HUNGARIAN AGRICULTURAL ENGINEERING
No 42/2023 5-13
Published online: http://hae-journals.org/
HU ISSN 0864-7410 (Print)
HU ISSN 2415-9751(Online)
DOI: 10.17676/HAE.2023.42.5
Received: 06.03.2023 - Accepted: 25.04.2023

PERIODICAL
OF THE COMMITTEE OF AGRICULTURAL AND BIOSYSTEM ENGINEERING OF THE HUNGARIAN ACADEMY OF SCIENCES
and
HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES
INSTITUTE OF TECHNOLOGY

RAPID DETECTION OF SOIL TEXTURE ATTRIBUTE BASED ON MID-INFRARED SPECTRAL LIBRARY IN SALT AFFECTED SOILS OF HUNGARY

Author(s):

M. A. MohammedZein[1,2,✉], E. Micheli[1], B. Rotich[1], P. N. Justine[1], A. E. E. Ahmed[3], H. Tharwat[1], A. Csorba[1]

Affiliation:

[1] Institute of Environmental Sciences – Hungarian University of Agriculture and Life Sciences, 2100 Gödöllő, Páter Károly u. 1., Hungary.
[2] Agricultural Research Corporation, Land and Water Research Centre, Wad Medani, Sudan
[3] Mechanical Engineering Doctoral School - Hungarian University of Agriculture and Life Sciences, 2100 Gödöllő, Páter Károly u. 1., Hungary.

Email address:

mohammdzain@yahoo.com; micheli.erika@uni-mate.hu; brotich@chuka.ac.ke; pjustine@mustnet.ac.tz; wadelawad2004@gmail.com; hanaatharwat@hotmail.com; csorba.adam@uni-mate.hu

Abstract: Quantifying among others the soil's physical properties is essential for the assessment of the diverse soil environmental functions including water balance of soils and pore structure, water erosion and various soil hydraulic properties. The mid-infrared (MIR) spectroscopy is a useful technique to predict soil attributes with high accuracy, efficiency and low cost. In this study, we examined the ability of our MIR soil spectral library in predicting the clay, silt, sand content of salt affected Hungarian soils. This research is part of a project to establish a MIR spectral library in the frame of the Hungarian Soil Information and Mentoring System (SIMS) survey. Salt affected soils type data was extracted from the spectral library then transformation of spectral reflectance values to absorbance values were performed. Moving average filtering method was applied to absorbance spectra before performing principal components analysis. To determine outlier samples and to select the proper samples for model calibration, Mahalanobis distance-based outlier detection method and Kennard-Stone Sampling selection method were applied on the principal component scores. Spectral and reference soil data were combined and split into training and testing datasets. MIR prediction models were built for sand, clay, and silt content using Partial Least Square Regression (PLSR) method. Coefficient determination, root mean square error and ratio performance to deviation were used to assess the models performance. The prediction accuracies of calibration sets for soil physical texture were excellent while the validation results were slightly lower but still with a good level of prediction.

Keywords: Kennard-Stone sampling, partial least square, Soil Information Monitoring System, Diffuse Reflectance Infrared Fourier Transform

1. Introduction

The soil surface layer's characteristics are important because they provide essential information for food production. Soil is a mixture of physical, chemical, mineralogical, and organic compounds, as well as water and air, and these properties have been degraded in many agricultural regions due to ineffective management [1]. Among the soil attributes, particle size distribution is an important soil property related to physical structure, and it is divided into three main fractions: clay (<0.002 mm), silt (0.002-0.02 mm), and sand (>0.02 mm). Soil physical attributes are required for different disciplines' study such as forest ecosystem, general agricultural production and long-term soil use [2]. One of the most essential soil physical features for determining infiltration rate, irrigation, and drainage practices is physical soil texture. It has a big impact on soil hydraulic characteristics including water permeability, soil water retention [3] and solute dispersion in the soil profile. Soil texture is therefore important for the environment, and land reclamation [4]. Soil texture also influences plant water uptake and the overall hydrological cycle [5], [6]. It has an impact on many

important soil attributes such as soil specific surface area and pore structure [7], [8]. On the other hand, convection attributable to upward water movement in reaction to evapotranspiration, diffusion due to a salinity gradient with depth, and limited drainage flow are the major causes of salt transfer to the soil surface [9]. Both soil water dynamics and salt accumulation phenomena are affected by physical soil texture. Furthermore, the relationship between some physical soil characteristics such as soil compaction, plasticity, consistency, mechanical resistance and air capacity are strongly correlated with soil particle size [10]. Additionally, soil mineral weathering rates, ion exchange and buffering capacity, and nitrogen and carbon sequestration are all affected by the relative content of particles within specific size ranges [11], [13]. Many soil processes, including pollutants and microbial activity, are governed by soil texture [14]. It has been used to aid in soil classification, management, and modeling of soil processes.

Knowing and analyzing the texture of the soil is vital to understanding how well it functions are related to plants and other soil processes. Various approaches can be used to identify the physical texture of soil. The two most important traditional assessment methods for soil texture are, hydrometer and the sieve-pipette, both are granulometric measurements of particle size using gravitational-sedimentation techniques. These methods are disadvantageous since they are extremely time-consuming and inaccurate e.g. under-estimate or overestimation of clay [15]. In addition, H_2O_2, HCl, $C_6H_5Na_3O_7$ and $NaHCO_3$ chemical compounds are necessary as pretreatment to remove soil organic matter, Fe oxides and carbonates from soil during measuring the soil physical texture. These compounds may generate toxic wastes that are environmentally harmful. Therefore, its application across large fields (e.g. soil survey activities and soil mapping) is impractical and expensive.

In contrast to the wet chemistry approaches, infrared spectroscopy has emerged as a feasible option for time and cost-effective solution for soil properties determination such soil texture. It is a low-cost and non-destructive method [16]. This approach is cheap, utilizes tiny subsamples and has the advantage that a single spectrum of soil sample integrates many attributes with highly precision [17], [18], do not require the use of chemical extracts that might harm the environment [19] and allows for the scanning of diverse soil types without samples dilution [20]. Fundamentally, soil infrared spectroscopy relies on the interplay of electromagnetic energy with matter to characterize samples' physical and biochemical composition. Fundamental molecular vibrations absorb electromagnetic radiation at specified wavelengths, resulting characteristic spectral fingerprints in mid-infrared (MIR) region (2.5 - 25 mm) which is sensitive to soils' organic and mineral components [14]. Several studies have shown that MIR across a wide range of soil types are more robust and provide accurate predictions of several soil properties such as clay and sand [21, 22, 23]. The reason for this is that the fundamental molecular vibrations of soil components that are absorbed at specific wavelengths of electromagnetic radiation occur in the absorbance MIR region. The MIR spectroscopy spectrum contains a high reflectivity, useful spectral features and gives greater information on soil attributes [24]. Various physical and chemical soil properties, including texture, have been detected using MIR spectroscopy [14], [25]. Nguyen [26] demonstrated Diffuse Reflectance Infrared Fourier Transform (DRIFT) MIR ability to distinguish diverse mineral components abundantly detected in Australian soils, such as kaolinite, quartz, carbonate, gibbsite, illite, and smectite minerals [27]. On the other hand, the soil science community has been recently working to create extensive soil mid-infrared spectral libraries on a national and global scale. Soil spectral libraries often contain significant amounts of soil samples that represent the diversity of soils in a given region. MIR spectral library has been shown to accurately estimate soil texture in addition to many soil attributes such as soil organic carbon, CEC, phosphorus and potassium content.

Due to the scatter effects caused by structure result in overlapping absorption features, diffuse reflectance spectra in soil are non-specific. To extract absorption patterns and correlate spectra with soil properties, multivariate techniques are required. Linear regression approaches for soil applications include stepwise multiple linear regression (SMLR), principal component regression (PCR), and partial least squares regression (PLSR). Thus, generation of prediction models based on the appropriate calibration dataset and robust algorithms is required for the accurate estimation of soil physical texture. In this regard, the PLSR is a powerful technique compared to other algorithms as it is easy to compute and interpret.

This study made use of data from the Hungarian MIR spectral library, which contained approximately 2200 MIR spectra collected on soils from Hungarian Soil Information and Mentoring System (SIMS). This massive database held data on soil samples analyzed using the same standard laboratory methods. As a result, we were able to determine which soil properties could be accurately predicted by MIR spectroscopy for assessing soil functions. The aims of this study therefore were to: a) build multivariate statistical models for

soil texture physical properties using PLSR and b) test the predictive capacity of MIR spectral library for sand, clay, and silt content in salt-affected soils types of Hungary.

2. Materials and Methods

2.1. Dataset

The soil samples spectral data utilized in this research were obtained from the MIR spectral library in Hungarian University of Agricultural and Life Sciences in Gödöllő which built based on the samples collected in frame of SIMS survey. The MIR spectral library database comprised measurements of about 2200 soil samples representing 10 Hungarian counties and five soil types. Salt affected soils type dataset was extracted from the MIR spectral library which contained about 100 soil samples.

2.2. Dataset preprocessing and outlier detection methods

Preprocessing methods for spectral dataset were used to enhance the accuracy of quantitative soil texture analysis**.** The Salt affected soils type spectra dataset were transformed from reflectance to absorbance value using the equation:

$$\text{Absorbance} = \log(1/\text{Reflectance}) \quad (1)$$

Absorbance spectra dataset were smoothed with a moving average window of 17 bands and Savitzky-Golay filtering methods to reduce and remove noise that represents random fluctuations in the signal. This noise may originate from the instrument or environmental laboratory conditions.

Principal Component Analysis (PCA) was applied to reduce the dimensionality of the spectral dataset, improve computational efficiency and to compress the spectral information into a few variables (Figure 1). Outlier detection was checked and calculated on PCs of spectral dataset using Mahalanobis distance method. The purpose of this methods was to identify samples that deviate from the average population of spectra[28]. Based on standard arbitrary threshold methods, the samples with a Mahalanobis dissimilarity larger than one were considered outliers.

2.3. Calibration sample selection and physical soil texture prediction models

In order to develop the best MIR spectral models for soil texture as well as to define how many observations (samples) should be listed as calibration dataset, Kennard-Stone sampling (KSS) selection method [29] was applied (Figure 2).

In terms of building soil texture models, salt affected soils processed dataset including reference soil data was split into training and testing datasets based on the KSS. Accordingly, 27 soil samples were selected for calibration dataset and the remaining samples were retained for the validation set (n = 63). In this study, MIR prediction models were built for sand, clay and silt content using PLSR [30] using calibration dataset as well as the highest number of principal components and oscorespls method [31].

Coefficient of determination (R^2), root mean square error (RMSE) and ratio performance to deviation (RPD) were used to assess the model's performance.

$$\boldsymbol{R2} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{y}_i)^2} \quad (2)$$

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - Y_i)^2} \quad (3)$$

$$\mathbf{RPD} = s_y/\boldsymbol{RMSE} \quad (4)$$

ŷ indicates the spectral library's predicted value, while **ȳ** and **y** represent the observed value average and observed value of reference soil database respectively **n** represents the sample number where ***I*** is equivalent to 1, 2, …, while, s_y the observed values' standard deviation.

RStudio software [32] was used for spectral displaying, analysis and modelling processes using several packages, functions and operators. Models development were performed using the caret package interface [33] and PLSR function from pls package [34].

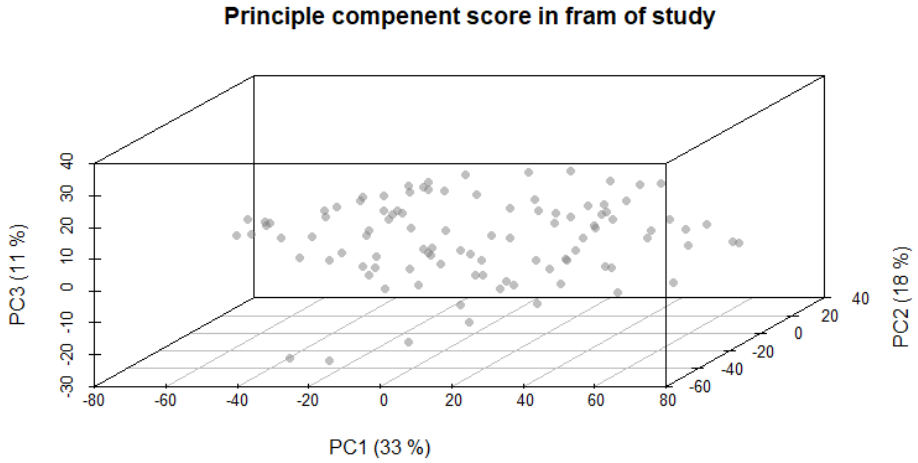

Figure 1. Principle component scores

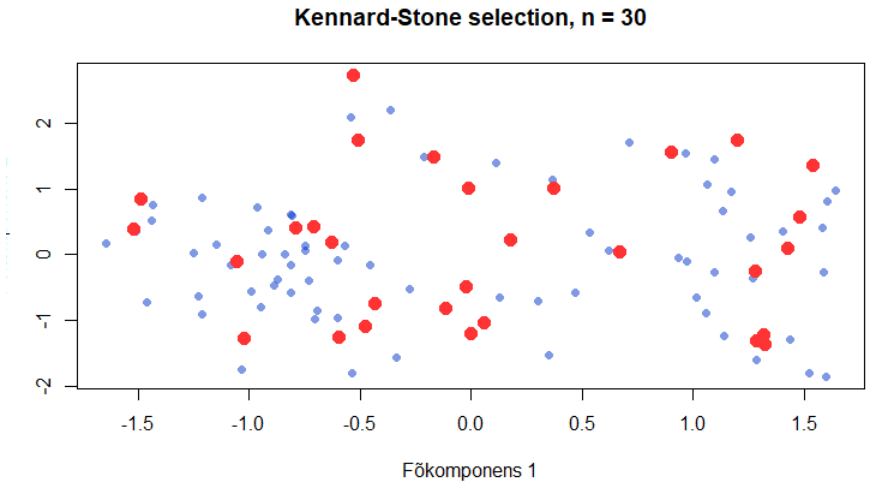

Figure 2. Kennard-stone sampling distributions

3. Results and Discussion

3.1. Mid-Infrared spectral signature

Generally, the MIR absorbances were caused by fundamental molecular vibrations, which were characterized by clearly identified peaks related to either organic or mineral compounds. Soil samples MIR spectra of salt affected soil dataset extracted from spectral library are given in (Figure 3). The general shape and the position of the absorption features are determined by the physical-chemical composition of the soil samples. Since the clay, silt and sand content refers to particle size classes that involves a wide range of mineral particles, the direct visual attribution of spectral features to these soil properties are limited. However, careful visual interpretation of the spectral data revealed that the spectral features of clay minerals were clear while the ones for quartz were less identifiable. For example, due to OH stretching, clays or aluminosilicates display

strong peaks around 3700 1/cm. Furthermore, the band about 1630 1/cm is usually assumed to be caused by water in the clay. The complex band at roughly 1048 1/cm may be due to clay mineral spectra, which is connected to the stretching vibrations of Si-O groups, similar finding were obtained by [35]. The hydroxyl stretching vibrations of kaolinite, smectite, and illite are thought to be responsible for the absorption bands amongst 3800 and 3600 (1/cm). More specifically, the absorption peak at 3620 (1/cm) might be due to clay minerals, a similar result was obtained by [26]. A sharp band at 798 1/cm with a shoulder around 779 1/cm prove the existence of quartz mixes [36]. According to Nguyen [26], weak spectra signatures near 1100–1000 cm^{-1} can also related to quartz. In addition, the clear band ranging between 2562 - 2480 1/cm may assigned to the vibration of molecules in quartz minerals.

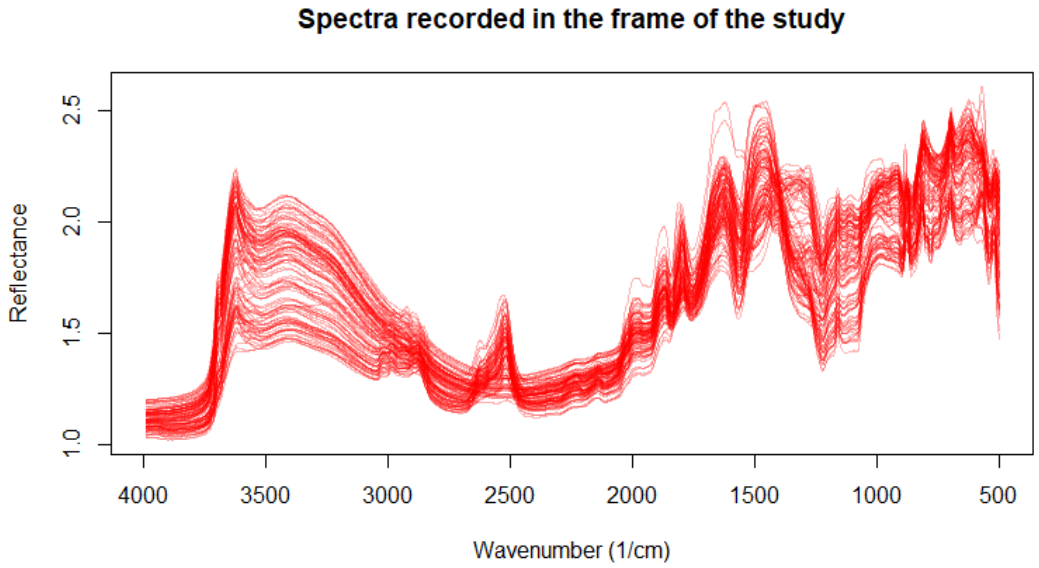

Figure 3. Mid-infrared spectra of salt affected soil dataset

3.2. Salt affected soil spectral dataset model performance

The pursuit of an efficient model for estimating soil texture is a common topic in soil science research [15]. Tables 1, 2 and 3 represent the test set validation and calibration of the spectral-based soil texture. Overall, components of physical soil texture were predicted excellent with the highest accuracy using the testing sets. Generally, the good performance models for sand, clay and silt content may be attributed to the high spectral activity of these materials in the MIR region.

3.2.1. Sand

Amongst all soil texture in this study, especially, sand content showed the highest prediction accuracy at training and testing datasets (Table 1). The coefficient determination was 0.88, ratio performance to deviation was 2.92, while root mean square error was 8.42 at testing dataset (Table 1). The model parameters for the testing set represent the real performance of the models. The high coefficients determination of the sand content predictive models are attributed to fundamental vibrations of associated minerals in the MIR regions [14]. According to Mohanty [37], the majority of the absorption peaks that are directly or indirectly related to SiO_2 fall in the MIR region. Thus, the MIR spectra predicted sand or SiO_2 with greater accuracy.

Table 1. Results of the prediction models.

	Training Dataset				**Testing Dataset**			
Sand %	**Mean**	**R^2**	**RMSE**	**RPD**	**Mean**	**R^2**	**RMSE**	**RPD**
	26.59	0.96	4.3	5.33	29.05	0.88	8.42	2.92

3.2.2. Clay

The testing datasets show that clay prediction accuracy is good but not as high as the sand component (Table 2). Total clay content had R^2 = 0.80, RMSE = 7.11 and RPD = 2.23 (Table 2). The high R^2 of clay content predictive models are attributed to specific strong absorption bands associated with chemical bonds [38] as well as fundamental vibrations of associated minerals in the MIR regions [14]. According to Urselmans [39], predicting clay content from MIR spectra was more direct because the absorption of the spectra was primarily concentrated in the mineral regions of the spectrum.

Table 2. Results of the prediction models.

	Training Dataset				**Testing Dataset**			
Clay %	**Mean**	**R^2**	**RMSE**	**RPD**	**Mean**	**R^2**	**RMSE**	**RPD**
	34.35	0.92	4.30	3.56	31.52	0.80	7.11	2.23

3.2.3. Silt

The prediction accuracy for total silt of the training dataset was high with the coefficient determination of 0.94 and ratio performance to deviation of 4.13 (Table 3). Whereas the root mean square error was 3.85. The prediction accuracy of testing set was good but slightly lower than the training set as well as sand content, but almost the same as clay content (Table 3). The total silt had (R^2 = 0.80, RMSE = 6.38, RPD = 2.27). The rather good silt prediction was surprising because it outperformed many previous findings, as reviewed by [14]. However, the achieved good results could be attributed to indirect effects on predicted silt. This assumption is supported by the strong negative correlation between conventionally measured silt and sand contents.

Table 3. Results of the prediction models.

	Training Dataset				**Testing Dataset**			
Silt %	**Mean**	**R^2**	**RMSE**	**RPD**	**Mean**	**R^2**	**RMSE**	**RPD**
	39.05	0.94	3.85	4.13	40.03	0.80	6.38	2.27

Generally, these results were similar to the findings by other researchers who achieved sand with R^2 of 0.94, silt with R^2 of 0.84 and clay with R^2 of 0.79 [40]. Thomas [41], showed good results in MIR-based predictions for clay with R^2 = 0.88 and sand with R^2 = 0.90 for soils from a Kenyan farm validation set as well as Madari [25] who obtained R^2 of 0.99, 0.8 and 0.96 for the estimation of sand, silt and clay, respectively. Hati [42], obtained R^2 of 0.79 for the sand and clay predictions, while an R^2 of 0.73 for the silt prediction from Eastern India soils which are lower prediction accuracy than our results. Similarly, Pirie [43] were unable to achieve high accuracy with predictions: clay (R^2 = 0.72), followed by sand (R^2 = 0.62) and silt (R^2 = 0.34).

4. Conclusions

The goal of this study is to predict sand, clay and silt from a salt-affected soil types dataset consisting of 100 soil samples extracted from the Hungarian MIR spectral library, using PLSR statistical model.
This study has demonstrated that MIR spectral libraries contain useful information related to soil texture and could be used as a cheap, fast and reliable alternative in the prediction of sand, clay and silt in salt-affected soil types in Hungary and elsewhere globally in soils with similar characteristics.

The PLSR model and technique outlined here can provide rapid predictions of physical soil texture in frame of these soil types.

References

[1] **S. Mantel, C. J. E. Schulp, and M. van den Berg**, "Modelling of soil degradation and its impact on ecosystem services globally," *Report; ISRIC—World Soil Inf. Wageningen, Netherlands*, 2014.

[2] **L. Duchesne and R. Ouimet**, "Digital mapping of soil texture in ecoforest polygons in Quebec, Canada," *PeerJ*, vol. 9, p. e11685, Jun. 2021, doi: 10.7717/peerj.11685.

[3] **J. H. M. Wösten, Y. A. Pachepsky, and W. J. Rawls**, "Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics," *J. Hydrol.*, vol. 251, no. 3–4, pp. 123–150, Oct. 2001, doi: 10.1016/S0022-1694(01)00464-4.

[4] **X. Li, S. X. Chang, and K. F. Salifu**, "Soil texture and layering effects on water and salt dynamics in the presence of a water table: a review," *Environ. Rev.*, vol. 22, no. 1, pp. 41–50, Mar. 2014, doi: 10.1139/er-2013-0035.

[5] **K. R. Hultine, D. F. Koepke, W. T. Pockman, A. Fravolini, J. S. Sperry, and D. G. Williams**, "Influence of soil texture on hydraulic properties and water relations of a dominant warm-desert phreatophyte," *Tree Physiol.*, vol. 26, no. 3, pp. 313–323, Mar. 2006, doi: 10.1093/treephys/26.3.313.

[6] **K. E. Saxton and W. J. Rawls**, "Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions," *Soil Sci. Soc. Am. J.*, vol. 70, no. 5, pp. 1569–1578, Sep. 2006, doi: 10.2136/sssaj2005.0117.

[7] **J. C. Fiès and A. Bruand**, "Particle packing and organization of the textural porosity in clay-silt-sand mixtures," *Eur. J. Soil Sci.*, vol. 49, no. 4, pp. 557–567, Dec. 1998, doi: 10.1046/j.1365-2389.1998.4940557.x.

[8] **J. C. Santamarina, K. A. Klein, Y. H. Wang, and E. Prencke**, "Specific surface: Determination and relevance," *Can. Geotech. J.*, vol. 39, no. 1, pp. 233–241, Feb. 2002, doi: 10.1139/t01-077.

[9] **S. Kessler, L. Barbour, K. C. J. Van Rees, and B. S. Dobchuk**, "Salinization of soil over saline-sodic overburden from the oil sands in Alberta," *Can. J. Soil Sci.*, vol. 90, no. 4, pp. 637–647, Dec. 2010, doi: 10.4141/CJSS10019.

[10] **C. T. Fongaro *et al.***, "Improvement of clay and sand quantification based on a novel approach with a focus on multispectral satellite images," *Remote Sens.*, vol. 10, no. 10, p. 1555, Sep. 2018, doi: 10.3390/rs10101555.

[11] **R. K. Kolka, D. F. Grigal, and E. A. Nater**, "Forest soil mineral weathering rates: Use of multiple approaches," *Geoderma*, vol. 73, no. 1–2, pp. 1–21, Sep. 1996, doi: 10.1016/0016-7061(96)00037-7.

[12] **E. de C. C. Telles *et al.***, "Influence of soil texture on carbon dynamics and storage potential in tropical forest soils of Amazonia," *Global Biogeochem. Cycles*, vol. 17, no. 2, p. n/a-n/a, Jun. 2003, doi: 10.1029/2002gb001953.

[13] **L. Wiklander**, "The role of neutral salts in the ion exchange between acid precipitation and soil," *Geoderma*, vol. 14, no. 2, pp. 93–105, 1975, doi: 10.1016/0016-7061(75)90068-3.

[14] J**. M. Soriano-Disla, L. J. Janik, R. A. Viscarra Rossel, L. M. MacDonald, and M. J. McLaughlin**, "The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties," *Appl. Spectrosc. Rev.*, vol. 49, no. 2, pp. 139–186, Feb. 2014, doi: 10.1080/05704928.2013.811081.

[15] **C. L. Thomas, J. Hernandez-Allica, S. J. Dunham, S. P. McGrath, and S. M. Haefele**, "A comparison of soil texture measurements using mid-infrared spectroscopy (MIRS) and laser diffraction analysis (LDA) in diverse soils," *Sci. Rep.*, vol. 11, no. 1, p. 16, Dec. 2021, doi: 10.1038/s41598-020-79618-y.

[16] **W. Ng, B. Minasny, S. H. Jeon, and A. McBratney**, "Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions," *Soil Secur.*, vol. 6, p. 100043, Mar. 2022, doi: 10.1016/j.soisec.2022.100043.

[17] **L. Raphael**, "Application of FTIR Spectroscopy to Agricultural Soils Analysis," *Fourier Transform. - New Anal. Approaches FTIR Strateg.*, 2011, doi: 10.5772/15732.

[18] **B. K. Waruru, K. D. Shepherd, G. M. Ndegwa, A. Sila, and P. T. Kamoni**, "Application of mid-infrared spectroscopy for rapid characterization of key soil properties for engineering land use," *Soils Found.*, vol. 55, no. 5, pp. 1181–1195, Oct. 2015, doi: 10.1016/j.sandf.2015.09.018.

[19] **R. A. Viscarra Rossel, R. N. McGlynn, and A. B. McBratney**, "Determining the composition of

mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy," *Geoderma*, vol. 137, no. 1–2, pp. 70–82, Dec. 2006, doi: 10.1016/j.geoderma.2006.07.004.

[20] **G. Siebielec, G. W. McCarty, T. I. Stuczynski, and J. B. Reeves**, "Near- and Mid-Infrared Diffuse Reflectance Spectroscopy for Measuring Soil Metal Content," *J. Environ. Qual.*, vol. 33, no. 6, pp. 2056–2069, Nov. 2004, doi: 10.2134/jeq2004.2056.

[21] **W. Ng *et al.***, "Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra," *Geoderma*, vol. 352, pp. 251–267, Oct. 2019, doi: 10.1016/j.geoderma.2019.06.016.

[22] **J. B. Reeves**, "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*, vol. 158, no. 1–2, pp. 3–14, Aug. 2010, doi: 10.1016/j.geoderma.2009.04.005.

[23] **R. A. Viscarra Rossel, D. J. J. Walvoort, A. B. McBratney, L. J. Janik, and J. O. Skjemstad**, "Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties," *Geoderma*, vol. 131, no. 1–2, pp. 59–75, Mar. 2006, doi: 10.1016/j.geoderma.2005.03.007.

[24] **K. D. Shepherd and M. G. Walsh**, "Infrared Spectroscopy—Enabling an Evidence-Based Diagnostic Surveillance Approach to Agricultural and Environmental Management in Developing Countries," *J. Near Infrared Spectrosc.*, vol. 15, no. 1, pp. 1–19, Feb. 2007, doi: 10.1255/jnirs.716.

[25] **B. E. Madari, J. B. Reeves, P. L. O. A. Machado, C. M. Guimarães, E. Torres, and G. W. McCarty**, "Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols," *Geoderma*, vol. 136, no. 1–2, pp. 245–259, Dec. 2006, doi: 10.1016/j.geoderma.2006.03.026.

[26] **T. T. Nguyen, L. J. Janik, and M. Raupach**, "Diffuse reflectance infrared fourier transform (Drift) spectroscopy in soil studies," *Aust. J. Soil Res.*, vol. 29, no. 1, pp. 49–67, 1991, doi: 10.1071/SR9910049.

[27] **L. J. Janik, R. H. Merry, S. T. Forrester, D. M. Lanyon, and A. Rawson**, "Rapid Prediction of Soil Water Retention using Mid Infrared Spectroscopy," *Soil Sci. Soc. Am. J.*, vol. 71, no. 2, pp. 507–514, Mar. 2007, doi: 10.2136/sssaj2005.0391.

[28] **B. K. Waruru, K. D. Shepherd, G. M. Ndegwa, P. T. Kamoni, and A. M. Sila**, "Rapid estimation of soil engineering properties using diffuse reflectance near infrared spectroscopy," *Biosyst. Eng.*, vol. 121, pp. 177–185, May 2014, doi: 10.1016/j.biosystemseng.2014.03.003.

[29] **R. W. Kennard and L. A. Stone**, "Computer Aided Design of Experiments," *Technometrics*, vol. 11, no. 1, pp. 137–148, Feb. 1969, doi: 10.1080/00401706.1969.10490666.

[30] **A. Lorber, L. E. Wangen, and B. R. Kowalski**, "A theoretical foundation for the PLS algorithm," *J. Chemom.*, vol. 1, no. 1, pp. 19–31, Jan. 1987, doi: 10.1002/cem.1180010105.

[31] **A. Wadoux, B. Malone, B. Minasny, M. Fajardo, and A. Mcbratney**, *Soil Spectral Inference With R*, vol. 49, no. 0. Cham: Springer International Publishing, 2020.

[32] **R Core Team**, "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.," 2022.

[33] **K. Max *et al.***, "Classification and Regression Training," *Packag. R CRAN*, p. 198, 2016, [Online]. Available: https://github.com/topepo/caret/%5CnBugReports.

[34] **B.-H. Mevik, R. Wehrens, and K. H. Liland**, "Partial Least Squares and Principal Component Regression," *Packag. R CRAN*, pp. 1–59, 2016, [Online]. Available: https://cran.r-project.org/web/packages/pls/pls.pdf.

[35] **A. Tinti, V. Tugnoli, S. Bonora, and O. Francioso**, "Recent applications of vibrational mid-Infrared (IR) spectroscopy for studying soil components: a review," *J. Cent. Eur. Agric.*, vol. 16, no. 1, pp. 1–22, 2015, doi: 10.5513/JCEA01/16.1.1535.

[36] **J. Madejová**, "FTIR techniques in clay mineral studies," *Vib. Spectrosc.*, vol. 31, no. 1, pp. 1–10, Jan. 2003, doi: 10.1016/S0924-2031(02)00065-6.

[37] **B. Mohanty, A. Gupta, and B. S. Das**, "Estimation of weathering indices using spectral reflectance over visible to mid-infrared region," *Geoderma*, vol. 266, pp. 111–119, Mar. 2016, doi: 10.1016/j.geoderma.2015.11.030.

[38] **R. A. V. Rossel and T. Behrens**, "Using data mining to model and interpret soil diffuse reflectance spectra," *Geoderma*, vol. 158, no. 1–2, pp. 46–54, Aug. 2010, doi: 10.1016/j.geoderma.2009.12.025.

[39] **T. Terhoeven-Urselmans, T.-G. Vagen, O. Spaargaren, and K. D. Shepherd**, "Prediction of Soil

Fertility Properties from a Globally Distributed Soil Mid-Infrared Spectral Library," *Soil Sci. Soc. Am. J.*, vol. 74, no. 5, pp. 1792–1799, Sep. 2010, doi: 10.2136/sssaj2009.0218.

[40] **L. J. Janik, R. H. Merry, and J. O. Skjemstad**, "Can mid infrared diffuse reflectance analysis replace soil extractions?," *Aust. J. Exp. Agric.*, vol. 38, no. 7, pp. 681–696, 1998, doi: 10.1071/EA97144.

[41] **C. L. Thomas, J. Hernandez-Allica, S. J. Dunham, S. P. McGrath, and S. M. Haefele**, "A comparison of soil texture measurements using mid-infrared spectroscopy (MIRS) and laser diffraction analysis (LDA) in diverse soils," *Sci. Rep.*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-020-79618-y.

[42] **K. M. Hati *et al.***, "Mid-Infrared Reflectance Spectroscopy for Estimation of Soil Properties of Alfisols from Eastern India," *Sustain.*, vol. 14, no. 9, p. 4883, Apr. 2022, doi: 10.3390/su14094883.

[43] **A. Pirie, B. Singh, and K. Islam**, "Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties," *Aust. J. Soil Res.*, vol. 43, no. 6, pp. 713–721, 2005, doi: 10.1071/SR04182.