least sensitive model type for OC (10th percentile 0.69) and clay (10th percentile 0.61).

3.4. Metadata analysis

It was possible to identify four main spectral clusters using the raw spectra (Supporting information Fig. S5). Most of the variation compressed by the first principal component before running k-means was due to baseline offset, despite column-normalization of the pooled datasets and the remaining variance retaining further components (total of 7). Two manual adjustments were made to this initial clustering: 1) instrument #17 (Bruker Invenio with HTS-XT accessory) was moved from cluster 2 (C2) to cluster (C4); and 2) instrument #14 (Thermo Fisher Nicolet with a Pike DRIFT accessory) was moved from C4 to C2 (Supporting information Fig. S6). After these two adjustments, cluster 1 (C1) was associated with one Bruker Alpha using the Bruker front-reflectance accessory. Cluster 3 (C3) was associated with the remaining five Bruker Alpha I and II instruments using the standard DRIFT accessory and one Vertex model also using a standard DRIFT accessory. Bruker Tensor, Vertex, and Invenio models all using the HTS-XT diffuse reflectance accessory and MCT detectors (n = 6) were grouped into cluster 4 (C4). In contrast, cluster 2 (C2) contained all the other manufacturers (Perkin Elmer, Thermo Fisher, Agilent) using various standard DRIFT accessories (n = 6). Additional details of the instruments in each cluster can be found in Supporting information Tables S1 and S2. For metadata with at least two levels of category, the correspondence analysis revealed significant associations (after the chi-squared

independence test at a 5 % significance level) with the defined clusters. Although some metadata information was heavily unbalanced, general associations could be drawn from it, especially considering that the primary instrument of this study belonged to C4. The clustering analysis is fully characterized in the Supporting information.

The calibration transfer performance assessed across the defined clusters revealed that the most consistent cluster was C4, which contained the KSSL Bruker Vertex #16 and six other instruments with similar configurations (Fig. 6a). Lin's CCC median and 10th percentile (in parenthesis) of this group were 0.93 (0.84), 0.81 (0.67), 0.88 (0.81), and 0.80 (0.58) for OC, clay, pH, and K, respectively. Cluster 2, on the other hand, had the poorest performance for all soil properties, reaching Lin's CCC of 0.81 (0.07), 0.69 (0.02), 0.71 (0.17), and 0.53 (0.07) for OC, clay, pH, and K, respectively. The remaining clusters had significantly different performances but always fell between the range of C2 and C4.

Considering the previous results where SNV and SST stood out from the other preprocessing and both MBL and Cubist were less sensitive to instruments variability, a detailed analysis of these factors with the defined cluster confirmed that SST was most capable of reducing spectral dissimilarity among calibration and application domains (Fig. 6b). SNV was not sufficiently efficient in dealing with the incompatibility issues present in C2, reaching an overall performance (both model types pooled together) of 0.83 (0.24), 0.78 (0.14), 0.65 (0.19), and 0.45 (0.09) for OC, clay, pH, and K, respectively. In contrast, spectral standardization with SST was able to improve performance in C2 to 0.91 (0.78), 0.77 (0.73), 0.84 (0.71), and 0.72 (0.56) for OC, clay, pH, and K, respectively. In comparison, C4, which includes six other instruments apart from the KSSL Bruker Vertex, achieved an SST performance of 0.91 (0.87), 0.76 (0.69), 0.88 (0.85), and 0.76 (0.73) for OC, clay, pH, and K, respectively. The small differences between median and 10th percentile suggests all instruments in cluster C1, C3 and C4 performed well after SST, with a significant reduction of inaccuracy for C2.

The instruments that had the highest influence on decreasing performance due to contrasting spectral responses (Lin's CCC lower than group 10th percentile), both in the preprocessing and model types comparison, were in most of the cases instruments #9 (Thermo Fisher Nicolet), #12 (Thermo Fisher Nicolet), and #13 (Perkin Elmer FT-IR II) (Supporting information Fig. S18), all belonging to cluster C2. A great difference, in turn, is that SST significantly improved the performance for all soil properties, e.g., Lin's CCC of OC from 0.02 to 0.80 for instrument #13 (Perkin Elmer FT-IR II), when compared to using SNV only (Supporting information Table S15). However, for other instruments belonging to C2 that were not too contrasting to instrument #16 (instruments #4 [Perkin Elmer Spectrum 100], #5 [Agilent 4300], and #14 [Thermo Fisher Nicolet] with low Euclidean dissimilarity), the spectral transformation enhancement was not significant, with some soil properties being impaired by SST (Supporting information Fig. S18).

4. Discussion

4.1. Internal calibration

Internal calibration performance was generally good across all instruments (Fig. 4). This indicates that despite having contrasting spectral responses for the same soil sample due to the particular characteristics of each instrument (Fig. 3a), models calibrated using their internal soil spectral library will deliver good estimates when new samples are measured by the same scanner and standard operating procedures (SOP). OC was the most consistent soil property that had minimal changes in performance regardless of the preprocessing and instrument employed (Fig. 4). In contrast, Clay was very sensitive to preprocessing (especially to BOC), and exchangeable K did not reach the same overall prediction capacity as OC, Clay, and pH. These overall performance findings are notably similar to a previous study that evaluated the MIR spectral range for the accurate measurement of soil
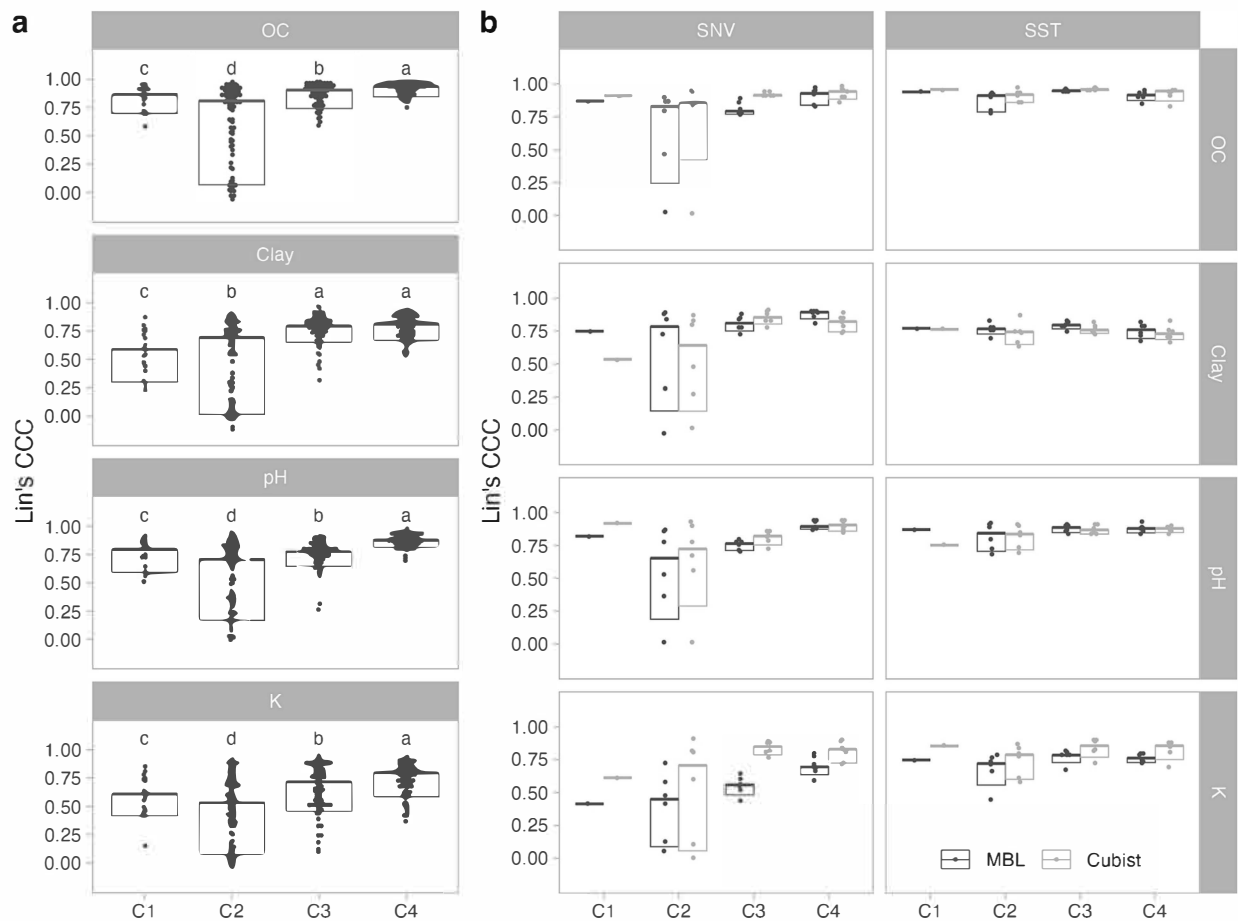
Fig. 6. Lin's concordance correlation coefficient (CCC) for (a) different spectra clusters after performing calibration transfer from the KSSL soil spectral library on ring trial instruments' spectra. Left panels include all model types, preprocessing, and instruments. Medians not sharing any letter are significantly different by permutation test at the 5% significance level. The top box notch refers to the median, while the box bottom notch represents the 10th percentile. MBL and Cubist model types, treated with SNV or SST preprocessing, are highlighted for different soil properties (b).

properties (Ng et al., 2022). Ng et al. (2022) grouped OC, Clay, and pH in a set of highly predictable soil properties (accuracy group A), while exchangeable K fell into the accuracy group C of a total of four levels (A, B, C, D).

Of all the preprocessing techniques, BOC produced the most variable results likely because the correction was implemented in this study with a very naive approach. It consisted of finding the lowest value for each spectrum and subtracting it across the whole spectral range, i.e., BOC was only capable of handling baseline offset. Light scattering effects mostly caused by particle size slightly distort the spectral responses and have caused a shift in the position of the lowest value across the spectrum (Johnston and Aochi, 2018). This was the case with sample 44 for instrument #6 (Bruker Alpha I, the most impacted by BOC), where the lowest value of sample 44 (sand content = 99.61 %) deviated from the other samples' position (Supporting information Fig. S15). It seems that some instrument configurations may be more sensitive and propagate the shift of BOC into the calibrated models, especially if a model of a soil property of interest relies on the variation around the lowest value position. Soil minerals, for example, have a vibrational group around 1200–970 cm^{-1} that is related to Si-O-Si stretching (Johnston and Aochi, 2018), the same region that was impacted by the lowest-value determination of BOC preprocessing of sample 44 in instrument #6 (Bruker Alpha I). This impact differed from the other preprocessing methods that found a way to manage sample 44 with extreme sand content and its impacts of light scattering before delivering a prediction. Therefore, users are recommended to be more cautious with the default operations of baseline offset correction in currently available spectroscopy software

and opt for methods that further account for scattering effects, such as SNV and multiplicative scatter correction (MSC).

The variability of clay prediction in the internal performance evaluation was further verified by the importance of spectral features used by the models. Considering the VIP metric, which calculates the importance of each original wavenumber to the orthogonal projection of PLSR, it seems that clay was more sensitive to offset and scattering effects than other soil properties (Supporting information Fig. S16). While OC had similar important features shared by either BOC or SNV, SNV's important features for clay were largely more distinct than BOC. SNV spectral features had a larger amplitude of VIP across the whole spectral range, while BOC had an overall parabolic shape with features being more influential around the limits of the spectral range, a possible effect of remaining light scattering and multicollinearity over important vibrational groups that was not properly accounted by BOC (Barnes et al., 1989).

Some instruments consistently outperformed others in the internal performance evaluation, especially instruments #16 (Bruker Vertex 70) and #18 (Burker Invenio S) for OC; instruments #16 (Bruker Vertex 70) and #17 (Burker Invenio R) for Clay; instruments #3 (Bruker Vertex 70) and #9 (Thermo Fisher Nicolet) for pH; and instruments #11 (Bruker Alpha II) and #16 (Bruker Vertex 70) for K (Fig. 4). This grouping is confirmed by the compact letter display from the permutation test, with all these instruments being at least classified with letter a (best) (Supporting information Fig. S17). Instrument #16 is the Bruker Vertex 70 from KSSL that follows rigid quality control and quality assurance for measuring MIR soil spectra (Method 7A7; Soil Survey Staff, 2022). The

other instruments that outperformed in the internal calibration performance fell into different clusters: C1 for instrument #11 (Bruker Alpha II), C2 for instrument #9 (Thermo Fisher Nicolet), C3 for instrument #17 (Burker Invenio R), and C4 with instruments #3 (Bruker Vertex 70), #16 (Bruker Vertex 70), and #18 (Bruker Inveio S), indicating that internal calibration performance seems to be independent of the results from the model transfer mode.

4.2. Model transfer

There was large variability in predictive performance across the 20 instruments when we applied the large KSSL models (calibration sample size = 15,000), with a number of instruments performing quite well with standard preprocessing techniques. However, a few instruments failed to produce reasonable results mostly due to the contrasting spectral patterns exemplified in Fig. 3. SNV consistently ranked the best or second best preprocessing and especially presented the highest 10th percentile among all of them. In fact, SNV was originally proposed to deal with multiplicative effects of particle size, scatter, and multicollinearity issues in diffuse reflectance spectroscopy (Barnes et al., 1989). Although the first derivative (i.e., SG1stDer) has been routinely used in soil spectroscopy studies (Barra et al., 2021), SNV is preferred especially because the centering and scaling operations do not affect the interpretation of the spectral features. SNV is a normalization method that brings the mean absorbance to the origin and rescales the amplitude of variation to one standard deviation, transforming it into a compressed or stretched curve. The combination of SNV with SG1stDer was clearly not a good choice for pH and K (Fig. 5a), although some instruments benefited from this special combination during their internal performance evaluation (Fig. 4). Despite SG1stDer having an overall lower dissimilarity (median = 0.58, IQR = 0.24) than SNV (median = 0.62, IQR = 0.35), results from Fig. 5 demonstrate that SNV pretreatment was on average always as good as or better than either of the SG1stDer pretreatment options. Spectral derivatives are complex, often difficult to interpret, and are dependent on the segment and gap size definitions (Barnes et al., 1989), therefore SNV only is instead recommended.

Averaged across all soil properties, Cubist with SNV is the best choice for C4 and equal to Cubist with SST for C3. However, SST gives a significant performance boost to C1 and C2, which were highly affected by instrument or sampling accessory incompatibility to KSSL Bruker Vertex. This pattern holds for SOC, pH and exchangeable K although with Exc. K, PLSR models with SNV outperformed Cubist with SNV. The pattern was less clear for clay content with SST still being critical for C2 but other pretreatments gave much better performance in other clusters. In turn, MBL yielded a higher 10th percentile compared to Cubist in all cases. Previous research with a Bruker Alpha similar to the one in used in C1 in this study (Sanderman et al., 2023) and a Bruker Vertex (instrument #8 of C3 in this study) (Dangal and Sanderman, 2020) have indicated that spectral standardization might not be necessary for some cases. Using the KSSL MIR spectral library and independent test sets, Dangal et al. (2019) found that Cubist and MBL outperformed PLSR and random forest models across 10 soil properties analyzed. Cubist is a hybrid rule-based and linear regression algorithm that is prone to overfitting when dealing with a high number of features, especially considering that spectral measurements are also multicollinear. To control this, dimensionality reduction via orthogonal projection or feature selection (Minasny and McBratney, 2008; Padarian et al., 2019) are highly recommended but may increase the complexity of operations. In this study, PCA was set to retain 99.99 % of the original cumulative variance of the training set to account for small changes in spectral absorption features of farther components. Therefore, it is recommended to find the tradeoff between compression (viable number of components) and spectral representation level (farther components may retain unique features, but increases the model complexity) using these model configurations.

4.3. Recommendations

Overall, the metadata analysis suggested that the major differences in spectral responses occurred due to manufacturer and model variations that ultimately represent different hardware and internal optics. In addition, due to the high occurrence of similar Bruker models in the ring trial, it seems that even minor changes in hardware, sampling accessories, and optical parts are enough to cause slight variations of spectral responses for the same set of soil properties that ultimately can lead to a significant spread in model performance (see C3 in Fig. 6a). However, for these Bruker models (C3 and C4), SNV alone was sufficient to correct baseline offset and light-scattering effects (Fig. 6b). SNV correction alone was sufficient for some other manufacturers from cluster C2 that did not have too contrasting spectral responses to instrument #16 (Bruker Vertex), i.e., instruments #4 (Perkin-Elmer Spectrum-100), #5 (Agilent 4300 Handheld FTIR), and #14 (Thermo Fisher Nicolet) (Supporting information Table S15). It is worth noting again that despite sometimes poor KSSL model transfer results, all the instruments were generally good when they were evaluated by internal calibration performance. When legacy soil spectral libraries or diverse instruments are combined together to form a larger and more robust database, or already-fitted models are leveraged to make new predictions outside the original purpose or feature space, then spectral standardization may be necessary.

While there is no one-size-fits-all approach to working with the KSSL library, some clear recommendations have come out of this study:

1. For Bruker instruments with the HTS-XT accessory and MCT detector, the KSSL library can be used directly with almost any mathematical pretreatment (either preprocessing or spectral standardization), although SNV preprocessing is still highly recommended due to the known issues found in FT-MIR (scattering and multicollinearity).
2. For Bruker Alpha instruments with the standard DRIFT accessory, SNV preprocessing performs equally as well as SST when Cubist is used for model building.
3. For instruments from other manufacturers using a standard DRIFT accessory, spectral standardization is necessary for producing reliable estimates from the KSSL library.
4. Given that there isn't always an obvious best choice for modeling, we recommend testing several models and several preprocessing methods. When validation samples aren't available for picking the best approach we recommend using an ensemble of these combinations.
5. Lastly, the quality of predictions also depends on the spectral library used for calibration. Therefore, when using the KSSL library, users need to ensure that the KSSL library is representative, based both on spectral similarity and range of soil properties of interest, of the target samples.

4.4. Further action

The results of this study highlight the importance of standardizing the spectral acquisition procedures across different laboratories and instruments to ensure the reliability and transferability of soil spectroscopy predictions. This is in line with the historical experience of wet-chemistry laboratories, which have developed quality control and certification mechanisms to reduce the variability and uncertainty in soil analysis. However, unlike wet-chemistry laboratories, spectral sensing offers a unique opportunity to leverage the power of open data and collaborative platforms to share and access soil information for the global good. One such initiative is the Open Soil Spectral Library (OSSL), which aims to provide a comprehensive and harmonized repository of soil spectra and metadata from various sources and instruments (Safanelli et al., 2023). The OSSL also offers support and training for users who want to generate predictions from their own spectra or use the

existing models from the library. By contributing to and using the OSSL, soil spectroscopists can benefit from the collective knowledge and expertise of the soil spectroscopy community, as well as enhance the richness and diversity of soil data for research purposes.

Our study also demonstrated that shared standard samples can form the basis of building calibration transfer among soil spectroscopy laboratories, especially in relation to the centralized USDA NSSC KSSL FT-MIR soil spectral library that can be leveraged to calibrate more robust prediction models, which is included in the OSSL. In addition, our study demonstrated a willingness to participate in the ring-trial research to extend the capability of the OSSL to generate soil information for the global good. Therefore, we encourage further action to promote the OSSL and its goals, as well as to explore the potential of hybrid soil laboratories that combine wet-chemistry and spectral methods to meet the growing demand for soil information (Demattê et al., 2019; Lal et al., 2021; Nocita et al., 2015; Shepherd et al., 2022).

5. Conclusion

For the first time, a large interlaboratory soil spectroscopy comparison was made using twenty FT-MIR laboratory instruments. By uniformly preparing and shipping several aliquots of the same soil samples, which reduced potential effects of variable soil analytical determinations, we quantified the amount of spectral variance among laboratories and determined the impact of model predictions by assessing the instruments' internal performance with those obtained by transferring models from a centralized spectral library, i.e., the USDA NCSS KSSL FT-MIR library. In this highly participatory study, we found that all instruments can deliver good predictions when calibrated internally, i.e., using the same instruments' characteristics and standard operating procedures for predicting new samples, by solely relying on regular spectral preprocessing that accounts for light scattering and multiplicative/additive effects (e.g., using standard normal variate - SNV). When performing model transfer from the large KSSL FT-MIR library, instruments' variability can be reduced by regular preprocessing (SNV) if the secondary instruments have the same instrument manufacturer as the KSSL FT-MIR library. Significant differences between the KSSL FT-MIR library and contrasting ring trial instruments responses were evident and confirmed by a semi-unsupervised spectral clustering. For heavily contrasting setups, spectral standardization is necessary before transferring prediction models. Non-linear model types like Cubist and MBL (local fitting) deliver more precise estimates because they seem to be less sensitive to spectral variations than global PLSR, although they are still affected by highly contrasting conditions. In summary, the results from this ring trial experiment may assist new laboratories in building spectroscopy capacity utilizing existing FT-MIR spectral libraries and support the recent global efforts to make soil spectroscopy universally accessible with centralized or shared operating procedures.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2023.116724.

References

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. Appl. Spectrosc. 43 (5), 772–777. https://doi.org/10.1366/0003702894202201.

Barra, I., Haefele, S.M., Sakrabani, R., Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. TrAC Trends Anal. Chem. 135, 116166 https://doi.org/10.1016/j.trac.2020.116166.

Ben-Dor, E., Ong, C., Lau, I.C., 2015. Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma 245–246, 112–124. https://doi.org/10.1016/j.geoderma.2015.01.002.

Chong, I.-G., Jun, C.-H., 2005. Performance of some variable selection methods when multicollinearity is present. Chemom. Intel. Lab. Syst. 78 (1–2), 103–112. https://doi.org/10.1016/j.chemolab.2004.12.011.

Cools, N., Delanote, V., Scheldeman, X., Quataert, P., De Vos, B., Roskams, P., 2004. Quality assurance and quality control in forest soil analyses: a comparison between European soil laboratories. Accred. Qual. Assur. 9 (11–12), 688–694. https://doi.org/10.1007/s00769-004-0856-4.

Dangal, S., Sanderman, J., 2020. Is Standardization Necessary for Sharing of a Large Mid-Infrared Soil Spectral Library? Sensors 20 (23), 6729. https://doi.org/10.3390/s20236729.

Dangal, S., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library. Soil Syst. 3 (1), 11. https://doi.org/10.3390/soilsystems3010011.

Demattê, J.A.M., Dotto, A.C., Bedin, L.G., Sayão, V.M., Souza, A.B.e., 2019. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. Geoderma 337, 111–121. https://doi.org/10.1016/j.geoderma.2018.09.010.

Dotto, A.C., Dalmolin, R.S.D., ten Caten, A., Grunwald, S., 2018. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for

multivariate prediction of soil organic carbon by Vis-NIR spectra. Geoderma 314, 262–274. https://doi.org/10.1016/j.geoderma.2017.11.006.

Dotto, A.C., Demattê, J.A.M., Viscarra Rossel, R.A., 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. https://doi.org/10.5194/soil-6-163-2020.

Du, W., Chen, Z.-P., Zhong, L.-J., Wang, S.-X., Yu, R.-Q., Nordon, A., Littlejohn, D., Holden, M., 2011. Maintaining the predictive abilities of multivariate calibration models by spectral space transformation. Anal. Chim. Acta 690 (1), 64–70. https://doi.org/10.1016/j.aca.2011.02.014.

Francos, N., Heller-Pearlshtien, D., Demattê, J.A.M., Van Wesemael, B., Milewski, R., Chabrillat, S., Tziolas, N., Sanz Diaz, A., Yagüe Ballester, M.J., Gholizadeh, A., Ben-Dor, E., 2023. A Spectral Transfer Function to Harmonize Existing Soil Spectral Libraries Generated by Different Protocols. Appl. Environ. Soil Sci. 2023, 1–17. https://doi.org/10.1155/2023/4155390.

Frei, R.W., MacNeil, J.D., 2019. Diffuse Reflectance Spectroscopy Environmental Problem Solving, 1st ed. CRC Press.

Gholizadeh, A., Neumann, C., Chabrillat, S., van Wesemael, B., Castaldi, F., Borůvka, L., Sanderman, J., Klement, A., Hohmann, C., 2021. Soil organic carbon estimation using VNIR–SWIR spectroscopy: The effect of multiple sensors and scanning conditions. Soil Tillage Res. 211, 105017 https://doi.org/10.1016/j.still.2021.105017.

FAO GLOSOLAN. (2023). Global Soil Partnership on Soil Spectroscopy. https://www.fao.org/global-soil-partnership/glosolan-old/soil-analysis/dry-chemistry-spectroscopy/en/.

Graves, S., Piepho, H., Selzer, L. (2019). multcompView: Visualizations of Paired Comparisons. In R package version 0.1-8. https://CRAN.R-project.org/package=multcompView.

Greenacre, M., 2016. Correspondence Analysis in Practice. Chapman and Hall/CRC. https://doi.org/10.1201/9781315369983.

Hollander, M., Wolfe, D.A., Chicken, E., 2013. Nonparametric Statistical Methods, 3rd ed. Wiley.

Johnston, C. T., & Aochi, Y. O. (2018). Fourier Transform Infrared and Raman Spectroscopy (pp. 269–321). https://doi.org/10.2136/sssabookser5.3.c10.

Kennard, R.W., Stone, L.A., 1969. Computer Aided Design of Experiments. Technometrics 11 (1), 137–148. https://doi.org/10.1080/00401706.1969.10490666.

Kuhn, M., Quinlan, J.R., 2022. Cubist: Rule- And Instance-Based Regression Modeling. In R Package Version (4), 1. https://CRAN.R-project.org/package=Cubist.

Kuhn, M., Vaughan, D., 2022. yardstick: Tidy Characterizations of Model Performance. In R Package Version 1. https://CRAN.R-project.org/package=yardstick.

Kuhn, M., Wickham, H., 2020. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. In. https://www.tidymodels.org.

Kuhn, M., Wickham, H., 2022. recipes: Preprocessing and Feature Engineering Steps for Modeling. In R Package Version 1, 1. https://CRAN.R-project.org/package=recipes.

Lal, R., Bouma, J., Brevik, E., Dawson, L., Field, D.J., Glaser, B., Hatano, R., Hartemink, A.E., Kosaki, T., Lascelles, B., Monger, C., Muggler, C., Ndzana, G.M., Norra, S., Pan, X., Paradelo, R., Reyes-Sánchez, L.B., Sandén, T., Singh, B.R., Zhang, J., 2021. Soils and sustainable development goals of the United Nations: An International Union of Soil Sciences perspective. Geoderma Reg. 25, e00398.

Li, S., Viscarra Rossel, R.A., Webster, R., 2022. The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon. Eur. J. Soil Sci. 73 (1) https://doi.org/10.1111/ejss.13202.

Liland, K., Mevik, B., & Wehrens, R. (2022). pls: Partial Least Squares and Principal Component Regression. In R package version 2.8-1. https://CRAN.R-project.org/package=pls.

Lin, L.-I.-K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45 (1), 255. https://doi.org/10.2307/2532051.

Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom. Intel. Lab. Syst. 94 (1), 72–79. https://doi.org/10.1016/j.chemolab.2008.06.003.

Minasny, B., Tranter, G., McBratney, A.B., Brough, D.M., Murphy, B.W., 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. Geoderma 153 (1–2), 155–162. https://doi.org/10.1016/j.geoderma.2009.07.021.

Ng, W., Minasny, B., Jeon, S.H., McBratney, A., 2022. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Security 6, 100043. https://doi.org/10.1016/j.soisec.2022.100043.

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., … Wetterlind, J. (2015). Soil spectroscopy: An alternative to wet chemistry for soil monitoring (pp. 139–159). https://doi.org/10.1016/bs.agron.2015.02.002.

Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. Geoderma Reg. 16, e00198.

Pimstein, A., Notesco, G., Ben-Dor, E., 2011. Performance of Three Identical Spectrometers in Retrieving Soil Reflectance under Laboratory Conditions. Soil Sci. Soc. Am. J. 75 (2), 746–759. https://doi.org/10.2136/sssaj2010.0174.

Pittaki-Chrysodonta, Z., Hartemink, A.E., Sanderman, J., Ge, Y., Huang, J., 2021. Evaluating three calibration transfer methods for predictions of soil properties using mid-infrared spectroscopy. Soil Sci. Soc. Am. J. 85 (3), 501–519. https://doi.org/10.1002/saj2.20225.

Poppiel, R.R., Paiva, A.F. da S., Demattê, J.A.M., 2022. Bridging the gap between soil spectroscopy and traditional laboratory: Insights for routine implementation. Geoderma 425, 116029. https://doi.org/10.1016/j.geoderma.2022.116029.

Quinlan, J.R., 1992. In: Learning with Continuous Classes. World Scientific, pp. 343–348.

Quinlan, J.R., 1993. Combining Instance-Based and Model-Based Learning. In: ICML'93: Proceedings of the Tenth International Conference on International Conference on Machine Learning, pp. 236–243.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M., Scholten, T., 2013. The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma 195–196, 268–279. https://doi.org/10.1016/j.geoderma.2012.12.014.

Ramirez-Lopez, L., Wadoux, A.-M.-J.-C., Franceschini, M.H.D., Terra, F.S., Marques, K.P. P., Sayão, V.M., Demattê, J.A.M., 2019. Robust soil mapping at the farm scale with vis–NIR spectroscopy. Eur. J. Soil Sci. 70 (2), 378–393. https://doi.org/10.1111/ejss.12752.

Ramirez-Lopez, L., Stevens, A., Viscarra Rossel, R., Lobsey, C., Wadoux, A., Breure, T., 2022. Resemble: regression and similarity evaluation for memory-based learning in spectral chemometrics. R Package Vignette R Package Version 2 (2), 1. https://cran.r-project.org/package=resemble.

Rayment, G.E., Miller, R.O., Sulaeman, E., 2000. Proficiency testing and other interactive measures to enhance analytical quality in soil and plant laboratories. Commun. Soil Sci. Plant Anal. 31 (11–14), 1513–1530. https://doi.org/10.1080/00103620009370523.

Rossel, R.A.V., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158 (1–2), 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025.

IEEE SA. (2023). P4005 - Standards and protocols for soil spectroscopy. https://sagroups.ieee.org/4005/.

Safanelli, J.L., Demattê, J.A.M., dos Santos, N.V., Rosas, J.T.F., Silvero, N.E.Q., Bonfatti, B.R., de Mendes, W. de S., 2021. Fine-scale soil mapping with Earth Observation data: a multiple geographic level comparison. Rev. Bras. Ciênc. Solo 45. https://doi.org/10.36783/18069657rbcs20210080.

Safanelli, J.L., Hengl, T., Sanderman, J., Parente, L., 2023. Open Soil Spectral Library. https://doi.org/10.5281/zenodo.7599269.

Sanderman, J., Savage, K., Dangal, S.R.S., 2020. Mid-infrared spectroscopy for prediction of soil health indicators in the United States. Soil Sci. Soc. Am. J. 84 (1), 251–261. https://doi.org/10.1002/saj2.20009.

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., Stewart, C., 2021. Can Agricultural Management Induced Changes in Soil Organic Carbon Be Detected Using Mid-Infrared Spectroscopy? Remote Sens. (Basel) 13 (12), 2265. https://doi.org/10.3390/rs13122265.

Sanderman, J., Gholizadeh, A., Pittaki-Chrysodonta, Z., Huang, J., Safanelli, J.L., Ferguson, R., 2023. Transferability of a large mid-infrared soil spectral library between two Fourier-transform infrared spectrometers. Soil Sci. Soc. Am. J. https://doi.org/10.1002/saj2.20513.

Saul, L.K., Roweis, S.T., 2003. Think globally, fit locally: unsupervised learning of low dimensional manifolds. J. Mach. Learn. Res. 4 (Jun), 119–155.

Seybold, C.A., Ferguson, R., Wysocki, D., Bailey, S., Anderson, J., Nester, B., Schoeneberger, P., Wills, S., Libohova, Z., Hoover, D., Thomas, P., 2019. Application of Mid-Infrared Spectroscopy in Soil Survey. Soil Sci. Soc. Am. J. 83 (6), 1746–1759. https://doi.org/10.2136/sssaj2019.06.0205.

Sharififar, A., Singh, K., Jones, E., Ginting, F.I., Minasny, B., 2019. Evaluating a low-cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models. Soil Use Manag. 35 (4), 607–616. https://doi.org/10.1111/sum.12537.

Shepherd, K.D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J., Ge, Y., 2022. A global soil spectral calibration library and estimation service. Soil Security 7, 100061. https://doi.org/10.1016/j.soisec.2022.100061.

Soil Survey Staff. (2022). Kellogg Soil Survey Laboratory methods manual. Soil Survey Investigations Report No. 42, Version 6.0. U.S. Department of Agriculture, Natural Resources Conservation Service. https://www.nrcs.usda.gov/resources/guides-and-instructions/kssl-guidance.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M. J., 2014. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. Appl. Spectrosc. Rev. 49 (2), 139–186. https://doi.org/10.1080/05704928.2013.811081.

Stevens, A., Ramirez-Lopez, L., 2022. An introduction to the prospectr package. In R Package Vignette. R Package Version 0.2.6.

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., Six, J., 2021. The central African soil spectral library: a new soil infrared repository and a geographical prediction analysis. SOIL 7 (2), 693–715. https://doi.org/10.5194/soil-7-693-2021.

Suvannang, N., & Hartmann, C. (2019). First Inter-laboratory Comparison Report of the Regional Soil Laboratory Network for Asia (SEALNET). https://www.fao.org/global-soil-partnership/glosolan-old/soil-analysis/quality-assurance-and-quality-control/en/.

Van Reeuwijk, L., Houba, V., 1998. Guidelines for quality management in soil and plant laboratories (FAO Soils Bulletin 74). Int. Soil Ref. and Info, Centre, FAOUN, Rome.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth Sci. Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Ge, Y., Gomez, C., Guerrero, C., Peng, Y., Ramirez-Lopez, L., Shi, Z., Stenberg, B., Webster, R., Winowiecki, L., Shen, Z., 2022. Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century. Eur. J. Soil Sci. 73 (4) https://doi.org/10.1111/ejss.13271.

Wijewardane, N.K., Ge, Y., Wills, S., Libohova, Z., 2018. Predicting Physical and Chemical Properties of US Soils with a Mid-Infrared Reflectance Spectral Library. Soil Sci. Soc. Am. J. 82 (3), 722–731. https://doi.org/10.2136/sssaj2017.10.0361.

Wolf, A.M., Miller, R.O., 1998. Development of a North American proficiency testing program for soil and plant analysis. Commun. Soil Sci. Plant Anal. 29 (11–14), 1685–1690. https://doi.org/10.1080/00103629809370059.