Geoderma 428 (2022) 116174

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Developing spectrotransfer functions (STFs) to predict basic physical and chemical properties of calcareous soils

Hasan Mozaffari [a], Ali Akbar Moosavi [a,*], Yaser Ostovari [b], Mohammad Amin Nematollahi [c], Mahrooz Rezaei [a,d]

[a] Department of Soil Science, College of Agriculture, Shiraz University, Shiraz, Iran
[b] Chair of Soil Science, Research Department of Ecology and Ecosystem Management, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
[c] Department of Biosystems Engineering, College of Agriculture, Shiraz University, Shiraz, Iran
[d] Meteorology and Air Quality Group, Wageningen University & Research, P.O. Box 47, 6700 AA Wageningen, the Netherlands

ARTICLE INFO

Handling Editor: Morgan Cristine L.S.

Keywords:
Low-cost procedure
Partial least square regression
Spectroscopy-based methods
Stepwise multiple linear regression
Vis-NIR spectral reflectance data

ABSTRACT

Spectroscopy in visible (Vis) and near-infrared (NIR) provides a rapid, timely, non-destructive, low-preparation, and less expensive analysis of soil samples in comparison with traditional laboratory analysis. Therefore, the objectives of this study were to predict basic physical and chemical properties of calcareous soils using Vis-NIR spectral reflectance data by applying partial least square regression (PLSR) and stepwise multiple linear regression (SMLR) approaches and finally to develop spectrotransfer functions (STFs). The target soil properties including sand, silt, clay, pH, electrical conductivity (EC), calcium carbonate equivalent (CCE), and water-soluble sodium (Na), potassium (K), calcium (Ca), and magnesium (Mg), in 234 soil samples, and DTPA-extractable iron (Fe), manganese (Mn), copper (Cu), and zinc (Zn), in 161 soil samples, were measured using standard laboratory procedures in calcareous soils of Fars province, Iran. The spectra of soil samples in Vis-NIR region (400–2500 nm) were collected using Rapid Content Analyzer spectrophotometer (NIRS XDS model) apparatus. Results revealed the better performance of developed PLSR models compared to STFs to predict most of the target soil properties; however, their capability differences were not significant. The STFs predicted sand, clay, and CCE with very good accuracy (0.78 $\leq$ determination coefficient of validation dataset, $R^2_{val} \leq 0.83$); pH, K, Mg, Fe, Mn, Cu, and Zn with good accuracy (0.65 $\leq R^2_{val} \leq 0.71$); and silt, EC, Na, and Ca with acceptable accuracy (0.55 $\leq R^2_{val} \leq 0.63$). It is strongly recommended to use the related STFs for predicting sand, clay, CCE, pH, K, Mg, Fe, Mn, Cu, and Zn of calcareous soils in order to save time and costs, less use of chemicals, and mapping large areas.

1. Introduction

Spectroscopy-based studies have recently become an interesting and useful topic in soil science (Vasava et al., 2019). In comparison with many traditional laboratory analysis, spectroscopy in the visible (Vis) and infrared (IR) light spectrum enables rapid, timely, non-destructive, low-preparation, and less expensive analysis of soil samples (Hobley and Prater, 2019; Zhao et al., 2020). In contrary, measuring soil physical and chemical properties by traditional methods is expensive and time and energy consuming, which needs to use chemicals that are harmful for human health and environment (Conforti et al., 2018; Adeline et al., 2017). Therefore, it seems that the prediction of basic soil physical and chemical properties using spectroscopy-based methods is useful and advantageous. There are many studies in literature that show the capability of spectroscopy-based methods to predict different soil properties (Khayamim et al., 2015; Conforti et al., 2018; Hobley and Prater, 2019; Ostovari et al., 2018; Guo et al., 2019; Munawar et al., 2020; Pyo et al., 2020).

One of the important aspects in predicting different soil properties using spectroscopy-based methods is the number of individual single bands as independent variables at a specific range of spectra. For instance, in Vis (350–700 nm) and near-infrared (NIR, 700–2500 nm) ranges, totally we are facing 2150 single bands (with data point interval of 1 nm). Analyzing and recognizing the effective single bands for predicting different soil properties by common methods is difficult due to multi-collinearity between spectra. Therefore, different methods like

* Corresponding author.
E-mail addresses: aamousavi@gmail.com, aamousavi@shirazu.ac.ir (A.A. Moosavi).

https://doi.org/10.1016/j.geoderma.2022.116174
Received 25 January 2022; Received in revised form 1 September 2022; Accepted 8 September 2022
Available online 19 September 2022

partial least square regression, PLSR (Ostovari et al., 2018), support vector machines, SVM (Xu et al., 2017), fuzzy rule-based models (Tsakiridis et al., 2019; Heil et al., 2019), and artificial neural networks, ANNs (Nawar and Mouazen, 2017) were employed to easily analyze the spectral data. It should be pointed out that some of the mentioned approaches (i.e., PLSR, SVM, and ANNs) have black box nature (Han et al., 2005; Park et al., 2010). Therefore, there is a lack of simple, easy to use, and applicable models for predicting soil properties by spectral data in literature. With applying PLSR (as a black box-nature approach) to analyze the spectral data, it is possible to recognize the effective and important single bands to predict different soil properties and to develop multiple linear models, the so-called spectrotransfer functions (STFs). It is noteworthy to mention that predicting soil properties by STFs as compared to predicting by PLSR models is i) simpler, ii) more applicable, iii) easier to use, and iv) no needs to use complicated computer software.

In literature, there are numerous attempts to predict different soil properties using spectroscopy data coupled with PLSR approach. For instance, organic carbon (OC) and extractable forms of potassium (K), sodium (Na), magnesium (Mg), and phosphorous (P) (Mouazen et al., 2010; Vohland et al., 2014); soil texture (Pinheiro et al., 2017; Zhang et al., 2017; Hobley and Prater, 2019; Xu et al., 2018a; Jaconi et al., 2019); bulk density (BD), pH, soil organic matter (SOM), total N, K, P; cation exchange capacity (CEC), free iron (Fe), water-soluble salts, available P, exchangeable aluminum (Al) and hydrogen, Al saturation, and base saturation (Vohland et al., 2014; Pinheiro et al., 2017; Zhang et al., 2017; Xu et al., 2018a, 2018b); calcium carbonate equivalent (CCE) and gypsum (Khayamim et al., 2015; Ostovari et al., 2018; Mina et al., 2021; Mozaffari et al., 2022c); microbial biomass-C (Vohland et al., 2014); and gravimetric and volumetric water contents (Zhang et al., 2017), among others, have been predicted with acceptable ($0.50 \leq R^2 < 0.65$), good ($0.65 \leq R^2 < 0.75$), or very good ($0.75 \leq R^2 < 0.90$) accuracies.

Although these studies are informative and improve our knowledge about spectroscopy-based approach, there is still lack of information regarding prediction of different soil properties in calcareous soils. In calcareous soils by using Vis-NIR spectral reflectance data, Ostovari et al. (2018) predicted soil erodibility, SOM, and $CaCO_3$ values with acceptable and good accuracies in southern part of Iran by PLSR method. Khayamim et al. (2015) reported reasonably good prediction of carbonates and gypsum by PLSR method in the central parts of Iran. Mina et al. (2021) used PLSR and SVM methods to predict clay, $CaCO_3$ and threshold friction velocity and predicted the mentioned properties with good accuracy in soils of Fars Province (southern parts of Iran). Furthermore, Mozaffari et al. (2022c) predicted CCE and SOM contents with excellent; clay with relatively well; and pH and near-saturated and saturated hydraulic conductivity with poor accuracies using PLSR method in the soils of Fars Province, Iran.

In addition to the mentioned limitation, there are a few studies in literature to develop STFs, which are simple, easy to use, and applicable to predict different soil properties e.g., soil bacterial abundance and diversity (Yang et al., 2019), hydraulic properties (Babaeian et al., 2015), erodibility factor (Ostovari et al., 2018; Salehi-Varnousfaderani et al., 2022), threshold friction velocity (Mina et al., 2021) and SOM (Mozaffari et al., 2022a). In fact, this gap emphasizes the essential demand for developing specific STFs to easily predict basic soil physical and chemical properties in order to save time and costs and lesser use of chemicals. Therefore, objectives of the present study were: i) to evaluate the capability of Vis-NIR spectral reflectance data to predict basic soil physical and chemical properties including sand, silt, clay, pH, electrical conductivity (EC), CCE, water-soluble Na, K, calcium (Ca) and Mg, and DTPA-extractable Fe, manganese (Mn), copper (Cu), and zinc (Zn) by PLSR method in calcareous soils, ii) to develop the STFs using the effective and important Vis-NIR spectral reflectance bands to predict the mentioned soil properties using stepwise multiple linear regression (SMLR) method, and iii) to compare the capability of PLSR method and SMLR-STFs to predict the mentioned soil properties in calcareous soils.

2. Material and methods

2.1. Study area

A total of 234 soil samples were collected from different locations of Fars Province, which is located in 50° 30′-55° 38′ E and 27° 03′-31° 42′ N in the south to southwest parts of Iran. The study area has an arid and semi-arid climate with the annual precipitation ranging from 50 to 1000 mm (MPB, 1994). The mean annual temperature of the study area is 17.5 °C (Gandomkar and Dehghani, 2012). The elevation of the southern and northern parts of the Fars Province varies between nearly 500 to 4000 m above the mean sea level, respectively (Khormali and Abtahi, 2003). The Fars Province experiences xeric, ustic, and aridic soil-moisture regimes along with mesic, thermic, and hyperthermic soil-temperature regimes (Banaei, 1998). Soils of the study area are calcareous and relatively calcareous (Mozaffari et al., 2021b; Ostovari et al., 2022), formed from soluble dolomite and calcite limestone with calcium carbonate (e.g., in the present study the soil CCE values were in range of 12.5–70.6 %). According to the Soil Taxonomy classification (Soil Survey Staff, 2014), the soils in the study area are mainly classified as Inceptisols (175 soils), Entisols (42 soils), and Aridisols (17 soils), which comprise 75, 18, and 7 % of the studied soils, respectively (Khormali and Abtahi, 2003; Natural Resources and Forestry Organization of Fars Province, 2010).

2.2. Determining physical and chemical properties

The soils (~2 kg) were randomly sampled by spade from the surface layer (0–20 cm depth) and different predominant land uses including cultivated fields of wheat, barley, corn, alfalfa, rice, and fallow fields (126 soil samples or 53.83 % of the total 234 soil samples), rangelands (69 soil samples or 29.5 % of the total samples), pear orchard garden (23 soil samples or 9.83 % of the total samples), and oak forests (16 soil samples or 6.84 % of the total samples) to cover a wide range of calcareous soils and land uses in the study area (Ostovari et al., 2021; Mozaffari et al., 2022b) during summer (August and September 2019). Samples were air dried, passed through a 2 mm sieve, and selected physical and chemical properties were determined using the common standard methods (similar to those used by Mozaffari et al., 2021a, 2022d), which are summarized in Table 1.

Table 1
Studied physical and chemical properties of the soil samples along with the methods used for their determination and the relevant references.

Property	Method	Reference
Texture (sand, silt, and clay contents) based on USDA classification system	Combination of the wet-sieving and sedimentation method by hydrometer	Gee and Bauder (1986)
pH of saturated paste by H_2O	Glass electrode pH-meter	Thomas (1996)
Electrical conductivity (EC) of saturated extract	EC-meter	Rhoades (1996)
Calcium carbonate equivalent (CCE)	Back titration with HCl	Loeppert and Suarez (1996)
Water-soluble sodium (Na) and potassium (K)	Flame photometry	Helmke and Sparks (1996)
Water-soluble calcium (Ca) and magnesium (Mg)	Titration with EDTA†	Richards (1954)
DTPA††-extractable iron (Fe), manganese (Mn), copper (Cu), and zinc (Zn)	Atomic absorption spectrometry	Lindsay and Norvell (1978)
Soil organic matter (SOM)	Walkley-Black wet combustion	Nelson and Sommers (1996)

†: Ethylene Diamine Tetraacetic Acid.
††: Diethylene Triamine Pentaacetic Acid.

2.3. Vis-NIR spectral data

Nearly 25 g of air-dried, 2 mm sieved, and thoroughly mixed soil samples were placed in the glassy container with about 4 cm diameter. Spectral absorbance of the soils were determined by a spectrophotometer apparatus (Rapid Content Analyzer, NIRS XDS model, Metrohm company, Switzerland) in 4200 wavelengths within a Vis-NIR range of 400–2500 nm with data point interval of 0.5 nm, wavelength accuracy of <0.05 nm, and band pass of 8.75 nm. In order to reduce the complexity, the spectra with 1 nm increment were considered as independent variables.

For each soil sample, three scans were taken to get an average absorbance spectrum. The spectral absorbance obtained at ranges of 400–417 nm and 2483–2500 nm were removed in order to eliminate noises. The reflectance (R) data were calculated by absorbance ones (A) and using a $1/10^{A}$ relation. Fig. 1 shows the mean, the mean ± standard deviation, the minimum, and the maximum of Vis-NIR reflectance spectra versus wavelengths along with four significant absorption bands in the studied soil samples. The pre-processing techniques including Savitzky-Golay derivative and smoothing methods (Savitzky and Golay, 1964) were used on spectral data with five smoothing points and zero-order Savitzky-Golay filter to eliminate fluctuations and noises and improve the quality of spectral data and predictions. It should be noted that all pre-processing procedures were performed after reducing the complexity and taking data with 1 nm increment. All of the pre-processing techniques were carried out using Unscrambler X v. 9.7 software package (CAMO, Technologies Inc., 2013).

2.4. Partial least squares regression (PLSR)

The PLSR is a multivariate statistical method commonly used for correlating soil spectral reflectance data to individual soil parameters and to predict the soil parameters, mainly due to its superiority over traditional methods in dealing with high dimensional multi-collinearity (Stenberg, 2010; Bilgili et al., 2011). In the PLSR method, the relationship between two matrices X (independent variables) and Y (dependent variables) is created through a multivariate linear relationship (Bilgili et al., 2011). The PLSR model decomposes the matrices of zero-mean variables X and Y into the following forms:

$$X = T\ P^T + E \tag{1}$$

$$Y = U\ Q^T + F \tag{2}$$

where the T and U are matrices of the extracted score vectors (components, latent vectors). The P and Q denote matrices of orthogonal loadings and E and F are the matrices of residuals. Regarding the weight vectors w and c, which present the corresponding score vectors by t = Xw and u = Yc, the PLSR method can be developed as an optimization problem as follows:

$$(w, c) = \underset{||w|| = ||c||=1}{\operatorname{argmax}} \{Cov(Xw,\ Yc)\}^2 \tag{3}$$

There are some algorithms such as NIPALS (nonlinear iterative PLS), SIMPLS (statistically inspired modification of the PLS), and kernels algorithms (Wold et al., 1983; de Jong, 1993; Rannar et al., 1994) for PLSR. The NIPALS and kernels algorithms are available in the Unscrambler software for this purpose. Having w and c, the response ($\hat{Y}$) can be estimated as follows:

$$\hat{Y} = X\ B \tag{4}$$

where

$$B = w\left(P^T\ w\right)^{-1} c = X^T\ U\left(T^T\ X\ X^T\ U\right)^{-1} T^T\ Y \tag{5}$$

In the present study, for developing different regression models, soil physical and chemical properties were considered as dependent variables (Y) and spectral reflectance bands in Vis-NIR region were considered as independent variables (X). All data were randomly divided into two subsets containing 75 % (calibration dataset) and 25 % (validation dataset) of the whole dataset. It should be pointed out that, the selected calibration and validation datasets were the same (constant) for predicting all of the studied soil properties. To the best of our knowledge, in PLSR analysis, selecting the best latent variables (LVs) is one of the most important steps in modeling procedure. The LVs are linear combinations of the original variables (Li et al., 2001), which are imported to PLSR models in the hidden layer. In order to optimize this hyper-parameter (number of LVs) for predicting a specific property, we developed PLSR models using calibration dataset with maximum 40 levels of LVs. The cross validation R^2 (R^2_{cv}) criterion was used to compare the accuracy of the developed PLSR models by each LVs level. The best number of LVs was the minimum one that resulted in the highest R^2_{cv} (Viscarra Rossel et al., 2006; Xu et al., 2018a). After optimizing the LVs for predicting a specific soil property, the PLSR model again was developed by calibration dataset and using the optimized number of LVs. After that, the models were tested using the validation datasets. All of the PLSR modeling procedures were performed in Unscrambler X v. 9.7 software package.

In order to recognize the important and effective Vis-NIR spectral reflectance single bands to predict a specific soil property, the plot of regression coefficient (B) values versus wavelengths obtained from PLSR analysis and Unscrambler X v. 9.7 software was used (see right hand side of Figs. 4, 6, 8, and 10). According to the addressed plots, there are several peaks showing the possibility of existing positive or negative correlations between spectral reflectance bands at recognized peak wavelengths and target properties. Therefore, all these recognized spectral reflectance values at peak wavelengths, as effective and important single bands, were imported to stepwise multiple linear regression (SMLR) models. Finally, some of the most important single bands (but not all of them) were appeared in the STFs (Ostovari et al., 2018; Mozaffari et al., 2022a). It should be noted that, all PLSR analysis, including developing the models and obtaining B plots to recognize the effective and important bands, for predicting the studied soil properties were performed using the processed reflectance spectra.

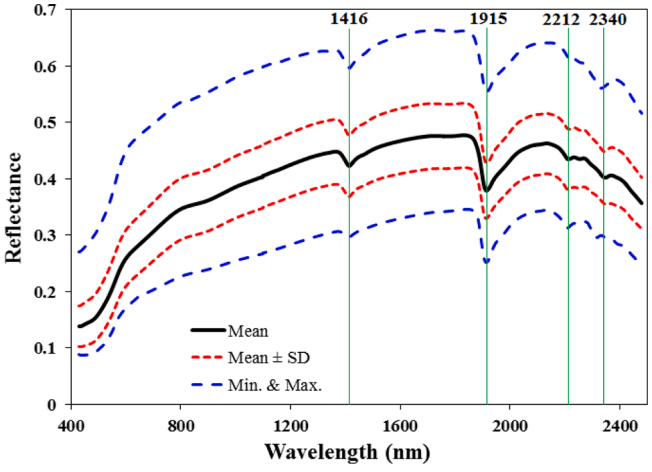

Fig. 1. The mean (Mean), mean ± standard deviation (Mean ± SD), minimum (Min.), and maximum (Max.) values of the reflectance spectra versus wavelengths in the studied soil samples.

2.5. Derivation of spectrotransfer functions (STFs)

The SMLR approach was used to develop STFs according to the maximum correlation between the studied soil properties and the effective spectral reflectance single bands. The following SMLR model was used to develop the STFs:

$$Soil\ attribute = a_0 + a_1R_{s1} + a_2R_{s2} + \ldots + a_mR_{sm} \quad (6)$$

where R_{s1} to R_{sm} are the effective and important spectral reflectance single bands, a_0 to a_m are regression coefficients, and m is the number of appeared spectral reflectance single bands in the STF. It should be noted that the input independent variables of the above-mentioned equation were the recognized effective spectral reflectance bands obtained from PLSR analysis. In the present study, the forward SMLR method was used. In this method, the variables were added according to a tolerance significance level, based on the *F*-test probability, which was set to 0.05. It should be noted that, the *F*-test is used for comparing variance of two statistical communities and testing whether the developed regression model between dependent and independent variables is significant or not. For performing this test, at first, the F statistic would be calculated based on the standard deviation of the observed and predicted values of a property. Then, the calculated F is compared with F statistic in the Fisher table at a specific probability level (0.05 in present study). If the calculated F be lesser than Fisher F statistic, the zero-significance assumption of regression model is accepted and vice versa. In other words, the developed regression model by independent variables can define the changes in dependent variable at a specific confidence level (i.e., 95 % in the present study). To put it simply, in STATISTICA software package (version 8.0) that we used for developing STFs, it is possible to easily recognize those X variables that have statistical significant relationship with Y variable in multiple linear regression by performing *F*-test. For instance, if the calculated p-level values be <0.05, for a specific X variable (i.e., appeared single bands in the developed STF), therefore the relationship between that X variable and the target attribute (Y) is statistically significant. All p-level values were calculated during developing the STFs in STATISTICA (version 8.0) software package. To develop STFs, the same (constant) calibration dataset with PLSR analysis, was used. After that, the developed models were tested using the validation dataset (same as the PLSR analysis). Generally, Fig. 2 illustrates the steps for preparing datasets, PLSR procedure, and developing STFs to predict soil properties.

2.6. Evaluation of the models

Some goodness of fit criteria including the determination coefficient (R^2), normalized root mean square error (NRMSE), Nash-Sutcliffe coefficient (NS), ratio of performance to interquartile range (RPIQ), and mean error (ME) were used to evaluate the models performance (Eqs. (7) to (11)).

$$R^2 = \frac{\left(\sum_{i=1}^{n}(O_i - \overline{O})(P_i - \overline{P})\right)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2 \sum_{i=1}^{n}(P_i - \overline{P})^2} \quad (7)$$

$$NRMSE\ (\%) = \frac{\sqrt{\sum_{i=1}^{n}\frac{(O_i - P_i)^2}{n}}}{\overline{O}} \times 100 \quad (8)$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \quad (9)$$

$$RPIQ = \frac{IQ}{RMSE} = \frac{Q_3 - Q_1}{\sqrt{\sum_{i=1}^{n}\frac{(O_i - P_i)^2}{n}}} \quad (10)$$

$$ME = \frac{\sum_{i=1}^{n}P_i - O_i}{n} \quad (11)$$

where O_i and P_i are the observed and predicted values, respectively; n is the number of data; and Q_3 and Q_1 are the third (75 %) and first (25 %) quartiles of samples, respectively. The NRMSE, which is a measure of error between the observed and the predicted values, varies between zero (the highest accuracy) to infinity. The NRMSE (%) ranges of 0–10, 10–20, 20–30 and > 30 are classified as excellent, good, fair, and poor classes, respectively (Bannayan and Hoogenboom, 2009). The NS statistic, which is a measure of model efficiency, changes from one (the highest accuracy) to negative infinity (Ritter and Muñoz-Carpena, 2013; Feng et al., 2017). According to Ritter and Muñoz-Carpena (2013), the NS values of 0.9–1, 0.8–0.9, 0.65–0.8, and <0.65 are classified as very good, good, acceptable, and unsatisfactory, respectively. The RPIQ is based on quartiles, which better represents the spread of the population.

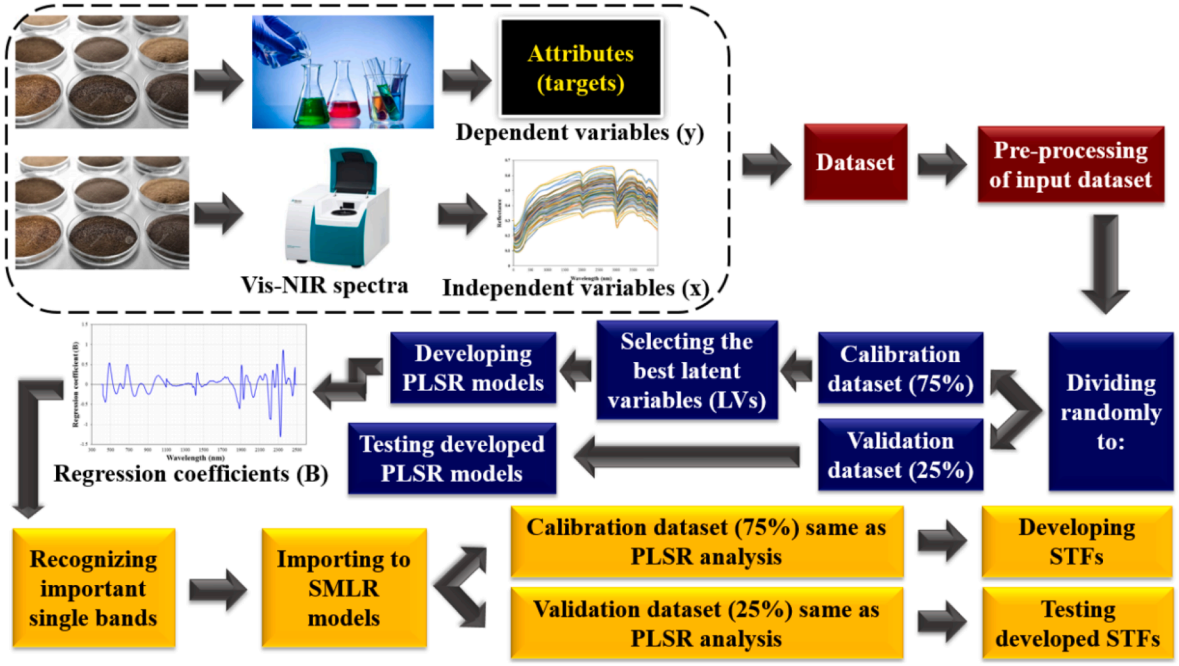

Fig. 2. Flowchart of the steps for preparing the dataset, the partial least square regression (PLSR) procedure, and developing spectrotransfer functions (STFs) to predict soil properties by Vis-NIR reflectance spectra.

This statistic is commonly used as a better indication for non-normal variables (Briedis et al., 2020) and should fall between 1.5 and 3 for an acceptable model (Jia et al., 2017). Generally, we considered five levels of accuracy based on R^2 values including: excellent ($0.90 \leq R^2 < 1$); very good ($0.75 \leq R^2 < 0.90$); good ($0.65 \leq R^2 < 0.75$); acceptable, moderate, or fair ($0.50 \leq R^2 < 0.65$); and poor ($R^2 < 0.50$). Excel (version 2013) software package was used for curves fitting and calculating the mentioned statistics.

3. Results

3.1. Descriptive statistics and correlation between properties

Table 2 shows the summary statistics and normality test criteria of the studied soil properties. The lowest and the highest coefficient of variations (CV) were related to pH (4.54 %) and Mg (393 %), respectively. According to Wilding (1985), pH and ln(Ca) fall in the low variability class ($0 < CV \leq 15$ %); $(sand)^{0.5}$, silt, CCE, ln(Fe), and ln(Zn) fall in the moderate variability class ($15 < CV \leq 35$ %); and the other studied soil properties fall in the high variability class ($CV > 35$ %). The CV values of 0.1 to 109 % (Moosavi and Sepaskhah, 2012) and 2 to 92 % (Rezaee et al., 2020a) have been reported for physicochemical/hydraulic properties of gravelly calcareous and paddy soils, respectively. Furthermore, Rezaee et al. (2020b) reported the CV values of 2 to 56 % for physicochemical attributes of their studied paddy soils.

As the Kolmogorov-Smirnov normality test shows, the sand, EC, Na, K, Ca, Mg, Fe, and Zn did not follow a normal distribution. They became close to normal distribution by the mentioned different transformations (Table 2).

Furthermore, Table 3 shows the Pearson correlation coefficients (r) among the studied soil properties. It should be noted that, in the present study, we performed Pearson correlation test between the normalized forms of variables, which is the first assumption for performing linear regression and correlation test. As can be seen, the ln(Fe), Mn, Cu, and ln(Zn) are positively correlated with SOM ($r > 0.4$) and negatively with CCE (with r values between −0.19 to −0.33). In addition, clay had a positive and significant correlation with Cu and ln(Fe) with r values of 0.27 and 0.45, respectively. According to literature, particle-size distribution (PSD), lime ($CaCO_3$) and SOM are well-known soil properties affecting the soil spectral reflectance data across the full range of spectra (Viscarra Rossel, 2007; Žížala et al., 2017; Ostovari et al., 2018). Therefore, it can be expected that the DTPA-extractable micronutrients be predicted with reasonable accuracy using spectroscopy-based methods.

3.2. PLSR analysis to predict soil properties using Vis-NIR reflectance spectra

3.2.1. Sand, silt, and clay contents

The relatively same standard deviation values of the calibration and the validation datasets for predicting the studied soil properties showed that the selected validation data were not concentrated in a specific small range and were able to test full range of the data. Fig. 3 shows the R^2_{cv} values of PLSR analysis for predicting $sand^{0.5}$, silt, and clay contents using Vis-NIR spectral reflectance bands by applying different values of LVs. The optimum numbers of LVs for predicting $sand^{0.5}$, silt, and clay contents were 11, 13, and 9, respectively.

Fig. 4 shows the predicted versus observed $sand^{0.5}$, silt, and clay contents using Vis-NIR spectral reflectance bands (left panel) along with their plots of regression coefficients (B) versus wavelengths (right panel) according to PLSR analysis. As can be seen, the $sand^{0.5}$, silt, and clay were predicted with R^2_{val} (the subscript val shows the validation dataset) values of 0.83, 0.67, and 0.81; $NRMSE_{val}$ values of 16.3, 11.1, and 17.0 %; NS_{val} values of 0.80, 0.67, and 0.81; $RPIQ_{val}$ values of 3.31, 2.32, and 4.26; and ME_{val} values of 0.18 $\%^{0.5}$, −0.47 %, and −0.62 %, respectively.

Table 2
Summary statistics and fitting parameters of the normal distribution for the studied soil properties.

Property†	Unit	N	Min	Max	Mean	SD	CV (%)	VC	SK	KR	KS††
Sand	%	234	2.28	76.8	17.4	11.9	68.4	High	1.59	4.75	0.110**
(Sand) $^{0.5}$	(%) $^{0.5}$	234	1.51	8.76	3.94	1.38	35.0	Moderate	0.36	0.30	0.087 ns
Silt	%	234	23.2	91.0	55.5	11.0	19.8	Moderate	0.42	1.06	0.086 ns
Clay	%	234	4.88	56.8	27.3	11.3	41.4	High	−0.14	−0.86	0.085 ns
pH	–	234	6.60	8.21	7.61	0.35	4.60	Low	−0.40	−0.38	0.081 ns
EC	$dS\ m^{-1}$	234	0.26	33.3	1.34	3.19	238	High	7.17	59.2	0.367**
EC $^{-1}$	$dS^{-1}\ m$	234	0.03	3.90	1.61	0.81	50.3	High	0.27	−0.02	0.053 ns
CCE	%	234	12.5	70.6	42.6	9.74	22.9	Moderate	0.23	0.81	0.081 ns
Na	$mg\ L^{-1}$	234	3.00	676	55.2	90.0	163	High	3.98	19.5	0.251**
ln(Na)	$ln(mg\ L^{-1})$	234	1.10	6.52	3.30	1.16	35.2	High	0.28	−0.07	0.058 ns
K	$mg\ L^{-1}$	234	1.32	135	16.4	19.0	116	High	3.18	13.4	0.218**
ln(K)	$ln(mg\ L^{-1})$	234	0.28	4.91	2.38	0.87	36.6	High	0.48	−0.07	0.082 ns
Ca	$mg\ L^{-1}$	234	32.0	944	140	131	93.6	High	3.43	14.4	0.219**
ln(Ca)	$ln(mg\ L^{-1})$	234	2.48	6.80	4.68	0.63	13.5	Low	0.94	1.03	0.077 ns
Mg	$mg\ L^{-1}$	234	4.80	3456	63.5	250	394	High	11.5	150	0.407**
ln(Mg)	$ln(mg\ L^{-1})$	234	1.57	8.15	3.21	1.13	35.2	High	0.57	1.40	0.111 ns
Fe	$mg\ kg^{-1}$	161	1.25	49.1	15.8	10.7	67.7	High	0.90	0.06	0.132**
ln(Fe)	$ln(mg\ kg^{-1})$	161	0.83	3.89	2.55	0.70	27.5	Moderate	−0.33	−0.41	0.070 ns
Mn	$mg\ kg^{-1}$	161	1.56	26.0	10.3	5.83	56.6	High	0.78	0.12	0.084 ns
Cu	$mg\ kg^{-1}$	161	0.21	3.77	1.82	0.83	45.6	High	0.42	−0.08	0.068 ns
Zn	$\times 10^2\ mg\ kg^{-1}$	161	9.20	243	52.4	40.5	77.3	High	2.17	5.98	0.166**
ln(Zn)	$ln(\times 10^2\ mg\ kg^{-1})$	161	2.22	5.49	3.73	0.66	17.7	Moderate	0.31	−0.21	0.081 ns
SOM	%	102	0.17	4.26	1.74	1.21	69.5	High	0.73	−0.06	0.104 ns

†: Sand: sand content; $Sand^{0.5}$: square root of sand content; Silt: silt content; Clay: clay content; pH: pH of saturated paste; EC: electrical conductivity of saturated extract; EC^{-1}: inverse of electrical conductivity of saturated extract; CCE: calcium carbonate equivalent (lime); Na, K, Ca, and Mg: water-soluble sodium, potassium, calcium, and magnesium, respectively; ln(Na), ln(K), ln(Ca), and ln(Mg): natural logarithms (ln) of water-soluble sodium, potassium, calcium, and magnesium, respectively; Fe, Mn, Cu, and Zn: DTPA-extractable iron, manganese, copper, and zinc, respectively; ln(Fe) and ln(Zn): natural logarithms of DTPA-extractable iron and zinc, respectively; SOM: soil organic matter.

††: N, Min., Max., Mean, SD, CV, VC, SK, KR, and KS are the number of soil samples, minimum, maximum, mean, standard deviation, coefficient of variations, variability classes according to Wilding (1985), skewness coefficient, kurtosis coefficient, and the statistics of Kolmogorov-Smirnov normality test, respectively. ns means lack of significant difference with the normal distribution, ** and * show significant differences with the normal distribution (departure from normal distribution) at the probability levels of 1 and 5 %, respectively (which became normal with the mentioned different transformation).

Table 3
Pearson correlation coefficients (r) between the studied soil properties.

	$(Sand)^{0.5}$	Silt	Clay	pH	EC^{-1}	CCE	ln(Na)	ln(K)	ln(Ca)	ln(Mg)	ln(Fe)	Mn	Cu	ln(Zn)
Silt	−0.83**													
Clay	−0.15	−0.39**												
pH	−0.01	0.03	−0.09											
EC^{-1}	0.12	−0.22*	0.27**	0.50**										
CCE	0.32**	−0.28**	−0.02	−0.08	−0.02									
ln(Na)	−0.10	0.23*	−0.31**	−0.36**	−0.75**	0.11								
ln(K)	0.09	0.05	−0.28**	−0.32**	−0.77**	0.04	0.66**							
ln(Ca)	−0.17	0.30**	−0.31**	−0.58**	−0.84**	0.02	0.89**	0.68**						
ln(Mg)	−0.14	0.24*	−0.27**	−0.49**	−0.66**	0.00	0.81**	0.48**	0.82**					
ln(Fe)	−0.16	−0.05	0.45**	−0.30**	0.19*	−0.26**	−0.38**	−0.36**	−0.25**	−0.18				
Mn	0.01	0.04	0.01	−0.30**	−0.27**	−0.33**	0.02	0.40**	0.21*	0.03	0.23*			
Cu	−0.10	−0.00	0.27**	−0.25*	−0.08	−0.32**	0.01	0.05	0.05	0.02	0.66**	0.33**		
ln(Zn)	0.06	0.06	−0.16	−0.18	−0.30**	−0.19	0.26**	0.56**	0.31**	0.15	0.02	0.49**	0.37**	
SOM	−0.02	−0.05	0.23*	−0.34**	−0.24*	−0.21*	−0.02	0.35**	0.12	−0.04	0.44**	0.74**	0.47**	0.54**

†: $Sand^{0.5}$: square root of sand content; Silt: silt content; Clay: clay content; pH: pH of saturated paste; EC^{-1}: inverse of electrical conductivity of saturated extract; CCE: calcium carbonate equivalent (lime); ln(Na), ln(K), ln(Ca), and ln(Mg): natural logarithms (ln) of water-soluble sodium, potassium, calcium, and magnesium, respectively; ln(Fe) and ln(Zn): natural logarithms of DTPA-extractable iron and zinc, respectively; Mn and Cu: DTPA-extractable manganese and copper, respectively; SOM: soil organic matter. ††: * and ** demonstrate the significant correlation between properties in probability levels of 5 and 1 %, respectively.

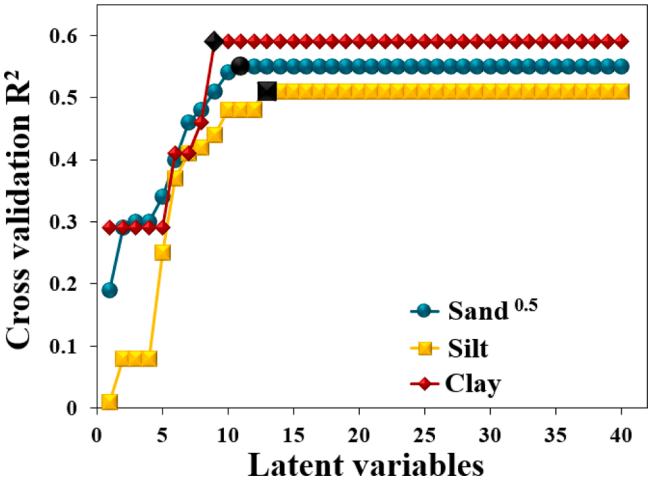

Fig. 3. The cross validation determination coefficient (R^2) values of the partial least square regression (PLSR) analysis for developing models to predict square root of sand content, $sand^{0.5}$; silt; and clay contents using Vis-NIR spectral reflectance bands by applying 1–40 latent variables (in each case the bold black spot shows the best number of latent variables).

3.2.2. pH, electrical conductivity (EC), and calcium carbonate equivalent (CCE)

According to Fig. 5, in PLSR analysis using Vis-NIR reflectance spectra, the optimum numbers of LVs to predict pH, EC^{-1}, and CCE were 17, 18, and 10, respectively.

The capability of PLSR method to predict pH, EC^{-1}, and CCE properties using Vis-NIR spectral reflectance bands along with the goodness of fit criteria and plots of B values versus wavelengths are shown in Fig. 6. The PLSR method predicted pH, EC^{-1}, and CCE by R^2_{val} values of 0.71, 0.63, and 0.82; $NRMSE_{val}$ values of 2.32, 35.8, and 9.67 %; NS_{val} values of 0.71, 0.63, and 0.80; $RPIQ_{val}$ values of 3.34, 2.17, and 2.52; and ME_{val} values of −0.02, 0.03 dS^{-1} m, and 0.72 %, respectively.

3.2.3. Water-soluble sodium (Na), potassium (K), calcium (Ca), and magnesium (Mg)

The R^2_{cv} values showed that the optimum numbers of LVs for predicting ln(Na), ln(K), ln(Ca), and ln(Mg) using PLSR and Vis-NIR spectral reflectance data were 19, 13, 24, and 11, respectively (Fig. 7).

According to Fig. 8, the PLSR method using Vis-NIR spectral reflectance bands predicted ln(Na), ln(K), ln(Ca), and ln(Mg) by R^2_{val} values of 0.74, 0.70, 0.62, and 0.71; $NRMSE_{val}$ values of 17, 20, 8.08, and 16.8 %; NS_{val} values of 0.73, 0.68, 0.62, and 0.67; $RPIQ_{val}$ values of 2.65, 2.43, 2.21, and 2.36; and ME_{val} values of −0.13, −0.11, −0.02, and −0.20 ln (mg L^{-1}), respectively.

3.2.4. DTPA-extractable iron (Fe), manganese (Mn), copper (Cu), and zinc (Zn)

As can be seen in Fig. 9, the optimum numbers of LVs for predicting ln(Fe), Mn, Cu, and ln(Zn) using PLSR method and Vis-NIR reflectance spectra were 17, 20, 15, and 11, respectively.

Fig. 10 demonstrates scatter plots of the observed versus predicted values of ln(Fe), Mn, Cu, and ln(Zn) and models performance using Vis-NIR spectral reflectance bands by applying PLSR method (left panel) along with plots of B values versus wavelengths (right panel). The models were developed by R^2_{val} values of 0.74, 0.70, 0.74, and 0.69; $NRMSE_{val}$ values of 14.8, 32.4, 22.7, and 10.4 %; NS_{val} values of 0.71, 0.70, 0.72, and 0.69; $RPIQ_{val}$ values of 2.96, 2.32, 3.37, and 2.70; and ME_{val} values of −0.03 ln(mg kg^{-1}), −0.02 mg kg^{-1}, −0.10 mg kg^{-1}, and −0.02 ln(mg kg^{-1}) for predicting ln(Fe), Mn, Cu, and ln(Zn), respectively.

3.3. Developing STFs to predict soil properties using Vis-NIR reflectance spectra

3.3.1. Sand, silt, and clay contents

The B values (Fig. 4), which were obtained from PLSR analysis, demonstrated several effective and important spectral reflectance bands in full range of spectra (420–2480 nm) to predict soil textural components (i.e., $sand^{0.5}$, silt, and clay contents) and to develop STFs. Based on the obtained effective bands from PLSR analysis, the following STFs (Eqs. (12) to (14)) were developed using 13, 17, and 13 spectral reflectance bands and SMLR method to predict $sand^{0.5}$, silt, and clay, respectively.

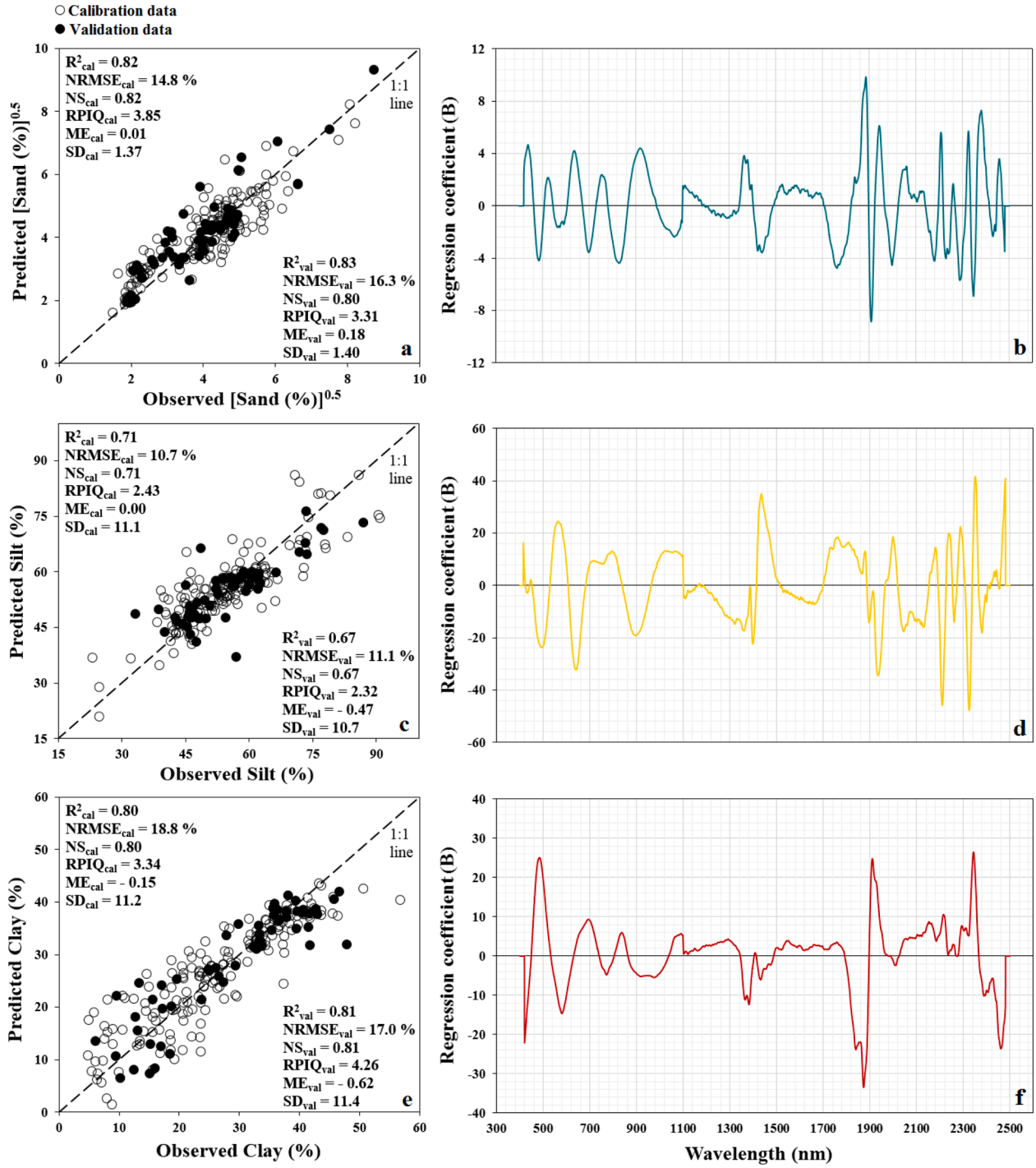

Fig. 4. The partial least square regression (PLSR) analysis results, including scatter plots, goodness of fit criteria, and plots of regression coefficients (B) versus wavelengths, to predict square root of sand content, sand$^{0.5}$ (a and b); silt (c and d); and clay (e and f) contents using Vis-NIR spectral reflectance data (the subscripts cal and val show the calibration and validation datasets, respectively).

$$[Sand\ (\%)]^{0.5} = 4.2 + 195\ (R_{438}) - 207\ (R_{485}) - 30.2\ (R_{525}) + 318\ (R_{636}) - 341\ (R_{698}) + 90.9\ (R_{919}) - 16.8\ (R_{1943}) + 49.2\ (R_{2207}) - 261\ (R_{2288}) + 234\ (R_{2325}) - 256\ (R_{2346}) + 278\ (R_{2380}) - 45.9\ (R_{2480}) \qquad \begin{matrix} i = 25 \\ m = 13 \end{matrix} \tag{12}$$

$$\begin{aligned} Silt\ (\%) = {} & 54.1 + 1885\ (R_{420}) - 2484\ (R_{496}) + 1545\ (R_{562}) - 795\ (R_{641}) \\ & +395\ (R_{794}) - 143\ (R_{897}) - 4545\ (R_{1398}) + 4347\ (R_{1434}) + 2475\ (R_{1762}) \quad i = 22 \\ & -2341\ (R_{1934}) + 882\ (R_{1998}) - 981\ (R_{2044}) - 2649\ (R_{2130}) + 634\ (R_{2180}) \quad m = 17 \\ & +530\ (R_{2351}) - 609\ (R_{2379}) + 2228\ (R_{2480}) \end{aligned} \tag{13}$$

$$\begin{aligned} Clay\ (\%) = {} & 26.5 - 1932\,(R_{420}) + 2495\,(R_{484}) - 1414\,(R_{580}) + 1009\,(R_{693}) \\ & -667\,(R_{836}) + 1096\,(R_{1381}) - 3554\,(R_{1872}) + 1032\,(R_{1910}) \quad i = 13 \\ & +3151\,(R_{2152}) - 502\,(R_{2214}) + 897\,(R_{2343}) - 349\,(R_{2388}) \quad m = 13 \\ & -1398\,(R_{2460}) \end{aligned} \tag{14}$$

where, R_s, i, and m show the spectral reflectance value at wavelength s, number of the selected effective single bands from PLSR analysis (B plot) which were imported to SMLR model, and number of the appeared effective single bands in the developed STF, respectively. Capability of the STFs to predict sand$^{0.5}$, silt, and clay is shown in Fig. 11. The developed STFs predicted sand$^{0.5}$, silt, and clay by R^2_{val} values of 0.78, 0.62, and 0.81; $NRMSE_{val}$ values of 17.3, 11.9, and 17.3 %; NS_{val} values of 0.77, 0.62, and 0.80; $RPIQ_{val}$ values of 3.11, 2.16, and 4.18; and ME_{val} values of 0.11 %$^{0.5}$, −0.43 %, and −0.62 %, respectively.

3.3.2. pH, EC, and CCE

As can be seen in Fig. 6 (right panel), the important and effective bands for predicting pH, EC^{-1}, and CCE were mostly in wavelength ranges of 420–1100 and 1900–2480 nm and around 1400 nm. Using the effective spectral reflectance bands based on the B values and SMLR method, the following STFs (Eqs. (15) to (17)) were developed to predict pH, EC^{-1}, and CCE. As it is demonstrated, 11, 16, and 15 spectral reflectance bands were appeared in the STFs to predict pH, EC^{-1}, and CCE, respectively.

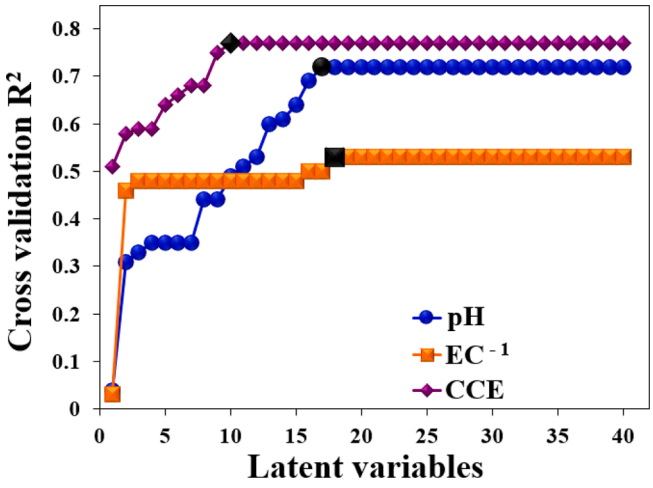

Fig. 5. The cross validation determination coefficient (R^2) values of the partial least square regression (PLSR) analysis for developing models to predict pH of saturated paste, pH; inverse of electrical conductivity of soil saturated extract, EC^{-1}; and calcium carbonate equivalent, CCE using Vis-NIR spectral reflectance bands by applying 1–40 latent variables (in each case the bold black spot shows the best number of latent variables).

$$\begin{aligned} pH = {} & 8.34 - 35.5\ (R_{442}) + 35.5\ (R_{488}) - 25.5\ (R_{637}) + 15.7\ (R_{680}) + 30.8\ (R_{1497}) \\ & -36.8\ (R_{1885}) + 8.96\ (R_{1967}) - 40.7\ (R_{2220}) + 76.2\ (R_{2290}) - 118\ (R_{2325}) \quad i = 24 \\ & +84.2\ (R_{2352}) \quad m = 11 \end{aligned} \tag{15}$$

$$\begin{aligned} EC^{-1}\ (dS^{-1}\ \mathrm{m}) = {} & 1.23 + 0.142\ (R_{450}) + 70\ (R_{682}) - 113\ (R_{757}) + 114\ (R_{1386}) + 116\ (R_{1426}) \\ & -92.2\ (R_{1494}) - 203\ (R_{1877}) + 15.5\ (R_{1911}) + 45\ (R_{1970}) - 133\ (R_{2035}) \quad i = 20 \\ & +231\ (R_{2149}) - 68.3\ (R_{2208}) + 121\ (R_{2290}) - 177\ (R_{2327}) + 141\ (R_{2351}) \quad m = 16 \\ & -68.9\ (R_{2452}) \end{aligned} \tag{16}$$

$$\begin{aligned} CCE\,(\%) = {} & 1.15 - 928\,(R_{420}) + 1017\,(R_{500}) - 615\,(R_{584}) + 1352\,(R_{684}) \\ & -1134\,(R_{789}) + 763\,(R_{1105}) - 1695\,(R_{1386}) + 1172\,(R_{1428}) \quad i = 18 \\ & +185\,(R_{1796}) - 387\,(R_{1927}) + 110\,(R_{2222}) - 343\,(R_{2335}) \quad m = 15 \\ & +80.7\,(R_{2372}) + 1650\,(R_{2412}) - 1267\,(R_{2480}) \end{aligned} \tag{17}$$

The scatter plots of the observed versus predicted values along with goodness of fit criteria for predicting pH, EC^{-1}, and CCE using STFs are shown in Fig. 12. These properties were predicted by their related STFs with R^2_{val} values of 0.66, 0.55, and 0.83; $NRMSE_{val}$ values of 2.51, 39.9, and 9.97 %; NS_{val} values of 0.66, 0.53, and 0.80; $RPIQ_{val}$ values of 3.09, 1.92, and 2.40; and ME_{val} values of −0.01, 0.10 dS^{-1} m, and 1.07 %, respectively.

3.3.3. Water-soluble Na, K, Ca, and Mg

There were several effective bands in full range of spectra for predicting natural logarithm forms of studied water-soluble cations when PLSR method was used (Fig. 8). According to Eqs. (18) to (21), the numbers of effective spectral reflectance bands appeared in STFs were 14, 15, 16, and 15 for predicting ln(Na), ln(K), ln(Ca), and ln(Mg), respectively.

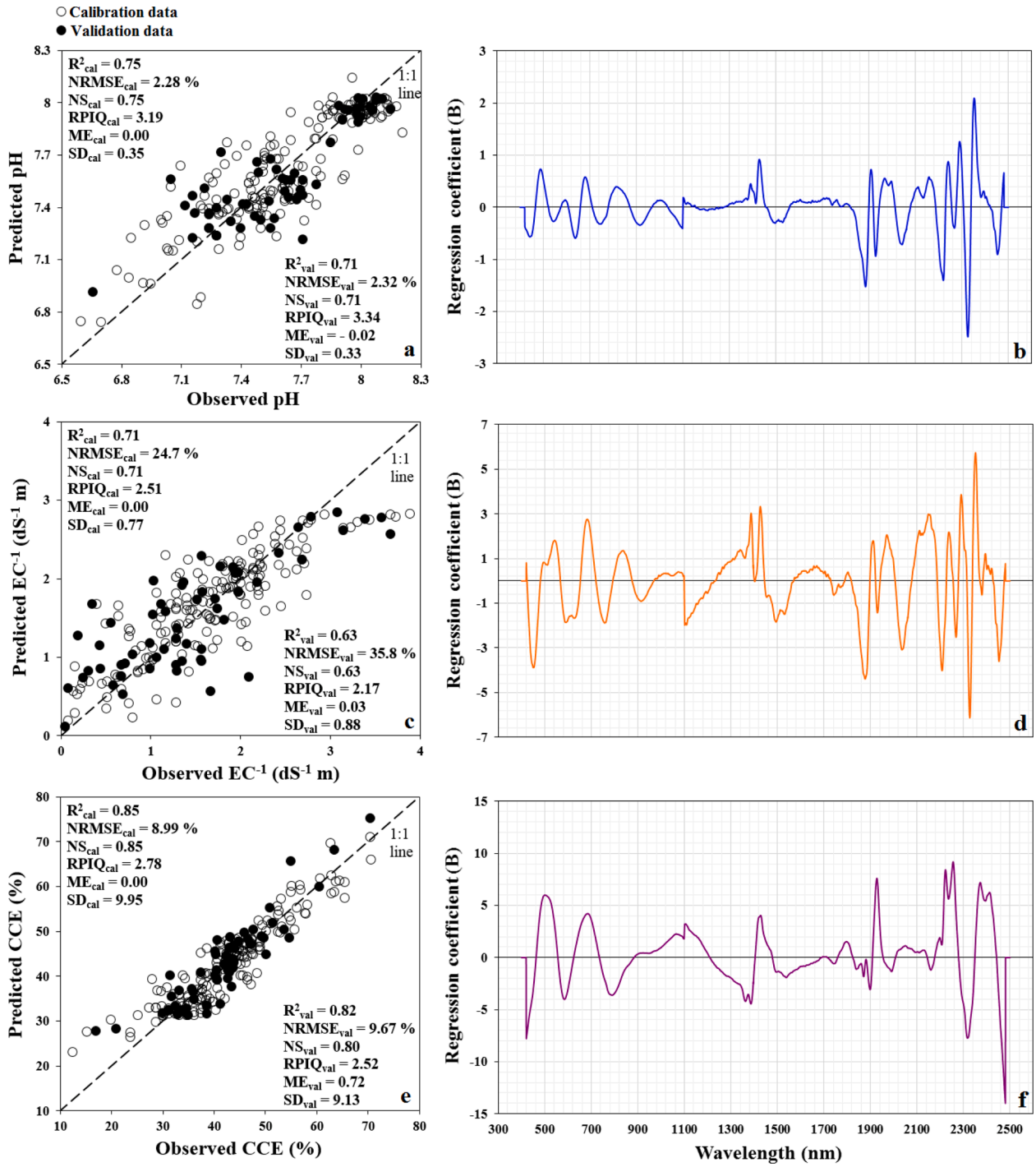

Fig. 6. The partial least square regression (PLSR) analysis results, including scatter plots, goodness of fit criteria, and plots of regression coefficients (B) versus wavelengths, to predict pH of saturated paste, pH (a and b); inverse of electrical conductivity of soil saturated extract, EC^{-1} (c and d); and calcium carbonate equivalent, CCE (e and f) values using Vis-NIR spectral reflectance data (the subscripts cal and val show the calibration and validation datasets, respectively).

$$\begin{aligned}\ln\left[Na\ \left(mg\ L^{-1}\right)\right] &= 2.69 + 12.9\ (R_{452}) + 122\ (R_{586}) - 498\ (R_{682}) + 416\ (R_{757}) \\ &+14\ (R_{905}) - 311\ (R_{1386}) + 225\ (R_{1534}) - 105\ (R_{1804}) \qquad i = 28 \\ &+57.8\ (R_{1920}) + 146\ (R_{2045}) - 79.9\ (R_{2352}) + 266\ (R_{2378}) \qquad m = 14 \\ &-302\ (R_{2417}) + 37.6\ (R_{2453})\end{aligned} \tag{18}$$

$$\ln\left[K\ \left(mg\ L^{-1}\right)\right] = -0.6 - 48.5\ (R_{969}) + 108\ (R_{1098}) - 108\ (R_{1431}) - 252\ (R_{1697}) + 370\ (R_{1863}) + 175\ (R_{1910}) - 110\ (R_{1922}) - 20.6\ (R_{1943}) + 107\ (R_{2163}) + 102\ (R_{2208}) - 276\ (R_{2226}) - 301\ (R_{2312}) + 395\ (R_{2328}) - 217\ (R_{2350}) + 79\ (R_{2425}) \quad i = 31,\ m = 15 \tag{19}$$

$$\ln\left[Ca\ \left(mg\ L^{-1}\right)\right] = 5.13 + 57.4\ (R_{448}) - 63\ (R_{489}) + 20.2\ (R_{634}) + 3.9\ (R_{685}) - 47.5\ (R_{1385}) + 177\ (R_{1403}) - 190\ (R_{1427}) - 118\ (R_{1803}) + 259\ (R_{1875}) - 44.6\ (R_{1972}) - 11.3\ (R_{2199}) - 33.9\ (R_{2270}) - 116\ (R_{2312}) + 210\ (R_{2327}) - 165\ (R_{2350}) + 65.7\ (R_{2453}) \quad i = 25,\ m = 16 \tag{20}$$

$$\ln\left[Mg\ \left(mg\ L^{-1}\right)\right] = 2.66 - 258\ (R_{420}) + 337\ (R_{477}) - 110\ (R_{538}) - 308\ (R_{699}) + 361\ (R_{743}) - 157\ (R_{975}) + 212\ (R_{1093}) - 416\ (R_{1384}) + 207\ (R_{1495}) + 373\ (R_{1748}) - 326\ (R_{1837}) + 146\ (R_{1902}) - 236\ (R_{1975}) + 181\ (R_{2048}) - 31.4\ (R_{2257}) \quad i = 24,\ m = 15 \tag{21}$$

To predict ln(Na), ln(K), ln(Ca), and ln(Mg) using the STFs, the R^2_{val} values were 0.63, 0.65, 0.60, and 0.66; $NRMSE_{val}$ values were 20.7, 21.5, 8.34, and 18.3 %; NS_{val} values were 0.60, 0.63, 0.60, and 0.61; $RPIQ_{val}$ values were 2.18, 2.26, 2.14, and 2.17; and ME_{val} values were −0.18, −0.12, 0.00, and −0.24 ln(mg L^{-1}), respectively (Fig. 13).

3.3.4. DTPA-extractable Fe, Mn, Cu, and Zn

The B values demonstrated the effective spectral reflectance bands for predicting ln(Fe), Mn, Cu, and ln(Zn) were mostly in the wavelength ranges of 420–900 and 1900–2480 nm and around 1400 nm. The developed STFs (Eqs. (22) to (25)) showed that the numbers of the most important spectral reflectance bands for predicting ln(Fe), Mn, Cu, and ln(Zn) were 12, 17, 15, and 11, respectively.

$$\ln\left[Fe\ \left(mg\ kg^{-1}\right)\right] = 3.95 + 65.5\ (R_{420}) - 84.8\ (R_{451}) + 77\ (R_{582}) - 86\ (R_{678}) + 58.2\ (R_{890}) - 83.4\ (R_{1426}) + 57.5\ (R_{1922}) - 54.8\ (R_{2166}) + 105\ (R_{2222}) + 58\ (R_{2315}) - 42\ (R_{2349}) - 76.1\ (R_{2480}) \quad i = 20,\ m = 12 \tag{22}$$

$$\mathrm{Mn}\left(\mathrm{mg\,kg^{-1}}\right) = 21.1 + 945\,(R_{420}) - 1136\,(R_{489}) + 608\,(R_{569}) - 520\,(R_{662}) + 564\,(R_{720}) - 543\,(R_{811}) + 553\,(R_{1154}) - 915\,(R_{1389}) + 891\,(R_{1601}) - 183\,(R_{1747}) + 834\,(R_{1894}) - 6.29\,(R_{1926}) - 943\,(R_{2077}) + 418\,(R_{2199}) - 746\,(R_{2255}) + 548\,(R_{2318}) - 261\,(R_{2371}) \quad i = 19,\ m = 17 \tag{23}$$

$$\mathrm{Cu}\ \left(\mathrm{mg\ kg^{-1}}\right) = 2.19 + 51.6\ (R_{447}) - 47.6\ (R_{497}) - 34.3\ (R_{786}) + 92.9\ (R_{882}) - 121\ (R_{1103}) + 201\ (R_{1404}) - 106\ (R_{1429}) - 45.4\ (R_{1779}) - 68.7\ (R_{2220}) - 57\ (R_{2243}) + 138\ (R_{2272}) + 111\ (R_{2340}) - 253\ (R_{2385}) + 330\ (R_{2431}) - 188\ (R_{2480}) \quad i = 26,\ m = 15 \tag{24}$$

$$\ln\left[Zn\ \left(\times 10^2\ \mathrm{mg\ kg^{-1}}\right)\right] = 7.39 + 130\ (R_{442}) - 137\ (R_{488}) + 27.7\ (R_{548}) - 18\ (R_{666}) - 6.79\ (R_{1405}) + 49.1\ (R_{1472}) + 54.2\ (R_{1900}) - 57.8\ (R_{1927}) - 54.8\ (R_{2155}) - 36.3\ (R_{2314}) + 58.8\ (R_{2480}) \quad i = 18,\ m = 11 \tag{25}$$

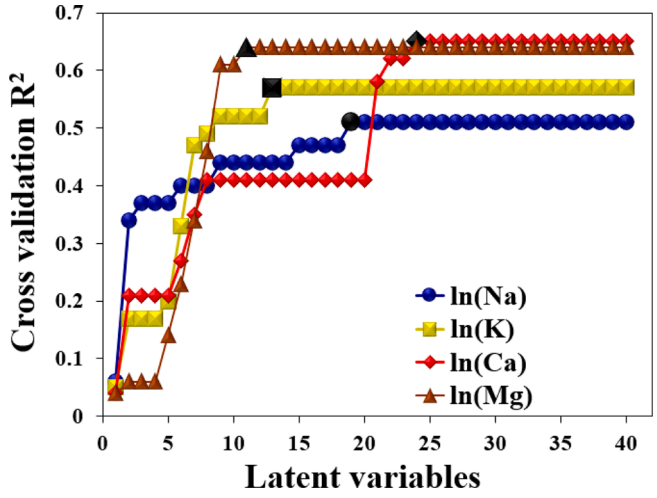

Fig. 7. The cross validation determination coefficient (R^2) values of the partial least square regression (PLSR) analysis for developing models to predict natural logarithm forms of water-soluble sodium, ln(Na); potassium, ln(K); calcium, ln (Ca); and magnesium, ln(Mg) using Vis-NIR spectral reflectance bands by applying 1–40 latent variables (in each case the bold black spot shows the best number of latent variables).

Results showed that accuracy of the STFs for predicting the mentioned DTPA-extractable micronutrients were close to PLSR results (Fig. 14). As can be seen, developed STFs had R^2_{val} values of 0.70, 0.70, 0.71, and 0.68; $NRMSE_{val}$ values of 15.2, 32.2, 23.4, and 10.6 %; NS_{val} values of 0.69, 0.70, 0.70, and 0.67; $RPIQ_{val}$ values of 2.87, 2.34, 3.27, and 2.65; and ME_{val} values of −0.04 ln(mg kg^{-1}), −0.28 mg kg^{-1}, −0.07 mg kg^{-1}, and −0.03 ln(mg kg^{-1}) for predicting ln(Fe), Mn, Cu, and ln(Zn), respectively.

3.4. Comparing capability of the developed PLSR models and STFs to predict soil properties

The R^2_{val} values of the developed PLSR models and STFs (Eqs. (12) to (25)) using Vis-NIR reflectance spectra for predicting the studied soil properties were compared in Fig. 15. The maximum value of R^2_{val} (0.83) was observed for predicting $sand^{0.5}$ (by PLSR) and CCE (by STF) and the minimum value of R^2_{val} (0.55) was obtained for predicting EC^{-1} (by STF).

4. Discussion

4.1. Prediction of soil properties using spectroscopy-based methods

4.1.1. Sand, silt, and clay contents

According to the NRMSE classification guideline, the PLSR models and STFs predicted $sand^{0.5}$, silt, and clay contents with good accuracy ($NRMSE_{val}$ values of 10–20) using Vis-NIR reflectance spectra. In addition, based on the NS classification, the $sand^{0.5}$ (by PLSR model) and clay (by both PLSR model and STF) were predicted with good (NS_{val} values of 0.80–0.90); $sand^{0.5}$ (by STF) and silt (by PLSR model) with acceptable (NS_{val} values of 0.65–0.80); and silt (by STF) with unsatisfactory ($NS < 0.65$) accuracies. Generally, the values of R^2_{val} and $RPIQ_{val}$ illustrated very good predictions of $sand^{0.5}$ and clay (by both PLSR models and STFs); good prediction of silt (by PLSR model); and acceptable (fair or moderate) prediction of silt (by STF) using Vis-NIR reflectance spectra. A bit overestimation of $sand^{0.5}$ and a bit underestimation of silt and clay were observed when these texture components were predicted by both mentioned spectroscopy-based approaches due to positive and negative values of ME_{val}, respectively.

Normally, high content of clay and low content of sand may cause darker color in soil due to different mineralogical structure. In general, an object (e.g., soil) with darker color tends to has lesser value of Albedo coefficient. This is noteworthy to mention, the Albedo coefficient represents the degree of brightness and changes between 0 (absolute darkness) to 1 (absolute brightness). Therefore, different mineralogy and physico-chemical properties in soil directly affect soil color and consequently affect the absorption/reflection of electromagnetic waves in Vis-NIR region. As can be seen from Fig. 1, the mean reflectance spectra of the soil samples showed four significant absorption bands around 1416, 1915, 2212, and 2340 nm wavelengths. Gomez et al. (2008) and Ostovari et al. (2018) stated that these absorption bands are mainly related to the overtone of OH stretches (moisture adsorbed to clay surface), combination of OH stretches and H-O-H in water molecules trapped in the clay crystal lattice, O-H-metal band and OH stretch combinations, and CO_3^{2-} groups in carbonate minerals, respectively. As can be seen in Fig. 4, there are several significant B values around 1416, 1915, and 2212 nm wavelengths to predict soil texture components by PLSR approach. According to Eqs. (12) to (14), different reflectance bands around these wavelengths were appeared in the developed STFs to predict soil texture components, i.e., at 1943 and 2207 nm for $sand^{0.5}$; at 1398, 1434, 1934, and 2180 nm for silt; and at 1381, 1910, and 2214 nm for clay contents. Although the clay content was predicted with very good accuracy by spectroscopy-based methods (R^2_{val} values of 0.81 for both developed PLSR model and STFs), but still it is not excellent prediction due to dispersion of the predicted versus the observed points around 1:1 line (Fig. 4e and Fig. 11c). It may be due to the type of clay minerals. According to Khormali and Abtahi (2003), the major clay minerals in the study area (Fars Province) are illite and chlorite with some evidences to form smectites (in well drained Alfisols, somewhat poorly drained Mollisols, and Calcixerepts with high precipitation). They reported the mean value of cation exchange capacity (CEC) in soils of Fars Province as 43 $cmol_{(+)}\ kg^{-1}$. The type of major clay minerals and relatively low CEC values in the soils of study area show low amount of negative charge at clay surface unit, low adsorption capacity for bipolar water molecules, and consequently, low absorption of Vis-NIR spectra (especially around 1416, 1915, and 2212 nm wavelengths) by soil. Generally, with having high CEC value and high adsorption of water molecules at clay surface, the absorption of spectra at wavelengths of 1416, 1915, and 2212 nm can be high and consequently the prediction of clay content by Vis-NIR spectra can be accompanied with high accuracy. According to Stenberg et al. (2010) absorption between 1300 and 1500 nm mainly corresponds to hydroxyl groups in clay minerals and water; the strong absorption near to 1900 nm is due to water (hygroscopic water and water held within clay mineral structures); and absorption between 2100 and 2500 nm is due to the clay minerals (e.g., kaolinite, illite, and smectite), carbonates, and organic compounds. Mina et al. (2021) and Salehi-Varnousfaderani et al. (2022) found that the soils with higher content of clay showed lower values of spectral reflectance in Vis-NIR region. Regarding applying PLSR method by using Vis-NIR spectral reflectance bands to predict soil texture, Pinheiro et al. (2017), in the central Amazon, Brazil, accurately predicted clay ($R^2_{val} = 0.78$), while moderately predicted sand ($R^2_{val} = 0.62$) and poorly predicted silt ($R^2_{val} = 0.36$). Xu et al. (2018a) reported acceptable to good predictions of soil texture components ($0.57 \leq R^2 \leq 0.71$) in southeast of China. According to Hobley and Prater (2019) in soils of Germany, the PLSR method has more capability to predict clay and sand fractions, whereas random forest regression performs better for predicting silt fraction using Vis-NIR spectra.

On the other hand, Bilgili et al. (2010) observed the highest correlation between clay content and spectral reflectance at 450 nm wavelength in Turkish soils, which is in line with our results. Our results are also in line with the findings of Xu et al. (2018a) who reported that the effective spectral bands for predicting soil texture components are near to 480, 920, 1910, and 2200 nm for sand; 780 nm for silt; and in range of 410–580 nm and at 1410, 1900, 2200, and 2400 nm for clay. Hobley and Prater (2019) also reported that the region around 1910 and 2190 nm are effective bands to predict clay content. Conforti et al. (2018) observed the largest correlation coefficient between sand, silt, and clay

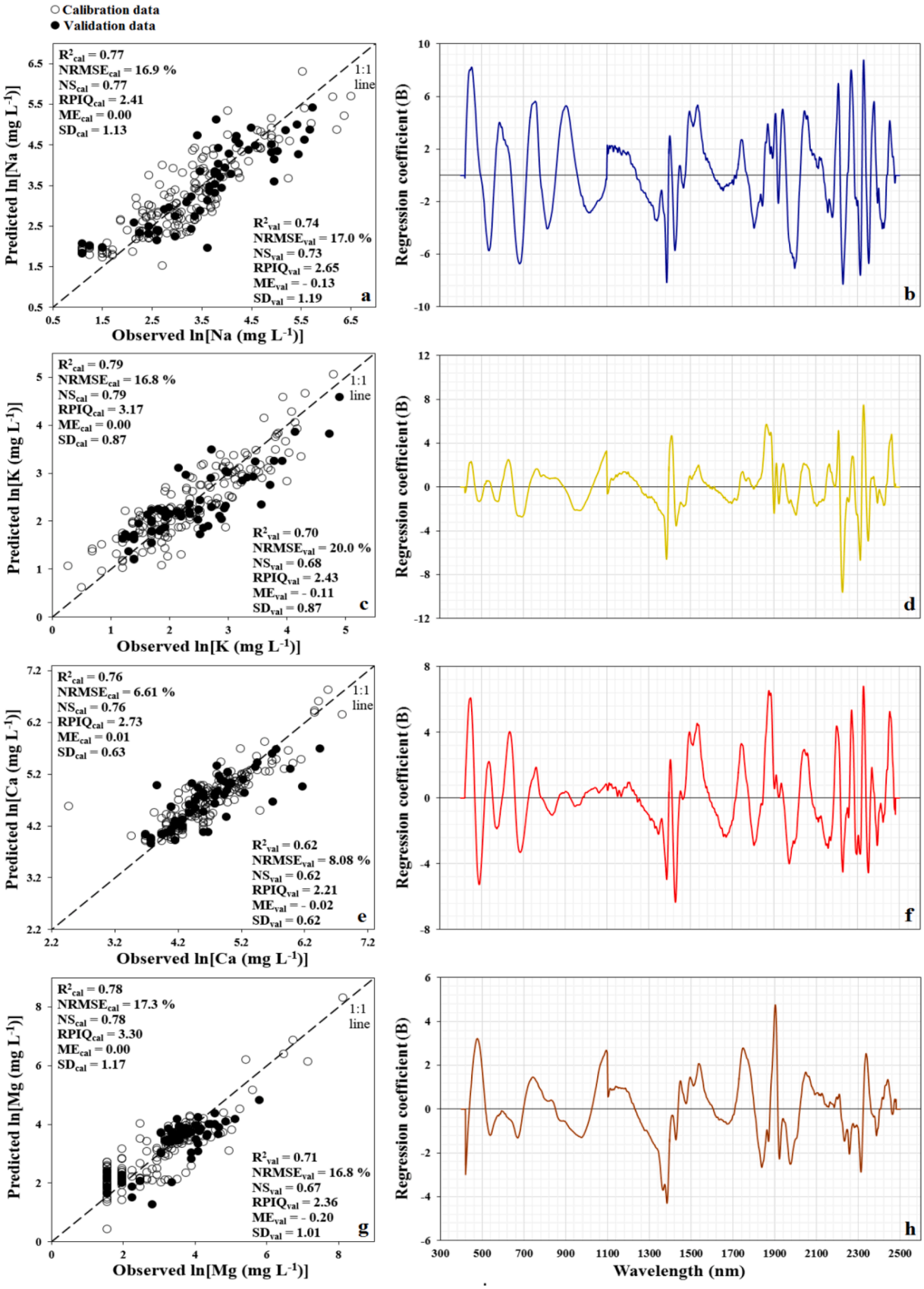

Fig. 8. The partial least square regression (PLSR) analysis results, including scatter plots, goodness of fit criteria, and plots of regression coefficients (B) versus wavelengths, to predict natural logarithm forms of water-soluble sodium, ln(Na) (a and b); potassium, ln(K) (c and d); calcium, ln(Ca) (e and f); and magnesium, ln (Mg) (g and h) values using Vis-NIR spectral reflectance data (the subscripts cal and val show the calibration and validation datasets, respectively).

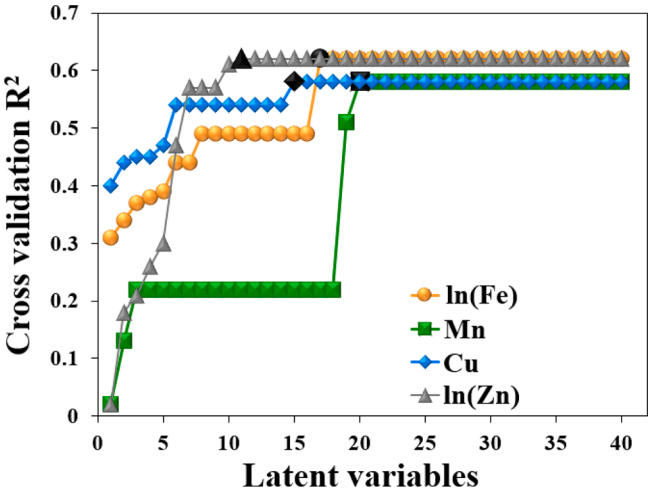

Fig. 9. The cross validation determination coefficient (R^2) values of the partial least square regression (PLSR) analysis for developing models to predict natural logarithm of DTPA-extractable iron, ln(Fe); DTPA-extractable manganese, Mn; and copper, Cu; and natural logarithm of DTPA-extractable zinc, ln(Zn) using Vis-NIR spectral reflectance bands by applying 1–40 latent variables (in each case the bold black spot shows the best number of latent variables).

and Vis-NIR spectra at 1810, 430, and 1840 nm, respectively in soils of southern Italy.

4.1.2. pH, EC, and CCE

The Vis-NIR reflectance spectra-derived PLSR model and STF predicted pH with excellent and acceptable accuracies based on the NRMSE and NS classification guidelines. Overall, according to R^2_{val} and $RPIQ_{val}$ values, the pH was predicted with good accuracy by both developed PLSR model and STF using Vis-NIR reflectance spectra as predictor. The negative values of ME_{val} (−0.02 and −0.01 for developed PLSR model and STF, respectively) showed a very low underestimation of pH by both mentioned spectroscopy-based approaches. According to Viscarra Rossel et al. (2006) and Vohland et al. (2014), intense fundamental molecular frequencies related to soil components occur in the mid-infrared, MIR (2500–25,000 nm) region. They stated that the MIR region is more capable to predict soil pH in comparison with Vis-NIR region. It should be noted that, although the pH was predicted with good accuracy (based on the calculated goodness of fit criteria) by the developed PLSR model and STF, the points of predicted versus observed values were not concentrated on 1:1 line and they were dispersed (Figs. 6a and 12a). This may correspond to low variability of pH in our soil samples. Gomez et al. (2012) and Pinheiro et al. (2017) poorly predicted the pH ($R^2 \leq 0.40$) in French and Brazilian soils, respectively, while Conforti et al. (2018) and Xu et al. (2018a) predicted pH with good ($R^2 = 0.70$) and very good ($R^2 = 0.86$) accuracies using Vis-NIR spectroscopy data and PLSR method. Our results were in accordance with that of Vohland et al. (2014) who reported some key wavelengths to predict pH in range of 1915–2465 nm. Xu et al. (2018a) also reported several effective spectral data around wavelengths of 480, 780, 1120, 1910, 2200, and 2390 nm for predicting pH.

The calculated goodness of fit criteria for validation dataset showed that the EC^{-1} was predicted with poor and unsatisfactory accuracies by both developed PLSR model and STF using Vis-NIR reflectance spectra based on the NRMSE (>30 %) and NS (<0.65) classification guidelines, respectively. Although the NRMSE and NS values were disagreed with good prediction of EC^{-1}, generally the mentioned parameter was predicted with acceptable accuracy by both applied spectroscopy-based methods according to R^2_{val} and $RPIQ_{val}$ values. In addition, the values of ME_{val} (0.03 and 0.1 dS^{-1} m) showed a bit overestimation of EC^{-1} when both PLSR model and STF were used. Electrical conductivity (EC) is a measure of salinity and water-soluble salts values. In calcareous soils of arid and semi-arid regions, alkaline water-soluble salts of NaCl, KCl, $CaCl_2$, and $MgCl_2$ are predominant and result in the lighter color of soil. Therefore, high light color water-soluble salts result in high EC values and consequently soil absorb less and reflect more electromagnetic waves in Vis-NIR range. The moderate (acceptable) accuracy for predicting EC^{-1} by spectroscopy-based methods in the present study can be due to the effects of water-soluble salts on soil color and the significant correlation of EC^{-1} with clay and SOM content (with r values of 0.27 and −0.24, respectively). There are several studies in literature, which predicted EC values moderately with R^2 values between 0.45 and 0.61 (Bilgili et al., 2010; Gholizade et al., 2013; Seifi et al., 2020) and accurately with R^2 values between 0.69 and 0.94 (Cozzolino et al., 2013; Nawar et al., 2014; Nawar et al., 2015; Feyziyev et al., 2016; Srivastava et al., 2017; Vibhute et al., 2018) using Vis-NIR spectra and PLSR method. In literature, several effective bands have been reported for predicting EC in wavelength ranges of 1062–1894 and 1914–2420 nm (Zovko et al., 2018); 400–1891, 2017–2165, and 2280–2359 nm (Seifi et al., 2020); 417–980 nm (Vibhute et al., 2018); and at wavelengths of 440 and 490 nm (Gholizade et al., 2013); and 1400 and 2400 nm (Srivastava et al., 2017), which mostly match with our results.

According to the NRMSE and NS classification guidelines, the CCE was predicted with excellent and good accuracies, respectively using Vis-NIR reflectance spectra as predictor and by the developed PLSR model and STF. Anyway, the CCE was accurately predicted (based on the R^2_{val} and $RPIQ_{val}$ values) by both developed PLSR model and STF. An overestimation was observed for predicting CCE by both mentioned spectroscopy-based approaches with positive values of ME_{val}. The carbonates, which is represented by CCE in the present study, is a well-known soil property that affects spectral reflectance across Vis-NIR region. Therefore, one of the most obvious absorption bands around 2340 nm (Fig. 1) is related to carbonates groups (Viscarra Rossel, 2007; Žížala et al., 2017; Ostovari et al., 2018). As it is demonstrated, the reflectance value at 2335 nm wavelength was appeared in the developed STF (Eq. (17)) for predicting CCE. In addition, there were 14 other single bands across the full range of spectra in Vis-NIR region that were included to develop CCE-STF. It may be due to significant effects of CCE on soil color. Generally, soils with high contents of sand and lime ($CaCO_3$) and low contents of clay and SOM have lighter color and consequently absorb less and reflect more electromagnetic waves in the Vis-NIR range (Mozaffari et al., 2022c). Very good prediction of CCE in a wide range (12.5–70.6 %) by using 15 single bands can be very useful in determining this property in calcareous soils with less usage of chemicals, while saving time and costs. Khayamim et al. (2015) and Bilgili et al. (2010) predicted CCE with acceptable accuracy in Iranian and Turkish soils (R^2 values of 0.54 and 0.65, respectively). While, Ostovari et al. (2018) and Mina et al. (2021) in calcareous soils of Iran, Summers et al. (2011) in Australian and Gomez et al. (2012) in French soils accurately predicted CCE ($0.69 \leq R^2 \leq 0.76$) using Vis-NIR spectroscopy combined with PLSR method which are in line with our result. Regarding the effective single bands to predict CCE, Bilgili et al. (2010) observed relatively high correlations between $CaCO_3$ and spectral reflectance bands in the Vis (350–700 nm) and at the end of NIR (near 2500 nm) ranges. Ostovari et al. (2018) highlighted $CaCO_3$ content has a significant ($p < 0.05$) correlation with some spectral bands at 571, 645, 805, 1418, and 2348 nm wavelengths. Summers et al. (2011) and Khayamim et al. (2015) also reported the high correlation between $CaCO_3$ and spectral data at 2325 and 2338 nm wavelengths. The appeared effective bands in Eq. (17) are close to those reported by the mentioned researchers.

4.1.3. Water-soluble Na, K, Ca, and Mg

The PLSR approach by using Vis-NIR reflectance spectra predicted ln (Na), ln(K), and ln(Mg) with good and acceptable accuracies, while predicted ln(Ca) with the excellent and unsatisfactory accuracies based on NRMSE and NS classification guidelines, respectively. Generally, the PLSR approach showed a good capability for predicting ln(Na), ln(K),

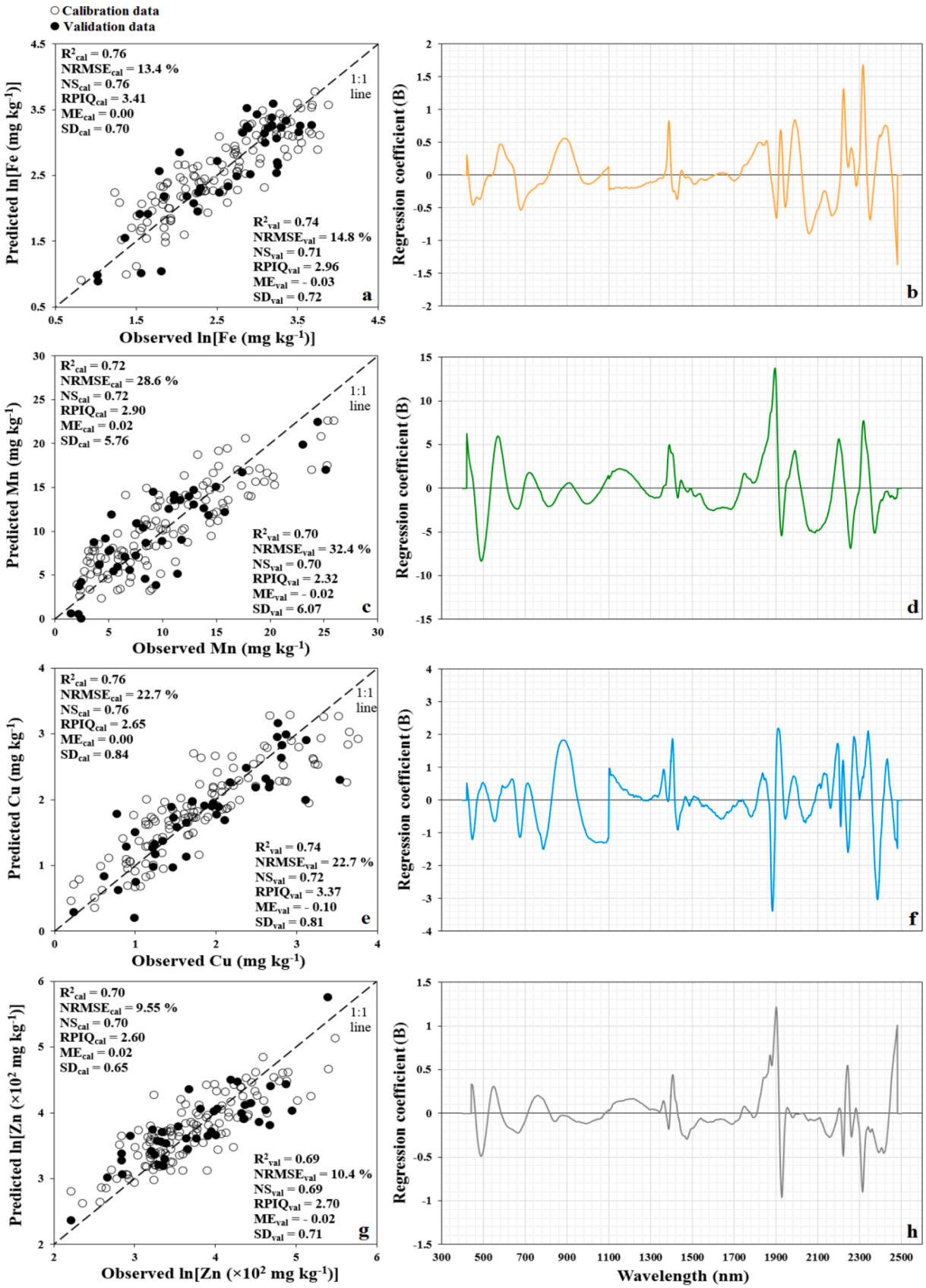

Fig. 10. The partial least square regression (PLSR) analysis results, including scatter plots, goodness of fit criteria, and plots of regression coefficients (B) versus wavelengths, to predict natural logarithm of DTPA-extractable iron, ln(Fe) (a and b); DTPA-extractable manganese, Mn (c and d); and copper, Cu (e and f); and natural logarithm of DTPA-extractable zinc, ln(Zn) (g and h) values using Vis-NIR spectral reflectance data (the subscripts cal and val show the calibration and validation datasets, respectively).

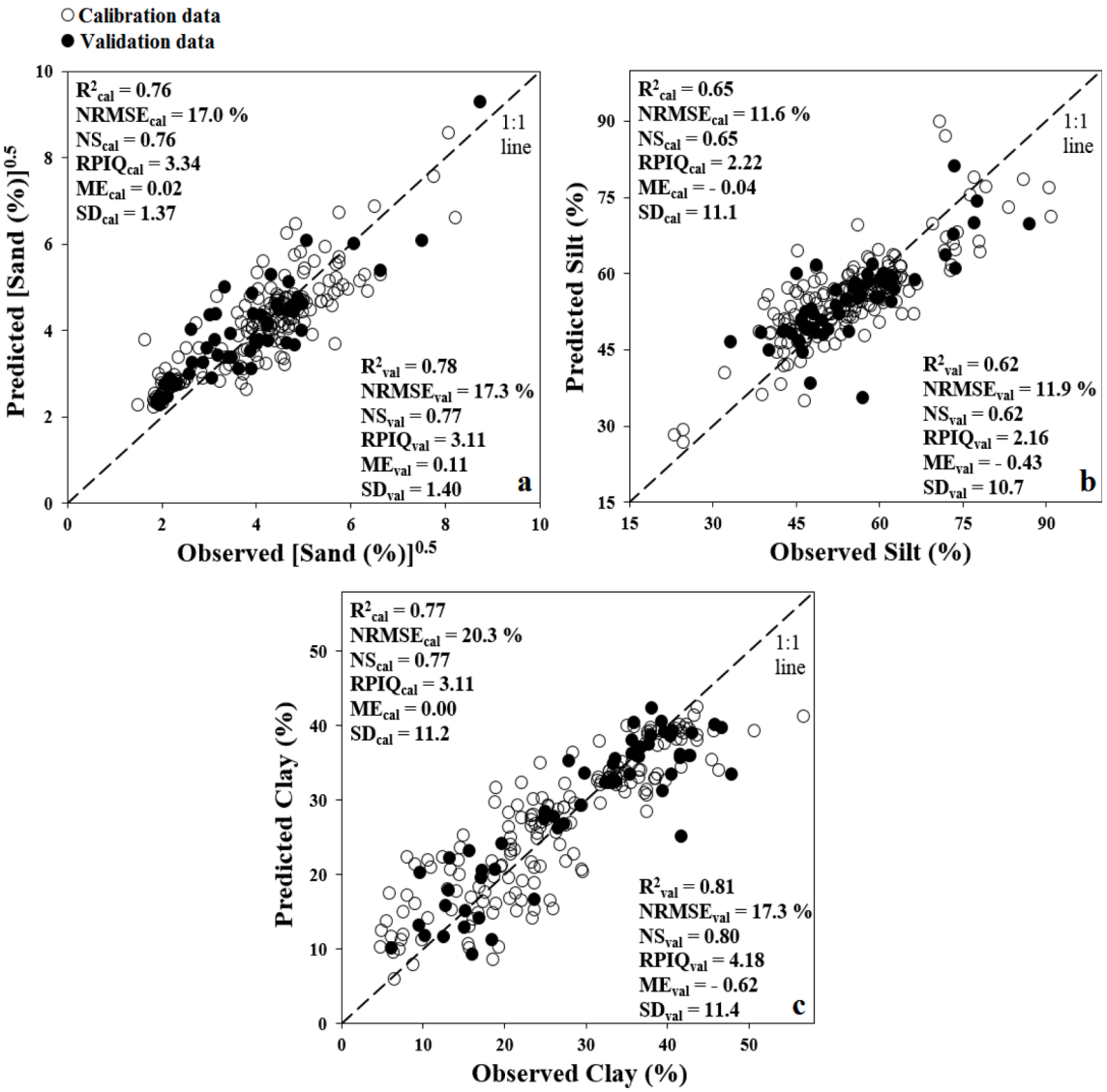

Fig. 11. Scatter plots of the observed versus predicted (a) square root of sand content, sand$^{0.5}$ (Eq. (12)); (b) silt (Eq. (13)); and (c) clay (Eq. (14)) contents using effective Vis-NIR spectral reflectance bands as predictors by applying the stepwise multiple linear regression (SMLR) method along with the goodness of fit criteria (the subscripts cal and val show the calibration and validation datasets, respectively).

and ln(Mg) and acceptable capability for predicting ln(Ca) according to R^2_{val} and $RPIQ_{val}$ values. Prediction of all natural logarithm forms of water-soluble cations by PLSR method was accompanied by underestimation with $-0.2 \leq ME_{val}$ ln(mg L^{-1}) $\leq$ -0.02. On the other hand, the developed STFs predicted ln(Na) and ln(K) with fair accuracy; ln(Ca) with excellent accuracy; and ln(Mg) with good accuracy based on the classification guideline of NRMSE. While, according to NS classification, all natural logarithm forms of water-soluble cations were predicted with unsatisfactory accuracy by the developed STFs. In general, there were an acceptable accuracy for predicting ln(Na) and ln(Ca) and good accuracy for predicting ln(K) and ln(Mg) based on R^2_{val} and $RPIQ_{val}$ criteria by STFs and Vis-NIR reflectance spectra. Similar to PLSR results, underestimations of ln(Na), ln(K), and ln(Mg) were observed with the negative values of ME_{val} when STFs were used, while the ME_{val} value for predicting ln(Ca) was zero.

The extractable Na, K, Ca, and Mg can be assessed with reflectance data based on clay, SOM, and CCE to which they are correlated, but with due consideration of the lack of physical causality (Bilgili et al., 2010). In the present study, the water-soluble forms of these base cations were investigated. Table 3 shows that the ln(Na), ln(K), ln(Ca), and ln(Mg) negatively correlated with clay content (r values of −0.31, −0.28, −0.31, and −0.27, respectively). It can be concluded that, in soils with high content of clay, these bases were predominantly adsorbed to negatively charged clay surfaces and lesser values remain in the soil solution. In addition, the ln(K) positively correlated with SOM content (r = 0.35). Furthermore, the mentioned cations are the predominant water-soluble bases in calcareous soils of the arid and semi-arid regions. These water-soluble salts can result in lighter color of soils that directly affect the absorption/reflection of spectra in Vis-NIR region. The acceptable to good prediction accuracies for these cations ($0.60 \leq R^2_{val} \leq 0.74$) using Vis-NIR reflectance spectra by applying PLSR models and STFs may be due to the mentioned reasons. It is expected that predicting exchangeable forms of the mentioned bases using spectroscopy-based methods be more accurate than that of their water-soluble forms due to more correlation between their exchangeable forms and clay and SOM contents which can be accurately assessed by Vis-NIR spectra. Regarding capability of PLSR using Vis-NIR spectra as predictor to predict alkaline cations, similar to our results, Mouazen et al. (2010), in soils of Belgium and Northern France, predicted exchangeable Na, K, and Mg with acceptable, good, and very good accuracies ($0.60 < R^2 < 0.80$) and reported some effective spectral bands in wavelength ranges of 453–900 and 1400–2448 nm for their prediction. While, Pinheiro

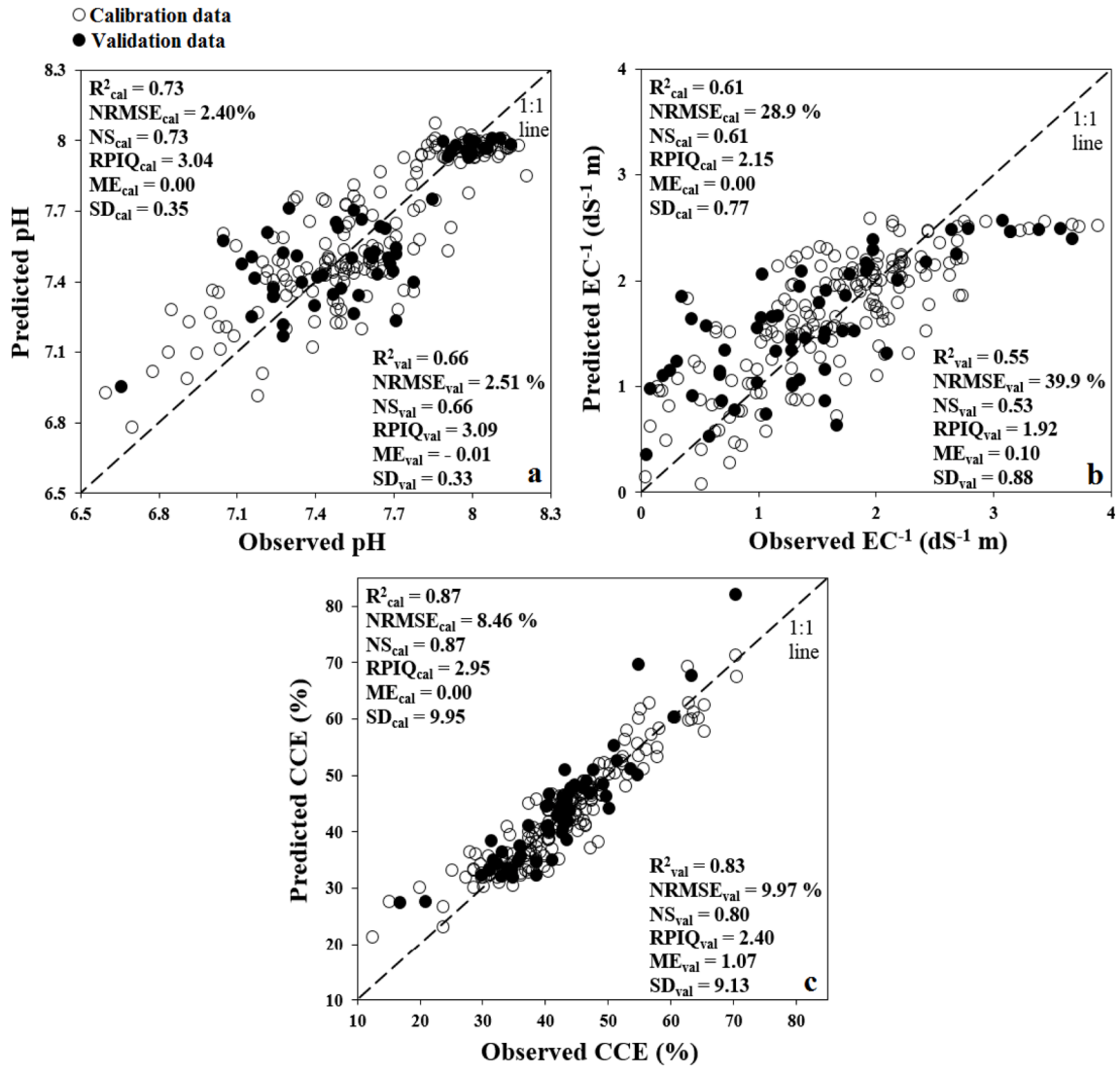

Fig. 12. Scatter plots of the observed versus predicted (a) pH of saturated paste, pH (Eq. (15)); (b) inverse of electrical conductivity of saturated extract, EC^{-1} (Eq. (16)); and (c) calcium carbonate equivalent, CCE (Eq. (17)) values using effective Vis-NIR spectral reflectance bands as predictors by applying the stepwise multiple linear regression (SMLR) method along with the goodness of fit criteria (the subscripts cal and val show the calibration and validation datasets, respectively).

et al. (2017) poorly predicted exchangeable Ca, by some effective spectral reflectance bands in wavelength ranges of 400–1000 and 2100–2500 nm, and Mg, by these in full range of spectra, (R^2 values of 0.40 and 0.33, respectively). Srivastava et al. (2017) in Indian soils accurately predicted water-soluble Na and Ca + Mg by R^2_{val} values of 0.86 and 0.83, respectively. Bilgili et al. (2010) poorly predicted exchangeable Na and K (R^2 values of 0.19 and 0.38, respectively), while predicted exchangeable Ca and Mg with good and acceptable accuracies (R^2 values of 0.68 and 0.55, respectively). They reported significant correlation between spectral reflectance data with exchangeable Na near 350 nm and in 1500–2000 nm range; with exchangeable K and Mg in 350–1500 nm range; and with exchangeable Ca in 350–500 and 1800–2500 nm ranges.

4.1.4. *DTPA-extractable Fe, Mn, Cu, and Zn*

The developed PLSR models and STFs using Vis-NIR reflectance spectra predicted ln(Fe) and ln(Zn) with good; Mn with poor; and Cu with fair accuracies based on the NRMSE. Furthermore, based on the NS classification guidelines the mentioned approaches predicted all of the mentioned soil properties, acceptably. Overall, according to the values of R^2_{val} (0.68–0.74) and $RPIQ_{val}$ (2.32–3.37), the ln(Fe), Mn, Cu, and ln (Zn) were predicted by good accuracy using the applied spectroscopy-based approaches and the spectral reflectance data in Vis-NIR region. In addition, the negative values of ME_{val} indicated that there was a bit underestimation for predicting of the studied soil properties using the mentioned approaches. The DTPA-extractable micronutrients can adsorb onto clay and SOM surfaces due to their positive charges. As can be seen in Table 3, the ln(Fe), Mn, Cu, and ln(Zn) are highly correlated with SOM with r values of 0.44, 0.74, 0.47, and 0.54, respectively. Single-site studies have shown that SOM enhances the metal concentration and bioavailability in soils (Marschner and Kalbitz, 2003; Loveland and Webb, 2003; Luo et al., 2016). The direct and positive effects of SOM on ln(Fe), Mn, Cu, and ln(Zn) availability may be attributed to the capacity of SOM to form soluble complexes with metals, thus decreasing their sorption and increasing their mobility (Habiby et al., 2014; Moreno-Jiménez et al., 2019; Tayebi et al. 2021). Furthermore, the positive correlation between clay content and ln(Fe) and Cu (with r values of 0.45 and 0.27, respectively) and negative correlation between CCE and ln(Fe), Mn, Cu, and ln(Zn) (with r values of −0.26, −0.33, −0.32, and −0.19, respectively) are other reasons for good predictions of DTPA-extractable micronutrients by spectral reflectance data in Vis-NIR range. In addition to these reasons, these micronutrients can directly affect soil color, i.e., Fe and Cu make soils to have stronger color (red to brown); Mn makes soils darker; and Zn makes

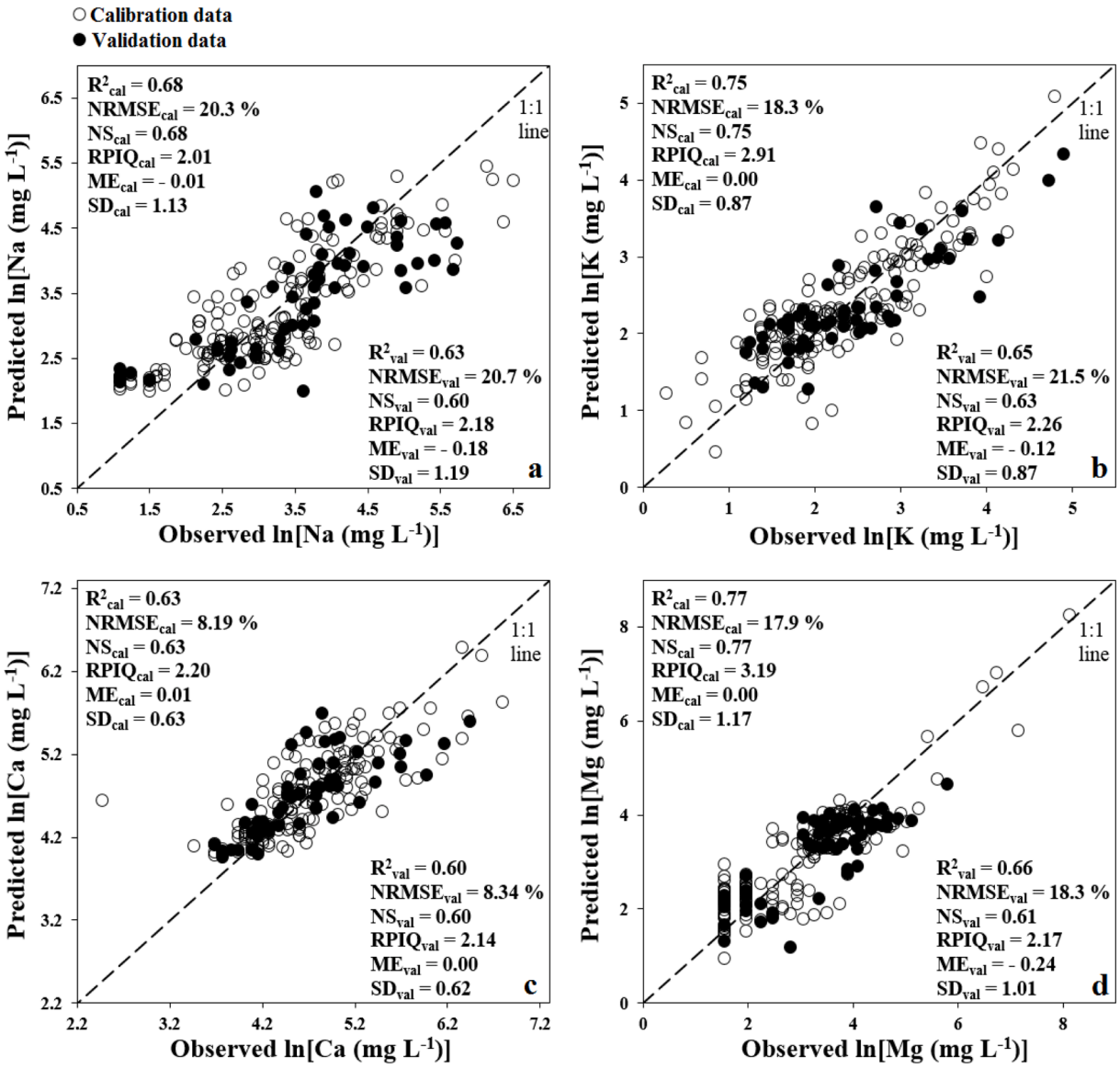

Fig. 13. Scatter plots of the observed versus predicted (a) natural logarithm forms of water-soluble sodium, ln(Na) (Eq. (18)); (b) potassium, ln(K) (Eq. (19)); (c) calcium, ln(Ca) (Eq. (20)); and (d) magnesium, ln(Mg) (Eq. (21)) values using effective Vis-NIR spectral reflectance bands as predictors by applying the stepwise multiple linear regression (SMLR) method along with the goodness of fit criteria (the subscripts cal and val show the calibration and validation datasets, respectively).

soils lighter. According to Table 3, the correlations of DTPA-extractable micronutrients with SOM content were ranked as: Mn > ln(Zn) > Cu ≈ ln(Fe); and with clay content were ranked as: ln(Fe) > Cu > ln(Zn) > Mn. In addition, the mentioned micronutrients were correlated with CCE as: Mn ≈ Cu > ln(Fe) > ln(Zn). Therefore, the combination of correlations between micronutrients and basic soil properties (i.e., clay, SOM, and CCE) and their direct effects on soil color has resulted in relatively same levels of accuracy for predicting studied micronutrients using spectroscopy-based methods and Vis-NIR spectral reflectance data. The developed STFs (Eqs. (22) to (25)) are simple and easy to use with good capability to predict micronutrients, which are highly recommended to apply in calcareous soils. Therefore, our expectations regarding prediction of the mentioned micronutrients were met. Regarding PLSR capability to predict micronutrients, Johnson et al. (2019) reported poorly prediction of Mehlich-3 extractable Fe, Mn, Cu, and Zn with R^2_{cv} values of 0.38, 0.32, 0.43, and 0.44 using NIR spectral reflectance bands for soils of 20 sub-Saharan African countries. Terra et al (2015) in Brazilian soils predicted DTPA-Triethanol Amine (TEA) extractable Fe and Zn with poor accuracy (R^2_{val} values of 0.39 and 0.26), Mn with moderate accuracy (R^2_{val} = 0.54), and Cu with good accuracy (R^2_{val} = 0.69) using spectral reflectance bands in Vis-NIR range. Singh et al. (2019) in soils of Papua New Guinea predicted DTPA-extractable Fe, Mn, and Cu with substantial accuracy and predicted Zn with poor accuracy using NIR diffuse reflectance spectroscopy. In addition, Moron and Cozzolino (2003) in soils of Uruguay reported some important and effective bands in range of 400–800 nm, around 1100 and 1400 nm, and in range of 1900–2500 nm to estimate Fe, Cu, and Zn. They stated that strong correlation between micronutrients and spectral reflectance data at 1400 and 1900 nm and between 2200 and 2400 nm is attributed to the adsorption features of micronutrients oxides and with secondary clay mineral adsorption. Singh et al. (2019) stated that the characteristic absorption features around 550 and 850 nm may be linked with electronic transitions of Fe^{2+} and Fe^{3+}. Wu et al. (2007) in eastern soils of China reported high correlation (r > 0.40) with spectral reflectance bands and values of Fe, Cu, and Zn (measured by inductively coupled plasma-atomic emission spectroscopy, ICP-AES method) in ranges of 400–600 and 1000–2500 nm.

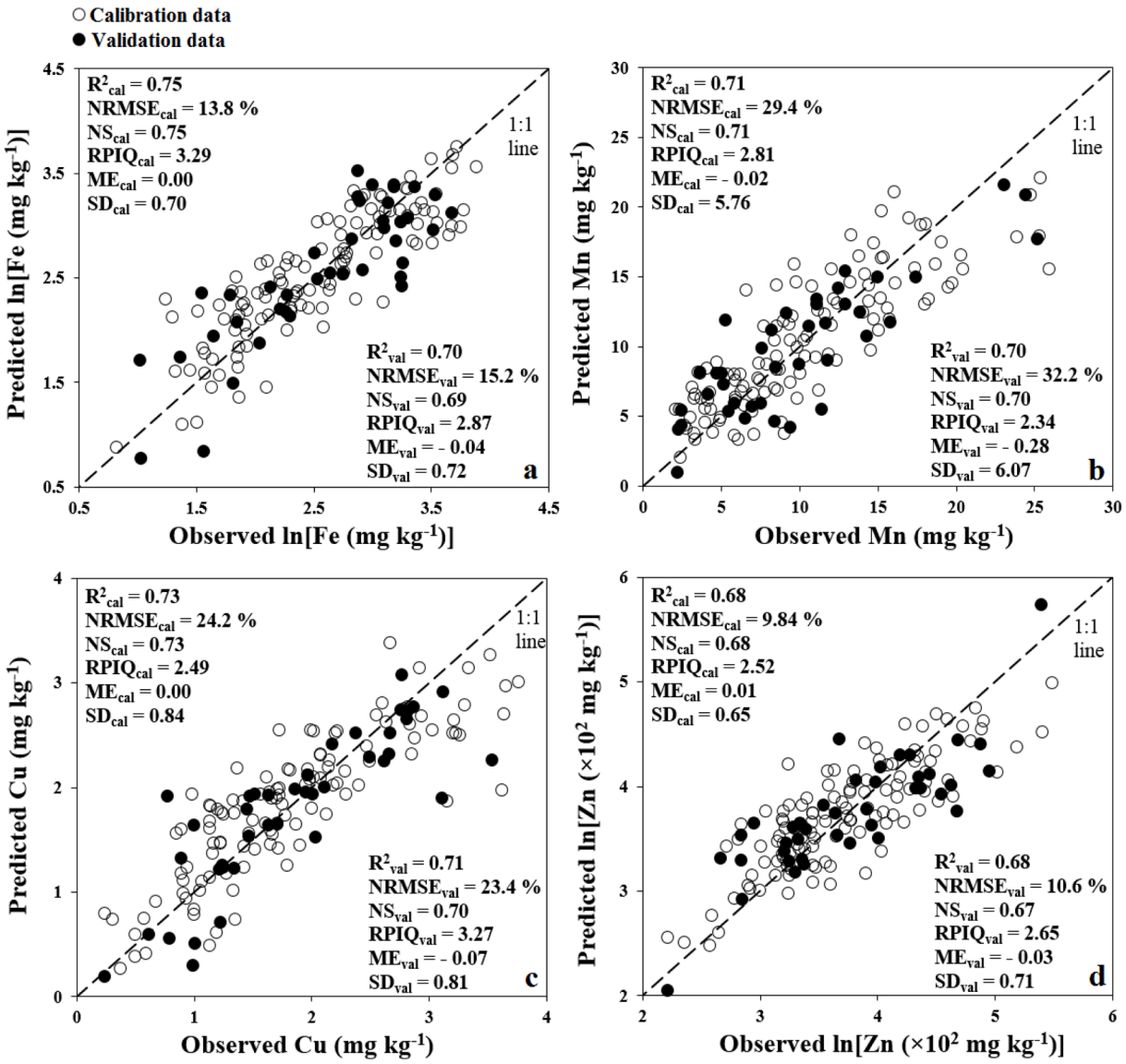

Fig. 14. Scatter plots of the observed versus predicted (a) natural logarithm of DTPA-extractable iron, ln(Fe) (Eq. (22)); (b) DTPA-extractable manganese, Mn (Eq. (23)); and (c) copper, Cu (Eq. (24)); and (d) natural logarithm of DTPA-extractable zinc, ln(Zn) (Eq. (25)) values using effective Vis-NIR spectral reflectance bands as predictors by applying the stepwise multiple linear regression (SMLR) method along with the goodness of fit criteria (the subscripts cal and val show the calibration and validation datasets, respectively).

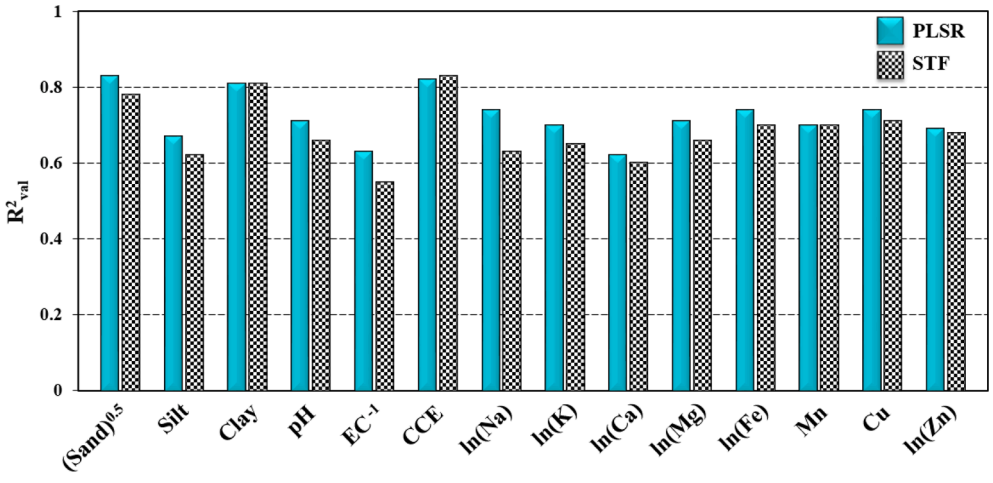

Fig. 15. Comparing the determination coefficient of validation dataset (R^2_{val}) values of the developed partial least square regression (PLSR) models and spectrotransfer functions (STFs) (Eqs. (12) to (25)) using Vis-NIR reflectance spectra to predict the studied soil properties ($Sand^{0.5}$: square root of sand content; Silt: silt content; Clay: clay content; pH: pH of saturated paste; EC^{-1}: inverse of electrical conductivity of saturated extract; CCE: calcium carbonate equivalent (lime); ln(Na), ln(K), ln(Ca), and ln(Mg): natural logarithms (ln) of water-soluble sodium, potassium, calcium, and magnesium, respectively; ln(Fe) and ln(Zn): natural logarithms of DTPA-extractable iron and zinc, respectively; Mn and Cu: DTPA-extractable manganese and copper, respectively).

4.2. Comparing capability of the developed PLSR models and STFs to predict soil properties

The PLSR models showed a better performance compared to STFs to predict sand$^{0.5}$, silt, pH, EC^{-1}, ln(Na), ln(K), ln(Ca), ln(Mg), ln(Fe), Cu, and ln(Zn) using Vis-NIR reflectance spectra with R^2_{val} values increases of 6.4, 8.1, 7.6, 14.5, 17.5, 7.7, 3.3, 7.6, 5.7, 4.2, and 1.5 %, respectively. In contrast, an increase of 1.2 % was observed in R^2_{val} of STF compared to PLSR model developed for predicting CCE. In addition, there were no differences in R^2_{val} values between the developed PLSR models and STFs to predict clay and Mn. Although the PLSR models resulted in better capabilities compared to STFs for predicting most of the studied soil properties, their capabilities differences were not significant. The maximum difference between capability of the developed PLSR model and STF was observed in predicting ln(Na) with a R^2_{val} difference of 0.11 (17.5 %). The better performance of PLSR models compared to STFs may be related to adding $\geq$ 10 LVs and including more bands in modeling procedure that have multi-collinearity. But the STFs are so easy to use and applicable with acceptable, good, and very good accuracies for predicting studied basic soil properties which shows our procedure to select the important and effective Vis-NIR reflectance single bands for developing STFs was sound and appropriate. All studied soil properties were predicted with R^2_{val} values $\geq$ 0.55 using Vis-NIR spectra by both of the applied spectroscopy-based methods. Generally, the developed STFs (Eqs. (12) to (25)) predicted sand$^{0.5}$, clay, and CCE with very good accuracy; pH, ln(K), ln(Mg), ln(Fe), Mn, Cu, and ln(Zn) with good accuracy; and silt, EC^{-1}, ln(Na), and ln(Ca) with acceptable accuracy using Vis-NIR reflectance spectra as predictor.

Overall, the following points are recommended for future studies: 1) the new STFs be developed for predicting other soil physical and chemical properties, which have not been investigated in the present study; 2) the new STFs with various pre-processing techniques (e.g., the first-order derivative) be developed for predicting studied soil properties to test if it yields better results; 3) our developed STFs be tested for predicting different soil physical and chemical properties in the other regions including non-calcareous soils; 4) our procedure for selecting effective spectral single bands be used for developing STFs and predicting soil properties; and more importantly 5) for predicting different soil properties using spectroscopy-based methods, researchers pay special attention to develop simple, easy to use, and applicable STFs.

5. Conclusion

Accurate prediction of basic soil physical and chemical properties using spectroscopy-based methods (as a rapid, timely, non-destructive, low-preparation, and less expensive analysis of soil samples) can be very valuable. The current study was carried out in order to predict basic soil properties of calcareous soils in Fars Province, Iran using Vis-NIR reflectance spectra. Results indicated that accuracy of the developed PLSR models were better than STFs for predicting most of the studied basic soil physical and chemical properties, however their performance differences were not significant. The STFs were developed using the minimum 11, for predicting pH and ln(Zn), and maximum 17, for predicting silt and Mn, effective spectral reflectance single bands in Vis-NIR region. The sand$^{0.5}$, clay, and CCE were predicted with very good accuracy ($0.78 \leq R^2_{val} \leq 0.83$); pH, ln(K), ln(Mg), ln(Fe), Mn, Cu, and ln (Zn) were predicted with good accuracy ($0.65 \leq R^2_{val} \leq 0.71$); and silt, EC^{-1}, ln(Na), and ln(Ca) were predicted with acceptable (fair or moderate) accuracy ($0.55 \leq R^2_{val} \leq 0.63$) by STFs. The developed STFs are highly recommended for predicting sand$^{0.5}$, clay, CCE, pH, ln(K), ln (Mg), ln(Fe), Mn, Cu, and ln(Zn) in calcareous soils in order to map soils in large scales as an easily approach without using chemicals for soil analysis, which are harmful for human health and environment, as well as saving time and money. The developed STFs may also work well in non-calcareous soils; however, their applicability should be tested before using.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgment

Authors appreciate financial supports from Shiraz University for carrying out the research. Authors also would like to thank dear Editor in Chief Prof. Cristine Morgan and three anonymous Reviewers for the very constructive and valuable comments that really improved quality of the manuscript.

References

Adeline, K.R.M., Gomez, C., Gorretta, N., Roger, J.M., 2017. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. Geoderma 288, 143–153.
Babaeian, E., Homaee, M., Montzka, C., Vereecken, H., Norouzi, A.A., 2015. Towards retrieving soil hydraulic properties by hyperspectral remote sensing. Vadose Zone J. 14, 1–17.
Banaei, M.H., 1998. Soil Moisture and Temperature Regime Map of Iran. Soil and Water Research Institute, Ministry of Agriculture, Tehran, Iran.
Bannayan, M., Hoogenboom, G., 2009. Using pattern recognition for estimating cultivar coefficients of a crop simulation model. Field Crops Res. 111 (3), 290–302.
Bilgili, V., Akbas, F., van ES, H., 2011. Combined use of hyperspectral VNIR reflectance spectroscopy and kriging to predict soil variables spatially. Precis. Agric. 12, 395–420.
Bilgili, A.V., van Es, H.M., Akbas, F., Durak, A., Hively, W.D., 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. J. Arid Environ. 74 (2), 229–238.
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.
CAMO, Technologies Inc., 2013. The Unscrambler appendices: method references. http://www.camo.com/TheUnscrambler/Appendices/The20Unscrambler%20Method%20References/ (Last verified: May 2, 2013).
Conforti, M., Matteucci, G., Buttafuoco, G., 2018. Using laboratory Vis-NIR spectroscopy for monitoring some forest soil properties. J. Soils Sediments. 18 (3), 1009–1019.
Cozzolino, D., Cynkar, W.U., Dambergs, R.G., Shah, N., Smith, P., 2013. In situ measurement of soil chemical composition by near-infrared spectroscopy: A tool toward sustainable vineyard management. Commun. Soil Sci. Plant Anal. 44 (10), 1610–1619.
de Jong, S., 1993. SIMPLS: An alternative approach to partial least squares regression. Chemom. Intell. Lab. Syst. 18 (3), 251–263.
Feng, Y., Cui, N., Gong, D., Zhang, Q., Zhao, L., 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. Agric. Water Manag. 193, 163–173.
Feyziyev, F., Babayev, M., Priori, S., L'Abate, G., 2016. Using visible-near infrared spectroscopy to predict soil properties of mugan plain, Azerbaijan. Open J. Soil Sci. 06 (03), 52–58.
Gandomkar, A., Dehghani, R., 2012. 'Study of Temperature Changes in Fars Province'. World Acad. Sci. Eng. Technol., Open Science Index 63, Int. J. Environ. Eng. 6, 127-129.
Gee, G.W., Bauder, J.W., 1986. Particle size analysis, hydrometer methods. In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods. American Society of Agronomy and Soil Science Society of America, Madison, Wisconsin, USA, pp. 383–411.
Gholizade, A., MohdSoom, M.A., Saberioon, M.M., Borůvka, L., 2013. Visible and near infrared reflectance spectroscopy to determine chemical properties of paddy soils. J. Food Agric. Environ. 11, 859–866.
Gomez, C., Viscarra Rossel, R.A., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field Vis-NIR spectroscopy: an Australian case study. Geoderma 146 (3-4), 403–411.
Gomez, C., Lagacherie, P., Coulouma, G., 2012. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data. Geoderma 189–190, 176–185.
Guo, L., Zhang, H., Shi, T., Chen, Y., Jiang, Q., Linderman, M., 2019. Prediction of soil organic carbon stock by laboratory spectral data and airborne hyperspectral images. Geoderma 337, 32–41.
Habiby, H., Afyuni, M., Khoshgoftarmanesh, A.H., Schulin, R., 2014. Effect of preceding crops and their residues on availability of zinc in a calcareous Zn-deficient soil. Biol. Fertil. Soils 50 (7), 1061–1067.

Han, I.-S., Han, C., Chung, C.-B., 2005. Melt index modeling with support vector machines, partial least squares, and artificial neural networks. J. Appl. Polym. Sci. 95 (4), 967–974.
Heil, J., Häring, V., Marschner, B., Stumpe, B., 2019. Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils. Geoderma 337, 11–21.
Helmke, P., Sparks, D.L., 1996. Lithium, sodium, potassium, rubidium, and cesium. In: Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, T., Sumner, M.E. (Eds.), Methods of Soil Analysis, Part 3, Chemical Methods. American Society of Agronomy and Soil Science Society of America, Madison, Wisconsin, USA, pp. 551–574.
Hobley, E.U., Prater, I., 2019. Estimating soil texture from vis-NIR spectra. Eur. J. Soil Sci. 70 (1), 83–95.
Jaconi, A., Vos, C., Don, A., 2019. Near infrared spectroscopy as an easy and precise method to estimate soil texture. Geoderma 337, 906–913.
Jia, X., Chen, S., Yang, Y., Zhou, L., Yu, W., Shi, Z., 2017. Organic carbon prediction in soil cores using VNIR and MIR techniques in an alpine landscape. Sci. Rep. 7, 2144.
Johnson, J.-M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K.D., Saito, K., 2019. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. Geoderma 354, 113840.
Khayamim, F., Wetterlind, J., Khademi, H., Robertson, A.H.J., Cano, A.F., Stenberg, B.o., 2015. Using visible and near infrared spectroscopy to estimate carbonates and gypsum in soils in arid and subhumid regions of Isfahan, Iran. J. Near Infrared Spectrosc. 23 (3), 155–165.
Khormali, F., Abtahi, A., 2003. Origin and distribution of clay minerals in calcareous arid and semi-arid soils of Fars Province, southern Iran. Clay Miner. 38 (4), 511–527.
Li, B., Martin, E., Morris, J., 2001. Latent variable selection in partial least squares modelling. IFAC Proc. 34 (25), 463–468.
Lindsay, W.L., Norvell, W.A., 1978. Development of a DTPA soil test for zinc, iron, manganese, and copper. Soil Sci. Soc. Amer. J. 42 (3), 421–428.
Loeppert, R.H., Suarez, D.L., 1996. Carbonate and gypsum. In: Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. (Eds.), Methods of Soil Analysis, Part 3, Chemical and Microbiological Properties. American Society of Agronomy and Soil Science Society of America, Madison, Wisconsin, USA, pp. 437–474.
Loveland, P., Webb, J., 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. Soil Tillage Res. 70, 1–18.
Luo, W., Sardans, J., Dijkstra, F.A., Peñuelas, J., Lü, X.-T., Wu, H., Li, M.-H., Bai, E., Wang, Z., Han, X., Jiang, Y., 2016. Thresholds in decoupled soil-plant elements under changing climatic conditions. Plant Soil. 409 (1-2), 159–173.
Marschner, B., Kalbitz, K., 2003. Controls of bioavailability and biodegradability of dissolved organic matter in soils. Geoderma 113, 211–235.
Mina, M., Rezaei, M., Sameni, A., Moosavi, A.A., Ritsema, C., 2021. Vis-NIR spectroscopy predicts threshold velocity of wind erosion in calcareous soils. Geoderma 401, 115163.
Moosavi, A.A., Sepaskhah, A.R., 2012. Determination of unsaturated soil hydraulic properties at different applied tensions and water qualities. Arch. Agron. Soil Sci. 58 (1), 11–38.
Moreno-Jiménez, E., Plaza, C., Saiz, H., Manzano, R., Flagmeier, M., Maestre, F.T., 2019. Aridity and reduced soil micronutrient availability in global drylands. Nat. Sustain. 2 (5), 371–377.
Moron, A., Cozzolino, D., 2003. Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils. J. Near Infrared Spectrosc. 11 (2), 145–154.
Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma 158 (1-2), 23–31.
Mozaffari, H., Moosavi, A.A., Sepaskhah, A.R., 2021a. Land use-dependent variation of near-saturated and saturated hydraulic properties in calcareous soils. Environ. Earth. Sci. 80, 769.
Mozaffari, H., Rezaei, M., Ostovari, Y., 2021b. Soil sensitivity to wind and water erosion as affected by land use in southern Iran. Earth 2, 287–302.
Mozaffari, H., Moosavi, A.A., Cornelis, W., 2022a. Vis-NIR-spectroscopy-and loss-on-ignition-based functions to estimate organic matter content of calcareous soils. Arch. Agron. Soil Sci. https://doi.org/10.1080/03650340.2022.2047944. In press.
Mozaffari, H., Moosavi, A.A., Dematte, J.A.M., 2022b. Estimating particle-size distribution from limited soil texture data: Introducing two new methods. Biosyst. Eng. 216, 198–217.
Mozaffari, H., Akbar Moosavi, A., Ostovari, Y., Cornelis, W., 2022c. Comparing visible-near-infrared spectroscopy with classical regression pedotransfer functions for predicting near-saturated and saturated hydraulic conductivity of calcareous soils. J. Hydrol. 613, 128412.
Mozaffari, H., Moosavi, A.A., Sepaskhah, A., Cornelis, W., 2022d. Long-term effects of land use type and management on sorptivity, macroscopic capillary length and water-conducting porosity of calcareous soils. Arid. Land Res. Manag. https://doi.org/10.1080/15324982.2022.2066582. In press.
MPB (Ministry of Programming and Budgeting), 1994. Economic and Social Status of Fars Province. Publication Centre for Informatics and Development Studies (In Persian).
Munawar, A.A., Yunus, Y., Devianti, S.P., 2020. Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties. Data Brief. 30, 105469.
Natural Resources and Forestry Organization of Fars Province, 2010. Detailed Report of the Dorudzan Watershed. Fars Regional Water Authority, Shiraz, Iran, pp. 15-125 (In Persian).
Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., 2014. Modeling and mapping of soil salinity with reflectance spectroscopy and Landsat data using two quantitative methods (PLSR and MARS). Remote Sens. 6, 10813–10834.
Nawar, S., Buddenbaum, H., Hill, J., 2015. Estimation of soil salinity using three quantitative methods based on visible and near-infrared reflectance spectroscopy: a case study from Egypt. Arab. J. Geosci. 8, 5127–5140.
Nawar, S., Mouazen, A.M., 2017. Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line Vis-NIR spectroscopy measurements of soil total nitrogen and total carbon. Sensors 17, 2428.
Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter, in: Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. (Eds.), Methods of Soil Analysis, Part 3, Chemical and Microbiological Properties. American Society of Agronomy and Soil Science Society of America, Madison, Wisconsin, USA, pp. 961-1010.
Ostovari, Y., Moosavi, A.A., Mozaffari, H., Poppiel, R.R., Tayebi, M., Demattê, J.A.M., 2022. Chapter 32-Soil erodibility and its influential factors in the Middle East, in: Pourghasemi, H.R. (Ed.). Computers in Earth and Environmental Sciences. Elsevier, pp. 441-454.
Ostovari, Y., Ghorbani-Dashtaki, S., Bahrami, H.-A., Abbasi, M., Dematte, J.A.M., Arthur, E., Panagos, P., 2018. Towards prediction of soil erodibility, SOM and $CaCO_3$ using laboratory Vis-NIR spectra: a case study in a semi-arid region of Iran. Geoderma 314, 102–112.
Ostovari, Y., Moosavi, A.A., Mozaffari, H., Pourghasemi, H.R., 2021. RUSLE model coupled with RS-GIS for soil erosion evaluation compared with T value in Southwest Iran. Arab. J. Geosci. 14, 110.
Park, T.C., Kim, U.S., Kim, L.H., Jo, B.W., Yeo, Y.K., 2010. Heat consumption forecasting using partial least squares, artificial neural network and support vector regression techniques in district heating systems. Korean J. Chem. Eng. 27, 1063–1071.
Pinheiro, É.F.M., Ceddia, M.B., Clingensmith, C.M., Grunwald, S., Vasques, G.M., 2017. Prediction of soil physical and chemical properties by visible and near-infrared diffuse reflectance spectroscopy in the central amazon. Remote Sens. 9, 293.
Pyo, J.C., Hong, S.M., Kwon, Y.S., Kim, M.S., Cho, K.H., 2020. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. Sci. Total Environ. 741, 140162.
Rannar, S., Lindgren, F., Geladi, P., Wold, S., 1994. A PLS kernel algorithm for datasets with many variables and fewer objects. Part 1: theory and algorithm. J. Chemom. 8, 111–125.
Rezaee, L., Moosavi, A.A., Davatgar, N., Sepaskhah, A.R., 2020a. Soil quality indices of paddy soils in Guilan province of northern Iran: Spatial variability and their influential parameters. Ecol. Indic. 117, 106566.
Rezaee, L., Moosavi, A.A., Davatgar, N., Sepaskhah, A.R., 2020b. Shrinkage-swelling characteristics and plasticity indices of paddy soils: spatial variability and their influential parameters. Arch. Agron. Soil Sci. 66, 2005–2025.
Rhoades, J.D., 1996. Salinity: Electrical conductivity and total dissolved salts, in: Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. (Eds.), Methods of Soil Analysis, Part 3, Chemical and Microbiological Properties. Madison, Wisconsin, USA: American Society of Agronomy and Soil Science Society of America, pp. 417-436.
Richards, L.A., 1954. Diagnosis and improvement of saline and alkali soils, in: Richards, L.A. (Ed.). U. S. Salinity Laboratory Staff, USDA Hand Book NO. 60. Washington, DC, USA, 160 P.
Ritter, A., Muñoz-Carpena, R., 2013. Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. J. Hydrol. 480, 33–45.
Salehi-Varnousfaderani, B., Honarbakhsh, A., Tahmoures, M., Akbari, M., 2022. Soil erodibility prediction by Vis-NIR spectra and environmental covariates coupled with GIS, regression and PLSR in a watershed scale, Iran. Geoderma Reg. 28, e00470.
Savitzky, A., Golay, J.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.
Seifi, M., Ahmadi, A., Neyshabouri, M.R., Taghizadeh-Mehrjardi, R., Bahrami, H.A., 2020. Remote and Vis-NIR spectra sensing potential for soil salinization estimation in the eastern coast of Urmia hyper saline lake, Iran. Remote Sens. Appl.: Soc. Environ. 20, 100398.
Singh, K., Majeed, I., Panigrahi, N., Vasava, H.B., Fidelis, C., Karunaratne, S., Bapiwai, P., Yinil, D., Sanderson, T., Snoeck, D., Das, B.S., Minasny, B., Field, D.J., 2019. Near infrared diffuse reflectance spectroscopy for rapid and comprehensive soil condition assessment in smallholder cacao farming systems of Papua New Guinea. Catena 183, 104185.
Soil Survey Staff, 2014. Keys to Soil Taxonomy, twelfth ed. USDA-Natural Resource Conservation Service, Washington, DC.
Srivastava, R., Sethi, M., Yadav, R.K., Bundela, D.S., Singh, M., Chattaraj, S., Singh, S.K., Nasre, R.A., Bishnoi, S.R., Dhale, S., Mohekar, D.S., Barthwal, A.K., 2017. Visible-near infrared reflectance spectroscopy for rapid characterization of salt-affected soil in the Indo-Gangetic plains of Haryana, India. J. Indian Soc. Remote Sens. 45, 307–315.
Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. Geoderma 158, 15–22.
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. Adv. Agron. 107, 163–215.
Summers, D., Lewis, M., Ostendorf, B., Chittleborough, D., 2011. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. Ecol. Indic. 11, 123–131.
Tayebi, M., Fim Rosas, J.T., Mendes, W.d.S., Poppiel, R.R., Ostovari, Y., Ruiz, L.F.C., dos Santos, N.V., Cerri, C.E.P., Silva, S.H.G., Curi, N., Quiñonez Silvero, N.E., Demattè, J.

A.M., 2021. Drivers of organic carbon stocks in different LULC history and along soil depth for a 30 years image time series. Remote Sens. 13, 2223.
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.
Thomas, G. W. (1996). Soil pH and soil asidity, in: Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. (Eds.), Methods of Soil Analysis, Part 3, Chemical and Microbiological Properties. American Society of Agronomy and Soil Science Society of America, Madison, Wisconsin, USA, pp. 475-490.
Tsakiridis, N.L., Theocharis, J.B., Panagos, P., Zalidis, G.C., 2019. An evolutionary fuzzy rule-based system applied to the prediction of soil organic carbon from soil spectral libraries. Appl. Soft Comput. 81, 105504.
Vasava, H.B., Gupta, A., Arora, R., Das, B.S., 2019. Assessment of soil texture from spectral reflectance data of bulk soil samples and their dry-sieved aggregate size fractions. Geoderma 337, 914–926.
Vibhute, A.D., Kale, K.V., Mehrotra, S.C., Dhumal, R.K., Nagne, A.D., 2018. Determination of soil physicochemical attributes in farming sites through visible, near-infrared diffuse reflectance spectroscopy and PLSR modeling. Ecol. Process. 7, 26.
Viscarra Rossel, R.A., 2007. Robust modelling of soil diffuse reflectance spectra by "bagging-partial least squares regression". J. Near Infrared Spectrosc. 15, 39–47.
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, 59–75.
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.
Wilding, L.G., 1985. Soil spatial variability: Its documentation, accommodation and implication to soil surveys, in: Nielsen, D.R., Bouma, J. (Eds.), Soil Spatial Variability Proceedings of a Workshop of the ISSS and the SSA. Las Vegas PUDOC, Wageningen, USA, pp. 166-187.
Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method, in: Kågström, B., Ruhe, A. (Eds.), Matrix Pencils, Lecture Notes in Mathematics. Springer, Heidelberg, Berlin, 973, pp. 286-293.
Wu, Y., Chen, J., Ji, J., Gong, P., Liao, Q., Tian, Q., Ma, H., 2007. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. Soil Sci. Soc. Am. J. 71, 918–926.
Xu, D., Ma, W., Chen, S., Jiang, Q., He, K., Shi, Z., 2018a. Assessment of important soil properties related to Chinese Soil Taxonomy based on vis-NIR reflectance spectroscopy. Comput. Electron. Agric. 144, 1–8.
Xu, C., Xu, X., Liu, M., Liu, W., Yang, J., Luo, W., Zhang, R., Kiely, G., 2017. Enhancing pedotransfer functions (PTFs) using soil spectral reflectance data for estimating saturated hydraulic conductivity in southwestern China. Catena 158, 350–356.
Xu, S., Zhao, Y., Wang, M., Shi, X., 2018b. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis-NIR spectroscopy. Geoderma 310, 29–43.
Yang, Y., Viscarra Rossel, R.A., Li, S., Bissett, A., Lee, J., Shi, Z., Behrens, T., Court, L., 2019. Soil bacterial abundance and diversity better explained and predicted with spectro-transfer functions. Soil Biol. Biochem. 129, 29–38.
Zhang, Y., Biswas, A., Ji, W., Adamchuk, V.I., 2017. Depth-specific prediction of soil properties in situ using vis-NIR spectroscopy. Soil Sci. Soc. Am. J. 81, 993–1004.
Zhao, L., Hong, H., Fang, G., Algeo, T.J., Wang, C., Li, M., Yin, K., 2020. Potential of VNIR spectroscopy for prediction of clay mineralogy and magnetic properties, and its paleoclimatic application to two contrasting Quaternary soil deposits. Catena 184, 104239.
Žížala, D., Zádorová, T., Kapička, J., 2017. Assessment of soil degradation by erosion based on analysis of soil properties using aerial hyperspectral images and ancillary data, Czech Republic. Remote Sens. 9, 28.
Zovko, M., Romić, D., Colombo, C., Di Iorio, E., Romić, M., Buttafuoco, G., Castrignanò, A., 2018. A geostatistical Vis-NIR spectroscopy index to assess the incipient soil salinization in the Neretva River valley, Croatia. Geoderma 332, 60–72.