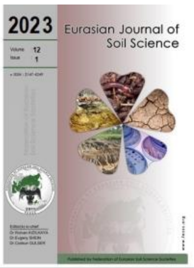
Eurasian Journal of Soil Science 2023, 12(4), 300 - 309

Eurasian Journal of Soil Science

Journal homepage : http://ejss.fesss.org



Prediction of some selected soil properties using the Hungarian Mid-infrared spectral library

Mohammed Ahmed MohammedZein [a,b,*], **Adam Csorba** [a], **Brian Rotich** [a], **Phenson Nsima Justin** [a], **Hanaa Tharwat Mohamed** [a], **Erika Micheli** [a]

[a] Institute of Environmental Sciences, Department of Soil Science, Hungarian University of Agriculture and Life Sciences, Páter Károly u. 1, 2100 Gödöllő, Hungary
[b] Land and Water Research Center, Agricultural Research Corporation, Sudan

Article Info
Received : 01.02.2023
Accepted : 20.05.2023
Available online : 05.06.2023

Author(s)
M.A.MohammedZein*
A.Csorba
B.Rotich
P.N.Justin
H.T.Mohamed
E.Micheli

* Corresponding author

Abstract

Routine soil chemical and physical laboratory analysis provides a better understanding of the soil by evaluating its quality and functions. Demands for the development of national Mid-infrared (MIR) spectral libraries for predicting soil attributes with high accuracy have risen substantially in the recent past. Such MIR spectral library is usually regarded as a fast, cheap and non-destructive technique for estimating soil properties compared to laboratory soil analysis. The main objective of this research was to assess the performance of the Hungarian MIR spectral library in estimating four soil properties namely: Cation Exchange Capacity (CEC), Exchangeable Mg and Ca and pH water at different scenarios. Archived soil samples were scanned and spectra data were saved in the Fourier transform infrared spectrometer OPUS software. Preprocessed filtering, outlier detection and calibration sample selection methods were applied to the spectral library. MIR calibration models were built for soil attributes using partial least square regression method and the models were validated with sample predictions. R2, RMSE and RPD were used to assess the goodness of calibration and validation models. MIR spectral library had the ability to estimate soil properties such as CEC and exchangeable Ca and Mg through various scale models (national, county and soil type). The findings showed that the Hungarian MIR spectral library for estimation of soil properties has the ability to provide good information on national, county and soil type scales at different levels of accuracy.

Keywords: Mid-infrared spectroscopy, soil information monitoring system, partial least square regression, fourier transform infrared spectrometer, coefficient determination.

Introduction

Soil is an important element of agricultural production especially in agroecology. Food security, water security, energy sustainability, climate stability, biodiversity preservation, and the provision of ecosystem services all depend on soil (McBratney et al., 2014). It is critical to recognize and monitor soil physical and chemical attributes using innovative approaches. Routine soil chemical and physical laboratory analysis must be performed to evaluate soil functions. Conventional laboratory techniques are widely regarded as accurate methods for characterizing soil attributes, however, they sometimes have been viewed as impractical due to their time-consuming, and occasional imprecision (Demattê et al., 2019). Many new soil analysis techniques have recently been developed, in particular, diffuse reflectance spectroscopy. In essence, an Infrared (IR) spectrum provides a chemical profile of the sample. Soil infrared spectroscopy techniques have demonstrated several advantages over wet chemistry methods. This approach is cheap, utilizes tiny subsamples and have the advantage that a single spectrum of soil sample integrates many attributes with high precision (Raphael, 2011; Waruru et al., 2015). It permits rapid acquiring of soil data, does not require the use of chemical extracts

: https://doi.org/10.18393/ejss.1309753
: http://ejss.fesss.org/10.18393/ejss.1309753
Publisher : Federation of Eurasian Soil Science Societies
e-ISSN : 2147-4249

that might harm the environment and allows for the scanning diverse of soil types without samples dilution (Siebielec et al., 2004; Viscarra Rossel et al., 2006; Seybold et al., 2019). The IR spectroscopy is a repeatable and reproducible analytical approach for predicting soil properties (Soriano-Disla et al., 2014).

The electromagnetic spectrum of infrared radiation ranges from 0.7 µm to 1 mm that contains: near-infrared (0.70 - 2.5 µm), mid-infrared (2.5 - 25 µm) and far-infrared (25 - 1000 µm) (Nocita et al., 2015). The two most important spectral ranges for soil investigation and analysis are mid-infrared and near-infrared (Wijewardane et al., 2018). The mid-infrared (MIR) spectroscopy spectrum contains a high reflectivity, useful spectral features and gives greater information on soil attributes (Shepherd and Walsh, 2007; Bo Stenberg et al., 2010). This is due to the fact that MIR range results are based on fundamental molecular vibrations that are absorbed at the specific wavelengths of electromagnetic radiation, while vis-NIR spectra result from overtones and combination bands which are complex and more difficult to describe than those recorded in the MIR region. The type of molecular motions, functional groups, or bonds present in the soil sample can be identified through mid-infrared spectroscopy since every frequency correlates to a certain quantity of energy and a specific molecular motion such asstretching, bending, etc (Tinti et al., 2015). The MIR range shows high-density peaks (Shepherd and Walsh, 2007; Soriano-Disla et al., 2014), containing much mineral composition information on soils such as Si-bearing minerals and iron forms. The basic vibrations of functional groups in minerals and organic matter of soil samples are used to explain the strong absorption of mid-infrared spectra (Shepherd and Walsh, 2007). MIR has been confirmed to show better results and high predictions for several soil properties across soil types in comparison to near-infrared spectroscopy (Pirie et al., 2005; Minasny and McBratney, 2008). In order to build predictive models, data from mid-infrared spectroscopy can be grouped into spectral libraries (harmonized point-dataset with coupled reflectance and analytical reference measurements) with the progress of mid-infrared spectroscopy in soil science. Many publications show soil attributes have been efficiently estimated based on the mid-infrared spectral library with high accuracy. It has been usefully applied to predict various physical (Kasprzhitskii et al., 2018), biological and chemical soil properties (Reeves and Smith, 2009; Acqui et al., 2010). On the other hand, traditional soil surveys and fresh soil sampling campaigns are costly and time-consuming. Legacy soil samples have an abundance of spectral information that can be utilized to improve the calibration models of the mid-infrared spectral library. The majority of large soil spectral databases are built from archived historical soil samples (Rossel and Webster, 2012).

Multivariate statistical techniques have given a powerful approach for soil component discrimination, such as partial least square regression (PLSR). PLSR has been used for soil attributes prediction from the spectral library and can quantify varied soil attributes with a high level of accuracy (Seybold et al., 2019). PLSR is easy to compute and understand (Wijewardane et al., 2018), and commonly integrates PCA and multiple regression (Wold et al., 2001).

Although soil spectroscopy methods have been used in previous years in literature to predict some soil attributes in Hungary, the potential use of an extensive national MIR spectral library that contains different soil types for estimating soil properties is yet to be intensively explored. Therefore, the present study objectives were: i) to build a multivariate statistical models using PLSR based on Hungarian MIR spectral libaray and ii) to test the predictive capacity of the Hungarian spectral library in the spectral based estimation of key some chemical soil properties (CEC, exchangeable Ca and Mg and pH in water).

Material and Methods

Hungarian MIR spectral library and soil samples

The spectral library consists of 2200 soil samples, corresponding to horizons of 543 soil profiles. The soil samples collected from the laboratory bank archives of Soil Information Conservation and Montoring System (SIMS), representing 10 Hungarian counties which are: Baranya, Fejer, Komarom_Esztergom, Nograd, Pest, Tolna, Bacs-Kiskun, Bekes, Csongrad and Jasz-Nagykun-Szolnok (Figure 1). These samples belong to the first SIMS project survey of 1992. The MIR spectral library was built at the Hungarian University of Agriculture and Life Sciences, Szent István Campus.

Previously, all soil samples were dried, mashed, and filtered via a two-millimetre sieve, with the remaining part stored in SIMS archives in plastic containers at room temperature. Three hundred gram from each sample were packaged in plastic sacs and shipped out to the Department of Soil Science, Godollo. Coning and quartering was done to obtain 20 g of soil subsamples, which were then grinded to less than 0.5 (fine powdered particle size between 20 and 53um) by hand using an agate pestle and mortar. Through a micro spatula, the fine soil samples were put into aluminium sample cups, and one by one the loaded samples were placed in the sample holding tray. Excess soil was removed to reduce sample surface roughness and the surface was leveled with a straight-edged tool. The Bruker Alpha II with a spectral range of 2500 – 25000 nm (4000 – 500 cm^{-1}) was used to scan the 2200 soil samples given for this study under DRIFT mode. A scan of the gold background was taken before the measurement of each sample to account for variations in temperature and moisture content. Gold background is used as a reference material in mid-infrared spectroscopy methods since it does not absorb infrared light (Nash, 1986). Every soil sample was read three times using three subsamples, and each spectrum was produced from 47 scans. Soil spectra were measured following the protocol proposed by the World Agroforestry Centre (Dickens, 2014). The collected information of all spectra was saved with the Fourier Transform Infrared Spectrometer (FTIR) spectrometer OPUS software.

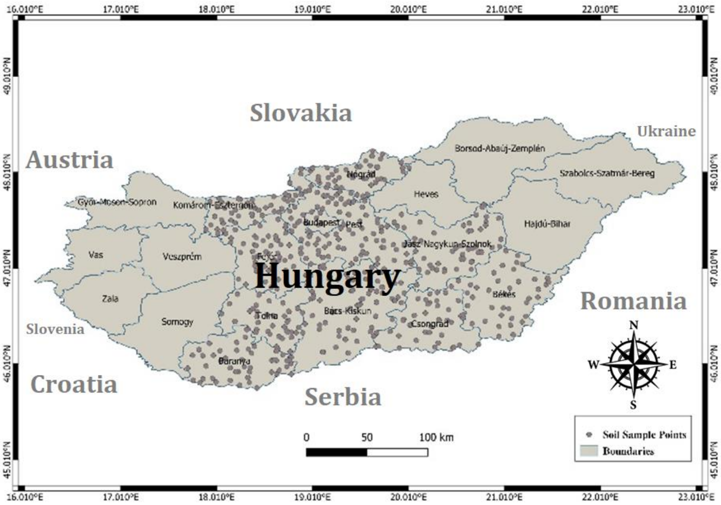


Figure 1. Spread of sampling points according to counties in Hungary

In terms of soil reference data, physical and chemical soil parameters were determined at the horizon level using conventional laboratory methods in the frame of the SIMS project and have been stored in the project database since 1992. TIM (1995) gave details for reference laboratory methods used in the conventional database of SIMS. The conventional database was subjected to quality and consistency checks before being used as soil reference data for calibration models.

Spectral data mangement

Initially, the transformation of measured spectral reflectance to absorbance value was performed using the equation:

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

Absorbance spectra were preprocessed with a moving average window of 17 bands and Savitzky-Golay filtering methods (Savitzky and Golay, 1964).

Principal Component Analysis (PCA) was applied to reduce the dimension of the spectral library and improve computational efficiency for different model scenarios of our data. Mahalanobis distance outlier detection method was carried out on principal component scores of spectral data to identify samples that deviate from the average population of spectra (Shepherd and Walsh, 2002; Waruru et al., 2014). Based on standard arbitrary threshold methods, the samples with a Mahalanobis dissimilarity larger than one were considered outliers. Detected outlier samples were filtered away from the mid-infrared spectral library dataset at different levels of the scenarios then further investigation and calibration were performed on the remaining soil samples.

Kennard-Stone Sampling method (Kennard and Stone, 1969) was applied to the spectral library data to define how many observations (samples) should be listed in calibration. Optimal calibration sample sets was selected and the remaining samples were retained for the validation set.

Spectroscopic modeling for soil properties prediction

The mid-infrared spectral library and soil reference data, including the depths of horizons, were merged into one dataset. Three modelling scenarios were used. Consequently, the dataset was split according to 10 counties, 6 soil types and the national scenario that included the whole dataset. Furthermore, depending on the KSS method, the dataset of each sub-scenario was split into a calibration dataset and validation datasets. In this research, PLSR (Lorber et al., 1987) statistical models were fitted between latent variables (mid-infrared spectral library) and response variables (soil attributes) based on calibration data using the highest number of principal components and oscorespls method (Wadoux et al., 2020). For each soil property, the PLSR regression coefficients were plotted using the number of components. The built PLSR models and the appropriate number of components were used to predict soil properties using spectra on the calibration and validation datasets. Four soil properties in the frame of this study were predicted, including, exchangeable calcium (Ca^{++}), exchangeable magnesium (Mg^{++}), pH water, and Cation Exchange Capacity (CEC). R software (R Core Team, 2022) was used for spectral displaying, analysis and modelling processes. Models development and predictions were performed using the caret package interface (Max et al., 2016) and PLSR function from pls package (Liland et al., 2016).

Models performance and accuracy

Coefficient of determination (R2), ratio performance to deviation (RPD) and root mean square error (RMSE) were used to determine the goodness and inaccuracy of the model's predictions.

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

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

$$\mathrm{RPD} = \quad \mathrm{s_y}/\mathrm{RMSE}$$

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.

eval function of R was used to derive the goodness measurement of prediction and validation models.

Results and Discussion

Summary statistics of spectral library soil attributes

In this study, the predictability of 4 soil attributes were assessed at different scenarios. Figure 2 shows the distribution of the dataset for soil attributes at the national level. The soil attributes of the spectral library dataset showed wide-ranging distributions, as well as based on frequency histograms, many of them were skewed from the normal distribution (Figure 2). These factors were expected in this database, since the samples were derived from different depths and horizons of soil types at wide spatial variability covering several variations of climatic conditions, geological formation and parent material, land cover and human activities.

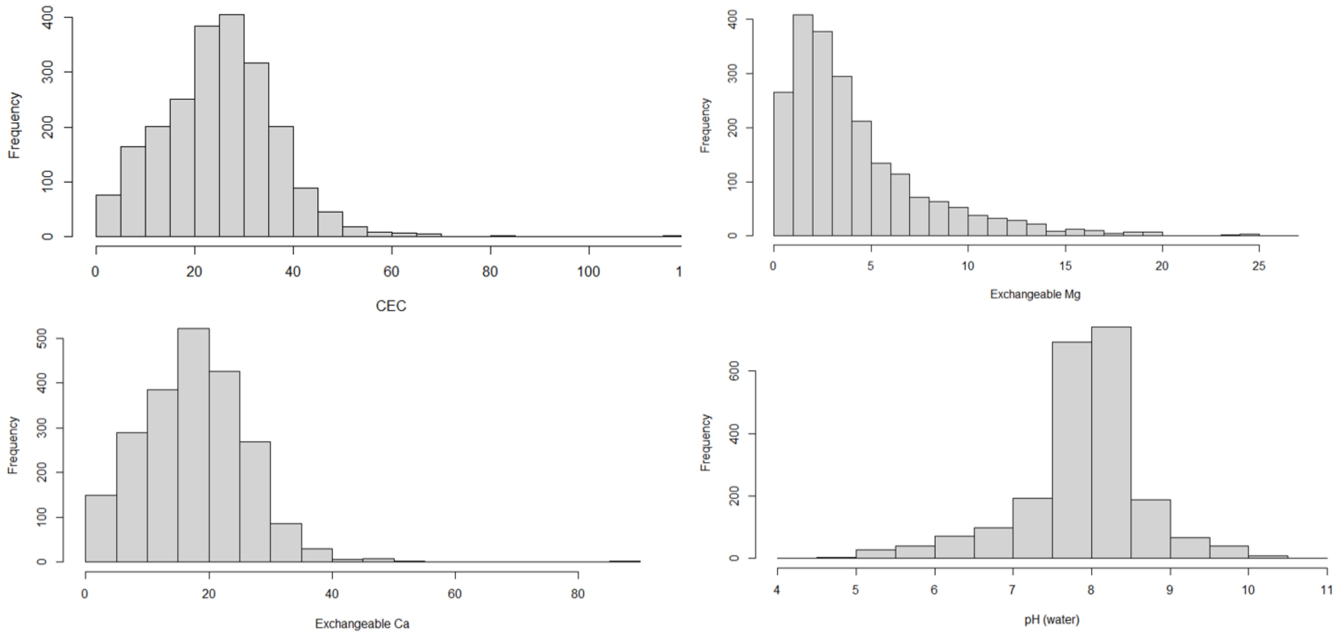


Figure 2. Distribution of dataset for soil properties

Mid-Infrared spectral signature and regression coefficient of PLSR models

Absorbance signatures in the MIR spectral library were due to fundamental molecular vibrations described by peaks related to different compounds (Figure 3), mainly minerals and organic constituents. Despite, the overlapping bands and exchangeable cations are not spectrally active, several absorption bands linked to certain functional groupings were identified (Figure 3). For example, many exchangeable cations influence the position and strength of the wide band around 3400 1/cm. Its position falls in the order of K$^+$, Na$^+$, Ca^{2+} and Mg^{2+}, which corresponds to the cation's increasing polarizing strength (charge/radius). These findings are in agreement with the results of some earlier authors such as Madejová (2003) and Tinti et al., (2015).

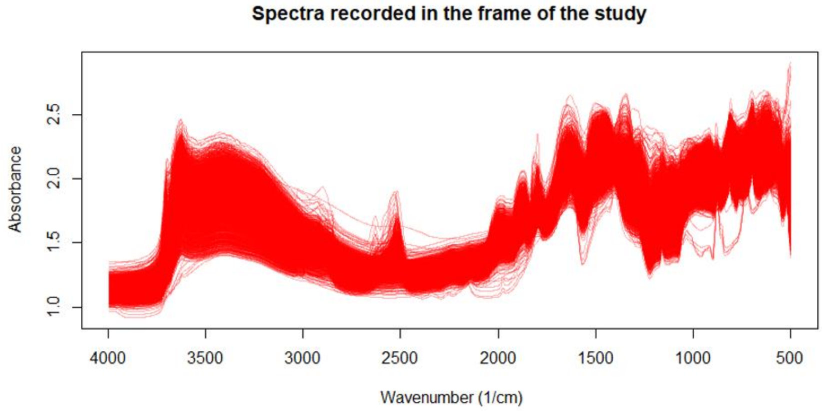


Figure 3. Absorbance mid-infrared spectral library of different Hungarian soil types using Bruker alfa II

On ther hand, the regression coefficient of PLSR illustrated the association between the mid-infrared frequencies and the soil constituents. The plots of PLSR regression coefficients vs wavelength for calibration models of the 4 soil attributes at national levels data are shown in figure 4. In this context Viscarra Rossel et al. (2006) stated that the positive peaks belong to the interest components, whilst negative peaks refer to interfering components. It's worthwhile to mention that some important wavelengths for the CEC prediction model are almost similar to those of the clay. For instance, the weak bands at 400 1/cm and significant broad wavelengths between 1000 to 1500 1/cm. The important bands for predicting exchangeable calcium are those near 400, 900, 1300, 720 and 1800 1/cm with the latter two bands attributed to diagnostic peaks for calcite (Figure 4) which may consistent with the result obtained by Nguyen et al. (1991). The peak bands for models prediction of exchangeable magnesium are those near 400 and 1200 1/cm, in addition to bands near 1440, 1470, and 875 1/cm which are representative of carbonate and may be caused by the presence of magnesium carbonate and dolomite (Figure 4). Figure 4 also showed that the regression coefficients for exchangeable Ca and Mg prediction models are identical in many spectral regions to those of the clay and organic matter, indicating that these soil properties are associated with each other.

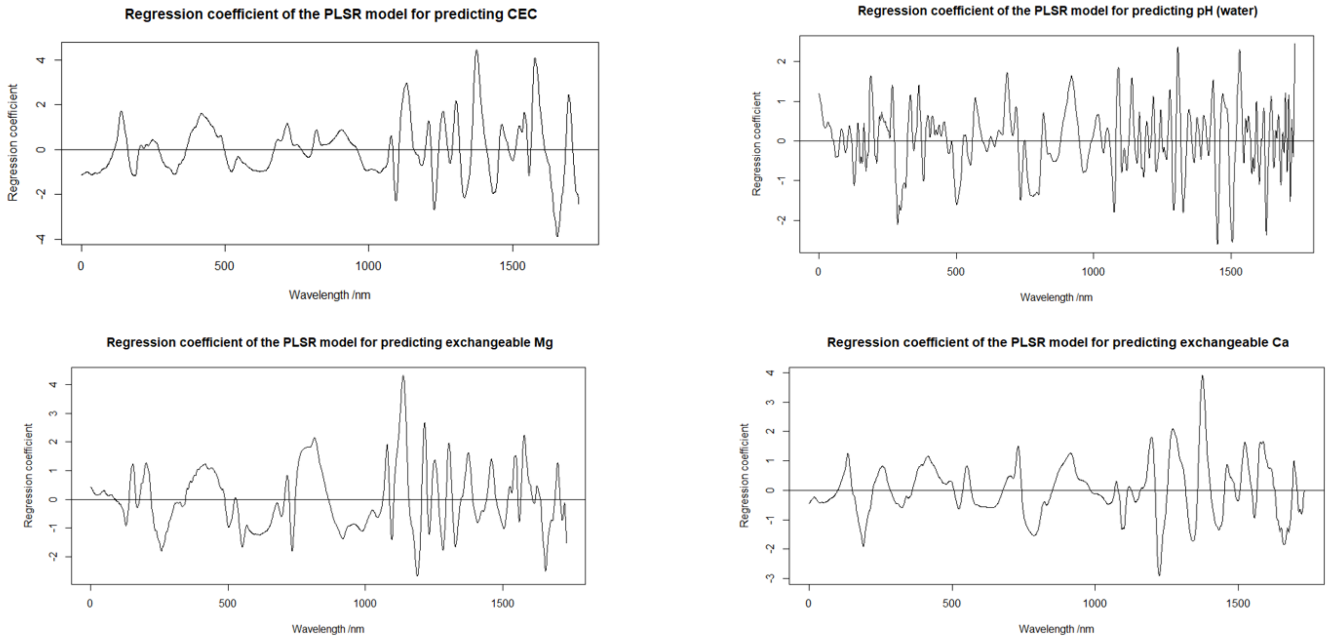


Figure 4. PLSR models' standard regression coefficient for predicting CEC, Exch. Mg, Ca, and pH water

Hungarain Spectral Library Model Performance

Cation Exchange Capacity

The calibration model of CEC at the national scale reached a R2 of 0.61 and RMSE of 8.24 and the validation set reached respective R2 and RMSE of 0.57 and 7.78 (Table 1). At the counties level, Baranya and Tolna showed a R2 ≥ 0.90, while Fejer had R2 of 0.83, and three counties showed R2 of 0.68 (Bekes, Csongrad and Jasz-Nagykun-Szolnok) while only one county showed R2 below 0.55 (Bacs-Kiskun) in the calibration models. Validation sets showed only four counties had R2 ≥ 0.60, while the remaining six counties had R2 ≤ .051. At the soil type scenarios, brown forest soils and alluvial and colluvial soils showed the best calibration results (R2 of 0.86 and RMSE of 3.96 and 4.29 respectively) whereas Chernozem soils had R2 of 0.47 and RMSE of 7.08 and were the worst result (Table 1). Validation sets showed two soil types had R2 ≥ 0.70 (brown forest and Skeletal soils). Four soil types showed R2 ≤ 0.50. The poor results were expected since CEC is not spectrally active, while the other good results were due to the contribution of clay minerals and organic carbon matter to the prediction of CEC and they are correlated with each other (Stenberg et al., 2010). Demattê et al. (2019) showed similar prediction accuracy ranges (R2 0.97 – 0.11) for CEC in the Brazilian spectral library. In addition, several studies with good predictions were observed by Pirie et al. (2005) that showed prediction reached a R2 of 0.82 in small spectral library (415 soil samples). Terhoeven-Urselmans et al. (2010) also achieved the good accuracy (R2 = 0.83) for 4438 global soil samples.

Table 1. PLSR model values, descriptive statistics and results of calibration and validation prediction models of CEC

		Calibration set							Validation set						
	CEC cmol(+)/kg	n	Min	Max	Mean	R2	RMSI	RPD	n	Min	Max	Mean	R2	RMSE	RPD
National		241	1.48	119.9	26.14	0.61	8.24	1.60	1959	1.64	116.5	24.94	0.57	7.78	1.53
Counties	Pest	98	2.38	59.63	22.69	0.76	5.68	2.05	294	2.15	67.40	25.14	0.65	7.00	1.70
	Baranya	70	3.85	67.94	25.05	0.90	3.39	3.24	141	5.61	42.61	24.07	0.80	2.67	2.23
	Fejer	49	4.76	66.80	27.76	0.83	5.35	2.42	186	8.34	83.12	27.74	0.38	8.25	1.27
	Komarom-Esztergom	35	7.73	60.28	23.13	0.65	6.38	1.72	125	8.39	46.40	22.03	0.61	4.90	1.60
	Nograd	55	3.33	57.22	28.56	0.77	6.82	2.11	88	2.95	49.82	27.64	0.73	5.42	1.93
	Tolna	39	5.50	119.9	29.48	0.96	4.73	4.96	153	5.55	53.00	24.86	0.51	5.41	1.44
	Bacs-Kiskun	98	2.25	54.47	16.44	0.50	7.71	1.42	186	1.48	84.21	11.63	0.28	7.58	1.18
	Bekes	70	11.2	57.66	34.09	0.68	5.51	1.77	132	3.41	58.39	33.71	0.45	6.72	1.35
	Csongrad	50	4.38	48.00	25.04	0.68	7.67	1.77	116	5.66	49.67	28.19	0.31	11.41	1.21
	Jasz-Nagykun-Szolnok	40	1.68	42.44	24.14	0.68	6.37	1.78	179	5.42	61.73	29.33	0.49	5.79	1.41
Soil types	Chernozem	149	2.89	46.40	23.56	0.47	7.08	1.38	530	3.41	61.73	26.99	0.32	6.93	1.22
	Brown forest	99	3.85	57.22	23.66	0.86	3.96	2.73	395	2.95	49.82	23.83	0.77	4.22	2.09
	Alluvial & colluvial	55	2.86	59.63	26.47	0.86	4.29	2.70	153	2.25	53.00	22.31	0.48	6.51	1.40
	Meadow	149	1.68	119.89	32.64	0.55	11.84	1.49	261	4.51	68.16	32.32	0.50	7.44	1.42
	Skeletal	99	2.38	61.57	16.45	0.50	8.25	1.43	200	1.48	49.33	11.31	0.70	4.84	1.84
	Salt-affected	27	6.70	66.83	32.51	0.68	8.11	1.81	64	4.20	84.21	29.75	0.04	13.6	1.03

Exchangeable Mg and Ca

The exchangeable Mg and Ca of both calibration and validation models showed decrease and variance results. The calibration results at national level were good for exchangeable Mg but were satisfactory for exchangeable Ca with respective R2 values being 0.77 and 0.54 and RPD values 2.09 and 1.48. Whereas validation model sets had R2 values of Mg and Ca of 0.52 and 0.48, respectively (Tables 2 and 3). Calibration prediction at county levels for exchangeable Mg showed 4 counties had R2 ≥ 0.90 and 3 counties had R2 lower than 0.55 (Table 3) while, exchangeable Ca showed 6 counties had R2 ≥ 0.80 and only Csongrad county had R2 lower than 0.55 (Table 2). However, the validation prediction results had R2 ranging from 0.14 to 0.66 for exchangeable Mg and ranging from 0.18 to 0.74 for exchangeable Ca (Tables 2 and 3).

Calibration predictions for exchangeable Mg were satisfactory (R2 lower than 0.75) for all soil types except Alluvial and colluvial soils (R2 of 0.94 and RPD of 4.01) and Meadow soils (R2 of 0.82 and RPD of 2.37; Table 9) whereas calibration predictions for exchangeable Ca were poorer (R2 ≤ 0.50 and RPD ≤ 1.42) for three soil types, but was excellent for brown forest soils (R2 of 0.96 and RMSE of 1.56) and alluvial and colluvial soils (R2 of 0.83 and RMSE of 3.32; Table 2). Validation results of soil types had R2 ranging from 0.33 to 0.60 for exchangeable Mg and ranging from 0.32 to 0.71 for exchangeable Ca except salt-affected soils had R2 of 0.01 (Tables 2 and 3).

The poor model results were not expected but we posit that exchangeable Ca and Mg may be present in low concentrations, not have particular MIR absorption features as well as the lack of correlation with spectrally active properties. Ng et al., (2022) concluded that the high correlation with spectrally active elements or the element concentration itself in soils is primarily responsible for prediction accuracy of the elemental concentrations. Furthermore, inverse links with carbon content may also justify the low prediction results, suggesting less sites for exchangeable cations on soil charges (from organic matter) that may be filled by H^+. TIM (1995) reported that soil conditions in Hungary show soil nutrient use stagnated between 1985 and 1990, and reduced sharply after 1990. Soil nutrient balance became negative compared to the period of 1981 to 1986. These reasons, in addition to different land nutrition management conditions, may justify the low exchangeable (Ca^{++} and Mg^{++}) predictions and CEC in various areas, counties and soil types in Hungary. Exchangeable Ca was predicted with fairly good accuracy (R2 = 0.85) by Rossel et al. (2008) followed by exchangeable Mg (R2 = 0.78). Similarly, study by Stenberg and Rossel (2010) observed good predictions for exchangeable Ca (R2 = 0.89) and Mg (R2 = 0.76). Pirie et al. (2005) however, reported lower performance for exchangeable Mg (R2 = 0.69) and exchangeable Ca (R2 = 0.64). Similarly, study by Terhoeven-Urselmans et al. (2010) observed lower predictions for exchangeable Mg (R2 = 0.54) and exchangeable Ca (R2 = 0.78).

Table 2. PLSR model values, descriptive statistics and results of calibration and validation prediction models of exchangeable Ca

		Calibration set							Validation set						
	Ca cmol(+)/kg	n	Min	Max	Mean	R2	RMSE	RPD	n	Min	Max	Mean	R2	RMSE	RPD
National		241	0.67	87.46	18.52	0.54	6.72	1.48	1959	0.60	85.52	17.54	0.48	6.21	1.39
Counties	Pest	0.87	49.06	16.34	0.75	4.51	2.00	294	0.67	45.89	18.39	0.63	5.15	1.66	1.70
	Baranya	2.00	54.03	18.13	0.91	2.74	3.36	141	4.44	35.05	17.21	0.74	2.65	1.98	2.23
	Fejer	3.29	48.05	18.58	0.91	2.67	3.36	186	5.36	53.92	20.92	0.37	5.53	1.26	1.27
	Komarom-Esztergom	5.18	45.79	17.41	0.63	5.29	1.67	125	5.30	34.85	16.57	0.59	3.99	1.57	1.60
	Nograd	1.67	38.40	18.56	0.80	4.60	2.26	88	1.35	30.72	18.04	0.73	3.74	1.95	1.93
	Tolna	3.83	87.46	21.78	0.95	3.84	4.46	153	3.79	39.07	19.33	0.42	4.65	1.32	1.44
	Bacs-Kiskun	1.51	40.89	11.36	0.84	2.86	2.54	186	0.82	59.07	8.71	0.36	5.25	1.26	1.18
	Bekes	4.96	41.52	21.97	0.98	1.23	6.48	132	5.74	42.32	22.57	0.18	6.78	1.11	1.35
	Csongrad	1.76	45.13	16.44	0.09	9.50	1.06	116	2.67	39.17	18.78	0.34	8.49	1.24	1.21
	Jasz-Nagykun-Szolnok	0.60	31.33	15.51	0.57	5.44	1.55	179	3.12	45.31	18.76	0.36	6.01	1.26	1.41
Soil types	Chernozem	1.54	33.76	17.50	0.34	6.15	1.23	530	4.46	45.31	21.13	0.40	5.51	1.29	1.22
	Brown forest	1.68	38.40	16.95	0.96	1.56	5.34	395	1.35	35.43	16.45	0.67	3.62	1.75	2.09
	Alluvial & colluvial	2.28	40.46	19.33	0.83	3.32	2.45	153	1.51	39.07	16.25	0.50	4.91	1.41	1.40
	Meadow	0.60	87.46	20.85	0.66	7.39	1.73	261	2.92	53.92	20.57	0.32	6.83	1.21	1.42
	Skeletal	0.87	45.13	12.11	0.50	6.00	1.42	200	0.67	40.24	8.65	0.71	3.67	1.86	1.84
	Salt-affected	2.91	45.89	17.04	0.43	8.07	1.35	64	2.07	59.07	14.31	0.01	9.51	0.96	1.03

Table 3. PLSR model values, descriptive statistics and results of calibration and validation prediction models of exchangeable Mg

		Calibration set							Validation set						
	Mg cmol(+)/kg	n	Min	Max	Mean	R2	RMSE	RPD	n	Min	Max	Mean	R2	RMSE	RPD
National		241	0.67	87.46	18.52	0.54	6.72	1.48	1959	0.60	85.52	17.54	0.48	6.21	1.39
Counties	Pest	0.87	49.06	16.34	0.75	4.51	2.00	294	0.67	45.89	18.39	0.63	5.15	1.66	1.70
	Baranya	2.00	54.03	18.13	0.91	2.74	3.36	141	4.44	35.05	17.21	0.74	2.65	1.98	2.23
	Fejer	3.29	48.05	18.58	0.91	2.67	3.36	186	5.36	53.92	20.92	0.37	5.53	1.26	1.27
	Komarom-Esztergom	5.18	45.79	17.41	0.63	5.29	1.67	125	5.30	34.85	16.57	0.59	3.99	1.57	1.60
	Nograd	1.67	38.40	18.56	0.80	4.60	2.26	88	1.35	30.72	18.04	0.73	3.74	1.95	1.93
	Tolna	3.83	87.46	21.78	0.95	3.84	4.46	153	3.79	39.07	19.33	0.42	4.65	1.32	1.44
	Bacs-Kiskun	1.51	40.89	11.36	0.84	2.86	2.54	186	0.82	59.07	8.71	0.36	5.25	1.26	1.18
	Bekes	4.96	41.52	21.97	0.98	1.23	6.48	132	5.74	42.32	22.57	0.18	6.78	1.11	1.35
	Csongrad	1.76	45.13	16.44	0.09	9.50	1.06	116	2.67	39.17	18.78	0.34	8.49	1.24	1.21
	Jasz-Nagykun-Szolnok	0.60	31.33	15.51	0.57	5.44	1.55	179	3.12	45.31	18.76	0.36	6.01	1.26	1.41
Soil types	Chernozem	1.54	33.76	17.50	0.34	6.15	1.23	530	4.46	45.31	21.13	0.40	5.51	1.29	1.22
	Brown forest	1.68	38.40	16.95	0.96	1.56	5.34	395	1.35	35.43	16.45	0.67	3.62	1.75	2.09
	Alluvial & colluvial	2.28	40.46	19.33	0.83	3.32	2.45	153	1.51	39.07	16.25	0.50	4.91	1.41	1.40
	Meadow	0.60	87.46	20.85	0.66	7.39	1.73	261	2.92	53.92	20.57	0.32	6.83	1.21	1.42
	Skeletal	0.87	45.13	12.11	0.50	6.00	1.42	200	0.67	40.24	8.65	0.71	3.67	1.86	1.84
	Salt-affected	2.91	45.89	17.04	0.43	8.07	1.35	64	2.07	59.07	14.31	0.01	9.51	0.96	1.03

pH in water

Overall, the predictions for soil chemical reaction within the different scenarios were poor. Soil pH in water at the national level had the poorest results in both groups of calibration and validation datasets (Table 4). In general, many counties pH models were better than the national and soil type levels. Four counties including Baranya, Bacs-Kiskun, Bekes and Jasz-Nagykun-Szolnok had high predictions (R2 = 0.91– 0.98 and RMSE = 0.12 – 0.32) in calibration sets, while two counties included Tolna and Csongrad represented the worst results (R2 = 0.09 and 0.04, respectively; Table 4) in the calibration data sets. Three counties had R2 raning from 0.59 to 0.78, while other counties had R2 ≤ 0.51 in validation sets.

With reference to the soil types and with regard to calibration sets only brown forest had the highest results (R2 of 0.94 and RMSE of 0.28). Salt-affected soils and alluvial and colluvial soils represented satisfactory models (R2 of 0.69 and 0.62, respectively; Table 4), while all the validation datasets results had R2 ≤ 0.38. The poor model results were expected because this attribute lacked direct spectral responses, while others good results may be due to correlation between pH and soil organic carbon and carbonates (Minasny et al., 2009; Reeves, 2010; Sarathjith et al., 2014). Terhoeven-Urselmans et al. (2010) obtained a higher prediction of water pH (R2 = 0.81) at a global level of spectral library compared to our results. Generally, Figure 4 and the descriptive statistics tables showed some soil attributes had small datasets that may have affected the prediction's accuracies. Reeves and Smith (2009) found that dataset diversity, parent materials, land uses, and climate such as our spectral library can lead to poor model prediction results.

Table 4. PLSR model values, descriptive statistics and results of calibration and validation prediction models of pH (water)

		Calibration set							Validation set						
	Mg cmol(+)/kg	n	Min	Max	Mean	R2	RMSE	RPD	n	Min	Max	Mean	R2	RMSE	RPD
National		241	4.80	9.84	7.90	0.29	1.17	1.19	1959	4.00	10.51	7.88	0.18	1.02	1.10
Counties	Pest	98	5.19	10.4	7.75	0.47	0.97	1.38	294	4.92	10.5	7.94	0.51	0.57	1.43
	Baranya	70	4.21	9.12	7.65	0.91	0.32	3.28	141	5.28	8.84	7.68	0.78	0.40	2.15
	Fejer	49	6.54	9.57	8.01	0.65	0.33	1.70	186	6.08	9.77	7.99	0.17	0.88	1.10
	Komarom-Esztergom	35	4.92	8.92	7.69	0.70	0.53	1.84	125	5.09	8.78	7.81	0.43	0.81	1.34
	Nograd	55	4.77	8.41	1.48	0.16	1.48	1.10	88	4.80	8.45	6.94	0.69	0.45	1.81
	Tolna	39	5.12	8.72	7.76	0.09	1.37	1.06	153	5.01	8.51	7.88	0.05	0.99	1.03
	Bacs-Kiskun	98	6.62	10.0	8.19	0.97	0.14	5.45	186	6.37	9.84	8.09	0.59	0.37	1.57
	Bekes	70	5.92	9.88	8.09	0.91	0.24	3.43	132	6.25	9.52	8.00	0.19	0.81	1.11
	Csongrad	50	6.87	9.90	8.32	0.04	2.28	1.03	116	4.00	10.1	8.17	0.13	2.76	0.94
	Jasz-Nagykun-Szolnok	40	6.14	9.92	8.01	0.98	0.12	6.67	179	5.88	9.96	7.93	0.32	0.57	1.21
Soil types	Chernozem	149	6.19	9.92	8.16	0.18	1.28	1.11	530	5.85	9.97	8.02	0.02	1.12	1.01
	Brown forest	99	4.21	9.12	7.41	0.94	0.28	3.94	395	4.77	8.73	7.26	0.38	0.96	1.28
	Alluvial & colluvial	55	6.65	9.57	7.99	0.62	0.31	1.63	153	5.50	9.28	7.94	0.33	0.49	1.23
	Meadow	149	6.52	10.1	8.13	0.13	1.05	1.08	261	4.00	9.88	7.99	0.16	1.04	1.10
	Skeletal	99	5.21	8.89	7.82	0.17	0.99	1.10	200	5.25	8.92	7.97	0.15	0.87	1.09
	Salt-affected	27	5.92	10.5	8.98	0.69	0.67	1.83	64	7.22	10.51	8.89	0.34	0.68	1.24

Conclusion

Based on the final findings of this study, the following points can be concluded:

1. The MIR spectral library reported with 2200 soil samples based on legacy soil samples of the Hungary SIMS project as well as, predicting an array of four soil chemical properties.
2. Models were built using PLSR for national level, ten counties and six soil types using the SIMS reference soil database and the spectral library data.
3. The results were logical for the CEC, exchangeable Ca and Mg which are not spectrally active but correlated with other active elements.
4. For soil properties that are not spectrally active with low content in the soil or have small sizes of samples, the prediction can turn out to be inaccurate, like pH water.
5. The results showed that legacy soil samples can be used to generate a spectral library with good quality information.
6. The developed Mid-infrared spectral library therefore can provides a way for rapid soil properties estimation at low cost and with short time compared to the conventional method.

References

D'Acqui, L.P., Pucci, A., Janik, L.J., 2010. Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *European Journal of Soil Science* 61(6): 865–876.

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

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

(BSSL): A general view, application and challenges. *Geoderma* 354*:* 113793.

Dickens, A.A.. 2014. Standard operating procedures. Method for analysing samples for spectral characteristics in mid IR range using alpha. Code: METH07V02. World Agroforestry Centre, Nairobi, Kenya.

Kasprzhitskii, A., Lazorenko, G., Khater, A., Yavna, V., 2018. Mid-infrared spectroscopic assessment of plasticity characteristics of clay soils. *Minerals* 8(5): 184.

Kennard, R.W., Stone, L.A., 1969. Computer Aided Design of Experiments. *Technometrics* 11(1): 137-148.

Lorber, A., Wangen, L.E., Kowalski, B.R., 1987. A theoretical foundation for the PLS algorithm. *Journal of Chemometrics* 1(1): 19–31.

Madejová, J., 2003. FTIR techniques in clay mineral studies. *Vibrational Spectroscopy* 31(1): 1–10.

Max, K., Weston, S., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, R. C., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., 2016. Caret: classification and regression training. R package version 6.0-71. Available at [Access date : 26.01.2023]: https://CRAN.R-project.org/package=caret

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213: 203–213.

Liland, K.H., Mevik, B.H., Wehrens, R., Hiemstra, P., 2016. Partial least squares and principal component regression. CRAN, 66p. Available at [Access date : 26.01.2023]: https://cran.r-project.org/web/packages/pls/pls.pdf

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences* 32(9): 1378–1388.

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.

Nash, D.B., 1986. Mid-infrared reflectance spectra (23–22 µm) of sulfur, gold, KBr, MgO, and halon. *Applied Optics* 25(14): 2427-2433.

Ng, W., Minasny, B., Jeon, S.H., McBratney, A., 2022. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. *Soil Security* 6: 100043.

Nguyen, T.T., Janik, L.J., Raupach, M., 1991. Diffuse reflectance infrared fourier transform (Drift) spectroscopy in soil studies. *Australian Journal of Soil Research* 29(1): 49–67.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E. Ben, Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J. A. M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., D. Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. In: Advances in Agronomy. Sparks, D.L. (Ed.). Vol. 132, pp. 139–159.

Pirie, A., Singh, B., Islam, K., 2005. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Australian Journal of Soil Research* 43(6): 713–721.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at [Access date : 26.01.2023]: https://www.R-project.org/

Raphael, L., 2011. Application of FTIR spectroscopy to agricultural soils analysis. In: Fourier Transforms - New Analytical Approaches and FTIR Strategies. Nikolic, G. (Ed.). InTech. pp. 385-404.

Reeves, J.B., 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.

Reeves, J.B., 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.

Rossel, R.A.V., 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.

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

Sarathjith, M.C., Das, B.S., Wani, S.P., Sahrawat, K.L., 2014. Dependency measures for assessing the covariation of spectrally active and inactive soil properties in diffuse reflectance spectroscopy. *Soil Science Society of America Journal* 78(5): 1522–1530.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36(8): 1627–1639.

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.

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.

Shepherd, K.D., Walsh, M.G., 2007. Infrared spectroscopy—enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *Journal of Near Infrared Spectroscopy* 15(1): 1–19.

Siebielec, G., McCarty, G.W., Stuczynski, T.I., Reeves, J.B., 2004. Near- and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. *Journal of Environmental Quality* 33(6): 2056–2069.

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.

Stenberg, B., Rossel, R.A.V., 2010. Diffuse reflectance spectroscopy for high-resolution soil sensing. In: Proximal Soil Sensing. Progress in Soil Science. Viscarra Rossel, R., McBratney, A., Minasny, B. (Eds.). Springer, Dordrecht. pp. 29–47.

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. In: Advances in Agronomy. Sparks, D.L. (Ed.). Vol. 107, pp. 163–215.

Terhoeven-Urselmans, T., Vagen, T.G., Spaargaren, O., Shepherd, K.D., 2010. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Science Society of America Journal* 74(5): 1792–1799.

TIM, 1995. Soil Conservation and Monitoring System. Ministry of Agriculture. Budapest, Hungary. Unpublished Report. [in Hungarian]

Tinti, A., Tugnoli, V., Bonora, S., Francioso, O., 2015. Recent applications of vibrational mid-Infrared (IR) spectroscopy for studying soil components: a review. *Journal of Central European Agriculture* 16(1): 1–22.

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.

Wadoux, A., Malone, B., Minasny, B., Fajardo, M., Mcbratney, A.B., 2020. Soil spectral inference with R: Analysing digital soil spectra using the R programming environment. Springer Cham. 247p.

Waruru, B.K., Shepherd, K.D., Ndegwa, G.M., Sila, A., Kamoni, P.T., 2015. Application of mid-infrared spectroscopy for rapid characterization of key soil properties for engineering land use. *Soils and Foundations* 55(5): 1181–1195.

Waruru, B.K., Shepherd, K.D., Ndegwa, G.M., Kamoni, P.T., Sila, A.M., 2014. Rapid estimation of soil engineering properties using diffuse reflectance near infrared spectroscopy. *Biosystems Engineering* 121: 177–185.

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.

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