5-6-2022

Calibration Set Optimization and Library Transfer for Soil Carbon Estimation Using Soil Spectroscopy—A Review

Minerva J. Dorantes
University of Arkansas, Fayetteville, mjdorant@uark.edu

Bryan A. Fuentes
University of Arkansas, Fayetteville

David M. Miller
University of Arkansas, Fayetteville

Citation

Dorantes, M. J., Fuentes, B. A., & Miller, D. M. (2022). Calibration Set Optimization and Library Transfer for Soil Carbon Estimation Using Soil Spectroscopy—A Review. *Soil Science Society of America Journal, 86* (4), 879-903. https://doi.org/10.1002/saj2.20435

Received: 24 September 2021 | Accepted: 14 April 2022 | Published online: 27 June 2022

DOI: 10.1002/saj2.20435

Soil Science Society of America Journal

SHORT CRITICAL REVIEWS OR ESSAYS

Calibration set optimization and library transfer for soil carbon estimation using soil spectroscopy—A review

Minerva J. Dorantes | **Bryan A. Fuentes** | **David M. Miller**

Dep. of Crop, Soil, and Environmental Sciences, Univ. of Arkansas, 115 Plant Sciences Building, Fayetteville, AR 72701, USA

Correspondence
Minerva J. Dorantes, Dep. of Crop, Soil, and Environmental Sciences, Univ. of Arkansas, 115 Plant Sciences Building, Fayetteville, AR 72701, USA.
Email: mjdorant@uark.edu

Assigned to Associate Editor Katherine Heckman.

Abstract

Resource-efficient techniques for accurate soil property estimation are necessary to satisfy the increasing demand for soil data to support environmental monitoring, precision agriculture, and spatial modeling. Over the last 30 yr, infrared soil spectroscopy has developed into a rapid, robust, and cost-effective technique for soil carbon analysis. Ongoing global efforts to make soil spectroscopy operational require the development of soil spectral libraries, which are the main source of data for the construction of calibration models. Understanding calibration optimization is important to ensure the efficient use of soil spectral libraries for the accurate estimation of soil carbon. Moreover, spectral library transfer can benefit new data collection, soil monitoring, and modeling efforts. This review presents techniques for optimization of calibration models and library transfer. Selection of calibration set size and subsetting are presented as current calibration optimization techniques. Moreover, spiking is discussed as an effective technique for spectral library transfer. Overall, studies have suggested that an increase in calibration size improves model performance and this continues until an optimal size is reached. Additionally, subsetting can improve model performance if the resulting subsets reduce the variability of spectrally active components. Studies have also suggested that spiking is effective when used in conjunction with subsetting techniques. These findings denote the current applicability and potential of optimization and library transfer techniques for the accurate estimation of soil carbon with soil spectroscopy. Future efforts should focus on refining

Abbreviations: ANN, artificial neural network; CARS, competitive adaptive reweighted sampling; CT, committee trees; ED, Euclidean distance; EDF, exponential decreasing function; GA-PLSR, genetic algorithm partial least squares regression; GPR, Gaussian process regression; HEM, heteroscedastic effects model; IQR, interquartile range; LW-PLSR, locally weighted partial least squares regression; LWR, locally weighted regression; MARS, multiplicative adaptive regression splines; MBL, memory-based learning; MD, Mahalanobis distance; MIR, mid-infrared; MLR, multiple linear regression; NIR, near-infrared; oPC-MD, optimized principal components Mahalanobis distance; OPS, ordered prediction selection; PAM, partitioning around medoids; PCA, principal components analysis; PC-MD, principal components Mahalanobis distance; PLSR, partial least squares regression; RMSE, root mean square error; RPD, ratio of performance to deviation; RPIQ, ratio of performance to interquartile range; SBL, spectrum-based learner; SEP, standard error of prediction; SOC, soil organic carbon; SPLSR, sparse partial least squares regression; SSL, soil spectral library; SVMR, support vector machine regression; VIP, variable importance for projection; VNIR, visible and near-infrared; VNIR-SWIR, visible and near-infrared and short-wave infrared.

optimization techniques to further expand the operability of soil spectroscopy for soil carbon estimation.

1 | INTRODUCTION

Measuring and monitoring soil carbon is fundamental to the management of food security, environmental health, and plant and animal welfare. There is increasing demand for soil carbon data to support carbon market monitoring, reporting, and verification, environmental monitoring, precision agriculture, and spatial modeling (Brown et al., 2006; Sanderman et al., 2021; Wijewardane et al., 2018). To satisfy this demand, resource-efficient techniques for accurate soil carbon estimation are necessary. The accurate estimation of soil carbon for the aforementioned efforts can be difficult to achieve due to costly, intrusive, and time-consuming traditional laboratory methods (Dotto et al., 2018; Smith et al., 2020). Regardless of the exact method of laboratory analysis, the monetary and environmental cost associated with quantifying soil carbon is a barrier to wide-scale monitoring and informed decision-making.

Over the past few decades, infrared soil spectroscopy has become prevalent as an alternative to traditional soil carbon analysis because it is fast, cost-effective, nondestructive, environmentally friendly, robust, and adaptable for use in the lab or in situ (Barra et al., 2021; Gholizadeh et al., 2013; Nocita et al., 2015; Viscarra Rossel et al., 2006, 2016). Soil spectroscopy is less destructive than traditional laboratory analysis as only a relatively small sample and minimal sample preparation are required. Samples may only need to be dried and ground prior to scanning and scanning may only take seconds leading to cost and time savings. Moreover, soil spectroscopy does not require the use of hazardous chemical extractants; therefore, it is less harmful to the environment (O'Rourke & Holden, 2011; Viscarra Rossel et al., 2006). Lastly, a single spectrum can be used to assess several soil properties, making it a robust analysis method (Comstock et al., 2019; McBratney et al., 2006; Viscarra Rossel et al., 2006, Viscarra Rossel et al., 2008).

The practical use of soil spectral data depends on the construction of a soil spectral library (SSL). A SSL is a database containing spectra and their corresponding soil property measurements determined by traditional methods, defined here as those other than spectral based. For a SSL to be useful, the soil property measurements (i.e., analyte data) and associated spectral data must be from a reliable laboratory procedure (Viscarra Rossel et al., 2008). Moreover, the SSL should contain sufficient soil samples to capture the expected soil variability in the area where it will be applied (Minasny et al., 2009; Reeves, 2010).

In soil spectroscopy, a calibration model relates the spectral data to the analyte data of soil samples to predict soil chemical or physical properties. An important process in the construction of calibration models is optimization. Optimization of calibration models focuses on reducing the statistical error of model estimates and helps ensure the efficient use of a SSL for the prediction of soil properties. Furthermore, optimized calibration models built from an existing SSL can be used to estimate soil properties at a new site through library transfer techniques. The review presented here is an effort to provide an overview of previous work and current trends in calibration optimization and library transfer techniques for soil spectroscopy. This work does not discuss spectral pre-processing techniques, nor does it intend to compare model performance across different spectral ranges (e.g., mid-infrared, near-infrared, etc.), both of which can influence soil property estimates. For more information on those topics, the reader is referred to Vasques et al. (2008) and Bellon-Maurel and McBratney (2011), respectively. Studies cited in this work are those pertinent to the estimation of soil carbon. This property was selected as the focus of this work given its importance to soil quality and soil health as well as the growing demand for soil carbon data for climate change monitoring (Lal, 2014; Smith et al., 2020).

2 | CALIBRATION MODELS AND OPTIMIZATION

The estimation of soil properties using soil spectroscopy is conducted through calibration models constructed from observations that relate analyte data (e.g., organic carbon concentration) to spectral data (e.g., absorbances across a spectral range). The spectral data and corresponding analyte data used to construct these models is often termed the "training set." Construction of a calibration model from a training set requires the application of statistical learning techniques that consist of computational-statistical procedures to construct estimation/prediction models with improved accuracy through iterative "learning" and fitting (Tibshirani et al., 2017). The accuracy of a calibration model is a measure of its systematic error, which is defined as the difference between the model estimates/predictions and the accepted true value of the soil property. In general, the assessment of calibration model accuracy should be conducted using an independent "validation set" (Bellon-Maurel & McBratney, 2011; Brown et al., 2005; Gemperline, 2006).

Calibration model optimization is a fundamental process in soil spectroscopy that focuses on improving overall model performance (e.g., reducing statistical error or bias). Calibration model optimization routines can determine the number of observations required to achieve an acceptable model accuracy, as well as improve the representativeness of the spectral data and their relationship to the analyte data. Moreover, some optimization routines consider the soil variability in the calibration set, which is important when observations are from soils developing under different environmental conditions, weathering stages, or soil depths. Calibration model optimization techniques are discussed next and where available, measurements of model error (e.g., RMSE) are presented. Unless otherwise stated, these error values are based on an independent validation set, as reported by the corresponding authors.

Core Ideas

- Effective optimization routines increase representativeness of a calibration set to unknowns.
- Subsetting by spectral similarity and local calibrations improve model accuracy.
- Spiking with subsetting for library transfer can improve model accuracy.

2.1 | Calibration set size

Calibration set size affects model performance. If infrared spectroscopy is to be considered a cost-efficient method of soil analysis, then it is important to determine the optimal number of samples required not only in terms of its effect on model performance, but also for its cost-savings potential. The calibration model should contain sufficient observations to capture the variability of the soils in the area where it will be applied (Viscarra Rossel et al., 2008). Studies have reported that model accuracy increases with calibration size until a point is reached when no additional significant improvement is achieved (see Figures 1 and 2) (Angelopoulou et al., 2020; Clairotte et al., 2016; Debaene et al., 2014; Gogé et al., 2014; Grinand et al., 2012; Lucà et al., 2017; Shepherd & Walsh, 2002). An optimal calibration size is one at which a good tradeoff between model accuracy and resource efficiency is found. However, determining the optimal calibration size is not straightforward. Building a calibration model from many soil samples is neither cost nor time efficient and it can lead to increased noise in the model. Furthermore, conducting statistical analysis on a set with a large number of observations can be computationally expensive (Debaene et al., 2014; Lucà et al., 2017). On the contrary, building a calibration model from a few soil samples may save time and money, but can lead to inaccurate predictions (Lucà et al., 2017).

Several studies have examined the effect that varying the calibration set size has on model performance. Shepherd and Walsh (2002) assessed the effect of decreasing the size of a highly diverse calibration set on soil organic carbon (SOC) model performance. The authors observed that the R^2 of the independent validation was less variable and thus more stable for models constructed using 20 and 30% of randomly selected observations from the calibration set. They noted that, when starting with a large set size, the predictive performance decreased gradually with decreasing sample size. Contrarily, when starting with a small set size (approximately <20% of total observations), the predictive performance decreased abruptly with decreasing sample size (Shepherd & Walsh, 2002). This indicates that the magnitude of influence of each observation in the calibration set is not constant, but rather is influenced by the initial calibration set size. Using a French national mid-infrared (MIR) database, Grinand et al. (2012) tested the effect of calibration set size by systematically increasing the proportion of total observations used for the calibration with the remaining observations used for validation. Similar to the study by Shepherd and Walsh (2002), these authors achieved stable validation results for SOC when calibration models were constructed with a random selection of 20% (R^2 = .89, ratio of performance to deviation, RPD = 3.00, standard error of prediction, SEP = 0.67%) and 30% (R^2 = .89, RPD = 3.10, SEP = 0.65%) of the total observations. Additionally, the authors noted that there was a significant increase in the RPD and R^2 when the calibration set size was increased from 10 to 20% (R^2 = .84 vs. .89, RPD = 2.50 vs. 3.00). Similarly, the SEP decreased from 0.80 to 0.67%, respectively. Contrarily, there was only a minimal decrease in error when the calibration set size was increased from 20% (SEP = 0.67%) to 80% (SEP = 0.59%), with reduced stability of validation metrics at larger calibration sizes. The authors attributed these results to the proportion of atypical observations in the calibration at larger calibration set sizes. These studies suggest that model accuracy increases with an increase in calibration set size, but the influence of additional soil samples for the calibration set is dependent on the initial calibration size and the proportion of atypical observations added to the calibration set.

Clairotte et al. (2016) tested the separate and combined use of visible and near-infrared (VNIR), near-infrared (NIR), and mid-infrared (MIR) spectra from a French national spectral database to determine the minimum calibration set intensity (i.e., optimal percentage of calibration observations) required to obtain an accurate prediction for an SOC dataset with a range of 0.2–6.3%. The authors tested 10 different calibration set intensities ranging from 10 to 100%, in 10% increments. Results of the randomly selected calibration models demonstrated that the RPD and ratio of performance to interquartile

range (RPIQ) increased gradually and the SEP decreased gradually with increasing calibration intensity but observed very little improvement above 60% intensity. Furthermore, they determined that the optimal calibration intensity was greater for the calibration set that only used MIR spectra (50% intensity, SEP = 0.63%), as compared to those sets that used VNIR (30% intensity, SEP = 0.92) and NIR spectra (30% intensity, SEP = 0.85%). Nevertheless, better predictions were achieved using only MIR spectra (lowest SEP = 0.60%) than using VNIR (lowest SEP = 0.87%) or NIR (lowest SEP = 0.82%). The authors suggested that VNIR and NIR contain less useful information than MIR for predicting SOC and thus, require less calibration observations to extract the useful information and achieve their best model performance (Clairotte et al., 2016).

Several studies have investigated calibration set size in conjunction with different sample selection schemes including random sampling, stratified random sampling, Kennard–Stone (Kennard & Stone, 1969), and analyte value range. Brown et al. (2005) used VNIR models to assess the effect of three sampling schemes and a varying percentage of total calibration observations (10–70%) on the prediction of SOC in north-central Montana. The sampling schemes were (a) random sampling, (b) stratified random sampling of soil profiles per site, and (c) spectrally stratified random sampling using partitioning around medoids (PAM) (Kaufman & Rousseeuw, 1990). They observed a decrease in RMSE with an increase in observations more than 20% of the total dataset (57 of 283) and predictions with RMSE <0.14% with at least 35% of the total dataset across all sampling schemes. However, model performance varied depending on the sampling scheme. The models constructed from spectrally stratified sampling outperformed those of the other sampling schemes and consistently resulted in lower maximum RMSE values when 20–35% of the total dataset was used, indicating that sample selection influences the results. In their study on the separate and combined use of VNIR, NIR, and MIR spectra to predict SOC, Clairotte et al. (2016) also tested the effect of Kennard–Stone sampling on optimal calibration intensity. The authors noted that the optimal calibration intensity was greater with Kennard–Stone selection of calibration samples than with random sampling (MIR: 60 vs. 50%; VNIR: 50 vs. 30%; NIR: 70 vs. 30%). Nevertheless, much better predictions were achieved by models constructed from Kennard–Stone samples as compared to those from random sampling (lowest SEP with Kennard–Stone vs. random sampling and MIR: 0.26 vs. 0.60%; VNIR: 0.48 vs. 0.87%; NIR: 0.44 vs. 0.82%).

Several studies have tested the effect of varying calibration set size for the prediction of soil carbon at local scales. Debaene et al. (2014) investigated the effect of VNIR calibration set size on model performance for the within-farm prediction of SOC concentration. Four sampling schemes were used to select the calibration set: (a) random sampling, (b) selective sampling by analyte value, (c) spectrally stratified random sampling using K-means clustering, and (d) spectrally stratified random sampling using principal components analysis (PCA) scores. The difference in lowest RMSE as well as the calibration size required to achieve the lowest error was small between the differently selected calibration models. Overall, random sampling achieved the smallest RMSE with the fewest observations. The RMSEs of the random sampling and analyte value models ranged from 0.12 to 0.18% and were achieved using approximately 60% of the calibration set observations. The K-means clustering models had the widest range in RMSE (0.12–0.27%) as well as the smallest proportion of the calibration set required to achieve this RMSE (57%). The PCA score models had RMSEs between 0.12 and 0.22% with the minimum achieved using approximately 67% of the calibration set. The authors determined that a minimum of 79 of the total 199 calibration observations (approximately 40%) were suitable to adequately predict SOC concentration with a RMSE of 0.13%. Using a French national spectral database in the VNIR range, Gogé et al. (2014) compared various strategies to predict SOC concentration for a local site. The authors observed the effect of calibration size on model accuracy and noted that model RMSE and bias decreased and R^2 increased as the number of observations, selected using the Kennard-Stone algorithm, increased.

The effect of calibration set size on total soil carbon prediction at a local scale using VNIR was tested by Lucà et al. (2017). Three calibration models, selected through stratified sampling by analyte value, were assessed and they each achieved different levels of performance depending on the calibration set size. In general, the RMSE decreased as the calibration set size increased. The best predictive performances were obtained using between 50 and 90% (72 and 130 of 144 observations) of the total calibration set. In addition to these studies, Ramirez-Lopez et al. (2014) investigated the combined effect of calibration set size and three calibration sampling algorithms: Kennard–Stone, conditioned Latin hypercube (McKay et al., 1979; Minasny & McBratney, 2006), and fuzzy c-means (de Gruijter et al., 2010). These authors found that the improvement in model performance by spectrally stratified random sampling depends on the calibration size. When models are small, the sampling algorithm significantly improves model accuracy; however, when the models are large, the sampling algorithm has little influence on model accuracy. Although random sampling is a statistically sound sample selection method, it is prone to select samples with little representativeness to the whole set, particularly when working with a large SSL composed of highly, pedologically diverse soil samples. In these cases, a spectrally stratified sampling approach such as Kennard–Stone, fuzzy c-means, or conditioned Latin hypercube sampling may be preferred.

Overall, these studies confirm previous findings that model accuracy increases with an increase in the calibration set size. Additionally, these studies demonstrate that the optimal calibration size depends on various factors, including the initial calibration set size, the sampling scheme used to select the calibration observations, and the spectral range of the calibration model. In addition to the aforementioned factors, the optimal calibration set size can vary depending on the mineralogical diversity and the geographical extent covered by the observations in the calibration (Clingensmith et al., 2019; Lucà et al., 2017; Ludwig et al., 2019).

2.2 | Sample representativeness

The representativeness of the calibration set is another optimization factor that influences model performance. The empirical nature of spectroscopic calibrations limits their prediction accuracy to how well the calibration observations represent the unknowns (Nocita et al., 2015). To construct a robust calibration model, the observations in the calibration set must be representative of the soils to which the model will be applied (Angelopoulou et al., 2020). Lucà et al. (2017) indicates that a representative sample set should be selected on the basis of spectral features or analytical properties. It is important to consider both the expected variability in soil chemical and physical properties that are spectrally active, as well as the expected distribution of the soil property values of the unknown observations. Additionally, limiting the range of variability in spectrally active properties and analyte concentrations of the calibration model can improve model performance. For example, NIRS studies on forages and grains obtained better results with calibration models developed for a limited, well-defined population (e.g., a specific varietal), as opposed to a universal calibration for all varieties (Murray et al., 1987; Roberts et al., 2004). Similarly, soil calibrations can perform better if constructed for a reduced spectral, pedologic, geographic, or analyte concentration range (Brown et al., 2005; Madari et al., 2005; Reeves & Smith, 2009).

Calibration models can be constructed to estimate a specific group of the prediction set, such as specific analyte value ranges or soil types. This is achieved by subsetting the SSL using calibration selection approaches (Soriano-Disla et al., 2014). One approach is to construct local calibration models using the nearest spectral neighbors of the prediction set (i.e., the local approach). Another approach is the stratification of the SSL to select subsets based on ancillary information or classification criteria to build targeted calibrations. Examples of selection criteria include soil types or factors known to influence soil properties and presumably also the spectral response. It is important to note that using this ancillary information is a cost-effective approach when soil information systems (e.g., a soil survey with taxonomic attributes that can be related to the collected soil samples) are readily available. This approach may not be feasible in scenarios where soil information systems are not available or in the appropriate scale for accurate representation of soil spatial variability. Once the targeted calibrations are constructed, they can be used to predict the target subset or group of unknowns (McDowell, Bruland, Deenik, & Grunwald, 2012; Soriano-Disla et al., 2014). The local model approach and the stratification approach, each hereafter referred to as subsetting, can be used either independently or simultaneously, as well as in conjunction with other optimization techniques to improve model performance (Lucà et al., 2017).

Table 1 summarizes some studies of spectroscopy for soil carbon estimation that have applied subsetting techniques. The example studies presented in Table 1 are not an exhaustive representation of the literature on calibration set subsetting. However, they are representative of subsetting criteria discussed in this paper and of the various techniques used in recent soil spectroscopy studies.

2.2.1 | Subsetting by analyte value

Several studies have explored the effect of subsetting by analyte value. Janik and Skjemstad (1995) split the total dataset into three subsets by range in SOC concentration (0–2.5%, 2.5–10%, 9–25%) to improve model accuracy of a partial least squares regression (PLSR) based on cross-validation. While the calibration models constructed from the lowest and narrower ranges (0–2.5% and 2.5–10%) resulted in a larger R^2 (.979), the highest and wider range model (9–25%) performed worse than the full set calibration model (R^2 of .892 vs. .975). This discrepancy may be due to the small calibration size of the highest range model. McDowell, Bruland, Deenik, & Grunwald et al. (2012) investigated the effect of subsetting a VNIR and MIR calibration set by various soil sample characteristics, among them total carbon concentration. The authors determined through preliminary analysis, that subsetting the calibration observations into low (0–10%) and high (10–55%) total carbon concentrations produced the best results for both spectral ranges. Consequently, they used 10% as the threshold value for subsetting the calibration set. The low carbon models, which also had a narrower range of analyte values, decreased in RPD (VNIR: 3.46 to 1.63; MIR: 4.07 to 2.34), RPIQ (VNIR: 3.19 to 2.12; MIR: 3.74 to 3.05) and R^2 (VNIR: 0.91 to 0.61; MIR: 0.94 to 0.82) as compared to the total set model. No improvement was observed with the high carbon model in comparison to the full set model.

In a study that used VNIR for SOC prediction, Vasques et al. (2010) developed different calibration models based on a general soil type, which inadvertently split the observations by lower and higher carbon concentration (0.01–14.7%

TABLE 1 Summary of soil spectroscopy studies that use subsetting for calibration set optimization and whether subsetting improved soil carbon prediction

Subsetting criteria	Technique	Total improvement over full set calibration[a]	Spectral range	Reference
Analyte value	SOC range	Yes	MIR	Janik & Skjemstad (1995)
Analyte value	Total carbon range	No	NIR and MIR	Madari et al. (2005)
Analyte value	SOC range	Yes	VNIR	Vasques et al. (2010)
Analyte value	Total carbon range	No	VNIR and MIR	McDowell, Bruland, Deenik, and Grunwald (2012)
Pedodiversity	Geographic extent	Yes	NIR	Sudduth & Hummel (1996)
Pedodiversity	Taxonomic soil class	No	NIR and MIR	Madari et al. (2005)
Pedodiversity	Textural group	Yes	NIR and MIR	Madari et al. (2005)
Pedodiversity	Textural group	Yes	NIR	Brunet et al. (2007)
Pedodiversity	Taxonomic soil order	Yes	VNIR	Vasques et al. (2010)
Pedodiversity	Taxonomic soil order, mineralogy, SOM	Yes	VNIR and MIR	McDowell, Bruland, Deenik, and Grunwald (2012)
Pedodiversity	Geographic extent	Yes	MIR	Baldock et al. (2013)
Pedodiversity	Geographic extent	Yes	VNIR	Peng et al. (2013)
Pedodiversity	Taxonomic soil order	No	MIR	Wijewardane et al. (2018)
Pedodiversity	Master horizon	Yes	MIR	Wijewardane et al. (2018)
Pedodiversity	Taxonomic soil order + land use	Yes	VNIR-SWIR	Moura-Bueno et al. (2019)
Pedodiversity	Physiographic region, land use, textural class	Yes	VNIR	Moura-Bueno et al. (2020)
Spectral similarity	Local model: LW-PLSR	Yes	NIR	Christy & Dyer (2006)
Spectral similarity	Local model: LOCAL	Yes	NIR	Fernández Pierna & Dardenne (2008)
Spectral similarity	Local model: LW-PLSR	No	NIR and MIR	Igne et al. (2010)
Spectral similarity	Local model: LOCAL	Yes	NIR	Genot et al. (2011)
Spectral similarity	Local model: LW-PLSR, LOCAL	No	VNIR	Ramirez-Lopez, Behrens, Schmidt, Stevens, et al. (2013)
Spectral similarity	Local model: LW-PLSR	Yes	VNIR	Nocita et al. (2014)
Spectral similarity	Local model: LW-PLSR	Yes	VNIR	Gupta et al. (2018)
Spectral similarity	Local model: SBL	Yes	VNIR	Ramirez-Lopez, Behrens, Schmidt, Stevens, et al. (2013)
Spectral similarity	Local model: SBL	Yes	MIR	Dangal et al. (2019)
Wavelength selection	GA-PLSR	Yes	VNIR	Vohland et al. (2011)
Wavelength selection	CARS-PLSR	Yes	VNIR and MIR	Vohland et al. (2014)
Wavelength selection	OPS-PLSR	Yes	VNIR	Sarathjith et al. (2016)
Wavelength selection	SPLSR, HEM,	Yes	VNIR	Clingensmith et al. (2019)
Wavelength selection	Automatic selection of wavenumber regions (Ludwig et al., 2019)	Yes	MIR and NIR	Ludwig et al. (2021)

Note. CARS-PLSR, competitive adaptive reweighted sampling partial least squares regression; GA-PLSR, genetic algorithm partial least squares regression; HEM, heteroscedastic effects model; LW-PLSR, locally weighted partial least squares regression; MIR, mid-infrared; NIR, near-infrared; OPS-PLSR: ordered prediction selection partial least squares regression; PCR, principal component regression; SBL, spectrum-based learner; SOC, soil organic carbon; SOM, soil organic matter; SPLSR, sparse partial least squares regression; VNIR, visible and near-infrared; VNIR-SWIR, visible and near-infrared and short-wave infrared.

[a]Total improvement measured as decrease in error of prediction between that of the full calibration set and at least one calibration subset.

and 13.52–57.54%, respectively). Contrary to the results from McDowell, Bruland, Deenik, & Grunwald (2012), both subset models achieved higher RPD (lower C model: 1.26, higher C model: 1.20) and R^2 (lower C: .41, higher C: .38) as compared to the full set model (RPD: 1.12 and R^2: .29). Madari et al. (2005) subset observations into three groups of varying range in SOC concentration (0.02–40.19%, 0.02–6.60%, and 0.02–3.00%) to construct calibration models. Through cross-validation, the full set models (i.e., 0.02–40.19%) resulted in a greater R^2 (MIR: 0.934, NIR: 0.809) than the subset models (MIR: 0.840 and 0.810; NIR: 0.726 and 0.712 for the lower and higher SOC subsets, respectively). In general, the MIR models outperformed the NIR models based on cross-validation results.

The results of these studies demonstrate that the effect of subsetting by analyte value varies and may be influenced by other factors. In general, subsetting a calibration set by analyte value alone is most useful for improving prediction accuracy when the overall variability in spectra is low. Accordingly, the greatest variability in the calibration model will result from the variability in the analyte values (Clingensmith et al., 2019). When both spectral and analyte value variability are high, subsetting to account for both sources of variability can lead to better model performance. This is an important consideration for deciding when and how to subset the calibration set. Additionally, an adequate statistical comparison of model performance across different ranges of analyte values should not only be based on RMSE, given that this will decrease as the analyte range decreases (Stenberg et al., 2010), but also requires a comparison of R^2, RPD, and RPIQ in the context of the respective interquartile range (IQR), if possible (Ludwig et al., 2021).

2.2.2 | Subsetting by pedodiversity

Several spectroscopic studies have used the variation in soil types and properties (i.e., pedodiversity) to subset the calibration set for SOC modeling. Subsetting criteria based on pedodiversity include taxonomic classification, soil-landscape/geographic region, and soil-forming factors. A soil taxonomic classification indicates a range of properties that are limited by the soil parent material, mineralogy, and climate (Seybold et al., 2019). Knowledge about the relationships between taxonomic units and soil properties has been used to relate SOC to soil-forming factors at the landscape scale (Wills et al., 2013).

Soil mineralogy and texture are spectrally active because their components interact with electromagnetic radiation and thus, cause variation in reflectance features (Moura-Bueno et al., 2019). Stenberg et al. (2010) argue that models are more robust and perform better when constructed from a large, heterogeneous calibration set from soils with diverse parent materials (Vašát et al., 2017). Parent materials contribute different minerals and particle sizes that can better represent the potential characteristics of the prediction set (Nawar & Mouazen, 2017; Stenberg et al., 2010). However, problems with a diverse calibration set can arise if the unknowns are very different from the calibration set in terms of property values and spectrally active properties (Bellon-Maurel & McBratney, 2011; Brown et al., 2005, 2006; Sankey et al., 2008; Wijewardane et al., 2018).

The high spatial variation of SOC is another important consideration when subsetting by pedodiversity (Schmidt et al., 2010). Identifying these patterns of variability is important as soils belonging to distinct patterns should be modeled separately (McBratney et al., 1991). One approach is to stratify the data by spatial units of similar landscape and soil-forming factors (i.e., soil-landscape units). Presumably, soil properties within a soil landscape will be less variable as compared to the soil population across the landscape due to interactions between soil-forming factors (McCarty & Reeves, 2006). In general, soil spectroscopy studies have demonstrated that models constructed through subsetting by spectral or pedologic criteria perform better than those that do not subset the calibration set (Ramirez-Lopez, Behrens, Schmidt, Stevens, et al., 2013).

Madari et al. (2005) used NIR and MIR spectral data from diverse Brazilian soils to model SOC. These authors subset their calibration sets by taxonomic soil class. Two subset models resulted, one for soils classified as Ferralsols and the other for Acrisols according to the World Reference Base (FAO, 1998). These subset models achieved lower R^2 values as well as lower RMSEs (MIR: R^2 = .862 and .905, RMSE = 0.545 and 0.449%; NIR: R^2 = .725 and .784, RMSE = 0.770 and 0.675% for Ferralsols and Acrisols models, respectively) as compared to the full set calibration models (MIR: R^2 = .934, RMSE: 1.088%; NIR: R^2 = .809, RMSE = 1.855%). The authors concluded that overall, the models fitted by taxonomic class did not outperform the full set model for MIR and NIR.

In a study to estimate SOC concentration of soils in Florida by VNIR spectroscopy, Vasques et al. (2010) tested the effect of subsetting the observations in the calibration set by soil order (Alfisols, Entisols, Histosols, Inceptisols, Mollisols, Spodosols, and Ultisols) on PLSR model performance. Additionally, the authors tested the performance of a committee trees (CT) model that included soil order as a categorical variable, and which was fitted with the full set and another fitted through subsetting by mineral vs. organic horizon. For the PLSR model, subsetting the observations by soil order improved the R^2, RPD, and RMSE for six of the seven soil orders, as compared to the full set PLSR model (Alfisols: 0.58/1.54/0.51%, Entisols: 0.50/1.36/0.93%, Inceptisols: 0.42/1.24/1.19%, Mollisols: 0.68/1.54/0.90%, Spodosols: 0.56/1.41/0.70%, Ultisols: 0.75/1.91/0.33%, and full set: 0.29/1.12/4.60% for R^2/RPD/RMSE). An

important consideration is that the values presented are the result of back-transformation of logSOC estimates to the original units, which the authors noted significantly reduced the quality of the PLSR models resulting in unreliable estimates. On the other hand, the R^2 and RPD of the CT model did not improve by including a categorical variable of soil orders (0.65/1.69/0.69% for R^2/RPD/RMSE) nor by subsetting by mineral/organic horizon type (Mineral: 0.66/1.70/0.70% and Organic: 0.35/1.23/10.23% for R^2/RPD/RMSE), as compared to the full set CT model (0.79/2.14/2.52% for R^2/RPD/RMSE). Moreover, the full set CT model outperformed the PLSR models that were subset by soil order. These findings pose an important consideration that the type of statistical learning model used can lessen the benefit of subsetting by pedodiversity.

Wijewardane et al. (2018) investigated whether subsetting an MIR calibration set by land use/cover, soil order, and soil master horizons improved the prediction accuracy of SOC. The authors developed calibration models for each subset using the PLSR and artificial neural network (ANN) models. On average, subsetting by all three criteria reduced the RMSE of the PLSR models as compared to that fitted with the full set. Moreover, subsetting by soil order and master horizon resulted in lower statistical error than subsetting by land use/cover. Most of the ANN models, including the full set model, outperformed the PLSR models (R^2/bias/RPD/RPIQ/RMSE of 0.95/0.00%/4.55/0.82/1.89% and 0.99/−0.01%/11.46/2.05/0.75% for the full-set PLSR and ANN models, respectively). These results can be attributed to the superior capacity of ANNs in modeling complex and nonlinear relationships between analyte value and spectra. Moreover, although the subset ANN models outperformed all the PLSR models, they did not outperform the full set ANN model. These results corroborate those of Vasques et al. (2010), who also found that the effectiveness of subsetting for reducing the error of calibration models depends on the statistical learning model. Models such as CTs and ANNs may not benefit from subsetting because they can handle complex relationships in high-dimensional feature space. Statistical models based on machine and deep learning handle relationships in a similar way to manual subsetting and thus, the improvement in model performance by subsetting is little to none (Viscarra Rossel & Behrens, 2010).

In a study of total carbon in Hawaiian soils, McDowell, Bruland, Deenik, & Grunwald (2012) fit MIR calibration models for broad soil groups. The soil groups were defined as sets of soil orders with similar clay mineralogy and soil organic matter concentration (Group 1: Andisols; Group 2: Aridisols, Entisols, Inceptisols, Mollisols, and Vertisols; Group 3: Oxisols and Ultisols; and Group 4: Histosols and Spodosols). The calibration model constructed for Group 2 (high-activity clay soil orders) resulted in greater accuracy (R^2: .96 and RPD: 5.57) than that of the full set calibration model (R^2: .94 and RPD: 4.07). An interesting result of this study was that the within-subset spectral variability was equally as high as that between subsets. The authors explained that soil taxonomic classification is based on properties that are often not spectrally active and that spectrally active properties, such as mineralogical properties, may not be exclusive to a taxonomic classification level (McDowell, Bruland, Deenik, Grunwald, & Knox, 2012). Consequently, subsets based on soil orders can contain spectral features that are not mutually exclusive, which negatively affects model performance. As presented in the study by McDowell, Bruland, Deenik, Grunwald, & Knox (2012), a limitation exists in using single-criterion taxonomic subsets such as soil orders or horizonation, given the high within-order variability present in pedologic conditions, such as highly dissimilar A and B horizons of a soil profile. In scenarios where high within-group variability is expected, multi-criteria subsetting, such as soil order coupled with horizonation, can be more useful.

Moura-Bueno et al. (2019) stratified a visible-near-infrared and short-wave infrared (VNIR-SWIR) spectral library of 810 observations using various combinations of two distinct soil classes and three land use types to construct calibration models for SOC% prediction. The full set was stratified into subsets based on the mean spectrum for each criterion and a quantitative analysis of the distribution of variance of the projected spectral data. Overall, in models with a sufficient calibration size ($n > 77$), subsetting by soil and land use type improved model performance. The subset models resulted in an R^2 of .42–.82, RMSE of 0.29–0.70%, and an RPIQ of 1.99–2.60. The best model performance was achieved by a single soil type-single land use subset (R^2 = .82, RMSE = 0.29%, and RPIQ = 2.60), which used 45% less observations than the full-set model (R^2 = .74, RMSE = 0.55%, and RPIQ = 2.16). The authors attributed the better performance of this subset model to a reduction in the spectral variance, soil textural variance and SOC concentration of the calibration subsets. Additionally, results of the best-performing subset models demonstrated that soil spectral and compositional characteristics had a greater effect on model performance than the calibration set size. The authors concluded that, while the spectral library was local, spectral variability was high and subsetting the library to reduce spectral and soil property variability was effective in improving model performance. Furthermore, they proposed that future studies should consider a weighted sampling approach for the construction of calibration models that assigns weights according to the spectral and compositional variation captured by each observation (Moura-Bueno et al., 2019).

Demattê and da Silva Terra (2014) examined the relationship between VNIR spectra and soil pedogenic properties along a toposequence (i.e., a soil catena). The authors observed that variations in reflectance intensity, specific wavelengths, and spectral shape enabled the detection of

distinct mineralogy and textures. The spectral variations observed across soil depth helped distinguish between soil classes. The authors concluded that soil spectroscopy was able to discriminate between weathering levels and the presumed pedogenic processes (Demattê & da Silva Terra, 2014). Although these authors did not perform subsetting for calibration models, their study suggests that subsetting spectral data by criteria associated with pedogenic processes can be useful for taxonomic purposes, especially at the soil catena scale.

Various studies have used soil texture as subsetting criteria. Typically, as clay increases, so does SOC; however, this relationship can be confounded in spectroscopy by the spectral response of sand (Soriano-Disla et al., 2014; Stenberg et al., 2010; Vasques et al., 2010). Sand particles in a sample can influence the spectral response of SOC features (Stenberg et al., 2010). Consequently, soil samples with high sand and low SOC concentration can be very similar to samples with low sand and high SOC concentration (Nocita et al., 2015). Therefore, including particle size or textural classes in soil spectroscopic models for SOC of soil samples presumed to have high sand, can result in better prediction accuracy (Vasques et al., 2010).

Madari et al. (2005) performed subsetting by soil textural groups for the prediction of SOC with NIR and MIR spectra. The textural groups were defined based on the following particle fractions: very clayey (>60% clay), clayey (35–60% clay), and medium textured (<35% clay and >15% sand). The NIR calibration models for the very clayey subset, resulted in better cross-validation model performance than their MIR counterparts (R^2 = .975 and .967, respectively). On the contrary, the MIR calibration model for the clayey and medium-textured subsets (R^2 = .962 and .917, respectively) outperformed its NIR counterpart (0.938 and 0.871). Overall, subsetting by textural class resulted in improved model performance over the full-set calibration (R^2 = .809 NIR and .934 MIR). The authors concluded that models based on NIR are better-suited for sets of observations with homogeneous textures, while MIR models are best for heterogeneous textures (Madari et al., 2005). Accordingly, a subsetting scheme based on soil texture should be complemented by the selection of the appropriate spectral range (NIR or MIR), if possible. A study by Brunet et al. (2007) assessed how the heterogeneity of the soil particle size affects the prediction of total carbon by NIR. These authors constructed calibration models for coarse-textured and clayey subsets. Subsetting the data resulted in improved prediction accuracy as compared to the full-set model (coarse-textured: R^2 = .96, SEP = 0.044%; clayey: 0.89 and 0.150%; full-set: 0.84 and 0.354%). Furthermore, the more heterogeneous, coarse-textured subset model outperformed the more homogeneous clayey subset model.

In a study that explored subsetting a subtropical, Brazilian VNIR spectral library using several criteria independently (i.e., three physiographic regions, three land use and land cover types, and four textural classes), Moura-Bueno et al. (2020) found that subsets that reduced the variance in SOC%, clay%, and spectral variance had an increase in accuracy of SOC predictions as compared to the full-set model. The authors noted that although the full-set model performed well (RMSE = 1.02%, R^2 = .76, bias = −0.22%, RPIQ = 1.51) considering the high variance in SOC% (standard deviation = 1.81%) and diversity in clay mineralogy, subsetting by all criteria, reduced the bias in model predictions (lowest bias = 0.01%). The greatest reduction in RMSE (34% reduction as compared to full-set model) was observed for the land use/cover models (RMSE range = 0.50–1.67%, R^2 = .70–.86, bias = −0.55–0.01%, RPIQ = 1.31–2.71), followed by a 32% reduction in RMSE achieved by the physiographic region models (RMSE range = 0.54–0.97%, R^2 = .53–.93, bias = −0.28–0.04%, RPIQ = 1.63–2.28), and a 5% reduction by the textural class models (RMSE range = 0.56–1.10%, R^2 = .22–.82, bias = −0.25–0.04%, RPIQ = 0.89–2.25).

An interesting finding by Moura-Bueno et al. (2020), was that the diversity of clay mineralogy had a greater effect on spectral variance of the subsets than the clay concentration. Moreover, the authors presented a decision-making flow chart with their strategy on when and how to subset spectral libraries to predict SOC concentration, which they based on their study findings. In general, any decision on whether to subset should begin with an assessment of the analyte and ancillary information available in the spectral database. Next, if significant environmental (i.e., physiographic) and pedologic diversity exists, the observations should first be stratified by physiographic region, then by land use/cover, spectral similarity, and finally by textural class. The final decision on the best subsetting criteria should be based on a reduction of the variance of SOC, clay mineralogy, and spectral variance as compared to the full set calibration model (Moura-Bueno et al., 2020).

The geographic extent of the observations that comprise a calibration model can affect its statistical performance. Some studies suggest that it is better to develop models for smaller areas than for larger areas (Gholizadeh et al., 2013). The assumption being that calibration observations from soils collected across smaller areas will exhibit less variation in soil properties due to soils having similar pedologic conditions, which results in reduced variation and thus more accurate predictions (Kuang & Mouazen, 2011; Shi et al., 2015). It is important to note, however, that a reduction in the geographic extent of a calibration model may or may not reduce the spectral feature space (Ramirez-Lopez, Behrens, Schmidt, Stevens, et al., 2013; Shi et al., 2015). Moreover, building several isolated, small SSLs may not be practical for large-scale modeling or operational purposes.

Several studies have explored the effect of geographic extent on calibration model performance. Sudduth and Hummel (1996) studied the effective geographic range of an NIR

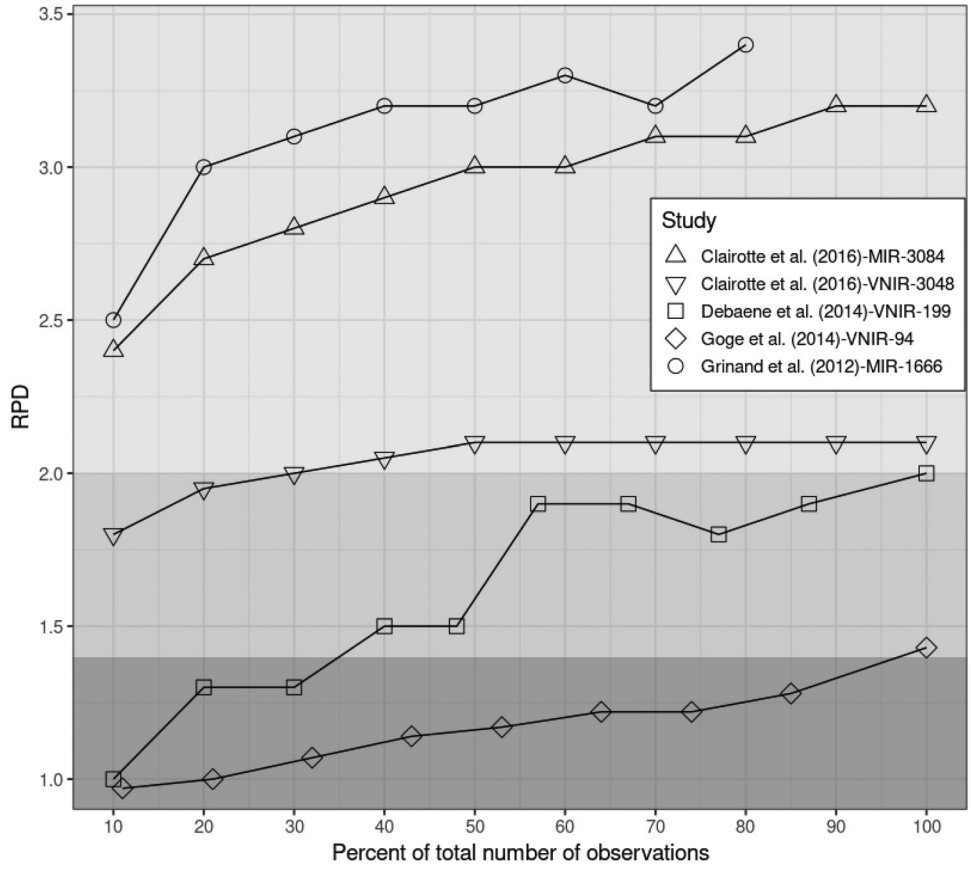

FIGURE 1 Plot of percentage of total number of calibration set observations used for modeling vs. ratio of performance to deviation (RPD) for the prediction of soil organic carbon using soil spectroscopy. The shading indicates distinct model reliability thresholds based on Chang et al. (2001). Dark gray shading represents the range in RPD considered to be an unreliable model, medium gray is for a fair model, and light gray is for a reliable model. The legend provides the study citation, the spectral range (visible and near-infrared [VNIR] or mid-infrared [MIR]) of the dataset, and the total number of calibration set observations in the dataset, in that order. In general, the RPD increases/improves as the percentage of total observations increases until a plateau is reached. The studies cited here are described in detail in this section and used an independent validation set

soil sensor for SOC prediction in the United States. These authors constructed a calibration model using data from soil samples collected in and around the state of Illinois and samples collected across the United States. The calibration model constructed using observations from Illinois and surrounding states was slightly less predictive than that constructed using only observations from Illinois. Furthermore, calibration models that used observations from a more extensive geographic range resulted in unacceptable predictions. The authors concluded that SOC predictions become increasingly less accurate as the geographic range represented by the observations increases (Sudduth & Hummel, 1996). Similarly, Vasques et al. (2008) demonstrated that a VNIR calibration model for SOC prediction in Florida performed better with data from soil samples confined to a watershed, as compared to statewide samples. Like Sudduth and Hummel (1996), the authors concluded that increasing the geographic extent of SOC spectroscopic models can reduce their quality, particularly if geographic-related soil variation is added to the calibration model (Vasques et al., 2010).

In a study to predict carbon and its fractions using MIR spectra, Baldock et al. (2013) found that regional models produced more accurate predictions with lower uncertainty for all analytes than a national calibration model. The RPDs and RMSEs calculated for soils in each regional model were higher and lower, respectively, as compared to the full set model (full set: RPD range = 1.3–4.6 and RMSE range = 0.684–0.240%; regional models: 2.8–4.7 and 0.573–0.185%). These authors observed that the major spectral differences between the observations in the regional and national calibrations were due to differences in mineral components (Baldock et al., 2013). Peng et al. (2013) developed VNIR calibration models for the prediction of SOC at the field scale. These authors subset the calibration set according to the geographic distance between each observation in a national spectral library and the field where the calibration models would be applied. Three calibrations models were developed using observations within 20, 30, and 40 km from the field site. Internal validation of the models revealed that the 30 km calibration subset outperformed the other two

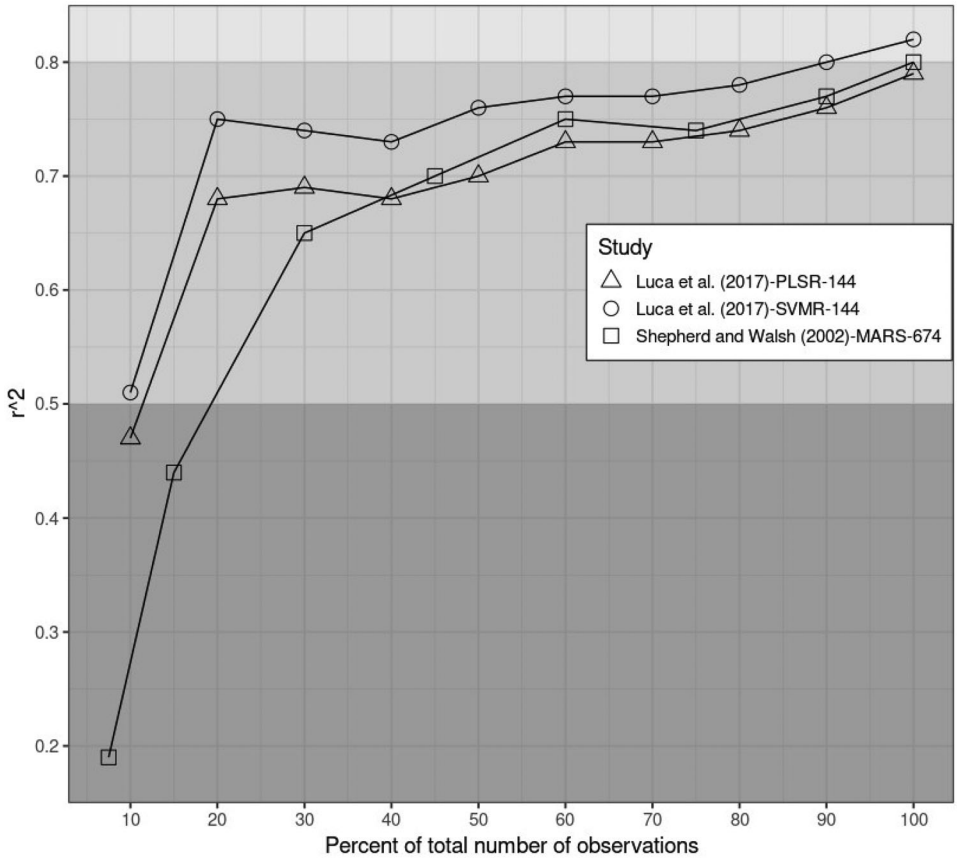

FIGURE 2 Plot of percent of total number of calibration set observations used for modeling vs. R^2 for the prediction of soil organic carbon using soil spectroscopy. The shading indicates distinct model reliability thresholds based on Chang et al. (2001). Dark gray shading indicates the range in R^2 considered to be an unreliable model, medium gray indicates a fair model, and light gray indicates a reliable model. The legend provides the study citation, the modeling approach used (partial least squares regression [PLSR], support vector machine regression [SVMR], multiplicative adaptive regression splines [MARS]), and the total number of calibration set observations in the dataset, in that order. In general, the R^2 increases/improves as the percentage of total observations increases. The studies cited here are described in detail in this section and used an independent validation set

subsets, as well as the full set calibration. The authors concluded that the 30 km calibration subset performed the best because soils within this distance had a similar landscape and parent material, particularly in terms of carbonate concentration, to soils in the field site and were therefore more spectrally similar.

2.2.3 | Subsetting by spectral similarity

In addition to subsetting by soil-related criteria, construction of calibration models can involve subsetting based on spectral similarity/dissimilarity metrics (Reeves & Smith, 2009). This approach aims to construct calibration models from observations that are representative of the spectral features and soil properties in the prediction set. Spectral similarity is defined as observations that are close to each other in the spectral feature space. The distance between the observations can be computed with any distance metric. The most applied distance metrics in soil spectroscopy are the Euclidean distance (ED) and the Mahalanobis distance (MD). The ED and MD can be measured in the spectral space or in a projected space, such as the principal component space. Different variations of the MD in the principal component space have been widely applied in soil spectroscopy. These variations include the principal components Mahalanobis distance (PC-MD) and the optimized principal components Mahalanobis distance (oPC-MD). For more information about these and other distance metrics, the reader is referred to Ramirez-Lopez, Behrens, Schmidt, Rossel, et al. (2013).

2.3 | Local calibrations

A commonly used technique based on spectral similarity is to construct calibration models using only spectral neighbors, which are the spectra most similar to those in the prediction set. Calibration models constructed from spectral neighbors are termed local calibration models. The prediction from a local model is conducted on a case-by-case basis, meaning

that spectral neighbors are found for each observation in the prediction set. This approach assumes that the relationship between spectral features and soil properties is locally stable (Nocita et al., 2014). In this context, a global calibration model refers to a model fitted using all the calibration observations, not only the spectral neighbors of the prediction set (Barthès et al., 2020; Gomez et al., 2020).

Local calibration models can be constructed using memory-based learning (MBL). The MBL approach is a data-driven statistical learning approach that offers instance-oriented models. This means that MBL derives a calibration for each new target spectrum requiring a soil property prediction. The MBL approach selects a relatively small subset of spectral neighbors to predict each unknown observation (Dangal et al., 2019; Lobsey et al., 2017). Four characteristics must be defined for any MBL algorithm: (a) the similarity/dissimilarity metric (i.e., spectral distance metric) used to find the spectral neighbors, (b) how the similarity/dissimilarity information will be used (e.g., used to assign weights, used as predictors, etc.), (c) how many spectral neighbors to consider, and (d) how to fit the local points (i.e., the target function) (Dangal et al., 2019; Ramirez-Lopez, Behrens, Schmidt, Rossel, et al., 2013).

Commonly used MBL models in soil spectroscopy are locally weighted regression (LWR; Naes et al., 1990) and the LOCAL algorithm of Shenk et al. (1997). Locally weighted partial least squares regression (LW-PLSR) is a local version of PLSR that first defines spectral neighbors through the MD in the principal component space (Nocita et al., 2014). These neighbors are then weighted using a function and according to their spectral similarity to the target spectrum. Next, a PLSR is performed for the response value of the target spectrum and its corresponding neighbors to obtain the model coefficients (Gupta et al., 2018; Lobsey et al., 2017; Nocita et al., 2014). As with global PLSR, the regression coefficients are used to predict response values associated with the target spectra. Like LW-PLSR, the LOCAL algorithm calibrates local PLSR models based on spectral similarity; however, there are important differences between the algorithms. First, the LOCAL algorithm uses correlation coefficients as similarity metrics to select spectral neighbors of a target spectrum (Nocita et al., 2014; Shenk et al., 1997). Secondly, the LOCAL algorithm does not apply weights to the spectral neighbors of a target spectrum. Lastly, the predicted response value for each target spectrum results from a weighted sum of the predicted values across all local PLSR models (Fernández Pierna & Dardenne, 2008; Nocita et al., 2014). Both the LW-PLSR and the LOCAL algorithm are better suited for nonlinear predictor-response relationships (Genot et al., 2011; Peng et al., 2013). Nevertheless, as with global PLSR, the principal component space must represent the target spectrum and its corresponding neighbors well to achieve accurate predictions (Naes et al., 1990).

A study by Christy and Dyer (2006) compared the effectiveness of LW-PLSR to predict total carbon using NIR data from seven agricultural fields in Iowa and Kansas. They compared LW-PLSR to three commonly used global regression models, namely multiple linear regression (MLR), using principal components as predictors, and PLSR. The LW-PLSR approach produced the lowest error predictions for total carbon. Genot et al. (2011) used a large NIR spectral library from Belgium to predict total carbon concentration. The authors tested PLSR and the LOCAL algorithm. Additionally, they investigated the effect of increasing the fixed correlation coefficient between the spectra to find the spectral neighbors for each LOCAL model. The LOCAL algorithm outperformed the PLSR and a correlation coefficient value fixed at 0.99 (i.e., the highest value tested), produced the most accurate predictions for the LOCAL algorithm.

Igne et al. (2010) used NIR and MIR spectra to compare the performance of the LW-PLSR against PLSR and support vector machine regression (SVMR) in the prediction of total carbon in Ultisols from a field in Maryland. The LW-PLSR resulted in smaller error than the SVMR but had a similar error to PLSR. The authors concluded that LW-PLSR is a good alternative to global PLSR; however, they stressed the importance of having a balanced number of observations across the value range of the soil property to be predicted. Ramirez-Lopez, Behrens, Schmidt, Stevens, et al. (2013) developed a novel type of MBL termed the spectrum-based learner (SBL). These authors stated that one advantage of the SBL algorithm over other MBL models is that it determines the optimal number of principal components and the number of spectral neighbors for each target spectrum. Moreover, the SBL algorithm, as offered by the R package resemble (Ramirez-Lopez et al., 2016), allows the construction of local models with PLSR, weighted-PLSR, and Gaussian process regression (GPR). The GPR uses a kernel-based function to predict the response value based on the spectral neighbors.

Ramirez-Lopez, Behrens, Schmidt, Stevens, et al. (2013) tested the predictive performance of the SBL approach against PLSR, SVMR, LW-PLSR, and LOCAL for the estimation of SOC concentration using two VNIR SSLs. The SBL algorithm outperformed all other models in terms of RMSE and R^2. The SBL also had much faster processing time than the LW-PLSR and LOCAL. The authors attributed the better performance of the SBL algorithm to its superior ability in selecting spectral neighbors and to the use of the resulting distance matrix as a predictor variable in each local model. The authors also noted that LW-PLSR and LOCAL did not outperform PLSR and SVMR. According to the authors, LW-PLSR and LOCAL performed an inadequate selection of spectral neighbors. The authors stated that like other MBLs, SBL should be used for modeling complex datasets where non-linear relationships exist and they should be avoided in datasets with low variability due to the selective nature of

the spectral neighbors approach (Ramirez-Lopez, Behrens, Schmidt, Stevens, et al., 2013).

Dangal et al. (2019) tested the SBL of Ramirez-Lopez, Behrens, Schmidt, Stevens, et al. (2013) for the prediction of SOC concentration using a continental MIR SSL. The results of the SBL were compared with those from Cubist (see Quinlan, 1993), PLSR, and random forests (Breiman, 2001) models. The SBL model outperformed all the others in terms of RPD and RMSE. Moreover, the SBL model resulted in a slightly greater mean error than the Cubist model, but smaller mean error than PLSR and RF models. The authors concluded that the SBL model is a superior model for large and complex datasets due to its narrower prediction interval and its ability to provide an estimate of prediction uncertainty. Gupta et al. (2018) evaluated the performance of different local modeling approaches and several distance metrics for the prediction of SOC using a small VNIR SSL from India. Among the modeling approaches, there was a LW-PLSR that used a correlation coefficient-based distance metric to weigh spectral neighbors. This model outperformed all the other approaches. The authors determined that the higher prediction accuracy of the LW-PLSR was due to the model assigning higher weights to spectral neighbors with the same mineralogy as the test observations.

Several authors have explored the combined utility of subsetting through local models and pedodiversity. Nocita et al. (2014) applied a modified LW-PLSR for the prediction of SOC from a large, international VNIR SSL. In addition to spectral similarity metrics, these authors also used sand concentration and the geographic coordinates of the calibration observations to find similar observations to those in the target area. These authors noted an inverse relationship between the standard deviation of sand and SOC concentrations. They attributed this relationship to higher texture variations at lower SOC concentrations. The results of their study demonstrated that using sand concentration to find the spectral neighbors produced the most accurate models. Accordingly, the authors suggested the use of sand concentration as subsetting criteria given that spectral differences due to variations in sand are more prominent in low SOC concentrations. Shi et al. (2015) tested the utility of a geographically constrained LW-PLSR for the prediction of soil organic matter concentration using a national, VNIR SSL from China. The resulting model outperformed the unconstrained LW-PLSR. The authors explained that the use of geographical information to select the calibration observations removed uninformative spectra from the calibration model and thus, improved its accuracy.

2.4 | Wavelength selection

Wavelength selection aims at finding and using only the most "informative" wavelengths from the calibration set, rather than using the full spectra. Wavelength selection can result in parsimonious calibration models with greater statistical performance and interpretability (Ng et al., 2019; Vohland et al., 2014). The selected wavelengths should have a good signal/noise ratio, they should be linear, and their spectral variation should be proportional to changes in the soil property of interest (Gemperline, 2006). Overall, wavelength selection is meant to remove uninformative wavelengths, improve model interpretability, and decrease time complexity for analyzing the spectral data (Ng et al., 2019).

Several wavelength selection approaches have been applied in soil spectroscopy. Viscarra Rossel et al. (2008) used the variable importance for projection (VIP) of Wold et al. (2001) coupled with PLSR coefficients to select MIR wavelengths for the prediction of SOC. These authors found that important wavelengths for SOC include those related to O-H and N-H bond stretching vibrations (~3,400 cm^{-1}); alkyl–CH_2 asymmetric and symmetric stretches (~2,930–2,850 cm^{-1}); carboxylic acid and ketones (~1,725 cm^{-1}); amides, aromatics, aliphatic acids, and alkyl groups of soil organic material (1,600–1,400 cm^{-1}); and those related to carbohydrates and sugars (~1,100 cm^{-1}). Vohland et al. (2011) coupled PLSR with feature selection based on a genetic algorithm (GA-PLSR) for the estimation of various carbon fractions and total SOC using VNIR. Genetic algorithms are metaheuristic solutions to optimization problems that have been widely applied in chemometrics (see Jouan-Rimbaud et al., 1995; Leardi & Lupiáñez González, 1998). The genetic algorithm used by Vohland et al. (2011) identified two peaks related to water absorption (1,400 nm) and the hydroxyl band (2,200 nm) as prominent features for the estimation of SOC, which was in accordance with other soil spectroscopy studies (e.g., Ben-Dor & Banin, 1995). The SOC was predicted with a PLSR, GA-PLSR, and SVMR model and all approaches resulted in an R^2 of .89 and RPDs of 2.68, 2.82, and 2.77, respectively. Although the GA-PLSR and SVMR predictions had a similar accuracy (RMSE = 0.27%), the authors considered the GA-PLSR model to be more reliable given its slightly better overall performance. In a study to predict SOC in smallholder farms in India using VNIR, Clingensmith et al. (2019) tested the utility of two multivariate variable reduction methods commonly applied in genomics, the sparse partial least squares regression (SPLSR, Chun & Keles, 2010) and the heteroscedastic effects model (HEM, Shen et al., 2014). Overall, the SPLSR (R^2 = .65, bias = −0.02%, RMSE = 0.42%, RPD = 1.69, RPIQ = 2.21) and HEM (R^2 = .63, bias = −0.04%, RMSE = 0.43%, RPD = 1.64, RPIQ = 2.14) models improved predictions over those of PLSR (R^2 = .53, bias = −0.03%, RMSE = 0.48%, RPD = 1.47, RPIQ = 1.92) models and were helpful for model interpretation. Additionally, the authors noted that the HEM and SPLSR algorithms could improve SOC predictions compared with PLSR with calibrations constructed from significantly fewer spectral predictors.

Other wavelength selection approaches include the competitive adaptive reweighted sampling (CARS) technique of Li et al. (2009). The CARS technique builds multiple PLSR models on observations selected randomly (~80–90% of the calibration set) using a Monte Carlo strategy. Wavelengths of relatively small PLSR coefficients are then removed by applying an exponential decreasing function (EDF). Subsequently, weights are calculated for each remaining wavelength according to the PLSR coefficients and adaptive reweighted sampling is conducted to further eliminate wavelengths in a competitive manner. Vohland et al. (2014) applied the CARS technique to build calibration models for the estimation of SOC using VNIR and MIR data and compared the results of cross-validation. The CARS-PLSR model was significantly more accurate than the full-spectrum PLSR model for both spectral ranges (CARS-PLSR: R^2 = .74 and .91; RPD = 1.98 and 3.37, RMSE = 0.16 and 0.1%; Full-spectrum: 0.60 and 0.78, 1.58 and 2.12, and 0.21 and 0.15% for VNIR and MIR, respectively). These authors suggested that CARS selects wavelengths that are physically reasonable in a parsimonious and statistically accurate way.

In accordance with Teófilo et al. (2009), Sarathjith et al. (2016) conducted an ordered prediction selection (OPS) coupled with an EDF and variable indicators (e.g., VIP) to estimate SOC using VNIR spectra. The variable indicator-based OPS approach followed by these authors successfully found those meaningful wavelength regions for the estimation of SOC. The regions identified include those related to the first overtone of O-H stretches (~1,400–1,900 nm), and combination of the metal–OH bend associated with clay minerals (Clark, 1999; Viscarra Rossel et al., 2006; Vohland et al., 2014). According to the authors, the OPS-PLSR improved the prediction accuracy of SOC as compared to the full spectrum approach but only slightly (Full-spectrum model for Alfisols: R^2 = .56, RPD = 1.53 and RMSE = 0.08%; OPS-PLSR for Alfisols: 0.57, 1.54, and 0.08%). Ludwig et al. (2021) investigated the effects of SOC% range, sample size, and wavenumber region selection on the RMSE and RPIQ. They used an automatic method to select optimal models from more than 17,800 combinations of nine spectral regions between 7,000 and 1,030 cm^{-1} (MIR and long-range NIR) and spectral preprocessing treatments. The regions included peaks between 6,250 and 5,888 cm^{-1}, 5,556 cm^{-1}, 5,000 cm^{-1}, and between 4,167 and 4,545 cm^{-1}, which are associated with organic matter. Other regions considered were those between 3,500 and 3,000 cm^{-1} (related to OH in water and O-H, N-H, and C-H bond stretching), 3,021 to 2,359 cm^{-1} (aliphatic CH stretching), 2,359 to 1,694 (vibrations of carboxylic groups), and 1,694 to 1,030 cm^{-1} (amides, associated water, carboxylate, and aromatic groups). All nine regions were used in at least one optimal model for SOC% indicating the wide range of useful information for the estimation of SOC% within the MIR to long-range NIR spectral region. The authors found that spectral pretreatment and wavenumber selection greatly improved the accuracy of SOC% estimates of PLSR models fitted with fewer observations (n = 71: RPIQ from 3.6 ± 0.3 to 5.4 ± 1.0 and n = 119: RPIQ from 3.9 ± 0.7 to 5.9 ± 0.8), but there was no overall benefit of these techniques for PLSR models fitted with more observations (n = 144 and n = 263). The authors determined that model performance was related to the calibration set variability, which had opposite effects on the RMSE and RPIQ. Lower RMSEs were associated with more homogeneous calibration models and higher RMSEs with more heterogeneous models; however, as Clingensmith et al. (2019) found, more heterogeneous models also had a wider IQR resulting in higher RPIQs. The authors cautioned that RPIQ and RMSE values should not be interpreted independently in infrared studies, but rather in the context of their associated IQR values (Ludwig et al., 2021).

3 | LIBRARY TRANSFER

There are several efforts around the world for the collection of soil spectral data and the application of this data for the assessment of soil carbon (see global: Brown et al., 2006; Viscarra Rossel et al., 2016; national: Dangal et al., 2019; Nocita et al., 2014; Wijewardane et al., 2018; regional: Demattê et al., 2016; Terra et al., 2015; Vasques et al., 2010; local: Dotto et al., 2018; Guerrero et al., 2016; Lucà et al., 2017; Moura-Bueno et al., 2019; Sanderman et al., 2021). A major reason for the construction and maintenance of a SSL is its utility for building calibration models. Currently, there is widespread interest in the development of SSLs; however, there is debate as to what scale is most useful for developing accurate calibrations. In this context, a *global* SSL refers to a dataset containing observations (i.e., soil analyte data and associated spectra) from around the world, including multiple continents. A local SSL is a field-scale dataset. A regional SSL has a greater geographic extent than a local library and its observations are typically limited to a physiographic or similar region (Brown et al., 2006; Sankey et al., 2008). A regional or global SSL will typically contain a large number of observations that represent heterogeneous soil types and properties, allowing for the construction of large calibration models. The large number of observations may improve a calibration model's ability to accurately predict soil properties across several geographic extents as compared to a calibration model developed from a local SSL; however, the large size of a calibration model does not guarantee good model performance at a local site because soil variability is not constant across sites. Moreover, a regional or global SSL may fail to adequately capture the site-specific variability (Brown et al., 2006; Guerrero et al., 2014; Lobsey et al., 2017; Shepherd & Walsh, 2002). An important consideration when comparing the performance of spectroscopic models developed from

TABLE 2 Summary of library transfer studies and whether library transfer resulted in an accurate prediction of the target soil carbon concentration

General SSL	Target area	Accurate prediction of target area soil C[a]	Spectral Range	Reference
Farm	Farm	Yes	MIR	Reeves et al. (2001)
Regional	State	Yes	MIR	McCarty et al. (2002)
Global	Global	No	VNIR	Brown et al. (2006)
Regional	State	No	MIR	Minasny et al. (2009)
Regional	Farm	Yes	VNIR	Kuang & Mouazen (2011)
National	Farm	Yes	VNIR	Peng et al. (2013)
National	Regional	No	VNIR	Gogé et al. (2014)
National	National	Yes	VNIR	Gomez et al. (2020)
National	National	Yes	MIR	Briedis et al. (2020)
National	Continental	Yes	MIR	Dangal & Sanderman (2020)
National	Farm	Yes	MIR	Sanderman et al. (2021)

Note. MIR, mid-infrared; SSL, soil spectral library; VNIR, visible and near-infrared.

[a]Accurate prediction determined based on a correlation coefficient ≥0.8 or a ratio of performance to deviation ≥2.0.

regional and global spectral libraries to site-specific models, is that the former typically contain observations with a wide range of analyte values, resulting in models that can lead to high prediction errors (Stenberg et al., 2010). Therefore, in addition to the prediction error, an objective evaluation and comparison of model performance also requires metrics like R^2, RPD, and RPIQ.

Several studies have investigated the success of using a SSL developed for one area to construct calibration models for a different area. Table 2 summarizes some of these studies. The application of an existing (i.e., general) SSL to a new area (i.e., target area) is often referred to as library transfer. Transferring a general SSL to a target area can result in accurate predictions if the observations in the SSL represent similar pedodiversity to that of the target area (Gogé et al., 2014; Janik et al., 2007; Wetterlind & Stenberg, 2010). Similar pedodiversity leads to greater mineralogical and chemical similarity between the calibration observations from the existing SSL and the unknowns from the new area, which results in greater model performance (Guerrero et al., 2014; Stenberg et al., 2010).

Reeves et al. (2001) performed library transfer of a local MIR SSL containing 180 observations from two fields in Maryland. The authors used a PLSR model constructed using observations from one field to predict total organic carbon for the other field. In both fields, the constructed calibration model resulted in accurate predictions (Reeves et al., 2001). Shepherd and Walsh (2002) used a VNIR SSL with more than 1,000 observations from one region of Africa to predict SOC across a different region, also in Africa. They obtained accurate calibrations using multiplicative adaptive regression splines (MARS) (Shepherd & Walsh, 2002). McCarty et al. (2002) compared the prediction of two PLSR models constructed using a MIR SSL with observations from eight states in the United States. One PLSR model was constructed using 257 observations from the general SSL to predict SOC for 16 unknowns from a new state. The other PLSR model was constructed using 177 observations from the general SSL to predict 60 randomly selected observations from the same SSL. The authors obtained slightly higher R^2 values (.98 vs. .94), but also a higher prediction error (0.60 vs. 0.32%) with the first model than with the second model (McCarty et al., 2002).

Minasny et al. (2009) tested the applicability of three state-wide calibration models developed from a regional Australian MIR SSL to predict soil carbon. Each state-wide model was used for prediction in the other two states. They determined that their calibration models were state-specific and nontransferable, as evidenced by the high prediction errors (mean of absolute error: 0.85 to 0.35%). These authors also created a single model by combining observations from all three states and used it to predict a subset of observations. They found that the state models (R^2: .79–.92, mean of absolute error: 0.29–0.24%) outperformed the combined model (R^2 = .74, mean of absolute error: 0.36%) (Minasny et al., 2009). Kuang and Mouazen (2011) constructed VNIR calibration models for three farms in Europe. They used a farm-specific SSL to construct calibration models for the prediction of SOC% across each farm. Additionally, they compiled observations from the three farm-specific SSLs to construct a single calibration model to predict SOC at each farm. The model developed from the combined SSL resulted in predictions with larger R^2 and RPD values, but also larger RMSE values than two of the three farm-specific calibration models (combined model: n = 408, R^2 = .83, RPD = 2.49, RMSE = 0.54%; farm-specific1: n = 205, R^2 = .12, RPD = 1.07, RMSE = 0.19%; farm-specific2: n = 128, R^2 = .75, RPD = 2.00, RMSE = 0.30%). The authors

attributed these results to SOC ranges being wider in the combined SSL than in the farm-specific SSLs. The farm-specific model constructed with the smallest number of observations ($n = 70$), resulted in the largest R^2 and RPD and largest RMSE (R^2 = .96, RPD = 4.95, RMSE = 0.62%). Gogé et al. (2014) constructed a calibration model from a national VNIR SSL to predict SOC for a small region in France. The national SSL contained observations from the small region; however, the region was under-represented. The resulting model did not accurately predict SOC of the small region (RPD < 1.4, RMSE = 0.733%, bias > −5.0%) (Gogé et al., 2014).

Using the same French national SSL as Gogé et al. (2014), Gomez et al. (2020) constructed a PLSR model to predict SOC of observations from Tunisia. Additionally, the authors constructed a PLSR model using only the spectral neighbors (subsetting by spectral similarity) of the French national SSL to the Tunisian observations. These two PLSR models also included a variation consisting of log-transformed SOC values, which resulted in a total of four PLSR models. For the full-set models, the log-transformed model (R^2 = .90, RMSE = 0.66%, bias = −0.01%, RPD = 2.9, RPIQ = 2.6) outperformed the untransformed model (R^2 = .88, RMSE = 0.72%, bias = −0.04%, RPD = 2.7, RPIQ = 2.4). Similarly, the authors found that for the spectral neighbors models, the log-transformed model (R^2 = .92, RMSE = 0.57%, bias = −0.01%, RPD = 3.4, RPIQ = 3.0) outperformed the untransformed model (R^2 = .93, RMSE = 0.54%, bias = −0.07%, RPD = 3.6, RPIQ = 3.2). Finally, the log-transformed, spectral neighbors model outperformed the log-transformed, full-set model (R^2: .92 vs. .90, RMSE: 0.57 vs. 0.66%, bias: −0.01 vs. 0.01%, RPD: 3.4 vs. 2.9, RPIQ: 3.0 vs. 2.6, respectively). The authors concluded that regardless of the model (full-set or spectral neighbors) using log-transformed SOC data improved the predictions. Briedis et al. (2020) compared the performance of three calibration models constructed from a national Australian SSL (n = 567) to a PLSR model constructed from a national Brazilian library (n = 402) to predict SOC of Brazilian soil samples. These authors tested PLSR, SBL, and Cubist calibration models. The PLSR model constructed from the Brazilian SSL (RPIQ = 5.86) outperformed all the calibration models constructed from the Australian SSL (average RPIQ = 2.96).

Dangal and Sanderman (2020) tested whether a PLSR, MBL, and Cubist calibration model constructed from an American SSL (n > 55,000) could predict, among other sets, a European dataset of 596 observations. Using calibration models of spectra preprocessed with a baseline offset transformation, all three models achieved a good fit according to the R^2 (> .85), RPIQ (0.72 –0.81), and RMSE (2.8–3.15%). Additionally, the best prediction was achieved by the Cubist model (R^2 = .95, RMSE = 2.80%, RPIQ = 0.81, and bias = −0.72%). Sanderman et al. (2021) performed a study to determine whether changes in SOC concentration due to management could be detected through MIR spectroscopy. They used an American SSL (n > 80,000) and MBL to predict values for seven long-term research field sites in the United States (smallest n = 28, largest n = 390) and consequently determine whether the changes in SOC detected through conventional laboratory analysis were also detected by spectroscopic analysis. The calibration model constructed from the national SSL was able to predict SOC values for most sites very well (R^2: .70–.94, RPD: 1.82–3.55, RMSE: 0.10–0.33%, and bias: 0.08–0.38%) with the lower performance of some sites likely due to a narrower range in SOC%. On average, results of their ensemble machine learning with MBL predictions were significantly lower than the observed SOC values (1.14 vs. 1.37%). Nonetheless, the spectroscopic models were able to detect changes in SOC similar enough to those measured through conventional analysis in five of the seven sites and reach the same conclusions on the effect of agricultural management on SOC concentration. The authors concluded that existing large MIR SSLs can be used by other laboratories for the purpose of carbon monitoring.

Different techniques have been proposed to optimize library transfer of general SSLs and thus, improve the prediction accuracy of calibration models constructed from them. Optimization techniques, such as adjusting the number of observations and subsetting, can be applied to calibration models for the purpose of library transfer. Additionally, incorporating target area observations into the calibration model can improve model performance and thus, benefit more from general library transfer for site-specific modeling (Barthès et al., 2020; Brown, 2007; Lobsey et al., 2017; Sankey et al., 2008; Shepherd & Walsh, 2002; Sila et al., 2016; Wetterlind & Stenberg, 2010; Wijewardane et al., 2018).

Adding observations from the target area to a calibration model constructed from a general SSL to predict new observations from the target area is referred to as spiking. Spiking involves three general steps: (a) soil samples from the target area are analyzed, using the same laboratory methods as the observations in the calibration set and their observations are recorded; (b) these target area observations (i.e., spiking set) are added to the initial calibration set; and (c) the calibration model is "recalibrated" (Guerrero et al., 2014). A variation of spiking involves the replication of observations in the spiking set, which is referred to as spiking with extra weighting. This technique involves adding multiple copies of the target area observations to the initial calibration set in order to increase the leverage of the target area observations in the calibration (Guerrero et al., 2014). Spiking can be performed in combination with any of the optimization techniques previously discussed. For example, a spiking set can be selected based on its analyte value, pedogenic, or spectral similarity to the target area set of unknowns, thus performing a subsetting-spiking routine. Likewise, the number or proportion of the spiking

TABLE 3 Summary of library transfer studies that use spiking, spiking with extra weighting, and spiking and subsetting for the construction of calibrations to predict soil carbon and whether at least one of these techniques resulted in decreased prediction error

Criteria	General SSL	Target area	Decreased prediction error compared to calibration from general SSL	Spectral range	Reference
Spiking	Global	Watershed	Yes	VNIR	Brown (2007)
Spiking	Global	U.S. state	Yes	VNIR	Sankey et al. (2008)
Spiking	National	Farm	Yes	VNIR	Peng et al. (2013)
Spiking	National	Watershed	Yes	VNIR	Gogé et al. (2014)
Spiking	National to farm	Farm to small region	Yes	NIR	Guerrero et al. (2016)
Spiking	Farm	Continental	Yes	VNIR	Nawar & Mouazen (2017)
Spiking + weighting	Global	U.S. state	Yes	VNIR	Sankey et al. (2008)
Spiking + weighting	National	Farm to small region	Yes	NIR	Guerrero et al. (2014)
Spiking + weighting	National to farm	Farm to small region	Yes	NIR	Guerrero et al. (2016)
Spiking + weighting	Global	Farm	Yes	VNIR	Lobsey et al. (2017)
Spiking + subsetting	National	Farm	Yes	NIR	Wetterlind & Stenberg (2010)
Spiking + subsetting	National	Farm to small region	Yes	NIR	Guerrero et al. (2014)
Spiking + subsetting	Global	Farm	Yes	VNIR	Lobsey et al. (2017)
Spiking + subsetting	National and regional	Small region	Yes	MIR	Briedis et al. (2020)
Spiking + subsetting	Large region	Small region	Yes	VNIR	Ng et al. (2022)
Spiking + subsetting + weighting	National	Small region	Yes	MIR	Barthès et al. (2020)

Note. MIR, mid-infrared; NIR, near-infrared; SSL, soil spectral library; VNIR, visible and near-infrared.

set can be varied, thus resulting in a calibration size-spiking approach.

In general, when performing spiking, only a relatively small number of target area observations are included in the spiking set for the calibration model. This ensures that the model contains observations representative of those that it will predict (Nocita et al., 2015). However, as with a typical calibration, the number of target area observations included in the spiking set can be adjusted to optimize model performance. Typically, the larger the spiking set, the greater the prediction accuracy of the spiked calibration model. However, a larger number of spiking observations implies a greater cost of analysis, which decreases the low-cost advantage of soil spectroscopy for soil analysis (Guerrero et al., 2014).

Table 3 summarizes some soil spectroscopy studies that have used spiking and a combination of spiking and subsetting techniques for library transfer. The example studies presented in Table 3 are not to be considered an exhaustive representation of the literature on library transfer optimization techniques. However, they are representative of techniques discussed in this paper and of the diversity of techniques used in recent studies.

McCarty and Reeves (2000) were some of the first to suggest that inclusion of only a few observations from the target area in the calibration set might improve model performance. Similarly, Brown et al. (2006) hypothesized that spiking could improve the effectiveness of library transfer. Moreover, they also hypothesized that spiking for library transfer could result in more accurate predictions than using only observations from the target area. These hypotheses were supported by the work of Brown (2007), who predicted SOC concentration for a Ugandan watershed through library transfer of a global VNIR SSL (n = 3,794) spiked with local observations ($n \leq$ 206). Brown (2007) found that spiking the calibration model constructed from the global SSL with observations from the watershed improved model performance and, in some cases, outperformed a calibration model constructed only from the watershed (i.e., target area) observations (RMSE = 0.53 and 0.59%, respectively, for model with n spiking and n watershed = 206).

Sankey et al. (2008) used the same global SSL as Brown (2007) to compare target area calibration models to global SSL models and global SSL models spiked with up to 234 observations. Using these models, the authors predicted SOC concentration for three sites in Montana. The best model performance for each site (SEP = 0.380, 0.770, and 2.62%) was obtained by the spiked global SSL calibration model. These authors also tested the influence of weighting in the

spiked calibration model by applying lower weight to the global observations than to the target area observations. Overall, this approach slightly improved SOC prediction accuracy as compared to the unweighted, spiked model. The authors suggested that the optimum weight for highest prediction accuracy depends on the variability of the target area and the soil property (Sankey et al., 2008).

Wetterlind and Stenberg (2010) compared the performance of several small, farm-level calibration models (n = 25) with those constructed from a national Swedish NIR SSL (n = 396) for the prediction of SOC. The national SSL models consisted of a full-set calibration and a spectral neighbors model (n = 50). Additionally, both full-set and spectral neighbors models were also tested in their spiked variant (spiked with ≤25 farm observations). The spectral neighbors model did not outperform the full-set model. The spiked variants of the full-set and spectral neighbors models outperformed their nonspiked counterparts. Moreover, both spiked variants resulted in comparable prediction accuracy to that of the farm-specific calibration models. Additionally, the spiked variant of the spectral neighbors model outperformed the spiked variant of the full-set model. They attributed these findings to the ability of the spectral neighbors model to integrate the target area observations more easily due to its smaller size, as compared to the full-set model.

Peng et al. (2013) compared the performance of calibration models constructed from a national Danish VNIR SSL (n = 2,688) to predict SOC for a field in Denmark. These authors constructed calibration models using subsets of the national SSL based on observations that were geographically closest (n = 84), pedologically most similar (n = 96), and spectrally most like those of the target area (n = 100). Additionally, they spiked the national SSL with a random set of 30 observations from the target area (n = 2,718). The best predictions on the target area unknowns were from the geographically closest subset as well as the spiked national calibration models (each with RMSE = 0.19% and RPD = 3.7). Additionally, the spiked calibration outperformed the full-set national SSL (RMSE = 0.19 and 0.22%, respectively) (Peng et al., 2013). Gogé et al. (2014) constructed a calibration model using a French national VNIR SSL (n = 2,126) to predict SOC for a watershed in France. Moreover, the authors tested the spiked version of this model with a spiking set ranging from 10 to 94 observations. Spiking the calibration model decreased the RMSE and increased the R^2 for SOC concentration as compared to the nonspiked calibration model (RMSE = 0.733%) with the lowest error achieved by the spiked calibration model with the largest spiking set (RMSE = 0.579%).

Guerrero et al. (2014) used a national SSL from Spain to construct calibration models for the prediction of SOC across sites in Spain, the United Kingdom, and Sweden. These authors tested the effect of spiking the initial calibration models with extra weighting. These authors also evaluated 13 different subsetting strategies to select the spiking set, as well as the effect of different numbers of observations used to construct the calibration models from the national SSL. Results of this study indicated that spiking improved the prediction accuracy of all models. Moreover, differences in performance of the spiked models were due to the subsetting approach used to select the spiking set. The best predictions were achieved when the spiking set was selected according to spectral neighbors. The accuracy of the predictions was further improved by extra weighting of the spiking set. Moreover, smaller spiked calibration models outperformed larger spiked models.

Guerrero et al. (2016) constructed calibration models from eight national, regional, and local SSLs from Spain and Sweden to predict SOC concentration for 10 sites in Spain and one in the United Kingdom. These authors observed that the fewer the observations used to construct the initial calibration models, the greater the effect of spiking. That is, there is an inverse relationship between the calibration size and the effect of spiking on model performance. These results are in accordance with those of Guerrero et al. (2014). Furthermore, the fewer the observations for the initial calibration model, the smaller the effect of spiking with extra weighting. Overall, the highest prediction accuracy resulted from calibration models with extra weighting. These authors explained that small SSLs can be just as effective in yielding high prediction accuracy through spiking with extra weighting, and thus, large SSLs are not needed for local assessment of SOC concentration (Guerrero et al., 2016).

Lobsey et al. (2017) combined spectral subsetting with spiking to improve the statistical performance of small calibration models for SOC concentration of two sites in Australia and New Zealand. These authors selected a subset of representative observations from the target area to spike a calibration model developed from a global VNIR SSL (n = 17,928). Results of this study showed that spiking the global SSL with as few as 20 target area observations was sufficient to yield an accurate prediction of SOC concentration at both sites (RMSE = 0.48 and 1.16%). The spiked calibration models performed as well or better than those containing only target area observations (n ≤ 300) (Lobsey et al., 2017). In a study by Briedis et al. (2020), using a national Australian SSL (n = 567) spiked with as few as 20 target area observations (8% of Brazilian regional SSL), improved the prediction accuracy of total OC over using only the Australian SSL and local-type models calibrated with spectrally similar observations. The highest prediction accuracy achieved was using the full, target area calibration model (RMSE = 0.317% and RPIQ = 5.86). Moreover, the spiked Australian SSL model performed similarly to a model constructed using only the spiking set of 20 target-area observations (RPIQ = 4.74 and 4.49, respectively). The authors concluded that a proper selection of a small, spectrally similar calibration set can result in

<table>
<tr><th colspan="4">Calibration Optimization Techniques for General Use</th></tr>
<tr><th colspan="2">Subsetting by Spectral Similarity</th><th>Subsetting by Pedodiversity</th><th>Subsetting by Analyte Value</th></tr>
<tr><td colspan="2">• High spectral variability in SSL
• Knowledge of variability expected in spectrally active properties and SOC concentration of unknowns</td><td rowspan="2">• Soil information systems available at appropriate scale
• Knowledge of variability in chemical/physical properties (e.g., mineralogy, sand content) and their relation to SOC</td><td rowspan="2">• High variance/wide range in SOC concentration</td></tr>
<tr><td>Local Modeling</td><td>Wavelength Selection</td></tr>
<tr><td>• Locally stable spectra-analyte relationship
• Balanced number of observations across SOC concentration range</td><td>• Interest in identifying informative spectral regions
• Interest in qualitative spectral analysis</td><td></td><td></td></tr>
<tr><th colspan="4">Calibration Optimization Techniques for Library Transfer</th></tr>
<tr><th colspan="2">Subsetting</th><th colspan="2">Spiking</th></tr>
<tr><th>Subsetting-Spiking</th><th>Subsetting Only</th><th>Spiking with Weighting</th><th>Spiking Only</th></tr>
<tr><td>• Target area observations available for calibration
• Spectral similarity of general SSL subset to target area
• Conditions met for subsetting techniques above</td><td>• Target area observations not available for calibration
• Conditions met for subsetting techniques above</td><td>• Few target area observations available for calibration in relation to general SSL
• Use of variable weights requires understanding of spectral variability and pedodiversity of target area</td><td>• < 30 target area observations available for calibration
• n of general SSL slightly greater than n of target area observations</td></tr>
</table>

FIGURE 3 General decision chart for the selection of optimization techniques for spectroscopic modeling of soil organic carbon (SOC) concentration. SSL, soil spectral library

accurate and cost-effective OC prediction using MIR (Briedis et al., 2020).

Barthès et al. (2020) used a French national MIR SSL to predict soil inorganic carbon (SIC) in a region of France. The authors used the SBL algorithm to select spectral neighbors and performed spiking with extra weighting to construct a calibration model. Using only observations from the national SSL yielded an accurate prediction (SEP = 0.5%). Nevertheless, the prediction accuracy was improved through spiking with 10 observations, extra-weighted 40 times (SEP = 0.33%). The calibration model constructed using only local target area observations yielded less accurate results than the spiked calibration (SEP = 0.36%). In a more recent study, Ng et al. (2022) compared the effectiveness of spiking and subsetting (MBL and a localized PLSR) for the prediction of SOC in small regions of Australia using a large regional VNIR SSL (n = 1,867). The localized PLSR models, constructed with ≥20 observations (n = 20; RPIQ: 0.23–0.71, RMSE: 0.38–1.07%, bias: −0.11 to −0.01%), outperformed the target area (n = 20; RPIQ: 0.23–0.67, RMSE: 0.36–1.31%, bias: −0.17 to −0.00%) and spiked regional models (n = 20; RPIQ: 0.19–0.63, RMSE: 0.32–1.41%, bias: −0.77 to −0.02%). The authors concluded that spiking is dependent on the spectral similarity between the general SSL and the target area observations. These authors also concluded that calibration models created through spiking were overall, not better than models constructed using only target area observations (Ng et al., 2022).

The studies described demonstrate that various factors influence the effectiveness of calibration optimization techniques. The success of calibration optimization to improve prediction accuracy depends on SOC concentration range, the sample selection scheme used to build the calibration set, the modeling approach, and the spectral variability related to the pedodiversity of the calibration set. Additionally, the effectiveness of optimization techniques is influenced by the size of the SSL available for calibration. When constructing a calibration model for library transfer, optimization can be performed through subsetting, spiking, or a combination of both; however, considerations for the proportion of representative observations in the calibration set must be made. Figure 3 provides a generalized decision chart for the appropriate optimization technique. The chart presents conditions and factors required for successful optimization using the techniques discussed. The reader should note that the general guidance provided here is based on studies presented in this work and that it may be necessary to consider other conditions before selecting a technique.

4 | CONCLUSIONS

The analysis of soil carbon through soil spectroscopy benefits from optimization procedures to improve the statistical performance of calibration models. The approaches for model optimization discussed in this work included the selection of calibration set size, the creation of targeted calibration models through subsetting, and spiking.

Calibration set size influences model performance and has implications for the cost-savings potential of soil

spectroscopy. Obtaining a large SSL is not always an option as studies may have limited resources for data collection and analysis, making it crucial to consider strategies that allow for a reduced number of observations without a decrease in model performance. In general, model performance improves with increasing calibration size, until it stabilizes and there is no significant improvement with additional observations. The optimal calibration size depends on the initial calibration set size and the sampling scheme used to select the calibration set. The reduction in prediction error by the addition of observations diminishes as the initial calibration size increases. That is, the added benefit of new observations is greater for smaller calibrations than larger ones. If affordability and computational efficiency are considered, starting with a smaller calibration set of at least 30 observations can be much more efficient than starting with a large set and may yield equally good results.

In scenarios where soil spectral libraries already exist, it can be useful to identify the best technique for selecting the calibration set. If the spectral library is homogeneous in terms of its spectral variability, then random sampling can perform as well as stratified sampling. However, when the spectral variability is large, spectrally stratified sampling generally improves model performance. The spectrally stratified sampling approach has a greater influence on model performance when the models are small, so it is worthwhile to combine this optimization technique with an approach to define an optimal calibration size.

Reducing the range of variability in analyte concentrations can improve model performance. Subsetting a calibration set by analyte value is an effective optimization technique when the spectral variability is low. Therefore, subsetting by analyte value should be avoided in SSLs derived from soil samples with highly diverse spectrally active physical and chemical properties. Additionally, statistical dispersion is known to influence model performance, with a smaller dispersion (i.e., narrower data range) resulting in a reduction in RMSE. Therefore, it is critical that authors used and present suitable metrics of statistical performance when comparing across models with calibration data of varying range (e.g., R^2, RPD, RPIQ).

If the spectral variability in the SSL is expected to be large, due to diverse mineralogy, large spatial extent, or other factors known to influence the analyte being assessed, then subsetting to reduce this variability within calibration sets can lead to better model performance. In these scenarios, utilizing criteria based on soil-forming factors that influence mineralogical properties, may be the most effective technique to improve model performance. The criteria used in these cases, should reduce within-subset and increase across-subset spectral and analyte variability. Subsets based on a single criterion, such as taxonomic soil order or horizonation, can contain spectral features that are not mutually exclusive; therefore, a multi-criteria approach can be more useful.

Subsetting by spectral similarity to the prediction set (i.e., through local modeling) is another effective technique for calibration optimization; however, as with soil-related criteria, it should be avoided in datasets with low spectral variability. Wavelength selection can result in parsimonious calibration models with better model performance and interpretability than the full-set models. Moreover, investigations on wavelength selection methods can guide the development of new spectroscopic instruments. The effectiveness of subsetting for improving model performance depends on the modeling approach. Utilizing a machine or deep learning, which can handle complex relationships in high-dimensional space, is generally as or more effective in improving model performance as compared to subsetting by analyte value, pedodiversity, or spectral similarity.

The capacity of an existing SSL to perform well in a new target area, depends on the spectral and analyte similarity to the target area unknowns. In library transfer, similar pedodiversity leads to greater mineralogical and chemical similarity, which in turn leads to greater spectral and analyte similarity between the calibration observations from the existing SSL and the target unknowns; thus, improving the statistical performance of the calibration models. Spiking can be performed in addition to or in combination with any of the other optimization techniques to improve model performance for library transfer. Spiking with representative target area observations improves model performance. Typically, the prediction accuracy of the spiked calibration model increases as the size or proportion of the spiking set increases, because a larger proportion of spiking observations results in greater representativeness of the target unknowns. Spiking is most useful in scenarios where target area SSLs are too small ($n < 30$) to produce accurate predictions. In these scenarios, using a spiked general SSL calibration model, outperforms the target area model. If target area observations are limited, spiking with extra weighting is a cost-effective method to improve model performance. Spiking with extra weighting reduces the need to add/collect new target observations because it duplicates existing target-area observations. Spiking with subsetting is most effective when using a criterion that best separates spectrally active features related to the soil property being predicted; thus, it is important to couple subsetting by spectral similarity with spiking, particularly when the SSL to be transferred is spectrally different from the target area.

Optimization techniques can further improve the efficiency and reduce the cost of soil spectroscopy for soil carbon analysis and should be studied further. These techniques are useful for improving the model performance of calibrations constructed from both small and large SSLs. In cases where a large SSL already exists, optimization techniques represent a cost-effective solution to improve the effectiveness of library transfer. In areas where SSLs are rare or absent, optimization

techniques can support new data collection efforts as well as the construction of more parsimonious calibration models.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support and constructive feedback provided by the editor and reviewers, which improved this manuscript. Funding for publication was provided by the University of Arkansas Graduate Professional Student Congress Tiffany Marcantonio Research Grant and the University of Arkansas Open Access Publishing Fund.

AUTHOR CONTRIBUTIONS

Minerva J. Dorantes: Conceptualization; Data curation; Formal analysis; Funding acquisition; Visualization; Writing – original draft; Writing – review & editing. Bryan A. Fuentes: Visualization; Writing – review & editing. David M. Miller: Writing – review & editing. All authors contributed to and approved the final version of this manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Minerva J. Dorantes https://orcid.org/0000-0002-2877-832X

Bryan A. Fuentes https://orcid.org/0000-0003-3506-7101

REFERENCES

Angelopoulou, T., Balafoutis, A., Zalidis, G., & Bochtis, D. (2020). From laboratory to proximal sensing spectroscopy for soil organic carbon estimation—A review. *Sustainability*, *12*(2), 443. https://doi.org/10.3390/su12020443

Baldock, J. A., Hawke, B., Sanderman, J., & Macdonald, L. M. (2013). Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. *Soil Research*, *51*(8), 577–583. https://doi.org/10.1071/SR13077

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. *Trends in Analytical Chemistry*, *135*, 116166. https://doi.org/10.1016/j.trac.2020.116166

Barthès, B. G., Kouakoua, E., Coll, P., Clairotte, M., Moulin, P., Saby, N. P. A., Le Cadre, E., Etayo, A., & Chevallier, T. (2020). Improvement in spectral library-based quantification of soil properties using representative spiking and local calibration – The case of soil inorganic carbon prediction by mid-infrared spectroscopy. *Geoderma*, *369*, 114272. https://doi.org/10.1016/j.geoderma.2020.114272

Bellon-Maurel, V., & McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. *Soil Biology and Biochemistry*, *43*(7), 1398–1410. https://doi.org/10.1016/j.soilbio.2011.02.019

Ben-Dor, E., & 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. https://doi.org/10.2136/sssaj1995.03615995005900020014x

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Briedis, C., Baldock, J., de Moraes Sá, J. C., dos Santos, J. B., & 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. https://doi.org/10.1016/j.geoderma.2020.114401

Brown, D. J. (2007). Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma*, *140*(4), 444–453. https://doi.org/10.1016/j.geoderma.2007.04.021

Brown, D. J., Bricklemyer, R. S., & Miller, P. R. (2005). Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma*, *129*(3–4), 251–267. https://doi.org/10.1016/j.geoderma.2005.01.001

Brown, D. J., Shepherd, K. D., Walsh, M. G., Dewayne Mays, M., & Reinsch, T. G. (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, *132*(3–4), 273–290. https://doi.org/10.1016/j.geoderma.2005.04.025

Brunet, D., Barthès, B. G., Chotte, J.-L., & Feller, C. (2007). Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. *Geoderma*, *139*, 106–117. https://doi.org/10.1016/j.geoderma.2007.01.007

Chang, C.-W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy–Principal components regression analyses of soil properties. *Soil Science Society of America Journal*, *65*(2), 480–490. https://doi.org/10.2136/sssaj2001.652480x

Christy, C. D., & Dyer, S. A. (2006). Estimation of soil properties using a combination of spectral and scalar sensor data. In *2006 IEEE Instrumentation and Measurement Technology Conference Proceedings* (pp. 729–734). IEEE. https://doi.org/10.1109/IMTC.2006.328147

Chun, H., & Keles, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(1), 3–25. https://doi.org/10.1111/j.1467-9868.2009.00723.x

Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N. P. A., Bernoux, M., & Barthès, B. G. (2016). National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. *Geoderma*, *276*, 41–52. https://doi.org/10.1016/j.geoderma.2016.04.021

Clark, R. N. (1999). Spectroscopy of rocks and minerals and principles of spectroscopy. In A. N. Rencz (Ed.), *Remote sensing for the earth sciences* (pp. 3–58). John Wiley & Sons.

Clingensmith, C. M., Grunwald, S., & Wani, S. P. (2019). Evaluation of calibration subsetting and new chemometric methods on the spectral prediction of key soil properties in a data-limited environment: Evaluation of subsetting and new chemometric methods. *European Journal of Soil Science*, *70*(1), 107–126. https://doi.org/10.1111/ejss.12753

Comstock, J. P., Sherpa, S. R., Ferguson, R., Bailey, S., Beem-Miller, J. P., Lin, F., Lehmann, J., & Wolfe, D. W. (2019). Carbonate determination in soils by mid-IR spectroscopy with regional and continental scale models. *PLOS ONE*, *14*(2), e0210235. https://doi.org/10.1371/journal.pone.0210235

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 Systems*, *3*(1), 11. https://doi.org/10.3390/soilsystems3010011

Dangal, S. R. 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

Debaene, G., Niedźwiecki, J., Pecio, A., & Żurek, 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. https://doi.org/10.1016/j.geoderma.2013.09.022

de Gruijter, J. J., McBratney, A. B., & Taylor, J. (2010). Sampling for high-resolution soil mapping. In R. A. Viscarra Rossel, A. B. McBratney, & B. Minasny (Eds.), *Proximal soil sensing* (pp. 3–14). Springer.

Demattê, J. A. M., Bellinaso, H., Araújo, S. R., Rizzo, R., & Souza, A. B. (2016). Spectral regionalization of tropical soils in the estimation of soil attributes. *Revista Ciencia Agronomica*, *47*, https://doi.org/10.5935/1806-6690.20160071

Demattê, J. A. M., & da Silva Terra, F. (2014). Spectral pedology: A new perspective on evaluation of soils along pedogenetic alterations. *Geoderma*, *217–218*, 190–200. https://doi.org/10.1016/j.geoderma.2013.11.012

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

FAO. (1998). World reference base for soil resources. *World Soil Resources Reports*, 84. FAO.

Fernández Pierna, J. A., & Dardenne, P. (2008). Soil parameter quantification by NIRS as a Chemometric challenge at 'Chimiométrie 2006'. *Chemometrics and Intelligent Laboratory Systems*, *91*(1), 94–98. https://doi.org/10.1016/j.chemolab.2007.06.007

Gemperline, P. (2006). *Practical guide to chemometrics*. CRC Press.

Genot, V., Colinet, G., Bock, L., Vanvyve, D., Reusen, Y., & Dardenne, P. (2011). Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. *Journal of Near Infrared Spectroscopy*, *19*(2), 117–138. https://doi.org/10.1255/jnirs.923

Gholizadeh, A., Borůvka, L., Saberioon, M., & Vašát, R. (2013). Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Applied Spectroscopy*, *67*(12), 1349–1362. https://doi.org/10.1366/13-07288

Gogé, F., Gomez, C., Jolivet, C., & Joffre, R. (2014). Which strategy is best to predict soil properties of a local site from a national Vis–NIR database? *Geoderma*, *213*, 1–9. https://doi.org/10.1016/j.geoderma.2013.07.016

Gomez, C., Chevallier, T., Moulin, P., Bouferra, I., Hmaidi, K., Arrouays, D., Jolivet, C., & Barthès, B. G. (2020). Prediction of soil organic and inorganic carbon concentrations in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library. *Geoderma*, *375*, 114469. https://doi.org/10.1016/j.geoderma.2020.114469

Grinand, C., Barthès, B. G., Brunet, D., Kouakoua, E., Arrouays, D., Jolivet, C., Caria, G., & Bernoux, M. (2012). Prediction of soil organic and inorganic carbon contents at a national scale (France) using mid-infrared reflectance spectroscopy (MIRS). *European Journal of Soil Science*, *63*(2), 141–151. https://doi.org/10.1111/j.1365-2389.2012.01429.x

Guerrero, C., Stenberg, B., Wetterlind, J., Viscarra Rossel, R. A., Maestre, F. T., Mouazen, A. M., Zornoza, R., Ruiz-Sinoga, J. D., & 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. *European Journal of Soil Science*, *65*(2), 248–263. https://doi.org/10.1111/ejss.12129

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A. M., Gabarrón-Galeote, M. A., Ruiz-Sinoga, J. D., Zornoza, R., & 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. https://doi.org/10.1016/j.still.2015.07.008

Gupta, A., Vasava, H. B., Das, B. S., & Choubey, A. K. (2018). Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. *Geoderma*, *325*, 59–71. https://doi.org/10.1016/j.geoderma.2018.03.025

Igne, B., Reeves, J. B., McCarty, G., Hively, W. D., Lund, E., & Hurburgh, C. R. (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*(3), 167–176. https://doi.org/10.1255/jnirs.883

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

Janik, L. J., Skjemstad, J. O., Shepherd, K. D., & Spouncer, L. R. (2007). The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Australian Journal of Soil Research*, *45*, 73–81. https://doi.org/10.1071/SR06083

Jouan-Rimbaud, D., Massart, D.-L., Leardi, R., & De Noord, O. E. (1995). Genetic algorithms as a tool for wavelength selection in multivariate calibration. *Analytical Chemistry*, *67*(23), 4295–4301. https://doi.org/10.1021/ac00119a015

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.

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

Kuang, B., & Mouazen, A. M. (2011). Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *European Journal of Soil Science*, *62*(4), 629–636. https://doi.org/10.1111/j.1365-2389.2011.01358.x

Lal, R. (2014). Societal value of soil carbon. *Journal of Soil and Water Conservation*, *69*(6), 186A–192A. https://doi.org/10.2489/jswc.69.6.186A

Leardi, R., & Lupiáñez González, A. (1998). Genetic algorithms applied to feature selection in PLS regression: How and when to use them. *Chemometrics and Intelligent Laboratory Systems*, *41*(2), 195–207. https://doi.org/10.1016/S0169-7439(98)00051-3

Li, H., Liang, Y., Xu, Q., & Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta*, *648*(1), 77–84. https://doi.org/10.1016/j.aca.2009.06.046

Lobsey, C. R., Viscarra Rossel, R. A., Roudier, P., & 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. https://doi.org/10.1111/ejss.12490

Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G., & Buttafuoco, G. (2017). Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. *Geoderma*, *288*, 175–183. https://doi.org/10.1016/j.geoderma.2016.11.015

Ludwig, B., Greenberg, I., Sawallisch, A., & Vohland, M. (2021). Diffuse reflectance infrared spectroscopy estimates for soil properties using multiple partitions: Effects of the range of contents, sample size, and algorithms. *Soil Science Society of America Journal*, *85*(3), 546–559. https://doi.org/10.1002/saj2.20205

Ludwig, B., Murugan, R., Parama, V. R. R., & Vohland, M. (2019). Accuracy of estimating soil properties with mid-infrared spectroscopy: Implications of different chemometric approaches and software packages related to calibration sample size. *Soil Science Society of America Journal*, *83*(5), 1542–1552. https://doi.org/10.2136/sssaj2018.11.0413

Madari, B. E., Reeves, J. B., Coelho, M. R., Machado, P. L. O. A., DePolli, H., Coelho, R. M., Benites, V. M., Souza, L. F., & McCarty, G. W. (2005). Mid- and near-infrared spectroscopic determination of carbon in a diverse set of soils from the Brazilian National Soil Collection. *Spectroscopy Letters*, *38*(6), 721–740. https://doi.org/10.1080/00387010500315876

McBratney, A. B., Hart, G. A., & McGarry, D. (1991). The use of region partitioning to improve the representation of geo statistically mapped soil attributes. *European Journal of Soil Science*, *42*(3), 513–532. https://doi.org/10.1111/j.1365-2389.1991.tb00427.x

McBratney, A. B., Minasny, B., & Viscarra Rossel, R. (2006). Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. *Geoderma*, *136*(1–2), 272–278. https://doi.org/10.1016/j.geoderma.2006.03.051

McCarty, G. W., Iii, J. B. R., Reeves, V. B., Follett, R. F., & Kimble, J. M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal*, *66*, 640–646. https://doi.org/10.2136/sssaj2002.6400a

McCarty, G. W., & Reeves, J. B. III (2000). Development of rapid instrumental methods for measuring soil organic carbon. In J. M. Kimble, R. F. Follett, & B. A. Stewart (Eds.), *Assessment methods for soil carbon* (pp. 371–380). Lewis Publishers.

McCarty, G. W., & Reeves, J. B. (2006). Comparison of near infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. *Soil Science*, *171*(2), 94–102. https://doi.org/10.1097/01.ss.0000187377.84391.54

McDowell, M. L., Bruland, G. L., Deenik, J. L., & Grunwald, S. (2012). Effects of subsetting by carbon content, soil order, and spectral classification on prediction of soil total carbon with diffuse reflectance spectroscopy. *Applied and Environmental Soil Science*, *2012*, 1–14. https://doi.org/10.1155/2012/294121

McDowell, M. L., Bruland, G. L., Deenik, J. L., Grunwald, S., & Knox, N. M. (2012). Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma*, *189–190*, 312–320. https://doi.org/10.1016/j.geoderma.2012.06.009

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245. https://doi.org/10.1080/00401706.1979.10489755

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, *32*(9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

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

Moura-Bueno, J. M., Dalmolin, R. S. D., ten Caten, A., Dotto, A. C., & Demattê, J. A. M. (2019). Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. *Geoderma*, *337*, 565–581. https://doi.org/10.1016/j.geoderma.2018.10.015

Moura-Bueno, J. M., Dalmolin, R. S. D., Horst-Heinen, T. Z., ten Caten, A., Vasques, G. M., Dotto, A. C., & 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. https://doi.org/10.1016/j.scitotenv.2020.139895

Murray, I., Williams, P., & Norris, K. (1987). *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists.

Naes, T., Isaksson, T., & Kowalski, B. (1990). Locally weighted regression and scatter correction for near-infrared reflectance data. *Analytical Chemistry*, *62*(7), 664–673.

Nawar, S., & Mouazen, A. M. (2017). Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *Catena*, *151*, 118–129. https://doi.org/10.1016/j.catena.2016.12.014

Ng, W., Minasny, B., Jones, E., & McBratney, A. (2022). To spike or to localize? Strategies to improve the prediction of local soil properties using regional spectral library. *Geoderma*, *406*, 115501. https://doi.org/10.1016/j.geoderma.2021.115501

Ng, W., Minasny, B., Malone, B. P., Sarathjith, M. C., & Das, B. S. (2019). Optimizing wavelength selection by using informative vectors for parsimonious infrared spectra modelling. *Computers and Electronics in Agriculture*, *158*, 201–210. https://doi.org/10.1016/j.compag.2019.02.003

Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., & Montanarella, L. (2014). Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biology and Biochemistry*, *68*, 337–347. https://doi.org/10.1016/j.soilbio.2013.10.022

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. *Advances in Agronomy*, *132*, 139–159. https://doi.org/10.1016/bs.sgron.2015.02.002

O'Rourke, S. M., & Holden, N. M. (2011). Optical sensing and chemometric analysis of soil organic carbon–A cost effective alternative to conventional laboratory methods? *Soil Use and Management*, *27*(2), 143–155. https://doi.org/10.1111/j.1475-2743.2011.00337.x

Peng, Y., Knadel, M., Gislum, R., Deng, F., Norgaard, T., de Jonge, L. W., Moldrup, P., & Greve, M. H. (2013). Predicting soil organic carbon at field scale using a national soil spectral library. *Journal of Near Infrared Spectroscopy*, *21*(3), 213–222. https://doi.org/10.1255/jnirs.1053

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Rossel, R. A. V., Demattê, J. A. M., & Scholten, T. (2013). Distance and similarity-search metrics for use with soil vis–NIR spectra. *Geoderma*, *199*, 43–53. https://doi.org/10.1016/j.geoderma.2012.08.035

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., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J. A. M., & Scholten, T. (2014). Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma*, *226–227*, 140–150. https://doi.org/10.1016/j.geoderma.2014.02.002

Ramirez-Lopez, L., Stevens, A., Viscarra Rossel, R., Lobsez, C., Wadoux, A., & Breure, T. (2016). *Resemble: Regression and similarity evaluation for memory-based learning in spectral chemometrics* (R package version, 1.2.2). https://cran.r-project.org/web/packages/resemble/

Reeves, J. B. III (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*, *158*(1–2), 3–14. https://doi.org/10.1016/j.geoderma.2009.04.005

Reeves, J. B. III, McCarty, G. W., & Reeves, V. B. (2001). Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. *Journal of Agricultural and Food Chemistry*, *49*(2), 766–772. https://doi.org/10.1021/jf0011283

Reeves, J. B. III, & Smith, D. B. (2009). The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Applied Geochemistry*, *24*(8), 1472–1481. https://doi.org/10.1016/j.apgeochem.2009.04.017

Roberts, C., Workman, J. J., & Reeves, J. B. III (2004). *NIR in agriculture*. ASA, CSSA, and SSSA.

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 Sensing*, *13*(12), 2265. https://doi.org/10.3390/rs13122265

Sankey, J. B., Brown, D. J., Bernard, M. L., & 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. https://doi.org/10.1016/j.geoderma.2008.09.019

Sarathjith, M. C., Das, B. S., Wani, S. P., & Sahrawat, K. L. (2016). Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma*, *267*, 1–9. https://doi.org/10.1016/j.geoderma.2015.12.031

Schmidt, K., Behrens, T., Friedrich, K., & Scholten, T. (2010). A method to generate soilscapes from soil maps. *Journal of Plant Nutrition and Soil Science*, *173*, 163–172. https://doi.org/10.1002/jpln.200800208

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 Science Society of America Journal*, *83*(6), 1746–1759. https://doi.org/10.2136/sssaj2019.06.0205

Shen, X., Li, Y., Rönnegård, L., Udén, P., & Carlborg, Ö. (2014). Application of a genomic model for high-dimensional chemometric analysis: A genomic model for chemometric analysis. *Journal of Chemometrics*, *28*(7), 548–557. https://doi.org/10.1002/cem.2614

Shenk, J. S., Westerhaus, M. O., & Berzaghi, P. (1997). Investigation of a LOCAL calibration procedure for near infrared instruments. *Journal of Near Infrared Spectroscopy*, *5*(4), 223–232. https://doi.org/10.1255/jnirs.115

Shepherd, K. D., & Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal*, *66*(3), 988–998. https://doi.org/10.2136/sssaj2002.9880

Shi, Z., Ji, W., Viscarra Rossel, R. A., Chen, S., & Zhou, Y. (2015). Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library. *European Journal of Soil Science*, *66*(4), 679–687. https://doi.org/10.1111/ejss.12272

Sila, A. M., Shepherd, K. D., & Pokhariyal, G. P. (2016). Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemometrics and Intelligent Laboratory Systems*, *153*, 92–105. https://doi.org/10.1016/j.chemolab.2016.02.013

Smith, P., Soussana, J., Angers, D., Schipper, L., Chenu, C., Rasse, D. P., Batjes, N. H., Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J. E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., & 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. https://doi.org/10.1111/gcb.14815

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. *Applied Spectroscopy Reviews*, *49*(2), 139–186. https://doi.org/10.1080/05704928.2013.811081

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107, 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7

Sudduth, K. A., & Hummel, J. W. (1996). Geographic operating range evaluation of a NIR soil sensor. *Transactions of the ASAE*, *39*(5), 1599–1604.

Teófilo, R. F., Martins, J. P. A., & Ferreira, M. M. C. (2009). Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *Journal of Chemometrics*, *23*(1), 32–48. https://doi.org/10.1002/cem.1192

Terra, F. S., Demattê, J. A. M., & Viscarra Rossel, R. A. (2015). Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis–NIR and mid-IR reflectance data. *Geoderma*, *255*(256), 81–93. https://doi.org/10.1016/j.geoderma.2015.04.017

Tibshirani, S., Friedman, H., & Hastie, T. (2017). *The elements of statistical learning series in statistics*. Springer.

Vašát, R., Kodešová, R., Klement, A., & Borůvka, L. (2017). Simple but efficient signal pre-processing in soil organic carbon spectroscopic estimation. *Geoderma*, *298*, 46–53. https://doi.org/10.1016/j.geoderma.2017.03.012

Vasques, G. M., Grunwald, S., & Harris, W. G. (2010). Spectroscopic models of soil organic carbon in Florida, USA. *Journal of Environmental Quality*, *39*(3), 923–934. https://doi.org/10.2134/jeq2009.0314

Vasques, G. M., Grunwald, S., & Sickman, J. O. (2008). Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma*, *146*(1–2), 14–25. https://doi.org/10.1016/j.geoderma.2008.04.007

Viscarra Rossel, R. A., & 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

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-Science Reviews*, *155*, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012

Viscarra Rossel, R. A., Jeon, Y. S., Odeh, I. O. A., & McBratney, A. B. (2008). Using a legacy soil sample to develop a mid-IR spectral library. *Australian Journal of Soil Research*, *46*(1), 1–16. https://doi.org/10.1071/SR07099

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*(1–2), 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007

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. https://doi.org/10.1016/j.geoderma.2011.08.001

Vohland, M., Ludwig, M., Thiele-Bruhn, S., & Ludwig, B. (2014). Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, *223*(225), 88–96. https://doi.org/10.1016/j.geoderma.2014.01.013

Wetterlind, J., & 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. https://doi.org/10.1111/j.1365-2389.2010.01283.x

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 Science Society of America Journal*, *82*(3), 722–731. https://doi.org/10.2136/sssaj2017.10.0361

Wills, S., Seybold, C., Chiaretti, J., Sequeira, C., & West, L. (2013). Quantifying tacit knowledge about soil organic carbon stocks using soil taxa and official soil series descriptions. *Soil Science Society of America Journal*, *77*(5), 1711–1723. https://doi.org/10.2136/sssaj2012.0168

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

How to cite this article: Dorantes, M. J., Fuentes, B. A., & Miller, D. M. (2022). Calibration set optimization and library transfer for soil carbon estimation using soil spectroscopy—A review. *Soil Science Society of America Journal*, *86,* 879–903. https://doi.org/10.1002/saj2.20435